Department of Molecular Medicine
Karolinska Institutet, Stockholm, Sweden

MOLECULAR GENETIC STUDIES OF COLORECTAL CANCER

XiaoLei Zhou

Stockholm 2005

Published and printed by Karolinska University Press
Box 200, SE-171 77 Stockholm, Sweden

ISBN 91-7140-489-9

To My wife and son

ABSTRACT

Colorectal cancer (CRC) is one of the leading causes of morbidity and mortality due to cancer in the Western countries. Epidemiological studies have shown that at least 20-30% of all CRCs have a potentially identifiable genetic cause. Inherited forms of CRC with Mendelian dominant inheritance have been a major focus of study and many high-penetrance susceptibility genes have been identified, such as *APC, hMLH1, hMSH2*. However these high-penetrance susceptibility genes can only account for 3%-5% of all CRCs, thus additional, so far unknown, genes for CRC predisposition remain to be identified. Recent evidence has shown that a large proportion of genetic variation in CRC risk is probably due to a combination of a large number of common alleles with low penetrance and shared environmental factors. So in order to plot the accurate genetic-risk profiles of CRC for future population-based diagnosis and prevention programs, it is as equally important to identify low penetrance alleles as to identify high penetrance alleles. In this thesis, we evaluated the contribution of *hMLH3*, *APC* and *MYH* as either high- or low-penetrance susceptibility genes for CRC in the Swedish population by the approaches of mutation screening and association study.

hMLH3, is a newly identified DNA Mismatch Repair (MMR) gene interacting with *hMLH1* and suggested to be a candidate gene for HNPCC. We screened 70 index patients suggestive of a genetic predisposition for germline mutations in the gene, and found one frameshift and 11 missense variants in 16 index patients (23%). The only frameshift mutation found, 885delG, seemed to segregate very well with cancer in the family as a monogenic high-penetrance allele. All the missense variants identified were not shown to segregate with cancer in the families, so *hMLH3* mutations were not important as a major predisposing factor for familial CRC. However, in one family, we observed a co-segregation of one *hMLH3* missense variant and one *hMSH2* missense variant with the disease phenotype, suggesting that *hMLH3* could work as a low-penetrance predisposing gene for CRC by working together with other low-penetrance genes in an additive or multiplicative manner. (**Paper I**)

The recent finding that one missense variant in *APC* (I1307K) led to an increased CRC risk in Askenazi Jews raises the possibility that there should exist other similar predisposing variants in this gene in different populations. We tested this possibility by performing the mutation screening of *APC* in cohorts of Swedish hereditary CRCs, sporadic CRCs and normal controls. Our results showed that some *APC* variants other than I1307K could act as low-penetrance alleles for CRC, for example, the variant 3' UTR 8636C>A. This variant was found to have an increased CRC risk of around two fold-a risk comparable to that of I1307K-though the association studies had statistically significant results only at the borderline (**Paper II**). In addition, the 3' UTR 8636C>A variant and the *APC* locus it implicates were incidentally found to be associated with autism spectrum disorder (ASD), substantiated by several follow-up association studies with statistically significant results (**Paper IV**).

MYH plays a significant role in the repair of mutations caused by 8-oxo-7,8-dihydroxy-2-deoxyguanosine (8-oxoG), one of the most stable product of oxidative DNA damage. Biallelic germline mutations of this gene have been shown to predispose to a proportion of multiple colorectal adenomas and cancer. We screened 84 unrelated Swedish non-FAP and non-HNPCC familial CRCs for germline mutations in the gene to evaluate the contribution of *MYH* mutations in CRC with few polyps. None of the cases was found to carry any pathogenic sequence change, indicating that *MYH* mutations are not likely to account for familial CRCs in the

absence of multiple adenomas. However, we could demonstrate that the two most common pathogenic variants Y165C and G382D found in Caucasians existed in the Swedish population as well, and both of them appeared to confer a slightly increased CRC risk in heterozygosity. Additional candidate low-penetrance alleles for CRC were also identified at the position of amino acid 423 (R423Q, R423P, and R423R). Whether mutations at position 423 have pathological relevance needs to be further studied. (**Paper III**)

Recently, there is increasing evidence to indicate that hereditary and sporadic MSI cancers evolve through different pathways. We wondered if microarray technology can help differentiate these two tumors into two distinct entities. We transcriptionally profiled six HNPCC and seven sporadic MSI CRCs using Affymetrix® Microarray HG-U95Av2 chips. Unsupervised hierarchical clustering analysis and principle component analysis (PCA) failed to distinguish hereditary MSI tumors from its sporadic counterparts, indicating that the molecular difference between these two phenotypes, if any, is insignificant. In addition, a gene called Tight Junction Protein 3 (*TJP3*), a novel member of the MAGUK protein family found at the cell tight junction, was shown to be downregulated by about 2 fold in transcription in a small subset of HNPCC tumors with suggested bad prognosis compared with other samples, and in a MSI cell line with high metastatic potential compared with its isogenic cell line with low metastatic potential. Therefore, *TJP3* would be a metastasis-associated gene in CRCs with MSI, which warrants further functional characterization. (**Paper V**)

Key words: colorectal cancer (CRC), high-penetrance alleles, low-penetrance alleles, autism spectrum disorder (ASD), MSI, microarray, metastasis, *hMLH3*, *APC*, *MYH*, *TJP3*.

PAPERS INCLUDED IN THE THESIS

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

I. Hong-Xu Liu,* **Xiao-lei Zhou***, Tao Liu, Barbro Werelius, Gudrun Lindmark, Niklas Dahl, and Annika Lindblom (2003). The role of hMLH3 in familial colorectal cancer. Cancer Res **63**(8): 1894-9. (* authors contributed equally)

II. **Xiao-Lei Zhou**, Ulrika Eriksson, Barbro Werelius, Ulf Kressner, Xiao-Feng Sun, and Annika Lindblom (2004). Definition of candidate low risk APC alleles in a Swedish population. Int J Cancer **110**(4): 550-7.

III. **X.-L. Zhou,** T. Djureinovic, B. Werelius, G. Lindmark, X.-F. Sun and A. Lindblom (2005). Germline mutations in the MYH gene in Swedish familial and sporadic colorectal cancer. Genet Test **9**(2): 147-51.

IV. **Xiao-Lei Zhou**, MaiBritt Giacobini, Britt-Marie Anderlid, Henrik Anckarsäter, Davood Omrani, Christopher Gillberg, Magnus Nordenskjöld, and Annika Lindblom (2005). A single nucleotide polymorphism in the adenomatous polyposis coli gene is associated with autism spectrum disorder (ASD). *submitted*

V. **Zhou X-L**, Christopher W, Jensen T, Fischer H, Werelius B, Lindmark G, Påhlman L, Lindforss U, Brunak S, Lindblom A (2005). Expression profiles of hereditary and sporadic colorectal cancers and cell lines with microsatellite instability (MSI). *Manuscript*

OTHER PAPERS ON RELATED TOPICS

X.-L. Zhou, B. Werelius and A. Lindblom (2004). A screen for germline mutations in the gene encoding CCCTC-binding factor (CTCF) in familial non-BRCA1/BRCA2 breast cancer. Breast Cancer Res **6**(3): R187-90.

A. Lindblom, **X. L. Zhou**, T. Liu, A. Liljegren, J. Skoglund and T. Djureinovic (2004). Colorectal cancer as a complex disease: defining at-risk subjects in the general population - a preventive strategy." Expert Rev Anticancer Ther **4**(3): 377-85.

Johanna Skoglund, Tatjana Djureinovic, **Xiaolei Zhou**, Jana Vandrovcova, Elise Renkonen, Lennart Iselius, Marie Luise Bisgaard, Päivi Peltomäki, Annika Lindblom (2005). 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. Journal of Medical Genetics. *In press.*

Tatjana Djureinovic*, Johanna Skoglund*, Jana Vandrovcova, **Xiao-Lei Zhou**, Antonia Kalushkova, Lennart Iselius and Annika Lindblom (2005). A genome-wide linkage analysis in Swedish families with hereditary non-FAP/non-HNPCC colorectal cancer. *Gut. In press.* (* authors contributed equally)

LIST OF ABBREVIATIONS

3’ UTR	3’ untranslating region
APC	adenomatous polyposis coli
BER	base excision repair
Bp	base pair
CIMP	CpG island methylator phenotype
CIN	chromosome instability
CRC	colorectal cancer
DCC	deleted in colorectal carcinoma
DNA	deoxyribonucleic acid
DPC4	deleted in pancreatic cancer locus 4
E2F4	E2F transcription factor 4
FAP	familial adenomatous polyposis coli
hMLH1	human MutL homolog 1 gene
hMLH3	human MutL homolog 3 gene
hMSH2	human MutS homolog 2 gene
hMSH3	human MutS homolog 3 gene
hMSH6	human MutS homolog 6 gene
HNPCC	hereditary non-polyposis colorectal cancer
hPMS1	human post meiotic segregation homolog 1 gene
hPMS2	human post meiotic segregation homolog 2 gene
IGFIIR	insulin-like growth factor type II receptor gene
Kb	kilo bases
KRAS	Kirsten rat sarcoma, viral oncogene homologue
LOH	loss of heterozygosity
MGMT	*O*-6-methylguanine DNA methyltransferase
MMR	microsatellite repair
MSI (MIN)	microsatellite instability
MSI-H	microsatellite instability-high
MSI-L	microsatellite instability-low
MSS	microsatellite stable
P16	cyclin-dependent kinase inhibitor 2A
P53	tumor protein 53 gene

PCR	polymerase chain reaction
PIK3CA	phosphoinositide-3-kinase, catalytic, alpha polypeptide
PTEN	phosphatase and tensin homolog
RASSF1	Ras association (RalGDS/AF-6) domain family 1
RASSF2	Ras association (RalGDS/AF-6) domain family 2
Rb1	retinoblastoma 1
RNA	ribonucleic acid
SMAD4	mothers against decapentaplegic homolog 4
STK11(LKB1)	serine/threonine kinase 11
TCF4	transcription factor 4
TGFßR2	transforming growth factor beta receptor 2

CONTENTS

1 INTRODUCTION

1.1 CANCER GENETICS

Cancer is, in essence, a genetic disease. This view could be traced back 100 years ago when Boveri first suggested aneuploidy due to chromosome missegregation as the fundamental basis of cancer (Boveri 1914). Since 1960, first cytogenetic and then molecular techniques have shown that tumors expand as a clone from a single altered cell, and that clinical 'progression' is the result of sequential somatic genetic or epigenetic changes followed by selection, generating increasingly aggressive subpopulations within the expanding clone (Nowell 1976; Nowell 2002). This prevailing model of tumorigenesis as somatic evolution has been confirmed by numerous molecular studies and it has been estimated that cells have to acquire five to seven successive mutations to allow tumor growth, invasion and metastasis (Fearon et al. 1990; Boland et al. 1999; Luebeck et al. 2002). Alternative theories of tumorigenesis also exist. For example, in a so-called "the tissue field theory", it is proposed that carcinogenesis and neoplasia occur entirely through emergent (supracellular) phenomena after disruption of the normal interactions that take place among cells in the parenchyma and stroma of an organ (Sonnenschein et al. 2000). Whatever the theory is, the pivotal role that cancer susceptibility genes have in tumor initiation and progression in either tumor cells themselves or the surrounding stromal cells is undisputable, which is best illustrated by the identification of sequential mutations of *APC*, *KRAS*, *SMAD4* and *p53* in a defined series of stages of from normal mucosa to carcinoma in colorectal tumorigenesis (Fearon et al. 1990; Vogelstein et al. 1993).

Multiple types of cancer susceptibility genes have been identified. Although they differ in different tumors, the loss or abnormal functions of them will render most cancers to acquire the same set of capabilities during their development. These capabilities are: (1) self-sufficiency in growth signals; (2) evading apoptosis; (3) insensitivity to antigrowth signals; (4) tissue invasion and metastasis; (5) sustained angiogenesis; and (6) limitless replicative potential (Hanahan et al. 2000). Cancer susceptibility genes can be generally categorized into three classes-gatekeepers, caretakers and landscapers.

1.1.1 Gatekeepers

Gatekeepers comprise oncogenes and tumor suppressor genes (Kinzler et al. 1997; Michor et al. 2004). The mutations of both oncogenes and tumor suppressor genes operate similarly at the physiological level: they drive the neoplastic process by increasing tumor cell number through the stimulation of cell birth or the inhibition of cell death or cell cycle arrest (Vogelstein et al. 2004).

1.1.1.1 Oncogenes

In 1969, Huebner and Todaro proposed the "oncogene" hypothesis of cancer based on the discovery that some endogenous viruses contained transforming elements, and the activation of these endogenous transforming elements could cause cancer (Huebner et al. 1969). The viral oncogene is designated v-*onc*, and the proto-oncogene is the normal cellular gene from which the oncogene, c-*onc*, is derived. Ever since the identification of the first proto-oncogene *c-src* more than a quarter of a century ago, approximately 200 proto-oncogenes have now been identified. These oncogenes can be classified into five broad classes based on their functional and biochemical properties: (1) secreted growth factors (e.g. *SIS*), (2) cell surface receptors (e.g. *ERBB*, *FMS*), (3) components of intracellular signal transduction systems (e.g. the RAS family, *ABL*), (4) DNA-binding nuclear proteins, including transcription factors (e.g. *MYC*, *JUN*), (5) components of the network of cyclins, cyclin-dependent kinases and kinases inhibitors that govern progress through the cell cycle (e.g. *MDM2*). In contrast to the original studies in animals, oncogenic activation through retroviral transmission does not appear to be a major factor in human tumorigenesis. However, oncogenes are frequently activated by gain of function mutations or fusions with other genes, or they are aberrantly expressed due to amplification, increased promoter activity, or protein stabilization (Munger 2002), and hence they play important roles in diverse signaling pathways that are involved in various stages of human cancers-tumor initiation, progression, angiogenesis and metastasis (Michor et al. 2004). An activating somatic mutation in one allele is generally enough to confer a selective growth advantage on the cell. A large number of cellular oncogenes have been identified to play a role in colorectal cancer, such as *KRAS*, *MYC*, *SRC*, *β-catenin*, *BRAF*, etc. Recently, systematic mutational analyses showed that a minimum of 30% of CRCs contain at least one mutation in the tyrosine kinase family (kinome) (Bardelli et al. 2003), and 32% of CRCs contain a mutation in the *PIK3CA* gene (Samuels et al. 2004).

1.1.1.2 Tumor suppressor genes

As the name indicates, tumor suppressor genes function to suppress neoplastic growth of cells. Mutations in these genes impair this growth-suppressor mechanism by removing the brakes from the cell and result in uncontrolled growth. Examples are the genes *Rb1* (the first isolated human tumor suppressor gene), *p53* and *p16*. We now recognize that tumor suppressor genes regulate diverse cell activities, including cell cycle checkpoint responses, detection and repair of DNA damage, protein ubiquitination and degradation, mitogenic signaling, cell specification, differentiation and migration, and tumor angiogenesis (Sherr 2004). It has been suggested that the p53 and Rb pathways are needed to be disrupted for all cancers to develop (Vogelstein et al. 2004). The inactivation of tumor suppressor genes in human tumors is usually achieved by having two "hits" or mutations targeting the same gene-Knudson's classical "two hit" model (Knudson 1971). This situation commonly arises through the deletion of one allele via a gross chromosomal event, such as loss of heterozygosity (LOH), coupled with an intragenic mutation of the other allele. Mutations (both point mutations and

deletions) are not the only way through which a tumor suppressor gene can be inactivated. In recent years, it has become obvious that epigenetic changes, and probably haploinsufficiency, can also serve to switch off tumor suppressor genes. For example, the loss for *p16* occurs through the independent mechanisms of deletions, point mutations, or promoter hypermethylation (Kamb et al. 1994; Herman et al. 1995). Functional haploinsufficiency refers to a process in which loss of one allele contributes to tumor progression even though a wildtype allele remains. This has been shown in mice with tumors carrying single, inactivated alleles of the genes encoding *p53, TGF-β1*, *27kip1*, and *DMP1* (Balmain et al. 2003). The evidence in the human may come from the study on chromosome 3p21.3 in lung cancer showing that several genes with tumor-suppressor activities are involved in the LOH event and that the remaining alleles are rarely mutated (Ji et al. 2002). Most of the identified susceptibility genes responsible for familial cancer syndromes by now are tumor suppressor genes. In addition, in the sporadic counterparts of these familial cancer syndromes, the inactivating somatic mutations of the same set of tumor suppressor genes are also frequently found (Gailani et al. 1996; Ponder 2001; Vogelstein et al. 2004). For example, *APC* mutations in the germline predispose to FAP, but are detected somatically in more than 80% of sporadic CRCs as well. There is a long list of tumor suppressor genes besides *APC* that are found involved in pathogenesis of CRC, including *p53, AXIN2, STK11(LKB1), PTEN, DCC, NM32,* etc.

1.1.2 Caretakers

Caretakers, or stability genes, function in maintaining the genomic integrity of the cell and regulate DNA repair mechanisms, chromosome segregation, and cell cycle checkpoints (Kinzler et al. 1997). Caretaker defects lead to genetic instabilities that contribute to the accumulation of mutations in other genes, such as oncogenes and tumor suppressor genes that directly affect cell proliferation and survival, thus promoting tumorigenesis indirectly (Lengauer et al. 1998; Rajagopalan et al. 2003; Iwasa et al. 2005). Because the problem of DNA damage has existed *ab initio*, DNA repair systems must have arisen early, and are highly conserved, in evolution (Hoeijmakers 2001). At least four main, partly overlapping DNA damage repair pathways operate in mammals: the nucleotide-excision repair (NER), base-excision repair (BER) and mismatch repair (MMR) systems are responsible for repairing subtle mistakes made during normal DNA replication or induced by exposure to mutagens (Hoeijmakers 2001; Vogelstein et al. 2004), while the homologous recombination and non-homologous end joining system takes on the task of repairing DNA damage involving large proportions of chromosomes, such as double-strand DNA breaks (DSBs) (Hoeijmakers 2001; Vogelstein et al. 2004). Each DNA repair system is a complex biochemical process that requires the participation of many genes, and still remains poorly understood.

The role of defective DNA repair systems in driving tumorigenesis is shown by the existence of human cancer syndromes with germline mutations in caretaker genes.

Inborn defects in NER cause syndromes of xeroderma pigmentosum, cockayne syndrome and trichothiodystrophy (TTD), all characterized by extreme sun sensitivity. Germline mutations in MMR genes, such as *hMLH1* and *hMSH2*, are well known to predispose to HNPCC. Recently, germline mutations in the *MYH* genes are found to account for a considerable proportion of multiple colorectal adenomas and cancer with autosomal recessive inheritance-the first human disorder identified to be caused by inherited BER deficiencies. There are also several established examples in which germline mutations of caretaker genes involved in the defective homologous recombination and non-homologous end joining system predispose for inherited syndromes, such as *ATM* for ataxia telangiectasia, *BRCA1* and *BRCA2* for breast cancer, and *NBS1* for Nijmegen breakage syndrome (NBS) (Hoeijmakers 2001). As with tumor suppressor genes, both alleles of caretaker genes generally need be inactivated for a pathological effect to result.

1.1.3 Landscaper

Landscaper defects do not directly affect cellular growth, but generate an abnormal stromal environment that contributes to the neoplastic transformation of cells (Michor et al. 2004; Iwasa et al. 2005). It is now well accepted that cancer is a communal affair, with malignant transformation of an epithelium occurring within the context of a dynamically evolving tissue stroma that is composed of multiple cell types surrounded by an extracellular matrix (Weaver et al. 2004). There are many examples to show that cancers can arise as a consequence of inherited or acquired genetic mutations in stromal cells. Barcellos-Hoff and colleagues demonstrated that irradiation of the mammary-gland stromal component contributed to neoplastic progression of non-irradiated epithelial cells in vivo (Barcellos-Hoff et al. 2000). Moinfar and colleagues reported that distinct genetic alterations and chromosomal rearrangements were absent in the malignant carcinoma cells, but present in neighboring stromal cells instead (Moinfar et al. 2000). Moreover, studies of the familial juvenile polyposis syndrome also indicate that it is actually the genetic alteration of the *SMAD4/DPC* gene in stromal cells that predispose colonic epithelial cells to carcinoma (Jacoby et al. 1997; Howe et al. 1998). The dysfunction or deregulation of landscaper genes can disrupt normal tissue homeostasis, reduce host immune surveillance and defense, induce angiogenesis and inflammation, and promote tumor growth and migration (Mueller et al. 2004). Some example landscaper genes are metalloproteinases (*MMPs*), uroplasminogen activator (*uPA*), tissue plasminogen activator (*tPA*), fibroblast growth factor-2 (*FGF-2*), and platelet-derived growth factor (*PDGF*).

1.2 GENERAL BACKGROUND OF CRC

Colorectal cancer (CRC) is the third most common cause of cancer-related death in the western world. Based on familial clustering studies, it is estimated that 20-30% of CRCs have a potentially identifiable genetic cause (Johns et al. 2001). Inherited forms of CRC with Mendelian dominant inheritance have been a major focus of study and

many high-penetrance susceptibility genes have been identified, such as *APC, hMLH1, hMSH2, LKB/STK11, SMAD4*. These genetic discoveries have been translated into clinical practice and have led to improved risk assessment through the use of genetic testing (Grady 2003). However these high-risk cancer genes can only account for 3%-5% of all CRCs, leaving the majority of CRCs molecularly unexplained. More and more evidence is emerging for the concept that a large proportion of genetic variation in cancer risk is probably due to more common but less penetrant alleles-the model of common disease/common variant (Balmain et al. 2003; Lohmueller et al. 2003). The successful sequencing and annotation of the human genome has provided tools for large-scale association studies that should facilitate the identification of low-penetrance susceptibility alleles (de la Chapelle 2004). By using known low-risk profiles in a nation, it will be possible to select a subpopulation at increased risk of CRC regardless of family history. This scenario will provide an opportunity to design tailored preventive programs for at-risk subjects and this will be a better alternative compared with general population screening to reduce the incidence of colorectal cancer (Lindblom et al. 2004).

1.2.1 Incidence

Every year, WHO estimates that more than 945 000 people develop colorectal cancer worldwide, accounting for 8.5% of all new cases, and around 492 000 patients die (Potter 1999; Weitz et al. 2005). Incidence rates vary approximately 20-fold around the world, being higher in industrialized regions of North America, Northern and Western Europe, New Zealand and Australia, lower in Asia, Africa and South America (Pisani et al. 1999; Potter 1999). In Western countries, the cumulative lifetime risk of CRC and death from CRC is approximately 5-6% and 2.5%, respectively (Sondergaard et al. 1991; Fuchs et al. 1994; Terdiman et al. 1999). Colon cancer occurs with approximately equal frequency in men and women (McMichael et al. 1980; Potter 1999); however in high-incidence areas such as North America and Australia, as well as in Japan and Italy where rates are rising rapidly, rates in men now exceed those in women by as much as 20%. Rectal cancer is up to twice as common in men as in women (Potter 1999). In Sweden, though the epidemiological data of cancer incidence vary among years, CRC always ranks second in incidence in both sexes, after breast cancer and prostate cancer, respectively (National Board of health and welfare, 1992, 1999, 2000, 2002, 2003). In 2003, 2765 male and 2612 female CRCs were newly diagnosed, representing 11.2% and 11.8% of all cancer diagnoses in each sex, respectively.

1.2.2 Etiology

1.2.2.1 Genetic factors and family history

CRC can be divided into familial (hereditary) and sporadic cases based on family history, with sporadic cases accounting for the majority. Epidemiological studies have long indicated that the predisposition for CRC in the population may account for a substantial fraction of CRCs. First-degree relatives (parents, siblings and children) of

patients with sporadic CRC have been shown to have a two to three-fold increased risk for both colorectal adenomas and CRC compared with the general population (Lovett 1976; Cannon-Albright et al. 1988; Dong et al. 2001; Johns et al. 2001). According to a recent twin study, 35% of variation in CRC was assigned to heritable factors (Lichtenstein et al. 2000). In Sweden, it has been observed that 12.5% of 0-66 year old patients with CRC have an affected parent, giving a population attributable proportion of 6.9%, equal to that in breast cancer (Hemminki et al. 2002).

Except in hereditary cancer syndromes, it is usually impossible to determine in a given case whether familial occurrence is due to shared predisposition or chance. Similarly, in a given "sporadic" case the apparent absence of a family history of cancer might not reflect the true situation (de la Chapelle 2004). There is increasing evidence to suggest that the occurrence of some "sporadic" CRCs result largely from the inheritance of many low-penetrance susceptibility alleles, and in rare cases even from recessive high risk alleles, such as *MYH*. Recently an epidemiological study based on the nationwide Swedish Family-Cancer database supported the existence of high risk recessive inheritance in 0.75% of all CRCs (Hemminki et al. 2005). Therefore, the distinction between sporadic and familial cases of CRC has become blurred.

1.2.2.1.1 High-risk familial CRC

Several high-risk inherited forms of colorectal cancer syndromes are known and the predisposing genes have been characterized for the majority of the cases. These syndromes include familial adenomatous polyposis (FAP) and hereditary non-polyposis colorectal cancer (HNPCC) (**Table 3**)

There are still a considerably large proportion of high-risk families whose molecular component remains to be identified. These families are collectively called hereditary colorectal cancer (HCRC) that consists of three or more first-degree relatives with CRC in at least two generations, and is likely to segregate monogenic, as yet unknown high-risk predisposing genes. Affected individuals from HCRC families have much fewer adenomas than FAP cases, and are not compatible with the diagnostic Amsterdam Criteria for HNPCC in family history either. They are characterized by MSI-negative tumors, later age of tumor onset, and equal distributions of the tumors in the bowel, sometimes with rectal tumors predominating (Lindgren et al. 2002). The risk individuals have a similar life-time cancer risk as those in HNPCC (Fuchs et al. 1994; Johns et al. 2001). A genome-wide linkage analysis has been performed in a subset of Swedish HCRC families by our group without finding linkage to any common locus, suggesting heterogeneity among these families (Djureinovic et al. 2005).

1.2.2.1.2 Low-risk familial CRC

Low-risk CRC families can be classified into two categories, two close relative (TCR) and one close relative (OCR). TCR refers to families having two affected first-degree relatives, and OCR refers to families having only one affected first-degree relative with

an early age of tumor onset. Similar to HCRC, TCR tumors are in general MSI-negative and have a late onset age, predominately in the distal colon (Lindgren et al. 2002; Liljegren et al. 2003). In addition, there is a tendency for risk individuals from TCR families to have higher prevalence of adenomas and hyperplastic polyps than the general population and HNPCC subjects (Liljegren et al. 2003). Epidemiological studies clearly demonstrated that risk individuals in TCR and OCR families have a lifetime risk of CRC of 10-20% and 20-40%, respectively (Lovett 1976; Houlston et al. 1990; St John et al. 1993). Kindred studies have suggested that the familial risk seen in TCR and OCR families probably arises from mildly to moderately penetrant inherited susceptibility factors (Cannon-Albright et al. 1988). The frequencies of different types of CRC in Sweden are presented in **Figure 1**.

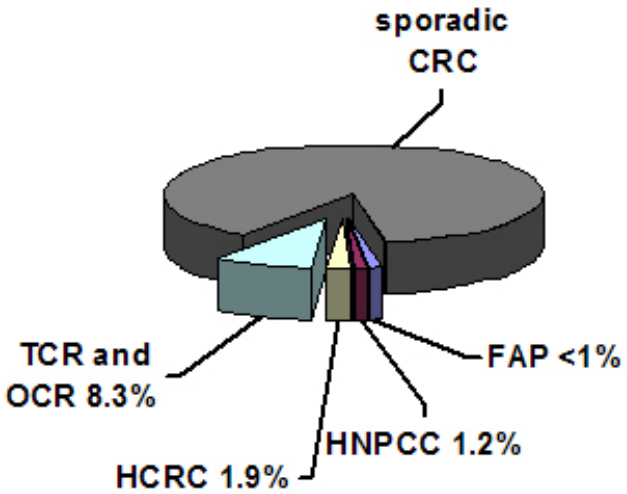

Figure 1. The frequencies of different types of CRC with respect to genetic background in Sweden (Olsson et al. 2003).

1.2.2.2 Non-genetic risk factors

Epidemiological studies have recognized the contribution of non-genetic factors in the etiology of CRC for many years. Compelling evidence is given by the observation that, when Japanese have moved from their home country of low incidence to the USA of high incidence, their CRC risk will increase to the same as that of their adopted country in 20-30 years. It should be pointed out the association of some dietary components (for example, fiber, meat and fat) with CRC risk is still controversial, which requires large meta-analyses and prospective cohort studies in the future. Some of defined risk factors for CRC development are summarized in **Table 1**.

1.3 MOLECULAR ASPECTS OF CRC

1.3.1 Genomic instability

Most human malignancies are recognized cytogenetically by marked aneuploidy and complex chromosomal rearrangements that include non-reciprocal translocations, DNA

Table 1. Non-genetic factors (including environmental factors)

Older age
Male sex
Cholecystectomy
Ureterocolic anastomosis
Ulcerative colitis
Crohn's colitis
Hormonal factors: nulliparity, late age at first pregnancy, early menopause
Diet rich in meat and fat, and poor in fibre, folate, and calcium
Sendentary lifestyle
Obesity
Diabetes mellitus
Smoking
Previous irradiation
Occupational hazards (eg, asbestos exposure)
High alcohol intake

Modified from (Weitz et al. 2005)

fragmentation and chromosome fusions (Feldser et al. 2003). This chromosome instability (CIN) can be found in approximately 85% of CRCs. The remaining 15% of CRCs show a different genomic instability that is characterized by mismatch repair (MMR) deficiency. As this instability was first found in short repetitive DNA sequences known as microsatellites, it is named microsatellite instability (MSI, also known as MIN) (Lindblom 2001). In addition, the third type of genomic instability, base excision repair (BER) instability, was identified recently to operate in a small proportion of CRCs (Al-Tassan et al. 2002).

At present there is still much controversy over whether or not genomic instability is necessary for driving tumor growth, and whether or not it is the usual initiating event in tumorigenesis. Some authors have argued that, although genomic instability can aid tumorigenesis, Darwinian natural selection is more likely to be the initiating and main driving force of both tumor initiation and progression considering increased cell turn-over rates of tumor cells (Sieber et al. 2003). However, the concept that genomic instability can greatly promote tumorigenesis, especially in cases with predisposed DNA repair deficiency, is not an issue for debate. The finding that CIN and MIN are mutually exclusive in CRC tissue samples and cell lines (Eshleman et al. 1995; Rowan et al. 2000; Abdel-Rahman et al. 2001) strengthens the idea that CIN and MIN are equivalently important mechanisms for acquiring the genetic alterations needed for tumorigenesis.

1.3.1.1 Chromosome instability (CIN)

More than a century ago, one of the first things that researchers noticed when looking at the cancer cells under microscope was that tumor cells often had excess chromosomes (Rajagopalan et al. 2004). It is a well-known fact that the karyotypes of solid tumors

that arise from the colon, breast and prostate are generally much more complex than those seen in haematological malignancies. For example, chronic myelogenous leukaemia (CML) has very few chromosomal aberrations, and often involving the nonrandom aberrant translocation between chromosomes 9 and 22 (the Philadelphia chromosome) (Nowell et al. 1960; Rowley 1975). This may reflect the possible case that solid tumors need more rounds of clonal expansions driven by genetic anomalies than haematological malignancies. The chromosomal aberrations found in CRC are so complex that it is too difficult to sum up all the chromosomal changes with a few representative karyotypes. CIN phenotype seems to be dominant, as it can be conferred on a chromosomal stable, diploid cell when it is fused with a CIN cell (Rajagopalan et al. 2003). Molecular analysis of extremely small polyps (2 mm, range 1-3 mm) has revealed a significantly high degree of allelic imbalance, indicating that CIN occurs very early in tumorigenesis (Shih et al. 2001). Conceivably, CIN could accelerate the rate of LOH for inactivating tumor suppressor genes and/or effectively amplify an oncogene by duplicating the chromosome on which it lies, thus contributing to tumorigenesis (Rajagopalan et al. 2004). Indeed, the most frequently observed LOH in CRC are in chromosomal regions of 5q, 17p and 18q on which tumor suppressor genes *APC*, *p53* and possibly *SMAD4* reside, respectively (Vogelstein et al. 1988), consistent with the role of CIN for the inactivation of tumor suppressor genes. In yeast, more than 100 different genes have been shown to cause CIN (Spencer et al. 1990), but the identification of key human genes responsible for CIN, either through abnormal expression or mutations, is only in its infancy (Rajagopalan et al. 2003). Genetic analyses have shown that mutations or deregulation in the genes involved in chromosome metabolism, spindle assembly and dynamics, cell-cycle regulation, and checkpoint control controlling cell cycle can result in human CIN (Jallepalli et al. 2001). Example genes are *BUB1*, *BUBR1*, *MAD2*, *Aurora A* (*STK15/BTAK*), *Securin*, *Survivin*, *Cyclin E*, *ZW10*, *p53*, *ATM*, *CHK2*, *BRCA*1 and *BRCA2* (Hollstein et al. 1991; Li et al. 1997; Zhou et al. 1998; Bell et al. 1999; Tutt et al. 1999; Boultwood 2001; Jallepalli et al. 2001; Weaver et al. 2002). However these genes, except *p53*, are very infrequently found to be associated with CIN in CRC. Recently a gene called *hCDC4* (also known as *FBW7* or *Archipelag*o) was linked with CIN in CRC and it was mutated at an early stage of tumorigenesis in a high percentage of CRCs (Rajagopalan et al. 2004). In addition, a large-scale mutation screening of human homologues of genes that cause CIN in *Saccharomyces cerevisiae* has implicated *MRE11, hZw10*, *hZwilch*/*FLJ10036*, *hRod*/*KNTC and ding* with CIN in CRC (Wang et al. 2004). Interestingly, the crucial gatekeeper gene *APC* in CRC tumorigenesis has been proposed as potential initiator of CIN as well (Fodde et al. 2001; Kaplan et al. 2001), though not universally supported in other studies (Eshleman et al. 1998; Abdel-Rahman et al. 2001). In addition, telomere dysfunction could also make an important contribution for initiating CIN in all cancers including CRC (Feldser et al. 2003).

1.3.1.2 Microsatellite instability (MSI)

The primary function of the post-replicative mismatch repair (MMR) is to eliminate base-base mismatches and insertion/deletions loops (IDLs) that arises as a consequent of DNA polymerase slippage during DNA synthesis (Peltomaki 2001; Peltomaki 2001). DNA mismatch repair is an evolutionarily conserved process and a prototype of this pathway has been the *Escherichia coli* system, in which early *in vitro* experiments established the necessary components. The principle players in mammalian MMR are the homologues of bacterial MutS and MutL proteins, which function in the form of heterodimers. Single base-base and IDL mismatches are recognized by MSH2-MSH6 (MutSα) heterodimers. In the absence of MSH6, the recognition of IDLs can be mediated by MSH2-MSH3 (MutSβ) heterodimers, indicating redundant functions in MSH3 and MSH6 (Marsischky et al. 1996; Das Gupta et al. 2000). Additional MutS homologs in the human are known that are required for functions other than MMR. These proteins include MSH4 and MSH5 that are necessary for meiotic (and possibly mitotic) recombination (Santucci-Darmanin et al. 2000; Svetlanov et al. 2004). On the other hand, MLH1 forms three different MutL heterodimers with PMS2, PMS1 and MLH3, named MutLα, MutLβ, and MutLγ, respectively. MutLα plays an essential role in coordinating the interplay between the mismatch recognition complex and other proteins necessary for MMR. The specific function of MutLβ and MutLγ in MMR are not confirmed, but MutLγ is predicted to address a subset of IDLs, and might have redundant functions with PMS2 in the MutLα complex (Jiricny 2000). MutLγ is also involved in meiotic recombination by interacting with MSH4 (Santucci-Darmanin et al. 2002).

At present, how MMR is executed is still poorly understood. Recently, a model called molecular switch was proposed based on studies on *Escherichia coli*. In this model, MutS binds to mismatched nucleotides as a mismatch sensor in an ADP bound state and provokes ADP→ATP exchange, resulting in formation of a stable hydrolysis-independent sliding clamp that is capable of diffusion for at least 1 kb along the DNA adjacent to the mismatch. This process is suggested to occur iteratively to load multiple ATP-bound MutS clamps that can interact with MutL, followed by interacting with and activating MutH endonuclease. Additionally, MutL promotes ATP binding-independent turnover of idle MutS sliding clamps (Acharya et al. 2003). Eukaryotes have no known homologs of *E. coli* MutH, and it is believed that in eukaryotic MMR the process of strand excision, resynthesis and ligation after activation of complexes of MutS and MutL will include at least the proliferating cell nuclear antigen (PCNA), exonucleases (such as, EXO1), DNA polymerases (δ and ε), single-stranded DNA binding protein and possibly helicases (Peltomaki 2003).

The hallmark of MSI is widespread mutation of insertions and deletions in repetitive DNA sequences known as microsatellites. Microsatellites are scattered throughout the human genome and comprise tandemly repeated DNA sequences of 1-6 bases, most commonly as $(CA)_n$, so are very prone to strand slippage and replication errors in

MMR-deficient cells (Hoang et al. 1997). MSI was initially described in association with HNPCC (Aaltonen et al. 1993), and soon found in 12-15% of sporadic CRCs (Ionov et al. 1993; Thibodeau et al. 1993; Kim et al. 1994). In addition, MSI is also detected in extra colonic cancers, such as gastric, breast, endometrial and upper urinary tract carcinomas (Han et al. 1993; Peltomaki et al. 1993; Risinger et al. 1993; Risinger et al. 1996). It is estimated that tumor cells with MSI typically have mutation rates at the nucleotide level 100-1,000-fold greater than normal cells (Bhattacharyya et al. 1994; Eshleman et al. 1995). This "mutator phenotype" conferred by MSI is believed to contribute to the initiation and promotion of multistage carcinogenesis (Yamamoto et al. 2002), but the simple inactivation of an MMR gene is not thought to be by itself a transforming event. There are many targets of MSI, especially those containing repetitive sequences in the coding region. In 1995, *TGFβRII* was reported as the first target gene for instability in MSI tumors (Markowitz et al. 1995). Within a short time, many other important target genes have been identified, including such as the apoptosis genes *BAX* and *caspase-5* (Rampino et al. 1997; Schwartz et al. 1999), the cell cycle regulator genes *E2F4*, *TCF4* and *PTEN* (Souza et al. 1997; Tashiro et al. 1997; Duval et al. 1999), the DNA repair genes *hMSH3* and *hMSH6* (Yamamoto et al. 1997), and the growth factor and receptor *IGFIIR* (Souza et al. 1996; Thorstensen et al. 2001). At present, at least 30 genes have been claimed to be MSI targeted genes and the gene number is still increasing with time; however only a few of them have been validated in replication (Duval et al. 2002). In addition, MSI has been suggested to have other functions in tumorigenesis by reducing apoptosis, transcription-coupled repair efficiency and chromosomal stability (de Wind et al. 1995; Hawn et al. 1995; Mellon et al. 1996; Leadon et al. 1998; Duckett et al. 1999; Hickman et al. 1999).

In 1997, a National Cancer Institute (NCI) Workshop meeting held in Bethesda proposed a panel of five markers for the uniform detection of MSI tumors. This panel of markers includes two mononucleotides, BAT25 and BAT26, and three dinucleotide repeats, D2S123, D5S346 and D17S250 (Boland et al. 1998). Tumors with instability at two or more of these markers were defined as being MSI-high (MSI-H), whereas those with instability at one repeat or showing no instability were defined as MSI-low (MSI-L) and microsatellite stable (MSS) tumors, respectively (Boland et al. 1998; Whitehall et al. 2001). Dinucleotide markers are sensitive to MSI-L status, whereas mononucleotide are relatively specific for MSI-H cancers (Dietmaier et al. 1997; Jass et al. 2000). It is not settled yet whether or not MSI-L tumors should be considered separate from MSS tumors, and conclusions drawn from different studies are partly conflicting (Peltomaki 2003). The commonly used diagnostic MSI test based on polymerase chain reaction (PCR) sometimes produces unexpected and overlooked result, due to the reasons such as contamination of the tumor with normal cells and intralesional heterogeneity (Muller et al. 2004). Another recently introduced rapid and cheap technique to test MSI is immunohistochemistry (IHC) staining of the tumors (Lindor et al. 2002). Tumors that have lost the function of one of the MMR genes show negative staining for the protein product of that gene by IHC. In almost all evaluation

studies using cases with known MSI status and known pathogenic mutations, IHC always showed a concordant detection rate of more than 90% (Hendriks et al. 2003; Hendriks et al. 2004; Muller et al. 2004). However, it should be kept in mind that IHC cannot completely replace MSI analysis based on PCR as long as the role of other putative MMR genes in CRC has not been elucidated (Vasen et al. 2004). MSI-H in HNPCC is caused by germline mutations in DNA MMR genes, *hMSH2*, *hMLH1*, *hMSH6*, and *hPMS2* (Umar et al. 2004). MSI-H and MSI-L in sporadic CRCs are attributable to methylation and inactivation of the DNA repair genes *hMLH1* and *MGMT*, respectively (Kane et al. 1997; Cunningham et al. 1998; Thibodeau et al. 1998; Kuismanen et al. 1999; Whitehall et al. 2001).

1.3.1.3 Base-excision Repair (BER) instability

The BER pathway plays a significant role in the repair of mutations caused by reactive oxygen species (ROS). ROS are generated in living cells during normal cellular metabolism and by exogenous sources such as ionizing radiation and various chemical oxidants (Hashiguchi et al. 2002). These ROS can react with DNA to produce a variety of genotoxic lesions that have been implicated in many degenerative diseases such as aging, cancer, and neurodegenerative disorders. One of the most common lesions is 8-oxo-7,8-dihydro-2'deoxyguanosine (8-oxoG) (Shigenaga et al. 1990). It readily mispairs with adenine residues, which, if not repaired, creates permanent G:C→T:A transversions in the subsequent DNA replication process (Wood et al. 1992). In *Escherichia Coli*, three glycosylases, MutT, MutM and MutY, involved in the BER pathway are found to protect cells from the deleterious effects of 8-oxoG (Michaels et al. 1992). Their human homologues, OGG, MYH, and MTH, have been identified with almost the same functions as in Escherichia Coli (Sakumi et al. 1993; Slupska et al. 1996). MTH removes oxidized precursor from the nucleotide pool prior to incorporation into DNA, OGG excises 8-oxoG from 8-oxoG:C base pairs, and MYH excises adenine from 8-oxoG:A base pairs. Of these three human BER components, only the *MYH* gene has been demonstrated to have a pathogenic role in CRC development (Al-Tassan et al. 2002). CRCs with germline biallelic mutations in *MYH* contain an excess of G:C→T:A transversions in *APC* or *KRAS* in tumor cells tested, consistent with defective BER function (Al-Tassan et al. 2002; Lipton et al. 2003).

1.3.2 Epigenetics

Epigenetic changes are modifications of the genome heritable during cell division that do not involve a change in DNA sequence (Feinberg 2004). Since its discovery in 1983, the epigenetics of human cancer has been in the shadows of human cancer genetics. But this area has become increasingly visible with a growing understanding of specific epigenetic mechanisms and their role in cancer, including hypomethylation, hypermethylation, loss of imprinting and chromatin modification (Feinberg et al. 2004). Now the genomic screening with a microarray-based strategy that combines gene expression status and epigenetic regulation has identified a large number of genes that are preferentially hypermethylated in CRC (Suzuki et al. 2002; Mori et al. 2004),

substantiating the idea that epigenetic events play a key role in colorectal tumorigenesis.

Global genomic hypomethylation (loss of methylation) at CpG dinucleotides is the first epigenetic abnormality identified in cancer cells and involves every tumor type studied, both benign and malignant (Feinberg et al. 2004). It has been shown that hypomethylation occurs at an early stage in colorectal neoplasia (Goelz et al. 1985; Feinberg et al. 1988; Hernandez-Blazquez et al. 2000). Hypomethylation can lead to gene activation by demethylating normally methylated CpG islands where genes reside. The first example comes from the oncogene *H-ras* (Feinberg et al. 1983). Later, many genes have been shown to be overexpressed in cancer due to hypomethylation, including the metastasis-associated gene *S100A4* in CRC (Nakamura et al. 1998). In addition, hypomethylation might favor mitotic recombination leading to LOH, as well as promoting karyotypically detectable rearrangement (Esteller et al. 2002).

CpG islands are GC- and CpG-rich areas of 0.5 to several Kb in size, usually located in the vicinity of genes and often found near the promoter of widely expressed genes (Jones et al. 1999). It has long been recognized that CpG island hypermethylation plays a key role in silencing genes that are as critical for, and possibly as frequent in, tumorigenesis as mutations in coding regions. In sporadic CRC, it has been shown that the interruption of Ras signaling can be achieved by either genetic activation of the *KRAS* oncogene or epigenetic silencing of RAS effectors *RASSF1* and *RASSF2* (van Engeland et al. 2002; Akino et al. 2005; Hesson et al. 2005). Almost half of the tumor-suppressor genes that cause familial cancers through germline mutations can be inactivated through promoter hypermethylation (Baylin et al. 2000). In CRC, at least three predisposing genes for hereditary cases, *MLH1*, *APC* and *LKB1*, can undergo transcriptional inactivation by promoter hypermethylation in sporadic cases (Hiltunen et al. 1997; Kane et al. 1997; Esteller et al. 2000; Esteller et al. 2000). Through the comparison of methylation profiles between MSI and MSS CRCs, a subset of colorectal tumors were shown to have a so-called CpG island methylator phenotype (CIMP), characterized by having higher incidence of hypermethylation at multiple promoter regions tested (Toyota et al. 1999). At present, epigenetic silencing has been recognized as a third way of satisfying Knudson' hypothesis that two hits are necessary for silencing of tumor suppressor genes (Jones et al. 1999).

Loss of imprinting (LOI) is the loss of parental origin-specific marks in cancer, leading either to aberrant activation of the normally silent copy of a growth-promoting gene, or the silencing of the normally expressed allele of a growth-inhibiting gene (Feinberg 2004). Most of the attention on imprinting and cancer has been focused on the insulin-like growth factor-II gene (*IGF2*), an important cancer gene in its own right (Feinberg 2004). LOI of *IGF2* is seen in several common cancers, including cancer of the large bowel (Cui et al. 1998). LOI of *IGF2* can also be found in normal colonic mucosa of CRC patients, and even circulating peripheral blood lymphocytes in many cases (Cui et

al. 1998). A very recent experiment using APC-mutant mouse models with or without LOI of *IGF2* gave the strong proof that epigenetic alterations of *IGF2* can promote intestinal tumorigenesis, possibly by causing a shift towards an undifferentiated cell fate to increase the rate of tumor initiation (Sakatani et al. 2005).

The interesting question is that if there is some underlying predisposition to epigenetic silencing. It has been shown that CRC patients with CIMP were 14 times more likely to have a family history of cancer than patients lacking CIMP (Frazier et al. 2003). A recent study provided some evidence that the hypermethylation status of *hMLH1* could be maintained in the germline to produce epigenetic inheritance (Suter et al. 2004).

1.3.3 Metastasis

Most deaths in human cancer including CRC are due to metastatic disease that remains resistant to chemo-radiation-therapy (Lindblom et al. 1996). Metastasis consists of a series of sequential steps, all of which must be successful completed. These include shedding of cells from a primary tumor into the circulation, survival of the cells in the circulation, arrest in a new organ, extravasation into the surrounding tissue, initiation and maintenance of growth, and vascularization of the metastatic tumor (Chambers et al. 2002). Despite the obvious importance of metastasis, the process remains incompletely characterized at the molecular and biochemical levels (Steeg 2003). In part, the knowledge is limited because metastasis is a "hidden" process which occurs inside the body and so is inherently difficult to observe (Chambers et al. 2002). It has long been recognized that certain tumor types tend to metastasize to specific organs. For example, breast cancer metastasizes to bone, liver, brain and lungs, while CRC preferentially spreads to liver (Chambers et al. 2002). In 1889, an English surgeon Stephen Paget proposed his famous theory of "seed and soil" to explain the observed organ-specific metastasis (Paget 1989). He believed that metastasis depends on cross-talk between selected cancer cells (the "seeds") and specific organ microenvironments (the "soil"), and metastasis can only form when the seed and soil are compatible (Fidler 2003). This theory was challenged in 1929 by James Ewing, who proposed that metastatic dissemination occurs by purely mechanical factors that are a result of the anatomical structure of the vascular system (Ewing 1928; Fidler 2003). Current biological evidence shows that these two theories are not mutually exclusive and play complementary roles in metastatic process. Mechanical factors influence the initial fate of cancer cells after they have left a primary tumor. Blood-flow patterns from the primary tumour determine which organ the cells travel to first. There, the relative sizes of cancer cells and capillaries lead to the efficient arrest of most circulating cancer cells in the first capillary bed that they encounter. For example, cancer cells leaving a primary colon tumor would enter the hepatic-portal system and be taken first to the capillaries of the liver and arrest there. But the next phase of metastatic colonization in the new organ is a very inefficient process, with only a small proportion of cancer cells growing into macroscopic metastasis. The outcome of metastasis is dictated by molecular interactions of metastasizing cells (the "seeds") with the microenvironment

of the new organ (the "soil") (Chambers et al. 2002; Fidler 2003). The organ microenvironment can markedly change the gene-expression patterns of cancer cells, and therefore their behavior in proliferation, angiogenesis, cohesion, motility and invasion (Fidler 2002). The contribution of chemokine receptor to organ-specific metastasis provides a convincing example. It has been shown that tumor cells express patterns of chemokine receptors that match chemokines specifically expressed in organs to which these cancers commonly metastasize. Both breast cancer cell lines and primary breast tumors express the chemokine receptors CXCR4 and CCR7 at high levels, and the specific ligands for these two receptors, CXCL12 and CCL21, are found at elevated levels in lymph nodes, lung, liver, and bone marrows to which these cancers often metastasize (Chambers et al. 2002). In primary CRC, CXCR4 gene expression has been demonstrated to be significantly associated with recurrence, survival, and liver metaststasis; furthermore, liver metastasis has been found to have a much higher level of CXCR4 expression than its localized counterpart (Kim et al. 2005). Another line of evidence for the idea that growth regulation in new organs of metastasizing tumor cells is a crucial regulator of metastatic capability is from the identification of metastasis suppressor genes. Metastasis suppressor genes are identified by their reduced expression in highly metastatic, compared with tumorigenic but poorly or non-metastatic, tumor cells (Steeg 2003). Eight metastasis suppressor genes have been identified so far, including *NM23*, *MKK4*, *KAI1*, *BRMS1*, *KiSS*1, *RHOGDI2*, *CRSP3* and *VDUP1* (Steeg 2003). At least *NM23* and *RHOGDI2* have been shown to play a role in CRC metastasis (Tagashira et al. 1998; Gildea et al. 2002; Suzuki et al. 2004).

Metastatic cells are believed to be rare in the primary tumor mass that happen to acquire metastatic capability during late stages of tumor progression in the traditional progression model (Poste et al. 1980; Fidler 2003). However this model was challenged by recent expression profiling studies on human tumors (Pantel et al. 2004). Expression profiling studies on breast cancer reported that in both lymph-node positive and negative breast tumors, there exists a gene expression signature strongly predictive of a short interval to distant metastases (van de Vijver et al. 2002; van 't Veer et al. 2002). Through comparing expression profiles of human primary tumors versus metastases, a common expression signature of only 17 genes was found to be associated with metastasis in different adenocarcinomas of diverse origin (lung, breast, prostate, colorectal, uterus, ovary) (Ramaswamy et al. 2003). An expression profiling study specific to CRC also identified a set of 194 known genes and 41 ESTs that discriminated well between samples with or without metastasis at diagnosis or during follow-up (Bertucci et al. 2004). These studies suggested that a gene-expression program of metastasis may already be present in the bulk of some primary tumors at the time of diagnosis and could therefore be used to predict which tumors would become metastatic (Ramaswamy et al. 2003). Weinberg et al. extrapolated that a subset of the mutant alleles acquired by incipient tumor cells early in tumorigenesis confer not only the selected replicative advantage, but also later in tumorigenesis, the proclivity to metastasize (Bernards et al. 2002). Genes and genetic changes specifically and

exclusively involved in orchestrating the process of metastasis do not exist. Instead, the genes for metastasis are largely those that cancer biologists have been studying intensively for a generation: the oncogenes and tumor suppressor genes. For example, the RAS signaling pathway has long been known to affect metastatic outcomes (McKenna et al. 1990; Chambers 1992). In an experimental animal model of liver metastasis, activated RAS signaling was shown to play a critical role in maintaining metastatic growth (Varghese et al. 2002). However the traditional progression model for metastasis does not lose continuing support from biological studies. It has been demonstrated using animal models that cell subpopulations that pre-exist within the same tumor have heterogonous metastatic potential (Fidler 2003). Genetic analysis on individual disseminated tumor cells using the combination of immunocytochemical staining and fluorescence *in situ* hybridization (FISH) demonstrated that breast cancer cells having migrated into the bone marrow are genetically heterogeneous (Klein et al. 2002). In addition, it was found that gene expression profiles of metastasis were distinct in different organs from the same primary tumors, indicating that the genetic progression of the disseminated cells must have been subjected to different selective pressure (Kang et al. 2003). It remains to be seen in the future how to incorporate these two seemingly conflicted progression models in tumor metastasis. Pantel et al. hypothesized that metastatic spread might follow the traditional and new models, which are complementary but emphasize specific routes, depending on whether cancer cells disseminate from the primary tumor to the lymph nodes or blood during the early stages of tumor growth (Pantel et al. 2004).

At present, the popular and efficient way of identifying genes associated with the spread of cancer is using microarray analysis, or rarely, serial analysis of gene expression (SAGE) technology, to compare gene expression profiles between primary tumors and metastasis, or between non-metastatic and metastatic cell lines. Genes with a different expression pattern between samples to be compared may have an important role in the progression from localized to disseminated cancer. These genomic approaches have helped to identify a large number of genes involved in CRC metastasis (Hegde et al. 2001; Saha et al. 2001; Li et al. 2004). In paper V, a similar approach was adopted to identify *TJP3* as a novel candidate metastasis-associated gene in CRCs with MSI.

1.3.4 Low-penetrance susceptibility alleles

There is increasing evidence to suggest that the inheritance of most common cancers is polygenic. Pharoah et al. found that the polygenic model is very compatible with the distribution of breast cancer risk in the population. He proposed according to his model that the breast cancer susceptibility in the population is mainly due to a large number of alleles each conferring a small genotypic risk (perhaps of the order of 1.5-2.0) combined additively or multiplicatively (Pharoah et al. 2002). Segregation analysis suggested that half of all breast cancer cases may occur within the most susceptible 12% of the population with an 11% or greater risk of breast cancer by age 70 (Pharoah

et al. 2002). Thus it is reasonable to expect that the construction and use of genetic risk profiles based on low-risk, or low-penetrance, alleles would provide significant improvements in the efficacy of population-based programs of intervention for cancers and other common diseases (Pharoah et al. 2002).

It is very likely that this polygenic model also applies to many other common cancers and diseases, including CRC. The number of low penetrance allele for any specific cancer is unknown, but might be in the hundreds or thousands (Pharoah et al. 2004). The candidate low-penetrance alleles for CRC may come from those genes involved in metabolic pathways, or in methylation, those modifying the colonic microenvironment, oncogenes, tumor suppressor genes, mismatch repair genes, and genes involved in immune response (de Jong et al. 2002). The strongest case for the existence of low-penetrance alleles for increased CRC risk is represented by the *APC* I1307K polymorphism. This variant was originally found in an Ashkenazi Jewish individual with a family history of CRC fortuitously due to the occurrence of a truncating APC product (Laken et al. 1997). It is found in 5-7% of the Ashkenazi population and 8-15% of Ashkenazi Jews with CRC, and the relative risk of CRC with this allele is estimated to be around two fold in several studies (Gryfe et al. 1998; Woodage et al. 1998; Rozen et al. 1999; Drucker et al. 2000; Zauber et al. 2003). Though many published case-control studies for studying low-penetrance alleles have often been criticized for lacking of reproducibility due to reasons such as insufficient statistical power and inappropriately chosen cases and controls, more and more low-penetrance alleles for CRC predisposition are found with good evidence, such as by pooled meta-analysis, well-designed case-control studies, and functional tests. The representative low-penetrance alleles for CRC are summarized in **Table 2**.

Recent data have suggested that epigenetic alterations in normal cells may as well be associated with common cancer risk. LOI of *IGF2* is found in 10% of the population and appears to be five-fold more common in patients with a family history of CRC, and 21-fold more common in patients with a personal history of CRC (Cui et al. 2003; Feinberg 2004). In addition, modifier genes that can influence the severity of a Mendelian syndrome should also, in principle, be classified as low-penetrance genes. A classical example is a modifier locus named Mom-1 (modifier of Min 1), able to modify the number of intestinal polyps in $Apc^{Min+/-}$ mouse. Later, this region was shown to contain a gene encoding a secretary phospholipase (*PLA2G2A*). However, the human orthologue of *PLA2G2A* does not seem to have a similar effect on FAP patients with *APC* mutations (Spirio et al. 1996).

1.3.5 Genetic alterations and tumorigenesis pathways

Tumorigenesis is a multistep process, each step reflecting genetic alterations that drive the progressive transformation of normal human cells into highly malignant derivatives (Hanahan et al. 2000). The research on CRC is leading the way in understanding cancer genetics and provides a useful model for other cancers. One reason is that most CRCs

Table 2. Representative low-penetrance alleles for CRC

Gene	polymorphism	Ratings	Refs
ADH3	Codon 350	*	(Houlston et al. 2001; de Jong et al. 2002)
APC	I1307K	**	(Laken et al. 1997; Woodage et al. 1998)
APC	E1317Q	*	(Houlston et al. 2001; de Jong et al. 2002)
APC	D1822V	*	(Slattery et al. 2001)
APC	8636C>A (3'UTR)	*	Paper II
APOE	ε2/3	*	(de Jong et al. 2002)
BLM	BLM^{Ash} (2281delATCTGAinsTAGATTC	*	(Gruber et al. 2002)
CCDN1	A870G	*	(Grieu et al. 2003; Le Marchand et al. 2003)
CDH1	–347 insA (promoter)	*	(Shin et al. 2004)
CHEK2	del1100C	*	(Meijers-Heijboer et al. 2003)
COX1	L15Δ	*	(Ulrich et al. 2004)
HRAS1	VNTR rare alleles	**	(Houlston et al. 2001; de Jong et al. 2002)
HFE	H63D, C282Y	*	(Shaheen et al. 2003)
IGF1	CA repeat	*	(Slattery et al. 2004)
IRS1	G972R	*	(Slattery et al. 2004)
IL6	–174 G>C	*	(Landi et al. 2003)
IL8	–251 T>A	*	(Landi et al. 2003)
MLH1	D132H	*	(Lipkin et al. 2004; Shin et al. 2005)
MLH3	W1276R	*	Paper I
MMP1	2G	*	(Hinoda et al. 2002)
MMP3	6A	*	(Hinoda et al. 2002)
MTHFR	C677T	**	(Houlston et al. 2001; de Jong et al. 2002)
MTHFR	A1298C	**	(Houlston et al. 2001; de Jong et al. 2002)
MTR	A66G	*	(Matsuo et al. 2002)
NAT2	Rapid acetylator alleles	**	(Houlston et al. 2001; de Jong et al. 2002)
PPARG	P10A	*	(Landi et al. 2003)
TGFBR1	*6A ($del(Ala)_3$)	**	(Pasche et al. 2004)
TNFA	–308 G>A	*	(Landi et al. 2003)
TS	2R/3R promoter	*	(Ulrich et al. 2002; Chen et al. 2003)
VDR	M1T	*	(Wong et al. 2003)
VDR	3'-UTR polyA short/long	*	(Slattery et al. 2004)
STK15	F31I	*	(Ewart-Toland et al. 2003)
EXO1	T439M	*	(Yamamoto et al. 2005)
EXO1	P757L	*	(Yamamoto et al. 2005)
MYH	Y165C	*	(Croitoru et al. 2004; Farrington et al. 2005), Paper III
MYH	G382D	*	(Croitoru et al. 2004; Farrington et al. 2005), paper III

modified from (Kemp et al. 2004)

*some reports of association, but early data or unconfirmed evidence; **good evidence of association. These classifications represent the author's personal opinions only.

are believed to occur through a well-known adenoma to carcinoma sequence identified

almost three decades ago; the other reason is the relatively easy availability of biopsies from the various stages of tumor development for genetic analysis (de la Chapelle 2004). In 1990, Fearon and Vogelstein proposed a stepwise genetic model of colorectal tumorigenesis (Fearon et al. 1990) (**Figure 2**). In this model, also dubbed as "Vogelgram", CRC was supposed to develop in a series of genetic alterations, corresponding with histological progression from normal colonic epithelium to adenomatous dysplasia through microinvasion, adenocarcinoma and, finally, metastasis. This model has been generally validated and some essential components of this model do seem to occur in a predictable manner, such as *APC*, *KRAS* and *p53*.

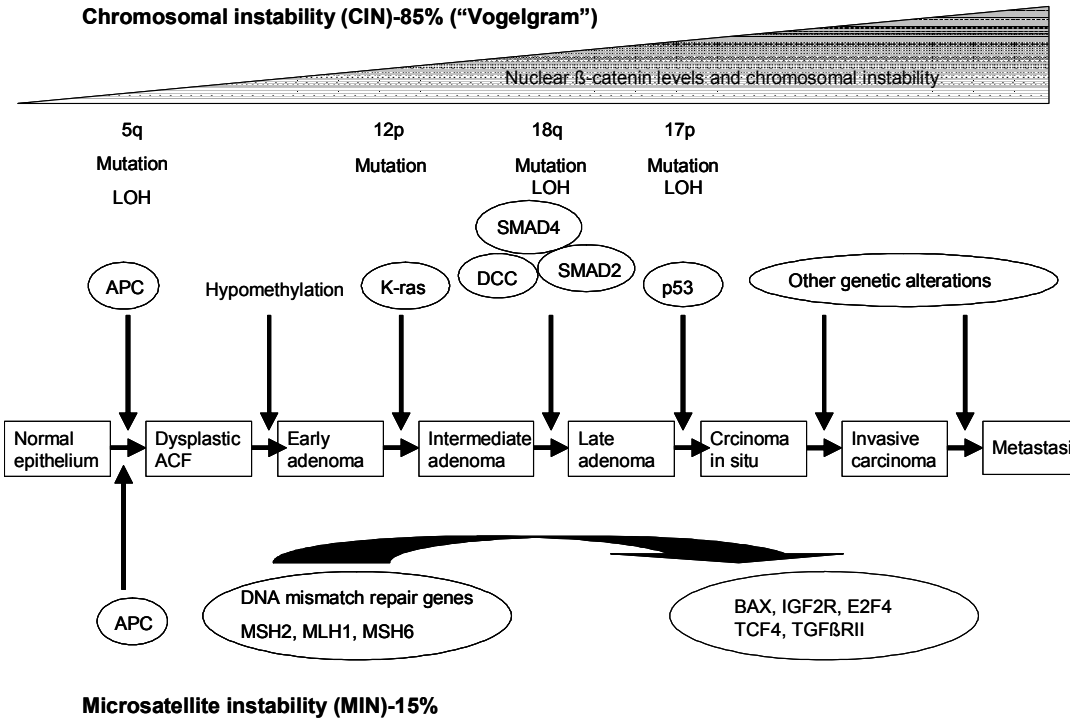

Figure 2. Sequential genetic and epigenetic changes leading to the evolution of colorectal cancer (Modified from Fearon et al. 1990, Chung 2000, and Fodde 2002).

The *APC* gene, localized on chromosome 5q21, is an important component of the canonical Wnt/wingless transduction pathway. In unstimulated cells, APC forms a complex with β-catenin, axin1, axin2 and glycogen synthase kinase (GSK3β), which is capable of earmarking free cytoplasmic β-catenin by phosphorylation for degradation through the ubiquitin/proteasome pathway (Fodde 2002). On inactivation of the two copies of *APC*, β-catenin accumulates in the cytoplasm and shuttles to the nucleus (Fodde 2002). Once in the nucleus, it binds the transcription factor TCF-4, leading to expression of a whole battery of genes that have the potential to change the proliferation and differentiation state of cells, including *c-myc*, *cyclin D*, *Ephrins* and many others (Nathke 2004). Mutations in *APC* have been demonstrated in the majority of FAP patients in the germline (Nakamura et al. 1991; Cottrell et al. 1992) and in more than 80% of sporadic CRC, somatically (Nagase et al. 1993; Kinzler et al. 1996). These

mutations are thought to occur in the earliest stage in colorectal tumorigenesis, and precede the other alterations observed during CRC formation (Vogelstein et al. 1988; Powell et al. 1992). It has been reported that even dysplastic aberrant foci, a presumptive precursor lesion to CRC, harbor *APC* mutations (Jen et al. 1994). In addition, LOH of 5q21 could be observed in at least 30% of colonic adenomas and adenocarcinoma, and it was presumed that *APC* was the most likely target of these events (Vogelstein et al. 1988) . Thus *APC* inactivation appears to be the initial step in colorectal tumorigenesis. In this context, *APC* can be classified as a gatekeeper gene.

Consistent with the important role of the Wnt/wingless signaling pathway in controlling colonic tumorigenesis, mutations in other members of the Wnt pathway have also been shown to be associated with CRC. Gain-of-function mutations in the *β-catenin* have been identified in as many as 50% of colon tumors with intact *APC* (Sparks et al. 1998). In addition, mutations in *Axin2* and *TCF4* have also been reported in a small percentage of CRCs (Duval et al. 1999; Liu et al. 2000).

Of note, the APC protein is a large complex protein with multiple functional domains that mediate both oligomerization and binding to many intracellular proteins. Besides β-catenin, these proteins include γ-catenin, glycogen synthase kinase (GSK3β), Axin, tubulin, EB1 and hDLG (Chung 2000). Thus APC could participate in colonic tumorigenesis via a number of other different mechanisms through modulating actin cytoskeletal integrity, cell-cell adhesion, cell migration, and microtubule assembly and chromosome segregation. However, for most of these mechanisms we are still lacking information.

KRAS, localized on chromosome 12p12, belongs, together with *HRAS* and *NRAS*, to a family of GTPases, which mediates cellular response to growth signals. In CRC, *KRAS* is the preferential target for genetic mutation compared with the other two family members. Only a small percentage of CRCs have been reported to harbor mutations in *N-RAS* (Delattre et al. 1989; Liu et al. 2003). *KRAS* is found mutated in approximately 50% of CRCs and a similar percentage of adenomas greater that 1 cm in size, almost uniformly occurring as activating point mutations in codons 12, 13, and to a lesser extent, 61, as in other human tumors (Bos 1988; Vogelstein et al. 1988; Chung 2000). It has been noticed that small adenomas with *APC* mutations carry *KRAS* mutations in approximately 20% of the tumors, whereas approximately 50% of more advanced adenomas have *KRAS* mutations (Bos et al. 1987), indicating *KRAS* mutations could promote clonal expansion during the early to advanced adenoma stage. However, *KRAS* mutations do not seem to initiate the tumorigenesis as *APC* mutations do. There have been several reports of *KRAS* mutations in histologically normal mucosa. In addition, it has been shown that *KRAS* mutations are not associated with dysplasia alone without the concurrence of *APC* mutations (Jen 1994).

LOH on chromosome 18q21 is found in about 50% of adenomas and 70% of CRCs (fearon 1990) and is often implicated as an indicator for poor prognosis. This minimally lost region contains several candidate tumor suppressor genes, including *DCC*, *DPC4/SMAD4* and *MADR2/SMAD2*. The fact that mutations of *DCC* are rarely found in CRCs and a murine model with inactivation of *DCC* does not affect mouse intestinal growth and differentiation has questioned the tumor suppressor function of *DCC* (Luongo et al. 1996). However, the recent observation that *DCC* triggers cell death and is a receptor for netrin-1 has prompted a renewal of interest in the role of *DCC* in tumorigenesis (Mazelin et al. 2004). Both *SMAD2* and *SMAD4* are intracellular transducers in the transforming growth factor (TGF-β) signaling pathway that inhibits the growth of colon epithelial cells. Mutations of *SMAD2* and *SMAD4* have been detected in several carcinomas including CRC, but are uncommon overall (Hahn et al. 1996). The occurrence of mutations of *SMAD4* in the germline of a subset of juvenile polyposis families further supports the notion that *SMAD4* acts as a tumor suppressor (Howe et al. 1998).

Mutations in the *p53* gene are the most common genetic alterations reported in a variety of human cancers. *p53*, as one of the crucial tumor suppressor genes in maintaining cell homeostasis, integrates numerous signalling pathways; however, the most important consequence of *p53* inactivation in tumorigenesis is probably a complete loss of the DNA damage checkpoints in controlling cell cycle and apoptosis, leading to widespread genomic instability. In CRC, *p53* mutations and allelic losses on 17p manifested as LOH are found in more than 75% of the tested cases (Vogelstein et al. 1988). Furthermore, mutation of *p53* coupled with LOH of 17p was found to coincide with the appearance of carcinoma in an adenoma, pointing to its role for malignant transformation as a late event (Baker et al. 1989; Baker et al. 1990).

Besides the representative genetic changes described above, the "Vogelgram" also stresses the participation of epigenetic alterations in colorectal tumorigenesis. There is a global hypomethylation that coexists with focal hypermethylation after *APC* inactivation. It is suggested that DNA hypomethylation and methylation are independent processes, leading to oncogenic activation and tumor suppressor gene silencing, respectively (Esteller et al. 2002).

Subsequent development of the "Vogelgram" encompasses the different levels of genomic instability (Chung 2000) (**Figure 2**). The "Vogelgram" basically outlines the development of tumors with CIN observed in 85% of CRCs-the CIN pathway. The remaining 15% of CRCs with MSI may follow an independent tumorigenesis pathway, the MSI pathway. The two inherited CRC syndromes, FAP and HNPCC, are examples of these two divergent pathways. As compared with CIN tumors, MSI tumors in the context of HNPCC can show differences in several respects: cytogenetically, MSI tumors have diploid DNA content and lack of LOH at loci harboring tumor suppressor genes such as *APC*, *p53*, and candidate loci on chromosome 18q; genetically, MSI

tumors have a higher frequency of mutations in genes with simple repeat sequences in the coding region, such as *TGFßRII*, *IGF2R*, *BAX*, *E2F-4*, and *Tcf-4*, supposed to surrogate for mutations of *KRAS* and *p53* with a lower frequency in MSI tumors; morphologically, MSI tumors are more likely to arise in the right colon, to have an associated lymphocytic infiltrate, to be more poorly differentiated, and to have a better prognosis (Lothe et al. 1993; Jass et al. 1998; Wright et al. 2000; Lindblom 2001). CIN and MSI tumors have been separated into two distinct clusters or principle components based on expression profiles in microarray analysis (Mori et al. 2003; Bertucci et al. 2004).

Evidence is accumulating that the traditional adenoma-carcinoma sequence as demonstrated in HNPCC and FAP cancers may not apply to sporadic MSI CRCs. Several reports have indicated that most, if not all, sporadic MSI CRCs might develop within serrated polyps comprising hyperplastic polyps, mixed polyps and serrated adenomas (Iino et al. 1999; Jass et al. 2002). Hyperplastic polyps have long been assumed to be innocent, however, recent observational evidence in the literature supports that variant hyperplastic polyps, such as serrated adenomas, do have neoplastic potential as foci of intramucosa cancer (Longacre et al. 1990). Jass et al proposed the serrated pathway to account for the development of CRCs with serrated polyps as suspected precursor (Jass et al. 2000; Jass et al. 2002). Molecular characterization of serrated polyps and sporadic MSI CRCs has shown that the serrated pathway is closely associated with MSI, CIMP and, more recently, activating mutations in the *BRAF* oncogene. *BRAF* participates in the RAS-RAF-MEK-ERK-MAP kinase pathway and is activated in a RAS-dependent manner. The mutation of *BRAF* (mainly V599E) is often found in sessile serrated adenomas, as well as classic serrated adenomas and mixed polyps. In addition, the mutation of *BRAF* is also frequently seen in sporadic CRCs of CIMP, particularly the MSI-H subgroup, but in high contrast, it is rarely detected in HNPCC cancers (Deng et al. 2004; Kambara et al. 2004; Koinuma et al. 2004). These findings highlighted the link between the serrated precursor lesion and its suggested driving forces in tumorigenesis. It is likely that *BRAF* mutation and DNA methylation are key and possibly synergistic events in initiation of the serrated neoplastic pathway. In addition, these findings also indicated that the serrated pathway could underline a proportion of MSI-L and CIN CRCs with CIMP besides sporadic MSI CRCs.

1.4 INHERITED FORMS OF CRC

The genes associated with several inherited syndromes predisposing for colon cancer have been identified. These molecularly characterized syndromes are conventionally divided into the polyposis syndromes and non-polyposis CRC (**Table 3**) (Allen et al. 2003). At present, clinical genetic testing approaches have been developed for most of these syndromes and are now a part of accepted clinical care.

Table 3. Inherited forms of syndromes predisposing for CRC

Syndrome	Gene	Location	Inheritance
Polyposis syndromes			
Adenomatous polyposis syndromes			
Familial adenomatous polyposis	*APC*	5q21	AD
Attenuated adenomatous polyposis coli	*APC*	5q21	AD
Gardner syndrome	*APC*	5q21	AD
Turcot syndrome	*APC*	5q21	AD
MYH-associated polyposis	*MYH*	1p34.3-p32.1	AR
Hamartomatous polyposis syndromes			
Familial juvenile polyposis syndrome	*SMAD4* (15%)	18q21.1	AD
	BMPR1A (25%)	10q22.3	AD
Cowden syndrome	*PTEN* (80%)	10q23.31	AD
Bannayan-ruvalcaba-riley syndrome	*PTEN* (~60%)	10q23.31	AD
Peutz-Jeghers syndrome	*STK11/LKB1* (50-70%)	19p13.3	AD
Non-polyposis syndromes			
Hereditary non-polyposis CRC (HNPCC)	*hMSH2*	2p21	AD
	hMLH1	3p21-23	AD
	hMSH6	2p21	AD
	hPMS2	7p22	AD
Muir-Torre syndrome	*hMSH2*	2p21	AD
	hMLH1	3p21-23	AD
Turcot syndrome	*hMSH2*	2p21	AD
	hMLH1	3p21-23	AD

1.4.1 Familial adenomatous polyposis syndromes (FAP)

FAP was first clearly described as a dominantly inherited Mendelian trait by Lockhart-Mummery in 1925 (Fearnhead et al. 2002). The incidence of FAP is about 1 in 10,000 individuals, accounting for less than 1% of CRCs (Bulow et al. 1996). In classical FAP, affected individuals develop hundreds to thousands of adenomatous polyps in the colon and rectum that usually appear during adolescence or the third decade of life. If left untreated, they will invariably develop CRC by the early forties at the latest. Approximately two-thirds of FAP patients will have congenital hypertrophy of the retinal pigment epithelium (CHRPE) (Blair et al. 1980). Other extracolonic manifestations of FAP include upper gastrointestinal tumors, mandibular osteomas, desmoid tumors, papillary carcinoma of the thyroid, medulloblastoma, hepatoblastoma and soft tissue tumors (Fearnhead et al. 2001).

FAP is caused by germline mutations of the tumor suppressor *APC* gene located on chromosome 5q21, which was identified by a positional cloning strategy in 1991

(Groden et al. 1991; Kinzler et al. 1991). It encodes a 2843 amino acid multidomain protein in its commonest isoform. Exon 15 comprises 75% of the coding sequence, thus being the most common target for both germline and somatic mutations (Beroud et al. 1996). The vast majority of *APC* mutations are nonsense or frameshift mutations that result in truncation of the APC product. A small number of FAP patients has been demonstrated to carry genomic rearrangements and deletions of *APC* (Su et al. 2000; Sieber et al. 2002; Su et al. 2002). Further information on these mutations can be found in the Human Genetic Disease Database and the *APC* database (Beroud et al. 1996). It has been noticed that there is a good genotype-phenotype correlation. Germline mutations between codons 168 and 1680 are associated with classical FAP (Miyoshi et al. 1992; Nagase et al. 1993) whereas germline mutations between codons 457 and 1444 are associated with CHRPE (Caspari et al. 1995). The occurrence of mandibular osteomas and desmoid tumors are often found to be associated with germline mutations between codons 1395 and 1560 (Wallis et al. 1999).

Routine mutation detection technique, such as the protein truncating test (PTT), denaturing gradient gel electrophoresis (DGGE) and DNA sequencing, can only identify germline mutations in approximately 70% of individuals with classical FAP (Sieber et al. 2002). Even with the use of conversion technology to separate alleles prior to mutation screening (Yan et al. 2000), the underlying germline mutations for some cases still remain elusive (Yan et al. 2002). In a recent study, one FAP case with unidentifiable *APC* mutation was demonstrated to have about 50% reduction in transcript expression from one allele (Yan et al. 2002). The observed allele-specific loss of expression could be shown to be inherited, and linked with the development of FAP (Yan et al. 2002). However the molecular basis for reduced expression of one allele remains undefined (Yan et al. 2002).

1.4.2 Attenuated adenomatous polyposis coli (AAPC)

AAPC, a variant of FAP, is characterized by greatly reduced number of polyps (<100), late onset of carcinoma and frequently, the absence of extracolonic features (Sieber et al. 2000). The mutations in *APC* associated with AAPC have mainly been detected in three parts of the gene: in the 5' end (the first 5 exons), in exon 9, and in the distal 3' end (Knudsen et al. 2003). These mutations are viewed as hypomorphic, encoding shorter than full-length proteins to explain the lower number of polyps in AAPC cases (de la Chapelle 2004). One possible mechanism for mutations occurring at the beginning of the gene to have hypomorphic APC is that they employ an in-frame initiation codon to reinitiate the translation (Heinen et al. 2002).

1.4.3 Gardner's syndrome

Gardner's syndrome, another variant of FAP, is characterized by the development of osteomas, desmoid tumors, epidermoid cysts, fibromas and odontomas in addition to colonic adenomas (Gardner et al. 1953). Individuals with Gardner's syndrome are at the same risk for CRC as classic FAP.

1.4.4 Turcot syndrome (TS)

TS is defined by the association of colorectal and central nervous system (CNS) tumors. It can be associated with two different types of germline genetic defects: mutation of the *APC* gene that is usually found in FAP, or mutation of a mismatch-repair gene (*hMLH1* or *hMSH2*) that is usually found in HNPCC (Hamilton et al. 1995). TS with *APC* mutations usually presents with childhood cerebellar medulloblastoma, while TS with mismatch repair gene mutations is often associated with gliomas arising in both children and adults (Lucci-Cordisco et al. 2003).

1.4.5 *MYH*-associated polyposis (MAP)

A recent finding is that a considerable proportion of FAP/AAPC cases with unidentifiable *APC* mutations and cases with multiple colorectal adenomas can be attributable to germline biallelic mutations in the BER gene *MYH*. The first evidence for MAP came from a study of a British Caucasian family in which three siblings were affected by multiple colorectal adenomas and carcinoma, but no detectable inherited *APC* mutations or microsatellite instability in their tumors. Analysis of mutations of *MYH* showed that all three affected siblings were compound heterozygotes for the missense mutations Y165C and G382D. Unaffected family members were either homozygous normal or heterozygous for one of the mutations (Al-Tassan et al. 2002), suggesting *MYH* as a recessive predisposition gene for CRC. Other investigators subsequently confirmed the association between these two and other biallelic germline *MYH* mutations and colorectal polyposis and CRC known not to have germline mutations in *APC* (Jones et al. 2002; Enholm et al. 2003; Sampson et al. 2003; Sieber et al. 2003; Gismondi et al. 2004; Vandrovcova et al. 2004). These reports showed that MAP accounts for 20-30% of patients with 15-100 colorectal adenomas and about 10% of patients with classic polyposis. Mutation screening studies on population-based series of CRCs demonstrated that MAP could be present in about 1% of all CRCs, indicating its important role in CRC predisposition as common as FAP (Enholm et al. 2003; Croitoru et al. 2004; Farrington et al. 2005). In addition, *MYH* mutations seems to be ethnic-specific. The Y165C and G382D mutations have been found to be the most common changes in Caucasian populations, accounting for 80% of all *MYH* variants reported to date. The Y90X mutations seems confined to patients of Pakistani origin, whereas E466X in Indian cases (Chow et al. 2004).

MAP is difficult to differentiate clinically from FAP and AAPC, though it tends to present later and be confined to the gastrointestinal tract (Chow et al. 2004). MAP individuals with early-onset CRC and few adenomas have also been reported (Enholm et al. 2003; Wang et al. 2004; Farrington et al. 2005). Thus mutation screening of *MYH* should be considered not only in patients with multiple polyps and a family history compatible with autosomal recessive inheritance, but also in patients with early-onset CRC and /or few adenomas.

1.4.6 Hereditary non-polyposis syndromes (HNPCC)

HNPCC is the most common inherited form of CRC, accounting for 1-3% of all CRC cases (Allen et al. 2003). HNPCC is also called the Lynch syndrome after Henry Lynch, an oncologist who pioneered the study of this disease (Lynch et al. 1971). This syndrome could be traced back to the study of a famous "Family G" by Aldred Warthin in 1913 (Warthin 1913). HNPCC is an autosomal dominant disorder characterized by early-onset CRC with MSI (often right-sided), few adenomas and specific extra-colonic cancers, including those of the endometrium, ovary, stomach, small bowel, uroepithelial tract, hepatobiliary and brain (Lynch et al. 1977). The cumulative cancer risk in HNPCC approaches about 80% and 50-60% for colorectal and endometrial cancer, respectively. The corresponding lifetime risks for other cancer types remain between 2% and 13% (Aarnio et al. 1999).

To create a set of consistent clinical features that define HNPCC and allow for sharing of data in multicenter studies, the International Collaborative Group on HNPCC (ICG-HNPCC) established the first diagnostic criteria based on CRC, known as "Amsterdam criteria I" (Vasen et al. 1991). However Amsterdam criteria I are very stringent, and a large number of individuals who may have HNPCC will be missed if these are the sole criteria used for considering the diagnosis in the clinical setting (Allen et al. 2003). Thus the revised Amsterdam criteria (Amsterdam criteria II) were developed to include extracolonic HNPCC-associated cancers, cancers of the endometrium, small bowel, ureter, and renal pelvis (Vasen et al. 1999) (**Table 4**). A different set of guidelines, the Bethesda criteria, are also developed, aiming to select patients to whom MSI analysis can be offered for identifying potential individuals with HNPCC (Umar et al. 2004) (Table 4)

Table 4. Amsterdam II and Bethesda criteria

Revised Amsterdam criteria (Amsterdam criteria II)
There should be at least three relatives with an HNPCC-associated cancer (cancer of the colorectum, endometrium, small bowel, ureter, or renal pelvis) and
1. One should be a first-degree relative to the other two
2. At least two successive generations should be affected
3. At least one should be diagnosed before age 50
4. Familial adenomatous polyposis should be excluded
5. Tumours should be verified by pathological examination
Bethesda guidelines (ver. 2004)
Tumours from individuals should be tested for MSI in the following situations:
1. Colorectal cancer diagnosed in a patient who is less than 50 years of age
2. Presence of synchronous, metachronous colorectal, or other HNPCC-associated tumors, regardless of age
3. Colorectal cancer with the MSI-H histology diagnosed in a patient who is less than 60 years of age
4. Colorectal cancer in a patient with one or more first-degree relatives with an HNPCC-related tumour, with one of the cancers being diagnosed under age 50 years
5. Colorectal cancer in a patient with two or more first- or second-degree relatives with HNPCC-related tumors, regardless of age

HNPCC is caused by germline mutations in one of the mismatch repair (MMR) genes, with more than 90% of the identifed mutations in *hMSH2* and *hMLH1*. Approaximately 5% of HNPCC families could be accounted for by a germline mutation in *hMSH6* (Jarvinen 2004). Germline mutations in *hPMS1* and *hPMS2* have been described (Nicolaides et al. 1994; Nicolaides et al. 1998), but have not been shown clearly to predispose to HNPCC (Liu et al. 2001). Over 400 different pathogenic mutations have been registered in the international HNPCC database (http://www.nfdht.nl). Mutations in MMR genes tend to be point mutations, resulting in frameshifts, truncations, or missense mutations (not priori deleterious). Genomic arrangement can be detected in *hMSH2* in 10-20% of HNPCC cases, more common than that of *hMLH1* and *hMSH6* (Di Fiore et al. 2004). There appears to be a genotype-phenotype correlation as with FAP. For example, *hMSH2* mutations are more common in patients with an excess of extra-colonic tumors. Women with *hMSH6* mutations appear to be more likely to develop endometrial cancer (Wijnen et al. 1999; Charames et al. 2000) and *hMSH6*-related tumors tend to show a lower degree of instability (Wijnen et al. 1999). As with FAP, allele-specific loss of expression of *hMLH1*, *hMSH2* and *hMSH6* is also observed in a small number of cases without sequence changes in coding regions (Renkonen et al. 2003). As the effect is restricted to the same allele in all cells, sequence changes in introns or control regions causing aberrant expression or degradation of the transcript are likely to underlie this effect (de la Chapelle 2004). Recently it was found that a transcribed paralogous gene to *hPMS2*, *PMS2CL*, could greatly obscure the mutation detection and interpretation of *hPMS2* (De Vos et al. 2004; Nakagawa et al. 2004). This indicates that actual mutations in *hPMS2* have probably been overlooked and are more common than previously thought (Nakagawa et al. 2004). Some studies have suggested the involvement of *hMLH3* in CRC predisposition including paper I in this thesis, although the functional and clinical significance of these mutations needs further investigation (Wu et al. 2001).

In addition to MMR genes, *AXIN2*, *TGFβR2*, and *POLD* have been reported to be causes for a very small proportion of dominantly inherited CRC (da Costa et al. 1995; Lu et al. 1998; Lammi et al. 2004). The likelihood of *EXO1* in HNPCC predisposition was brought into question recently, because of the identification of similar variants in a control population (Jagmohan-Changur et al. 2003).

1.4.7 Muir-Torre syndrome (MTS)

MTS is a rare inherited disorder characterized by predisposing to both sebaceous skin tumors (especially sebaceous gland tumors) and internal neoplasm (mostly CRC) (Schwartz et al. 1995). In a significant proportion of MTS patients, skin tumors exhibited microsatellite instability that is a hallmark of HNPCC, thus it has been suggested that MTS represents a rare variant of HNPCC. In MTS, the vast majority of germline mutations have been identified in *hMSH2*, with a few exceptions in *hMLH1*(Lucci-Cordisco et al. 2003).

1.4.8 Familial juvenile polyposis syndrome (JPS)

JPS is a dominantly inherited disorder with an incidence of 1 in 100,000 to 1 in 160,000 births. It is the most common form of the hamartomatous syndrome and characterized by multiple, hamartomatous polyps affecting the colon and rectum. The incidence of colon cancer is 17-22% by age 35 and approaches 68% by age 60 (Schreibman et al. 2005). Germline mutations of *SMAD4* are identified in 15% of JPS cases, and those of *BMPR1A* in 25% of cases (Sayed et al. 2002). Both genes are involved in bone morphogenetic protein (BMP) mediated signalling and are members of the TGF-ß superfamily. In addition, 5% of claimed JPS cases are identified with a mutation in the *PTEN* gene (phosphatase and tensin homolog), responsible for the Cowden's syndrome. However, upon further review, these patients in fact had the Cowden's syndrome (Schreibman et al. 2005).

1.4.9 Cowden's syndrome (CS)

CS, also known as multiple hamartoma syndrome, is an autosomal dominant cancer condition associated with numerous possible clinical manifestations. Commonly present are skin changes, including acral keratoses and facial trichilemmomas, as well as oral papillomas and scrotal tongue. There is an increased risk of malignancies including breast, thyroid, and endometrial cancers. Hamartomas may affect multiple systems including the skin, gastrointestinal tract, central nervous system, breast, and thyroid (Hanssen et al. 1995; Delatycki et al. 2003). CS is rarer than JPS with a prevalence of 1 in 200,000. At present it is questionable whether individuals with CS are at increased risk for intestinal cancer as previously believed (Schreibman et al. 2005). Approximately 80% of CS patients are found to harbour germline mutations in the *PTEN* gene (Marsh et al. 1998). *PTEN* has been suggested to be a tumor suppressor gene, negatively regulating the PI3k pathway by inhibiting the phosphorylation as a lipid phosphatase (Weng et al. 1999).

1.4.10 Bannayan-ruvalcaba-riley syndrome (BRR)

Based on the clinical similarities between the three autosomal dominant hamartomatous polyp syndromes, Bannayan-zonana syndrome, Riley-Smith syndrome, and Ruvalcaba-Myhre-Smith syndrome, geneticists began to accept the notion of combing these three disorders into a single entity, Bannayan-ruvalcaba-riley syndrome (BRR) (Gorlin et al. 1992). The shared clinical manifestations may include macrocephaly, multiple subcutaneous and visceral lipomas and hemangiomas, hamartomatous intestinal polyposis, and pigmentary spotting of the penis. Among these manifestations, intestinal polyposis could affect up to 45% of these patients, mainly limited to the distal ileum and colon (Schreibman et al. 2005). The *PTEN* gene has been identified as a susceptibility gene (Marsh et al. 1999).

1.4.11 Peutz-Jeghers syndrome (PJS)

PJS is another rare autosomal dominant hamartomatous condition, occurring in approximately 1:200,000 live births. It is characterized by mucocutaneous melanotic pigmentation, gastrointestinal hamartomatous polyposis, and an increased risk for the development of both gastrointestinal and extra-intestinal malignancies including colorectal, oesophageal, gastric, small intestinal, breast, ovary and pancreatic cancer (Boardman et al. 1998). In contract to JPS, the hamartomatous polyps in individuals with PJS are most prevalent in the small intestine, but may also be present in the stomach and large bowel (Boardman et al. 1998). To date, the only identifiable susceptibility gene for PJS is *STK11* (serine/threonine-protein kinase 11, also known as *LKB1*), though linkage studies have suggested other loci (Olschwang et al. 1998). PJS is a candidate tumor suppressor gene that encodes a multifunctional serine-threonine kinase involved in the transduction of intracellular growth signals (Schreibman et al. 2005). Genetic studies have showed that germline *STK11* mutations can be identified in 50-70% of PJS cases with a family history and there is a high frequency of de novo mutations in sporadic cases (Westerman et al. 1999).

1.5 IDENTIFICATION OF CRC SUSCEPTIBILITY GENES

1.5.1 Identification of high-penetrance genes

Family-based linkage analysis has been the foundation for the many successes in mapping of genes associated with Mendelian disorders and several rare subtypes of common cancer with the near-Mendelian inheritance mode, such as, *APC* (Groden et al. 1991; Kinzler et al. 1991), *hMLH1* (Lindblom et al. 1993) and *hMSH2* (Peltomaki et al. 1993) in colorectal cancer. Thus, for those families showing a dominant inheritance mode of unknown high-penetrance genes, linkage analysis is still the first choice. Recently, a linkage study of 53 kindreds affected with CRC or large adenomas provided evidence for a previously unidentified susceptibility locus on chromosome 9q. Interestingly, our linkage data on a large single CRC family pointed to the same region (Skoglund et al. 2005). It should be noted that when linkage analysis is applied to complex disorders such as CRC, it is inherently afflicted by any or all of the following factors: high population frequency, incomplete penetrance, phenocopies, genetic heterogeneity, possibly epistasis, and pleiotropy. The problem is even more exaggerated when many small-sized families are used in a single study. As a consequence, the application of linkage analysis to complex disorders has shown limited success so far and replication of any positive results published is very difficult. At the present, non-parametric linkage (NPL) analysis appears to be the method of choice for complex diseases. The other alternative is to use classical parametric linkage analysis in large single families. In addition, the map location identified in linkage studies of complex diseases is usually rather broad, and can encompass hundreds or even thousands of possible genes across many megabases of DNA. Within such a region, the identification of a causative gene represents a formidable challenge.

1.5.2 Identification of low-penetrance genes

The association study is to compare the frequency of a genetic variant in diseased individuals (cases) and individuals without the disease (controls), and has been heralded as a promising method for identification of low-penetrance alleles (Risch et al. 1996; Risch 2000; Cardon et al. 2001). Allelic association is present when the distribution of genotypes differs in cases and controls (Pharoah et al. 2004). The case-control study is the most commonly used population-based study design for allelic association. The advantages to this approach are that cases are readily obtained and can be efficiently genotyped and compared with control populations (Cardon et al. 2001). Despite its ease, however, this approach has been most prone to spurious association results. Alternative approaches include using family-based controls to avoid the potential problem of population stratification, such as the transmission disequilibrium test (TDT). Though this approach has been popularized by application to many diseases, it is rarely used for cancer because recruitment of parents of diseased individuals is almost impossible in large numbers, and the use of other family members severely reduces power (Pharoah et al. 2004). Conventional case-control gene-association studies have a long history of false-positive results (Risch et al. 1996). Either they cannot be replicated or they lack the corroboration by linkage analysis (Cardon et al. 2001). Two published systematic reviews on specific polymorphisms and CRC risk can also illustrate this apparent failure. Only a few polymorphisms (see Table 2) out of dozens are reported in more than one report and most reports had less than 80% power to detect a twofold difference in cancer risk at the 0.05 significance level (Houlston et al. 2001; de Jong et al. 2002). This is, of course, not saying that all published associations are incorrect, merely that adequately powered studies have not been carried out to confirm them (Pharoah et al. 2004). The common errors encountered in association studies of complex diseases are summarized in **Table 5** (Cardon et al. 2001).

Table 5. Common errors in association studies

● Small sample size
● Subgroup analysis and multiple testing
● Random error
● Poorly matched control group (population admixture)
● Failure to detect linkage disequilibrium with adjacent loci
● Overinterpreting results and positive publication bias
● Unwarranted “candidate gene” declaration after identifying association in arbitrary genetic region

Despite these known limitations, the power of association analysis to detect genetic contributions to complex diseases can be much greater than that of linkage studies (Risch et al. 1996). To detect a gene with frequency 0.1, conferring a two fold increase in risk by linkage would require about 10 000 affected sibling pairs. In contrast, it should be detectable through association with only 500 unselected cases and 500 controls (Houlston et al. 2004). Some guidelines and solutions have been proposed to attenuate some of the most common limitations of the association study, such as inadequate samples and population stratification. The lack of samples can be resolved

by forming multicenter collaborations. One example of this type of collaboration is the Consortium of Cohorts, which was formed by National Cancer Institute to address the need for large-scale collaborations for the study of gene-gene and gene-environment interactions in the aetiology of cancer and has more than 20 cohorts participating (Pharoah et al. 2004). Another solution is to use early-onset cases or cases with affected relatives. It has been shown that, relative to a standard case-control association study with cases unselected for family history, the sample size required to detect a common disease-susceptibility allele is typically reduced by more than twofold if cases with one affected first-degree relatives are used, and by more than fourfold if cases with two affected first-degree relatives are used (Antoniou et al. 2003). Population stratification has been a well-known explanation for spurious association results. However, recent investigation suggested that population stratification does not seem to be a serious problem for large association studies if cases and controls are well matched (Cardon et al. 2003). Some investigators have begun to explore the use of unlinked markers to detect population stratification and even correct it when it is present (Pritchard et al. 1999; Bacanu et al. 2000; Pritchard et al. 2000). These unlinked markers used for testing stratification are chosen from a panel known to exhibit differences in allele frequency between populations (Pritchard et al. 1999).

So far, association studies based on candidate genes have been the only practical alternative to linkage analysis and almost all studies reported to date have evaluated only one or two SNPs at each candidate loci (Houlston et al. 2004). The chance of success for candidate-gene association studies can be greatly improved by careful selection of both candidate genes and candidate polymorphisms (Tabor et al. 2002). Candidate genes can be chosen based on known or predicted function, or from linkage peaks generated by family-based studies. Other possibilities include use of animal models or searching for intermediate phenotypes that are strongly heritable and also associated with the disease phenotype of interest (Pharoah et al. 2004). The traditional way to discover candidate polymorphisms is re-sequencing a sample of the relevant population across the genomic region of interest. This is an laborious, time-consuming and costly procedure. The past few years have seen rapid acquisition of data on the occurrence of common SNPs, making it possible to reach nearly all gene variants without formal re-sequencing (Pharoah et al. 2004). The dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP) now contains nearly 9 million SNPs, including most of the around 11 million SNPs with minor allele frequency of 1% or greater that are estimated to exist in the human genome (Kruglyak et al. 2001). It has been known that many SNPs have alleles that show strong linkage disequilibrium (LD) with other nearby SNP alleles. One SNP can thereby serve as a proxy, tagSNP, for many others in an association screen. Once the patterns of LD are known for a given region, a few tagSNPs can be chosen, individually or in multimarker combinations (haplotypes), to capture most of the common variation within that region (Johnson et al. 2001; Gabriel et al. 2002). Characterization of patterns of LD across the human genome is now an area of highly active research, led by a large international initiative to

evaluate LD patterns in many populations, the International HapMap Project (http://www.hapmap.org) (Couzin 2002). At present, there is emerging literature in which to use the multilocus genotype data to estimate and compare haplotype frequencies in cases and controls (Barnby et al. 2005; Ma et al. 2005; Pimm et al. 2005). It has been shown that, for case-control studies, haplotype-based methods can be more powerful than single-locus analyses when the SNPs are in LD with a causative diallelic locus (Xiong et al. 2000).

The candidate-gene association study will, at best, identify only a fraction of genetic risk factors even for diseases in which the pathophysiology is relatively well understood (Hirschhorn et al. 2005). To fully understand the allelic variation that underlies complex diseases, complete genome sequencing for many individuals with and without disease is required (Wang et al. 2005). Recently it has become possible to carry out partial surveys of the genome by genotyping large numbers of common SNPs in genome-wide association studies with the advent of high-throughput SNP analytical platforms, such as SNP microarrays (Wang et al. 2005). A more controversial area in the optimal design of genome-wide association studies revolves around the question of how many SNPs to use and whether the approach should be haplotype-map or sequence based (Kemp et al. 2004). The haplotype-map based approach would involve first identifying all the common variants in the genome and then selecting an appropriate set of tagging SNPs for genotyping in a case-control set (Pharoah et al. 2004). Such a study is not yet feasible because a high-density SNP map, for example from the International HapMap Project, that covers the whole genome is not yet available (Pharoah et al. 2004). It has been estimated that 200,000-500,000 SNPs will be needed to adequately tag all SNPs with a minor-allele frequency of 5% or more (Kruglyak 1999). Advocates of a sequence-based approach proposed that genome-wide studies should be focused on missense SNPs (Botstein et al. 2003). This strategy would require the genotyping of only 30,000-60,000 SNPs and would potentially detect lower frequency alleles (1-20%) (Botstein et al. 2003; Kemp et al. 2004). However, the rationale that underlies this sequence-based approach has been questioned a lot. It may be true that causal alleles for monogenic disorders include a large proportion of missense variants (Hirschhorn et al. 2002), but it may not be case for complex diseases. Clearly, the causal alleles for complex diseases are far less likely to be subject to strong negative selection and might therefore comprise different types of variants to those that underlie monogenic disorders. Therefore, the causal alleles for complex diseases are more likely to include non-coding regulatory variants with a modest impact on expression (Hirschhorn et al. 2005). Despite the ambiguities outlined above, there are good reasons to expect that genome-wide association studies for uncovering the genetic components of common diseases will be carried out in the near future as evidenced by the rapid evolution of dense SNP maps and high-throughput genotyping technologies in the past few years.

The current generation of association studies depends on the “common disease, common variant (CDCV)” hypothesis, because it involves testing common

polymorphisms or haplotypes (Pharoah et al. 2004). It is however, equally possible that disease susceptibility is due to distinct genetic variants in different individuals and disease-susceptibility alleles have very low population frequencies, the "multiple rare-variant hypothesis" (Terwilliger et al. 1998; Wang et al. 2005). A typical example is the low-penetrance variant of 1100delC in *CHEK2* that predisposes to breast cancer, and possibly CRC, whose population frequency is as low as 1% (Meijers-Heijboer et al. 2002; de Jong et al. 2005). Rare alleles as such are unlikely to be found by typing arbitrarily selected SNPs in a genome-wide screen even in regions of strong LD (Houlston et al. 2004). A mutation screening strategy based on candidate genes (and followed, if possible, by a suitable association study) is more likely to be a better choice (Kemp et al. 2004).

2 AIMS OF THE STUDY

1. To determine the contribution of *hMLH3*, *APC* and *MYH* as either high- or low-penetrance susceptibility genes for CRC in the Swedish population (Papers I, II, III).

2. To explore the possibility of *APC* in predisposing autism spectrum disorder (ASD) using association studies (Paper IV).

3. To generate expression profiles for hereditary and sporadic MSI CRCs using Affymetrix® Microarray, with the aim 1) to see if these two subtypes of tumor could be differentiated based on expression profiles, 2) to identify genes that are associated with a specific clinicopathological trait, for example, metastasis.

3 MATERIALS AND METHODS

3.1 MATERIALS

3.1.1 Samples used in Papers I, II, III

Patients with suggested familial predisposition were recruited from a large cohort of consecutive patients who underwent oncogenetic counseling at the Cancer Family Clinic from the year 1990 to 1999 at Karolinska Hospital, Stockholm, Sweden. Family history was obtained and all diagnoses were ascertained through medical records, pathology reports or in very few cases, death certificates. All members at increased risk for CRC from each family were counseled and offered regular colonoscopy surveillance. Based on the family history, these patients can be classified into groups of HNPCC, HCRC, TCR and OCR (see the Introduction part for the definition of each group). HNPCC cases occupied a small proportion of all cases, and were only included for studying in Paper I. MSI status had been valuated in tumors for all patients. All HCRC, TCR, and OCR patients, as well as five HNPCC patients meeting the Amsterdam I criteria in Paper I, had MSI negative tumors. In addition, germline mutations in *hMLH1, hMSH2* and *hMSH6* had been excluded in all but two cases with MMR gene missense variants of unclear pathogenicity in paper I (Wahlberg et al. 1999; Huang et al. 2001).

The exact number of sporadic CRC cases and normal controls used in association studies in different papers varied with the experimental design. Sporadic CRCs were available either from the Department of Surgery at Uppsala University Hospital (Sweden) or from the Department of Pathology at Linköping Hospital (Sweden). The first source of normal controls was unrelated healthy relatives of patients undergoing genetic testing at the Department of Clinical Genetics of Karolinska Hospital or at Department of Pathology of Linköping Hospital. The medical information of these controls could be available if needed. The second source of normal controls was subjects randomly chosen from a large cohort of healthy blood donors in the Stockholm region. Information about family history of colorectal cancer or polyps, age, and gender was not available for these subjects.

3.1.2 Samples used in Paper IV

There were two groups of ASD cases of different ethnic origin used in the association studies (**Tables 6a and 6b**). To match the cases in ethnicity, there were also two groups of normal controls. One group consisted of 67 Iranians and the other group consisted of 476 Swedes.

Table 2a. Characterization of 39 Middle Eastern and African ASD cases
(diagnosed at Department of Clinical Genetics at Karolinska Hospital)

Primary Diagnosis	*No of subjects*
Autism	39
Comorbidity and associated handicaps	
Mental retardation	29
ADHD (Attention-Deficit/Hyperactivity Disorder)	1
Epilepsy	2
Country of Origin	
Iran	6
Iraq	5
Uganda	5
Ethiopia	4
Eritrea	3
Syria	3
Somalia	2
Turkey	2
Gambia	1
Morocco	1
Ghana	1
Other/unknown	6

The 39 cases include 29 males and 10 females and the ages of whom range between 7 and 18 years old (median 11.5).

Table 2b. Characterization of 75 Swedish ASD cases
(diagnosed at Department of Child and Adolescent Psychiatry at Göteborg University)

Primary Diagnosis	*No of subjects*
Autism	26
Asperger syndrome	22
Atypical autism	21
Autistic traits	6
Comorbidity and associated handicaps	
Mental retardation	21
Chronic tics	14
ADHD (Attention-Deficit/Hyperactivity Disorder)	33
Bipolar disorder	4
Epilepsy	1
Country of Origin	
Sweden	75

The 75 cases include 51 males and 24 females and the ages of whom range between 4 and 55 years old (median 25.0).

3.1.3 Samples used in Paper V

3.1.3.1 Tumor materials

A total of 14 tissue samples from 13 CRC patients used in microarray analysis were taken from Department of Surgery either at Huddinge University Hospital or Uppsala University Hospital. The samples were stored at -80^{o}C after snap freezing in liquid nitrogen within a short time after surgery. All the samples had been tested to be MSI high. Seven of the 14 samples were defined as hereditary tumors derived from patients the family history of whom met the Amsterdam Criteria I or II, and all found to carry a germline mutation in mismatch repair genes of either *hMLH1* or *hMSH2* at our Cancer Family Clinic. The remaining seven samples were defined as sporadic tumors derived from patients over 56 years of age with no family history of CRC.

3.1.3.2 Cell lines

HCT116 and HCT116M2 are isogenic cell lines, evidenced by their very similar karyotypes based on G-banding performed (data not shown). HCT116 was purchased from the American Type Culture Collection (Rockville, MD). HCT116M2 was a kind gift from Professor Vogelstein at Johns Hopkins University School of Medicine, USA. HCT116M2 has acquired high metastatic capability as compared to HCT116 with low metastatic capability.

3.2 METHODS

3.2.1 Denaturing high-performance liquid chromatography (DHPLC) (paper I, II, III, IV)

DHPLC has emerged as one of the most versatile technologies for the analysis of genetic variations. It is based on the principle that heteroduplexes formed between mutated DNA and normal DNA molecules can be distinguished from homoduplexes consisting entirely of normal DNA molecules by a different elution profile at a given optimized temperature. The experimental process includes basically two steps. The first step is to bind nucleic acids (for example, PCR products) to the column with the use of TEAA (triethylamine acetate). The binding capability of nucleic acids to the column is proportional to the number of ion pairs formed between the negatively charged nucleic acids and the positively charged triethylammonium ions adsorbed to the stationary phase. The second step is to elute nucleic acids bound to the column with the use of Acetonitrile. Acetonitrile does so by competing with hydrophobic interactions between the column and TEAA. The heteroduplexes contain a disruption of the helical nature of the duplex due to the presence of incorrect base pairing at the site of the mutation, leading to fewer ion-paring bonds with the column. Thus, compared with homoduplexes, heteroduplexes will be eluted earlier, i.e., at a lower concentration of Acetonitrile (**Figure 3**). DHPLC is capable of detecting all combinations of nucleotide mismatches, independent of the location of the mismatch within the fragment. DNA fragments up to 2,000 base pairs can be analyzed with this method. Using appropriate analysis temperatures, sensitivity of detection approaches 100%, as demonstrated in

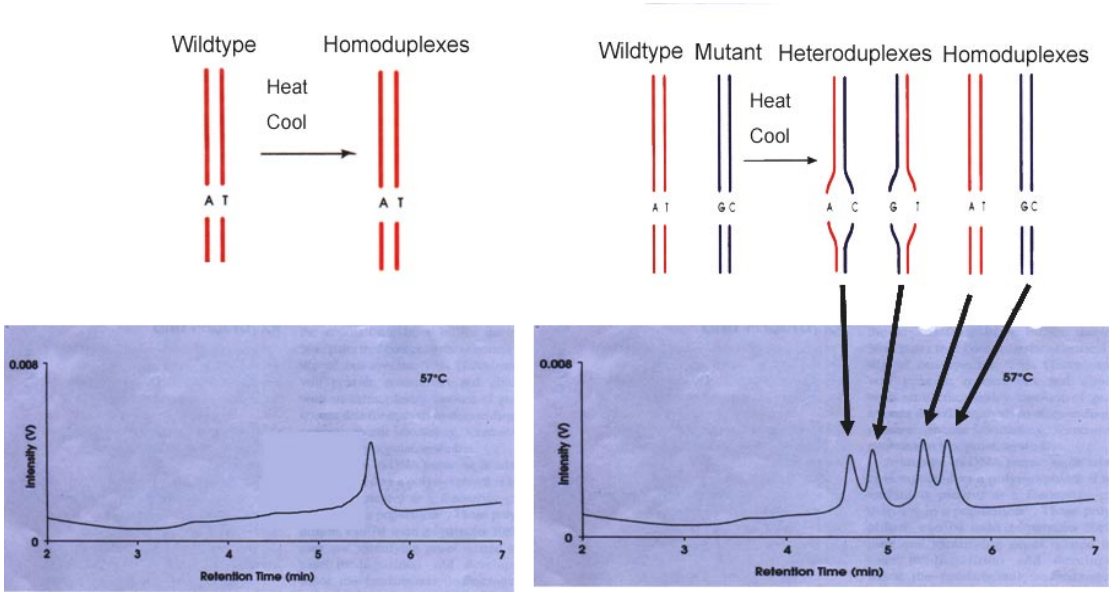

Figure 3. Schematic representation of DHPLC

many publications comparing it with SSCP and DGGE commonly used in mutation detection (Choy et al. 1999; Gross et al. 1999; Bunn et al. 2002). However, DHPLC can only reveal the presence of a mismatch, the location and chemical nature of which has to be established by sequencing.

In practical experimental procedure, DHPLC was carried out in a Transgenomic Wave DNA Fragment Analysis System, an automated instrument equipped with a DNASep column (Transgenomic, Crewe, United Kingdom). The PCR product can be loaded onto the instrument directly for analysis without any purification. The mobile phase gradient composed of TEAA and Acetonitrile and the running column temperature for optimal hetereoduplex separation were determined using the software WAVERMAKER 3. 4 provided with the instrument. Abnormal elution profiles were identified by visual inspection of the chromatogram.

3.2.2 Direct DNA sequencing (paper I, II, III)

The most popular method of DNA sequencing is known as chain termination sequencing, dideoxy sequencing, or Sanger sequencing after its inventor biochemist Frederick Sanger (Sanger et al. 1977). The key to the method is the use of modified bases called dideoxynucleotides. When a piece of DNA is being replicated and a dideoxynucleotide base is incorporated into the new chain, it stops the replication reaction at positions where a particular nucleotide is used, thus resulting in a series of related DNA fragments. The fragments are then size-separated by electrophoresis on a polyacrylamide gel. A variation of this method, cycle sequencing have been developed for automated sequencing machines. Cycle sequencing has the advantage of yielding large amounts of product from relatively little template and being well tolerant of sequencing templates with some impurities, for example, PCR products. In automatic sequencing, the dideoxynucleotides are tagged with different colored fluorescent dyes,

thus all four parallel base-specific reactions can occur in the same tube and be separated in the same lane on the gel. As each labeled DNA fragment passes a detector at the bottom of the gel, the color is recorded and the sequence is reconstructed from the pattern of colors representing each nucleotide in the sequence. In Papers I to III, genomic DNA was reamplified for all samples exhibiting abnormal DHPLC profiles with the same primer pairs and thermocycling conditions as used for DHPLC analysis. DNA sequencing of the purified PCR product was performed with an ABI PRISM 377 DNA Sequencer (PE Applied Biosystems, Foster City, CA) using BigDye™ Terminator V3.0 Ready Reaction Cycle Sequencing Kit (PE Applied Biosystems, Foster City, CA).

3.2.3 Pyrosequencing (Paper IV)

Pyrosequencing is a real-time bioluminometric technique for determination of nucleic acid sequence (Ronaghi et al. 1996). The pyrosequencing enzymatic cascade starts with the release of pyrophosphate (PPi) as a result of nucleotide incorporation by DNA polymerase. The released PPi is subsequently converted to ATP by ATP sulfurylase, which provides the energy for luciferase to oxidize luciferin and generate light seen as a peak in the pyrogram. Unincorporated nucleotides are degraded by apyrase prior to addition of the next nucleotide, allowing iterative addition of nucleotides. Since the added nucleotide is known, the sequence of the template can be determined (Fakhrai-Rad et al. 2002) (**Figure 4**).

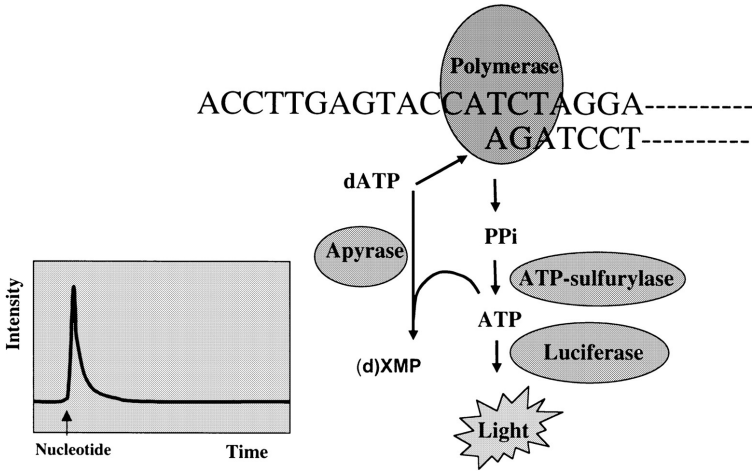

Figure 4. Schematic representation of the progress of the enzyme reaction in pyrosequencing.

One major application of pyrosequencing is genotyping of SNPs. For analysis of SNPs, the 3' end of a primer is designed to hybridize one or a few bases before the polymorphic position. In a single tube, all the different variations can be determined as the region is sequenced. A striking feather of pyrogram readouts for SNP analysis is the clear distinction between the various genotypes (**Figure 5**).

The experimental procedure in Paper IV for genotyping SNPs of interest is briefed as follows: a short fragment of 100-300 bp in size covering the target sequence was PCR amplified with one of the primer pairs biotinylaed at the 5' end. The biotinylated PCR

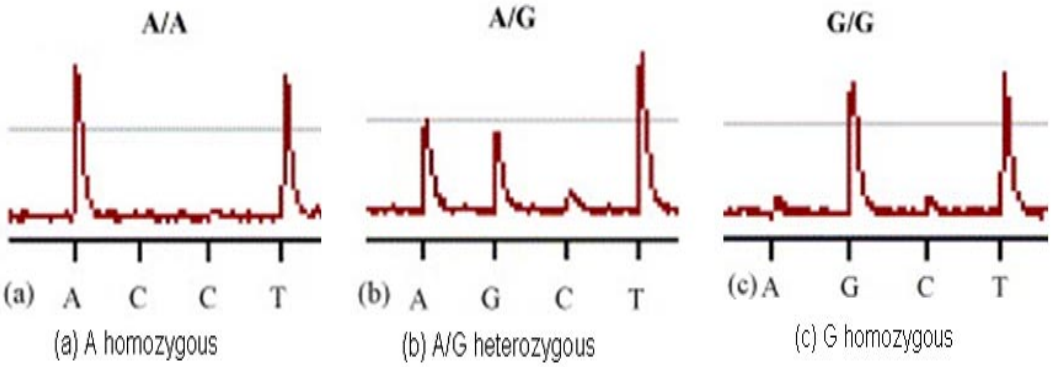

Figure 5. Three different SNPs genotyped by pyrosequencing

products were then purified, denatured, washed, and annealed to sequencing primers after immobilization to streptavidin coated sepharose beads. The generated samples were analyzed on a PSQ HS 96A instrument, using SNP reagents (Biotage, AB) according to standard protocols. The order of nucleotide dispensations was determined by the PSQ HS 96 SNP software (Biotage AB), which was also used for automatic assay evaluation and genotype scoring. This technology has the advantage of high-throughput, accuracy, and ease-of-use.

3.2.4 Microarray analysis (paper V)

DNA microarrays afford the ability to profile the expression of thousands of genes in a single experiment. It is being used more and more widely for the study of clinical medical diseases, particularly cancer, and has provided valuable insights into disease classification and understanding of the mechanisms of onset and progression. Two types of DNA microarrays have been utilized to profile gene expression, cDNA arrays and oligonucleotide arrays. For cDNA microarrays, probes of interest are produced by PCR amplification of cDNA inserts, which are then printed by high-speed robotics onto microscope slides. Recently, a preferable alternative is to use long oligonucleotide (50–70 mer) to be printed onto microscope slides, because the hybridization characteristics of oligonucleotides are generally better than PCR-derived cDNA clones. Each microarray measures two samples with different colors, so that after hybridization, the two colors are scanned separately and relative expression of each RNA molecule is determined by comparing intensities. Oligonucleotide arrays are available from several commercial companies, and Affymetrix® GeneChip probe arrays developed and produced by the Affymetrix Inc. (Santa Clara, CA) are the most widely used. Affymetrix® microarrays contain between 11 and 20 pairs of oligonucleotide probes for a target RNA, for which one of the pairs (perfect match probe, PM) is the reverse complement to an ideally unique 25-mer in the RNA and the other (mismatch probe, MM) contains a mutated middle base pair and serves as a measure of stray signal (Butte 2002).

In paper V, a group of hereditary and sporadic MSI CRCs, as well as two CRC cell lines were transcriptionally profiled using Affymetrix® Microarray HG-U95Av2 chips

that contain probe sets for analysis of 12,625 transcripts. The experimental procedure is illustrated in **Figure 6**.

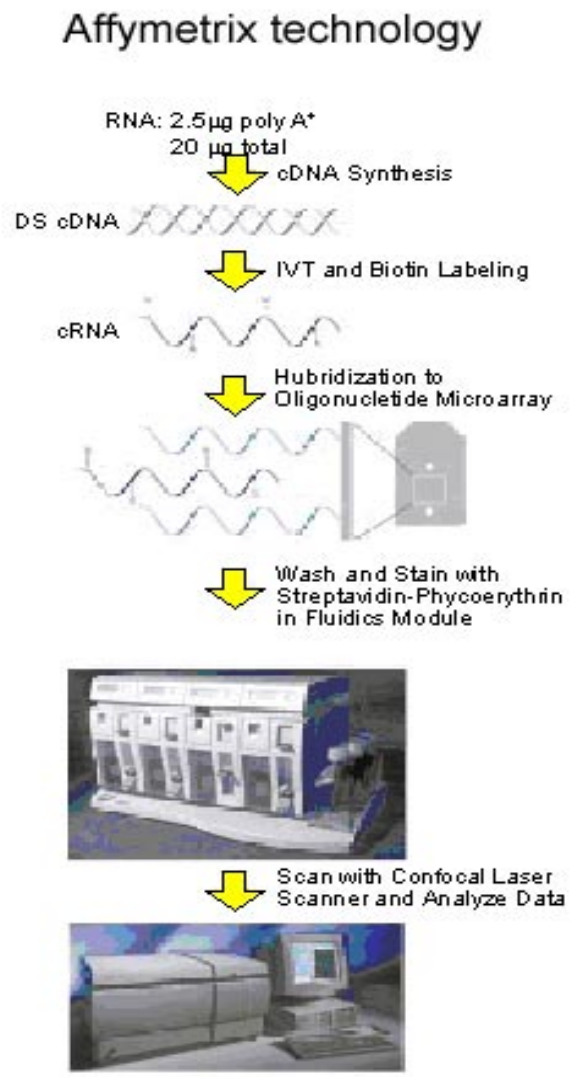

Figure 6. Schematic representation of the experimental procedure using Affymetrix® Microarray chips.

3.2.5 Northern blot (paper V)

Northern blot analysis provides a direct relative comparison of message abundance between samples on a single membrane. It is also the preferred method for determining transcript size and for detecting alternatively spliced transcripts. In this procedure, RNA fragments isolated by restriction endonuclease digestion of whole RNA samples are first fractionated by size via electrophoresis on an agarose gel under denaturing conditions. The size-separated RNA molecules are blotted onto a Nylon filter by capillary action with a high salt buffer that is passed through the gel from a buffer-saturated paper sheet. The blotted nucleic acids are then immobilized on the membrane by a fixation step using baking or UV crosslinking, followed by hybridization to a radiolabelled single-stranded DNA molecule, named probe. The signal corresponding to the probe hybridized to the restriction fragment containing the sequence of interest is

visualized by either photographic film or a phosphorimager (Ferrari et al. 1996). Northern blot was used in Paper V to confirm those differentially expressed genes between compared samples from microarray analysis. The gene glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) was chosen as a reference control, due to its little variation in expression across all the microarrays.

3.2.6 Statistical analysis (Papers I to V)

The Chi-square test was used to compare the difference in genotype or haplotype frequencies between cases and controls in the association studies. P value <0.05 was considered to be statistically significant (Paper I to IV). In Paper IV, haplotypes and LD values between SNPs of the *APC* gene were determined using the software Haploview V3.1.1 (http://www.broad.mit.edu/mpg/haploview/). In Paper V, a two-tailed unpaired student-t test with P value <0.05 was employed to help determine the list of genes that were regarded as differentially expressed between compared samples.

4 RESULTS AND DISCUSSION

4.1 PAPER I

The role of hMLH3 in familial colorectal cancer

hMLH3 is a newly identified DNA MMR gene, and suggested to be another candidate gene for HNPCC (Lipkin et al. 2000). To investigate a possible role of *hMLH3* in predisposition to CRC, we screened *hMLH3* for germ-line mutations in index patients from 70 families suggestive of a genetic predisposition for CRC. None of the families had FAP or AAPC. MSI status of all tumors had been established previously, and germline mutations in *hMLH1, hMSH2*, and *hMSH6* were excluded in all but two families (families 21 and 199) (Wahlberg et al. 1999; Huang et al. 2001).

In total, 1 frameshift variant and 11 missense variants were identified in 16 index patients (23%). The only frameshift variant identified, *hMLH3* 885delG, led to a truncation very early in the protein. In family 141 where the variant was found, the sequence change seemed to segregate quite well with the disease (**Figure 7**). Two cases with CRC, one woman with HNPCC-related endometrial cancer, and one of three unaffected relatives all aged above 75 were variant carriers. In addition, this mutation was not detected in any sporadic CRC cases or normal controls. A different frameshift mutation was reported in a Dutch HNPCC case (Wu et al. 2001), and this variant was absent in 188 Dutch normal controls, and 707 Finnish cancer-free controls (Hienonen et al. 2003). Taking these evidence and ours together, we concluded that *hMLH3* might cause CRC by having some high-penetrance pathogenic alleles, but with imcomplete penetrance. Since there are only two frameshift mutations with obvious pathogenic effect that have been identified by now worldwide, *hMLH3* is not likely to play a major role in predisposition for CRC. In addition, tumors in patient with frameshift mutations were clearly MSI negative (this study) or MSI-low (Wu et al. 2001), suggesting the mechanism in carcinogenesis not likely to involve a deficient MMR function. This is in agreement with the studies on yeast and mice that MLH3 has a minor role in MMR but a more important function in meiotic recombination (Lipkin et al. 2002; Santucci-Darmanin et al. 2002).

For all the other missense variants found, they did not segregate well with disease in families. Thus these variants are not likely to be full explanation for the familial aggregation of CRC. We noticed that these variants were mainly found in low risk families of TCR and OCR, prompting us to think that some of these variants might constitute low-penetrance alleles. In similarity to what has been suggested for breast cancer (Pharoah et al. 2002), a majority of CRCs would be caused by a number of low-penetrance alleles, which act on their own or interact with each other in a multiplicative or additive manner. The study of family 199 could lend some evidence to support this hypothesis (**Figure 8**). In this family, an *hMSH2* missense variant, E198G, was identified in all four sibs and their father in a previous study (Wahlberg et al. 1999).

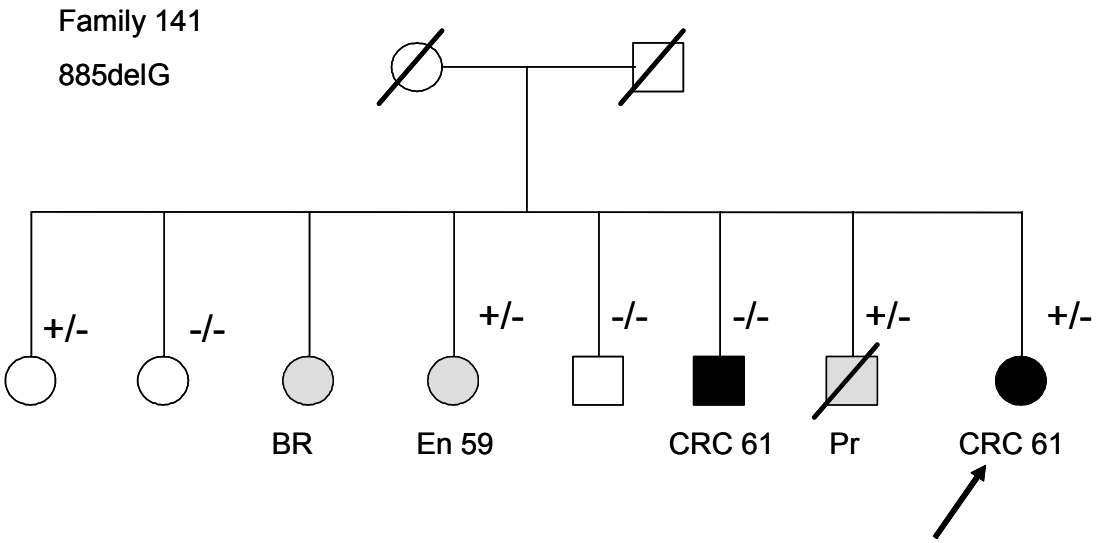

Figure 7. Arrow, index patients; black symbols, colorectal cancers; gray symbols, other cancers or adenomas; CRC, colorectal cancer; Pr, prostate cancer; Br, breast cancer; En, endometrial cancer; Numbers next to diagnosis denote age at onset; genotypes are to the top right of family member symbols. +, variant carrier; -, nonvariant carrier.

The fact that all tumors were MSI negative and the father displayed a somewhat weaker phenotype (a tubular adenoma with cancer at age 80), as well as the fact that missense variations in *hMSH2* were not often reported to be pathogenic, made us question the consequence of this *hMSH2* variant alone. The missing link could be the missense *hMLH3* variant, W1276R, found in all of the four sibs and inherited from their mother. It appears that variants in the two different genes have contributed together to the disease phenotype of the four sibs as the chance for the co-inheritance of two variant alleles from each parent to all of the four sibs is only 1 in 256. Two recent association studies (Hienonen et al. 2003; de Jong et al. 2004) have found no associations between *hMLH3* missense variants and increased CRC risk. However both studies lacked a statistical power to find any genuine association with the use of only a few hundred samples, as is the case with our small-scaled pilot association study.

4.2 PAPER II

Definition of candidate low risk APC alleles in a Swedish population

The *APC* gene has been found to underlie the syndrome of AAPC, a variant of FAP, characterized by a low number (<100) of colonic polyps and a later age of cancer onset (Knudsen et al. 2003). In extreme cases, AAPC patients can only have a few, even none of the polyps. One previous finding of ours that risk subjects from molecularly undetermined families of HCRC, TCR and OCR had more polyps and later age of cancer onset on average than the general population and HNPCC subjects (Lindgren et al. 2002) justifies the *APC* gene as a good susceptibility gene to test. We screened 91 subjects from families belonging to HCRC, TCR and OCR for germline mutations in the *APC* gene. Nearly all these subjects manifested with a family history of vertical transmission of cancer and the number of adenomas each patient had was much lower

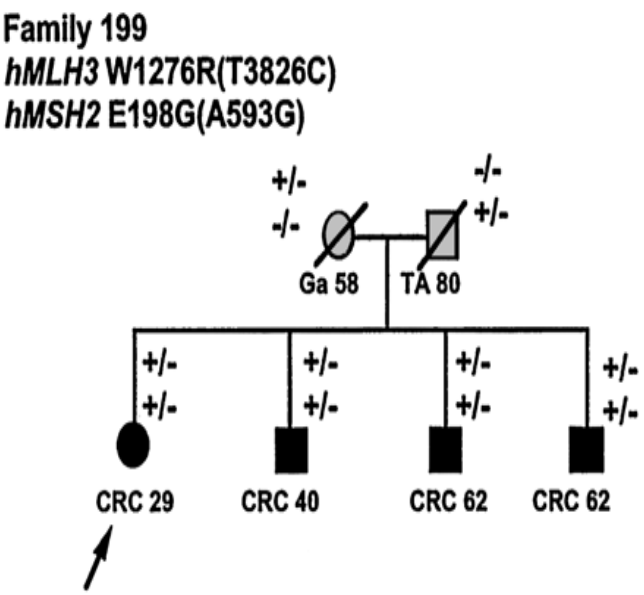

Figure 8. Symbols and abbreviations used are denoted as in Figure 7. Ga, gastric cancer. TA, tabular adenoma

than 10, thus precluding the involvement of mutations in the *MYH* gene. In addition, germline mutations in *hMLH1*, *hMSH2* and *hMSH6* were excluded previously for these subjects (Wahlberg et al. 1999; Huang et al. 2001).

In total, one clearly pathogenic nonsense mutation and 11 putative pathogenic variants (10 different missense variants and one 3' UTR variant) were detected in 20 out of 91 index patients screened (22%). The nonsense mutation, L334X, was found in a risk individual displaying 27 adenomas at age 35, whose mother and brother died from early onset CRC. L334X was never reported before, but located very close to the reported R332X mutation in the alternatively spliced region of exon 9. This exon has been demonstrated to harbor mutations related to AAPC. Thus this mutation most certainly constitutes a high-risk allele for CRC and strengthens the idea that the carrier of this mutation with the most number of adenomas of all screened subjects belongs to an AAPC family.

No other putative pathogenic variants found in this analysis showed clear evidence of being associated with cancer. They did not segregate well with the disease phenotype in their respective families and/or they were equally prevalent among sporadic cases and normal controls. Due to the limited number of samples in the association studies, the mildly increased CRC risk of some variants as low-penetrance alleles is not conclusive. The strongest evidence for *APC* to harbor low-penetrance alleles for CRC in the literature until now is the finding of I1307K, which has convincingly been shown to confer a 1.3-1.9 fold of increased CRC risk in Ashkenazim (Woodage et al. 1998). Since it is ethnic-specific, we did not find it in the Swedish population. However, we did find two of the variants in this study that have been related to a variation in CRC

risk, E1317Q (Frayling et al. 1998; Lamlum et al. 2000) and D1822V (Slattery et al. 2001), though based on weak evidence.

The 3' UTR 8636C>A found in this study seems to be most associated with an increased risk of CRC. This variant segregated quite well, although not perfectly, with adenoma in one family, but not in two high-risk families, indicating its possible role as a low-penetrance allele. In this case, a perfect segregation of risk alleles with disease would not be observed. Theoretically, the 5' or 3' UTR region harbor some sequence elements that can participate in control of posttranscriptional gene expression, so the 3' UTR 8636C>A may possibly confer a direct pathogenic effect. A previous finding reported one 5' UTR variant in a *APC*-mutation negative polyposis family, which segregated very well with the disease and was absent in a series of 146 Caucasian controls (Heinimann et al. 2001). In this study, 3' UTR 8636C>A was first found to be more prevalent in sporadic cases than in controls (13/182, 7.1% *vs.* 10/461, 2.2%; OR, 3.4, 95% CI, 1.5-7.9). However, using another 247 sporadic CRC cases and 476 normal controls from Linköping Hospital in an extended association study, the combined result was at the borderline of statistical significance (18/429, 4.2% *vs.* 22/927, 2.4%; OR, 1.8, 95% CI, 0.96 –3.40). Therefore, to define the role of this variant in the Swedish population, a large number of cases and controls will be needed for an association study with adequate statistical power. Since 3' UTR 8636C>A has such a low odds ratio and population frequency, at least five to six thousand samples are to be required for cases and controls, respectively (Pharoah et al. 2004). The collection of such large number of samples in a short time is obviously a difficult task. One solution is through the collaboration of multiple labs. In addition, with more and more functional information being disclosed for the *APC* gene, it is likely in the near future that we can evaluate the possible functional importance of the variant directly.

In this study, we also carried out the mutation screening of *APC* in sporadic CRC cases, normal controls, and family members of the index patients using the same sensitive technique, DHPLC, as for the familial cases. Only one additional missense variant was found, indicating that we probably have identified most of the common *APC* variants in the Swedish population. These variants can serve as a starting point for evaluating the contribution of the *APC* gene in CRC predisposition on the population level.

4.3 PAPER IV

Association of adenomatous polyposis coli gene polymorphisms with autism spectrum disorder (ASD)

While we first tested the association between the *APC* 3' UTR 8636C>A variant with the CRC risk in Paper II, we found that one of the four sets of controls differed from the other three by showing a higher frequency of the tested variant, similar to that found in the sporadic colorectal cancer cases. They were anonymous non-cancer patients, many with suspected mental retardation and/or autism spectrum disorder (ASD) undergoing genetic testing of Fragile-X Syndrome at the Department of Clinical

Genetics, Karolinska Hospital. A Gut-Brain axis has long been discussed in relation to ASD (Wakefield 2002) and, in a patient originally referred for autism, a deletion of the *APC* gene and occurrence of adenomatous polyposis and rectal cancer were reported (Barber et al. 1994). This prompted us to think that there might exist a possible association between the ASD symptoms and the *APC* variant overrepresented among the patients undergoing genetic testing for Fragile X syndrome.

We did three confirmatory association studies on the possible association between the *APC* 3' UTR 8636C>A variant and ASD. The first was a retrospective case-control study including 177 patients with ASD symptoms and 476 healthy Swedish controls. To exclude the influence of population stratification, we did additional two ethnic-specific association studies. One study used 39 ASD patients with Middle Eastern / African origin and 67 Iranian controls, and the second used 75 Swedish ASD patients and the same 476 Swedish controls as used in the first retrospective association study. All the three association studies showed a statistically significant association between the variant in allelic frequency and ASD (P=0.024, 0.01, and 0.014, respectively). In addition, the last two ethnic-specific association studies also showed a statistically significant association between the variant in genotypic frequency and ASD (P=0.001 and 0.015, respectively). To provide more statistical evidence, we next did the haplotype association analysis by genotyping additional four adjacent SNPs (rs2229992, rs42427, rs459552 and rs465899) in the *APC* gene. All these four SNPs were determined to be within a single haplotype block with D' values ranging from 0.86 to 0.98 using the software Haploview v3.1.1. Among all the common individual haplotypes (the frequencies > 0.05) constructed from the four SNPs, one haplotype TGAG was found to be statistically associated with ASD (P=0.0061).

In the present study, we have found preliminary statistical evidence for an ASD susceptibility region at the *APC* locus. This study is limited by the assessment of only a few SNPs covering a very small genomic region and the haplotype boundary constructed from the SNPs is not well defined, so it is needed to extend into the neighboring regions of the *APC* locus to do the fine mapping in future studies with the ultimate goal of identifying causative variants. At the moment it is impossible to know whether the causative variants are within the *APC* gene, or from a different gene in linkage disequilibrium with *APC*. However, from many respects, *APC* is obviously a good candidate ASD predisposing gene for a closer scrutiny. For example, APC can be involved in regulating structural aspects of synaptic contacts, the defects of which are postulated as an etiological mechanism for autism. It has been reported that mutations of neuroligins NLGN3 and NLGN4 encoding cell adhesion molecules present at the postsynaptic side of the synapse can cause autism (Jamain et al. 2003). Association studies have implicated one component of the Wnt pathway of which APC is also involved, WNT2, with autism (Wassink et al. 2001), though not replicated in follow up studies (McCoy et al. 2002; Li et al. 2004). The likely general and important role that APC has in the development of neurodevelopmental disorders including ASD is also

suggested by a recent study in which *APC* was found to be associated with susceptibility to schizophrenia (Cui et al. 2005).

4.4 PAPER III

Germline Mutations in the *MYH* Gene in Swedish Familial and Sporadic Colorectal Cancer

Biallelic germline mutations in the base excision repair gene *MYH* have been shown to predispose to a proportion of multiple colorectal adenomas and cancer (Al-Tassan et al. 2002; Sampson et al. 2003; Sieber et al. 2003). This conclusion was mainly based on the study of cases selected for polyposis or multiple adenomas, thus the role of the *MYH* germline mutations in familial colorectal cancer associated with few polyps still needs further clarification. To address this question, we conducted germline mutation screening of *MYH* in a series of 84 Swedish non- FAP and non-HNPCC families with increased risk for colorectal cancer based on the family history.

No obvious pathogenic monoallelic or biallelic *MYH* mutations, even the two most common pathogenic variants Y165C and G382D occurring in the Caucasians (Al-Tassan et al. 2002; Sampson et al. 2003; Sieber et al. 2003) were found in this patient cohort. We did find three missense variants, V22M, Q324H, and S501F, but they were the previously described polymorphisms in different studies. The absence of *MYH* mutations in our familial colorectal cancers with few polyps supports the established idea that pathogenic *MYH* mutations are mainly associated with the phenotype of multiple adenomas, with or without cancer. Though it has been observed that biallelic *MYH* carriers could have only a few or none of the adenomas (Enholm et al. 2003; Wang et al. 2004; Farrington et al. 2005), the occurrence of such cases is infrequent.

We also examined the prevalence of the two most common pathogenic *MYH* variants found in Caucasians, Y165C and G382D, in 450 Swedish sporadic CRCs and 480 Swedish normal controls. One incentive was to know if these two pathogenic variants might also exist in the Swedish population; the second incentive was to see if there is any evidence to suggest that carriers of heterozygous *MYH* mutations are at increased CRC risk, a long-existing puzzle. The heterozygous *MYH* Y165C was found in 0.91% of cases and 0.43% of controls; and the heterozygous *MYH* G382D in 0.45% of cases and 0.21% of controls. The heterozygosity frequencies of both variants in the Swedish population were similar to those reported in other Caucasian populations (Enholm et al. 2003; Wang et al. 2004; Farrington et al. 2005). The genotypic distributions of the two mutations were in Hardy-Weinberg equilibrium among controls, but not among cases. It appears that both the mutations were overrepresented in cases by around 2 fold as compared with controls, albeit not statistically significant due to limited sample size. Very recently, several association studies involving a large number of cases and controls have been published (Croitoru et al. 2004; Wang et al. 2004; Farrington et al. 2005). The heterozygous *MYH* mutations are shown to cause an increased CRC risk with a similar magnitude as we saw in this study. In addition, we found three novel

heterozygous sequence changes at the same position of amino acid 423 in exon 13 of the *MYH* gene, R423Q, R423P and R423R. The combined occurrence of sequence variations at this position was overrepresented in cases compared to controls (5/447 vs. 0/478, P=0.02, Chi-square test). Whether mutations at position 423 have pathological relevance needs to be further studied.

The finding of heterozygous pathogenic *MYH* mutations in the Swedish population makes us believe that *MYH* mutations could underlie a small proportion of familial colorectal cancers in Sweden. The negative finding of bi-allelic *MYH* mutations in Swedish cases may only reflect the fact that the mutation-screening studies missed those potential *MYH*-associated families. One reason is that true *MYH*-associated families may have a prevalence as low as FAP in the population (Enholm et al. 2003); the other reason is that *MYH*-associated families are easily mixed with sporadic CRCs, due to the autosomal recessive nature of the *MYH* mutations.

4.5 PAPER V

Expression profiles of hereditary and sporadic colorectal cancers and cell lines with microsatellite instability (MSI)

HNPCC and sporadic MSI CRC have been viewed as the familial and sporadic counterparts of the same underlying pathological process. Recently, however, there is increasing evidence to indicate that hereditary and sporadic MSI cancers could differ in terms of clinical, morphological, and molecular features (Hawkins et al. 2002; Jass 2004). Since microarray technology has shown many successes in tumour classification, we wondered if this technology could help differentiate hereditary and sporadic MSI CRCs based on genetic expression profiles, lending further evidence to the hypothesis that these two phenotypes develop through different genetic pathways. Therefore, expression profiling was performed on six HNPCCs and seven sporadic MSI CRCs using Affymetrix® HG-U95Av2 chips. The generated data were then analyzed with two unsupervised analytical methods, hierarchical clustering and principal component analysis (PCA).

Both unsupervised hierarchical clustering and PCA failed to distinguish the hereditary and sporadic MSI CRCs into distinct categories. Indeed many clinical, morphological, and molecular differences have been described in the literature. For example, compared with its hereditary counterpart, sporadic MSI CRCs are more age-related, more common among females, have a greater predilection for the proximal colon, have a low frequency of dedifferentiation or tumor budding at the advancing tumour margin, have a lower frequency of mutation of *APC*, *β-catenin* and *KRAS*, and have a much higher frequency of DNA methylation (Hawkins et al. 2002; Jass et al. 2002). However, none of these features could serve as an independent discriminator for the definite distinction between sporadic and hereditary MSI CRCs and might influence the profiles of genetic changes only to a mild degree (Jass 2004). So we thought that this might be a major

reason why unsupervised hierarchical clustering and PCA in this study failed to separate hereditary and sporadic MSI CRCs into two categories.

By now, the most clear-cut molecular difference that has emerged between hereditary and sporadic MSI CRCs is the mutation status of the *BRAF* gene. The mutation of *BRAF*, mainly V599E, is frequently present in sporadic MSI CRC, but not in HNPCC (Deng et al. 2004; Kambara et al. 2004). *BRAF* is a serine/threonine kinase of the RAF family, which participates in RAS-RAF-MEK-ERK-MAP kinase pathway, mediating cellular response to growth signals (Mercer et al. 2003). Studies have shown that *BRAF* is activated through somatic mutations and *BRAF* mutations are, to some extent, biologically similar to *KRAS* mutations in CRC pathogenesis because both occur at approximately the same stage of the adenoma-carcinoma sequence (Yuen et al. 2002). Thus, the difference in *BRAF* mutation status cannot be taken as an independent discriminator in the context of expression profiling to differentiate sporadic and hereditary MSI CRC.

The classification of tumors based on gene expression profiles is further complicated by tissue heterogeneity. In the case of colorectal cancer, cancer cells within a tumor are among themselves heterogeneous and are also contaminated by normal endothelium, different types of stromal cells, and inflammatory cells. So expression signals from the tumor cells can easily be masked by signals from other cell types. It would be interesting to see in future studies if the use of a homogeneous population of CRC cells, such as cells procured by laser-assisted microdissection, can better classify hereditary and sporadic MSI CRCs based on expression profiles.

However, both unsupervised hierarchical clustering and PCA analyses suggested that the samples seemed to be clustered depending on the prognosis potential of the tumors. The three samples, H-JAL, H-ML, and H-SOU, with the suggested bad prognosis had closest relationship in both analyses. Of note, H-ML and H-Jaal had much shorter survival time after surgery compared with other samples. The comparison between these three samples with the others in respect to expression profiles revealed a set of 296 genes that were differentially expressed by greater than two fold at lower 90% confidence bound. Among this set of deregulated genes, many had been defined to be related to the invasion and metastasis of CRC, such as *CXCL12, IGF1*, *Stratifin* and *NME1* (Le Roith et al. 1999; Bhatia et al. 2003; Garinis et al. 2003; Smith et al. 2005). To gain more evidence that some of differentially expressed genes between the three samples with suggested bad prognosis and the others might be specifically correlated with the metastatic potential of CRC with MSI, we also compared expression profiles of two isogenic cell lines with MSI, HCT116 and HCT116M2, with low and high metastatic capability, respectively. There were two genes, tripartite motif-containing 2 (*TRIM2*) and tight junction protein 3 (*TJP3*), that were found in common to be downregulated by around two fold in both tumors and cell lines with high metastatic

potential. However, only *TJP3* could be confirmed by Northern Blot analysis on cell lines to have a similar result as in the microarray analysis.

TJP3 is a novel member of the MAGUK protein family found at the cell tight junction (Haskins et al. 1998). The tight junction functions as an intercellular barrier and intramembrane diffusion fence (Wong et al. 1997), thereby constituting the first blockade for cancer cells to overcome in order to metastasize. Loss of tight junction molecules, *ZO-1* and *MUPP-1*, in breast cancer tissues has been associated with a poor prognosis in patients with breast cancer (Martin et al. 2004). It is a possibility that, like *ZO-1* and *MUPP-1*, the reduced expression of *TJP3* in CRC tissues with MSI could enhance the capability of CRC cells to invade surrounding normal tissues and stromal elements, which warrants future functional characterization.

5 CONCLUSIONS

We showed that *hMLH3* could harbor high-penetrance mutations causing familial CRC; however, the occurrence of mutations of this type is a rare event. Therefore, *hMLH3* is not important as a major predisposing factor in CRC. In addition, we have provided evidence to show that some missense variants in *hMLH3* could act as low-penetrance alleles that work in a multiplicative or additive manner with other low penetrance alleles.

We provided some preliminary evidence that there could be additional low-penetrance alleles in *APC* except the well-studied I1307K in Askenazi Jews, illustrated by the finding of 3' UTR 8636C>A. We also defined the mutation spectrum of *APC* in the Swedish population, providing a rather comprehensive list of candidate *APC* low-penetrance alleles for future characterization. In addition, the sequence variant 3' UTR 8636C>A and the *APC* gene locus it implicates were fortuitously found to be associated with ASD, and substantiated by follow-up association studies with statistically significant results.

We demonstrated that bi-allelic germline *MYH* mutations are not likely to account for familial CRCs in the absence of multiple adenomas. However the finding of the two most common pathogenic variants (Y165C and G382D) related to Caucasians in the Swedish population makes it very likely that *MYH* mutations could also underlie a small proportion of familial CRCs in Sweden. In addition, we also provided the evidence that heterozygotic pathogenic alleles in *MYH* could confer a slightly increased CRC risk.

The Unsupervised hierarchical clustering analysis and PCA based on expression profiles using Affymetrix® HG-U95Av2 chips did not distinguish between hereditary and sporadic MSI CRCs, meaning that the molecular difference between these two phenotypes, if any, is insignificant. In addition, we provided the preliminary evidence that one gene, *TJP3*, a novel member of the MAGUK protein family found at the cell tight junction, would be associated with the metastatic potential of CRCs with MSI.

6 ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to all those who gave me the possibility to complete this thesis, especially,

Prof. **Annika Lindblom,** my supervisor, for accepting me as a PhD student in your group, for your constant support, understanding, trust, encouragement and sometimes, forgiveness when I made mistakes, and for many visits to your house with warmth and an unforgettable journey to your beautiful hometown. Thank you so much for not only allowing me to share your wide knowledge in science, but also teaching me to always have an enthusiastic and optimistic attitude toward work and life. "Never give up" is the key I learned from you for becoming a good scientist. You have always been so much than a supervisor to me and will continue to be so.

Dr. **Jindong Chen**, for introducing me to such a wonderful research team at Karolinska Institute, for accompanying me and guiding me to get through the initial difficult adaptation period in the new lab when I just came to Sweden. You were my teacher when I was studying in a college in China. Luckily I still had you as my teacher even after I came to Sweden for a PhD degree. You have helped me move forward the most important steps in my whole life.

All present and former colleagues in Annika's group: **Barbro Werelius** for preparing everything in the lab and countless help in every respect, **Paula Maguire**, **Johanna Skoglund**, **Tatjana Djureinovic**, **Jana Vandrovcova**, **Barbara Kremeyer** and **Ulrika Eriksson** for always being ready to help with the issues inside or outside the lab and for the pleasant working atmosphere and laughter in the lab. Thank you so much, girls! I will cherish all the memorable time we enjoyed together in the lab, restaurants, and pubs forever. And **Moraima Zelada-Hedman**, **Brita Arver**, **Tao Li**u, **Sima Salahshor**, **Heléne Fischer**, **Liping Luo**, **Quan Du**, **Hongxu Liu**, **Annelie Liljegren**, **Sara Margolin**, **Berith Wejderot, and Nahla Mohamed** for good advice and help during my studies.

All of my collaborators in Sweden, Denmark and the USA for their help in the various projects and their valuable discussions and suggestions in my writing of manuscripts: **Magnus Nordenskjöld**, **Christopher Gillberg**, **XiaoFeng Sun**, **Christopher Workman**, **Gudrun Lindmark**, **Niklas Dahl**, **Ulf Kressner**, **Maibritt Giacobini**, **Britt-Marie Anderlid**, **Herik Anckarsäter**, **Davood Omrani**, **Thomas Jensen**, **Ulrik Lindforss**, **Lars Påhlman** and **Lennart Iselius**.

All members presently or previously working at CMM, especially, **Wen-onn Lui**, **Louise Bylund**, **Selim Sengül**, **Anna Hammarsjö**. Thank you all so much for all the agreeable talks, assistance, encouragements, and sharing your technical knowledge.

Everybody at the routine lab of clinical genetic testing, especially **Kerstin Holmberg** for karyotyping two cell lines for me.

Prof. **Magnus Nordenskjöld**, for creating an excellent working environment, for a good collaboration, and for many short instructive chats with humour in the kitchen or the corridor.

Dr. **Günter Weber** for always providing the answers to my various questions kindly and patiently, for solving the problems I had in the experiment, and for being as a marvellous organiser of our journal club from which I have benefited a lot.

Dr. **Fabio Sánchez**, for always having time to answer my questions, for the wonderful time we had in the jogging and football, and for the serious and relaxing talks on the topics of life, culture, civilization, future……Dr. **Keng-Ling Wallin**, for always being like my elder sister to care for me in both work and life.

Profs. **Anna Wedell**, **Catharina Larsson**, and **Agneta Nordenskjöld**, for always being helpful whenever needed.

Lennart Helleday for always calming down my crazy computer, which sometimes drove me mad.

Yvonne Cowan, **Gunilla Risberg**, **Majalena Granqvist, Delphi Post** and **Kerstin Florell** for professional help in administrative issues.

Ann-Christin Thelander, **Sigrid Sahlén**, **Margareta Tapper-Persson** for the care and kindness to me, for always having time to help.

Annika Lindblom, **Paula Maguire**, **Johanna Skoglund**, **Tatjana Djureinovic**, for their careful reading of my thesis and their good suggestions.

All my other friends who are present and have left Sweden: **Zhiping Zhang, Changjuan Pan, YunShang Piao, Danhua Pan, Fanqing Xia, Quan Du, Jue Wang, Fengqing Xiang, Weihua Dong, Jinjing Pei, Huixin Wang, Shuhua Ma, Yinghui Liu, Yang Yang, Shaozheng Wei, Yi Zhang, Hanjin Xie, Hui Gao, Wen Cai, Hao Mo, Hao Wu, Yongtao Xue, Mats Franzén, Qianren Jin, Xinjun Li, Yaofeng Zhao, Peipei Zhao, Sicheng Wen, Jing Ding, Wenpeng Niu, Qinyang Wu, Xiaojing Hao, Zheng Zhou, Yu Shi, Zhong He, Xupeng Ge, Guojun Chen, Hong Fang, Liping Luo**, **Ke Yang**, **Fangyuan Li**, **Qiuping Yuan**, **Jianming Wang**, **Biying Zheng**, **Chengyun Zheng**, **Wen-Hui Weng**, **ShaSha Lu**, **Chen Zhao**, **Liming Bu**, **Luosheng Li**, **Hong Jiao**, **Zhiguo Chen**, **Weihua Zhan**g, **Li Xu**, **Haixin Lei**, **Suya Liu**, **Yaping Jin,** and many more for whom I did not list their names here. Thank you all so much for friendship, help and the enjoyable time we spent together.

Prof. **Lishan Zhang**, my supervisor for my Master degree in China, for opening the research door to me, for introducing me to study abroad, for always treating me as her son, and for enormous inpiration and endless support.

My parents-in-law: **Zemin Zhang** and **Yuelan Fang**, for their selfless love and support.

My parents**, Jianguo Zhou and Lihua Lu,** and my younger brother, **Xiaolin Zhou**, for your true love, encouragement, and for always giving me confidence.爸爸, 妈妈, 和弟弟, 谢谢你们给我的爱和支持.

Lastly, my wife **Pu Zhang** and my lovely son, **Ziyue Zhou**, for your endless love, support and understanding, for always providing a peaceful harbor for me to anchor, for sharing happiness as well as sadness with me, and for the many moments I was deeply moved.

The work was financially supported by the Swedish Cancer Society, Ragnar and Torsten Söderbergs stiftelse, the Swedish Scientific Council, the Stockholm Cancer Society, and Nilson-Ehle Foundation.

7 REFERENCES

Aaltonen, L. A., Peltomaki, P., et al. (1993). "Clues to the pathogenesis of familial colorectal cancer." Science **260**(5109): 812-6.

Aarnio, M., Sankila, R., et al. (1999). "Cancer risk in mutation carriers of DNA-mismatch-repair genes." Int J Cancer **81**(2): 214-8.

Abdel-Rahman, W. M., Katsura, K., et al. (2001). "Spectral karyotyping suggests additional subsets of colorectal cancers characterized by pattern of chromosome rearrangement." Proc Natl Acad Sci U S A **98**(5): 2538-43.

Acharya, S., Foster, P. L., et al. (2003). "The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair." Mol Cell **12**(1): 233-46.

Akino, K., Toyota, M., et al. (2005). "The Ras effector RASSF2 is a novel tumor-suppressor gene in human colorectal cancer." Gastroenterology **129**(1): 156-69.

Allen, B. A. and Terdiman, J. P. (2003). "Hereditary polyposis syndromes and hereditary non-polyposis colorectal cancer." Best Pract Res Clin Gastroenterol **17**(2): 237-58.

Al-Tassan, N., Chmiel, N. H., et al. (2002). "Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors." Nat Genet **30**(2): 227-32.

Antoniou, A. C. and Easton, D. F. (2003). "Polygenic inheritance of breast cancer: Implications for design of association studies." Genet Epidemiol **25**(3): 190-202.

Bacanu, S. A., Devlin, B., et al. (2000). "The power of genomic control." Am J Hum Genet **66**(6): 1933-44.

Baker, S. J., Fearon, E. R., et al. (1989). "Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas." Science **244**(4901): 217-21.

Baker, S. J., Preisinger, A. C., et al. (1990). "p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis." Cancer Res **50**(23): 7717-22.

Balmain, A., Gray, J., et al. (2003). "The genetics and genomics of cancer." Nat Genet **33 Suppl**: 238-44.

Barber, J. C., Ellis, K. H., et al. (1994). "Adenomatous polyposis coli and a cytogenetic deletion of chromosome 5 resulting from a maternal intrachromosomal insertion." J Med Genet **31**(4): 312-6.

Barcellos-Hoff, M. H. and Ravani, S. A. (2000). "Irradiated mammary gland stroma promotes the expression of tumorigenic potential by unirradiated epithelial cells." Cancer Res **60**(5): 1254-60.

Bardelli, A., Parsons, D. W., et al. (2003). "Mutational analysis of the tyrosine kinome in colorectal cancers." Science **300**(5621): 949.

Barnby, G., Abbott, A., et al. (2005). "Candidate-gene screening and association analysis at the autism-susceptibility locus on chromosome 16p: evidence of association at GRIN2A and ABAT." Am J Hum Genet **76**(6): 950-66.

Baylin, S. B. and Herman, J. G. (2000). "DNA hypermethylation in tumorigenesis: epigenetics joins genetics." Trends Genet **16**(4): 168-74.

Bell, D. W., Varley, J. M., et al. (1999). "Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome." Science **286**(5449): 2528-31.

Bernards, R. and Weinberg, R. A. (2002). "A progression puzzle." Nature **418**(6900): 823.

Beroud, C. and Soussi, T. (1996). "APC gene: database of germline and somatic mutations in human tumors and cell lines." Nucleic Acids Res **24**(1): 121-4.

Bertucci, F., Salas, S., et al. (2004). "Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters." Oncogene **23**(7): 1377-91.

Bhatia, K., Siraj, A. K., et al. (2003). "The tumor suppressor gene 14-3-3 sigma is commonly methylated in normal and malignant lymphoid cells." Cancer Epidemiol Biomarkers Prev **12**(2): 165-9.

Bhattacharyya, N. P., Skandalis, A., et al. (1994). "Mutator phenotypes in human colorectal carcinoma cell lines." Proc Natl Acad Sci U S A **91**(14): 6319-23.

Blair, N. P. and Trempe, C. L. (1980). "Hypertrophy of the retinal pigment epithelium associated with Gardner's syndrome." Am J Ophthalmol **90**(5): 661-7.

Boardman, L. A., Thibodeau, S. N., et al. (1998). "Increased risk for cancer in patients with the Peutz-Jeghers syndrome." Ann Intern Med **128**(11): 896-9.

Boland, C. R. and Ricciardiello, L. (1999). "How many mutations does it take to make a tumor?" Proc Natl Acad Sci U S A **96**(26): 14675-7.

Boland, C. R., Thibodeau, S. N., et al. (1998). "A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer." Cancer Res **58**(22): 5248-57.

Bos, J. L. (1988). "The ras gene family and human carcinogenesis." Mutat Res **195**(3): 255-71.

Bos, J. L., Fearon, E. R., et al. (1987). "Prevalence of ras gene mutations in human colorectal cancers." Nature **327**(6120): 293-7.
Botstein, D. and Risch, N. (2003). "Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease." Nat Genet **33 Suppl**: 228-37.
Boultwood, J. (2001). "Ataxia telangiectasia gene mutations in leukaemia and lymphoma." J Clin Pathol **54**(7): 512-6.
Boveri, T. (1914). "Zur Frage der Entstehung Maligner Tumoren." Jena: Verlag Gustav Fisher.
Bulow, S., Faurschou Nielsen, T., et al. (1996). "The incidence rate of familial adenomatous polyposis. Results from the Danish Polyposis Register." Int J Colorectal Dis **11**(2): 88-91.
Bunn, C. F., Lintott, C. J., et al. (2002). "Comparison of SSCP and DHPLC for the detection of LDLR mutations in a New Zealand cohort." Hum Mutat **19**(3): 311.
Butte, A. (2002). "The use and analysis of microarray data." Nat Rev Drug Discov **1**(12): 951-60.
Cannon-Albright, L. A., Skolnick, M. H., et al. (1988). "Common inheritance of susceptibility to colonic adenomatous polyps and associated colorectal cancers." N Engl J Med **319**(9): 533-7.
Cardon, L. R. and Bell, J. I. (2001). "Association study designs for complex diseases." Nat Rev Genet **2**(2): 91-9.
Cardon, L. R. and Palmer, L. J. (2003). "Population stratification and spurious allelic association." Lancet **361**(9357): 598-604.
Caspari, R., Olschwang, S., et al. (1995). "Familial adenomatous polyposis: desmoid tumours and lack of ophthalmic lesions (CHRPE) associated with APC mutations beyond codon 1444." Hum Mol Genet **4**(3): 337-40.
Chambers, A. F. (1992). "Mechanisms of oncogene-mediated alterations in metastatic ability." Biochem Cell Biol **70**(9): 817-21.
Chambers, A. F., Groom, A. C., et al. (2002). "Dissemination and growth of cancer cells in metastatic sites." Nat Rev Cancer **2**(8): 563-72.
Charames, G. S., Millar, A. L., et al. (2000). "Do MSH6 mutations contribute to double primary cancers of the colorectum and endometrium?" Hum Genet **107**(6): 623-9.
Chen, J., Hunter, D. J., et al. (2003). "Polymorphism in the thymidylate synthase promoter enhancer region modifies the risk and survival of colorectal cancer." Cancer Epidemiol Biomarkers Prev **12**(10): 958-62.
Chow, E., Thirlwell, C., et al. (2004). "Colorectal cancer and inherited mutations in base-excision repair." Lancet Oncol **5**(10): 600-6.
Choy, Y. S., Dabora, S. L., et al. (1999). "Superiority of denaturing high performance liquid chromatography over single-stranded conformation and conformation-sensitive gel electrophoresis for mutation detection in TSC2." Ann Hum Genet **63 (Pt 5)**: 383-91.
Chung, D. C. (2000). "The genetic basis of colorectal cancer: insights into critical pathways of tumorigenesis." Gastroenterology **119**(3): 854-65.
Cottrell, S., Bicknell, D., et al. (1992). "Molecular analysis of APC mutations in familial adenomatous polyposis and sporadic colon carcinomas." Lancet **340**(8820): 626-30.
Couzin, J. (2002). "Genomics. New mapping project splits the community." Science **296**(5572): 1391-3.
Couzin, J. (2002). "Human genome. HapMap launched with pledges of $100 million." Science **298**(5595): 941-2.
Croitoru, M. E., Cleary, S. P., et al. (2004). "Association between biallelic and monoallelic germline MYH gene mutations and colorectal cancer risk." J Natl Cancer Inst **96**(21): 1631-4.
Cui, D. H., Jiang, K. D., et al. (2005). "The tumor suppressor adenomatous polyposis coli gene is associated with susceptibility to schizophrenia." Mol Psychiatry **10**(7): 669-77.
Cui, H., Cruz-Correa, M., et al. (2003). "Loss of IGF2 imprinting: a potential marker of colorectal cancer risk." Science **299**(5613): 1753-5.
Cui, H., Horon, I. L., et al. (1998). "Loss of imprinting in normal tissue of colorectal cancer patients with microsatellite instability." Nat Med **4**(11): 1276-80.
Cunningham, J. M., Christensen, E. R., et al. (1998). "Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability." Cancer Res **58**(15): 3455-60.
da Costa, L. T., Liu, B., et al. (1995). "Polymerase delta variants in RER colorectal tumours." Nat Genet **9**(1): 10-1.
Das Gupta, R. and Kolodner, R. D. (2000). "Novel dominant mutations in Saccharomyces cerevisiae MSH6." Nat Genet **24**(1): 53-6.
de Jong, M. M., Hofstra, R. M., et al. (2004). "No association between two MLH3 variants (S845G and P844L)and colorectal cancer risk." Cancer Genet Cytogenet **152**(1): 70-1.
de Jong, M. M., Nolte, I. M., et al. (2002). "Low-penetrance genes and their involvement in colorectal cancer susceptibility." Cancer Epidemiol Biomarkers Prev **11**(11): 1332-52.

de Jong, M. M., Nolte, I. M., et al. (2005). "Colorectal cancer and the CHEK2 1100delC mutation." Genes Chromosomes Cancer **43**(4): 377-82.
de la Chapelle, A. (2004). "Genetic predisposition to colorectal cancer." Nat Rev Cancer **4**(10): 769-80.
de Wind, N., Dekker, M., et al. (1995). "Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer." Cell **82**(2): 321-30.
De Vos, M., Hayward, B. E., et al. (2004). "Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome." Am J Hum Genet **74**(5): 954-64.
Delattre, O., Olschwang, S., et al. (1989). "Multiple genetic alterations in distal and proximal colorectal cancer." Lancet **2**(8659): 353-6.
Delatycki, M. B., Danks, A., et al. (2003). "De novo germline PTEN mutation in a man with Lhermitte-Duclos disease which arose on the paternal chromosome and was transmitted to his child with polydactyly and Wormian bones." J Med Genet **40**(8): e92.
Deng, G., Bell, I., et al. (2004). "BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer." Clin Cancer Res **10**(1 Pt 1): 191-5.
Di Fiore, F., Charbonnier, F., et al. (2004). "Screening for genomic rearrangements of the MMR genes must be included in the routine diagnosis of HNPCC." J Med Genet **41**(1): 18-20.
Dietmaier, W., Wallinger, S., et al. (1997). "Diagnostic microsatellite instability: definition and correlation with mismatch repair protein expression." Cancer Res **57**(21): 4749-56.
Dong, C. and Hemminki, K. (2001). "Modification of cancer risks in offspring by sibling and parental cancers from 2,112,616 nuclear families." Int J Cancer **92**(1): 144-50.
Drucker, L., Shpilberg, O., et al. (2000). "Adenomatous polyposis coli I1307K mutation in Jewish patients with different ethnicity: prevalence and phenotype." Cancer **88**(4): 755-60.
Duckett, D. R., Bronstein, S. M., et al. (1999). "hMutSalpha- and hMutLalpha-dependent phosphorylation of p53 in response to DNA methylator damage." Proc Natl Acad Sci U S A **96**(22): 12384-8.
Duval, A., Gayet, J., et al. (1999). "Frequent frameshift mutations of the TCF-4 gene in colorectal cancers with microsatellite instability." Cancer Res **59**(17): 4213-5.
Duval, A. and Hamelin, R. (2002). "Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability." Cancer Res **62**(9): 2447-54.
Enholm, S., Hienonen, T., 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): 827-32.
Eshleman, J. R., Casey, G., et al. (1998). "Chromosome number and structure both are markedly stable in RER colorectal cancers and are not destabilized by mutation of p53." Oncogene **17**(6): 719-25.
Eshleman, J. R., Lang, E. Z., et al. (1995). "Increased mutation rate at the hprt locus accompanies microsatellite instability in colon cancer." Oncogene **10**(1): 33-7.
Esteller, M., Avizienyte, E., et al. (2000). "Epigenetic inactivation of LKB1 in primary tumors associated with the Peutz-Jeghers syndrome." Oncogene **19**(1): 164-8.
Esteller, M. and Herman, J. G. (2002). "Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours." J Pathol **196**(1): 1-7.
Esteller, M., Sparks, A., et al. (2000). "Analysis of adenomatous polyposis coli promoter hypermethylation in human cancer." Cancer Res **60**(16): 4366-71.
Ewart-Toland, A., Briassouli, P., et al. (2003). "Identification of Stk6/STK15 as a candidate low-penetrance tumor-susceptibility gene in mouse and human." Nat Genet **34**(4): 403-12.
Ewing, J. (1928). Neoplastic Diseases edn6, W.B. Saunders, Philadelphia.
Fakhrai-Rad, H., Pourmand, N., et al. (2002). "Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms." Hum Mutat **19**(5): 479-85.
Farrington, S. M., Tenesa, A., et al. (2005). "Germline susceptibility to colorectal cancer due to base-excision repair gene defects." Am J Hum Genet **77**(1): 112-9.
Fearnhead, N. S., Britton, M. P., et al. (2001). "The ABC of APC." Hum Mol Genet **10**(7): 721-33.
Fearnhead, N. S., Wilding, J. L., et al. (2002). "Genetics of colorectal cancer: hereditary aspects and overview of colorectal tumorigenesis." Br Med Bull **64**: 27-43.
Fearon, E. R. and Vogelstein, B. (1990). "A genetic model for colorectal tumorigenesis." Cell **61**(5): 759-67.
Feinberg, A. P. (2004). "The epigenetics of cancer etiology." Semin Cancer Biol **14**(6): 427-32.
Feinberg, A. P., Gehrke, C. W., et al. (1988). "Reduced genomic 5-methylcytosine content in human colonic neoplasia." Cancer Res **48**(5): 1159-61.
Feinberg, A. P. and Tycko, B. (2004). "The history of cancer epigenetics." Nat Rev Cancer **4**(2): 143-53.
Feinberg, A. P. and Vogelstein, B. (1983). "Hypomethylation of ras oncogenes in primary human cancers." Biochem Biophys Res Commun **111**(1): 47-54.

Feldser, D. M., Hackett, J. A., et al. (2003). "Telomere dysfunction and the initiation of genome instability." Nat Rev Cancer **3**(8): 623-7.
Ferrari, M., Cremonesi, L., et al. (1996). "Molecular diagnosis of genetic diseases." Clin Biochem **29**(3): 201-8.
Fidler, I. J. (2002). "The organ microenvironment and cancer metastasis." Differentiation **70**(9-10): 498-505.
Fidler, I. J. (2003). "The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited." Nat Rev Cancer **3**(6): 453-8.
Fodde, R. (2002). "The APC gene in colorectal cancer." Eur J Cancer **38**(7): 867-71.
Fodde, R., Kuipers, J., et al. (2001). "Mutations in the APC tumour suppressor gene cause chromosomal instability." Nat Cell Biol **3**(4): 433-8.
Frayling, I. M., Beck, N. E., et al. (1998). "The APC variants I1307K and E1317Q are associated with colorectal tumors, but not always with a family history." Proc Natl Acad Sci U S A **95**(18): 10722-7.
Frazier, M. L., Xi, L., et al. (2003). "Association of the CpG island methylator phenotype with family history of cancer in patients with colorectal cancer." Cancer Res **63**(16): 4805-8.
Fuchs, C. S., Giovannucci, E. L., et al. (1994). "A prospective study of family history and the risk of colorectal cancer." N Engl J Med **331**(25): 1669-74.
Gabriel, S. B., Schaffner, S. F., et al. (2002). "The structure of haplotype blocks in the human genome." Science **296**(5576): 2225-9.
Gailani, M. R., Stahle-Backdahl, M., et al. (1996). "The role of the human homologue of Drosophila patched in sporadic basal cell carcinomas." Nat Genet **14**(1): 78-81.
Gardner, E. J. and Richards, R. C. (1953). "Multiple cutaneous and subcutaneous lesions occurring simultaneously with hereditary polyposis and osteomatosis." Am J Hum Genet **5**(2): 139-47.
Garinis, G. A., Manolis, E. N., et al. (2003). "High frequency of concomitant nm23-H1 and E-cadherin transcriptional inactivation in primary non-inheriting colorectal carcinomas." J Mol Med **81**(4): 256-63.
Gildea, J. J., Seraj, M. J., et al. (2002). "RhoGDI2 is an invasion and metastasis suppressor gene in human cancer." Cancer Res **62**(22): 6418-23.
Gismondi, V., Meta, M., et al. (2004). "Prevalence of the Y165C, G382D and 1395delGGA germline mutations of the MYH gene in Italian patients with adenomatous polyposis coli and colorectal adenomas." Int J Cancer **109**(5): 680-4.
Goelz, S. E., Vogelstein, B., et al. (1985). "Hypomethylation of DNA from benign and malignant human colon neoplasms." Science **228**(4696): 187-90.
Gorlin, R. J., Cohen, M. M., Jr., et al. (1992). "Bannayan-Riley-Ruvalcaba syndrome." Am J Med Genet **44**(3): 307-14.
Grady, W. M. (2003). "Genetic testing for high-risk colon cancer patients." Gastroenterology **124**(6): 1574-94.
Grieu, F., Malaney, S., et al. (2003). "Lack of association between CCND1 G870A polymorphism and the risk of breast and colorectal cancers." Anticancer Res **23**(5b): 4257-9.
Groden, J., Thliveris, A., et al. (1991). "Identification and characterization of the familial adenomatous polyposis coli gene." Cell **66**(3): 589-600.
Gross, E., Arnold, N., et al. (1999). "A comparison of BRCA1 mutation analysis by direct sequencing, SSCP and DHPLC." Hum Genet **105**(1-2): 72-8.
Gruber, S. B., Ellis, N. A., et al. (2002). "BLM heterozygosity and the risk of colorectal cancer." Science **297**(5589): 2013.
Gryfe, R., Di Nicola, N., et al. (1998). "Somatic instability of the APC I1307K allele in colorectal neoplasia." Cancer Res **58**(18): 4040-3.
Hahn, S. A., Schutte, M., et al. (1996). "DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1." Science **271**(5247): 350-3.
Hamilton, S. R., Liu, B., et al. (1995). "The molecular basis of Turcot's syndrome." N Engl J Med **332**(13): 839-47.
Han, H. J., Yanagisawa, A., et al. (1993). "Genetic instability in pancreatic cancer and poorly differentiated type of gastric cancer." Cancer Res **53**(21): 5087-9.
Hanahan, D. and Weinberg, R. A. (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.
Hanssen, A. M. and Fryns, J. P. (1995). "Cowden syndrome." J Med Genet **32**(2): 117-9.
Hashiguchi, K., Zhang, Q. M., et al. (2002). "Characterization of 2-hydroxyadenine DNA glycosylase activity of Escherichia coli MutY protein." Int J Radiat Biol **78**(7): 585-92.
Haskins, J., Gu, L., et al. (1998). "ZO-3, a novel member of the MAGUK protein family found at the tight junction, interacts with ZO-1 and occludin." J Cell Biol **141**(1): 199-208.
Hawkins, N. J., Bariol, C., et al. (2002). "The serrated neoplasia pathway." Pathology **34**(6): 548-55.

Hawn, M. T., Umar, A., et al. (1995). "Evidence for a connection between the mismatch repair system and the G2 cell cycle checkpoint." Cancer Res **55**(17): 3721-5.
Hegde, P., Qi, R., et al. (2001). "Identification of tumor markers in models of human colorectal cancer using a 19,200-element complementary DNA microarray." Cancer Res **61**(21): 7792-7.
Heinen, C. D., Goss, K. H., et al. (2002). "The APC tumor suppressor controls entry into S-phase through its ability to regulate the cyclin D/RB pathway." Gastroenterology **123**(3): 751-63.
Heinimann, K., Thompson, A., et al. (2001). "Nontruncating APC germ-line mutations and mismatch repair deficiency play a minor role in APC mutation-negative polyposis." Cancer Res **61**(20): 7616-22.
Hemminki, K. and Chen, B. (2005). "Familial risks for colorectal cancer show evidence on recessive inheritance." Int J Cancer **115**(5): 835-8.
Hemminki, K. and Czene, K. (2002). "Attributable risks of familial cancer from the Family-Cancer Database." Cancer Epidemiol Biomarkers Prev **11**(12): 1638-44.
Hendriks, Y., Franken, P., et al. (2003). "Conventional and tissue microarray immunohistochemical expression analysis of mismatch repair in hereditary colorectal tumors." Am J Pathol **162**(2): 469-77.
Hendriks, Y. M., Wagner, A., et al. (2004). "Cancer risk in hereditary nonpolyposis colorectal cancer due to MSH6 mutations: impact on counseling and surveillance." Gastroenterology **127**(1): 17-25.
Herman, J. G., Merlo, A., et al. (1995). "Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers." Cancer Res **55**(20): 4525-30.
Hernandez-Blazquez, F. J., Habib, M., et al. (2000). "Evaluation of global DNA hypomethylation in human colon cancer tissues by immunohistochemistry and image analysis." Gut **47**(5): 689-93.
Hesson, L. B., Wilson, R., et al. (2005). "CpG island promoter hypermethylation of a novel Ras-effector gene RASSF2A is an early event in colon carcinogenesis and correlates inversely with K-ras mutations." Oncogene **24**(24): 3987-94.
Hickman, M. J. and Samson, L. D. (1999). "Role of DNA mismatch repair and p53 in signaling induction of apoptosis by alkylating agents." Proc Natl Acad Sci U S A **96**(19): 10764-9.
Hienonen, T., Laiho, P., et al. (2003). "Little evidence for involvement of MLH3 in colorectal cancer predisposition." Int J Cancer **106**(2): 292-6.
Hiltunen, M. O., Alhonen, L., et al. (1997). "Hypermethylation of the APC (adenomatous polyposis coli) gene promoter region in human colorectal carcinoma." Int J Cancer **70**(6): 644-8.
Hinoda, Y., Okayama, N., et al. (2002). "Association of functional polymorphisms of matrix metalloproteinase (MMP)-1 and MMP-3 genes with colorectal cancer." Int J Cancer **102**(5): 526-9.
Hirschhorn, J. N. and Daly, M. J. (2005). "Genome-wide association studies for common diseases and complex traits." Nat Rev Genet **6**(2): 95-108.
Hirschhorn, J. N., Lohmueller, K., et al. (2002). "A comprehensive review of genetic association studies." Genet Med **4**(2): 45-61.
Hoang, J. M., Cottu, P. H., et al. (1997). "BAT-26, an indicator of the replication error phenotype in colorectal cancers and cell lines." Cancer Res **57**(2): 300-3.
Hoeijmakers, J. H. (2001). "Genome maintenance mechanisms for preventing cancer." Nature **411**(6835): 366-74.
Hollstein, M., Sidransky, D., et al. (1991). "p53 mutations in human cancers." Science **253**(5015): 49-53.
Houlston, R. S., Murday, V., et al. (1990). "Screening and genetic counselling for relatives of patients with colorectal cancer in a family cancer clinic." Bmj **301**(6748): 366-8.
Houlston, R. S. and Peto, J. (2004). "The search for low-penetrance cancer susceptibility alleles." Oncogene **23**(38): 6471-6.
Houlston, R. S. and Tomlinson, I. P. (2001). "Polymorphisms and colorectal tumor risk." Gastroenterology **121**(2): 282-301.
Howe, J. R., Roth, S., et al. (1998). "Mutations in the SMAD4/DPC4 gene in juvenile polyposis." Science **280**(5366): 1086-8.
Huang, J., Kuismanen, S. A., et al. (2001). "MSH6 and MSH3 are rarely involved in genetic predisposition to nonpolypotic colon cancer." Cancer Res **61**(4): 1619-23.
Huebner, R. J. and Todaro, G. J. (1969). "Oncogenes of RNA tumor viruses as determinants of cancer." Proc Natl Acad Sci U S A **64**(3): 1087-94.
Iino, H., Jass, J. R., et al. (1999). "DNA microsatellite instability in hyperplastic polyps, serrated adenomas, and mixed polyps: a mild mutator pathway for colorectal cancer?" J Clin Pathol **52**(1): 5-9.
Ionov, Y., Peinado, M. A., et al. (1993). "Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis." Nature **363**(6429): 558-61.

Iwasa, Y., Michor, F., et al. (2005). "Population genetics of tumor suppressor genes." J Theor Biol **233**(1): 15-23.
Jacoby, R. F., Schlack, S., et al. (1997). "A juvenile polyposis tumor suppressor locus at 10q22 is deleted from nonepithelial cells in the lamina propria." Gastroenterology **112**(4): 1398-403.
Jagmohan-Changur, S., Poikonen, T., et al. (2003). "EXO1 variants occur commonly in normal population: evidence against a role in hereditary nonpolyposis colorectal cancer." Cancer Res **63**(1): 154-8.
Jallepalli, P. V. and Lengauer, C. (2001). "Chromosome segregation and cancer: cutting through the mystery." Nat Rev Cancer **1**(2): 109-17.
Jamain, S., Quach, H., et al. (2003). "Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism." Nat Genet **34**(1): 27-9.
Jarvinen, H. J. (2004). "Hereditary cancer: guidelines in clinical practice. Colorectal cancer genetics." Ann Oncol **15 Suppl 4**: iv127-31.
Jass, J. R. (2004). "HNPCC and sporadic MSI-H colorectal cancer: a review of the morphological similarities and differences." Fam Cancer **3**(2): 93-100.
Jass, J. R., Do, K. A., et al. (1998). "Morphology of sporadic colorectal cancer with DNA replication errors." Gut **42**(5): 673-9.
Jass, J. R., Whitehall, V. L., et al. (2002). "Emerging concepts in colorectal neoplasia." Gastroenterology **123**(3): 862-76.
Jass, J. R., Young, J., et al. (2000). "Hyperplastic polyps and DNA microsatellite unstable cancers of the colorectum." Histopathology **37**(4): 295-301.
Jen, J., Powell, S. M., et al. (1994). "Molecular determinants of dysplasia in colorectal lesions." Cancer Res **54**(21): 5523-6.
Ji, L., Nishizaki, M., et al. (2002). "Expression of several genes in the human chromosome 3p21.3 homozygous deletion region by an adenovirus vector results in tumor suppressor activities in vitro and in vivo." Cancer Res **62**(9): 2715-20.
Jiricny, J. (2000). "Mediating mismatch repair." Nat Genet **24**(1): 6-8.
Johns, L. E. and Houlston, R. S. (2001). "A systematic review and meta-analysis of familial colorectal cancer risk." Am J Gastroenterol **96**(10): 2992-3003.
Johnson, G. C., Esposito, L., et al. (2001). "Haplotype tagging for the identification of common disease genes." Nat Genet **29**(2): 233-7.
Jones, P. A. and Laird, P. W. (1999). "Cancer epigenetics comes of age." Nat Genet **21**(2): 163-7.
Jones, S., Emmerson, P., et al. (2002). "Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G:C-->T:A mutations." Hum Mol Genet **11**(23): 2961-7.
Kamb, A., Gruis, N. A., et al. (1994). "A cell cycle regulator potentially involved in genesis of many tumor types." Science **264**(5157): 436-40.
Kambara, T., Simms, L. A., et al. (2004). "BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum." Gut **53**(8): 1137-44.
Kane, M. F., Loda, M., et al. (1997). "Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines." Cancer Res **57**(5): 808-11.
Kang, Y., Siegel, P. M., et al. (2003). "A multigenic program mediating breast cancer metastasis to bone." Cancer Cell **3**(6): 537-49.
Kaplan, K. B., Burds, A. A., et al. (2001). "A role for the Adenomatous Polyposis Coli protein in chromosome segregation." Nat Cell Biol **3**(4): 429-32.
Kemp, Z., Thirlwell, C., et al. (2004). "An update on the genetics of colorectal cancer." Hum Mol Genet **13 Spec No 2**: R177-85.
Kim, H., Jen, J., et al. (1994). "Clinical and pathological characteristics of sporadic colorectal carcinomas with DNA replication errors in microsatellite sequences." Am J Pathol **145**(1): 148-56.
Kim, J., Takeuchi, H., et al. (2005). "Chemokine receptor CXCR4 expression in colorectal cancer patients increases the risk for recurrence and for poor survival." J Clin Oncol **23**(12): 2744-53.
Kinzler, K. W., Nilbert, M. C., et al. (1991). "Identification of FAP locus genes from chromosome 5q21." Science **253**(5020): 661-5.
Kinzler, K. W. and Vogelstein, B. (1996). "Lessons from hereditary colorectal cancer." Cell **87**(2): 159-70.
Kinzler, K. W. and Vogelstein, B. (1997). "Cancer-susceptibility genes. Gatekeepers and caretakers." Nature **386**(6627): 761, 763.
Klein, C. A., Blankenstein, T. J., et al. (2002). "Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer." Lancet **360**(9334): 683-9.
Knudsen, A. L., Bisgaard, M. L., et al. (2003). "Attenuated familial adenomatous polyposis (AFAP). A review of the literature." Fam Cancer **2**(1): 43-55.

Knudson, A. G., Jr. (1971). "Mutation and cancer: statistical study of retinoblastoma." Proc Natl Acad Sci U S A **68**(4): 820-3.
Koinuma, K., Shitoh, K., et al. (2004). "Mutations of BRAF are associated with extensive hMLH1 promoter methylation in sporadic colorectal carcinomas." Int J Cancer **108**(2): 237-42.
Kruglyak, L. (1999). "Prospects for whole-genome linkage disequilibrium mapping of common disease genes." Nat Genet **22**(2): 139-44.
Kruglyak, L. and Nickerson, D. A. (2001). "Variation is the spice of life." Nat Genet **27**(3): 234-6.
Kuismanen, S. A., Holmberg, M. T., et al. (1999). "Epigenetic phenotypes distinguish microsatellite-stable and -unstable colorectal cancers." Proc Natl Acad Sci U S A **96**(22): 12661-6.
Laken, S. J., Petersen, G. M., et al. (1997). "Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC." Nat Genet **17**(1): 79-83.
Lamlum, H., Al Tassan, N., et al. (2000). "Germline APC variants in patients with multiple colorectal adenomas, with evidence for the particular importance of E1317Q." Hum Mol Genet **9**(15): 2215-21.
Lammi, L., Arte, S., et al. (2004). "Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer." Am J Hum Genet **74**(5): 1043-50.
Landi, S., Moreno, V., et al. (2003). "Association of common polymorphisms in inflammatory genes interleukin (IL)6, IL8, tumor necrosis factor alpha, NFKB1, and peroxisome proliferator-activated receptor gamma with colorectal cancer." Cancer Res **63**(13): 3560-6.
Le Marchand, L., Seifried, A., et al. (2003). "Association of the cyclin D1 A870G polymorphism with advanced colorectal cancer." Jama **290**(21): 2843-8.
Le Roith, D., Karas, M., et al. (1999). "The role of the insulin-like growth factors in cancer." Isr Med Assoc J **1**(1): 25-30.
Leadon, S. A. and Avrutskaya, A. V. (1998). "Requirement for DNA mismatch repair proteins in the transcription-coupled repair of thymine glycols in Saccharomyces cerevisiae." Mutat Res **407**(2): 177-87.
Lengauer, C., Kinzler, K. W., et al. (1998). "Genetic instabilities in human cancers." Nature **396**(6712): 643-9.
Li, J., Nguyen, L., et al. (2004). "Lack of evidence for an association between WNT2 and RELN polymorphisms and autism." Am J Med Genet B Neuropsychiatr Genet **126**(1): 51-7.
Li, M., Lin, Y. M., et al. (2004). "Genes associated with liver metastasis of colon cancer, identified by genome-wide cDNA microarray." Int J Oncol **24**(2): 305-12.
Li, Y., Gorbea, C., et al. (1997). "MAD2 associates with the cyclosome/anaphase-promoting complex and inhibits its activity." Proc Natl Acad Sci U S A **94**(23): 12431-6.
Lichtenstein, P., Holm, N. V., 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., 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): 1140-7.
Lindblom, A. (2001). "Different mechanisms in the tumorigenesis of proximal and distal colon cancers." Curr Opin Oncol **13**(1): 63-9.
Lindblom, A. and Linder, S. (1996). "The metastatic phenotype--prognostic implications." Crit Rev Oncol Hematol **24**(2): 71-96.
Lindblom, A., Tannergard, P., et al. (1993). "Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer." Nat Genet **5**(3): 279-82.
Lindblom, A., Zhou, X. L., et al. (2004). "Colorectal cancer as a complex disease: defining at-risk subjects in the general population - a preventive strategy." Expert Rev Anticancer Ther **4**(3): 377-85.
Lindgren, G., Liljegren, A., et al. (2002). "Adenoma prevalence and cancer risk in familial non-polyposis colorectal cancer." Gut **50**(2): 228-34.
Lindor, N. M., Burgart, L. J., et al. (2002). "Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors." J Clin Oncol **20**(4): 1043-8.
Lipkin, S. M., Moens, P. B., et al. (2002). "Meiotic arrest and aneuploidy in MLH3-deficient mice." Nat Genet **31**(4): 385-90.
Lipkin, S. M., Rozek, L. S., et al. (2004). "The MLH1 D132H variant is associated with susceptibility to sporadic colorectal cancer." Nat Genet **36**(7): 694-9.
Lipkin, S. M., Wang, V., et al. (2000). "MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability." Nat Genet **24**(1): 27-35.
Lipton, L., Halford, S. E., et al. (2003). "Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway." Cancer Res **63**(22): 7595-9.

Liu, H. X., Zhou, X. L., et al. (2003). "The role of hMLH3 in familial colorectal cancer." Cancer Res **63**(8): 1894-9.
Liu, T., Yan, H., et al. (2001). "The role of hPMS1 and hPMS2 in predisposing to colorectal cancer." Cancer Res **61**(21): 7798-802.
Liu, W., Dong, X., et al. (2000). "Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating beta-catenin/TCF signalling." Nat Genet **26**(2): 146-7.
Lohmueller, K. E., Pearce, C. L., et al. (2003). "Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease." Nat Genet **33**(2): 177-82.
Longacre, T. A. and Fenoglio-Preiser, C. M. (1990). "Mixed hyperplastic adenomatous polyps/serrated adenomas. A distinct form of colorectal neoplasia." Am J Surg Pathol **14**(6): 524-37.
Lothe, R. A., Peltomaki, P., et al. (1993). "Genomic instability in colorectal cancer: relationship to clinicopathological variables and family history." Cancer Res **53**(24): 5849-52.
Lovett, E. (1976). "Family studies in cancer of the colon and rectum." Br J Surg **63**(1): 13-18.
Lu, S. L., Kawabata, M., et al. (1998). "HNPCC associated with germline mutation in the TGF-beta type II receptor gene." Nat Genet **19**(1): 17-8.
Lucci-Cordisco, E., Zito, I., et al. (2003). "Hereditary nonpolyposis colorectal cancer and related conditions." Am J Med Genet A **122**(4): 325-34.
Luebeck, E. G. and Moolgavkar, S. H. (2002). "Multistage carcinogenesis and the incidence of colorectal cancer." Proc Natl Acad Sci U S A **99**(23): 15095-100.
Luongo, C. and Dove, W. F. (1996). "Somatic genetic events linked to the Apc locus in intestinal adenomas of the Min mouse." Genes Chromosomes Cancer **17**(3): 194-8.
Lynch, H. T., Guirgis, H. A., et al. (1977). "Familial cancer syndromes: a survey." Cancer **39**(4 Suppl): 1867-81.
Lynch, H. T. and Krush, A. J. (1971). "Cancer family "G" revisited: 1895-1970." Cancer **27**(6): 1505-11.
Ma, D. Q., Whitehead, P. L., et al. (2005). "Identification of Significant Association and Gene-Gene Interaction of GABA Receptor Subunit Genes in Autism." Am J Hum Genet **77**(3): 377-88.
Markowitz, S., Wang, J., et al. (1995). "Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability." Science **268**(5215): 1336-8.
Marsh, D. J., Coulon, V., 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): 507-15.
Marsh, D. J., Kum, J. B., et al. (1999). "PTEN mutation spectrum and genotype-phenotype correlations in Bannayan-Riley-Ruvalcaba syndrome suggest a single entity with Cowden syndrome." Hum Mol Genet **8**(8): 1461-72.
Marsischky, G. T., Filosi, N., et al. (1996). "Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair." Genes Dev **10**(4): 407-20.
Martin, T. A., Watkins, G., et al. (2004). "Loss of tight junction plaque molecules in breast cancer tissues is associated with a poor prognosis in patients with breast cancer." Eur J Cancer **40**(18): 2717-25.
Matsuo, K., Hamajima, N., et al. (2002). "Methionine Synthase Reductase Gene A66G Polymorphism is Associated with Risk of Colorectal Cancer." Asian Pac J Cancer Prev **3**(4): 353-359.
Mazelin, L., Bernet, A., et al. (2004). "Netrin-1 controls colorectal tumorigenesis by regulating apoptosis." Nature **431**(7004): 80-4.
McCoy, P. A., Shao, Y., et al. (2002). "No association between the WNT2 gene and autistic disorder." Am J Med Genet **114**(1): 106-9.
McKenna, W. G., Weiss, M. C., et al. (1990). "The role of the H-ras oncogene in radiation resistance and metastasis." Int J Radiat Oncol Biol Phys **18**(4): 849-59.
McMichael, A. J. and Potter, J. D. (1980). "Reproduction, endogenous and exogenous sex hormones, and colon cancer: a review and hypothesis." J Natl Cancer Inst **65**(6): 1201-7.
Meijers-Heijboer, H., van den Ouweland, A., et al. (2002). "Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations." Nat Genet **31**(1): 55-9.
Meijers-Heijboer, H., Wijnen, J., et al. (2003). "The CHEK2 1100delC mutation identifies families with a hereditary breast and colorectal cancer phenotype." Am J Hum Genet **72**(5): 1308-14.
Mellon, I., Rajpal, D. K., et al. (1996). "Transcription-coupled repair deficiency and mutations in human mismatch repair genes." Science **272**(5261): 557-60.
Mercer, K. E. and Pritchard, C. A. (2003). "Raf proteins and cancer: B-Raf is identified as a mutational target." Biochim Biophys Acta **1653**(1): 25-40.
Michaels, M. L. and Miller, J. H. (1992). "The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine)." J Bacteriol **174**(20): 6321-5.

Michor, F., Iwasa, Y., et al. (2004). "Dynamics of cancer progression." Nat Rev Cancer **4**(3): 197-205.
Miyoshi, Y., Ando, H., et al. (1992). "Germ-line mutations of the APC gene in 53 familial adenomatous polyposis patients." Proc Natl Acad Sci U S A **89**(10): 4452-6.
Moinfar, F., Man, Y. G., et al. (2000). "Concurrent and independent genetic alterations in the stromal and epithelial cells of mammary carcinoma: implications for tumorigenesis." Cancer Res **60**(9): 2562-6.
Mori, Y., Selaru, F. M., et al. (2003). "The impact of microsatellite instability on the molecular phenotype of colorectal tumors." Cancer Res **63**(15): 4577-82.
Mori, Y., Yin, J., et al. (2004). "Identification of genes uniquely involved in frequent microsatellite instability colon carcinogenesis by expression profiling combined with epigenetic scanning." Cancer Res **64**(7): 2434-8.
Mueller, M. M. and Fusenig, N. E. (2004). "Friends or foes - bipolar effects of the tumour stroma in cancer." Nat Rev Cancer **4**(11): 839-49.
Muller, A., Edmonston, T. B., et al. (2004). "MSI-testing in hereditary non-polyposis colorectal carcinoma (HNPCC)." Dis Markers **20**(4-5): 225-36.
Munger, K. (2002). "Disruption of oncogene/tumor suppressor networks during human carcinogenesis." Cancer Invest **20**(1): 71-81.
Nagase, H. and Nakamura, Y. (1993). "Mutations of the APC (adenomatous polyposis coli) gene." Hum Mutat **2**(6): 425-34.
Nakagawa, H., Lockman, J. C., et al. (2004). "Mismatch repair gene PMS2: disease-causing germline mutations are frequent in patients whose tumors stain negative for PMS2 protein, but paralogous genes obscure mutation detection and interpretation." Cancer Res **64**(14): 4721-7.
Nakamura, N. and Takenaga, K. (1998). "Hypomethylation of the metastasis-associated S100A4 gene correlates with gene activation in human colon adenocarcinoma cell lines." Clin Exp Metastasis **16**(5): 471-9.
Nakamura, Y., Nishisho, I., et al. (1991). "Mutations of the adenomatous polyposis coli gene in familial polyposis coli patients and sporadic colorectal tumors." Princess Takamatsu Symp **22**: 285-92.
Nathke, I. S. (2004). "The adenomatous polyposis coli protein: the Achilles heel of the gut epithelium." Annu Rev Cell Dev Biol **20**: 337-66.
Nicolaides, N. C., Littman, S. J., et al. (1998). "A naturally occurring hPMS2 mutation can confer a dominant negative mutator phenotype." Mol Cell Biol **18**(3): 1635-41.
Nicolaides, N. C., Papadopoulos, N., et al. (1994). "Mutations of two PMS homologues in hereditary nonpolyposis colon cancer." Nature **371**(6492): 75-80.
Nowell, P. C. (1976). "The clonal evolution of tumor cell populations." Science **194**(4260): 23-8.
Nowell, P. C. (2002). "Tumor progression: a brief historical perspective." Semin Cancer Biol **12**(4): 261-6.
Nowell, P. C. and Hungerford, D. A. (1960). "Chromosome studies on normal and leukemic human leukocytes." J Natl Cancer Inst **25**: 85-109.
Olschwang, S., Markie, D., et al. (1998). "Peutz-Jeghers disease: most, but not all, families are compatible with linkage to 19p13.3." J Med Genet **35**(1): 42-4.
Paget, S. (1989). "The distribution of secondary growths in cancer of the breast. 1889." Cancer Metastasis Rev **8**(2): 98-101.
Pantel, K. and Brakenhoff, R. H. (2004). "Dissecting the metastatic cascade." Nat Rev Cancer **4**(6): 448-56.
Pasche, B., Kaklamani, V., et al. (2004). "TGFBR1*6A and cancer: a meta-analysis of 12 case-control studies." J Clin Oncol **22**(4): 756-8.
Peltomaki, P. (2001). "Deficient DNA mismatch repair: a common etiologic factor for colon cancer." Hum Mol Genet **10**(7): 735-40.
Peltomaki, P. (2001). "DNA mismatch repair and cancer." Mutat Res **488**(1): 77-85.
Peltomaki, P. (2003). "Role of DNA mismatch repair defects in the pathogenesis of human cancer." J Clin Oncol **21**(6): 1174-9.
Peltomaki, P., Aaltonen, L. A., et al. (1993). "Genetic mapping of a locus predisposing to human colorectal cancer." Science **260**(5109): 810-2.
Peltomaki, P., Lothe, R. A., et al. (1993). "Microsatellite instability is associated with tumors that characterize the hereditary non-polyposis colorectal carcinoma syndrome." Cancer Res **53**(24): 5853-5.
Pharoah, P. D., Antoniou, A., et al. (2002). "Polygenic susceptibility to breast cancer and implications for prevention." Nat Genet **31**(1): 33-6.
Pharoah, P. D., Dunning, A. M., et al. (2004). "Association studies for finding cancer-susceptibility genetic variants." Nat Rev Cancer **4**(11): 850-60.

Pimm, J., McQuillin, A., et al. (2005). "The Epsin 4 gene on chromosome 5q, which encodes the clathrin-associated protein enthoprotin, is involved in the genetic susceptibility to schizophrenia." Am J Hum Genet **76**(5): 902-7.
Pisani, P., Parkin, D. M., et al. (1999). "Estimates of the worldwide mortality from 25 cancers in 1990." Int J Cancer **83**(1): 18-29.
Ponder, B. A. (2001). "Cancer genetics." Nature **411**(6835): 336-41.
Poste, G. and Fidler, I. J. (1980). "The pathogenesis of cancer metastasis." Nature **283**(5743): 139-46.
Potter, J. D. (1999). "Colorectal cancer: molecules and populations." J Natl Cancer Inst **91**(11): 916-32.
Powell, S. M., Zilz, N., et al. (1992). "APC mutations occur early during colorectal tumorigenesis." Nature **359**(6392): 235-7.
Pritchard, J. K. and Rosenberg, N. A. (1999). "Use of unlinked genetic markers to detect population stratification in association studies." Am J Hum Genet **65**(1): 220-8.
Pritchard, J. K., Stephens, M., et al. (2000). "Inference of population structure using multilocus genotype data." Genetics **155**(2): 945-59.
Rajagopalan, H., Jallepalli, P. V., et al. (2004). "Inactivation of hCDC4 can cause chromosomal instability." Nature **428**(6978): 77-81.
Rajagopalan, H. and Lengauer, C. (2004). "Aneuploidy and cancer." Nature **432**(7015): 338-41.
Rajagopalan, H., Nowak, M. A., et al. (2003). "The significance of unstable chromosomes in colorectal cancer." Nat Rev Cancer **3**(9): 695-701.
Ramaswamy, S., Ross, K. N., et al. (2003). "A molecular signature of metastasis in primary solid tumors." Nat Genet **33**(1): 49-54.
Rampino, N., Yamamoto, H., et al. (1997). "Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype." Science **275**(5302): 967-9.
Renkonen, E., Zhang, Y., et al. (2003). "Altered expression of MLH1, MSH2, and MSH6 in predisposition to hereditary nonpolyposis colorectal cancer." J Clin Oncol **21**(19): 3629-37.
Risch, N. and Merikangas, K. (1996). "The future of genetic studies of complex human diseases." Science **273**(5281): 1516-7.
Risch, N. J. (2000). "Searching for genetic determinants in the new millennium." Nature **405**(6788): 847-56.
Risinger, J. I., Barrett, J. C., et al. (1996). "Molecular genetic evidence of the occurrence of breast cancer as an integral tumor in patients with the hereditary nonpolyposis colorectal carcinoma syndrome." Cancer **77**(9): 1836-43.
Risinger, J. I., Berchuck, A., et al. (1993). "Genetic instability of microsatellites in endometrial carcinoma." Cancer Res **53**(21): 5100-3.
Ronaghi, M., Karamohamed, S., et al. (1996). "Real-time DNA sequencing using detection of pyrophosphate release." Anal Biochem **242**(1): 84-9.
Rowan, A. J., Lamlum, H., et al. (2000). "APC mutations in sporadic colorectal tumors: A mutational "hotspot" and interdependence of the "two hits"." Proc Natl Acad Sci U S A **97**(7): 3352-7.
Rowley, J. D. (1975). "Nonrandom chromosomal abnormalities in hematologic disorders of man." Proc Natl Acad Sci U S A **72**(1): 152-6.
Rozen, P., Shomrat, R., et al. (1999). "Prevalence of the I1307K APC gene variant in Israeli Jews of differing ethnic origin and risk for colorectal cancer." Gastroenterology **116**(1): 54-7.
Saha, S., Bardelli, A., et al. (2001). "A phosphatase associated with metastasis of colorectal cancer." Science **294**(5545): 1343-6.
Sakatani, T., Kaneda, A., et al. (2005). "Loss of imprinting of Igf2 alters intestinal maturation and tumorigenesis in mice." Science **307**(5717): 1976-8.
Sakumi, K., Furuichi, M., et al. (1993). "Cloning and expression of cDNA for a human enzyme that hydrolyzes 8-oxo-dGTP, a mutagenic substrate for DNA synthesis." J Biol Chem **268**(31): 23524-30.
Sampson, J. R., Dolwani, S., et al. (2003). "Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH." Lancet **362**(9377): 39-41.
Samuels, Y., Wang, Z., et al. (2004). "High frequency of mutations of the PIK3CA gene in human cancers." Science **304**(5670): 554.
Sanger, F., Nicklen, S., et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-7.
Santucci-Darmanin, S., Neyton, S., et al. (2002). "The DNA mismatch-repair MLH3 protein interacts with MSH4 in meiotic cells, supporting a role for this MutL homolog in mammalian meiotic recombination." Hum Mol Genet **11**(15): 1697-706.
Santucci-Darmanin, S., Walpita, D., et al. (2000). "MSH4 acts in conjunction with MLH1 during mammalian meiosis." Faseb J **14**(11): 1539-47.

Sayed, M. G., Ahmed, A. F., et al. (2002). "Germline SMAD4 or BMPR1A mutations and phenotype of juvenile polyposis." Ann Surg Oncol **9**(9): 901-6.
Schreibman, I. R., Baker, M., et al. (2005). "The hamartomatous polyposis syndromes: a clinical and molecular review." Am J Gastroenterol **100**(2): 476-90.
Schwartz, R. A. and Torre, D. P. (1995). "The Muir-Torre syndrome: a 25-year retrospect." J Am Acad Dermatol **33**(1): 90-104.
Schwartz, S., Jr., Yamamoto, H., 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.
Shaheen, N. J., Silverman, L. M., et al. (2003). "Association between hemochromatosis (HFE) gene mutation carrier status and the risk of colon cancer." J Natl Cancer Inst **95**(2): 154-9.
Sherr, C. J. (2004). "Principles of tumor suppression." Cell **116**(2): 235-46.
Shigenaga, M. K., Park, J. W., et al. (1990). "In vivo oxidative DNA damage: measurement of 8-hydroxy-2'-deoxyguanosine in DNA and urine by high-performance liquid chromatography with electrochemical detection." Methods Enzymol **186**: 521-30.
Shih, I. M., Zhou, W., et al. (2001). "Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis." Cancer Res **61**(3): 818-22.
Shin, B. Y., Chen, H., et al. (2005). "Low Allele Frequency of MLH1 D132H in American Colorectal and Endometrial Cancer Patients." Dis Colon Rectum.
Shin, Y., Kim, I. J., et al. (2004). "A functional polymorphism (-347 G-->GA) in the E-cadherin gene is associated with colorectal cancer." Carcinogenesis **25**(11): 2173-6.
Sieber, O. M., Heinimann, K., et al. (2003). "Genomic instability--the engine of tumorigenesis?" Nat Rev Cancer **3**(9): 701-8.
Sieber, O. M., Lamlum, H., et al. (2002). "Whole-gene APC deletions cause classical familial adenomatous polyposis, but not attenuated polyposis or "multiple" colorectal adenomas." Proc Natl Acad Sci U S A **99**(5): 2954-8.
Sieber, O. M., Lipton, L., et al. (2003). "Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH." N Engl J Med **348**(9): 791-9.
Sieber, O. M., Tomlinson, I. P., et al. (2000). "The adenomatous polyposis coli (APC) tumour suppressor--genetics, function and disease." Mol Med Today **6**(12): 462-9.
Skoglund, J., Djureinovic, T., et al. (2005). "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." Journal of Medical Genetics **accepted**.
Slattery, M. L., Neuhausen, S. L., et al. (2004). "Dietary calcium, vitamin D, VDR genotypes and colorectal cancer." Int J Cancer **111**(5): 750-6.
Slattery, M. L., Samowitz, W., et al. (2001). "A molecular variant of the APC gene at codon 1822: its association with diet, lifestyle, and risk of colon cancer." Cancer Res **61**(3): 1000-4.
Slattery, M. L., Samowitz, W., et al. (2004). "Associations among IRS1, IRS2, IGF1, and IGFBP3 genetic polymorphisms and colorectal cancer." Cancer Epidemiol Biomarkers Prev **13**(7): 1206-14.
Slupska, M. M., Baikalov, C., et al. (1996). "Cloning and sequencing a human homolog (hMYH) of the Escherichia coli mutY gene whose function is required for the repair of oxidative DNA damage." J Bacteriol **178**(13): 3885-92.
Smith, J. M., Johanesen, P. A., et al. (2005). "CXCL12 activation of CXCR4 regulates mucosal host defense through stimulation of epithelial cell migration and promotion of intestinal barrier integrity." Am J Physiol Gastrointest Liver Physiol **288**(2): G316-26.
Sondergaard, J. O., Bulow, S., et al. (1991). "Cancer incidence among parents of patients with colorectal cancer." Int J Cancer **47**(2): 202-6.
Sonnenschein, C. and Soto, A. M. (2000). "Somatic mutation theory of carcinogenesis: why it should be dropped and replaced." Mol Carcinog **29**(4): 205-11.
Souza, R. F., Appel, R., et al. (1996). "Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumours." Nat Genet **14**(3): 255-7.
Souza, R. F., Yin, J., et al. (1997). "Frequent mutation of the E2F-4 cell cycle gene in primary human gastrointestinal tumors." Cancer Res **57**(12): 2350-3.
Sparks, A. B., Morin, P. J., et al. (1998). "Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer." Cancer Res **58**(6): 1130-4.
Spencer, F., Gerring, S. L., et al. (1990). "Mitotic chromosome transmission fidelity mutants in Saccharomyces cerevisiae." Genetics **124**(2): 237-49.
Spirio, L. N., Kutchera, W., et al. (1996). "Three secretory phospholipase A(2) genes that map to human chromosome 1P35-36 are not mutated in individuals with attenuated adenomatous polyposis coli." Cancer Res **56**(5): 955-8.

St John, D. J., McDermott, F. T., et al. (1993). "Cancer risk in relatives of patients with common colorectal cancer." Ann Intern Med **118**(10): 785-90.
Steeg, P. S. (2003). "Metastasis suppressors alter the signal transduction of cancer cells." Nat Rev Cancer **3**(1): 55-63.
Su, L. K., Kohlmann, W., et al. (2002). "Different familial adenomatous polyposis phenotypes resulting from deletions of the entire APC exon 15." Hum Genet **111**(1): 88-95.
Su, L. K., Steinbach, G., et al. (2000). "Genomic rearrangements of the APC tumor-suppressor gene in familial adenomatous polyposis." Hum Genet **106**(1): 101-7.
Suter, C. M., Martin, D. I., et al. (2004). "Germline epimutation of MLH1 in individuals with multiple cancers." Nat Genet **36**(5): 497-501.
Suzuki, E., Ota, T., et al. (2004). "nm23-H1 reduces in vitro cell migration and the liver metastatic potential of colon cancer cells by regulating myosin light chain phosphorylation." Int J Cancer **108**(2): 207-11.
Suzuki, H., Gabrielson, E., et al. (2002). "A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer." Nat Genet **31**(2): 141-9.
Svetlanov, A. and Cohen, P. E. (2004). "Mismatch repair proteins, meiosis, and mice: understanding the complexities of mammalian meiosis." Exp Cell Res **296**(1): 71-9.
Tabor, H. K., Risch, N. J., et al. (2002). "Opinion: Candidate-gene approaches for studying complex genetic traits: practical considerations." Nat Rev Genet **3**(5): 391-7.
Tagashira, H., Hamazaki, K., et al. (1998). "Reduced metastatic potential and c-myc overexpression of colon adenocarcinoma cells (Colon 26 line) transfected with nm23-R2/rat nucleoside diphosphate kinase alpha isoform." Int J Mol Med **2**(1): 65-8.
Tashiro, H., Blazes, M. S., et al. (1997). "Mutations in PTEN are frequent in endometrial carcinoma but rare in other common gynecological malignancies." Cancer Res **57**(18): 3935-40.
Djureinovic, T., * Skoglund, J.,* et al. (2005). "A genome-wide linkage analysis in Swedish families with hereditary non-FAP/non-HNPCC colorectal cancer." *Gut. In press.* (* authors contributed equally)
Terdiman, J. P., Conrad, P. G., et al. (1999). "Genetic testing in hereditary colorectal cancer: indications and procedures." Am J Gastroenterol **94**(9): 2344-56.
Terwilliger, J. D. and Weiss, K. M. (1998). "Linkage disequilibrium mapping of complex disease: fantasy or reality?" Curr Opin Biotechnol **9**(6): 578-94.
Thibodeau, S. N., Bren, G., et al. (1993). "Microsatellite instability in cancer of the proximal colon." Science **260**(5109): 816-9.
Thibodeau, S. N., French, A. J., et al. (1998). "Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1." Cancer Res **58**(8): 1713-8.
Thorstensen, L., Diep, C. B., et al. (2001). "WNT1 inducible signaling pathway protein 3, WISP-3, a novel target gene in colorectal carcinomas with microsatellite instability." Gastroenterology **121**(6): 1275-80.
Toyota, M., Ahuja, N., et al. (1999). "CpG island methylator phenotype in colorectal cancer." Proc Natl Acad Sci U S A **96**(15): 8681-6.
Tutt, A., Gabriel, A., et al. (1999). "Absence of Brca2 causes genome instability by chromosome breakage and loss associated with centrosome amplification." Curr Biol **9**(19): 1107-10.
Ulrich, C. M., Bigler, J., et al. (2002). "Thymidylate synthase promoter polymorphism, interaction with folate intake, and risk of colorectal adenomas." Cancer Res **62**(12): 3361-4.
Ulrich, C. M., Bigler, J., et al. (2004). "Polymorphisms in PTGS1 (=COX-1) and risk of colorectal polyps." Cancer Epidemiol Biomarkers Prev **13**(5): 889-93.
Umar, A., Boland, C. R., et al. (2004). "Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability." J Natl Cancer Inst **96**(4): 261-8.
Umar, A., Risinger, J. I., et al. (2004). "Testing guidelines for hereditary non-polyposis colorectal cancer." Nat Rev Cancer **4**(2): 153-8.
Wahlberg, S., Liu, T., et al. (1999). "Various mutation screening techniques in the DNA mismatch repair genes hMSH2 and hMLH1." Genet Test **3**(3): 259-64.
Wakefield, A. J. (2002). "The gut-brain axis in childhood developmental disorders." J Pediatr Gastroenterol Nutr **34 Suppl 1**: S14-7.
Wallis, Y. L., Morton, D. G., et al. (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.
van de Vijver, M. J., He, Y. D., et al. (2002). "A gene-expression signature as a predictor of survival in breast cancer." N Engl J Med **347**(25): 1999-2009.

van Engeland, M., Roemen, G. M., et al. (2002). "K-ras mutations and RASSF1A promoter methylation in colorectal cancer." Oncogene **21**(23): 3792-5.
van 't Veer, L. J., Dai, H., et al. (2002). "Gene expression profiling predicts clinical outcome of breast cancer." Nature **415**(6871): 530-6.
Vandrovcova, J., Stekrova, J., et al. (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.
Wang, L., Baudhuin, L. M., et al. (2004). "MYH mutations in patients with attenuated and classic polyposis and with young-onset colorectal cancer without polyps." Gastroenterology **127**(1): 9-16.
Wang, W. Y., Barratt, B. J., et al. (2005). "Genome-wide association studies: theoretical and practical concerns." Nat Rev Genet **6**(2): 109-18.
Wang, Z., Cummins, J. M., et al. (2004). "Three classes of genes mutated in colorectal cancers with chromosomal instability." Cancer Res **64**(9): 2998-3001.
Varghese, H. J., Davidson, M. T., et al. (2002). "Activated ras regulates the proliferation/apoptosis balance and early survival of developing micrometastases." Cancer Res **62**(3): 887-91.
Warthin, A. (1913). "Heredity with reference to carcinoma." Arch Intern Med **12**: 546-555.
Vasen, H. F., Hendriks, Y., et al. (2004). "Identification of HNPCC by molecular analysis of colorectal and endometrial tumors." Dis Markers **20**(4-5): 207-13.
Vasen, H. F., Mecklin, J. P., et al. (1991). "The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC)." Dis Colon Rectum **34**(5): 424-5.
Vasen, H. F., Watson, P., et al. (1999). "New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC." Gastroenterology **116**(6): 1453-6.
Wassink, T. H., Piven, J., et al. (2001). "Evidence supporting WNT2 as an autism susceptibility gene." Am J Med Genet **105**(5): 406-13.
Weaver, V. M. and Gilbert, P. (2004). "Watch thy neighbor: cancer is a communal affair." J Cell Sci **117**(Pt 8): 1287-90.
Weaver, Z., Montagna, C., et al. (2002). "Mammary tumors in mice conditionally mutant for Brca1 exhibit gross genomic instability and centrosome amplification yet display a recurring distribution of genomic imbalances that is similar to human breast cancer." Oncogene **21**(33): 5097-107.
Weitz, J., Koch, M., et al. (2005). "Colorectal cancer." Lancet **365**(9454): 153-65.
Weng, L. P., Smith, W. M., et al. (1999). "PTEN suppresses breast cancer cell growth by phosphatase activity-dependent G1 arrest followed by cell death." Cancer Res **59**(22): 5808-14.
Westerman, A. M. and Wilson, J. H. (1999). "Peutz-Jeghers syndrome: risks of a hereditary condition." Scand J Gastroenterol Suppl **230**: 64-70.
Whitehall, V. L., Walsh, M. D., et al. (2001). "Methylation of O-6-methylguanine DNA methyltransferase characterizes a subset of colorectal cancer with low-level DNA microsatellite instability." Cancer Res **61**(3): 827-30.
Wijnen, J., de Leeuw, W., et al. (1999). "Familial endometrial cancer in female carriers of MSH6 germline mutations." Nat Genet **23**(2): 142-4.
Vogelstein, B., Fearon, E. R., et al. (1988). "Genetic alterations during colorectal-tumor development." N Engl J Med **319**(9): 525-32.
Vogelstein, B. and Kinzler, K. W. (1993). "The multistep nature of cancer." Trends Genet **9**(4): 138-41.
Vogelstein, B. and Kinzler, K. W. (2004). "Cancer genes and the pathways they control." Nat Med **10**(8): 789-99.
Wong, H. L., Seow, A., et al. (2003). "Vitamin D receptor start codon polymorphism and colorectal cancer risk: effect modification by dietary calcium and fat in Singapore Chinese." Carcinogenesis **24**(6): 1091-5.
Wong, V. and Gumbiner, B. M. (1997). "A synthetic peptide corresponding to the extracellular domain of occludin perturbs the tight junction permeability barrier." J Cell Biol **136**(2): 399-409.
Wood, M. L., Esteve, A., et al. (1992). "Genetic effects of oxidative DNA damage: comparative mutagenesis of 7,8-dihydro-8-oxoguanine and 7,8-dihydro-8-oxoadenine in Escherichia coli." Nucleic Acids Res **20**(22): 6023-32.
Woodage, T., King, S. M., et al. (1998). "The APCI1307K allele and cancer risk in a community-based study of Ashkenazi Jews." Nat Genet **20**(1): 62-5.
Wright, C. M., Dent, O. F., et al. (2000). "Prognostic significance of extensive microsatellite instability in sporadic clinicopathological stage C colorectal cancer." Br J Surg **87**(9): 1197-202.
Wu, Y., Berends, M. J., et al. (2001). "A role for MLH3 in hereditary nonpolyposis colorectal cancer." Nat Genet **29**(2): 137-8.
Xiong, M., Akey, J., et al. (2000). "The haplotype linkage disequilibrium test for genome-wide screens: its power and study design." Pac Symp Biocomput: 675-86.

Yamamoto, H., Hanafusa, H., et al. (2005). "Single nucleotide polymorphisms in the EXO1 gene and risk of colorectal cancer in a Japanese population." Carcinogenesis **26**(2): 411-6.
Yamamoto, H., Imai, K., et al. (2002). "Gastrointestinal cancer of the microsatellite mutator phenotype pathway." J Gastroenterol **37**(3): 153-63.
Yamamoto, H., Sawai, H., et al. (1997). "Frameshift somatic mutations in gastrointestinal cancer of the microsatellite mutator phenotype." Cancer Res **57**(19): 4420-6.
Yan, H., Dobbie, Z., et al. (2002). "Small changes in expression affect predisposition to tumorigenesis." Nat Genet **30**(1): 25-6.
Yan, H., Papadopoulos, N., et al. (2000). "Conversion of diploidy to haploidy." Nature **403**(6771): 723-4.
Yuen, S. T., Davies, H., et al. (2002). "Similarity of the phenotypic patterns associated with BRAF and KRAS mutations in colorectal neoplasia." Cancer Res **62**(22): 6451-5.
Zauber, N. P., Sabbath-Solitare, M., et al. (2003). "The characterization of somatic APC mutations in colonic adenomas and carcinomas in Ashkenazi Jews with the APC I1307K variant using linkage disequilibrium." J Pathol **199**(2): 146-51.
Zhou, H., Kuang, J., et al. (1998). "Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation." Nat Genet **20**(2): 189-93.
Zondervan, K. T. and Cardon, L. R. (2004). "The complex interplay among factors that influence allelic association." Nat Rev Genet **5**(2): 89-100.