MOLECULAR ALTERATIONS IN FAMILIAL COLORECTAL CANCER

Tuija Hienonen

Helsinki 2005

MOLECULAR ALTERATIONS IN FAMILIAL COLORECTAL CANCER

Tuija Hienonen

Department of Medical Genetics
Molecular and Cancer Biology Programme
Haartman Institute and Biomedicum Helsinki
University of Helsinki
Finland

Academic dissertation

To be publicly discussed, with the permission of the Medical Faculty of the University of Helsinki, in the lecture hall 2, Biomedicum Helsinki, Haartmaninkatu 8, on November 25th 2005, at 12 noon

Helsinki 2005

Supervised by

Lauri A. Aaltonen, MD, PhD
Research Professor of the Finnish Academy of Sciences
Department of Medical Genetics
Haartman Institute and Biomedicum Helsinki
University of Helsinki
Finland

Auli Karhu, PhD
Department of Medical Genetics
Haartman Institute and Biomedicum Helsinki
University of Helsinki
Finland

Reviewed by

Raquel Seruca, MD, PhD
Institute of Molecular Pathology and Immunology
University of Porto
Portugal

Outi Monni, PhD
Docent
Biomedicum Biochip Center
Institute of Biomedicine
University of Helsinki
Finland

Official opponent

Johanna Schleutker, PhD
Docent, principal investigator
Laboratory of Cancer Genetics
Institute of Medical Technology
University of Tampere and
Tampere University Hospital
Finland

ISBN 952-91-9348-3 (paperback)
ISBN 952-10-2730-4 (PDF)
http://ethesis.helsinki.fi
Yliopistopaino
Helsinki 2005

CONTENTS

ABBREVIATIONS

A	adenine
ACF	aberrant crypt focus
AFAP	attenuated familial adenomatous polyposis
AI	allelic imbalance
APC	adenomatous polyposis coli
AURKA	aurora kinase A
BAX	BLC2-associated X protein
bp	base pair
BER	base excision repair
BRAF	v-raf murine sarcoma viral oncogene homolog B1
BRCA2	breast cancer 2
BUB1	budding uninhibited by benzimidazoles 1 (yeast) homolog
BUB1B	budding uninhibited by benzimidazoles 1 (yeast) homolog beta
C	cytosine
CDK4	cyclin-dependent kinase 4
CDKN2A	cyclin-dependent kinase inhibitor 2A
cDNA	complementary deoxyribonucleic acid
CGH	comparative genomic hybridization
CIN	chromosomal instability
cM	centiMorgan
c-Myc	v-myc avian myelocytomatosis viral oncogene homolog
CRC	colorectal cancer
CS	Cowden syndrome
Cyr61	cysteine-rich, angiogenic inducer, 61
DCC	deleted in colorectal carcinoma
dHPLC	denaturing high-performance liquid chromatography
DNA	deoxyribonucleic acid
DSBR	double strand break repair
EST	expressed sequence tag
EXO1	exonuclease 1
FAP	familial adenomatous polyposis
FMM	familial malignant melanoma
G	guanine
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GI	gastrointestinal
GIST	gastrointestinal stromal tumour
GSTT1	glutathione S-transferase theta 1
HNPCC	hereditary non-polyposis colorectal cancer
HPRC	hereditary papillary renal carcinoma
HRAS	Harvey rat sarcoma viral oncogene homolog
IDL	insertion/deletion loop
IGFII	insulin-like growth factor II
JP	juvenile polyposis
kb	kilobase
KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog

KRAS	Kirsten rat sarcoma viral oncogene homolog
LD	linkage disequilibrium
LEF	lymphoid enhancer factor, transcription factor
LKB1	serine/threonine kinase 11 (*STK11*)
LOH	loss of heterozygosity
LOI	loss of imprinting
MDM2	double minute 2 (mouse) homolog
MEN2	multiple endocrine neoplasia type 2
MET	met proto-oncogene (hepatocyte growth factor receptor)
MGMT	O^6-methylguanine-DNA methyltransferase
Min	multiple intestinal neoplasia
MLH1, -3	MutL (*E. coli*) homologs 1 and 3
MMR	mismatch repair
Mom1, -2	modifier of multiple intestinal neoplasia 1 and 2
MRE11	meiotic recombination protein 11 (*S. cerevisiae*) homolog
mRNA	messenger ribonucleic acid
MSH2, -3, -6	MutS (*E. coli*) homologs 2, 3, and 6
MSI	microsatellite instability, microsatellite-unstable
MSI-H	high microsatellite instability
MSI-L	low microsatellite instability
MSS	microsatellite stability, microsatellite stable
MTHFR	5,10-methylenetetrahydrofolate reductase
MTS	Muir-Torre syndrome
MYCL	v-myc avian myelocytomatosis viral oncogene homolog, lung carcinoma-derived
MYH	MutY (*E. coli*) homolog
NAT1, -2	N-acetyltransferases 1 and 2
NER	nucleotide excision repair
NF1	neurofibromin 1
NMD	nonsense-mediated decay
p	short arm of a chromosome
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PJS	Peutz-Jeghers syndrome
PLA2G2A	phospholipase A2, group IIA
PMS1, -2	postmeiotic segregation increased 1 and 2
PRL-3	protein tyrosine phosphatase type IVA, member 3
PTEN	phosphatase and tensin homolog
q	long arm of a chromosome
qPCR	quantitative polymerase chain reaction
QTL	quantitative trait locus
RB	retinoblastoma
RET	ret proto-oncogene
RNA	ribonucleic acid
SEMG1	semenogelin I
SMAD2, -4	Mad (*D. melanogaster*) homologs 2 and 4
SNP	single nucleotide polymorphism
T	thymine
TCF	T-cell factor, transcription factor
TGF-β	transforming growth factor-β

TGFβRII	transforming growth factor-β receptor type II
TP53	tumour protein p53
TS	Turcot syndrome
VNTR	variable number of tandem repeats
WNT	wingless-type
XP	xeroderma pigmentosum
ZNF217	zinc finger protein 217

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original articles referred to in the text by Roman numerals I-IV.

I Laiho P*, **Hienonen T***, Karhu A, Lipton L, Aalto Y, Thomas HJ, Birkenkamp-Demtröder K, Hodgson S, Salovaara R, Mecklin JP, Järvinen H, Knuutila S, Halford S, Ørntoft TF, Tomlinson I, Launonen V, Houlston R, Aaltonen LA: Genome-wide allelotyping of 104 Finnish colorectal cancers reveals an excess of allelic imbalance in chromosome 20q in familial cases. *Oncogene* 2003; 22(14):2206-14.

II **Hienonen T**, Salovaara R, Mecklin JP, Järvinen H, Karhu A, Aaltonen LA: Preferential amplification of *AURKA* 91A (Ile31) in familial colorectal cancers. *International Journal of Cancer* 2005; [Epub ahead of print].

III **Hienonen T**, Laiho P, Salovaara R, Mecklin JP, Järvinen H, Sistonen P, Peltomäki P, Lehtonen R, Nupponen NN, Launonen V, Karhu A, Aaltonen LA: Little evidence for involvement of *MLH3* in colorectal cancer predisposition. *International Journal of Cancer* 2003; 106(2):292-6.

IV **Hienonen T**, Sammalkorpi H, Enholm S, Alhopuro P, Barber TD, Lehtonen R, Nupponen NN, Lehtonen H, Salovaara R, Mecklin JP, Järvinen H, Koistinen R, Arango D, Launonen V, Vogelstein B, Karhu A, Aaltonen LA: Mutations in two short noncoding mononucleotide repeats in most microsatellite-unstable colorectal cancers. *Cancer Research* 2005; 65(11):4607-13.

* Equal contribution

Publication I was included in the thesis of Päivi Laiho (Molecular classification of colorectal cancer, Helsinki 2005).

ABSTRACT

The genetic aetiology of colorectal cancer (CRC) is one of the best characterized of all human cancer types. Genes involved in cell cycle regulation or DNA error repair systems have been shown to predispose to familial CRC. Most of the identified susceptibility alleles are rare in the general population and may thus explain only a minority of all CRC cases. However, the more prevalent but less well characterized low-penetrance susceptibility alleles could associate with a majority of heritable cases. More information on the specific low-penetrance susceptibility alleles is thus required.

CRC progresses via two main pathways: the chromosomal instability (CIN) and microsatellite instability (MSI) pathways. Most CRCs display CIN with gross chromosomal alterations but the exact mechanisms initiating the instability phenotype are not well characterized. A smaller subset of the CRCs display MSI, i.e. small deletions or insertions in short repetitive regions of the genome, as a result of defective DNA mismatch repair (MMR) system. This study aimed at identifying the molecular mechanisms involved in CRC predisposition and progression in both of the instability pathways.

Novel susceptibility loci were sought by identifying commonly deleted or amplified regions in familial CIN CRC cases. Especially chromosome 20q13 was seen to harbour amplifications more frequently in familial than in sporadic CRCs indicating the possible existence of novel predisposing oncogene(s) within that amplicon. A 20q13 gene *AURKA* was chosen for further analysis since it has been suggested to function as a low-penetrance susceptibility gene. A 91A-specific amplification pattern was observed in the familial cases. Furthermore, a trend between younger age at diagnosis and 91A was observed, supporting the role of this allele in CRC predisposition.

Around 15% of all CRCs display MSI and are associated with inactive MMR genes *MLH1* or *MSH2*, and less frequently *PMS2* or *MSH6*. Causative mutation has not been identified in a subset of the MSI cases, indicating the possible involvement of other genes. We investigated the putative role of MMR gene *MLH3* in CRC predisposition. A patient with mild CRC clustering in the family harboured a germline missense mutation that was not found in the cancer-free controls. The variant is not an attractive disease-causing candidate, thus *MLH3* does not seem to play a major role in CRC predisposition.

We investigated the possible role of *SEMG1* as a novel target for microsatellite mutations in MSI CRCs. Although frequently mutated, the *SEMG1* microsatellite mutations do not seem to be selected for during tumorigenesis. Intergenic microsatellites were studied to obtain information on the general background mutation frequency of MSI CRCs. One microsatellite was frequently mutated, indicating that mutation frequency alone cannot be utilized to identify true MSI target genes.

REVIEW OF THE LITERATURE

1. Cancer genetics and epigenetics

Cancer originates from a single cell that acquires a clonal growth advantage over the surrounding cells. Additional genetic and epigenetic changes occurring in the progeny cells facilitate malignant transformation by producing genetically and biologically abnormal cells. This clonal progression model of cancer indicates that cancer formation requires several changes in a given cell (Nowell 1976). Thus age is a prominent factor in cancer susceptibility (Miller 1980) but cancer risk is higher in individuals with an inherited genetic predisposition.

Malignant transformation is a result of exogenous and endogenous factors disrupting cellular homeostasis normally controlled by the cell cycle, differentiation, and apoptosis pathways. It has been suggested that almost all cancer types display the following six features untypical of normal cells: self-sufficiency in growth signals, insensitivity to growth inhibitory signals, evasion of apoptosis, limitless replication potential, sustained angiogenesis, and metastasis (Hanahan and Weinberg 2000). These abnormal cellular features are caused by alterations in tumour suppressor genes and oncogenes that control the cellular homeostasis in an antagonist manner.

1.1 Tumour suppressor genes

Tumour suppressor genes inhibit abnormal cells from progressing in the cell cycle. Inactivation of both alleles of a tumour suppressor gene is required for uncontrolled cell growth, thus tumour suppressor genes are recessive at the cellular level (Knudson's two-hit hypothesis) (Knudson 1971). Individuals with an inherited tumour suppressor gene mutation are at a higher cancer risk because all cells harbour one inactive allele. Inherited mutations in tumour suppressor genes are typically small changes such as point mutations or small insertions/deletions. The second alteration affecting the wild-type allele is often larger, such as deletion of a whole chromosome.

A haploinsufficiency model suggests that inactivation of one allele of certain tumour suppressor genes is enough to promote tumorigenesis. Evidence of some haploinsufficient tumour suppressor genes have been obtained, for example the colorectal cancer (CRC)-associated genes *PTEN* and *LKB1* (Di Cristofano *et al.* 1998, Miyoshi *et al.* 2002, Rossi *et al.* 2002). Loss of the wild-type allele might provide a further growth advantage to the affected cells as suggested by the more frequent *LKB1* loss of heterozygosity (LOH) observed in carcinomas than in polyps (Entius *et al.* 2001).

Tumour suppressor genes can be divided into three groups: gatekeepers, landscapers, and caretakers (Kinzler and Vogelstein 1997, Kinzler and Vogelstein 1998). Gatekeepers control cell growth by inhibiting cell cycle progression and promoting apoptosis of abnormal cells. Landscaper gene mutations affect the cellular microenvironment (stromal cells) rather than growth pattern of the neoplastic cells themselves. Caretaker tumour suppressor genes are involved in controlling genomic

stability by repairing DNA errors (Kinzler and Vogelstein 1997, Kinzler and Vogelstein 1998).

1.1.1 DNA repair genes

DNA repair genes are caretaker tumour suppressor genes involved in controlling genomic stability. Exogenous and endogenous factors, such as radiation, reactive oxygen species, and DNA replication and recombination, constantly cause DNA sequence alterations that need to be corrected in order to maintain proper cellular functions. An array of DNA repair systems have been identified with at least five main DNA error correction mechanisms functioning in human cells: 1. base excision repair (BER) corrects residues modified by endogenous agents, 2. nucleotide excision repair (NER) restores mainly helix-distorting errors caused by environmental mutagens, 3. mismatch repair (MMR) corrects nucleotide mismatches and small insertion/deletion loops (IDLs) generated during DNA replication or recombination, 4. direct repair reverses DNA damage, and 5. double-strand break repair (DSBR) corrects breakage in the double helix (Hoeijmakers 2001, Friedberg 2003). Mutations in BER, NER, MMR, and DSBR genes have been shown to be causative in cancer predisposition syndromes, such as *MYH*-associated colorectal polyposis, xeroderma pigmentosum (XP), hereditary non-polyposis colorectal cancer (HNPCC), and breast cancer, respectively (Al-Tassan *et al.* 2002, Friedberg 2003).

1.2 Oncogenes

Oncogenes are excessively active counterparts of cellular proto-oncogenes functioning as, for example, growth factors, growth factor receptors, intracellular messengers, or transcription factors. Oncogenes promote cell proliferation by inducing mitosis and inhibiting differentiation and apoptosis. Gain of function mutations, such as point mutations, amplifications, or chromosomal rearrangements transform proto-oncogenes into excessively active oncogenes. Oncogenes function in a dominant manner, hence activating mutation in the other allele is sufficient to promote tumorigenesis (Bishop 1991).

Inherited oncogene mutations are rare, possibly because of negative selection pressure during embryonic development. Some known exceptions are germline mutations in *RET*, *MET*, *CDK4*, and *KIT* that have been found causative in cancer predisposition syndromes multiple endocrine neoplasia type 2 (MEN2) (Donis-Keller *et al.* 1993, Mulligan *et al.* 1993), hereditary papillary renal carcinoma (HPRC) (Zuo *et al.* 1996), familial malignant melanoma (FMM) (Schmidt *et al.* 1997), and familial gastrointestinal stromal tumour (GIST) syndrome (Nishida *et al.* 1998), respectively.

1.3 Epigenetic factors

Tumorigenic processes can also advance through epigenetic changes that affect gene expression levels without directly altering the DNA sequence. Epigenetic events such as altered methylation patterns, post-translational modification of histones, or arrangement of nucleosomes can lead to inactivation of tumour suppressor genes,

activation of oncogenes, or altered imprinting patterns (Lund and van Lohuizen 2004).

A prime example of an epigenetic event affecting tumour formation is the hypermethylation of *MLH1* promoter frequently observed in sporadic CRCs with microsatellite instability (MSI) (Kane *et al.* 1997, Cunningham *et al.* 1998, Herman *et al.* 1998, Veigl *et al.* 1998). Hypermethylation seems to be the main mechanism of *MLH1* inactivation in sporadic MSI CRCs (Kane *et al.* 1997, Cunningham *et al.* 1998, Herman *et al.* 1998, Veigl *et al.* 1998). *MLH1* inactivation abrogates normal MMR functions and results in a mutator phenotype due to the accumulation of microsatellite mutations. Other genes, for example *MGMT* and *CDKN2A* are also frequently hypermethylated already in adenomas of the colon (Petko *et al.* 2005).

Regulation of gene expression via loss of imprinting (LOI) has also been implicated in CRC (Kinouchi *et al.* 1996). Retained biallelic expression of *IGF2* has been linked to uncontrolled cell growth (Ogawa *et al.* 1993) and increased CRC risk (Cui *et al.* 2003).

2. Cancer susceptibility

Although a majority of cancer cases occur sporadically, hereditary counterparts of many cancer types have been identified (Vogelstein and Kinzler 2004). The risk of developing cancer is higher in individuals with an inherited predisposition. The total risk is dependent on the mode of inheritance, penetrance, contributing genetic factors (modifier genes), as well as on dietary, lifestyle, and environmental factors. Most of the cancer susceptibility syndromes have an autosomal dominant mode of inheritance but some rare examples of autosomal recessive syndromes exist, such as the *MYH*-associated colorectal polyposis (Al-Tassan *et al.* 2002). Inheriting a high-penetrance susceptibility allele increases cancer risk *per se*. Cancer families with mutations in high-penetrance genes are often characterized by: 1. younger age at onset than patients with sporadic disease, 2. bilateral disease or multiple disease sites in one organ, 3. multiple primary malignancies, 4. occurrence of cancer in the less usually affected sex, 5. cancer clustering in the family, and 6. cancer patients displaying other rare conditions (Marsh and Zori 2002). In comparison, the risk conferred by a low-penetrance susceptibility allele is influenced more by other genetic alterations and environmental factors. Therefore cancer usually manifests later and the degree of familial clustering is lower than for high-penetrance mutations (Kemp *et al.* 2004).

A single gene can harbour both high- and low-penetrance susceptibility alleles, as exemplified by the different sequence alterations in *APC* associated with diverse CRC risks and phenotypic manifestations. Different *APC* alterations can predispose either to: 1. a numerous amount of adenomatous polyps in familial adenomatous polyposis (FAP) patients with a near 100% risk of developing CRC if untreated (Kinzler and Vogelstein 1996), 2. a less severe polyposis phenotype and CRC risk in attenuated familial adenomatous polyposis (AFAP) patients (Kinzler and Vogelstein 1996), or 3. a modest, 1.5-2 fold increased CRC risk without an associated polyposis phenotype for example in Ashkenazi Jewish populations (Locker and Lynch 2004).

First-degree relatives of cancer patients with many common cancer types are at a 2-3 fold increased risk of developing cancer at the same site (Peto and Houlston 2001). Much of this familial aggregation is due to inherited susceptibility (Lichtenstein *et al.* 2000). Lichtenstein *et al.* (2000) estimated the heritability of cancer at 28 anatomical sites by studying almost 45000 twin pairs. The contribution of heritable factors was highest in prostate, colorectal, and breast cancer. An estimated 42% of the total prostate cancer risk was attributed to heritable factors, and the estimates for colorectal and breast cancer were 35% and 27%, respectively (Lichtenstein *et al.* 2000).

3. Familial colorectal cancer

Colorectal cancer (CRC) is the third most common cancer type and the fourth most common cause of cancer deaths worldwide (Parkin *et al.* 2005). CRC is more prevalent in the Western countries possibly due to environmental and dietary factors (Potter 1999). Hereditary factors have an important role in CRC with an estimated 35% of the total risk attributed to inherited susceptibility (Lichtenstein *et al.* 2000).

To date, a number of CRC susceptibility syndromes have been characterized that can be divided into two main groups based on the presence or absence of polyposis. Polyposis syndromes can be further divided into adenomatous and hamartomatous polyposis based on the disease phenotype. *MYH*-associated polyposis, familial adenomatous polyposis (FAP), and its variants attenuated familial adenomatous polyposis (AFAP), Gardner syndrome, and Turcot syndrome (TS) are characterized by adenomatous polyposis. Hamartomatous polyposis syndromes juvenile polyposis (JP), Peutz-Jeghers syndrome (PJS), Cowden syndrome (CS), and its variant Bannayan-Riley-Ruvalcaba syndrome have lower cancer risk than in adenomatous polyposis syndromes. The non-polyposis syndromes hereditary non-polyposis colorectal cancer (HNPCC) and its variants Muir-Torre syndrome (MTS) and Turcot syndrome are typically devoid of florid polyposis.

The most common and best characterized hereditary CRC syndromes FAP and HNPCC will be described in more detail in the following sections.

3.1 Familial adenomatous polyposis, FAP

FAP is characterized by the formation of hundreds, even thousands of adenomatous polyps and microadenomas in the large intestine already in adolescence or early adulthood. If untreated, colorectal cancer formation is inevitable by the early forties due to florid polyposis (Fearnhead *et al.* 2001). FAP is caused by mutations in *APC* (Groden *et al.* 1991, Nishisho *et al.* 1991). APC acts as a negative regulator of β-catenin in the WNT signalling pathway (Munemitsu *et al.* 1995). β-catenin in turn interacts with TCF and LEF transcription factors that control the expression levels of many genes, including oncogene *c-Myc* (He *et al.* 1998). Most *APC* mutations result in C-terminally truncated protein products with abrogated β-catenin binding (Kinzler and Vogelstein 1996). This leads to constitutive upregulation of the WNT signalling target genes (Korinek *et al.* 1997). Furthermore, truncated APC cannot promote the assembly of microtubules (Munemitsu *et al.* 1994, Smith *et al.* 1994) and may thus contribute to chromosomal instability (Fodde *et al.* 2001, Kaplan *et al.* 2001).

FAP patients display highly variable clinicopathological characteristics. For example, the number of polyps, age at cancer onset, and extracolonic manifestations such as the occurrence of desmoid tumours can vary between but also within families. This can be partly explained by the mutation spectrum of *APC*. For example, mutations leading to severe polyposis phenotype and early age at onset occur mostly in the central region of the gene (Nagase *et al.* 1992, Caspari *et al.* 1994, Gayther *et al.* 1994). In comparison, attenuated FAP is associated with mutations in the first or last third of *APC* (Spirio *et al.* 1993, Friedl *et al.* 1996, van der Luijt *et al.* 1996). The phenotype of FAP patients with the same *APC* mutation may also vary within and between families. This could be accounted for by inaccurate diagnosis, variability in the micro- or macroenvironmental conditions, or modifier genes (Houlston *et al.* 2001, Crabtree *et al.* 2002).

3.1.1 FAP modifier genes

Crabtree *et al.* (2002) compared phenotypes between close and more distant relatives with FAP and found evidence to support the existence of genetic modifiers of the disease. The modifier hypothesis is further supported by studies on *Min* (multiple intestinal neoplasia) mouse model of FAP. *Mom1* and *Mom2*, the modifier loci of *Min*, have been localized by linkage analysis to mouse chromosomes 4 and 18, respectively (Dietrich *et al.* 1993, Silverman *et al.* 2002). No evident mouse modifiers in *Mom2* locus have been identified to date (Silverman *et al.* 2003) but in *Mom1* locus, the variants of *Pla2g2a* (*Pla2s*) and *Myh* have been seen to modify polyp number (MacPhee *et al.* 1995, Sieber *et al.* 2004). However, the human homologs *PLA2G2A* or *MYH* do not seem to modify the FAP phenotype (Dobbie *et al.* 1996, Tomlinson *et al.* 1996, Plasilova *et al.* 2004, Kairupan *et al.* 2005). In humans, variants of N-acetyltransferases *NAT1* and *NAT2* have been suggested to modify the FAP phenotype (Crabtree *et al.* 2004).

3.2 Hereditary non-polyposis colorectal cancer, HNPCC

HNPCC patients are typically devoid of florid polyposis and adenoma development seems to be as frequent as in the general population. The aggressive adenoma theory of Jass (1995) suggests that the adenomas progress into carcinomas more rapidly and/or more often than in the general population. The culprit behind this accelerated tumour formation in HNPCC patients is defective mismatch repair (MMR) system unable to correct the small insertion/deletion mutations frequently generated during the replication of repeat sequences. Because the mutations cannot be corrected, malignant transformation is promoted because of accumulating microsatellite mutations (microsatellite instability, MSI) in genes involved in cell growth, differentiation, and apoptosis.

HNPCC is caused by monoallelic germline mutations of MMR genes and subsequent somatic hits that render the MMR system inactive. Germline mutations in MMR genes *MLH1*, *MSH2*, *PMS2*, and *MSH6* have been found causative in HNPCC (Fishel *et al.* 1993, Leach *et al.* 1993, Bronner *et al.* 1994, Nicolaides *et al.* 1994, Papadopoulos *et al.* 1994, Akiyama *et al.* 1997a, Miyaki *et al.* 1997, Worthley *et al.* 2005). A majority of HNPCC cases are caused by mutations in MMR genes *MLH1*

and *MSH2* (Peltomäki and Vasen 2004). Although most mutations (81%) are unique and specific to each family (Peltomäki and Vasen 2004), there are some founder mutations in, for example, the Ashkenazi Jews (Foulkes *et al.* 2002), Finns (Nyström-Lahti *et al.* 1995), and Newfoundlanders (Froggatt *et al.* 1999). Altogether 13% of the mutation positive cases seem to be associated with three of the most common *MLH1* and *MSH2* mutations (Peltomäki and Vasen 2004).

HNPCC is a somewhat misleading designation due to the occurrence of cancer in a number of other tissue types as well. Hence, the name Lynch syndrome has been used in parallel. The syndrome is associated with an increased cancer risk of, for example endometrium, stomach, ovary, ureter/renal pelvis, small bowel, and hepatobiliary tract (Watson and Lynch 1993, Aarnio *et al.* 1995). The risk of developing CRC or endometrial cancer seems to be highest. Cumulative CRC risk values ranging from 54% to 100% have been obtained, partly depending on the mutated gene in question, and endometrial cancer risk estimations have varied between 24% and 62% (Quehenberger *et al.* 2005). Quehenberger *et al.* (2005) suggest, however, that the previously presented CRC risks were overestimations due to the inclusion of mostly high risk families in the analyses.

Phenotypic manifestations in a given HNPCC family seem to be dependent on the mutant MMR gene in question. *MSH2* mutation carriers might be at higher risk of developing cancer at any site than *MLH1* mutation carriers (Vasen *et al.* 1996, Vasen *et al.* 2001). CRC risk associated with *MSH6* seems to be lower than for *MLH1* or *MSH2*, but the risk of endometrial cancer could be higher (Wijnen *et al.* 1999, Hendriks *et al.* 2004). Mutations in *PMS2* have been previously identified mostly in families with features of Turcot syndrome (Peltomäki and Vasen 2004) but were recently identified also in a *bone fide* HNPCC family (Worthley *et al.* 2005). Similar to FAP patients, HNPCC patients with identical mutations can display different disease phenotypes indicating the importance of environmental factors and/or modifier genes.

The role of MMR genes other than *MLH1*, *MSH2*, *PMS2*, and *MSH6* has been considered since a fraction of the HNPCC cases do not seem to associate with these genes (Liu *et al.* 1996). The implications of *PMS1*, *MLH3*, and *MSH3* mutations are however not clear. It has also been postulated that using expression-based mutation detection methods, the known predisposition genes might be found to be causative in a fraction of the previously seemingly mutation negative cases (Renkonen *et al.* 2003).

3.2.1 Amsterdam and Bethesda criteria

Amsterdam criteria I were proposed in 1991 to facilitate the identification of HNPCC families (Vasen *et al.* 1991). In 1999, the Amsterdam criteria I were supplemented to contain the extracolonic features associated with HNPCC (Vasen *et al.* 1999). The Amsterdam criteria II are as follows:

> At least three family members should be affected with an HNPCC-associated cancer (CRC, cancer of the endometrium, ureter, renal pelvis, or small bowel) and in addition, the following criteria should be met:

- one patient should be a first-degree relative of the other two patients
- at least two successive generations should be affected
- at least one of the associated cancers should be diagnosed before age 50
- FAP should be excluded
- pathological examinations should verify the tumours

The Bethesda guidelines were introduced in 1997 to help specifying the tumours that should be tested for MSI, a typical characteristic of HNPCC. The Bethesda criteria are based on age at diagnosis, pathological characteristics, family history, and the presence of extracolonic tumours (Rodriguez-Bigas *et al.* 1997). In 1998, a Bethesda panel consisting of five microsatellite markers BAT25, BAT26, D5S346, D2S123, and D17S250 was introduced to unify MSI-testing (Boland *et al.* 1998). Interpretation of the results obtained with the Bethesda panel was proposed as following:

- if two or more markers display instability, high level MSI (MSI-H) is scored
- if instability is detected with one marker, low level MSI (MSI-L) is scored
- if all markers are stable, microsatellite stability (MSS) is scored

In case of MSI-L, additional markers should be analyzed. MSI-L could be scored if less than 30% of the markers show MSI (Boland *et al.* 1998).

The revised Bethesda guidelines introduced in 2004 suggest that more mononucleotide repeats could be analyzed to increase the detection sensitivity of MSI-H tumours (Umar *et al.* 2004). Furthermore, MSI-L specific markers, such as *MYCL*, could be used. However, the division of CRCs into three distinct groups based on the degree of microsatellite instability has been a matter of debate (Umar *et al.* 2004). This issue will be discussed in the following chapter.

4. Genetic instability in colorectal cancer

CRC arises from a stem cell in the colonic crypt and the earliest identifiable lesion is an aberrant crypt focus (ACF). The progression into carcinoma proceeds via two main pathways, chromosomal instability (CIN) pathway or the microsatellite instability (MSI) pathway (Vogelstein *et al.* 1988, Fearon and Vogelstein 1990, Aaltonen *et al.* 1993, Ionov *et al.* 1993, Thibodeau *et al.* 1993). Typically around 85% of all CRC cases display CIN with gross chromosomal aberrations whereas the remaining 15% display MSI with more subtle aberrations in microsatellites. FAP is a typical example of a CRC syndrome displaying CIN whereas MSI is a typical feature of HNPCC.

A mild MSI phenotype (MSI-L) characterizes a subset of CRC cases and they have been suggested to constitute a distinct entity. Attempts to differentiate MSI-L from MSS tumours according to their clinicopathological or molecular features have produced conflicting results (Thibodeau *et al.* 1998, Jass *et al.* 1999, Halford *et al.* 2002, Laiho *et al.* 2002). It has also been suggested that if enough microsatellite markers are analyzed, almost all CRCs display MSI-L (Halford *et al.* 2002, Laiho *et al.* 2002). However, a subset of MSI-L CRCs seems to display distinct morphological features including a typical saw-toothed or serrated epithelial structure. This epithelial morphology characterizes also a small subset of MSS tumours suggesting the heterogeneity of the so-called serrated pathway (Sawyer *et al.* 2002). The pathway

depicts the progression of hyperplastic polyps to serrated adenomas and further to serrated CRCs (Iino *et al.* 1999, Jass 1999, Hawkins and Ward 2001). High incidence of *BRAF* mutations in serrated polyps proposes the early involvement of *BRAF* in the serrated neoplasia pathway (Chan *et al.* 2003). An expression analysis on serrated CRCs suggests that they form not only morphologically but also biologically distinct class of CRCs (Laiho *et al*, unpublished data).

4.1 Features of colorectal tumours arising through the CIN and MSI pathways

MSS or CIN CRCs are typically aneuploid, left-sided, behave aggressively, and harbour characteristic mutations in, for example *KRAS*, *APC*, and *TP53*. In contrast, MSI CRCs are predominantly diploid, right-sided, poorly differentiated, often characterized by a strong lymphocyte infiltration, and harbour distinct mutations in, for example *TGFβRII* and *BAX* (Smyrk and Lynch 1999). Most of the published studies suggest a better prognosis for patients with MSI CRC (Pawlik *et al.* 2004).

The clinicopathological and molecular differences between proximal and distal CRCs might be partly explained by the different embryological origin and blood vasculature of the colonic components. Differences in the colonic microenvironment might cause a different cellular response to different genetic changes and environmental factors (Bufill 1990, Bardelli *et al.* 2001, Glebov *et al.* 2003).

4.2 Adenoma-carcinoma sequence

The clonal progression of CIN CRCs is depicted by the adenoma-carcinoma sequence (Figure 1) (Vogelstein *et al.* 1988, Fearon and Vogelstein 1990). Multiple alterations in both oncogenes and tumour suppressor genes are required for malignant transformation. Although the alterations often seem to follow a distinct sequence, the number of changes seems to be the critical aspect rather than the actual order of the molecular alterations (Fearon and Vogelstein 1990).

The primary alterations in the adenoma-carcinoma sequence affect *APC*. *APC* mutations or 5q LOH targeting *APC* have been detected already in a large proportion of early adenomas, but equally frequently also in carcinomas (Vogelstein *et al.* 1988, Powell *et al.* 1992, Leslie *et al.* 2002). Mutations in oncogene *KRAS* are found equally frequently in large adenomas and carcinomas, but less frequently in small adenomas suggesting the role of *KRAS* in promoting growth rather than initiating tumorigenesis (Vogelstein *et al.* 1988, Leslie *et al.* 2002). KRas is involved in controlling cellular proliferation and differentiation pathways.

In addition to 5q deletions, 18q LOH is a frequent event in CRC tumorigenesis. Approximately 70% of carcinomas and a smaller proportion of early adenomas harbour 18q deletions (Vogelstein *et al.* 1988, Boland *et al.* 1995, Leslie *et al.* 2002). The most likely target genes for 18q deletions are *SMAD2* and *SMAD4* or the previously proposed *DCC* (Leslie *et al.* 2002). Both SMADs are involved in the TGF-β signalling pathway regulating for example cell growth, differentiation, and apoptosis (Heldin *et al.* 1997). Loss of chromosome 17p that targets *TP53* is also frequently observed in CRCs (Vogelstein *et al.* 1988, Leslie *et al.* 2002). *TP53*

alterations are more frequent in carcinomas than in adenomas indicating the possible association of defective p53 in adenoma-carcinoma transition. p53 maintains genomic stability by controlling cell cycle progression and apoptosis.

Progression into metastatic CRC requires additional molecular changes in order for the tumour to invade surrounding tissues, be released into the circulation, and colonize new sites such as the liver. The exact molecular events controlling CRC metastasis are not fully known. The involvement of, for example, PRL-3 and multiple factors in the WNT/β-catenin pathway has been suggested (Brabletz *et al.* 2001, Rivat *et al.* 2003, Zeng *et al.* 2003, Pai *et al.* 2004, Dhawan *et al.* 2005). Furthermore, recent protein expression comparisons between primary CRCs and hepatic metastases identified 9 differentially expressed proteins, such as proapolipoprotein and beta-globin, with potential involvement in carcinogenesis and metastatic processes (Yu *et al.* 2004).

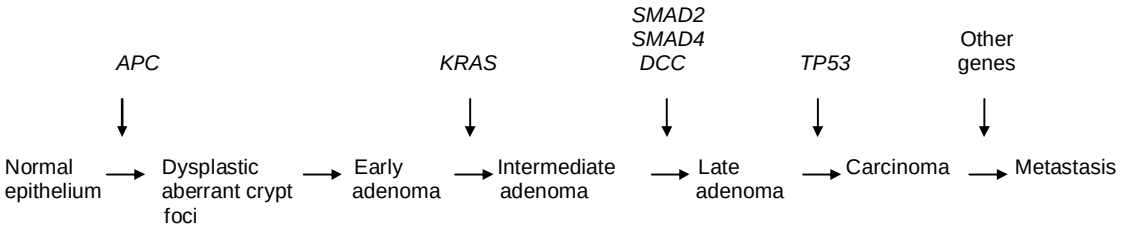

Figure 1. The sequential progression model of CRC tumorigenesis. Modified from Fearon and Vogelstein (1990) and Kinzler and Vogelstein (1996).

The exact mechanisms initiating CIN are not well known and over 100 candidate genes have been proposed in yeast (Lengauer *et al.* 1998, Jallepalli and Lengauer 2001). Many of the candidate genes are involved in cell cycle checkpoints and some of these genes, for example *RAD17* and *AURKA* have been linked to tumorigenic processes based on the observed overexpression in CRCs (Bischoff *et al.* 1998, Bao *et al.* 1999). Furthermore, *BUB1* and *BUBR1* mutations have been observed in CRCs (Cahill *et al.* 1998). In addition to the checkpoint genes, telomere integrity is known to suppress CIN.

4.3 Microsatellite instability pathway

An alternative pathway of tumorigenesis is the microsatellite instability (MSI) pathway caused by defective DNA mismatch repair system (MMR) (Figure 2). Cells with inactive MMR genes, mostly *MLH1* and *MSH2*, have approximately 100-fold higher mutation frequencies (Ionov *et al.* 1993, Strand *et al.* 1993, Bhattacharyya *et al.* 1994, Shibata *et al.* 1994). The mutations following MMR inactivation can affect genes important in for example cell cycle regulation and DNA repair. These events are required for malignant transformation and occur before visible tumour formation (Tsao *et al.* 2000, Shibata 2001). The number of microsatellite mutations can reflect the pattern of tumour progression and serve as molecular tumour clocks (Shibata *et al.* 1996).

There are several potential mutational targets in MMR deficient cells since microsatellites are abundantly found in coding and non-coding regions of the human genome (Toth *et al.* 2000). Microsatellite mutations have been observed in a number of putative MSI target genes and the tumorigenic implications of these mutations have been presented in some cases, such as *TGFβRII* and *BAX* (Wang *et al.* 1995, Ionov *et al.* 2000). Current knowledge on MSI target genes will be discussed in more detail in section 4.3.2.

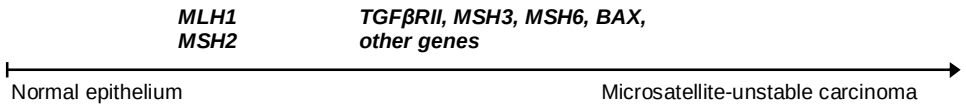

Figure 2. Microsatellite-unstable colorectal cancer progression.

4.3.1 DNA mismatch repair, MMR

MMR system maintains genomic stability mainly by correcting single base mismatches and small insertion/deletion loops (IDLs) that escape the nucleotide selection and proofreading activities of DNA polymerases during DNA replication. MMR proteins are also involved in: 1. inhibiting recombination of divergent DNA sequences, 2. correcting errors caused by mutagens such as alkylating agents, 3. participating in double-strand break repair as well as transcription coupled repair together with NER proteins, and 4. regulating DNA damage signalling and apoptosis (Harfe and Jinks-Robertson 2000).

MMR mechanisms are highly conserved in eukaryotes and prokaryotes. The *E.coli* MMR system is the best characterized and multiple MutS, MutL, and MutH proteins are known to function in mismatch recognition, mediating recognition and repair processes, and targeting repair to the newly synthesized DNA strand, respectively. All eukaryotic organisms studied to date have been shown to possess multiple MutS homologs (MSH proteins) and MutL homologs (MLH proteins) but no convincing MutH homologs have been identified (Harfe and Jinks-Robertson 2000).

In human error repair systems, mismatches are recognized by the MSH2/MSH6 complex (MutSα) and IDL recognition is mediated through MutSα and the MSH2/MSH3 (MutSβ) complex. Subsequently, MutS complexes interact with MutL complexes MLH1/PMS2 (MutLα) and occasionally with MLH1/MLH3 or MLH1/PMS1 (MutLβ). The MutL complexes mediate the error recognition to the downstream events involved in error repair (Kolodner and Marsischky 1999, Marra and Schar 1999, Harfe and Jinks-Robertson 2000) (Figure 3). The mechanism used to recognize the newly synthesized strand in human cells is not known. In *E.coli* the synthesized strand is transiently undermethylated but in humans the methylation patterns are irregular. It has been postulated that replication of the lagging strand as Okazaki fragments and formation of nicks during the process could aid in the discrimination of the DNA strands. The leading strand could be identified by the growing 3’ end. The role of PCNA in strand recognition has also been suggested due to the observed interaction with MMR proteins.

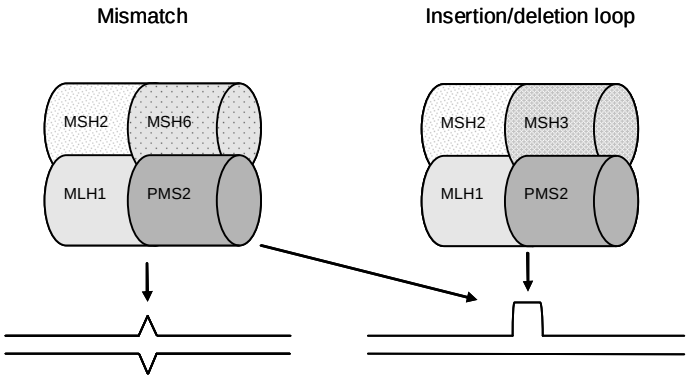

Figure 3. Principal proteins involved in DNA mismatch repair in humans. Interactions between MutS (MSH2/MSH6/MSH3) and MutL (MLH1/PMS2) homologs mediate the recognition of base mismatches and insertion/deletion loops. Modified from Marra and Schär (1999).

The excision of the newly synthesized DNA fragment containing the error involves DNA exonuclease(s) and at least *EXO1* has been linked to human MMR system (Tishkoff *et al.* 1997). Resynthesis of the excised strand is mediated by DNA polymerases and replication factors. DNA ligases subsequently complete the MMR processes.

The functional redundancy of MMR proteins in error recognition has an impact on the MSI status of tumours with mutations in different MMR genes. MLH1 or MSH2 deficient cells display high degree of MSI because of the involvement of these proteins in all MutL and MutS complexes, respectively. However, the variable degree of MSI in MSH6 deficient tumours is explained by the ability of MSH2/MSH3 complex to compensate for the IDL correction.

Interestingly, homozygous germline mutations of MMR genes *MLH1* and *MSH2* have been linked to neurofibromatosis type 1 and early onset haematological malignancy (Ricciardone *et al.* 1999, Wang *et al.* 1999, Whiteside *et al.* 2002). Defective MMR has been shown to render *NF1* susceptible to somatic frameshift mutations (Wang *et al.* 2003). Furthermore, a woman heterozygous for two different *MLH1* missense mutations developed breast cancer at the age of 35. The mild phenotype observed in the patient might be due to a residual MLH1 activity from the other allele (Hackman *et al.* 1997).

4.3.2 MSI target genes

Cells with defective MMR system accumulate microsatellite mutations in both noncoding and coding regions of the genome. If genes involved in controlling cell growth or maintaining genomic integrity are affected, tumour progression may be promoted. Generally these consequential mutations occur in coding microsatellites, leading to the formation of altered protein products or their degradation via nonsense-mediated decay (NMD). In addition to coding region mutations, mutations occurring

in regulatory non-coding regions may also have an impact on the gene functions. Usually, however, non-coding microsatellite mutations are thought to be inconsequential background events.

High mutation frequency has been considered as a useful criterion for defining true MSI target genes (Boland *et al.* 1998). Other criteria suggested by Boland *et al.* (1998) included functional evidence, biallelic inactivation, role in a growth suppressor pathway, and inactivation of the same growth suppressor pathway in MSS tumours. Some of these criteria have been questioned since. Monoallelic mutations might suffice in case of haploinsufficiency or dominant negative effect (Liu *et al.* 2000, Yamamoto *et al.* 2000). The affected genes might also have roles other than tumour suppressor functions, as exemplified by somatic microsatellite mutations observed in MMR genes *MSH3* and *MSH6*. Of note, somatic inactivation of *MSH3* and *MSH6* has been shown to further increase the instability phenotype (Akiyama *et al.* 1997b, Baranovskaya *et al.* 2001, Duval *et al.* 2001). Furthermore, the pathways involved in MSS and MSI tumours seem to differ. Hence, the most widely accepted criteria for defining true MSI target genes have been high mutation frequency and functional evidence.

Duval and Hamelin (2002) have suggested that genes mutated in MSI tumours could be divided into four groups based on the observed mutation frequency and the roles of the protein products. Survivor genes (1.) with essential cellular roles would be rarely mutated due to negative selection pressure. Hibernator genes (2.) with unimportant functions would accumulate mutations at the background level. The effect of co-operator gene (3.) mutations would be emphasized when occurring in parallel with mutations in other co-operator genes. Their mutation frequency would therefore vary depending on the context in which they occur. The highest mutation frequencies would be observed in transformator genes (4.) because the mutations confer a growth advantage (Duval and Hamelin 2002).

The mutation frequencies of true MSI target genes can thus vary extensively depending on their role in the affected cells. Furthermore, the mutation frequency of a given microsatellite seems to be dependent on the repeat type and length, as well as yet unknown influence from the surrounding sequence (Zhang *et al.* 2001, Suzuki *et al.* 2002, Vilkki *et al.* 2002). It is important to study the general background mutation frequency of MSI tumours to obtain information on the possible sequence elements affecting replication fidelity. In addition, functional evidence has to be provided to prove the importance of a given gene in MSI tumorigenesis (Perucho 2003).

To date, MSI tumours have been analyzed for microsatellite mutations in a large number of genes, but only a handful of genes have been shown to contribute to the MSI pathway of tumour progression. The first gene identified with somatic microsatellite mutations was *TGFβRII*. Subsequently, around 90% of MSI CRCs have been shown to harbour *TGFβRII* mutations, often in a biallelic form (Parsons *et al.* 1995). Functional studies have indicated a loss of *TGFβRII* tumour suppressor function as a result of these repeat mutations (Wang *et al.* 1995). Furthermore, *TGFβRII* mutations seem to be an early event in MSI tumorigenesis (Grady *et al.* 1998, Duval *et al.* 2001). In addition to *TGFβRII*, mutations in for example proapoptotic factor *BAX* and growth factor receptor *IGFRII* have been shown to contribute to MSI tumorigenesis (Souza *et al.* 1999, Ionov *et al.* 2000).

5. Prospects of CRC studies

The thus far identified hereditary polyposis and non-polyposis syndromes together account for approximately 5% of all CRC cases (Burt and Neklason 2005). Novel CRC susceptibility genes are still to be identified and evidence for putative loci on 9q and 15q has been obtained (Tomlinson *et al.* 1999, Jaeger *et al.* 2003, Wiesner *et al.* 2003). The underlying high penetrant mutations are most likely rare in the general population and are therefore causative in a minority of all CRC cases. In comparison, the more frequent low-penetrance susceptibility alleles might contribute to a larger proportion of all CRC cases.

Low-penetrance genes confer a modestly increased cancer risk and mild clustering of disease in the family. Therefore, the identification of causative genetic alterations needs to be pursued by methods other than linkage, a successfully utilized method in the identification of high-penetrance susceptibility genes (Houlston and Peto 2004). Allelic association analyses of low-penetrance genes can be performed with much smaller sample size than linkage analyses (Risch and Merikangas 1996, Camp 1997) and the number of cases needed for association analysis would be even lower if affected relatives were included in the analyses (Houlston and Peto 2003). In fact, association analysis has been the most common method in identifying low-penetrance susceptibility alleles (Houlston and Tomlinson 2001).

5.1 Candidate low-penetrance CRC susceptibility alleles

To date, multiple association analyses have been performed to estimate the CRC risk conferred by sequence variants in different genes. With increasing numbers of association analyses, some pooled analyses have also been undertaken to increase the number of samples per polymorphism analyzed (Houlston and Tomlinson 2001, de Jong *et al.* 2002, Chen *et al.* 2005).

One meta-analysis suggested *APC* I1307K as a candidate low-penetrance susceptibility allele (Houlston and Tomlinson 2001). The *APC* I1307K change is more frequent in CRC patients than in healthy controls in Ashkenazi Jewish populations. The amino acid change creates a hypermutable poly-A tract that is easily affected by a deletion or amplification event (Laken *et al.* 1997). In addition to *APC* I1307K, an association between *HRAS*-VNTR as well as *MTHFR* variants and CRC risk has been suggested (Houlston and Tomlinson 2001, de Jong *et al.* 2002). *HRAS* oncogene is involved in the mitogenic signal transduction and differentiation. Variable number of tandem repeats (VNTR) located ~1kb downstream of *HRAS* has been shown to interact with transcriptional regulatory elements and control the expression levels of nearby genes. MTHFR in turn catalyzes the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate which is the primary circulatory form of folate. The availability of folate and other methyl group donors has been suggested to affect methylation (Kim *et al.* 1996).

The role of *GSTT1* genotype and *NAT2* phenotype has also been suggested in two meta-analyses (de Jong *et al.* 2002, Chen *et al.* 2005). GSTT1 and NAT2 are carcinogen metabolism enzymes mediating defenses against a number of chemical carcinogens and environmental toxins (Upton *et al.* 2001, Hayes *et al.* 2005).

5.2 Identification of novel low-penetrance susceptibility alleles

A number of association studies have been performed to identify low-penetrance CRC susceptibility alleles. The relative risk attributed to a given polymorphism varies between the analyses and the meta-analyses discussed above. For example, the role of *MTHFR* alleles in CRC was supported by Houlston and Tomlinson (2001) and de Jong *et al.* (2002) but questioned by Chen *et al.* (2005). This may be due to different composition of studies included in the meta-analysis. Furthermore, there may be differences in the variables, for example ethnicity, gender, and tumour localization, that were taken into account when assessing the relative risks. When assessing the importance of a given polymorphism based on current literature, it has to be remembered that publication bias might have hindered the publication of negative results.

The limitation of association analysis is that in the absence of a good candidate polymorphism, the number of associations tested can be prohibitively large and the problem of multiple testing also rises. The number of markers needed for association analysis with sufficient power is dependent on the extent of linkage disequilibrium (LD) between a marker and a causal locus. Typically LD extends ~50-60kb in the human genome but even further in populations with homogeneous gene pool (Reich *et al.* 2001). The extent of LD can also be higher in young admixed populations that have been formed by gene flow from two or more genetically distinct founding populations (Stephens *et al.* 1994). Mixing of genetic material creates LD between linked but also unlinked loci that have different allele frequencies in the founding populations. Therefore, methods to discriminate between spurious and true associations need to be utilized (Pfaff *et al.* 2001).

Identification of low-penetrance cancer susceptibility genes can benefit from advances in gene-mapping techniques in mice. Crosses between susceptible and resistant strains have produced information on several mouse tumour susceptibility loci (Demant 2003). Furthermore, studying the interactions between multiple polymorphisms is much more straightforward in mice (Mao and Balmain 2003). However, information obtained from mouse studies might not be applicable in humans due to inter-species differences. This is exemplified by the discrepancies between mouse and human studies on modifying factors of FAP, as discussed in section 3.1.1.

For the reasons stated above, the identification of low-penetrance susceptibility alleles is understandably a laborious effort. Since these alleles can occur frequently in the general population, they can have a vast impact on general public health. It is therefore important to study the genotype-phenotype correlations and utilize novel technologies, for example dense SNP maps, SNP arrays, and array-CGH in the identification of the low-penetrance susceptibility alleles. Furthermore, prior information on potentially interesting genomic positions or gene functions would decrease the number of associations tested. A candidate genomic region may be assigned by utilizing chromosomal clues of for example commonly deleted or amplified genomic regions. Emerging novel technologies combined with novel information from cellular pathways can facilitate the identification of low-penetrance susceptibility alleles.

AIMS OF THE STUDY

Novel genes and loci associated with microsatellite stable colorectal cancer predisposition

I Identification of novel susceptibility loci in familial colorectal cancer.

II Role of *AURKA* in colorectal cancer susceptibility: a candidate gene selected based on study I.

Causes and characteristics of colorectal cancers displaying microsatellite instability

III Evaluation of the putative role of *MLH3* in microsatellite-unstable colorectal cancer.

IV Assessment of *SEMG1* as a candidate microsatellite instability target gene and evaluation of the general background mutation frequency of microsatellite-unstable colorectal cancers.

MATERIALS AND METHODS

1. Colorectal cancer patients and DNA samples

The colorectal carcinoma and corresponding normal tissue samples utilized in studies I-IV were chosen from sample series collected since 1994 (Aaltonen *et al.* 1998, Salovaara *et al.* 2000, Laiho *et al*. unpublished data). The series contain samples from altogether 1579 patients (at the time of the thesis' last project). Information on the family background was obtained from official population registries and cancer occurrence in the family from the Finnish Cancer Registries. A majority (86%) of the patients have no first-degree relatives affected with CRC. Familial cases were defined as having at least one first-degree relative affected with CRC. Most of the families display a mild CRC clustering. The samples and patient information were obtained with approval from the Ethics committees of the Department of Medical Genetics, University of Helsinki, Helsinki University Central Hospital, and Hospital District of Helsinki and Uusimaa.

The normal tissue DNA was extracted from blood or normal colonic epithelium distant from the tumour margins. The fresh frozen tumour tissue samples were histologically evaluated by a pathologist before DNA extraction. Each tumour sample undergoes an MSI analysis and of the currently established MSI cases, 22% (35/162) harbour germline *MLH1*, *MSH2*, or *MSH6* mutations (Aaltonen *et al.* 1998, Salovaara *et al.* 2000, Loukola *et al.* 2001).

1.1 Common regions of allelic imbalance (I)

Normal and tumour tissue samples were analyzed with microsatellite markers to detect common genomic regions of loss of heterozygosity/amplification (allelic imbalance, AI). Samples from altogether 99 familial and 186 sporadic Finnish CRC patients were utilized in study I. Polyposis cases (FAP, PJS, and JP) as well as HNPCC and other cases with MSI were excluded from the study. All tumour samples displayed at least 50%, and typically ~75%, of carcinoma tissue.

Additional normal tissue and CRC samples obtained from collaborators in Denmark and UK were analyzed for AI with microsatellite marker D20S178. Samples from altogether 18 familial and 24 sporadic cases from Denmark, and 67 familial and 96 sporadic cases from the UK were included in study I.

1.2 Genetic analyses of *AURKA* (II)

Normal and tumour tissue DNA samples from 125 familial and 110 sporadic Finnish MSS CRC cases were analyzed. Polyposis cases (FAP, PJS, JP, and *MYH*-associated polyposis) as well as HNPCC and other cases displaying MSI were excluded from the study. A majority (68%) of the cases had been analyzed for AI in study I. A majority of the tumour samples (91%) displayed at least 60% carcinoma tissue.

In addition, samples from 94 anonymous cancer-free blood donors provided by the Finnish Red Cross Blood Transfusion Centre were utilized.

1.3 *MLH3* mutation analysis (III)

Normal tissue DNA samples from 30 Finnish CRC patients displaying MSI were included in study III. The samples were chosen based on the molecular features suggested to be causally associated with *MLH3* defects (Wu *et al.* 2001). Namely, samples from 17 familial and 6 sporadic CRC patients were included, all of which were stable when analyzed with mononucleotide repeat marker BAT26. In addition, samples from all 7 available mutation-negative HNPCC patients were analyzed. The mutation negative HNPCC samples were provided by Professor Päivi Peltomäki. A tumour tissue sample from a patient with germline *MLH3* Met809Val mutation was analyzed. The tumour tissue sample displayed 70% carcinoma tissue.

In addition, samples from 713 anonymous cancer-free blood donors provided by the Finnish Red Cross Blood Transfusion Centre were utilized.

1.4 Genetic analyses of *SEMG1* and intergenic T9 repeats (IV)

Tumour tissue samples from 146 MSI CRC patients were utilized in study IV. The respective normal tissue samples available from 144 patients were also analyzed. In addition, MSS CRC samples from 238 patients and normal tissue samples from 50 cancer-free blood donors from the Finnish Red Cross Blood Transfusion Centre were utilized.

Most of the tumour samples used in study IV (92%) contained ≥ 50% carcinoma cells. Altogether 20% of the MSI CRC patients included in study IV harbour germline *MLH1* or *MSH2* mutations (Aaltonen *et al.* 1998, Salovaara *et al.* 2000, Loukola *et al.* 2001).

2. Cancer cell lines (IV)

In study IV, we utilized 31 cancer cell lines originating from colon (n=28), prostate (n=2), and endometrium (n=1). Of the CRC cell lines, 10 displayed MSI and 18 were MSS. Both prostate cancer cell lines as well as the endometrial cancer cell line displayed MSI.

3. Analysis methods

3.1 Common regions of allelic imbalance (I)

3.1.1 Microsatellite marker analysis

The AI frequency of familial and sporadic tumour sets was determined by using microsatellite marker analysis. Normal and tumour tissue samples from 29 familial and 75 sporadic CRC patients were analyzed with 372 microsatellite markers (ABI PRISM Linkage Mapping Set MD-10, P/N 450067, 10 cM density, Applied Biosystems (AB) Division, Foster City, CA). The DNA samples were amplified using fluorescent labelled primers and run on an ABI PRISM 377 DNA Sequencer (AB) according to manufacturer's instructions. The fragments were analyzed using GeneScan 3.1 and Genotyper 2.5 programs (AB). The samples displaying AI were determined using the previously published mathematical model (Canzian *et al.* 1996). In effect, AI was scored if the other allele emitted at least 40% less fluorescence. From this primary dataset of 372 markers, those markers showing the most promising association of AI and familial CRC were further analyzed using additional normal and tumour tissue samples from 70 familial and 111 sporadic CRC patients.

The samples obtained from collaborators in Denmark and UK were analyzed with one specific microsatellite marker, D20S178, to determine the AI frequencies in the familial and sporadic groups.

3.1.2 Comparative genomic hybridization

Data from a previously performed, unpublished CGH analysis was utilized to obtain information on the mechanisms underlying AI. Tumour samples from 26 of the familial patients were fluorescent labelled and co-hybridized with a differently labelled normal tissue DNA to metaphase chromosomes.

3.1.3 Quantitative PCR

Quantitative PCR (qPCR) experiments with TaqMan chemistry were performed on tumour tissue DNA samples from 26 familial and 26 stage and sex matched sporadic cases. The familial cases were already included in the CGH analysis. DNA was amplified using primers for microsatellite marker D20S178 and the PCR products were run on a GeneAmp 5700 Sequence detection system (AB). The relative copy numbers on chromosome 20 were determined relative to three normal tissue samples. As a reference, we used microsatellite markers D10S586, D11S1315, and D21S1904 from genomic regions stable according to the AI analysis.

3.2 Genetic analyses of *AURKA* (II)

3.2.1 *AURKA* sequencing

The *AURKA* coding region and exon-intron borders were sequenced in the normal tissue DNA from 10 familial CRC patients. The DNA samples were amplified and run on an ABI3730 automatic DNA sequencer according to manufacturer's instructions (AB). The frequency of the observed sequence variants was assessed by sequencing a

larger normal tissue sample set from additional familial and sporadic CRC patients. The occurrence of the sequence variants was also assessed in Finnish cancer-free controls by sequencing.

3.2.2 Amplification analysis by sequencing and fragment analysis

After identifying *AURKA* sequence variants, we analyzed the possible preferential amplification of either allele in the heterozygous individuals by sequencing or fragment analysis. The PCR fragments were run on an ABI3730 automatic DNA sequencer according to manufacturer's instructions (AB). The sequence graphs were analyzed with Chromas 2.21 (Technelysium Pty Ltd, Queensland, Australia) and the fragment analysis graphs with GeneMapper 3.0 (AB). Amplification scoring was performed according to Canzian *et al.* (1996).

3.3 *MLH3* mutation analysis (III)

3.3.1 *MLH3* sequencing and denaturing high-performance liquid chromatography (dHPLC) analysis

MLH3 sequence variants were screened in the normal tissue DNA of 30 CRC patients by sequencing the coding region and exon-intron borders. The PCR fragments were run on an ABI3100 according to manufacturer's instructions (AB). The possible occurrence of the detected sequence variants was assessed in Finnish cancer-free controls by dHPLC. For dHPLC, the PCR fragments were denatured and reannealed, and the heteroduplex analysis was performed using automated HPLC instrumentation with an Agilent 2G experimental dsDNA 2.1 X 75 mm 3.5µ column (Agilent, Palo Alto, CA). Samples displaying aberrant dHPLC graphs were sequenced to verify the alterations.

The previously published *MLH3* germline alterations (Wu *et al.* 2001) were also screened in the Finnish cancer-free controls by dHPLC.

3.4 Genetic analyses of *SEMG1* and intergenic T9 repeats (IV)

3.4.1 *SEMG1* sequencing and denaturing high-performance liquid chromatography (dHPLC) analysis

The occurrence of *SEMG1* IVS2+5(T)9 mutations was determined by sequencing and dHPLC. Tumour samples from 121 MSI and 21 MSS CRC patients were analyzed by sequencing and a set of 93 MSI and 230 MSS CRCs were analyzed by dHPLC. Some of the tumour samples were analyzed using both methods to confirm their *SEMG1* status. Altogether 146 MSI and 238 MSS tumour samples were analyzed. As controls, we assessed the normal tissue DNA of 144 MSI CRC patients and samples from 50 Finnish cancer-free controls by sequencing.

The coding region and exon-intron borders of *SEMG1* were screened for other possible sequence variants by sequencing (exons 2 and 3) and dHPLC (exon 1). We

analyzed 20 MSI CRCs, 20 MSS CRCs, and 7 cell lines including 5 MSI CRC, 1 MSS CRC, and 1 MSI endometrial cancer cell line. Samples producing aberrant peaks in dHPLC were sequenced.

3.4.2 Evaluation of *SEMG1* mRNA levels by quantitative PCR

The possible effect of the *SEMG1* IVS2+5(T)9 mutation at the RNA level was assessed by qPCR with SYBR Green chemistry. RNA from six CRC cell lines was extracted and the generated cDNA was amplified using cDNA-specific primers. The PCR amplifications were monitored using a GeneAmp 5700 Sequence detection system (AB). The *SEMG1* expression levels were determined relative to the *GAPDH* expression levels using the formula $2^{-\Delta\Delta Ct}$ (ABI PRISM 7700 Sequence Detection System User Bulletin #2, AB).

3.4.3 Evaluation of SEMG1 protein levels by Western blotting

The putative effect of the *SEMG1* IVS2+5(T)9 mutation at the protein level was assessed by Western blotting. Total protein was extracted from cell lysates and cell culture media of six CRC cell lines. From each cell line, equal amounts of total protein were loaded into a 10% Tris-HCl gel. Total protein was blotted on a polyvinyl membrane and SEMG1 was detected using a polyclonal SEMG1 antibody (Biocarta, San Diego, CA). Total protein staining and the GAPDH protein levels were used as running controls. The protein band intensities were measured using a FluorChem8800 imaging system (Alpha Innotech, San Leandro, CA).

3.4.4 Sequencing of intergenic T9 repeats

The general background mutation frequency of MSI CRCs was assessed by sequencing a panel of 29 MSI CRCs for mutations in ten intergenic T9 repeats. To exclude polymorphisms, the respective normal tissue samples of patients harbouring mutations in the T9 repeats were sequenced.

RESULTS

1. Common regions of allelic imbalance (I)

After analyzing the 29 familial and 75 sporadic Finnish CRC cases with a genome-wide panel of 372 microsatellite markers, the AI frequencies for each marker in familial and sporadic patient groups were calculated. Markers indicating an association between familial cancer and AI were chosen for further analysis. Namely, 10 markers from separate genomic areas were chosen, 7 of which showed at least 10% units more AI in the familial group than in the sporadic group and the adjacent marker showed such a difference as well. The 7 markers selected from each cluster were the ones showing the biggest difference in the AI frequencies between familial and sporadic groups. Furthermore, 3 markers were chosen that harboured near 30% units more AI in the familial group although the adjacent markers showed equal AI frequencies in both groups.

In the second round of analysis, new samples from 59 familial and 64 sporadic cases were analyzed with the 10 markers. Combining results from the two rounds indicated an association between AI on chromosome 20 and familial cancer. The remaining available familial cases (n=11) and a set of sporadic cases (n=47) were then analyzed with the chromosome 20 markers. Altogether, the 99 familial and 186 sporadic cases indicated a significant difference ($p \leq 0.01$) in the AI frequencies between familial and sporadic patients on 20q13 marker D20S178.

Subsequently, the frequency of AI on D20S178 was determined in sample sets from other populations. The small Danish set of 18 familial and 24 sporadic cases showed a significant difference (p=0.002) between the two groups. However, AI was observed equally frequently in the groups consisting of 67 familial and 96 sporadic cases from the UK (p=0.68).

Based on CGH analysis performed on 26 familial CRCs, amplifications rather than deletions were underlying AI on chromosome 20q. Amplifications on 20q were found in 85% of the tumours. qPCR experiments on 26 familial and 26 sporadic cases indicated that amplifications were equally common in familial (72%) and in sporadic (73%) CRCs.

2. Analyses of *AURKA* (II)

We sequenced the coding region and exon-intron borders of *AURKA*, a candidate CRC susceptibility gene on chromosome 20q13. Sequencing of normal tissue DNA samples from 10 Finnish familial CRC patients revealed 3 missense variants (Phe31Ile, Val57Ile, and Met373Val) and a 61bp duplication in intron 4 (IVS4+2-62dup). To determine the allele frequencies of these sequence variants, we analyzed altogether 125 familial and 110 sporadic samples for the occurrence of Phe31Ile, Val57Ile, and IVS4+2-62dup variants. The Met373Val variant was screened in 42 familial and 46 sporadic cases. The allele frequencies of each sequence alteration

were similar in the familial and sporadic patient groups, as well as in the 94 cancer-free controls (Table 1).

Table 1. *AURKA* allele frequencies in the Finnish CRC patients and cancer-free controls.

Sequence alteration	Allele frequency % (no. of cases)		
	Familial CRCs	Sporadic CRCs	Cancer-free controls
IVS4+2-62dup	18/82	17/83	20/80
(++/+-/--)	(3/39/83)	(2/33/75)	(6/26/62)
91T>A, Phe31Ile	74/26	70/30	72/28
(TT/TA/AA)	(70/44/11)	(52/50/8)	(46/43/5)
169G>A, Val57Ile	82/18	83/17	80/20
(GG/GA/AA)	(83/39/3)	(75/33/2)	(62/26/6)
1683A>G, Met373Val	91/9	95/5	90/10
(AA/AG/GG)	(34/7/0)	(40/4/0)	(76/18/0)

The allele specific amplification patterns of the four sequence alterations were assessed in the heterozygous individuals. The amplification frequencies of 91T>A (Phe31Ile) alleles were significantly different from the expected (p=0.03) (Table 2), whereas the other variants did not show allele specific amplification patterns. The amplification of 91A was more pronounced in the familial group (p=0.03) than in the sporadic group (p=0.36) (Table 2). Subsequently, the clinicopathological characteristics of the familial and sporadic CRC patients were compared with respect to the genotypes at each variant. The only correlation observed was a trend between younger age at diagnosis in the group consisting of familial heterozygous carriers and homozygous individuals for the more rare 91A allele (p=0.06).

Table 2. Frequencies of *AURKA* 91T>A genotypes in the Finnish familial and sporadic CRC cases. The proportion of amplified alleles in the tumour samples is depicted.

Samples analyzed (no. of samples)	Genotype frequencies TT/TA/AA	A amplified	T amplified	p
Familial CRCs (125)	70/44/11	20/44 (45%)	5/44 (11%)	0.03
Sporadic CRCs (110)	52/50/8	18/50 (36%)	11/50 (22%)	0.36
Combined (235)	122/94/19	38/94 (40%)	16/94 (17%)	0.03

We assessed aneuploidy of the tumours by combining results from studies I and II. Aneuploidy was scored in a given chromosome arm if AI was detected in at least 50% of the informative markers. In chromosomes 9q, 11q, 13q, and 20q, more AI was observed in tumours from 91T>A heterozygous than 91T homozygous individuals with p-values of 0.04, 0.02, 0.01, and 0.007, respectively. However, AI on 2q was observed more frequently in tumours from 91T homozygous individuals with a p-value of 0.01.

3. *MLH3* mutation analysis (III)

Sequencing of the *MLH3* coding region and exon-intron borders in the normal tissue DNA of 30 Finnish CRC patients revealed 5 missense variants: Phe390Ile, Val420Ile, Glu624Gln, Met809Val, and Asp1105Glu. The allele frequency of each variant was 1.7% (Table 3). Variants Phe390Ile, Val420Ile, Glu624Gln, and Asp1105Glu were detected in the Finnish cancer-free controls with allele frequencies of 0.6%, 1.1%, 1.1%, and 0.2%, respectively. The Met809Val variant found in a familial case with mild MSI was not found in the 713 cancer-free controls analyzed by dHPLC (Table 3). No LOH was observed in the tumour tissue of the patient harbouring the Met809Val change.

In the cancer-free control samples, 3 missense variants Asp385His, Thr942Ile, and Ser966Pro were detected with allele frequencies of 0.6%, 0.9%, and 0.9%, respectively (Table 3). These variants were not identified in the CRC patients' normal tissue DNA.

Table 3. Allele frequencies of *MLH3* germline variants identified in Finnish CRC patients and cancer-free controls.

***MLH3* variant**	**No. of CRC patients (allele frequency)**	**No. of cancer-free controls (allele frequency)**
Asp385His	0/30	1/90 (0.6%)
Phe390Ile	1/30 (1.7%)	1/90 (0.6%)
Val420Ile	1/30 (1.7%)	2/90 (1.1%)
Glu624Gln	1/30 (1.7%)	4/180 (1.1%)
Met809Val	1/30 (1.7%)	0/713
Thr942Ile	0/30	3/174 (0.9%)
Ser966Pro	0/30	3/174 (0.9%)
Asp1105Glu	1/30 (1.7%)	1/268 (0.2%)

The previously published 10 *MLH3* variants were also screened in the cancer-free controls with an emphasis on the frameshift mutation 2578delA (Wu *et al.* 2001). One of the published variants, Glu624Gln, was observed in one normal tissue sample from a Finnish CRC patient and was also found in the cancer-free controls with an allele frequency of 1.1%. The other sequence variants detected by Wu *et al.* (2001) were not detected in the Finnish cancer-free controls. A total of 707 control samples were analyzed for the 2578delA alteration.

4. Genetic analyses of *SEMG1* and intergenic T9 repeats (IV)

SEMG1 intron 2 T9 repeat [IVS2+5(T)9] mutations were observed in 51% (75/146) of the MSI tumours analyzed. None of the successfully analyzed respective normal tissue samples (n=141) displayed the mutations. Furthermore, no alterations were detected in the 238 MSS tumours or in the 50 cancer-free controls. In the 31 cell lines studied, only MSI cell lines were affected with a mutation frequency of 62% (8/13) (Table 4).

A possible second hit was screened by sequencing the *SEMG1* coding region and exon-intron borders in 20 MSI CRCs, 20 MSS CRCs, and 7 cell lines including 5 MSI CRC, 1 MSS CRC, and 1 MSI endometrial cancer cell line. We observed 5 sequence variants C1386T, Arg447His, Arg457Gln, C1417T, and IVS2+30A>G that were present also in the respective normal tissue DNA. The sequence alterations were also found in the cancer-free controls with similar allele frequencies.

The putative effects of the IVS2+5(T)9 mutations were analyzed at the RNA and protein levels. No differences were detected in the RNA or protein levels or sizes between the six CRC cell lines with different *SEMG1* IVS2+5(T)9 status.

The clinicopathological characteristics of the MSI CRC patients with (n=75) and without (n=71) the IVS2+5(T)9 mutations were compared. No correlations were observed between the mutation status and age at onset, sex distribution, location or staging of the tumour, or presence of germline MMR mutation.

Subsequently, the general background mutation frequency of MSI CRCs was assessed by sequencing 29 MSI CRCs for mutations in ten intergenic T9 repeats. Polymorphisms were excluded by screening the respective normal tissue samples from those patients displaying mutations in the tumour samples. After analyzing the tumour and normal tissue samples, we detected mutation frequencies of 0% for AL161657, 7% for AC103870, 10% for AL445240, 10% for AC108706, 27% for AP002759, 28% for AP002801, 29% for AC008163/1, 30% for AC008163/2, 27% for AC004006, and 86% for AC027013. Due to the high mutation frequency in the T9 repeat in AC027013, we extended our analysis to include a total of 145 MSI CRCs and respective normal tissue samples. Mutation frequency of 70% (102/145) was observed in the tumour samples, whereas none of the 143 available and successfully analyzed normal tissue samples displayed the mutation (Table 4).

Table 4. Frequency of *SEMG1* and AC027013 T9 mutations.

Samples analyzed	**Mutation frequency, % (no. of cases)**	
	SEMG1	**AC027013**
MSI CRCs	51 (75/146)	70 (102/145)
MSS CRCs	0 (0/238)	nd
Cancer cell lines		
MSI CRC	70 (7/10)	nd
MSS CRC	0 (0/18)	nd
MSI prostate	0 (0/2)	nd
MSI endometrial	100 (1/1)	nd
Normal tissue samples	0 (0/141)	0 (0/143)
Cancer-free controls	0 (0/50)	nd

nd=not determined

DISCUSSION

1. Novel colorectal cancer susceptibility loci (I)

The known polyposis and non-polyposis CRC syndromes together account for approximately 5% of all CRC cases. Since an estimated 35% of all CRCs can be attributed to heritable factors, the existence of novel susceptibility genes is evident (Lichtenstein *et al.* 2000). Linkage studies performed in the past decade further support the existence of novel genes predisposing to CRC. For example, linkage against known CRC susceptibility loci was observed in 10 families with excess clustering of the disease (Lewis *et al.* 1996). Furthermore, linkage to novel loci on 9q and 15q has been obtained by analyzing affected siblings or Ashkenazi families, respectively (Tomlinson *et al.* 1999, Jaeger *et al.* 2003, Wiesner *et al.* 2003). Most of the familial CRC clustering might be explained by a number of more prevalent low-penetrance susceptibility alleles. Implications for the general public health can be vast thus making the identification of these novel susceptibility genes important.

There are typically only two affected core family members in the non-syndromic CRC families of our sample collection. The underlying predisposing genes therefore confer a modestly increased cancer risk. Identification of these genes is best performed by means other than linkage analysis, especially since samples are mostly available only from the proband. Therefore, a genome-wide microsatellite marker analysis was performed to obtain information on commonly deleted or amplified regions, to form a basis for candidate gene analysis. This approach is based on the notion that AI can be detected more frequently in the familial cases at a given susceptibility locus. This allelotyping approach has the advantage that it does not require multiple samples from affected relatives. Although this method may be hampered by genetic heterogeneity, it is feasible in homogeneous populations. The Finnish population is ideal for this kind of studies due to founder effect, population isolates, and genetic drift that together have formed the gene pool homogeneous. Furthermore, this approach identified 13q deletions especially in familial breast cancer cases prior localizing the 13q gene *BRCA2* associated with breast cancer susceptibility (Thorlacius *et al.* 1991, Wooster *et al.* 1995).

We used a genome-wide microsatellite marker panel to allelotype Finnish familial CRCs with mild clustering of cancer in the family, as well as Finnish sporadic CRCs. All samples used in this study were from patients with MSS CRC. Furthermore, known polyposis cases were excluded from the analyses. We identified a significant difference in the AI frequencies between familial and sporadic cases on chromosome 20q13. From the microsatellite marker analysis graphs, it is difficult to interpret weather deletions or amplifications are causing the difference between the allele intensities. Based on our CGH experiments, the imbalances detected on 20q13 were caused by amplifications indicating a potential oncogenic mode of action for the putative CRC susceptibility gene within the affected region.

When assessing the frequency and amplitude of amplifications in the familial and sporadic CRCs by qPCR, both familial and sporadic cases seemed to harbour amplifications equally frequently. The qPCR method utilized here cannot separate the

alleles. Therefore, both alleles of sporadic CRCs might be frequently amplified whereas amplifications seem to preferentially target the other allele in the familial cases. Similar phenomenon has been seen with preferential amplification of a mutant *MET* allele in papillary renal cell cancer indicating the importance of mutant *MET* allele copy number in cancer formation (Fischer *et al.* 1998).

In previous studies, 20q amplifications have been detected in sporadic CRC cases from different populations but this is the first time amplifications were seen in excess in familial cases. Subsequently, sample sets obtained from CRC patients from Denmark and UK were analyzed for imbalances on 20q13. The small Danish set gave similar results but the UK set showed no differences between the familial and sporadic groups. These results indicate the possibility of inter-population variation due to the involvement of different susceptibility genes or modifying factors (genetic or environmental) in different populations.

The p-values obtained from the AI frequency comparisons between familial and sporadic groups have not been corrected for multiple testing. It is acknowledged that the probability of finding significant correlations increases with increasing numbers of comparisons made. However, the utility of, for example Bonferroni adjustments for multiple testing can be questioned. Although it protects from the type I errors (false positives), typically type II errors increase and true associations can be missed especially if the number of tests is high.

The identification of 20q13 amplifications preferentially in the familial CRCs suggests the presence of putative oncogene(s) predisposing to CRC within the amplicon. Some candidate CRC genes reside within the amplicon. For example, *AURKA* has recently been suggested to function as a low-penetrance susceptibility gene (Ewart-Toland *et al.* 2003) and *ZNF217* copy number changes have been linked to poorer survival and metastatic potential (Hidaka *et al.* 2000, Rooney *et al.* 2004). Furthermore, the potential tumorigenic role of other less well characterized genes within the 20q13 amplicon cannot be overlooked. Identification of the causative gene alterations would benefit from a more detailed delineation of the amplified region by using for example a more dense microsatellite marker map or array-CGH. Furthermore, expression analysis of genes within the amplicon and discovery of sequence alterations within candidate genes followed by association analyses might facilitate the identification of causative alterations.

2. *AURKA*, a novel low-penetrance susceptibility gene (II)

In study I, allele specific amplifications of 20q13 were seen more frequently in familial than in sporadic CRC cases indicating the possible existence of novel susceptibility gene(s) within the amplicon. A recently published association analysis suggested a role for 20q13 gene *AURKA* in CRC predisposition (Ewart-Toland *et al.* 2003). *AURKA* located within a genomic region orthologous to a mouse quantitative trait locus (QTL) associated with cancer susceptibility. Analyses on unselected CRC and healthy control individuals of Northern European ancestry indicated a trend between 91T>A (Phe31Ile) change and higher CRC risk. The potential tumorigenic role of the sequence alteration was highlighted by a stronger transforming potential of the 91A allele in cell culture experiments. Furthermore, a preferential amplification of

91A allele was detected together with a more aneuploid phenotype of the tumours from heterozygous individuals. The 91T>A change could therefore affect the normal function of AURKA in chromosome segregation (Zhou *et al.* 1998, Ewart-Toland *et al.* 2003). AURKA is also involved in the p53 pathway by controlling the degradation of p53 together with MDM2 (Katayama *et al.* 2004). AURKA overexpression leads to an increased degradation of p53 which in turn leads to downregulation of cell cycle checkpoint pathways and oncogenic transformation of affected cells (Katayama *et al.* 2004).

Since the study of Ewart-Toland *et al.* (2003), association analyses on other cancer types have produced similar results emphasizing the role of *AURKA* as a low-penetrance susceptibility gene (Dai *et al.* 2004, DiCioccio *et al.* 2004, Miao *et al.* 2004, Sun *et al.* 2004). The results have been somewhat conflicting as to weather the 91T>A change contributes solely or in combination with other *AURKA* changes. However, a recently published meta-analysis on several cancer types revealed a significantly increased cancer risk in 91T>A heterozygous individuals (Ewart-Toland *et al.* 2005)

In study II, the presence of possible *AURKA* sequence changes and their preferential amplification patterns were evaluated in samples from CRC patients with and without family history of the disease. The samples used in this study were from patients with non-syndromic MSS CRC. Prior to this study, confirmation to the preferential amplification of 91A had not been pursued and the role of *AURKA* had not been evaluated in familial cancer.

Sequencing of the *AURKA* coding region and exon-intron borders revealed 3 missense changes (Phe31Ile, Val57Ile, and Met373Val) and a 61bp duplication in intron 4 (IVS4+2-62dup). The frequency of these alterations was assessed in a larger set of normal tissue samples from familial and sporadic patients and the amplification patterns were evaluated in the heterozygous individuals. The frequency of each sequence variant was similar in both familial and sporadic groups as well as in cancer-free controls. The amplification patterns of 169G>A (Val57Ile), 1683A>G (Met373Val), and IVS4+2-62dup did not differ from the expected. Consistent with previous findings, the 91A allele was more frequently amplified than the 91T allele in the heterozygous individuals ($p=0.03$). As a novel finding, a difference between the familial and sporadic groups was detected: in the familial group, the amplification of 91A allele was significantly more frequent ($p=0.03$) whereas in the sporadic group a non-significant p-value of 0.36 was obtained.

Patients with different *AURKA* genotypes were compared within the familial and sporadic groups to find possible genotype-phenotype correlations. Comparisons were carried out between patients homozygous for the common allele and patient group containing both heterozygous carriers and homozygous patients for the rarer allele. No indications were obtained for the tumorigenic role of Val57Ile, Met373Val, or IVS4+2-62dup changes from comparisons of age at diagnosis, sex, tumour location, or staging of the tumour. However, familial 91T>A heterozygotes and 91A homozygotes were diagnosed at a younger age than familial patients with the more common 91T genotype ($p=0.06$). This trend was not seen in the sporadic group.

Ewart-Toland *et al.* (2003) further observed that tumours from 91T>A heterozygous individuals were more aneuploid, a finding that has not been confirmed to date. We combined the results from studies I and II to evaluate the ploidy of each tumour with respect to the *AURKA* genotype. We obtained evidence in favour of more frequent chromosomal imbalances on 9q, 11q, 13q, and 20q in heterozygous individuals. However, chromosome 2q was more aneuploid in tumours from 91T homozygous individuals.

In study II we sought for a potential target gene for the 20q13 amplification observed in study I. We evaluated the most promising candidate *AURKA* but, understandably, other possible amplification targets may exist. Other putative candidates, such as *ZNF217* could be analyzed to further elucidate the implications of the observed amplification. However, the results produced in study II indicate that the 20q13 amplifications target *AURKA* in an allele-specific manner and support the role of the 91T>A change in CRC predisposition. The difference in the amplification patterns of *AURKA* 91T and 91A alleles between the familial and sporadic groups indicates that the 91A allele could have a more pronounced effect in cases with family history of the disease. A modifying genetic factor could further emphasize the effect of 91A change in the familial patients. Further analyses are thus warranted.

3. Role of *MLH3* in microsatellite-unstable colorectal cancer (III)

Mutations in the MMR genes *MLH1* and *MSH2* have been shown to underlie a majority of the HNPCC cases whereas a smaller proportion is due to mutations in MMR genes *PMS2* and *MSH6* (Peltomäki and Vasen 2004). Causative mutations have not been detected in a subset of the families meeting the Amsterdam criteria, indicating the possible involvement of additional gene(s) in MSI CRC (Liu *et al.* 1996). In 2000, the identification of a human homolog of the *S. cerevisiae* Mlh3p was published (Lipkin *et al.* 2000). Mlh3p had previously been shown to function in MMR in yeast (Flores-Rozas and Kolodner 1998). Lipkin *et al.* (2000) suggested interactions between human MLH3 and MLH1 and, furthermore, a link between MLH3 defects and MSI. These findings prompted several genetic analyses on the role of *MLH3* in CRC predisposition. The first studies revealed no germline mutations or reported missense type of changes that did not present themselves as attractive disease-causing candidates (Loukola *et al.* 2000, Lipkin *et al.* 2001).

In 2001, Wu and co-workers suggested that *MLH3* could have a more pronounced role in HNPCC than had previously been considered (Wu *et al.* 2001). They identified 10 germline *MLH3* variants, including a one bp deletion and 9 missense variants, that were absent in the cancer-free controls. All variants were detected in patients suspected of HNPCC. Segregation analyses of the observed variants could not be performed. MSI analysis of tumours from patients with *MLH3* variants indicated more pronounced MSI in dinucleotide and tetranucleotide repeats than in mononucleotide repeats, atypical of MSI tumours caused by mutations in other MMR genes (Wu *et al.* 2001).

To further assess the role of *MLH3* in CRC predisposition, we selected CRC cases for germline mutation analysis based on the *MLH3* defect features suggested by Wu *et al.* (2001). *MLH3* germline mutations were screened by sequencing the coding region

and exon-intron borders of all available mutation negative HNPCC cases (n=7), as well as familial (n=17) and sporadic (n=6) CRC patients with more pronounced MSI in dinucleotide and tetranucleotide repeats than in mononucleotide repeats.

We identified five missense mutations (Phe390Ile, Val420Ile, Glu624Gln, Met809Val, and Asp1105Glu) with allele frequencies of 1.7%. Of note, the Glu624Gln change had been previously published by Wu *et al.* (2001) as a potential CRC-associated mutation. We detected all variants except Met809Val also in the cancer-free controls with similar allele frequencies as in CRC patients. The Met809Val change was not found in 713 cancer-free controls analyzed. However, it seems to affect a non-conserved residue located outside functionally relevant domains (Kondo *et al.* 2001).

In study III, 3 missense variants (Asp385His, Thr942Ile, and Ser966Pro) were detected in the Finnish cancer-free controls with allele frequencies varying between 0.6 and 0.9%. These variants were not present in the CRC cases analyzed. The cancer-free controls were analyzed for the occurrence of previously identified *MLH3* variants with a main focus on the most promising disease-associated mutation, 2578delA (Wu *et al.* 2001). One of the previously identified missense variants (Glu624Gln) was found in the controls. The other sequence changes, including the 2578delA, were not detected in the controls or in CRC cases analyzed.

Recently de Jong *et al.* (2004) examined the allele frequencies of a subset of the *MLH3* exon 1 variants published by Wu *et al.* (2001). The frequencies of the variants were determined in sporadic CRC patients and controls. A Ser817Gly missense variant reported by Wu *et al.* (2001) with an allele frequency of 0.2% was identified in the sporadic CRC patients and controls with similar allele frequencies (0.3% and 0.2%, respectively). Another missense variant affecting the same amino acid residue (Ser817Arg) was identified by de Jong *et al.* (2004) in the sporadic CRC patients. This variant was absent in the cancer-free controls. The truncating 2578delA mutation reported by Wu *et al.* (2001) was not found in 467 sporadic CRC patients or 497 controls. Furthermore, the Met809Val missense variant identified in study III was found in the Dutch sporadic CRC patients and control individuals with allele frequencies of 0.2% and 0.3%, respectively (de Jong *et al.* 2004).

Liu *et al.* (2003) have also analyzed the occurrence of *MLH3* variants in 70 unrelated individuals from families suggestive of a genetic predisposition to CRC. They identified one frameshift mutation (885delG) and 11 missense mutations in 16 index patients with MSI-negative tumours. The results therefore suggest mechanisms other than deficient MMR for *MLH3*-associated CRC. Liu and co-workers (2003) are the first to study the segregation of detected *MLH3* variants in the cancer families. Segregation analyses supported the role of 3 *MLH3* variants (885delG, G2221T, and T3826C) as causative for the disease, but most probably representing low risk alleles. They hypothesized that *MLH3* could contribute to CRC risk together with other low-penetrance genes (Liu *et al.* 2003).

Taken together, *MLH3* seems to harbour a wide range of sequence alterations. Most of the variants seem to lack apparent functional significance because of their localization in non-conserved residues, lack of clear segregation, and similar allele frequencies in familial and sporadic cancer cases as well as in controls (Loukola *et al.*

2000, Lipkin *et al.* 2001, Wu *et al.* 2001, Liu *et al.* 2003, de Jong *et al.* 2004, study III). The role of *MLH3* as a low-penetrance susceptibility gene needs to be studied further by association analyses. Since there seems to be inter-population differences in the *MLH3* variants detected, population specific association analyses would have to be performed.

4. High frequency of microsatellite mutations in *SEMG1* and intergenic T9 repeats (IV)

Mutations in the MMR genes lead to general instability phenomenon characterized by the accumulation of microsatellite mutations. These mutations affect tumour growth when occurring within genes having growth related or caretaker functions. A number of genes have been studied for MSI and mutations have been detected in the coding region of genes coding for growth factors, growth factor receptors, transcription factors, and apoptotic factors (Woerner *et al.* 2003). In 1998, Boland *et al.* suggested criteria that could be used to ascertain real MSI target genes. Two of the criteria, high mutation frequency and functional evidence, have been the most widely accepted.

Most of the MSI target gene studies published to date have relied solely on mutation frequency data in defining true target genes, partly because obtaining and interpreting functional evidence can be a laborious task. In the absence of functional evidence, the importance of control samples accentuates. To avoid miscalling of mutations, normal tissue DNA samples and MSS samples need to be analyzed. Furthermore, to obtain more reliable estimation on the true mutation frequency, primary tumours need to be assessed since estimations based on MSI cell lines can lead to overestimations (Duval *et al.* 2001, Suzuki *et al.* 2002). To facilitate true target gene identification, mutation frequency cut-off levels have been suggested (Duval *et al.* 2001, Woerner *et al.* 2003). However, it has also been presented that the tumorigenic role of a given gene can be difficult to ascertain solely from the mutation frequency data (Duval and Hamelin 2002, Perucho 2003). For example, mutations in survivor genes with fundamental roles in cellular homeostasis can be negatively selected for. In addition, mutations in co-operator genes can be sparse and have a mild but nonetheless significant effect that is further emphasized when occurring together with mutations in other genes operating in the same pathway (Yamamoto *et al.* 2000).

Generally non-coding microsatellite mutations are thought to be inconsequential background events and the mutation frequencies are therefore considered to be low. However, the functional importance of repeat elements outside the coding regions is not fully understood. They have been suggested to play a role in genome evolution, chromatin organization, and in the regulation of gene activity and DNA metabolic processes (Kashi *et al.* 1997, Li *et al.* 2002). In fact, some non-coding microsatellite mutations have been shown to have functional consequences. For example, 93% of primary MSI CRCs were seen to harbour mutations in *MRE11* intron 4 acceptor site that lead to a partial splicing defect and reduced mRNA and protein levels (Giannini *et al.* 2002). Furthermore, the ability of *MRE11* to function in S-phase checkpoint was impaired (Giannini *et al.* 2002). In addition, evidence for the role of CA-repeats and CA-rich elements as splicing regulators has recently been obtained (Hui *et al.* 2003, Hui *et al.* 2005). CA-elements were seen to control splicing by acting either as enhancers or silencers depending on the distance between the elements and the

alternative 5' splice sites (Hui *et al.* 2005). In another recently published study, CA-repeats were seen to control the promoter activity of *Cyr61* (Wang *et al.* 2005). These studies highlight the potential importance of non-coding microsatellites as, for example, gene expression regulators and indicate the need for further, large-scale studies of non-coding microsatellite mutations and their implications.

When assessing the background mutation frequency of MSI CRCs, we observed a relatively high mutation frequency in a T9 repeat in *SEMG1* intron 2. Subsequently we extended the analysis to include a larger number of MSI CRCs and a mutation frequency of 51% was observed. No *SEMG1* T9 repeat alterations were detected in the MSS CRCs or respective normal tissue controls analyzed, thus the *SEMG1* mutations seem to associate with MMR deficiency. The mutation frequency observed here is much higher than the cut-off levels used to determine true MSI target genes (Duval *et al.* 2001, Woerner *et al.* 2003) and much higher than in some of the proposed true target genes. To search for a possible effect of the *SEMG1* intron 2 mutations, we conducted RNA and protein level analyses but observed no differences between CRC cell lines with different *SEMG1* status. This indicates either that the mutation has an effect that cannot be detected by the methods used here, or that the mutations are bystander events without functional consequence. If the latter hypothesis holds, it shows that extremely high mutation frequencies can be detected in non-coding microsatellites even without selection pressure.

To assess the background mutation frequency of MSI CRCs further, we analyzed a panel of 29 primary MSI CRCs for mutations in 10 intergenic T9 repeats. The repeats were chosen from genomic regions distant from known genes or ESTs. The T9 repeats were altered in 0 – 86% of the studied CRCs. To exclude germline polymorphisms, we analyzed the normal tissue DNA of patients showing alterations in the tumour sample. Only one of the repeats showed alterations in the normal tissue samples. However, the 86% mutation frequency in a T9 repeat in genomic contig AC027013 persisted.

Subsequently, we extended our analysis to the whole set of 145 MSI CRCs and observed a mutation frequency of 71% without mutations in the respective normal tissue DNA samples. This result further indicates that high mutation frequencies can be found without any apparent functional consequence of the mutations. Of note, after scrutinizing the latest Celera database update, two novel transcripts were detected in the AC027013 T9 repeat region. With respect to the first transcript, the T9 repeat locates in the first intron, 2.7 and 16.9kb away from the exon-intron borders. The other transcript was located downstream of the T9 repeat with 1.1kb between the repeat and transcription start site: putative promoter regions were sought by several programs but none were found. Thus these observations speak against a functional consequence of the T9 repeat mutations. Due to the constantly growing knowledge on human transcriptomics, the intergenic nature of the T9 repeats studied here will need to be re-evaluated. To date, new transcripts have not been presented in the vicinity of AC027013 T9 repeat.

The variability observed in the mutation frequencies between non-coding repeats of the same composition and size might be at least partly explained by mechanisms controlling DNA replication. Replication accuracy is controlled by DNA polymerase fidelity as well as DNA sequence context. Polymerase fidelity is dependent on the

geometric selection of the correct nucleotide for insertion into DNA as well as on the 3'-5' exonuclease activity of the polymerase (Goodman 1997). Mechanisms underlying DNA context-dependent replication accuracy are less well understood but they may depend on DNA duplex stability and base-stacking properties. Furthermore, GC content or point substitution rates of flanking sequences have been suggested to affect microsatellite stability (Brock *et al.* 1999, Santibanez-Koref *et al.* 2001).

Taken together, the results provided by study IV emphasize the need of thorough large-scale analyses of microsatellite mutations in non-coding regions of the genome. Information on intronic and intergenic repeats from several genomic regions are needed since the mutation frequencies seem to vary quite extensively between repeats of the same base composition and size. Results produced by these studies would benefit the ongoing MSI target gene analyses and provide material for sequence comparisons to identify sequence elements controlling replication fidelity. As indicated by the results in study IV, mutation frequency alone cannot be utilized to identify putative MSI target genes. Thus, the tumorigenic role of the previously published candidate MSI target genes needs to be interpreted with caution in the absence of functional evidence.

CONCLUSIONS AND FUTURE PROSPECTS

Many colorectal cancer (CRC) predisposition syndromes have been characterized to date and the molecular changes driving CRC progression are perhaps the best characterized of all human cancer types. However, many questions remain and accumulating evidence suggests the existence of novel CRC susceptibility genes, especially of low-penetrance. This study aimed at identifying novel genes and loci associated with CRC predisposition and progression in both chromosomal instability (CIN) and microsatellite instability (MSI) pathways.

1. Identification of novel CRC susceptibility loci was pursued by assessing common regions of allelic imbalance in CIN CRCs. Chromosome 20q13 was seen to harbour amplifications more frequently in cases with family history of the disease, indicating the potential existence of novel predisposing oncogene(s) within the amplicon. Further investigations on the importance of the amplification are warranted. The putative tumorigenic role of one 20q13 candidate gene, *AURKA*, was analyzed in study II.

2. A candidate CRC susceptibility gene on chromosome 20q13, *AURKA*, was assessed in CIN CRC cases with and without family history of the disease. The observed preferential amplification of 91A allele in the familial cases and association with younger age at diagnosis support the role of *AURKA* as a novel CRC susceptibility gene. Further analyses could elucidate the potential modifying factors of *AURKA*-associated CRC. Studies I and II indicate the feasibility of using chromosomal clues in the identification of candidate loci and causative cancer genes.

3. A subset of MSI CRCs cannot be ascribed to the MMR genes currently associated with CRC predisposition. The role of MMR gene *MLH3* was assessed in mutation negative HNPCC cases, as well as in familial and sporadic CRC patients displaying distinct MSI features. Five missense mutations were identified, four of which were found in cancer-free controls as well. The remaining missense mutation is not an attractive disease-causing candidate. The results provided by study III and other recently published *MLH3* studies suggest a minor role for *MLH3* in CRC predisposition, to be evaluated further by association analyses.

4. MSI target genes have previously been proposed mainly based on high mutation frequency data. In study IV, two non-coding microsatellites were seen to display mutation frequencies higher than those observed in the coding regions of many suggested MSI target genes. Thus, mutation data alone is not sufficient to indicate the role of a given gene in MSI tumorigenesis and functional evidence should therefore be provided. MSI target gene studies would benefit from evaluations on the general background mutation frequency of MSI CRCs. Furthermore, information provided therein could elucidate the mechanisms governing the sequence context dependence of DNA replication fidelity.

ACKNOWLEDGEMENTS

This study was carried out at the Department of Medical Genetics, Haartman Institute and Biomedicum Helsinki, University of Helsinki, during 2001-2005. The former and present heads of the Department of Medical Genetics, Pertti Aula, Anna-Elina Lehesjoki, Leena Palotie, Kristiina Aittomäki, and Päivi Peltomäki are thanked for providing excellent research facilities.

I am sincerely grateful to my supervisor Lauri Aaltonen for his support and guidance during these years and for giving me the opportunity to work in his excellent research group. The atmosphere in the lab has been truly remarkable and inspiring thanks to Lauri. I believe that Lauri's enthusiasm and innovative approach has taught a great deal to all of us that have had the pleasure to work in his team.

I am also very grateful to my other supervisor Auli Karhu for her excellent and never-ending support as well as for her valuable ideas during the course of this work. Furthermore, I wish to thank Auli for the discussions we have had - both scientific and otherwise - and for her great sense of humour that always kept me going.

I thank the reviewers of my thesis, professor Raquel Seruca and docent Outi Monni for encouraging comments and excellent ideas on how to improve the thesis.

I thank all the co-authors and collaborators that have greatly facilitated this work: Lara Lipton, Yan Aalto, Huw Thomas, Karin Birkenkamp-Demtröder, Shirley Hodgson, Jukka-Pekka Mecklin, Heikki Järvinen, Sakari Knuutila, Sarah Halford, Torben Ørntoft, Ian Tomlinson, Richard Houlston, Pertti Sistonen, Päivi Peltomäki, Thomas Barber, Riitta Koistinen, and Bert Vogelstein.

A huge thanks to the rest of the Aaltonen Team, you have made these years very enjoyable: the PhD students Antti, Heli L, Heli S, Marianna, Pia, Päivi, Rainer, Sakari, Sanna, and Taru - a special thanks to my roommates Sanna, both Helis, and Pia for the time spent discussing various extracurricular topics and science, to Päivi for guiding me during the early days when things seemed bewildering, to Rainer for showing how to party like nothing else mattered; the postdocs Virpi and Pia - a special thanks to Virpi for her excellent support and for sharing her valuable ideas on numerous projects; the skillful lab team Anniina, Iina, Inga-Lill, and Mikko - a sincere thanks for your patient assistance with the projects that sometimes seemed endless; our pathologist Reijo for his valuable assistance with the slides; our research nurses Sini and Katja - a special thanks to Sini for her excellent help with the samples and taking care of routine business when I was busy with other things; the newcomers in the team Eeva-Mari, Miranda, Päivi, and Silva; and finally the past members Annika, Diego, Kirsi, Maija, Nina, Susa, Taija, and Ullis - a special thanks to Susa for the enjoyable and educative times during the shared projects, to Taija for inspiring discussions, and to Diego for his valuable views during the confusing times of project IV. A hearty thanks to all you guys. I will never forget these years or the numerous parties - and the pictures won't let me!

I also wish to acknowledge the personnel of the Finnish Genome Center and the sequencing and fragment analysis facilities in Biomedicum and Haartman Institute whose contribution has been of great importance during the course of this work.

A huge thanks to all my friends for the excellent times spent in various locations - for the unforgettable get-togethers, parties, sporting events, and board games that have kept my head together and balanced my life in a profound way.

And of course - a warm and sincere thanks to my amazing family for the endless support and for understanding my hectic schedule. I am truly grateful to Tomi for his love and support during these years - he has kept me on track and reminded me of the things that really matter in life. And of course my sweet Mom Sirja and Dad Pentti, my superb sisters Kirsi, Saara, and Virpi, and brothers-in-law Make, Peter, and Sami, and their lovely and energetic children Tino, Kaapo&Sofia&the new baby boy, and Nea&Niko, and last but not least my second family Ritva, Lauri, and Jani - you are all simply the best!

This study was supported by grants from the Maud Kuistila Memorial Foundation, Ida Montin Foundation, and Paulo Foundation.

REFERENCES

Aaltonen LA, Peltomäki P, Leach FS, Sistonen P, Pylkkänen L, Mecklin JP, Järvinen H, Powell SM, Jen J, Hamilton SR, Petersen GM, Kinzler KW, Vogelstein B, and de la Chapelle A. Clues to the pathogenesis of familial colorectal cancer. *Science* (1993). 260, 812-6.

Aaltonen LA, Salovaara R, Kristo P, Canzian F, Hemminki A, Peltomäki P, Chadwick RB, Kääriäinen H, Eskelinen M, Järvinen H, Mecklin JP, and de la Chapelle A. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N.Engl.J.Med.* (1998). 338, 1481-7.

Aarnio M, Mecklin JP, Aaltonen LA, Nyström-Lahti M, and Järvinen HJ. Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (HNPCC) syndrome. *Int.J.Cancer* (1995). 64, 430-3.

Akiyama Y, Sato H, Yamada T, Nagasaki H, Tsuchiya A, Abe R, and Yuasa Y. Germ-line mutation of the hMSH6/GTBP gene in an atypical hereditary nonpolyposis colorectal cancer kindred. *Cancer Res.* (1997a). 57, 3920-3.

Akiyama Y, Tsubouchi N, and Yuasa Y. Frequent somatic mutations of hMSH3 with reference to microsatellite instability in hereditary nonpolyposis colorectal cancers. *Biochem.Biophys.Res.Commun.* (1997b). 236, 248-52.

Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, Hodges AK, Davies DR, David SS, Sampson JR, and Cheadle JP. Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. *Nat.Genet.* (2002). 30, 227-32.

Bao S, Chang MS, Auclair D, Sun Y, Wang Y, Wong WK, Zhang J, Liu Y, Qian X, Sutherland R, Magi-Galluzi C, Weisberg E, Cheng EY, Hao L, Sasaki H, Campbell MS, Kraeft SK, Loda M, Lo KM, and Chen LB. HRad17, a human homologue of the Schizosaccharomyces pombe checkpoint gene rad17, is overexpressed in colon carcinoma. *Cancer Res.* (1999). 59, 2023-8.

Baranovskaya S, Soto JL, Perucho M, and Malkhosyan SR. Functional significance of concomitant inactivation of hMLH1 and hMSH6 in tumor cells of the microsatellite mutator phenotype. *Proc.Natl.Acad.Sci.U.S.A.* (2001). 98, 15107-12.

Bardelli A, Cahill DP, Lederer G, Speicher MR, Kinzler KW, Vogelstein B, and Lengauer C. Carcinogen-specific induction of genetic instability. *Proc.Natl.Acad.Sci.U.S.A.* (2001). 98, 5770-5.

Bhattacharyya NP, Skandalis A, Ganesh A, Groden J, and Meuth M. Mutator phenotypes in human colorectal carcinoma cell lines. *Proc.Natl.Acad.Sci.U.S.A.* (1994). 91, 6319-23.

Bischoff JR, Anderson L, Zhu Y, Mossie K, Ng L, Souza B, Schryver B, Flanagan P, Clairvoyant F, Ginther C, Chan CS, Novotny M, Slamon DJ, and Plowman GD. A homologue of Drosophila aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J.* (1998). 17, 3052-65.

Bishop JM. Molecular themes in oncogenesis. *Cell* (1991). 64, 235-48.

Boland CR, Sato J, Appelman HD, Bresalier RS, and Feinberg AP. Microallelotyping defines the sequence and tempo of allelic losses at tumour suppressor gene loci during colorectal cancer progression. *Nat.Med.* (1995). 1, 902-9.

Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, and Srivastava S. 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.* (1998). 58, 5248-57.

Brabletz T, Jung A, Reu S, Porzner M, Hlubek F, Kunz-Schughart LA, Knuechel R, and Kirchner T. Variable beta-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. *Proc.Natl.Acad.Sci.U.S.A.* (2001). 98, 10356-61.

Brock GJ, Anderson NH, and Monckton DG. Cis-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. *Hum.Mol.Genet.* (1999). 8, 1061-7.

Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, Earabino C, Lipford J, Lindblom A, Tannergård P, Bollag RJ, Godwin AR, Ward DC, Nordenskjøld M, Fishel R, Kolodner R, and Liskay RM. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* (1994). 368, 258-61.

Bufill JA. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Ann.Intern.Med.* (1990). 113, 779-88.

Burt R and Neklason DW. Genetic testing for inherited colon cancer. *Gastroenterology* (2005). 128, 1696-716.

Cahill DP, Lengauer C, Yu J, Riggins GJ, Willson JK, Markowitz SD, Kinzler KW, and Vogelstein B. Mutations of mitotic checkpoint genes in human cancers. *Nature* (1998). 392, 300-3.

Camp NJ. Genomewide transmission/disequilibrium testing--consideration of the genotypic relative risks at disease loci. *Am.J.Hum.Genet.* (1997). 61, 1424-30.

Canzian F, Salovaara R, Hemminki A, Kristo P, Chadwick RB, Aaltonen LA, and de la Chapelle A. Semiautomated assessment of loss of heterozygosity and replication error in tumors. *Cancer Res.* (1996). 56, 3331-7.

Caspari R, Friedl W, Mandl M, Møslein G, Kadmon M, Knapp M, Jacobasch KH, Ecker KW, Kreissler-Haag D, and Timmermanns G. Familial adenomatous polyposis: mutation at codon 1309 and early onset of colon cancer. *Lancet* (1994). 343, 629-32.

Chan TL, Zhao W, Leung SY, Yuen ST, and Cancer Genome Project. BRAF and KRAS mutations in colorectal hyperplastic polyps and serrated adenomas. *Cancer Res.* (2003). 63, 4878-81.

Chen K, Jiang QT, and He HQ. Relationship between metabolic enzyme polymorphism and colorectal cancer. *World J.Gastroenterol.* (2005). 11, 331-5.

Crabtree MD, Tomlinson IP, Hodgson SV, Neale K, Phillips RK, and Houlston RS. Explaining variation in familial adenomatous polyposis: relationship between genotype and phenotype and evidence for modifier genes. *Gut* (2002). 51, 420-3.

Crabtree MD, Fletcher C, Churchman M, Hodgson SV, Neale K, Phillips RK, and Tomlinson IP. Analysis of candidate modifier loci for the severity of colonic familial adenomatous polyposis, with evidence for the importance of the N-acetyl transferases. *Gut* (2004). 53, 271-6.

Cui H, Cruz-Correa M, Giardiello FM, Hutcheon DF, Kafonek DR, Brandenburg S, Wu Y, He X, Powe NR, and Feinberg AP. Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science* (2003). 299, 1753-5.

Cunningham JM, Christensen ER, Tester DJ, Kim CY, Roche PC, Burgart LJ, and Thibodeau SN. Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. *Cancer Res.* (1998). 58, 3455-60.

Dai Q, Cai QY, Shu XO, Ewart-Toland A, Wen WQ, Balmain A, Gao YT, and Zheng W. Synergistic effects of STK15 gene polymorphisms and endogenous estrogen exposure in the risk of breast cancer. *Cancer Epidemiol.Biomarkers Prev.* (2004). 13, 2065-70.

de Jong MM, Nolte IM, te Meerman GJ, van der Graaf WT, de Vries EG, Sijmons RH, Hofstra RM, and Kleibeuker JH. Low-penetrance genes and their involvement in colorectal cancer susceptibility. *Cancer Epidemiol.Biomarkers Prev.* (2002). 11, 1332-52.

de Jong MM, Hofstra RM, Kooi KA, Westra JL, Berends MJ, Wu Y, Hollema H, van der Sluis T, van der Graaf WT, de Vries EG, Schaapveld M, Sijmons RH, te Meerman GJ, and Kleibeuker JH. No association between two MLH3 variants (S845G and P844L)and colorectal cancer risk. *Cancer Genet.Cytogenet.* (2004). 152, 70-1.

Demant P. Cancer susceptibility in the mouse: genetics, biology and implications for human cancer. *Nat.Rev.Genet.* (2003). 4, 721-34.

Dhawan P, Singh AB, Deane NG, No Y, Shiou SR, Schmidt C, Neff J, Washington MK, and Beauchamp RD. Claudin-1 regulates cellular transformation and metastatic behavior in colon cancer. *J.Clin.Invest.* (2005). 115, 1765-76.

Di Cristofano A, Pesce B, Cordon-Cardo C, and Pandolfi PP. Pten is essential for embryonic development and tumour suppression. *Nat.Genet.* (1998). 19, 348-55.

DiCioccio RA, Song H, Waterfall C, Kimura MT, Nagase H, McGuire V, Hogdall E, Shah MN, Luben RN, Easton DF, Jacobs IJ, Ponder BA, Whittemore AS, Gayther SA, Pharoah PD, and Kruger-Kjaer S. STK15 polymorphisms and association with risk of invasive ovarian cancer. *Cancer Epidemiol.Biomarkers Prev.* (2004). 13, 1589-94.

Dietrich WF, Lander ES, Smith JS, Moser AR, Gould KA, Luongo C, Borenstein N, and Dove W. Genetic identification of Mom-1, a major modifier locus affecting Min-induced intestinal neoplasia in the mouse. *Cell* (1993). 75, 631-9.

Dobbie Z, Muller H, and Scott RJ. Secretory phospholipase A2 does not appear to be associated with phenotypic variation in familial adenomatous polyposis. *Hum.Genet.* (1996). 98, 386-90.

Donis-Keller H, Dou S, Chi D, Carlson KM, Toshima K, Lairmore TC, Howe JR, Moley JF, Goodfellow P, and Wells SA,Jr. Mutations in the RET proto-oncogene are associated with MEN 2A and FMTC. *Hum.Mol.Genet.* (1993). 2, 851-6.

Duval A, Rolland S, Compoint A, Tubacher E, Iacopetta B, Thomas G, and Hamelin R. Evolution of instability at coding and non-coding repeat sequences in human MSI-H colorectal cancers. *Hum.Mol.Genet.* (2001). 10, 513-8.

Duval A and Hamelin R. Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. *Cancer Res.* (2002). 62, 2447-54.

Entius MM, Keller JJ, Westerman AM, van Rees BP, van Velthuysen ML, de Goeij AF, Wilson JH, Giardiello FM, and Offerhaus GJ. Molecular genetic alterations in hamartomatous polyps and carcinomas of patients with Peutz-Jeghers syndrome. *J.Clin.Pathol.* (2001). 54, 126-31.

Ewart-Toland A, Briassouli P, de Koning JP, Mao JH, Yuan J, Chan F, MacCarthy-Morrogh L, Ponder BA, Nagase H, Burn J, Ball S, Almeida M, Linardopoulos S, and Balmain A. Identification of Stk6/STK15 as a candidate low-penetrance tumor-susceptibility gene in mouse and human. *Nat.Genet.* (2003). 34, 403-12.

Ewart-Toland A, Dai Q, Gao YT, Nagase H, Dunlop MG, Farrington SM, Barnetson RA, Anton-Culver H, Peel D, Ziogas A, Lin D, Miao X, Sun T, Ostrander EA, Stanford JL, Langlois M, Chan JM, Yuan J, Harris C, Bowman E, Clayman GL, Lippman SM, Lee JJ, Zheng W, and Balmain A. Aurora-A/STK15 T+91A is a general low penetrance cancer susceptibility gene: a meta-analysis of multiple cancer types. *Carcinogenesis* (2005). 26, 1368-73

Fearnhead NS, Britton MP, and Bodmer WF. The ABC of APC. *Hum.Mol.Genet.* (2001). 10, 721-33.

Fearon ER and Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* (1990). 61, 759-67.

Fischer J, Palmedo G, von Knobloch R, Bugert P, Prayer-Galetti T, Pagano F, and Kovacs G. Duplication and overexpression of the mutant allele of the MET proto-oncogene in multiple hereditary papillary renal cell tumours. *Oncogene* (1998). 17, 733-9.

Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, Kane M, and Kolodner R. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* (1993). 75, 1027-38.

Flores-Rozas H and Kolodner RD. The Saccharomyces cerevisiae MLH3 gene functions in MSH3-dependent suppression of frameshift mutations. *Proc.Natl.Acad.Sci.U.S.A.* (1998). 95, 12404-9.

Fodde R, Kuipers J, Rosenberg C, Smits R, Kielman M, Gaspar C, van Es JH, Breukel C, Wiegant J, Giles RH, and Clevers H. Mutations in the APC tumour suppressor gene cause chromosomal instability. *Nat.Cell Biol.* (2001). 3, 433-8.

Foulkes WD, Thiffault I, Gruber SB, Horwitz M, Hamel N, Lee C, Shia J, Markowitz A, Figer A, Friedman E, Farber D, Greenwood CM, Bonner JD, Nafa K, Walsh T, Marcus V, Tomsho L, Gebert J, Macrae FA, Gaff CL, Paillerets BB, Gregersen PK, Weitzel JN, Gordon PH, MacNamara E, King MC, Hampel H, De La Chapelle A, Boyd J, Offit K, Rennert G, Chong G, and Ellis NA. The founder mutation MSH2*1906G-->C is an important cause of hereditary nonpolyposis colorectal cancer in the Ashkenazi Jewish population. *Am.J.Hum.Genet.* (2002). 71, 1395-412.

Friedberg EC. DNA damage and repair. *Nature* (2003). 421, 436-40.

Friedl W, Meuschel S, Caspari R, Lamberti C, Krieger S, Sengteller M, and Propping P. Attenuated familial adenomatous polyposis due to a mutation in the 3' part of the APC gene. A clue for understanding the function of the APC protein. *Hum.Genet.* (1996). 97, 579-84.

Froggatt NJ, Green J, Brassett C, Evans DG, Bishop DT, Kolodner R, and Maher ER. A common MSH2 mutation in English and North American HNPCC families: origin, phenotypic expression, and sex specific differences in colorectal cancer. *J.Med.Genet.* (1999). 36, 97-102.

Gayther SA, Wells D, SenGupta SB, Chapman P, Neale K, Tsioupra K, and Delhanty JD. Regionally clustered APC mutations are associated with a severe phenotype and occur at a high frequency in new mutation cases of adenomatous polyposis coli. *Hum.Mol.Genet.* (1994). 3, 53-6.

Giannini G, Ristori E, Cerignoli F, Rinaldi C, Zani M, Viel A, Ottini L, Crescenzi M, Martinotti S, Bignami M, Frati L, Screpanti I, and Gulino A. Human MRE11 is inactivated in mismatch repair-deficient cancers. *EMBO Rep.* (2002). 3, 248-54.

Glebov OK, Rodriguez LM, Nakahara K, Jenkins J, Cliatt J, Humbyrd CJ, DeNobile J, Soballe P, Simon R, Wright G, Lynch P, Patterson S, Lynch H, Gallinger S, Buchbinder A, Gordon G, Hawk E, and Kirsch IR. Distinguishing right from left colon by the pattern of gene expression. *Cancer Epidemiol.Biomarkers Prev.* (2003). 12, 755-62.

Goodman MF. Hydrogen bonding revisited: geometric selection as a principal determinant of DNA replication fidelity. *Proc.Natl.Acad.Sci.U.S.A.* (1997). 94, 10493-5.

Grady WM, Rajput A, Myeroff L, Liu DF, Kwon K, Willis J, and Markowitz S. Mutation of the type II transforming growth factor-beta receptor is coincident with the transformation of human colon adenomas to malignant carcinomas. *Cancer Res.* (1998). 58, 3101-4.

Groden J, Thliveris A, Samowitz W, Carlson M, Gelbert L, Albertsen H, Joslyn G, Stevens J, Spirio L, Robertson M, Sargeant L, Krapcho K, Wolff E, Burt R, Hughes JP, Warrington J, McPherson J, Wasmuth J, Le Paslier D, Abderrahim H, Cohen D, Leppert M, and White R. Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* (1991). 66, 589-600.

Hackman P, Tannergard P, Osei-Mensa S, Chen J, Kane MF, Kolodner R, Lambert B, Hellgren D, and Lindblom A. A human compound heterozygote for two MLH1 missense mutations. *Nat.Genet.* (1997). 17, 135-6.

Halford S, Sasieni P, Rowan A, Wasan H, Bodmer W, Talbot I, Hawkins N, Ward R, and Tomlinson I. Low-level microsatellite instability occurs in most colorectal cancers and is a nonrandomly distributed quantitative trait. *Cancer Res.* (2002). 62, 53-7.

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

Harfe BD and Jinks-Robertson S. DNA mismatch repair and genetic instability. *Annu.Rev.Genet.* (2000). 34, 359-99.

Hawkins NJ and Ward RL. Sporadic colorectal cancers with microsatellite instability and their possible origin in hyperplastic polyps and serrated adenomas. *J.Natl.Cancer Inst.* (2001). 93, 1307-13.

Hayes JD, Flanagan JU, and Jowsey IR. Glutathione transferases. *Annu.Rev.Pharmacol.Toxicol.* (2005). 45, 51-88.

He TC, Sparks AB, Rago C, Hermeking H, Zawel L, da Costa LT, Morin PJ, Vogelstein B, and Kinzler KW. Identification of c-MYC as a target of the APC pathway. *Science* (1998). 281, 1509-12.

Heldin CH, Miyazono K, and ten Dijke P. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature* (1997). 390, 465-71.

Hendriks YM, Wagner A, Morreau H, Menko F, Stormorken A, Quehenberger F, Sandkuijl L, Møller P, Genuardi M, Van Houwelingen H, Tops C, Van Puijenbroek M, Verkuijlen P, Kenter G, Van Mil A, Meijers-Heijboer H, Tan GB, Breuning MH, Fodde R, Wijnen JT, Bröcker-Vriends AH, and Vasen H. Cancer risk in hereditary nonpolyposis colorectal cancer due to MSH6 mutations: impact on counseling and surveillance. *Gastroenterology* (2004). 127, 17-25.

Herman JG, Umar A, Polyak K, Graff JR, Ahuja N, Issa JP, Markowitz S, Willson JK, Hamilton SR, Kinzler KW, Kane MF, Kolodner RD, Vogelstein B, Kunkel TA, and Baylin SB. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc.Natl.Acad.Sci.U.S.A.* (1998). 95, 6870-5.

Hidaka S, Yasutake T, Takeshita H, Kondo M, Tsuji T, Nanashima A, Sawai T, Yamaguchi H, Nakagoe T, Ayabe H, and Tagawa Y. Differences in 20q13.2 copy number between colorecctal cancers with and without liver metastasis. *Clin. Cancer Res.* (2000). 6, 2712-7.

Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. *Nature* (2001). 411, 366-74.

Houlston R, Crabtree M, Phillips R, Crabtree M, and Tomlinson I. Explaining differences in the severity of familial adenomatous polyposis and the search for modifier genes. *Gut* (2001). 48, 1-5.

Houlston RS and Tomlinson IP. Polymorphisms and colorectal tumor risk. *Gastroenterology* (2001). 121, 282-301.

Houlston RS and Peto J. The future of association studies of common cancers. *Hum.Genet.* (2003). 112, 434-5.

Houlston RS and Peto J. The search for low-penetrance cancer susceptibility alleles. *Oncogene* (2004). 23, 6471-6.

Hui J, Stangl K, Lane WS, and Bindereif A. HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat.Struct.Biol.* (2003). 10, 33-7.

Hui J, Hung LH, Heiner M, Schreiner S, Neumuller N, Reither G, Haas SA, and Bindereif A. Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J.* (2005). 24, 1988-98.

Iino H, Jass JR, Simms LA, Young J, Leggett B, Ajioka Y, and Watanabe H. DNA microsatellite instability in hyperplastic polyps, serrated adenomas, and mixed polyps: a mild mutator pathway for colorectal cancer? *J.Clin.Pathol.* (1999). 52, 5-9.

Ionov Y, Peinado MA, Malkhosyan S, Shibata D, and Perucho M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* (1993). 363, 558-61.

Ionov Y, Yamamoto H, Krajewski S, Reed JC, and Perucho M. Mutational inactivation of the proapoptotic gene BAX confers selective advantage during tumor clonal evolution. *Proc.Natl.Acad.Sci.U.S.A.* (2000). 97, 10872-7.

Jaeger EE, Woodford-Richens KL, Lockett M, Rowan AJ, Sawyer EJ, Heinimann K, Rozen P, Murday VA, Whitelaw SC, Ginsberg A, Atkin WS, Lynch HT, Southey MC, Debinski H, Eng C, Bodmer WF, Talbot IC, Hodgson SV, Thomas HJ, and Tomlinson IP. An ancestral Ashkenazi haplotype at the HMPS/CRAC1 locus on 15q13-q14 is associated with hereditary mixed polyposis syndrome. *Am.J.Hum.Genet.* (2003). 72, 1261-7.

Jallepalli PV and Lengauer C. Chromosome segregation and cancer: cutting through the mystery. *Nat.Rev.Cancer.* (2001). 1, 109-17.

Jass JR. Colorectal adenoma progression and genetic change: is there a link? *Ann.Med.* (1995). 27, 301-6.

Jass JR. Serrated adenoma and colorectal cancer. *J.Pathol.* (1999). 187, 499-502.

Jass JR, Biden KG, Cummings MC, Simms LA, Walsh M, Schoch E, Meltzer SJ, Wright C, Searle J, Young J, and Leggett BA. Characterisation of a subtype of colorectal cancer combining features of the suppressor and mild mutator pathways. *J.Clin.Pathol.* (1999). 52, 455-60.

Kairupan CF, Meldrum CJ, Crooks R, Milward EA, Spigelman AD, Burgess B, Groombridge C, Kirk J, Tucker K, Ward R, Williams R, and Scott RJ. Mutation analysis of the MYH gene in an Australian series of colorectal polyposis patients with or without germline APC mutations. *Int.J.Cancer* (2005). 116, 73-7.

Kane MF, Loda M, Gaida GM, Lipman J, Mishra R, Goldman H, Jessup JM, and Kolodner R. 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.* (1997). 57, 808-11.

Kaplan KB, Burds AA, Swedlow JR, Bekir SS, Sorger PK, and Näthke IS. A role for the Adenomatous Polyposis Coli protein in chromosome segregation. *Nat.Cell Biol.* (2001). 3, 429-32.

Kashi Y, King D, and Soller M. Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.* (1997). 13, 74-8.

Katayama H, Sasai K, Kawai H, Yuan ZM, Bondaruk J, Suzuki F, Fujii S, Arlinghaus RB, Czerniak BA, and Sen S. Phosphorylation by aurora kinase A induces Mdm2-mediated destabilization and inhibition of p53. *Nat.Genet.* (2004). 36, 55-62.

Kemp Z, Thirlwell C, Sieber O, Silver A, and Tomlinson I. An update on the genetics of colorectal cancer. *Hum.Mol.Genet.* (2004). 13 Spec No 2, R177-85.

Kim YI, Pogribny IP, Salomon RN, Choi SW, Smith DE, James SJ, and Mason JB. Exon-specific DNA hypomethylation of the p53 gene of rat colon induced by dimethylhydrazine. Modulation by dietary folate. *Am.J.Pathol.* (1996). 149, 1129-37.

Kinouchi Y, Hiwatashi N, Higashioka S, Nagashima F, Chida M, and Toyota T. Relaxation of imprinting of the insulin-like growth factor II gene in colorectal cancer. *Cancer Lett.* (1996). 107, 105-8.

Kinzler KW and Vogelstein B. Lessons from hereditary colorectal cancer. *Cell* (1996). 87, 159-70.

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

Kinzler KW and Vogelstein B. Landscaping the cancer terrain. *Science* (1998). 280, 1036-7.

Knudson AG,Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc.Natl.Acad.Sci.U.S.A.* (1971). 68, 820-3.

Kolodner RD and Marsischky GT. Eukaryotic DNA mismatch repair. *Curr.Opin.Genet.Dev.* (1999). 9, 89-96.

Kondo E, Horii A, and Fukushige S. The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. *Nucleic Acids Res.* (2001). 29, 1695-702.

Korinek V, Barker N, Morin PJ, van Wichen D, de Weger R, Kinzler KW, Vogelstein B, and Clevers H. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. *Science* (1997). 275, 1784-7.

Laiho P, Launonen V, Lahermo P, Esteller M, Guo M, Herman JG, Mecklin JP, Järvinen H, Sistonen P, Kim KM, Shibata D, Houlston RS, and Aaltonen LA. Low-level microsatellite instability in most colorectal carcinomas. *Cancer Res.* (2002). 62, 1166-70.

Laken SJ, Petersen GM, Gruber SB, Oddoux C, Ostrer H, Giardiello FM, Hamilton SR, Hampel H, Markowitz A, Klimstra D, Jhanwar S, Winawer S, Offit K, Luce MC, Kinzler KW, and Vogelstein B. Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat.Genet.* (1997). 17, 79-83.

Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomäki P, Sistonen P, Aaltonen LA, Nyström-Lahti M, Guan XY, Zhang J, Meltzer PS, Yu JW, Kao FT, Chen DJ, Cerosaletti KM, Fournier REK, Todd S, Lewis T, Leach RJ, Naylor SL, Weissenbach J, Mecklin JP, Järvinen H, Petersen GM, Hamilton SR, Green J, Jass J, Watson P, Lynch HT, Trent JM, de la Chapelle A, Kinzler KW, and Vogelstein B. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* (1993). 75, 1215-25.

Lengauer C, Kinzler KW, and Vogelstein B. Genetic instabilities in human cancers. *Nature* (1998). 396, 643-9.

Leslie A, Carey FA, Pratt NR, and Steele RJ. The colorectal adenoma-carcinoma sequence. *Br.J.Surg.* (2002). 89, 845-60.

Lewis CM, Neuhausen SL, Daley D, Black FJ, Swensen J, Burt RW, Cannon-Albright LA, and Skolnick MH. Genetic heterogeneity and unmapped genes for colorectal cancer. *Cancer Res.* (1996). 56, 1382-8.

Li YC, Korol AB, Fahima T, Beiles A, and Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol.Ecol.* (2002). 11, 2453-65.

Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, and Hemminki K. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N.Engl.J.Med.* (2000). 343, 78-85.

Lipkin SM, Wang V, Jacoby R, Banerjee-Basu S, Baxevanis AD, Lynch HT, Elliott RM, and Collins FS. MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. *Nat.Genet.* (2000). 24, 27-35.

Lipkin SM, Wang V, Stoler DL, Anderson GR, Kirsch I, Hadley D, Lynch HT, and Collins FS. Germline and somatic mutation analyses in the DNA mismatch repair gene MLH3: Evidence for somatic mutation in colorectal cancers. *Hum.Mutat.* (2001). 17, 389-96.

Liu B, Parsons R, Papadopoulos N, Nicolaides NC, Lynch HT, Watson P, Jass JR, Dunlop M, Wyllie A, Peltomäki P, de la Chapelle A, Hamilton SR, Vogelstein B, and Kinzler KW. Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. *Nat.Med.* (1996). 2, 169-74.

Liu HX, Zhou XL, Liu T, Werelius B, Lindmark G, Dahl N, and Lindblom A. The role of hMLH3 in familial colorectal cancer. *Cancer Res.* (2003). 63, 1894-9.

Liu W, Dong X, Mai M, Seelan RS, Taniguchi K, Krishnadath KK, Halling KC, Cunningham JM, Boardman LA, Qian C, Christensen E, Schmidt SS, Roche PC, Smith DI, and Thibodeau SN. Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating beta-catenin/TCF signalling. *Nat.Genet.* (2000). 26, 146-7.

Locker GY and Lynch HT. Genetic factors and colorectal cancer in Ashkenazi Jews. *Fam.Cancer.* (2004). 3, 215-21.

Loukola A, Vilkki S, Singh J, Launonen V, and Aaltonen LA. Germline and somatic mutation analysis of MLH3 in MSI-positive colorectal cancer. *Am.J.Pathol.* (2000). 157, 347-52.

Loukola A, Eklin K, Laiho P, Salovaara R, Kristo P, Järvinen H, Mecklin JP, Launonen V, and Aaltonen LA. Microsatellite marker analysis in screening for hereditary nonpolyposis colorectal cancer (HNPCC). *Cancer Res.* (2001). 61, 4545-9.

Lund AH and van Lohuizen M. Epigenetics and cancer. *Genes Dev.* (2004). 18, 2315-35.

MacPhee M, Chepenik KP, Liddell RA, Nelson KK, Siracusa LD, and Buchberg AM. The secretory phospholipase A2 gene is a candidate for the Mom1 locus, a major modifier of ApcMin-induced intestinal neoplasia. *Cell* (1995). 81, 957-66.

Mao JH and Balmain A. Genomic approaches to identification of tumour-susceptibility genes using mouse models. *Curr.Opin.Genet.Dev.* (2003). 13, 14-9.

Marra G and Schär P. Recognition of DNA alterations by the mismatch repair system. *Biochem.J.* (1999). 338 (Pt 1), 1-13.

Marsh DJ and Zori RT. Genetic insights into familial cancer - update and recent discoveries. *Cancer Lett.* (2002). 181, 125-64.

Miao X, Sun T, Wang Y, Zhang X, Tan W, and Lin D. Functional STK15 Phe31Ile polymorphism is associated with the occurrence and advanced disease status of esophageal squamous cell carcinoma. *Cancer Res.* (2004). 64, 2680-3.

Miller DG. On the nature of susceptibility to cancer. The presidential address. *Cancer* (1980). 46, 1307-18.

Miyaki M, Konishi M, Tanaka K, Kikuchi-Yanoshita R, Muraoka M, Yasuno M, Igari T, Koike M, Chiba M, and Mori T. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat.Genet.* (1997). 17, 271-2.

Miyoshi H, Nakau M, Ishikawa TO, Seldin MF, Oshima M, and Taketo MM. Gastrointestinal hamartomatous polyposis in Lkb1 heterozygous knockout mice. *Cancer Res.* (2002). 62, 2261-6.

Mulligan LM, Kwok JB, Healey CS, Elsdon MJ, Eng C, Gardner E, Love DR, Mole SE, Moore JK, Papi L, Ponder MA, Telenius Hakan, Tunnacliffe A, and Ponder BAJ. Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. *Nature* (1993). 363, 458-60.

Munemitsu S, Souza B, Muller O, Albert I, Rubinfeld B, and Polakis P. The APC gene product associates with microtubules in vivo and promotes their assembly in vitro. *Cancer Res.* (1994). 54, 3676-81.

Munemitsu S, Albert I, Souza B, Rubinfeld B, and Polakis P. Regulation of intracellular beta-catenin levels by the adenomatous polyposis coli (APC) tumor-suppressor protein. *Proc.Natl.Acad.Sci.U.S.A.* (1995). 92, 3046-50.

Nagase H, Miyoshi Y, Horii A, Aoki T, Ogawa M, Utsunomiya J, Baba S, Sasazuki T, and Nakamura Y. Correlation between the location of germ-line mutations in the APC gene and the number of colorectal polyps in familial adenomatous polyposis patients. *Cancer Res.* (1992). 52, 4055-7.

Nicolaides NC, Papadopoulos N, Liu B, Wei YF, Carter KC, Ruben SM, Rosen CA, Haseltine WA, Fleischmann RD, Fraser CM, Adams MD, Venter JC, Dunlop MG, Hamilton SR, Petersen GM, de la Chapelle A, Vogelstein B, and Kinzler KW. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* (1994). 371, 75-80.

Nishida T, Hirota S, Taniguchi M, Hashimoto K, Isozaki K, Nakamura H, Kanakura Y, Tanaka T, Takabayashi A, Matsuda H, and Kitamura Y. Familial gastrointestinal stromal tumours with germline mutation of the KIT gene. *Nat.Genet.* (1998). 19, 323-4.

Nishisho I, Nakamura Y, Miyoshi Y, Miki Y, Ando H, Horii A, Koyama K, Utsunomiya J, Baba S, and Hedge P. Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science* (1991). 253, 665-9.

Nowell PC. The clonal evolution of tumor cell populations. *Science* (1976). 194, 23-8.

Nyström-Lahti M, Kristo P, Nicolaides NC, Chang SY, Aaltonen LA, Moisio AL, Järvinen HJ, Mecklin JP, Kinzler KW, and Vogelstein B. Founding mutations and Alu-mediated recombination in hereditary colon cancer. *Nat.Med.* (1995). 1, 1203-6.

Ogawa O, Becroft DM, Morison IM, Eccles MR, Skeen JE, Mauger DC, and Reeve AE. Constitutional relaxation of insulin-like growth factor II gene imprinting associated with Wilms' tumour and gigantism. *Nat.Genet.* (1993). 5, 408-12.

Pai R, Tarnawski AS, and Tran T. Deoxycholic acid activates beta-catenin signaling pathway and increases colon cell cancer growth and invasiveness. *Mol.Biol.Cell* (2004). 15, 2156-63.

Papadopoulos N, Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen CA, Haseltine WA, Fleischmann RD, Fraser CM, Adams MD, Venter JC, Hamilton SR, Petersen GM, Watson P, Lynch HT, Peltomäki P, Mecklin JP, de la Chapelle A, Kinzler KW, and Vogelstein B. Mutation of a mutL homolog in hereditary colon cancer. *Science* (1994). 263, 1625-9.

Parkin DM, Bray F, Ferlay J, and Pisani P. Global cancer statistics, 2002. *CA Cancer.J.Clin.* (2005). 55, 74-108.

Parsons R, Myeroff LL, Liu B, Willson JK, Markowitz SD, Kinzler KW, and Vogelstein B. Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. *Cancer Res.* (1995). 55, 5548-50.

Pawlik TM, Raut CP, and Rodriguez-Bigas MA. Colorectal carcinogenesis: MSI-H versus MSI-L. *Dis.Markers* (2004). 20, 199-206.

Peltomäki P and Vasen H. Mutations associated with HNPCC predisposition -- Update of ICG-HNPCC/INSiGHT mutation database. *Dis.Markers* (2004). 20, 269-76.

Perucho M. Tumors with microsatellite instability: many mutations, targets and paradoxes. *Oncogene* (2003). 22, 2223-5.

Petko Z, Ghiassi M, Shuber A, Gorham J, Smalley W, Washington MK, Schultenover S, Gautam S, Markowitz SD, and Grady WM. Aberrantly methylated CDKN2A, MGMT, and MLH1 in colon polyps and in fecal DNA from patients with colorectal polyps. *Clin.Cancer Res.* (2005). 11, 1203-9.

Peto J and Houlston RS. Genetics and the common cancers. *Eur.J.Cancer* (2001). 37 Suppl 8, S88-96.

Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, Hutchinson RG, Ferrell RE, Boerwinkle E, and Shriver MD. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am.J.Hum.Genet.* (2001). 68, 198-207.

Plasilova M, Russell AM, Wanner A, Wolf A, Dobbie Z, Muller HJ, and Heinimann K. Exclusion of an extracolonic disease modifier locus on chromosome 1p33-36 in a large Swiss familial adenomatous polyposis kindred. *Eur.J.Hum.Genet.* (2004). 12, 365-71.

Potter JD. Colorectal cancer: molecules and populations. *J.Natl.Cancer Inst.* (1999). 91, 916-32.

Powell SM, Zilz N, Beazer-Barclay Y, Bryan TM, Hamilton SR, Thibodeau SN, Vogelstein B, and Kinzler KW. APC mutations occur early during colorectal tumorigenesis. *Nature* (1992). 359, 235-7.

Quehenberger F, Vasen HF, and van Houwelingen HC. Risk of colorectal and endometrial cancer for carriers of mutations of the hMLH1 and hMSH2 gene: correction for ascertainment. *J.Med.Genet.* (2005). 42, 491-6.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, and Lander ES. Linkage disequilibrium in the human genome. *Nature* (2001). 411, 199-204.

Renkonen E, Zhang Y, Lohi H, Salovaara R, Abdel-Rahman WM, Nilbert M, Aittomäki K, Järvinen HJ, Mecklin JP, Lindblom A, and Peltomäki P. Altered expression of MLH1, MSH2, and MSH6 in predisposition to hereditary nonpolyposis colorectal cancer. *J.Clin.Oncol.* (2003). 21, 3629-37.

Ricciardone MD, Özcelik T, Cevher B, Özdag H, Tuncer M, Gurgey A, Uzunalimoglu Ö, Cetinkaya H, Tanyeli A, Erken E, and Özturk M. Human MLH1 deficiency predisposes to hematological malignancy and neurofibromatosis type 1. *Cancer Res.* (1999). 59, 290-3.

Risch N and Merikangas K. The future of genetic studies of complex human diseases. *Science* (1996). 273, 1516-7.

Rivat C, Le Floch N, Sabbah M, Teyrol I, Redeuilh G, Bruyneel E, Mareel M, Matrisian LM, Crawford HC, Gespach C, and Attoub S. Synergistic cooperation between the AP-1 and LEF-1 transcription factors in activation of the matrilysin promoter by the src oncogene: implications in cellular invasion. *FASEB J.* (2003). 17, 1721-3.

Rodriguez-Bigas MA, Boland CR, Hamilton SR, Henson DE, Jass JR, Khan PM, Lynch H, Perucho M, Smyrk T, Sobin L, and Srivastava S. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. *J.Natl.Cancer Inst.* (1997). 89, 1758-62.

Rooney PH, Boonsong A, McFadyen MC, McLeod HL, Cassidy J, Curran S, and Murray GI. The candidate oncogene ZNF217 is frequently amplified in colon cancer. *J. Pathol.* (2004). 204, 282-8.

Rossi DJ, Ylikorkala A, Korsisaari N, Salovaara R, Luukko K, Launonen V, Henkemeyer M, Ristimäki A, Aaltonen LA, and Mäkelä TP. Induction of cyclooxygenase-2 in a mouse model of Peutz-Jeghers polyposis. *Proc.Natl.Acad.Sci.U.S.A.* (2002). 99, 12327-32.

Salovaara R, Loukola A, Kristo P, Kääriäinen H, Ahtola H, Eskelinen M, Härkönen N, Julkunen R, Kangas E, Ojala S, Tulikoura J, Valkamo E, Järvinen H, Mecklin JP, Aaltonen LA, and de la Chapelle A. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. *J.Clin.Oncol.* (2000). 18, 2193-200.

Santibanez-Koref MF, Gangeswaran R, and Hancock JM. A relationship between lengths of microsatellites and nearby substitution rates in mammalian genomes. *Mol.Biol.Evol.* (2001). 18, 2119-23.

Sawyer EJ, Cerar A, Hanby AM, Gorman P, Arends M, Talbot IC, and Tomlinson IP. Molecular characteristics of serrated adenomas of the colorectum. *Gut* (2002). 51, 200-6.

Schmidt L, Duh FM, Chen F, Kishida T, Glenn G, Choyke P, Scherer SW, Zhuang Z, Lubensky I, Dean M, Allikmets R, Chidambaram A, Bergerheim UR, Feltis JT, Casadevall C, Zamarron A, Bernues M, Richard S, Lips CJ, Walther MM, Tsui LC, Geil L, Orcutt ML, Stackhouse T, Lipan J, Slife L, Brauch H, Decker J, Niehans G, Hughson MD, Moch H, Storkel S, Lerman MI, Linehan WM, and Zbar B. Germline and somatic mutations in the tyrosine kinase domain of the MET proto-oncogene in papillary renal carcinomas. *Nat.Genet.* (1997). 16, 68-73.

Shibata D, Peinado MA, Ionov Y, Malkhosyan S, and Perucho M. Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. *Nat.Genet.* (1994). 6, 273-81.

Shibata D, Navidi W, Salovaara R, Li ZH, and Aaltonen LA. Somatic microsatellite mutations as molecular tumor clocks. *Nat.Med.* (1996). 2, 676-81.

Shibata D. Talking tumours: histories from sequences. *Ann.Med.* (2001). 33, 148-52.

Sieber OM, Howarth KM, Thirlwell C, Rowan A, Mandir N, Goodlad RA, Gilkar A, Spencer-Dene B, Stamp G, Johnson V, Silver A, Yang H, Miller JH, Ilyas M, and Tomlinson IP. Myh deficiency enhances intestinal tumorigenesis in multiple intestinal neoplasia (ApcMin/+) mice. *Cancer Res.* (2004). 64, 8876-81.

Silverman KA, Koratkar R, Siracusa LD, and Buchberg AM. Identification of the modifier of Min 2 (Mom2) locus, a new mutation that influences Apc-induced intestinal neoplasia. *Genome Res.* (2002). 12, 88-97.

Silverman KA, Koratkar RA, Siracusa LD, and Buchberg AM. Exclusion of Madh2, Madh4, and Madh7 as candidates for the modifier of Min 2 (Mom2) locus. *Mamm.Genome* (2003). 14, 119-29.

Smith KJ, Levy DB, Maupin P, Pollard TD, Vogelstein B, and Kinzler KW. Wild-type but not mutant APC associates with the microtubule cytoskeleton. *Cancer Res.* (1994). 54, 3672-5.

Smyrk TC and Lynch HT. Microsatellite instability: impact on cancer progression in proximal and distal colorectal cancers. *Eur.J.Cancer* (1999). 35, 171-2.

Souza RF, Wang S, Thakar M, Smolinski KN, Yin J, Zou TT, Kong D, Abraham JM, Toretsky JA, and Meltzer SJ. Expression of the wild-type insulin-like growth factor II receptor gene suppresses growth and causes death in colorectal carcinoma cells. *Oncogene* (1999). 18, 4063-8.

Spirio L, Olschwang S, Groden J, Robertson M, Samowitz W, Joslyn G, Gelbert L, Thliveris A, Carlson M, and Otterud B. Alleles of the APC gene: an attenuated form of familial polyposis. *Cell* (1993). 75, 951-7.

Stephens JC, Briscoe D, and O'Brien SJ. Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. *Am.J.Hum.Genet.* (1994). 55, 809-24.

Strand M, Prolla TA, Liskay RM, and Petes TD. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* (1993). 365, 274-6.

Sun T, Miao X, Wang J, Tan W, Zhou Y, Yu C, and Lin D. Functional Phe31Ile polymorphism in Aurora A and risk of breast carcinoma. *Carcinogenesis* (2004). 25, 2225-30.

Suzuki K, Dai T, Suzuki I, Dai Y, Yamashita K, and Perucho M. Low mutation incidence in polymorphic noncoding short mononucleotide repeats in gastrointestinal cancer of the microsatellite mutator phenotype pathway. *Cancer Res.* (2002). 62, 1961-5.

Thibodeau SN, Bren G, and Schaid D. Microsatellite instability in cancer of the proximal colon. *Science* (1993). 260, 816-9.

Thibodeau SN, French AJ, Cunningham JM, Tester D, Burgart LJ, Roche PC, McDonnell SK, Schaid DJ, Vockley CW, Michels VV, Farr GH,Jr, and O'Connell MJ. Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1. *Cancer Res.* (1998). 58, 1713-8.

Thorlacius S, Jonasdottir O, and Eyfjord JE. Loss of heterozygosity at selective sites on chromosomes 13 and 17 in human breast carcinoma. *Anticancer Res.* (1991). 11, 1501-7.

Tishkoff DX, Boerger AL, Bertrand P, Filosi N, Gaida GM, Kane MF, and Kolodner RD. Identification and characterization of Saccharomyces cerevisiae EXO1, a gene encoding an exonuclease that interacts with MSH2. *Proc.Natl.Acad.Sci.U.S.A.* (1997). 94, 7487-92.

Tomlinson IP, Beck NE, Neale K, and Bodmer WF. Variants at the secretory phospholipase A2 (PLA2G2A) locus: analysis of associations with familial adenomatous polyposis and sporadic colorectal tumours. *Ann.Hum.Genet.* (1996). 60 (Pt 5), 369-76.

Tomlinson I, Rahman N, Frayling I, Mangion J, Barfoot R, Hamoudi R, Seal S, Northover J, Thomas HJ, Neale K, Hodgson S, Talbot I, Houlston R, and Stratton MR. Inherited susceptibility to colorectal adenomas and carcinomas: evidence for a new predisposition gene on 15q14-q22. *Gastroenterology* (1999). 116, 789-95.

Toth G, Gaspari Z, and Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* (2000). 10, 967-81.

Tsao JL, Yatabe Y, Salovaara R, Järvinen HJ, Mecklin JP, Aaltonen LA, Tavare S, and Shibata D. Genetic reconstruction of individual colorectal tumor histories. *Proc.Natl.Acad.Sci.U.S.A.* (2000). 97, 1236-41.

Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, Jass J, Lindblom A, Lynch HT, Peltomäki P, Ramsey SD, Rodriguez-Bigas MA, Vasen HF, Hawk ET, Barrett JC, Freedman AN, and Srivastava S. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J.Natl.Cancer Inst.* (2004). 96, 261-8.

Upton A, Johnson N, Sandy J, and Sim E. Arylamine N-acetyltransferases - of mice, men and microorganisms. *Trends Pharmacol.Sci.* (2001). 22, 140-6.

van der Luijt RB, Meera Khan P, Vasen HF, Breukel C, Tops CM, Scott RJ, and Fodde R. Germline mutations in the 3' part of APC exon 15 do not result in truncated proteins and are associated with attenuated adenomatous polyposis coli. *Hum.Genet.* (1996). 98, 727-34.

Vasen HF, Mecklin JP, Khan PM, and Lynch HT. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis.Colon Rectum* (1991). 34, 424-5.

Vasen HF, Wijnen JT, Menko FH, Kleibeuker JH, Taal BG, Griffioen G, Nagengast FM, Meijers-Heijboer EH, Bertario L, Varesco L, Bisgaard ML, Mohr J, Fodde R, and Khan PM. Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis. *Gastroenterology* (1996). 110, 1020-7.

Vasen HF, Watson P, Mecklin JP, and Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology* (1999). 116, 1453-6.

Vasen HF, Stormorken A, Menko FH, Nagengast FM, Kleibeuker JH, Griffioen G, Taal BG, Moller P, and Wijnen JT. MSH2 mutation carriers are at higher risk of cancer than MLH1 mutation carriers: a study of hereditary nonpolyposis colorectal cancer families. *J.Clin.Oncol.* (2001). 19, 4074-80.

Veigl ML, Kasturi L, Olechnowicz J, Ma AH, Lutterbaugh JD, Periyasamy S, Li GM, Drummond J, Modrich PL, Sedwick WD, and Markowitz SD. Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc.Natl.Acad.Sci.U.S.A.* (1998). 95, 8698-702.

Vilkki S, Launonen V, Karhu A, Sistonen P, Vastrik I, and Aaltonen LA. Screening for microsatellite instability target genes in colorectal cancers. *J.Med.Genet.* (2002). 39, 785-9.

Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM, and Bos JL. Genetic alterations during colorectal-tumor development. *N.Engl.J.Med.* (1988). 319, 525-32.

Vogelstein B and Kinzler KW. Cancer genes and the pathways they control. *Nat.Med.* (2004). 10, 789-99.

Wang J, Sun L, Myeroff L, Wang X, Gentry LE, Yang J, Liang J, Zborowska E, Markowitz S, and Willson JK. Demonstration that mutation of the type II transforming growth factor beta receptor inactivates its tumor suppressor activity in replication error-positive colon carcinoma cells. *J.Biol.Chem.* (1995). 270, 22044-9.

Wang Q, Lasset C, Desseigne F, Frappaz D, Bergeron C, Navarro C, Ruano E, and Puisieux A. Neurofibromatosis and early onset of cancers in hMLH1-deficient children. *Cancer Res.* (1999). 59, 294-7.

Wang Q, Montmain G, Ruano E, Upadhyaya M, Dudley S, Liskay RM, Thibodeau SN, and Puisieux A. Neurofibromatosis type 1 gene as a mutational target in a mismatch repair-deficient cell type. *Hum.Genet.* (2003). 112, 117-23.

Wang B, Ren J, Ooi LL, Chong SS, and Lee CG. Dinucleotide repeats negatively modulate the promoter activity of Cyr61 and is unstable in hepatocellular carcinoma patients. *Oncogene* (2005). 24, 3999-4008.

Watson P and Lynch HT. Extracolonic cancer in hereditary nonpolyposis colorectal cancer. *Cancer* (1993). 71, 677-85.

Whiteside D, McLeod R, Graham G, Steckley JL, Booth K, Somerville MJ, and Andrew SE. A homozygous germ-line mutation in the human MSH2 gene predisposes to hematological malignancy and multiple cafe-au-lait spots. *Cancer Res.* (2002). 62, 359-62.

Wiesner GL, Daley D, Lewis S, Ticknor C, Platzer P, Lutterbaugh J, MacMillen M, Baliner B, Willis J, Elston RC, and Markowitz SD. A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. *Proc.Natl.Acad.Sci.U.S.A.* (2003). 100, 12961-5.

Wijnen J, de Leeuw W, Vasen H, van der Klift H, Møller P, Stormorken A, Meijers-Heijboer H, Lindhout D, Menko F, Vossen S, Møslein G, Tops C, Bröcker-Vriends A, Wu Y, Hofstra R, Sijmons R, Cornelisse C, Morreau H, and Fodde R. Familial endometrial cancer in female carriers of MSH6 germline mutations. *Nat.Genet.* (1999). 23, 142-4.

Woerner SM, Benner A, Sutter C, Schiller M, Yuan YP, Keller G, Bork P, Doeberitz MK, and Gebert JF. Pathogenesis of DNA repair-deficient cancers: a statistical meta-analysis of putative Real Common Target genes. *Oncogene* (2003). 22, 2226-35.

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, and Micklem G. Identification of the breast cancer susceptibility gene BRCA2. *Nature* (1995). 378, 789-92.

Worthley DL, Walsh MD, Barker M, Ruszkiewicz A, Bennett G, Phillips K, and Suthers G. Familial mutations in PMS2 can cause autosomal dominant hereditary nonpolyposis colorectal cancer. *Gastroenterology* (2005). 128, 1431-6.

Wu Y, Berends MJ, Sijmons RH, Mensink RG, Verlind E, Kooi KA, van der Sluis T, Kempinga C, van dDer Zee AG, Hollema H, Buys CH, Kleibeuker JH, and Hofstra RM. A role for MLH3 in hereditary nonpolyposis colorectal cancer. *Nat.Genet.* (2001). 29, 137-8.

Yamamoto H, Gil J, Schwartz S,Jr, and Perucho M. Frameshift mutations in Fas, Apaf-1, and Bcl-10 in gastro-intestinal cancer of the microsatellite mutator phenotype. *Cell Death Differ.* (2000). 7, 238-9.

Yu B, Li SY, An P, Zhang YN, Liang ZJ, Yuan SJ, and Cai HY. Comparative study of proteome between primary cancer and hepatic metastatic tumor in colorectal cancer. *World J.Gastroenterol.* (2004). 10, 2652-6.

Zeng Q, Dong JM, Guo K, Li J, Tan HX, Koh V, Pallen CJ, Manser E, and Hong W. PRL-3 and PRL-1 promote cell migration, invasion, and metastasis. *Cancer Res.* (2003). 63, 2716-22.

Zhang L, Yu J, Willson JK, Markowitz SD, Kinzler KW, and Vogelstein B. Short mononucleotide repeat sequence variability in mismatch repair-deficient cancers. *Cancer Res.* (2001). 61, 3801-5.

Zhou H, Kuang J, Zhong L, Kuo WL, Gray JW, Sahin A, Brinkley BR, and Sen S. Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat.Genet.* (1998). 20, 189-93.

Zuo L, Weger J, Yang Q, Goldstein AM, Tucker MA, Walker GJ, Hayward N, and Dracopoli NC. Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma. *Nat.Genet.* (1996). 12, 97-9.

ISBN 952-91-9348-3 (paperback)
ISBN 952-10-2730-4 (PDF)
http://ethesis.helsinki.fi
Yliopistopaino
Helsinki 2005