Genetics and Tumor Genomics in Familial Colorectal Cancer

Genetics and Tumor Genomics in Familial Colorectal Cancer

Printing: Ipskamp Drukkers
Layout cover: Ipskamp Drukkers

The studies described in this thesis were performed at the Department of Pathology and the Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands.

The research described in this thesis was financially supported by the Dutch Cancer Society (UL2005-3247) and the NutsOhra Foundation (SNO-T-07-092).

The printing of this thesis was financially supported by the Dutch Cancer Society, Stichting Nationaal Fonds tegen Kanker - voor onderzoek naar reguliere en alternatieve therapieën, and the J.E. Jurriaanse Stichting.

Genetics and Tumor Genomics in Familial Colorectal Cancer

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 14 oktober 2010
klokke 15.00 uur

door

Janneke Willemijn Middeldorp
geboren te Deventer
in 1980

Promotiecommissie

Promotores:	Prof. dr. H. Morreau
	Prof. dr. P. Devilee
Copromotores:	Dr. T. van Wezel
	Dr. J. Wijnen
Overige leden:	Dr. F.J. Hes
	Prof. dr. N.C. Hoogerbrugge (Radboud University Medical Center)
	Prof. dr. G.A. Meijer (VU Medical Center)

Contents

Aims and outline of this thesis

The aim of the work described in this thesis was to identify novel genes that predispose to colorectal cancer (CRC). Colorectal cancer is one of the most common malignancies in the Western world. The lifetime risk in the Netherlands for developing CRC is about 6%, with a 5-year survival of about 55%. The risk of CRC can be increased both by genetic and environmental factors. In twin studies it was estimated that in up to 30% of all CRC cases inherited predisposition plays a role. However, only 6% of all cases is explained by the known syndromes, like Lynch syndrome, Familial Adenomatous Polyposis (FAP) and *MUTYH*-associated polyposis (MAP). In the remaining familial cases the underlying genetics remain elusive. In the cases that show a strong familial history of CRC, a high penetrance genetic factor can be expected to be responsible for the increased CRC risk. In other cases with a less strong family history but with an early onset of the disease, the increased cancer risk might be explained by common genetic variants, each conferring a small CRC risk. It is of great importance to identify both the low and high risk factors, because this knowledge can aid in obtaining further insight in the etiology of the disease. Moreover, identifying high risk factors will aid in identifying individuals at significant increased risk for CRC. These individuals can then be offered a regular screening of their colon to detect precursor lesions, and thereby prevent the disease from developing into a malignant lesion.

Chapter 1 provides a general introduction in the genetic and environmental factors that influence the risk of CRC. Moreover, genetic loci that have been identified to possibly harbor a high risk gene are discussed as well as low level risk loci that have been identified. Colorectal tumorigenesis is briefly discussed and the genetic en genomic instability seen in colorectal tumors is described, both for sporadic and familial tumors. Finally, new fields in CRC research, prevention, and treatment are reviewed.

Different approaches were used to identify novel genetic factors predisposing to CRC. These approaches can be broadly divided into germ-line genetic approaches and somatic genomic approaches. **Chapter 2 and 3** describe the germ-line approach to identify genetic loci harboring high or moderate penetrance risk factors. Linkage analysis was used in large families affected with CRC. The linkage scan was performed using 10K SNP arrays. With this high number of markers and large pedigrees the computational burden of the linkage analysis was high. The procedure we developed to handle this complexity is described in **Chapter 2**. The results of the linkage analysis in seven large CRC families are described in **Chapter 3**.

In the search for low risk factors we replicated the association of six loci, identified in large genome-wide association studies, in a Dutch clinical-based cohort of 995 familial CRC patients. We studied the possible association of these loci with several clinico-pathological parameters. The results of this study are described in **Chapter 4**.

The third approach we used was to study the genomic profile of the tumors familial CRC patients develop. With profiling the tumors from patients with familial CRC we aimed on one side at stratifying the different families based on the genomic profile of their tumors and moreover,

we aimed at identifying a locus that is frequently affected in these tumors indicating that there might be a gene important for the development of the disease located in that region. To properly interpret the profile of the tumors from patients with an unknown cause of their increased CRC predisposition, we first generated the profile for known CRC syndromes. The results of the profiling of tumors from patients with *MUTYH*-associated polyposis are described in **Chapter 5**. We also studied the profile of tumors from Lynch syndrome patients in the context of another PhD thesis. For the familial cases with unknown cause we studied the profile of 30 tumors originating from 15 families (2 tumors per family were studied). The results of this study are described in **Chapter 6**. And in **Chapter 7**, an improved method for tumor profiling is described, with which the allelic state of the chromosomes can be determined.
In **Chapter 8**, all results described in this thesis are discussed and perspectives for future research are given. **Chapter 9** provides a summary of this thesis.

Chapter 1

General Introduction

Colorectal cancer (CRC) is one of the most common cancers in the Western world.[1] In the Netherlands, each year around 11,000 patients are diagnosed with colorectal cancer.[2] The lifetime risk in the Netherlands for developing colorectal cancer is approximately 6%, with a five-year survival of about 55%.[1] The age distribution is wide; however, over half of the patients are diagnosed with colorectal cancer above 70 years of age.

The risk of CRC is influenced by both genetic and environmental factors. Although most colorectal cancer arises on a sporadic basis, in 10-30% of the cases inherited predisposition plays a role, as was estimated in twin studies.[3] First-degree relatives of CRC patients are at increased risk of developing colorectal cancer, with a relative risk of about 2.2. This risk is strongly correlated with the number of affected family members and an early onset of disease.[4] With two or more affected family members, the relative risk increases to about 4.0.[5,6] Individuals with a first-degree relative affected with colorectal adenomas are also at increased risk of CRC, with a relative risk of about 2.0.[6] The risk increases when the age at diagnosis of colorectal adenomas of the first-degree relative decreases.[7,8] About 6% of the CRC cases can be explained by several colorectal cancer syndromes, including Lynch syndrome, familial adenomatous polyposis (FAP), and *MUTYH*-associated polyposis (MAP). However, for the other familial cases, the underlying genetics remain elusive. High or moderate risk factors could play a role in families affected with colorectal cancer. Moreover, co-inheritance of several low risk factors could explain the excess risk in such cases.

Tumorigenesis

The development of colorectal cancer is a multistep process that involves somatic genetic and epigenetic changes. Several genes acquire a mutation that provides the cell with a growth advantage. The transcription of genes and microRNAs is dysregulated in cancers; tumor suppressor genes are shut down and oncogenes are activated.

Colorectal cancer generally develops from normal epithelium through different adenoma stages with increasing dysplasia into carcinoma.[9] A genetic model for the development of sporadic CRC was first postulated by Fearon and Vogelstein in 1990 (Figure 1).[10] They described the genetic changes of colorectal tumors along their development from adenoma to carcinoma and they assumed that – although the accumulation of mutations seems most important - the genetic changes occur in a specific order in most CRCs. Later, more genes and genetic aberrations were added to this model.[11,12]

Recent advances have provided more insight in the biology of the colonic crypts and their relation to tumorigenesis (reviewed by [13]). Colonic crypts are finger-like invaginations of the colonic epithelial layer in the underlying connective tissue of the lamina propria. At the base of each crypt stem cells are located that are capable of regenerating all intestinal cell types. When a stem cell acquires a mutation it can, through a selective advantage or genetic drift, fill

[1] http://www.ikcnet.nl

up the whole crypt. Subsequently, crypt fission can lead to spreading of the mutation in neighboring epithelium. Such aberrant crypt foci (ACF), consisting of a cluster of a small number of abnormal crypts, are thought to be an early step in the formation of adenomatous polyps.[13] Colorectal cancer development is often initiated by mutations in *APC*, which leads to an increased proliferation of the cell. *APC* mutations and/or loss of heterozygosity at the *APC* locus are found in over 80% of all colorectal cancers.[14-16] Mutations in *CTNNB1*, encoding β-catenin, have also been found to occur in CRC.[17] *KRAS* mutations are seen with progression of the adenoma. Later, in the progression from adenoma to carcinoma, additional mutations are acquired in for example *SMAD2* and *SMAD4*, and *p53*, often accompanied by loss of heterozygosity (LOH) of 17p and 18q.[11,18]
Recent studies show that the majority of mutations in a tumor occur in genes other than these commonly mutated genes, leading to a unique mutational signature for each tumor.[19,20] However, the role of these other mutations in tumorigenesis still needs to be elucidated.
The carcinogenesis model described above applies mainly to sporadic CRC. In hereditary syndromes, the carcinogenesis model is different depending on the germ-line defect.

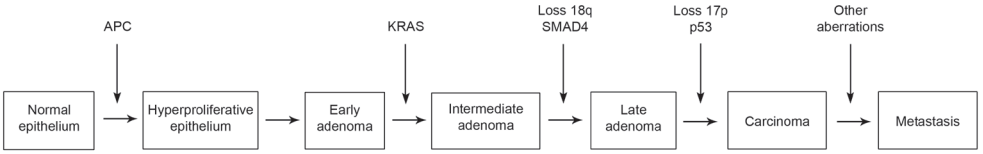

Figure 1. Genetic model of colorectal tumorigenesis as proposed by Fearon and Vogelstein [10]

Signaling Pathways

Several signaling pathways are involved in colorectal tumorigenesis. These pathways become dysregulated via mutational activation or inactivation of one of its proteins or via other ways of (in)activation like methylation.

Wnt signaling pathway
Activation of the Wnt signaling cascade is considered to be the initiating event in colorectal cancer. In the Wnt signaling cascade, β-catenin functions as a transcription factor upon binding to nuclear proteins of the Tcf family and regulates genes involved in cellular activation. APC is a key regulator of β-catenin because it regulates the levels of cytoplasmic β-catenin. APC forms a complex with GSK3β, axin, and β-catenin, that modulates the cytoplasmic β-catenin levels by degradation. Mutational inactivation of APC is observed in over 80% of the CRCs and leads to accumulation of β-catenin in the cytoplasm.[14-16] This accumulation leads to

dysregulation of the transcriptional targets of β-catenin/Tcf.[21] Fifty percent of the tumors that lack mutations in *APC*, display mutations in the β-catenin gene *CTNNB1*.[22] Examples of genes that are regulated by β-catenin/Tcf include cell cycle regulator *c-MYC*, G1/S-regulating *cyclin D1*, and *MMP-7* (matrilysin), a matrix-degrading metalloproteinase.[23-25] Germ-line mutations in the *APC* gene give rise to an inherited cancer predisposing syndrome, familial adenomatous polyposis (discussed below).

TGF-β signaling pathway

The transforming growth factor β (TGF-β) signaling pathway regulates the proliferation and differentiation of cells, embryonic development, wound healing, and angiogenesis.[26] TGF- β can bind three high-affinity cell-surface receptors (type I, II, and III). The intra-cellular domains of receptor type I and II contain serine-threonine protein kinases that initiate phosphorylation of SMAD transcription factors. Receptor type III binds TGF-β and transfers it to the signaling receptors type II. Upon binding of TGF-β, receptor type II forms a complex with receptor type I and phosphorylates the receptor, thereby stimulating the kinase-activity of the receptor.[27] The activated receptor type I phosphorylates SMAD2 and SMAD3 which then bind to SMAD4. This SMAD-complex then translocates to the nucleus, where it acts as a transcription factor. In epithelial cells, TGF-β inhibits cellular proliferation, which explains its tumor suppressor function.

Mutation inactivation of the TGF-β pathway is frequent in colorectal tumors. About 30% of the sporadic CRCs have a mutation in the *SMAD4* gene.[28] Moreover, the TGF-β receptor type II (TGFRB2) is frequently mutated in tumors with DNA mismatch repair deficiency (discussed below) and about 15% of DNA mismatch repair proficient colorectal carcinomas display mutational inactivation of TGFBR2.[29,30] Germ-line mutations in *SMAD4* are found in patients with Juvenile Polyposis (discussed below).[31]

Bone morphogenetic proteins (BMPs) also act in the TGF-β signaling pathway. Similar to TGF-β, BMPs bind to serine-threonine kinase receptors type 1 and 2 (BMPR1 and BMPR2). BMPR2 then phosphorylates and activates receptor type I. The activated BMPR1 subsequently phosphorylates SMAD1, SMAD5, and SMAD8, which associate with SMAD4 and translocate to the nucleus (reviewed by [32]). In addition to *SMAD4* mutations, germ-line mutations in *BMPR1A* are also found in Juvenile Polyposis patients.[33]

p53 signaling pathway

The p53 pathway is an important pathway that is involved in cell-cycle arrest upon cellular stress. It thereby acts as a tumor suppressor.[34] When cellular stresses occur, including DNA damage or hypoxia, p53 is stabilized and binds to the DNA to act as a transcription regulator. Genes regulated by p53 regulates are involved in important cellular processes like DNA repair, cell-cycle arrest, senescence, and apoptosis. Genes that are transcriptionally activated by p53 include *p21*, *MDM2*, *GADD45*, and *Bax*.[35] The pathway is often inactivated in CRC by an inactivating (missense) mutation in the p53 gene and subsequent deletion of the second

allele of the *p53* gene through physical loss of chromosome arm 17p.[36,37] Inactivation of *p53* coincides with the transition of an adenoma to the carcinoma stage.[38]

MAPK signaling pathway

Mitogen-activated protein kinases (MAPK) cascade is an important signaling pathway involved in cellular proliferation. Three major subfamilies of MAP kinases include the extra-cellular-signal-regulated kinases, stress-activated protein kinases, and MAPK14. The subfamily of extracellular-signal-regulated kinases (ERK MAPK), including Ras, Raf, MEK, and ERK, is important for intestinal epithelial differentiation.[39] Upon binding of extracellular signal proteins the Ras/Raf/MEK/ERK cascade transmits the signal to the nucleus where it regulates the transcription of genes involved in proliferation and differentiation. The cascade includes several proto-oncogenes and is dysregulated in about 30% of all cancers. Activating mutations in *KRAS* are found in approximately 38% of colorectal cancer.[40] *BRAF* mutations are observed in about 10% of colorectal cancers and are particularly observed in tumors with *MLH1* promoter hypermethylation.[41,42] Mutations in *KRAS* and *BRAF* appear to be mutually exclusive.[42,43]

Activation of the epidermal growth factor receptor (EGFR) stimulates the MAPK signaling pathway and is a target for cancer treatment. Oncogenic activation of the Ras/Raf/MEK/ERK cascade in colorectal cancer is strongly correlated with impaired response to anti-EGFR treatments like panitumumab and cetuximab.[44]

PI3K signaling pathway

Phosphatidylinositol 3-kinase (PI3K) is an enzyme that phosphorylates inositol phospholipids (IPs). The phosphorylated IPs (PIPs) subsequently activate the downstream AKT pathway. Activation of AKT leads to increased cell survival, growth and cellular proliferation (reviewed by [45]). Protein targets of AKT include mTor, Bad, Caspase 9, Tuberin, and GSK3β.[46] About 32% of colorectal tumors carry an activating mutation in *PIK3CA*, encoding the catalytic subunit.[47] PTEN functions as an inhibitor of the PI3K pathway because it dephosphorylates the inositol phospholipids. The PI3K pathway is interlinked with the MAPK pathway and is also stimulated upon EGFR activation. Consequently, inactivating PTEN mutations or activating PI3K mutations have been associated with reduced response to anti-EGFR treatment.[48,49]

Genetic instability

In addition to acquiring mutations in genes, colorectal tumors become genetically instable. The different forms of genetic instability involved in colorectal tumorigenesis are discussed below.

Microsatellite instability

Microsatellites are short repetitive sequences within the DNA. An accumulation of mutations in microsatellites, that make them shorter or longer, is called microsatellite instability (MSI or MIN). This microsatellite instability is caused by a defective DNA mismatch repair system, which fails to repair errors during DNA replication. MSI-H (MSI-high) is typically seen in tumors from patients with Lynch syndrome, which carry mutations in DNA mismatch repair genes *MLH1*, *MSH2*, *MSH6*, or *PMS2*. However, MSI is also observed in sporadic colorectal cancers due to hypermethylation of the promoter region of *MLH1*.[50-52] This is observed in about 15% of all colorectal cancers.[53,54] *BRAF* mutations are common in sporadic MSI-H tumors, whereas no *BRAF* mutations are observed in MSI-H Lynch tumors.[42]

Mutation rates are at least 100-fold increased in MMR deficient cells as compared to MMR proficient cells.[55,56] Via this pathway several genes that contain repeat sequences are targeted. Frequently targeted genes include growth factor receptors *TGFB-RII* and *IGFRII*, pro-apoptotic gene *BAX*, and *Caspase 5*.[30,57-59] MSI tumors are generally diploid or near-diploid and gross chromosomal gains and losses are absent.[60-62] However, some studies reported the presence of copy number aberrations in a minor fraction of MSI-H tumors, including gains of chromosomes 4, 8, 12, 13, and 20 and losses of chromosomes 1, 9, 11, and 15.[63,64]

Chromosomal instability

Chromosomal instability (CIN) is seen in the majority of CRCs and is characterized by numerous chromosomal gains and losses and copy neutral loss of heterozygosity (cnLOH). Many studies analyzed the genomic profile of tumors for gains and losses of chromosomes or chromosome arms using array comparative genomic hybridization. With recent advances in technology, genome-wide LOH studies have become feasible using SNP arrays. These SNP arrays not only provide genome-wide LOH information, but can also detect LOH in the absence of copy number aberrations (copy neutral LOH) which for example can arise via mitotic recombination or a physical loss followed by reduplication. CIN tumors are mostly aneuploid, reflecting their chromosomal instability.

Based on a comprehensive meta-analysis of tumor genome profiling studies, a model of genetic colorectal tumor progression has been established.[65] Losses of chromosomes 17p and 18 and gains of chromosomes 8q, 13q, and 20 occur early during tumor development. Losses of chromosomes 4p and 8p, and gains of 7p and 17q are correlated with the transition from primary tumor to liver metastasis. Loss of chromosome 14q and gains of chromosomes 1q, 11, 12p, and 19 are considered to be late events in tumor progression.[65]

Table 2 provides an overview of genomic profiling studies that have been performed since this meta-analysis was published. Most studies used metaphase-based CGH or array CGH to study genomic aberrations and only a minority of the studies used SNP arrays, which also provide information on LOH. Moreover, in most studies sporadic colorectal tumors were analyzed. Specific subgroups of colorectal cancer – for example MSS familial tumors or tumours

from Lynch patients - were studied to a much lesser extent.
Although the molecular basis for CIN is still unknown, several genes have been associated with the CIN pathway. Mitotic checkpoint genes *hBUB1* and *hBUBR1* have been described to contribute to chromosomal instability.[66,67] Tumor suppressor gene *p53*, involved in G1 arrest and apoptosis, is mutated in about 50% of sporadic colorectal cancers.[36] Inactivation of *p53* has been associated with the development of aneuploidy in cancers, because of loss of the arrest at the G1 checkpoint.[68] Mutations in the adenomatous polyposis coli gene *APC*, a member of the Wnt signal transduction pathway, occur early in tumorigenesis and have been linked to chromosomal instability.[69] Other genes that are frequently targeted by point mutations in CIN tumors are oncogene *KRAS*, whose activation leads to growth promotion, and *SMAD4*, a component of the TGF-β pathway.

CpG Island Methylator Phenotype
A third type of genetic instability observed in colorectal cancers is the occurrence of aberrant methylation of CpG islands, leading to the CpG island methylator phenotypes (CIMP).[70] CpG islands are short DNA sequences that are rich in CpG dinucleotides. Such islands are found in the 5' region of about half of all human genes.[71]. Hypermethylation of the cytosines in these CpG islands leads to transcriptional inactivation of such genes. Simultaneous hypermethylation of CpG islands of several genes has been termed the CIMP phenotype. Two subclasses can be distinguished: CIMP1 or CIMP-high with intense methylation of multiple genes and CIMP2 or CIMP-low with methylation of a limited number of genes. CIMP1 tumors are generally MSI-H and have a high frequency of *BRAF* mutations, whereas CIMP2 tumors are generally MSS and have a high frequency of *KRAS* mutations.[72] The cause of CIMP is currently unknown.

Hereditary Colorectal Cancer Syndromes

Several high penetrance colorectal cancer syndromes have been identified in the past two decades. The genes responsible for these syndromes have been identified using linkage analysis, deletion mapping, positional cloning, and by exploring tumor characteristics (Figure 2). The different CRC syndromes are briefly described below. An overview of all CRC syndromes as well as of polyposis syndromes is provided in Table 1. For several polyposis syndromes, including for example Cowden syndrome and hereditary mixed polyposis syndrome, a firm association with CRC has not yet been identified.

Lynch syndrome
Lynch syndrome, formerly called hereditary non-polyposis colorectal cancer or HNPCC, is the most prevalent CRC syndrome. It is estimated that Lynch syndrome accounts for at least 3% of all colorectal cancers.[73,74] Familial clustering of colorectal cancer, stomach, and uterine

cancer was already described by Warthin et al. in 1913.[75] Later, Lynch et al. described two additional families with clustering of colorectal cancer and uterine cancer.[76] In the early nineties, the genes predisposing to Lynch syndrome were identified. The syndrome is caused by germ-line mutations the DNA mismatch repair genes *MLH1*, *MSH2*, *MSH6*, and *PMS2*.[77-82] In addition, germ-line epigenetic inactivation of *MLH1*, by hypermethylation of the promoter of *MLH1*, leads to Lynch syndrome.[83] Recently, it was shown that a deletion of the last exons of *TACSTD1*, upstream of *MSH2*, also predisposes individuals to Lynch syndrome, because this deletion leads to epigenetic inactivation of *MSH2* in *TACSTD1*-expressing tissues.[84]
Lynch syndrome is characterized by the development of colorectal carcinomas at a mean age of 42 years for men and 47 years for women. The mean age at diagnosis of endometrial carcinomas in Lynch syndrome patients is approximately 47
years.[85] The colorectal cancers predominantly develop at the right side of the colon.[86] Other manifestations include cancers of the stomach, small bowel, ovaries, and the urinary tract.[87,88] The lifetime risk for developing CRC for carriers of amutation in one of the mismatch repair genes is approximately 66% for men and approximately 43% for women. The cumulative risk of endometrial cancer or CRC in women is approximately 73%.[85]

Autosomal dominant inheritable CRC without polyps	**Genes**	**MIM No**
Lynch syndrome	*MLH1, MSH2, MSH6, PMS2*	120435
Autosomal dominant inheritable CRC with adenomatous polyps		
Familial adenomatous polyposis (FAP)	*APC*	175100
Attenuated FAP (AFAP)	*APC*	175100
Autosomal dominant inheritable CRC with hamartomatous/mixed/ hyperplastic polyps		
Peutz-Jeghers syndrome (PJS)	*STK11 (LKB1)*	175200
Juvenile polyposis syndrome (JPS)	*SMAD4, BMPR1A*	174900
Hereditary hemorrhagic telangiectasia syndrome (HHT) *	*ENG, ACVRL1*	187300
Hyperplastic polyposis syndrome (HPT) *	*MUTYH, MBD4*	unassigned
Hereditary mixed polyposis syndrome (HMPS) *		601228
Cowden disease (CD) *	*PTEN*	158350
Birt-Hogg-Dube syndrome (BHT) *	*FLCN*	135150
Autosomal recessive inheritable CRC with adenomatous, serrated adenomas and hyperplastic polyps		
MUTYH-associated polyposis (MAP)	*MUTYH*	608456

Table 1. Overview of colorectal cancer syndromes and polyposis syndromes
* CRC risk unknown

Mutations in a DNA mismatch repair gene lead to a failure to repair errors during DNA replication, especially concerning mismatches and insertion/deletion loops. As a consequence, the hallmark of tumors from Lynch patients is an accumulation of errors in short repetitive sequences, so-called microsatellite instability (MSI).[54,89-91]

In 1990, clinical guidelines were established to identify Lynch syndrome families for research purposes. These so-called Amsterdam Criteria were based on family characteristics, age at diagnosis and the type of cancer.[92] However, because these criteria did not account for extracolonic cancers, new criteria (Amsterdam Criteria II) were established in 1999.[93] The Amsterdam Criteria have a high specificity, but a lower sensitivity to detect Lynch syndrome families. About 40% of the tumors from families that fulfill the Amsterdam Criteria II do not show an MSI phenotype.[94] Therefore, in 1997 the Bethesda guidelines were developed to identify tumors that should be tested for microsatellite instability, as a marker of patients that should be screened for germ-line mutations in one of the DNA mismatch repair genes.[95] These guidelines also select smaller families. In 2004, the Bethesda guidelines were revised to further improve the identification of Lynch syndrome patients.[96] Other criteria, which do not take family history into account, but select patients only on basis of early onset of colorectal adenomas and carcinomas, or recurrent disease, have been tested as well. These criteria prove a sensitive strategy to identify Lynch syndrome patients.[97] Additionally, immunohistochemical screening of all CRC patients for MLH1, MSH2, MSH6, and PMS2 has also been proposed and has so far only been implemented in Denmark. Using this approach, Lynch patients in small families and patients diagnosed at older age or with de novo mutations can also be identified.[98,99]

When a MMR gene mutation is identified in a family, colonoscopic surveillance is offered to the family. This surveillance facilitates early detection of precursor lesions, thereby preventing it from developing into a malignant lesion. In the Netherlands, large scale surveillance of Lynch syndrome families was introduced in the late 1980s.[100] It has been shown that the mortality because of CRC has decreased since the introduction of colonoscopic surveillance.[101]

Recently, two genetic factors have been found to modify the colorectal cancer risk in Lynch Syndrome families. It has been shown that two common low penetrance risk alleles (discussed below), located on 8q23.3 and 11q23.1, are associated with increased colorectal cancers risks. The latter association (11q23.1) was found in female Lynch syndrome patients only.[102]

Bi-allelic MMR mutations

Heterozygous mutations in the mismatch repair genes give rise to Lynch syndrome; however, rare cases of bi-allelic MMR mutations have also been described. Patients with homozygous or compound heterozygous mutations in *MLH1*, *MSH2*, *PMS2*, and *MSH6* develop juvenile leukemia and/or lymphoma associated with neurofibromatosis type 1.[103-109]

Muir-Torre syndrome

The Muir-Torre syndrome (MTS) is characterized by a combination of cutaneous lesions (multiple keratoacanthomas and sebaceous gland tumors) and colorectal, endometrial, urological, and upper gastrointestinal tumors.[110] Most MTS patients carry a germ-line mutation in *MSH2* or *MLH1*. Therefore, the syndrome is considered to be a clinical variant of the Lynch syndrome. In MTS families, mutations in *MSH2* are more frequent than mutations in *MLH1*, as compared to Lynch syndrome families.[111] Recently it was shown that sebaceous gland tumors can also occur in *MUTYH*-associated polyposis.[112]

Turcot's syndrome

Turcot's syndrome (TS) is a rare disorder that is clinically characterized by primary tumors of the central nervous system and colorectal polyposis. The phenotypic spectrum is broad, including various types of central nervous tumors and single adenomas are observed as well as adenomatous polyposis. In patients affected with this syndrome, germ-line mutations were identified in the *APC* gene or in one of the mismatch repair genes. The disease phenotype for patients with an *APC* mutation differs from that of patients with a mutation in one if the mismatch repair genes.[113]

Familial adenomatous polyposis

Familial adenomatous polyposis (FAP) is an autosomal dominantly inherited disease which hallmark is the development of a hundred to thousands of adenomatous polyps. FAP accounts for approximately 1% of all colorectal cancers.[114] The polyps develop already during adolescence, but symptoms usually present in the third decade of life.[115-117] The risk of developing carcinomas is nearly 100% if patients are not treated.[115] Extracolonic manifestations include epidermoid cysts, facial osteomas, thyroid carcinoma, duodenal carcinoma, and malignancies of the biliary tract.[115,116] Moreover, approximately 10% of FAP patients develop desmoid tumors, also referred to as aggressive fibromatoses.[118,119]

Germ-line mutations in the adenomatous polyposis coli (*APC*) gene were identified to be responsible for FAP.[120-124] *APC* is involved in the Wnt signaling pathway and its mutational inactivation leads to activation of the Wnt signaling through reduced degradation of β-catenin. The majority of the *APC* mutations in FAP patients concern nonsense mutations or frameshift mutations that lead to a truncated protein. Mutations at codons 1061 and 1309 account for about a third of all germ-line mutations. The tumors from FAP patients carry an additional somatic mutation in *APC.* The nature of this second-hit appears to be dependent on the type and location of the germ-line mutation.[125] The 'just-right' signaling model has been put forward as an explanation for this phenomenon. This model proposes that selection of *APC* genotypes occurs to retain some downregulation of β-catenin signaling rather than a constitutive activation of β-catenin signaling.[126]

Attenuated FAP

Attenuated FAP is a phenotypic variant of classical FAP, in which patients develop less adenomatous polyps and have a later onset of colorectal cancer. The number of polyps that patients develop can be very variable; within one family some affected members have few polyps while other family members have several hundred polyps at young age.[127-129] Attenuated FAP seems to be associated with mutations in the 3' and 5' end and exon 9 of *APC*.[129-132]

***MUTYH*-associated polyposis**

MUTYH-associated polyposis (MAP) is the only colorectal cancer syndrome described with a recessive mode of inheritance. It is caused by bi-allelic mutations in the base excision repair gene *MUTYH*.[133-135] MAP patients develop multiple polyps and cancer in their colon. However, the number of polyps is generally lower as compared to FAP patients.[134,136,137] The MUTYH protein plays a role in the repair of oxidative DNA damage: upon oxidative DNA damage, MUTYH removes incorrectly incorporated adenines opposite to an 8-oxo-guanine. As a consequence of MUTYH deficiency, somatic G:C>T:A transversions are seen in MAP patients in critical genes such as *APC* and *KRAS*. In *APC*, these transversions occur primarily in GAA DNA sequences.[133,134] In *KRAS*, a specific GGT>TGT mutation (c.34 G>T, p.Gly12Cys) is found in about 64% of MAP carcinomas.[138]

The penetrance of colorectal cancer development in MAP patients is nearly 100% at the age of 60 years.[139]

The clinical relevance of heterozygous mutations in *MUTYH* is still under debate. Recent association studies in large series using different approaches found some evidence for a modest and late onset increase in CRC risk for heterozygous mutation carriers (odds ratios 1.5-1.7).[139-141] This risk is comparable to the risk of CRC that first-degree relatives of individuals with CRC have. Most studies, however, fail to find significant associations between heterozygous *MUTYH* mutations and CRC risk. Similarly, two meta-analyses do not find evidence for such association.[142,143] However, these meta-analyses analyzed only the two hotspot mutations Y179C and G396D.

It has also been hypothesized that *MUTYH* mutations act as phenotypical modifiers in MMR mutation carriers, especially concerning *MSH6* mutation carriers. Enrichment of *MUTYH* mutations in carriers of missense mutations in *MSH6* has been described, but could not be confirmed in a second study. Furthermore, a strikingly mild phenotype was observed in a patient with a bi-allelic mutation in *MUTYH* and a *MSH6* mutation.[144-146]

Juvenile polyposis

Juvenile polyposis syndrome (JPS) is characterized by the development of multiple juvenile hamartomatous polyps in the gastrointestinal tract and an increased risk of cancer. It is a rare autosomal inherited disease with a penetrance of up to 70%.[147] About 1 in 100,000 individuals are affected with JPS and the mean age at diagnosis is 16 years.[148] JPS is caused by germ-line mutations in *SMAD4* and *BMPR1A*.[31,33] Mutations in these genes explain,

however, only about 40% of JPS patients. SMAD4 and BMPR1A are both members of the transforming growth factor-β (TGF-β) superfamily. The BMP receptor type 1A activates the transcription factor SMAD4 to downregulate cellular growth and division.

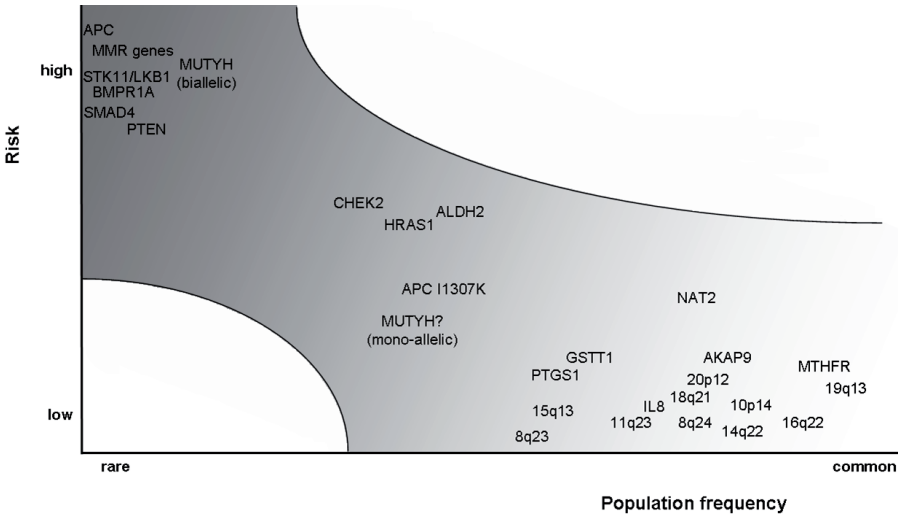

Figure 2. Genetic risk factors for colorectal cancer

This graph shows the known CRC risk factors. Rare variants with a low CRC risk are very difficult to find (lower left white area) and common variants that confer a high CRC risk are believed not to exist (upper right white area).

The MMR genes include *MLH1*, *MSH2*, *MSH6*, and *PMS2*. For the risk variant identified in genome-wide association studies their respective chromosomal locations are depicted in the graph: 8q23, rs16892766; 8q24, rs6983267; 10p14, rs10795668; 11q23, rs3802842; 14q22, rs4444235; 15q13, rs4779584; 16q22, rs9929218; 18q21, novel 1; 19q13, rs10411210; 20p14, rs961253.

Peutz-Jeghers syndrome

The Peutz-Jeghers syndrome (PJS) is an autosomal dominant inherited disease that predisposes to hamartomatous polyposis, affecting mostly the small bowel. It affects approximately 1 in 200,000 individuals and has a high penetrance. In around 50% of the patients, the syndrome is caused by germ-line mutations in *STK11* (also named *LKB1*), encoding serine/threonine protein kinase 11.[149] STK11 forms a complex with Ste adaptor protein STRAD and MO25.[150,151] This complex is involved in the regulation of cellular responses tot energy stress and cell polarity.[152] In addition to (benign) hamartomatous polyps, PJS patients develop mucocutaneous pigmentation and have an increased risk of developing malignancies in the gastrointestinal tract.[153]

Research into the characterization of the pathways in which STK11 is involved have not yet

provided a conclusive answer to the question how STK11 exerts its tumor suppressor function.[128] STK11 has been described to have a regulatory role in control of cell-cycle arrest, p53-mediated apoptosis, the Wnt signaling cascade, the ras-pathway, TGF-β signaling, and energy metabolism (reviewed by [154]).

Familial Colorectal Cancer of unknown cause

As briefly mentioned above, heritable factors may play a role in up to 30% of all CRC.[3] Family history is a strong risk factor for the development of colorectal cancer. About 6% of all CRC is explained by the colorectal cancer syndromes described above. However, for the other familial cases the underlying genetics remain elusive. Among these familial cases are large families with a clear positive family history of CRC. Such Amsterdam Criteria positive families without a mismatch repair defect have been termed 'Familial Colorectal Cancer Type X' or mismatch repair proficient CRC.[155] It has been estimated that about forty percent of the families fulfilling the Amsterdam Criteria is not affected with the Lynch syndrome, suggesting that other yet to be identified genetic factors predispose these families to CRC.[94] Moreover, it can be calculated from a Finnish study that only about 12% of the excess CRC risk associated with a family history of mismatch repair proficient CRC is explained by the known genes.[4,156]

In mismatch repair proficient CRC families, the cancer incidence is generally lower than in Lynch families.[4,155] Segregation analysis performed by Lubbe and colleagues suggested that aggregation of CRC in mismatch repair proficient families follows a recessive inheritance model.[4] Jenkins et al. also reported a role for recessively inherited factors in familial CRC.[157] The results of the study of Lindor et al. on the other hand, supported a dominant model of inheritance.[155] These differences in the suggested model of inheritance might be explained by differences in patient ascertainment. Lindor et al. included Amsterdam Criteria positive families, whereas Lubbe et al. included CRC patients diagnosed below 70 years of age. Finally, Jenkins et al. included only CRC patients with an age at diagnosis below 24 years.

High penetrance risk loci

Generally, linkage analysis is adopted to identify rare alleles that confer a high risk of familial colorectal cancer. Several loci have been identified using this approach; however, none of these have yet led to the identification of the genes responsible for the increased CRC risk in FCC families.

Wiesner and colleagues reported genetic linkage of CRC to chromosome 9q22.2-31.2 in a set of 74 affected sibling pairs.[158] Two other studies provided support for the presence of a CRC susceptibility locus in this region.[159,160] Kemp and colleagues described linkage to

colorectal cancer susceptibility of a region located on chromosome 3q21-q24.[161] In a later extension of this study including 34 additional families, linkage to this region was confirmed. However, mutation screening of 30 genes failed to identify pathogenic variants.[162] Further support for this region harboring a CRC susceptibility locus was provided by a Swedish linkage study of Picelli et al.[163] A linkage scan in 18 Swedish colorectal cancer families revealed linkage on chromosome 11, 14 and 22.[164] And finally, linkage to chromosome 7q31 has been identified using an affected sibling approach.[165]
Recently, a 111 kb copy number variable region on 3q26 was proposed to contain a regulatory element for *PPMIL* which could cause CRC susceptibility in *APC* mutation-negative polyposis families.[166] This region was identified using copy number analysis in polyps of these patients and subsequent gene expression analysis in the candidate region.

Low Penetrance Risk Alleles

Two approaches are generally used to identify low penetrance colorectal cancer risk variants. The first method is the candidate gene analysis. In candidate gene approaches, the following groups of genes are most frequently studied: genes involved in carcinogen metabolism, genes involved in methylation, genes encoding DNA repair proteins, microenvironmental modifiers, and oncogenes and tumor suppressor genes.[167,168] Two large meta-analyses of studies using the candidate gene approach, identified several polymorphisms that are associated with an increased or decreased colorectal cancer risk (Figure 2). The C677T polymorphism in *MTHFR* (methylene tetrahydrofolate reductase) is associated with a decreased CRC risk for homozygous carriers of the variant allele. *HRAS1* is a proto-oncogene that contains a variable number of tandem repeats region (VNTR). Rare alleles of this VNTR are associated with a moderately increased colorectal cancer risk. For the *NAT2* gene, only a phenotypic association was identified, i.e. a fast acetylatorship has been associated with an increased CRC risk. However, genotypic associations could not be identified. GSTT1, a detoxification enzyme, is associated with a small increase in colorectal cancer risk for the null genotype (homozygous deletion). ALDH2, a mitochondrial enzyme responsible for oxidation of acetaldehyde, is associated with in increased risk both in heterozygous and homozygous carriers of the variant allele. And finally, the variant I1307K in *APC* is associated with an increased CRC risk.[167,168] Other low risk variants that have been associated with CRC include *CHEK2* 1100delC, *AKAP9* M463I, *PTGS1* G213G, *IL8* c.-352T>A, *MTHFR* A429E.[169-171]
The second approach to identify low risk variants is genome-wide association studies, which have become feasible with the availability of high density single nucleotide polymorphism (SNP) arrays. Recent genome-wide association studies have successfully identified several loci that are associated with an increased risk of developing colorectal cancer. Risk loci were identified on chromosomes 8q24.21 (rs6983267), 18q21.1 (rs4939827, rs12953717 and rs4464148), chromosome 15q13.3 (rs4779584), 11q23.3 (rs3802842), 10p14 (rs10795668),

8q23.3 (rs16892766), 20p12.3 (rs961253), 14q22.2 (rs4444235), 16q22.1 (rs9929218), and 19q13.1 (rs10411210).[172-181] The next step is to study the mechanisms that underlie CRC susceptibility associated with the different loci. Some of the loci are in high linkage disequilibrium with genes that are strong candidates for the causal relation with colorectal cancer. However, for other loci no genes are in the direct vicinity, which makes the interpretation of these risk factors much more difficult. For rs6983267 (8q24.21), it has now been shown that it is located in a transcriptional enhancer and that it affects the DNA-binding affinity of Wnt-regulated transcription factor TCF7L2 (or TCF4). The enhancer element interacts with the *MYC* promoter; however, no correlation was observed between the rs6983267 genotype and *MYC* expression.[182,183] Additionally, a role in tumor evolution was suggested for rs6983267; tumor studies showed that the risk allele was favored in about 66% of the tumors with allelic imbalance at the locus of rs6983267.[184] For the locus on 18q21.1, the causal SNP has been identified by resequencing the genomic region. This SNP - named Novel 1 - also maps to a transcription factor binding site and has been associated with reduced expression of *SMAD7*. No relation with copy number changes on chromosome 18q21.1 was identified for Novel 1 in tumors.[185]

Environmental Factors

Several environmental factors can influence the risk of the development of colorectal cancer. First, diet is an important factor that influences the CRC risk. Intake of vegetables and fruits is found to reduce the risk of CRC. The effect of raw vegetables, green vegetables, and cruciferous vegetables has been particularly consistent in different studies.[186] The degree of risk reduction however, varies in the different studies. Fiber intake has also been suggested to decrease colorectal cancer risk, however existing data are inconsistent.[186] Consumption of red meat and processed meat has, on the other hand, been associated with an increased risk of colorectal cancer.[186,187] Moreover, alcohol consumption has been associated with an increased colorectal cancer risk.[186,188] Secondly, physical activity and body mass can influence CRC risk. Physical activity is known to reduce the risk of colon cancer, whereas obesity increases the risk of colon cancer. However, both factors do not seem to influence the risk of rectal cancer.[186,188,189] Thirdly, non-steroidal anti-inflammatory drugs (NSAIDs) and hormone replacement therapy (HRT) are described to lower the risk of colorectal cancer.[186] Fourthly, smoking of cigarettes, cigars, and pipes increases the risk of developing colorectal cancer.[186,188] And finally, recent studies in mice suggest that commensal colonic bacteria can promote colorectal tumorigenesis via inflammation.[190]

The influence of environmental factors differs per individual, based on their individual lifestyle. However, environmental factors can in part be shared by individuals from the same family. For example, dietary habits are part of a shared environment or ‘inherited environment’ within a family. In a study that analyzed cancer risk in monozygotic and dizygotic pairs of twins it was

estimated that shared environmental factors contribute up to 5% to the total colorectal cancer risk.[3]

New fields in CRC research, prevention, and treatment

Recent studies that identified colon cancer tumor-initiating cells, provide support for the cancer stem cell hypothesis.[191-193] The stem cell model proposes that tumorigenesis is initiated by dysregulation of the process of self-renewal of colonic stem cells and that as a consequence tumors contain a subcomponent that retains stem cell properties.[194] These cancer stem cells have been suggested to exhibit a greater plasticity that the stem cells populating the normal crypts. The stem cell properties ("stemness") of certain subpopulations of cancer cells can therefore also be seen as an extreme plasticity of tumors to adapt to and survive variable and constantly changing environmental conditions.[195] The concept of cancer stem cells has several consequences for cancer prevention and treatment. Cancer stem cells are more resistant to cytotoxic chemotherapy, because an intrinsic property of stem cells is that they promote survival and are resistant to apoptosis. Additionally, stem cells express multidrug resistant genes, including multifunctional efflux transporters that have an important role in drug distribution.[196] This indicates that novel treatments should be developed that can eradicate the cancer stem cells. Moreover, it has been proposed that stem cells are capable of metastasizing and can remain quiet until the appropriate signals activate them to develop into a macroscopic metastasis.[194]

Current topic in cancer prevention is the possible installation of population-based screening for colorectal cancer. The fecal occult blood test (FOBT) provides a non-invasive and sensitive method (sensitivity: ~65%) for detecting invasive colorectal cancer.[197] The Health Council of the Netherlands has advised to implement biennial immunochemical FOBT in men and women of 55-75 years of age as a nationwide screening program.[198] Another potential non-invasive screening method is a DNA test in stool, in which a panel of genetic markers is tested.[199]

New strategies in the diagnosis and treatment of colorectal cancer have been developed in recent years. For rectum cancer a more standardized way of treatment by the so-called total mesorectal excision (TME) in combination with preoperative radiotherapy has led to a reduction of local recurrence.[200] Furthermore, transanal endoscopic microsurgery (TEM) was introduced. Initially it was solely used as an approach to curatively remove large sessile adenomas of the rectum.[201] However, trials are now underway to also use the TEM technique in combination with neoadjuvant (preoperative) radio-chemotherapy to treat early rectal cancer. Biomarkers and imaging modules are being further developed to predict and monitor therapy response.

Surgeons more and more use endoscopic techniques to remove colorectal cancer with great benefit for the patients in terms of reduction of morbidity and decrease of hospital stay.

Whereas 15 years ago a patient with liver metastases would soon die of this condition, nowadays partial liver resections are being performed, sometimes in combination with isolated liver perfusions and/or radiofrequency ablation (RFA).[202,203] These developments have thereby increased the lifespan of such CRC patients.[204] Furthermore, the use of new chemotherapeutic regimens such as the addition of oxaliplatin and irinotecan has led to improved patient outcomes.[205-207] The introduction of orally given 5-fluorouracil (5-FU) derivatives is of major benefit for patients in terms of tolerability.[208] As a second line of therapy the use of targeted drugs i.e. inhibitors of cancer transduction pathways have shown its effects in certain groups of CRC patients.[209] Biomarkers are identified that predict treatment response. An example for the latter is that the presence of a somatic *KRAS* mutation predicts a lack of response upon treatment with epidermal growth factor receptor (EGFR) inhibitors.[210] There is a peculiar subgroup of CRC patients with peritoneal carcinomatosis patterns that benefit from cytoreductive surgery in combination with hyperthermic intraperitoneal chemotherapy (HIPEC). This therapy had shown to increase the overall survival of such patients.[211] More recently, it has been discussed whether HIPEC should be offered as a prophylactic treatment to prevent the progression of carcinomatosis.[212]

Table 2. Overview of published studies with comparative genomic hybridization analysis of colorectal tumors.

Study	Technique/platform	Samples	Conclusions
Diep et al.[65]	Meta-analysis of 31 CGH studies	373 primary CRCs and 102 liver metastases	Results from the combined analyses suggest that losses at 17p and 18 and gains of 8q, 13q, and 20 occur early in the establishment of primary CRCs, whereas loss of 4p is associated with the transition from Dukes' A to B-D. Deletion of 8p and gains of 7p and 17q are correlated with the transition from primary tumor to liver metastasis, whereas losses of 14q and gains of 1q, 11, 12p, and 19 are late events.
Scheffer et al.[213]	Affymetrix GeneChip Human Mapping 50K SNP arrays	130 colorectal tumors at different stages	Amplifications on chromosomes 7, 8q, 13q, 20, and X. Deletions on chromosomes 4, 8p, 14q, 15q, 17p, 18, 20p, and 22q. Samples with simultaneous deletions in 18q, 8p, 4p, and 15q have a particularly poor prognosis.
Darbary et al.[214]	Affymetrix 10K SNP arrays	13 sporadic colorectal cancers	Uniparental disomy was occurring in many colorectal cancers and often coordinately involved chromosome 14 and 18.
Postma et al.[215]	In house printed 30K oligonucleotide-based array	32 primary tumors of patients with advanced CRC (16 with good and 16 with poor response to chemotherapy)	Responders overall had more chromosomal alterations than nonresponders, especially loss of chromosome 18.
Coss et al.[216]	GenoSensor Array 300 (aCGH)	39 primary colorectal cancers	Amplifications: 20q (79%), 7q (46%), 2p (44%), 7p (41%), 13q (36%), 11p (31%). Deletions: 1p (46%), 17p (36%), 6q (31%), 18q (28%).
Nakao et al.[217]	MacArray Karyo4000 (4k BAC array)	77 sporadic colorectal cancers	Gain: 20q (70%), 7p, 8q, 13. Loss: 18q (68%), 8p, 17p. Aberrations on 3q, 10q, 11q, 15q, and Xp were linked to lymph node metastasis.
Melcher et al.[218]	Spectral karyotyping and SNP array	15 primary MSI tumors and 15 primary CIN tumors	The combination of spectral karyotyping and SNP-array analysis permits the detection of uniparental disomy.
Van Puijenbroek et al.[62]	Illumina 6K SNP arrays (Beadarray)	37 MSI-H colorectal tumors (31 from patients with familial MMR deficiency and 10 sporadic MLH1 hypermethylation)	All carcinomas showed few chromosomal aberrations. MSI-H carcinomas of MMR_UV carriers present more aberrations.
Al-Mulla et al.[219]	Metaphase-based CGH	65 MSS colorectal cancers	In MSS CRCs, the number of chromosomal losses is inversely proportional to Raf kinase inhibitor protein (RKIP) expression levels.
Camps et al.[220]	Agilent oligo microarray and chromosome 8 Human BAC microarray	51 primary adenocarcinomas of the colon	Low-level copy number changes of chromosomes 7, 8, 13, 18, and 20. A significant association of chromosomal breakpoints with structural variants in the human genome was observed.
Lips et al.[221]	Illumina 6K SNP arrays (Beadarray)	36 rectal carcinomas, with both an adenoma and a carcinoma fraction	Five specific chromosomal aberrations, in combination with immunohistochemistry for p53 and SMAD4, can predict possible progression of sessile rectal adenoma to early rectal carcinomas.
Derks et al.[222]	Metaphase-based CGH (n=51) and 5K BAC array (n=20)	71 colorectal carcinomas	Promoter methylation of pivotal tumor suppressor and DNA repair genes is associated with specific patterns of chromosomal changes in CRC, which are different from methylation patterns in MSI tumors.

Table 2 continued. Overview of published studies with comparative genomic hybridization analysis of colorectal tumors.

Study	Technique/platform	Samples	Conclusions
Diep et al.[65]	Meta-analysis of 31 CGH studies	373 primary CRCs and 102 liver metastases	Results from the combined analyses suggest that losses at 17p and 18 and gains of 8q, 13q, and 20 occur early in the establishment of primary CRCs, whereas loss of 4p is associated with the transition from Dukes' A to B-D. Deletion of 8p and gains of 7p and 17q are correlated with the transition from primary tumor to liver metastasis, whereas losses of 14q and gains of 1q, 11, 12p, and 19 are late events.
Scheffer et al.[213]	Affymetrix GeneChip Human Mapping 50K SNP arrays	130 colorectal tumors at different stages	Amplifications on chromosomes 7, 8q, 13q, 20, and X. Deletions on chromosomes 4, 8p, 14q, 15q, 17p, 18, 20p, and 22q. Samples with simultaneous deletions in 18q, 8p, 4p, and 15q have a particularly poor prognosis.
Darbary et al.[214]	Affymetrix 10K SNP arrays	13 sporadic colorectal cancers	Uniparental disomy was occurring in many colorectal cancers and often coordinately involved chromosome 14 and 18.
Postma et al.[215]	In house printed 30K oligonucleotide-based array	32 primary tumors of patients with advanced CRC (16 with good and 16 with poor response to chemotherapy)	Responders overall had more chromosomal alterations than nonresponders, especially loss of chromosome 18.
Coss et al.[216]	GenoSensor Array 300 (aCGH)	39 primary colorectal cancers	Amplifications: 20q (79%), 7q (46%), 2p (44%), 7p (41%), 13q (36%), 11p (31%). Deletions: 1p (46%), 17p (36%), 6q (31%), 18q (28%).
Nakao et al.[217]	MacArray Karyo4000 (4k BAC array)	77 sporadic colorectal cancers	Gain: 20q (70%), 7p, 8q, 13. Loss: 18q (68%), 8p, 17p. Aberrations on 3q, 10q, 11q, 15q, and Xp were linked to lymph node metastasis.
Melcher et al.[218]	Spectral karyotyping and SNP array	15 primary MSI tumors and 15 primary CIN tumors	The combination of spectral karyotyping and SNP-array analysis permits the detection of uniparental disomy.
Van Puijenbroek et al.[62]	Illumina 6K SNP arrays (Beadarray)	37 MSI-H colorectal tumors (31 from patients with familial MMR deficiency and 10 sporadic MLH1 hypermethylation)	All carcinomas showed few chromosomal aberrations. MSI-H carcinomas of MMR_UV carriers present more aberrations.
Al-Mulla et al.[219]	Metaphase-based CGH	65 MSS colorectal cancers	In MSS CRCs, the number of chromosomal losses is inversely proportional to Raf kinase inhibitor protein (RKIP) expression levels.
Camps et al.[220]	Agilent oligo microarray and chromosome 8 Human BAC microarray	51 primary adenocarcinomas of the colon	Low-level copy number changes of chromosomes 7, 8, 13, 18, and 20. A significant association of chromosomal breakpoints with structural variants in the human genome was observed.
Lips et al.[221]	Illumina 6K SNP arrays (Beadarray)	36 rectal carcinomas, with both an adenoma and a carcinoma fraction	Five specific chromosomal aberrations, in combination with immunohistochemistry for p53 and SMAD4, can predict possible progression of sessile rectal adenoma to early rectal carcinomas.
Derks et al.[222]	Metaphase-based CGH (n=51) and 5K BAC array (n=20)	71 colorectal carcinomas	Promoter methylation of pivotal tumor suppressor and DNA repair genes is associated with specific patterns of chromosomal changes in CRC, which are different from methylation patterns in MSI tumors.
Diep et al.[65]	Meta-analysis of 31 CGH studies	373 primary CRCs and 102 liver metastases	Results from the combined analyses suggest that losses at 17p and 18 and gains of 8q, 13q,

			and 20 occur early in the establishment of primary CRCs, whereas loss of 4p is associated with the transition from Dukes' A to B-D. Deletion of 8p and gains of 7p and 17q are correlated with the transition from primary tumor to liver metastasis, whereas losses of 14q and gains of 1q, 11, 12p, and 19 are late events.
Scheffer et al.[213]	Affymetrix GeneChip Human Mapping 50K SNP arrays	130 colorectal tumors at different stages	Amplifications on chromosomes 7, 8q, 13q, 20, and X. Deletions on chromosomes 4, 8p, 14q, 15q, 17p, 18, 20p, and 22q. Samples with simultaneous deletions in 18q, 8p, 4p, and 15q have a particularly poor prognosis.
Darbary et al.[214]	Affymetrix 10K SNP arrays	13 sporadic colorectal cancers	Uniparental disomy was occurring in many colorectal cancers and often coordinately involved chromosome 14 and 18.
Postma et al.[215]	In house printed 30K oligonucleotide-based array	32 primary tumors of patients with advanced CRC (16 with good and 16 with poor response to chemotherapy)	Responders overall had more chromosomal alterations than nonresponders, especially loss of chromosome 18.
Coss et al.[216]	GenoSensor Array 300 (aCGH)	39 primary colorectal cancers	Amplifications: 20q (79%), 7q (46%), 2p (44%), 7p (41%), 13q (36%), 11p (31%). Deletions: 1p (46%), 17p (36%), 6q (31%), 18q (28%).
Nakao et al.[217]	MacArray Karyo4000 (4k BAC array)	77 sporadic colorectal cancers	Gain: 20q (70%), 7p, 8q, 13. Loss: 18q (68%), 8p, 17p. Aberrations on 3q, 10q, 11q, 15q, and Xp were linked to lymph node metastasis.
Melcher et al.[218]	Spectral karyotyping and SNP array	15 primary MSI tumors and 15 primary CIN tumors	The combination of spectral karyotyping and SNP-array analysis permits the detection of uniparental disomy.
Van Puijenbroek et al.[62]	Illumina 6K SNP arrays (Beadarray)	37 MSI-H colorectal tumors (31 from patients with familial MMR deficiency and 10 sporadic MLH1 hypermethylation)	All carcinomas showed few chromosomal aberrations. MSI-H carcinomas of MMR_UV carriers present more aberrations.
Al-Mulla et al.[219]	Metaphase-based CGH	65 MSS colorectal cancers	In MSS CRCs, the number of chromosomal losses is inversely proportional to Raf kinase inhibitor protein (RKIP) expression levels.
Camps et al.[220]	Agilent oligo microarray and chromosome 8 Human BAC microarray	51 primary adenocarcinomas of the colon	Low-level copy number changes of chromosomes 7, 8, 13, 18, and 20. A significant association of chromosomal breakpoints with structural variants in the human genome was observed.
Lips et al.[221]	Illumina 6K SNP arrays (Beadarray)	36 rectal carcinomas, with both an adenoma and a carcinoma fraction	Five specific chromosomal aberrations, in combination with immunohistochemistry for p53 and SMAD4, can predict possible progression of sessile rectal adenoma to early rectal carcinomas.
Derks et al.[222]	Metaphase-based CGH (n=51) and 5K BAC array (n=20)	71 colorectal carcinomas	Promoter methylation of pivotal tumor suppressor and DNA repair genes is associated with specific patterns of chromosomal changes in CRC, which are different from methylation patterns in MSI tumors.

A PubMed search was performed on 22 June 2009. Search terms: "colorectal neoplasms/genetics[mesh] AND (SNP array OR CGH OR comparative genomic hybridization)". Limits: language: English. Only studies that were published after the comprehensive meta-analysis of Diep et al. were listed.[65] Only studies including human colorectal tumors with genome wide genomic data were included. Case reports and studies including less than 10 tumors were excluded. The study of Lips et al.[221] did not appear in the search results, but was included upon literature review.

Reference List

1. Parkin DM, Bray F, Ferlay J et al. Global cancer statistics, 2002. CA Cancer J Clin (2005), 55: 74-108.
2. Kiemeney LA, Lemmers FA, Verhoeven RH et al. [The risk of cancer in the Netherlands]. Ned Tijdschr Geneeskd (2008), 152: 2233-2241.
3. Lichtenstein P, Holm NV, Verkasalo PK et al. 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.
4. Lubbe SJ, Webb EL, Chandler IP et al. Implications of familial colorectal cancer risk profiles and microsatellite instability status. J Clin Oncol (2009), 27: 2238-2244.
5. Butterworth AS, Higgins JP, Pharoah P. Relative and absolute risk of colorectal cancer for individuals with a family history: a meta-analysis. Eur J Cancer (2006), 42: 216-227.
6. Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. Am J Gastroenterol (2001), 96: 2992-3003.
7. Ahsan H, Neugut AI, Garbowski GC et al. Family history of colorectal adenomatous polyps and increased risk for colorectal cancer. Ann Intern Med (1998), 128: 900-905.
8. Winawer SJ, Zauber AG, Gerdes H et al. Risk of colorectal cancer in the families of patients with adenomatous polyps. National Polyp Study Workgroup. N Engl J Med (1996), 334: 82-87.
9. Kumar V, Abbas AK, Fausto N, Aster J: Pathologic Basis of Disease, 8 edn. 2010.
10. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell (1990), 61: 759-767.
11. Vogelstein B, Kinzler KW. The multistep nature of cancer. Trends Genet (1993), 9: 138-141.
12. Houlston RS. What we could do now: molecular pathology of colorectal cancer. Mol Pathol (2001), 54: 206-214.
13. Humphries A, Wright NA. Colonic crypt organization and tumorigenesis. Nat Rev Cancer (2008), 8: 415-424.
14. Powell SM, Zilz N, Beazer-Barclay Y et al. APC mutations occur early during colorectal tumorigenesis. Nature (1992), 359: 235-237.
15. Miyoshi Y, Nagase H, Ando H et al. Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene. Hum Mol Genet (1992), 1: 229-233.
16. Miyaki M, Konishi M, Kikuchi-Yanoshita R et al. Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors. Cancer Res (1994), 54: 3011-3020.
17. Morin PJ, Sparks AB, Korinek V et al. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science (1997), 275: 1787-1790.
18. Vogelstein B, Fearon ER, Hamilton SR et al. Genetic alterations during colorectal-tumor development. N Engl J Med (1988), 319: 525-532.
19. Sjoblom T, Jones S, Wood LD et al. The consensus coding sequences of human breast and colorectal cancers. Science (2006), 314: 268-274.
20. Wood LD, Parsons DW, Jones S et al. The genomic landscapes of human breast and colorectal cancers. Science (2007), 318: 1108-1113.
21. Goss KH, Groden J. Biology of the adenomatous polyposis coli tumor suppressor. J Clin Oncol (2000), 18: 1967-1979.
22. Sparks AB, Morin PJ, Vogelstein B et al. Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. Cancer Res (1998), 58: 1130-1134.
23. Crawford HC, Fingleton BM, Rudolph-Owen LA et al. The metalloproteinase matrilysin is a target of beta-catenin transactivation in intestinal tumors. Oncogene (1999), 18: 2883-2891.
24. He TC, Sparks AB, Rago C et al. Identification of c-MYC as a target of the APC pathway. Science (1998), 281: 1509-1512.
25. Tetsu O, McCormick F. Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. Nature (1999), 398: 422-426.

26. Blobe GC, Schiemann WP, Lodish HF. Role of transforming growth factor beta in human disease. N Engl J Med (2000), 342: 1350-1358.
27. Wrana JL, Attisano L, Wieser R et al. Mechanism of activation of the TGF-beta receptor. Nature (1994), 370: 341-347.
28. Thiagalingam S, Lengauer C, Leach FS et al. Evaluation of candidate tumour suppressor genes on chromosome 18 in colorectal cancers. Nat Genet (1996), 13: 343-346.
29. Grady WM, Myeroff LL, Swinler SE et al. Mutational inactivation of transforming growth factor beta receptor type II in microsatellite stable colon cancers. Cancer Res (1999), 59: 320-324.
30. Markowitz S, Wang J, Myeroff L et al. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science (1995), 268: 1336-1338.
31. Howe JR, Roth S, Ringold JC et al. Mutations in the SMAD4/DPC4 gene in juvenile polyposis. Science (1998), 280: 1086-1088.
32. Hardwick JC, Kodach LL, Offerhaus GJ et al. Bone morphogenetic protein signalling in colorectal cancer. Nat Rev Cancer (2008), 8: 806-812.
33. Howe JR, Bair JL, Sayed MG et al. Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. Nat Genet (2001), 28: 184-187.
34. Baker SJ, Markowitz S, Fearon ER et al. Suppression of human colorectal carcinoma cell growth by wild-type p53. Science (1990), 249: 912-915.
35. Levine AJ. p53, the cellular gatekeeper for growth and division. Cell (1997), 88: 323-331.
36. Baker SJ, Fearon ER, Nigro JM et al. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. Science (1989), 244: 217-221.
37. Petitjean A, Achatz MI, Borresen-Dale AL et al. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. Oncogene (2007), 26: 2157-2165.
38. Baker SJ, Preisinger AC, Jessup JM et al. p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. Cancer Res (1990), 50: 7717-7722.
39. Taupin D, Podolsky DK. Mitogen-activated protein kinase activation regulates intestinal epithelial differentiation. Gastroenterology (1999), 116: 1072-1080.
40. Andreyev HJ, Norman AR, Cunningham D et al. Kirsten ras mutations in patients with colorectal cancer: the multicenter “RASCAL” study. J Natl Cancer Inst (1998), 90: 675-684.
41. Davies H, Bignell GR, Cox C et al. Mutations of the BRAF gene in human cancer. Nature (2002), 417: 949-954.
42. Deng G, Bell I, Crawley S et al. BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer. Clin Cancer Res (2004), 10: 191-195.
43. Rajagopalan H, Bardelli A, Lengauer C et al. Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. Nature (2002), 418: 934.
44. Siena S, Sartore-Bianchi A, Di NF et al. Biomarkers predicting clinical outcome of epidermal growth factor receptor-targeted therapy in metastatic colorectal cancer. J Natl Cancer Inst (2009), 101: 1308-1324.
45. Vivanco I, Sawyers CL. The phosphatidylinositol 3-Kinase AKT pathway in human cancer. Nat Rev Cancer (2002), 2: 489-501.
46. Samuels Y, Ericson K. Oncogenic PI3K and its role in cancer. Curr Opin Oncol (2006), 18: 77-82.
47. Samuels Y, Wang Z, Bardelli A et al. High frequency of mutations of the PIK3CA gene in human cancers. Science (2004), 304: 554.
48. Frattini M, Saletti P, Romagnani E et al. PTEN loss of expression predicts cetuximab efficacy in metastatic colorectal cancer patients. Br J Cancer (2007), 97: 1139-1145.
49. Sartore-Bianchi A, Martini M, Molinari F et al. PIK3CA mutations in colorectal cancer are associated with clinical resistance to EGFR-targeted monoclonal antibodies. Cancer Res (2009), 69: 1851-1857.
50. Kane MF, Loda M, Gaida GM et al. Methylation of the hMLH1 promoter correlates with lack of ex-

pression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. Cancer Res (1997), 57: 808-811.
51. Veigl ML, Kasturi L, Olechnowicz J et al. Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. Proc Natl Acad Sci U S A (1998), 95: 8698-8702.
52. Herman JG, Umar A, Polyak K et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proc Natl Acad Sci U S A (1998), 95: 6870-6875.
53. Kim H, Jen J, Vogelstein B et al. Clinical and pathological characteristics of sporadic colorectal carcinomas with DNA replication errors in microsatellite sequences. Am J Pathol (1994), 145: 148-156.
54. Aaltonen LA, Peltomaki P, Leach FS et al. Clues to the pathogenesis of familial colorectal cancer. Science (1993), 260: 812-816.
55. Bhattacharyya NP, Skandalis A, Ganesh A et al. Mutator phenotypes in human colorectal carcinoma cell lines. Proc Natl Acad Sci U S A (1994), 91: 6319-6323.
56. Eshleman JR, Lang EZ, Bowerfind GK et al. Increased mutation rate at the hprt locus accompanies microsatellite instability in colon cancer. Oncogene (1995), 10: 33-37.
57. Souza RF, Appel R, Yin J et al. Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumours. Nat Genet (1996), 14: 255-257.
58. Rampino N, Yamamoto H, Ionov Y et al. Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. Science (1997), 275: 967-969.
59. Schwartz S Jr, Yamamoto H, Navarro M et al. Frameshift mutations at mononucleotide repeats in caspase-5 and other target genes in endometrial and gastrointestinal cancer of the microsatellite mutator phenotype. Cancer Res (1999), 59: 2995-3002.
60. Kouri M, Laasonen A, Mecklin JP et al. Diploid predominance in hereditary nonpolyposis colorectal carcinoma evaluated by flow cytometry. Cancer (1990), 65: 1825-1829.
61. Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. Nature (1997), 386: 623-627.
62. van Puijenbroek M, Middeldorp A, Tops CM et al. Genome-wide copy neutral LOH is infrequent in familial and sporadic microsatellite unstable carcinomas. Fam Cancer (2008), 7: 319-330.
63. Trautmann K, Terdiman JP, French AJ et al. Chromosomal instability in microsatellite-unstable and stable colon cancer. Clin Cancer Res (2006), 12: 6379-6385.
64. Li LS, Kim NG, Kim SH et al. Chromosomal imbalances in the colorectal carcinomas with microsatellite instability. Am J Pathol (2003), 163: 1429-1436.
65. Diep CB, Kleivi K, Ribeiro FR et al. The order of genetic events associated with colorectal cancer progression inferred from meta-analysis of copy number changes. Genes Chromosomes Cancer (2006), 45: 31-41.
66. Cahill DP, Lengauer C, Yu J et al. Mutations of mitotic checkpoint genes in human cancers. Nature (1998), 392: 300-303.
67. Perez dC, I, de Carcer G, Malumbres M. A census of mitotic cancer genes: new insights into tumor cell biology and cancer therapy. Carcinogenesis (2007), 28: 899-912.
68. Shackney SE, Shankey TV. Common patterns of genetic evolution in human solid tumors. Cytometry (1997), 29: 1-27.
69. Fodde R, Kuipers J, Rosenberg C et al. Mutations in the APC tumour suppressor gene cause chromosomal instability. Nat Cell Biol (2001), 3: 433-438.
70. Toyota M, Ahuja N, Ohe-Toyota M et al. CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci U S A (1999), 96: 8681-8686.
71. Bird AP. CpG-rich islands and the function of DNA methylation. Nature (1986), 321: 209-213.
72. Shen L, Toyota M, Kondo Y et al. Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. Proc Natl Acad Sci U S A (2007), 104: 18654-18659.
73. Aaltonen LA, Salovaara R, Kristo P et al. Incidence of hereditary nonpolyposis colorectal cancer

and the feasibility of molecular screening for the disease. N Engl J Med (1998), 338: 1481-1487.
74. Salovaara R, Loukola A, Kristo P et al. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. J Clin Oncol (2000), 18: 2193-2200.
75. Warthin AS. Heredity with reference to carcinoma. Arch Intern Med (1913), 9: 546-555.
76. Lynch HT, Shaw MW, Magnuson CW et al. Hereditary factors in cancer. Study of two large midwestern kindreds. Arch Intern Med (1966), 117: 206-212.
77. Fishel R, Lescoe MK, Rao MR et al. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell (1993), 75: 1027-1038.
78. Lindblom A, Tannergard P, Werelius B et al. Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. Nat Genet (1993), 5: 279-282.
79. Peltomaki P, Aaltonen LA, Sistonen P et al. Genetic mapping of a locus predisposing to human colorectal cancer. Science (1993), 260: 810-812.
80. Bronner CE, Baker SM, Morrison PT et al. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature (1994), 368: 258-261.
81. Nicolaides NC, Papadopoulos N, Liu B et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature (1994), 371: 75-80.
82. Miyaki M, Konishi M, Tanaka K et al. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. Nat Genet (1997), 17: 271-272.
83. Suter CM, Martin DI, Ward RL. Germline epimutation of MLH1 in individuals with multiple cancers. Nat Genet (2004), 36: 497-501.
84. Ligtenberg MJ, Kuiper RP, Chan TL et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. Nat Genet (2009), 41: 112-117.
85. Stoffel E, Mukherjee B, Raymond VM et al. Calculation of risk of colorectal and endometrial cancer among patients with Lynch syndrome. Gastroenterology (2009), 137: 1621-1627.
86. Mecklin JP, Jarvinen HJ. Clinical features of colorectal carcinoma in cancer family syndrome. Dis Colon Rectum (1986), 29: 160-164.
87. Vasen HF, Offerhaus GJ, Hartog Jager FC et al. The tumour spectrum in hereditary non-polyposis colorectal cancer: a study of 24 kindreds in the Netherlands. Int J Cancer (1990), 46: 31-34.
88. Watson P, Lynch HT. Extracolonic cancer in hereditary nonpolyposis colorectal cancer. Cancer (1993), 71: 677-685.
89. Ionov Y, Peinado MA, Malkhosyan S et al. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature (1993), 363: 558-561.
90. Thibodeau SN, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon. Science (1993), 260: 816-819.
91. Peltomaki P, Lothe RA, Aaltonen LA et al. Microsatellite instability is associated with tumors that characterize the hereditary non-polyposis colorectal carcinoma syndrome. Cancer Res (1993), 53: 5853-5855.
92. Vasen HF, Mecklin JP, Khan PM et al. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum (1991), 34: 424-425.
93. Vasen HF, Watson P, Mecklin JP et al. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology (1999), 116: 1453-1456.
94. de Jong AE, van Puijenbroek M, Hendriks Y et al. Microsatellite instability, immunohistochemistry, and additional PMS2 staining in suspected hereditary nonpolyposis colorectal cancer. Clin Cancer Res (2004), 10: 972-980.
95. Rodriguez-Bigas MA, Boland CR, Hamilton SR et al. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. J Natl Cancer Inst (1997), 89: 1758-1762.
96. Umar A, Boland CR, Terdiman JP et al. Revised Bethesda Guidelines for hereditary nonpolyposis

colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst (2004), 96: 261-268.
97. Kievit W, de Bruin JH, Adang EM et al. Cost effectiveness of a new strategy to identify HNPCC patients. Gut (2005), 54: 97-102.
98. Hampel H, Frankel WL, Martin E et al. Feasibility of screening for Lynch syndrome among patients with colorectal cancer. J Clin Oncol (2008), 26: 5783-5788.
99. Jensen LH, Lindebjerg J, Byriel L et al. Strategy in clinical practice for classification of unselected colorectal tumours based on mismatch repair deficiency. Colorectal Dis (2008), 10: 490-497.
100. Vasen HF, Hartog Jager FC, Menko FH et al. Screening for hereditary non-polyposis colorectal cancer: a study of 22 kindreds in The Netherlands. Am J Med (1989), 86: 278-281.
101. de Jong AE, Hendriks YM, Kleibeuker JH et al. Decrease in mortality in Lynch syndrome families because of surveillance. Gastroenterology (2006), 130: 665-671.
102. Wijnen JT, Brohet RM, van Eijk R et al. Chromosome 8q23.3 and 11q23.1 variants modify colorectal cancer risk in Lynch syndrome. Gastroenterology (2009), 136: 131-137.
103. Ricciardone MD, Ozcelik T, Cevher B et al. Human MLH1 deficiency predisposes to hematological malignancy and neurofibromatosis type 1. Cancer Res (1999), 59: 290-293.
104. Wang Q, Lasset C, Desseigne F et al. Neurofibromatosis and early onset of cancers in hMLH1-deficient children. Cancer Res (1999), 59: 294-297.
105. De Rosa M, Fasano C, Panariello L et al. Evidence for a recessive inheritance of Turcot's syndrome caused by compound heterozygous mutations within the PMS2 gene. Oncogene (2000), 19: 1719-1723.
106. Whiteside D, McLeod R, Graham G et al. A homozygous germ-line mutation in the human MSH2 gene predisposes to hematological malignancy and multiple cafe-au-lait spots. Cancer Res (2002), 62: 359-362.
107. De Vos M, Hayward BE, Picton S et al. Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome. Am J Hum Genet (2004), 74: 954-964.
108. Menko FH, Kaspers GL, Meijer GA et al. A homozygous MSH6 mutation in a child with cafe-au-lait spots, oligodendroglioma and rectal cancer. Fam Cancer (2004), 3: 123-127.
109. Ostergaard JR, Sunde L, Okkels H. Neurofibromatosis von Recklinghausen type I phenotype and early onset of cancers in siblings compound heterozygous for mutations in MSH6. Am J Med Genet A (2005), 139A: 96-105.
110. Ponti G, Ponz dL. Muir-Torre syndrome. Lancet Oncol (2005), 6: 980-987.
111. Suspiro A, Fidalgo P, Cravo M et al. The Muir-Torre syndrome: a rare variant of hereditary nonpolyposis colorectal cancer associated with hMSH2 mutation. Am J Gastroenterol (1998), 93: 1572-1574.
112. Vogt S, Jones N, Christian D et al. Expanded extracolonic tumor spectrum in MUTYH-associated polyposis. Gastroenterology (2009), 137: 1976-1985.
113. Hamilton SR, Liu B, Parsons RE et al. The molecular basis of Turcot's syndrome. N Engl J Med (1995), 332: 839-847.
114. Jarvinen HJ. Epidemiology of familial adenomatous polyposis in Finland: impact of family screening on the colorectal cancer rate and survival. Gut (1992), 33: 357-360.
115. Bulow S. Clinical features in familial polyposis coli. Results of the Danish Polyposis Register. Dis Colon Rectum (1986), 29: 102-107.
116. Aaltonen LA. Hereditary intestinal cancer. Semin Cancer Biol (2000), 10: 289-298.
117. Nagy R, Sweet K, Eng C. Highly penetrant hereditary cancer syndromes. Oncogene (2004), 23: 6445-6470.
118. Bertario L, Russo A, Sala P et al. Genotype and phenotype factors as determinants of desmoid tumors in patients with familial adenomatous polyposis. Int J Cancer (2001), 95: 102-107.
119. Clark SK, Phillips RK. Desmoids in familial adenomatous polyposis. Br J Surg (1996), 83: 1494-1504.

120. Bodmer WF, Bailey CJ, Bodmer J et al. Localization of the gene for familial adenomatous polyposis on chromosome 5. Nature (1987), 328: 614-616.
121. Groden J, Thliveris A, Samowitz W et al. Identification and characterization of the familial adenomatous polyposis coli gene. Cell (1991), 66: 589-600.
122. Kinzler KW, Nilbert MC, Su LK et al. Identification of FAP locus genes from chromosome 5q21. Science (1991), 253: 661-665.
123. Leppert M, Dobbs M, Scambler P et al. The gene for familial polyposis coli maps to the long arm of chromosome 5. Science (1987), 238: 1411-1413.
124. Nishisho I, Nakamura Y, Miyoshi Y et al. Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. Science (1991), 253: 665-669.
125. Lamlum H, Ilyas M, Rowan A et al. The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis. Nat Med (1999), 5: 1071-1075.
126. Albuquerque C, Breukel C, van der Luijt R et al. The 'just-right' signaling model: APC somatic mutations are selected based on a specific level of activation of the beta-catenin signaling cascade. Hum Mol Genet (2002), 11: 1549-1560.
127. Brensinger JD, Laken SJ, Luce MC et al. Variable phenotype of familial adenomatous polyposis in pedigrees with 3' mutation in the APC gene. Gut (1998), 43: 548-552.
128. Hezel AF, Bardeesy N. LKB1; linking cell structure and tumor suppression. Oncogene (2008), 27: 6908-6919.
129. Soravia C, Berk T, Madlensky L et al. Genotype-phenotype correlations in attenuated adenomatous polyposis coli. Am J Hum Genet (1998), 62: 1290-1301.
130. Spirio L, Otterud B, Stauffer D et al. Linkage of a variant or attenuated form of adenomatous polyposis coli to the adenomatous polyposis coli (APC) locus. Am J Hum Genet (1992), 51: 92-100.
131. Spirio L, Olschwang S, Groden J et al. Alleles of the APC gene: an attenuated form of familial polyposis. Cell (1993), 75: 951-957.
132. van der Luijt RB, Meera KP, Vasen HF et al. 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: 727-734.
133. Al Tassan N, Chmiel NH, Maynard J et al. Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. Nat Genet (2002), 30: 227-232.
134. Jones S, Emmerson P, Maynard J et al. Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G:C-->T:A mutations. Hum Mol Genet (2002), 11: 2961-2967.
135. Sampson JR, Dolwani S, Jones S et al. Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH. Lancet (2003), 362: 39-41.
136. Nielsen M, Franken PF, Reinards TH et al. Multiplicity in polyp count and extracolonic manifestations in 40 Dutch patients with MYH associated polyposis coli (MAP). J Med Genet (2005), 42: e54.
137. Sieber OM, Lipton L, Crabtree M et al. Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. N Engl J Med (2003), 348: 791-799.
138. Lipton L, Halford SE, Johnson V et al. Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway. Cancer Res (2003), 63: 7595-7599.
139. Farrington SM, Tenesa A, Barnetson R et al. Germline susceptibility to colorectal cancer due to base-excision repair gene defects. Am J Hum Genet (2005), 77: 112-119.
140. Cleary SP, Cotterchio M, Jenkins MA et al. Germline MutY human homologue mutations and colorectal cancer: a multisite case-control study. Gastroenterology (2009), 136: 1251-1260.
141. Jones N, Vogt S, Nielsen M et al. Increased colorectal cancer incidence in obligate carriers of heterozygous mutations in MUTYH. Gastroenterology (2009), 137: 489-94, 494.
142. Lubbe SJ, Di Bernardo MC, Chandler IP et al. Clinical implications of the colorectal cancer risk associated with MUTYH mutation. J Clin Oncol (2009), 27: 3975-3980.
143. Webb EL, Rudd MF, Houlston RS. Colorectal cancer risk in monoallelic carriers of MYH variants.

Am J Hum Genet (2006), 79: 768-771.
144. Steinke V, Rahner N, Morak M et al. No association between MUTYH and MSH6 germline mutations in 64 HNPCC patients. Eur J Hum Genet (2008), 16: 587-592.
145. Niessen RC, Sijmons RH, Ou J et al. MUTYH and the mismatch repair system: partners in crime? Hum Genet (2006), 119: 206-211.
146. van Puijenbroek M, Nielsen M, Reinards TH et al. The natural history of a combined defect in MSH6 and MUTYH in a HNPCC family. Fam Cancer (2007), 6: 43-51.
147. Jarvinen H, Franssila KO. Familial juvenile polyposis coli; increased risk of colorectal cancer. Gut (1984), 25: 792-800.
148. Sayed MG, Ahmed AF, Ringold JR et al. Germline SMAD4 or BMPR1A mutations and phenotype of juvenile polyposis. Ann Surg Oncol (2002), 9: 901-906.
149. Hemminki A, Markie D, Tomlinson I et al. A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. Nature (1998), 391: 184-187.
150. Baas AF, Boudeau J, Sapkota GP et al. Activation of the tumour suppressor kinase LKB1 by the STE20-like pseudokinase STRAD. EMBO J (2003), 22: 3062-3072.
151. Boudeau J, Baas AF, Deak M et al. MO25alpha/beta interact with STRADalpha/beta enhancing their ability to bind, activate and localize LKB1 in the cytoplasm. EMBO J (2003), 22: 5102-5114.
152. Baas AF, Kuipers J, van der Wel NN et al. Complete polarization of single intestinal epithelial cells upon activation of LKB1 by STRAD. Cell (2004), 116: 457-466.
153. Giardiello FM, Welsh SB, Hamilton SR et al. Increased risk of cancer in the Peutz-Jeghers syndrome. N Engl J Med (1987), 316: 1511-1514.
154. Baas AF, Smit L, Clevers H. LKB1 tumor suppressor protein: PARtaker in cell polarity. Trends Cell Biol (2004), 14: 312-319.
155. Lindor NM, Rabe K, Petersen GM et al. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. JAMA (2005), 293: 1979-1985.
156. Aaltonen L, Johns L, Jarvinen H et al. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. Clin Cancer Res (2007), 13: 356-361.
157. Jenkins MA, Baglietto L, Dite GS et al. After hMSH2 and hMLH1--what next? Analysis of three-generational, population-based, early-onset colorectal cancer families. Int J Cancer (2002), 102: 166-171.
158. Wiesner GL, Daley D, Lewis S et al. A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. Proc Natl Acad Sci U S A (2003), 100: 12961-12965.
159. Kemp ZE, Carvajal-Carmona LG, Barclay E et al. Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United kingdom. Cancer Res (2006), 66: 5003-5006.
160. Skoglund J, Djureinovic T, Zhou XL et al. Linkage analysis in a large Swedish family supports the presence of a susceptibility locus for adenoma and colorectal cancer on chromosome 9q22.32-31.1. J Med Genet (2006), 43: e7.
161. Kemp Z, Carvajal-Carmona L, Spain S et al. Evidence for a colorectal cancer susceptibility locus on chromosome 3q21-q24 from a high-density SNP genome-wide linkage scan. Hum Mol Genet (2006), 15: 2903-2910.
162. Papaemmanuil E, Carvajal-Carmona L, Sellick GS et al. Deciphering the genetics of hereditary non-syndromic colorectal cancer. Eur J Hum Genet (2008), 16: 1477-1486.
163. Picelli S, Vandrovcova J, Jones S et al. Genome-wide linkage scan for colorectal cancer susceptibility genes supports linkage to chromosome 3q. BMC Cancer (2008), 8: 87.
164. Djureinovic T, Skoglund J, Vandrovcova J et al. A genome wide linkage analysis in Swedish families with hereditary non-familial adenomatous polyposis/non-hereditary non-polyposis colorectal cancer. Gut (2006), 55: 362-366.
165. Neklason DW, Kerber RA, Nilson DB et al. Common familial colorectal cancer linked to chromosome 7q31: a genome-wide analysis. Cancer Res (2008), 68: 8993-8997.

166. Thean LF, Loi C, Ho KS et al. Genome-wide scan identifies a copy number variable region at 3q26 that regulates PPM1L in APC mutation-negative familial colorectal cancer patients. Genes Chromosomes Cancer (2010), 49: 99-106.
167. de Jong MM, Nolte IM, te Meerman GJ et al. Low-penetrance genes and their involvement in colorectal cancer susceptibility. Cancer Epidemiol Biomarkers Prev (2002), 11: 1332-1352.
168. Houlston RS, Tomlinson IP. Polymorphisms and colorectal tumor risk. Gastroenterology (2001), 121: 282-301.
169. Wasielewski M, Vasen H, Wijnen J et al. CHEK2 1100delC is a susceptibility allele for HNPCC-related colorectal cancer. Clin Cancer Res (2008), 14: 4989-4994.
170. Webb EL, Rudd MF, Sellick GS et al. Search for low penetrance alleles for colorectal cancer through a scan of 1,467 non-synonymous SNPs in 2,575 cases and 2,707 controls with validation by kin-cohort analysis of 14,704 first-degree relatives. Hum Mol Genet (2006).
171. Kury S, Buecher B, Robiou-du-Pont S et al. Low-penetrance alleles predisposing to sporadic colorectal cancers: a French case-controlled genetic association study. BMC Cancer (2008), 8: 326.
172. Broderick P, Carvajal-Carmona L, Pittman AM et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet (2007), 39: 1315-1317.
173. Gruber SB, Moreno V, Rozek LS et al. Genetic Variation in 8q24 Associated with Risk of Colorectal Cancer. Cancer Biol Ther (2007), 6: 1143-1147.
174. Houlston RS, Webb E, Broderick P et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet (2008), 40: 1426-1435.
175. Jaeger E, Webb E, Howarth K et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet (2008), 40: 26-28.
176. Pittman AM, Webb E, Carvajal-Carmona L et al. Refinement of the basis and impact of common 11q23.1 variation to the risk of developing colorectal cancer. Hum Mol Genet (2008), 17: 3720-3727.
177. Poynter JN, Figueiredo JC, Conti DV et al. Variants on 9p24 and 8q24 are associated with risk of colorectal cancer: results from the Colon Cancer Family Registry. Cancer Res (2007), 67: 11128-11132.
178. Tenesa A, Farrington SM, Prendergast JG et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet (2008), 40: 631-637.
179. Tomlinson I, Webb E, Carvajal-Carmona L et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet (2007), 39: 984-988.
180. Tomlinson IP, Webb E, Carvajal-Carmona L et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet (2008), 40: 623-630.
181. Zanke BW, Greenwood CM, Rangrej J et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet (2007), 39: 989-994.
182. Pomerantz MM, Ahmadiyeh N, Jia L et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat Genet (2009), 41: 882-884.
183. Tuupanen S, Turunen M, Lehtonen R et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nat Genet (2009), 41: 885-890.
184. Tuupanen S, Niittymaki I, Nousiainen K et al. Allelic imbalance at rs6983267 suggests selection of the risk allele in somatic colorectal tumor evolution. Cancer Res (2008), 68: 14-17.
185. Pittman AM, Naranjo S, Webb E et al. The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. Genome Res (2009), 19: 987-993.
186. Potter JD. Colorectal cancer: molecules and populations. J Natl Cancer Inst (1999), 91: 916-932.
187. Larsson SC, Wolk A. Meat consumption and risk of colorectal cancer: a meta-analysis of prospec-

tive studies. Int J Cancer (2006), 119: 2657-2664.
188. Huxley RR, Ansary-Moghaddam A, Clifton P et al. The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. Int J Cancer (2009), 125: 171-180.
189. Moghaddam AA, Woodward M, Huxley R. Obesity and risk of colorectal cancer: a meta-analysis of 31 studies with 70,000 events. Cancer Epidemiol Biomarkers Prev (2007), 16: 2533-2547.
190. Wu S, Rhee KJ, Albesiano E et al. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. Nat Med (2009), 15: 1016-1022.
191. Dalerba P, Dylla SJ, Park IK et al. Phenotypic characterization of human colorectal cancer stem cells. Proc Natl Acad Sci U S A (2007), 104: 10158-10163.
192. O'Brien CA, Pollett A, Gallinger S et al. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature (2007), 445: 106-110.
193. Ricci-Vitiani L, Lombardi DG, Pilozzi E et al. Identification and expansion of human colon-cancer-initiating cells. Nature (2007), 445: 111-115.
194. Wicha MS, Liu S, Dontu G. Cancer stem cells: an old idea--a paradigm shift. Cancer Res (2006), 66: 1883-1890.
195. Gupta PB, Chaffer CL, Weinberg RA. Cancer stem cells: mirage or reality? Nat Med (2009), 15: 1010-1012.
196. Dean M. ABC transporters, drug resistance, and cancer stem cells. J Mammary Gland Biol Neoplasia (2009), 14: 3-9.
197. Morikawa T, Kato J, Yamaji Y et al. A comparison of the immunochemical fecal occult blood test and total colonoscopy in the asymptomatic population. Gastroenterology (2005), 129: 422-428.
198. Health Council of the Netherlands. A national colorectal cancer screening programme. publication no. 2009/13E. 2009. The Hague.
199. Davies RJ, Miller R, Coleman N. Colorectal cancer screening: prospects for molecular stool analysis. Nat Rev Cancer (2005), 5: 199-209.
200. Kapiteijn E, Marijnen CA, Nagtegaal ID et al. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer. N Engl J Med (2001), 345: 638-646.
201. Buess G, Mentges B, Manncke K et al. Technique and results of transanal endoscopic microsurgery in early rectal cancer. Am J Surg (1992), 163: 63-69.
202. Rothbarth J, Tollenaar RA, Schellens JH et al. Isolated hepatic perfusion for the treatment of colorectal metastases confined to the liver: recent trends and perspectives. Eur J Cancer (2004), 40: 1812-1824.
203. McGahan JP, Brock JM, Tesluk H et al. Hepatic ablation with use of radio-frequency electrocautery in the animal model. J Vasc Interv Radiol (1992), 3: 291-297.
204. Stang A, Fischbach R, Teichmann W et al. A systematic review on the clinical benefit and role of radiofrequency ablation as treatment of colorectal liver metastases. Eur J Cancer (2009), 45: 1748-1756.
205. de GA, Figer A, Seymour M et al. Leucovorin and fluorouracil with or without oxaliplatin as first-line treatment in advanced colorectal cancer. J Clin Oncol (2000), 18: 2938-2947.
206. Cunningham D, Pyrhonen S, James RD et al. Randomised trial of irinotecan plus supportive care versus supportive care alone after fluorouracil failure for patients with metastatic colorectal cancer. Lancet (1998), 352: 1413-1418.
207. Rougier P, Van CE, Bajetta E et al. Randomised trial of irinotecan versus fluorouracil by continuous infusion after fluorouracil failure in patients with metastatic colorectal cancer. Lancet (1998), 352: 1407-1412.
208. Hoff PM, Ansari R, Batist G et al. Comparison of oral capecitabine versus intravenous fluorouracil plus leucovorin as first-line treatment in 605 patients with metastatic colorectal cancer: results of a randomized phase III study. J Clin Oncol (2001), 19: 2282-2292.
209. Ortega J, Vigil CE, Chodkiewicz C. Current progress in targeted therapy for colorectal cancer.

Cancer Control (2010), 17: 7-15.
210. Karapetis CS, Khambata-Ford S, Jonker DJ et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. N Engl J Med (2008), 359: 1757-1765.
211. Verwaal VJ, Bruin S, Boot H et al. 8-year follow-up of randomized trial: cytoreduction and hyperthermic intraperitoneal chemotherapy versus systemic chemotherapy in patients with peritoneal carcinomatosis of colorectal cancer. Ann Surg Oncol (2008), 15: 2426-2432.
212. Verwaal VJ. Long-term results of cytoreduction and HIPEC followed by systemic chemotherapy. Cancer J (2009), 15: 212-215.
213. Sheffer M, Bacolod MD, Zuk O et al. Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. Proc Natl Acad Sci U S A (2009), 106: 7131-7136.
214. Darbary HK, Dutt SS, Sait SJ et al. Uniparentalism in sporadic colorectal cancer is independent of imprint status, and coordinate for chromosomes 14 and 18. Cancer Genet Cytogenet (2009), 189: 77-86.
215. Postma C, Koopman M, Buffart TE et al. DNA copy number profiles of primary tumors as predictors of response to chemotherapy in advanced colorectal cancer. Ann Oncol (2009), 20: 1048-1056.
216. Coss A, Tosetto M, Fox EJ et al. Increased topoisomerase IIalpha expression in colorectal cancer is associated with advanced disease and chemotherapeutic resistance via inhibition of apoptosis. Cancer Lett (2009), 276: 228-238.
217. Nakao M, Kawauchi S, Furuya T et al. Identification of DNA copy number aberrations associated with metastases of colorectal cancer using array CGH profiles. Cancer Genet Cytogenet (2009), 188: 70-76.
218. Melcher R, Zopf W, Hartmann E et al. Spectral karyotyping and SNP microarray analysis define uniparental disomy (UPD) as a novel mutational mechanism in MSI- and CSI-colorectal cancers. Cell Oncol (2008), 30: 507.
219. Al Mulla F, Hagan S, Al Ali W et al. Raf kinase inhibitor protein: mechanism of loss of expression and association with genomic instability. J Clin Pathol (2008), 61: 524-529.
220. Camps J, Grade M, Nguyen QT et al. Chromosomal breakpoints in primary colon cancer cluster at sites of structural variants in the genome. Cancer Res (2008), 68: 1284-1295.
221. Lips EH, van Eijk R, de Graaf EJ et al. Progression and tumor heterogeneity analysis in early rectal cancer. Clin Cancer Res (2008), 14: 772-781.
222. Derks S, Postma C, Carvalho B et al. Integrated analysis of chromosomal, microsatellite and epigenetic instability in colorectal cancer identifies specific associations between promoter methylation of pivotal tumour suppressor and DNA repair genes and specific chromosomal alterations. Carcinogenesis (2008), 29: 434-439.
223. Blaker H, Mechtersheimer G, Sutter C et al. Recurrent deletions at 6q in early age of onset non-HNPCC- and Non-FAP-Associated Intestinal Carcinomas. Evidence for a Novel Cancer Susceptibility Locus at 6q14-q22. Genes Chromosomes Cancer (2008), 47: 159-164.
224. Martin ES, Tonon G, Sinha R et al. Common and distinct genomic events in sporadic colorectal cancer and diverse cancer types. Cancer Res (2007), 67: 10736-10743.
225. Fijneman RJ, Carvalho B, Postma C et al. Loss of 1p36, gain of 8q24, and loss of 9q34 are associated with stroma percentage of colorectal cancer. Cancer Lett (2007), 258: 223-229.
226. Jones AM, Thirlwell C, Howarth KM et al. Analysis of copy number changes suggests chromosomal instability in a minority of large colorectal adenomas. J Pathol (2007), 213: 249-256.
227. Xiao XY, Zhou XY, Yan G et al. Chromosomal alteration in Chinese sporadic colorectal carcinomas detected by comparative genomic hybridization. Diagn Mol Pathol (2007), 16: 96-103.
228. Lips EH, de Graaf EJ, Tollenaar RA et al. Single nucleotide polymorphism array analysis of chromosomal instability patterns discriminates rectal adenomas from carcinomas. J Pathol (2007), 212: 269-277.
229. Fensterer H, Radlwimmer B, Strater J et al. Matrix-comparative genomic hybridization from multi-

center formalin-fixed paraffin-embedded colorectal cancer tissue blocks. BMC Cancer (2007), 7: 58.
230. De Angelis PM, Stokke T, Beigi M et al. Chromosomal 20q gain in the DNA diploid component of aneuploid colorectal carcinomas. Int J Cancer (2007), 120: 2734-2738.
231. Grade M, Hormann P, Becker S et al. Gene expression profiling reveals a massive, aneuploidy-dependent transcriptional deregulation and distinct differences between lymph node-negative and lymph node-positive colon carcinomas. Cancer Res (2007), 67: 41-56.
232. Bartos JD, Gaile DP, McQuaid DE et al. aCGH local copy number aberrations associated with overall copy number genomic instability in colorectal cancer: coordinate involvement of the regions including BCR and ABL. Mutat Res (2007), 615: 1-11.
233. Ghadimi BM, Grade M, Monkemeyer C et al. Distinct chromosomal profiles in metastasizing and non-metastasizing colorectal carcinomas. Cell Oncol (2006), 28: 273-281.
234. Derks S, Postma C, Moerkerk PT et al. Promoter methylation precedes chromosomal alterations in colorectal cancer development. Cell Oncol (2006), 28: 247-257.
235. Liu XP, Kawauchi S, Oga A et al. Chromosomal aberrations detected by comparative genomic hybridization predict outcome in patients with colorectal carcinoma. Oncol Rep (2007), 17: 261-267.
236. Kim MY, Yim SH, Kwon MS et al. Recurrent genomic alterations with impact on survival in colorectal cancer identified by genome-wide array comparative genomic hybridization. Gastroenterology (2006), 131: 1913-1924.
237. Habermann JK, Paulsen U, Roblick UJ et al. Stage-specific alterations of the genome, transcriptome, and proteome during colorectal carcinogenesis. Genes Chromosomes Cancer (2007), 46: 10-26.
238. Unotoro J, Kamiyama H, Ishido Y et al. Analysis of the relationship between sex and chromosomal aberrations in colorectal cancer by comparative genomic hybridization. J Int Med Res (2006), 34: 397-405.
239. Yaginuma Y, Unotoro J, Kamiyama H et al. Genomic copy-number aberrations related to lymph-node metastasis of colon cancer. J Int Med Res (2006), 34: 390-396.
240. Rodriguez J, Frigola J, Vendrell E et al. Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers. Cancer Res (2006), 66: 8462-9468.
241. Swede H, Bartos JD, Chen N et al. Genomic profiles of colorectal cancers differ based on patient smoking status. Cancer Genet Cytogenet (2006), 168: 98-104.
242. Alrawi SJ, Carroll RE, Hill HC et al. Genomic instability of human aberrant crypt foci measured by inter-(simple sequence repeat) PCR and array-CGH. Mutat Res (2006), 601: 30-38.
243. Andersen CL, Wiuf C, Kruhoffer M et al. Frequent occurrence of uniparental disomy in colorectal cancer. Carcinogenesis (2007), 28: 38-48.
244. Al Mulla F, Behbehani AI, Bitar MS et al. Genetic profiling of stage I and II colorectal cancer may predict metastatic relapse. Mod Pathol (2006), 19: 648-658.
245. Cardoso J, Molenaar L, de Menezes RX et al. Chromosomal instability in MYH- and APC-mutant adenomatous polyps. Cancer Res (2006), 66: 2514-2519.
246. Tsafrir D, Bacolod M, Selvanayagam Z et al. Relationship of gene expression and chromosomal abnormalities in colorectal cancer. Cancer Res (2006), 66: 2129-2137.
247. Mehigan BJ, Ashman JN, Baker RP et al. Mismatch repair, p53 and chromosomal aberrations in primary colorectal carcinomas. Acta Oncol (2006), 45: 61-66.
248. Grade M, Ghadimi BM, Varma S et al. Aneuploidy-dependent massive deregulation of the cellular transcriptome and apparent divergence of the Wnt/beta-catenin signaling pathway in human rectal carcinomas. Cancer Res (2006), 66: 267-282.

Chapter 2

A procedure for the detection of linkage with high density SNP arrays in a large pedigree with colorectal cancer

BMC Cancer (2007) 7:6

A procedure for the detection of linkage with high density SNP arrays in a large pedigree with colorectal cancer

Anneke Middeldorp[1], Shantie Jagmohan-Changur[2], Quinta Helmer[3], Heleen M van der Klift[4], Carli MJ Tops[4], Hans FA Vasen[5], Peter Devilee[2], Hans Morreau[1], Jeanine J Houwing-Duistermaat[3], Juul T Wijnen[2,4] and Tom van Wezel*[1]

Address: [1]Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands, [2]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands, [3]Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands, [4]Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands and [5]The Netherlands Foundation for the Detection of Hereditary Tumours, Leiden, The Netherlands

Email: Anneke Middeldorp - j.w.middeldorp@lumc.nl; Shantie Jagmohan-Changur - s.c.jagmohan@lumc.nl; Quinta Helmer - q.helmer@lumc.nl; Heleen M van der Klift - h.m.van_der_klift@lumc.nl; Carli MJ Tops - c.m.j.tops@lumc.nl; Hans FA Vasen - hfavasen@stoet.nl; Peter Devilee - p.devilee@lumc.nl; Hans Morreau - j.morreau@lumc.nl; Jeanine J Houwing-Duistermaat - j.j.houwing@lumc.nl; Juul T Wijnen - j.wijnen@lumc.nl; Tom van Wezel* - t.van_wezel@lumc.nl

* Corresponding author

Published: 12 January 2007

BMC Cancer 2007, **7**:6 doi:10.1186/1471-2407-7-6

Received: 07 August 2006
Accepted: 12 January 2007

This article is available from: http://www.biomedcentral.com/1471-2407/7/6

Abstract

Background: The apparent dominant model of colorectal cancer (CRC) inheritance in several large families, without mutations in known CRC susceptibility genes, suggests the presence of so far unidentified genes with strong or moderate effect on the development of CRC. Linkage analysis could lead to identification of susceptibility genes in such families. In comparison to classical linkage analysis with multi-allelic markers, single nucleotide polymorphism (SNP) arrays have increased information content and can be processed with higher throughput. Therefore, SNP arrays can be excellent tools for linkage analysis. However, the vast number of SNPs on the SNP arrays, combined with large informative pedigrees (e.g. >35–40 bits), presents us with a computational complexity that is challenging for existing statistical packages or even exceeds their capacity. We therefore setup a procedure for linkage analysis in large pedigrees and validated the method by genotyping using SNP arrays of a colorectal cancer family with a known *MLH1* germ line mutation.

Methods: Quality control of the genotype data was performed in Alohomora, Mega2 and SimWalk2, with removal of uninformative SNPs, Mendelian inconsistencies and Mendelian consistent errors, respectively. Linkage disequilibrium was measured by SNPLINK and Merlin. Parametric linkage analysis using two flanking markers was performed using MENDEL. For multipoint parametric linkage analysis and haplotype analysis, SimWalk2 was used.

Results: On chromosome 3, in the *MLH1*-region, a LOD score of 1.9 was found by parametric linkage analysis using two flanking markers. On chromosome 11 a small region with LOD 1.1 was also detected. Upon linkage disequilibrium removal, multipoint linkage analysis yielded a LOD score of 2.1 in the *MLH1* region, whereas the LOD score dropped to negative values in the region on chromosome 11. Subsequent haplotype analysis in the *MLH1* region perfectly matched the mutation status of the family members.

Conclusion: We developed a workflow for linkage analysis in large families using high-density SNP arrays and validated this workflow in a family with colorectal cancer. Linkage disequilibrium has to be removed when using SNP arrays, because it can falsely inflate the LOD score. Haplotype analysis is adequate and can predict the carrier status of the family members.

Background

Colorectal cancer (CRC) is the one of the most common malignancies in the Western world. Already in 1913, familial aggregation of CRC was described by Warthin [1] and later Lynch *et al*. described an additional family with clustering of colorectal and endometrial cancer [2]. Clinical definition of Lynch syndrome, or HNPCC, in 1991 [3,4] was instrumental for linkage analysis, and ultimately for the identification of the underlying gene defects in HNPCC families. The first HNPCC loci were mapped to chromosomes 2 and 3 using microsatellite markers [5,6]. This eventually led to the identification of germ line mutations in *MSH2* [7] and *MLH1* [8], respectively. Later, *PMS2* [9], *MSH6* [10,11] and recently *MutYH* [12] were identified as CRC susceptibility genes. However, the so far identified CRC susceptibility genes can only explain up to 5% of all cases [13], while in ∼35% of all colorectal cancer cases familial clustering is seen [14]. Furthermore, it is shown that first degree relatives of patients with colorectal cancer have a relative risk of 2.3 to develop the disease [15]. This indicates that still some genes with strong or moderate effect on CRC development remain to be identified. In order to identify these genes, linkage analysis in families could point to the loci where unknown susceptibility genes may reside. Indeed, different linkage analysis studies revealed potentially interesting regions on chromosomes 3q, 9q, 11q, 14q, 15q and 22q [16-20].

Families with a clustering of colorectal cancer but without germ line mutations in CRC genes have been under surveillance in Leiden since the 1980s. Due to the long period of follow-up, with three to four affected generations, these Dutch HNPCC-like families have become informative for linkage analysis.

Traditionally, linkage analysis is performed with multiallelic microsatellite markers. Recently, however, the more advanced single nucleotide polymorphism (SNP) arrays were brought into use for linkage analysis. It was shown that the information content of a dense SNP map is significantly and uniformly higher than that of a genome wide microsatellite marker map [21]. Several studies conducting linkage analysis on genotype data from SNP arrays appeared in recent years [22-24]. In these studies non-parametric as well as parametric linkage analysis was performed in sib pairs or in small to moderate size pedigrees. However, to date, no studies have been published on linkage analysis using SNPs in large pedigrees (e.g. >35–40 bits).

Studying large families with thousands of SNPs results in a computational complex analysis that is challenging for existing statistical packages and that may even exceed their capacity. Current linkage analysis programs can handle either large pedigrees or large numbers of markers, depending on the underlying algorithm. In order to perform linkage analysis in large pedigrees using SNP arrays, we explored the possibilities of currently available linkage analysis software. Most currently available programs are based on the Lander-Green or the Elston-Stewart algorithm or both. The computation time of the former algorithm increases exponentially with the number of bits (2n - f, where 'n' is the number of non-founders and 'f' the number of founders) in a pedigree, whereas the latter scales exponentially with the number of markers. To perform multipoint linkage analysis in a large family with SNP arrays in one run would probably take several months computation time, if at all possible.

Several programs are suitable for linkage analysis with bi-allelic markers. Genehunter and Merlin can handle a relative large numbers of markers, however the analysis is restricted to pedigrees of up to ∼30-bits [25,26]. Both programs are based on the Lander-Green Hidden Markov Model algorithm and can perform non-parametric as well as parametric linkage analysis. In Genehunter, the Elston-Stewart algorithm is also implemented, allowing the performance of simultaneous analysis of several markers as well as analysis of pedigrees of moderate size. A third program based on the Lander-Green Hidden Markov Model algorithm is Allegro 2. This program can handle large pedigrees (up to ∼40 bits), although the computational costs increase substantially when not all genotype information of the family is available [27,28]. Allegro calculates parametric LOD scores as well as NPL scores and allele-sharing LOD scores. Another program, SNPLINK [28,29] can perform automated linkage analysis with LD removal using either Allegro or Merlin. However, for all the above mentioned programs the different branches of large families (i.e. >35–40 bits) need to be analyzed separately. This will lead to substantial loss of information and potential undetected linkage.

MENDEL [30] is a program that is suitable for linkage analysis with SNPs in large pedigrees. It allows adjusting the maximum number of meioses, though the computation time will increase in that case. Both parametric and non-parametric linkage analysis can be performed in MENDEL. The program will either use the Lander-Green or the Elston-Stewart algorithm, depending on whichever is more efficient for the pedigree. SimWalk2 is a program that can perform multipoint parametric linkage analysis, haplotype analysis and a few other analyses in large pedigrees using bi-allelic markers. It uses Markov chain Monte Carlo methods to compute the likelihood [31]. Simwalk2 uses the MENDEL program for computing location scores. With the aim to detect linkage in CRC families exceeding 40 bits we established a procedure using freely available software packages and validated this in a large colorectal

cancer family, with a known causal *MLH1* germ line mutation on chromosome 3.

Methods

Patients

A large colorectal cancer family (Figure 1) with a recently identified mutation in the *MLH1* gene (c.1046dupT, p.Pro350fs) was studied. Nine family members are affected with colorectal cancer. Another two family members are affected with polyps and three cases with skin cancer (non-specified) and one case with endometrium cancer (non-specified) are seen as well. Peripheral blood lymphocytes were collected from the family members. DNA was extracted using standard procedures. A total of thirteen family members were genotyped on Affymetrix GeneChip Human Mapping 10K 2.0 SNP arrays. The arrays were processed according to the instructions of the manufacturer. The mean SNP call rate was 96.3% (89.0%-98.5%).

The study was approved by the Medical Ethical Committee of the LUMC (protocol P01-019).

Workflow

We processed the data according to the following workflow: 1) First, the genotype data were generated by GeneChip DNA Analysis Software (GDAS) from Affymetrix. 2) These genotype data were combined with the pedigree and the marker information in Alohomora. 3) In this program the uninformative SNPs were removed as well. 4) To be able to perform linkage analysis in the desired program, the output files (in Merlin-format) of Alohomora were by Mega2 converted to the proper format. 5) Mega2 also removed the Mendelian inconsistent errors. 6) The files were then ready to perform parametric linkage analysis using 2 flanking markers in MENDEL; affected-only analysis as well as parametric linkage analysis using liability classes was performed. 7) Based on the second analysis, regions of interest were defined that were further tested for Mendelian consistent errors and 8) possible linkage disequilibrium was removed in SNPLINK. 9) Multipoint parametric linkage analysis using the liability classes was then performed in Simwalk2 for the ROIs and 10) finally, the haplotypes were inferred in Simwalk2.

Data formatting and quality control

Genotype data of the individual family members were generated using GeneChip DNA Analysis Software (GDAS) from Affymetrix. In the Alohomora program [32] the pedigree information, allele frequencies and map position of the SNPs were combined with the genotype data generated by GDAS. The uninformative SNPs in this pedigree, that show either only A alleles and No Calls or only B alleles and No Calls, were removed from further analysis by Alohomora. The data files were exported in Merlin format. Subsequently, in Mega2 [33] these Alohomora files were converted into the appropriate format for the programs used for linkage analysis, i.e. either the Mendel 5 format or the SimWalk2 format. Mendelian inconsistent errors were removed from analysis with Mega2 by setting all genotypes of these SNPs to unknown.

Mendelian consistent errors

Mendelian consistent errors were identified by mistyping analysis. Since this analysis is computationally complex and therefore time consuming (2 1/4 hours for 35 SNPs), only the regions of interest were analyzed for Mendelian consistent errors. All chromosomal regions with LOD scores exceeding 1 and lacking negative LOD scores were defined as regions of interest (ROI). SimWalk2 [31] was used to check all ROI for Mendelian consistent errors by performing mistyping analysis. An error model with a uniform error rate for all mistypings was used. The overall rate of mistyping was set at 0.004 [34,35]. The threshold for the posterior probability of mistyping was set at 0.5 [36].

Linkage disequilibrium estimation

In the ROI the pair-wise correlation coefficient r^2, as a measure of linkage disequilibrium (LD) between adjacent SNPs, was estimated using SNPLINK [29] and Merlin [26]. Since we are only interested in estimates of r^2, we split the large family into nuclear families. In addition to the family under study, genotypes from 12 Dutch nuclear families from other studies (unpublished results) were used to calculate LD. The program SNPLINK provides a list of SNPs to be removed. We used as cut off value for LD removal an $r^2 \geq 0.4$. The information content was computed before and after removal of the SNPs using Merlin.

Linkage analysis

To determine the power to detect linkage in the *MLH1* family, we performed a simulation study using Simlink [37] under the assumption of a dominant trait with a piecewise linear penetrance. Subsequently, we performed an affected-only linkage analysis and modeled a dominant trait with an allele frequency of 0.001. For parametric linkage analysis, the proper assignment of affected status to family members is crucial since, due to the surveillance of the families, adenomas will be detected and removed before they can develop into a carcinoma. Additionally, the risk of cancer increases with age. And the risk of developing an adenoma is different from the risk of developing a carcinoma. To adjust for these phenomena, we defined 10 liability classes: four classes were defined with different penetrances for colorectal cancer; four classes for polyp carriers and two more liability classes for spouses, that carry a population risk of developing polyps or colorectal cancer and one for the family members of which the disease status is not known. These liability classes are based

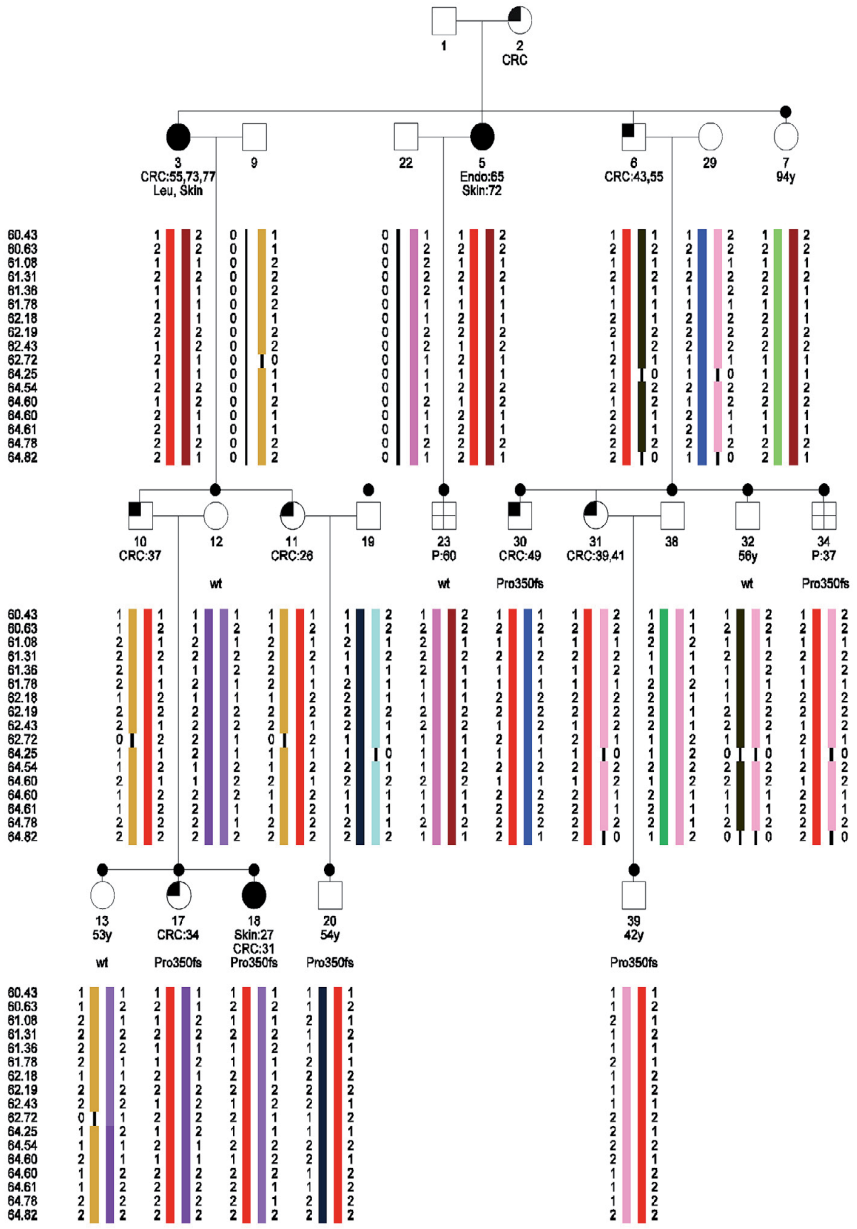

Figure 1

Haplotype analysis in a HNPCC family segregating the *MLH1* Pro350fs mutation. The haplotypes were constructed in SimWalk2 and subsequently visualized with HaploPainter [39]. CRC:55, colorectal cancer diagnosed at age 55; Endo, endometrial cancer; Skin, skin cancer; P, polyps; Pro350fs, carrier of the Pro350fs mutation in *MLH1*; wt, non-carrier; black dot, DNA of this family member has been typed on a 10K SNP array.

on the incidences of CRC and adenomas in the members of HNPCC families in the Netherlands, that do not carry the disease causing mutation [38].

In MENDEL [30], an affected-only parametric linkage analysis was performed using two flanking markers (computation time: ~20 sec per chromosome). In this analysis only family members with colorectal cancer were defined as affected and all other persons were set to unknown. Parametric linkage analysis with liability classes was performed thereafter, using two flanking markers (computation time: ~20 sec per chromosome). Cancers other than colorectal cancer were not considered to be part of the syndrome. In the ROIs appearing from this linkage analysis, possible Mendelian consistent errors were removed as well as the possible presence of linkage disequilibrium. Subsequently, multipoint parametric linkage analysis was performed in SimWalk2 [31], using the ten liability classes. In this multipoint analysis no more than 30 SNPs were analyzed, limited by the computational complexity (analysis time: 1 3/4 hours for 30 SNPs).

Haplotype analysis

Haplotype analysis was performed in the ROI, using SimWalk2. All SNPs in the region of interest (~18) were included in this analysis (computation time: 1 1/3 hours for 18 SNPs). The results of the haplotyping were visualized in HaploPainter [39]. The haplotype segregation in the family could then be compared to the segregation of the mutation in *MLH1* in this family.

Results and discussion

Linkage analysis using bi-allelic genotype data from SNP arrays and large families is a computational challenge using commonly used, freely available analysis software. For the different steps of the linkage analysis; e.g. data formatting, detection of Mendelian inconsistencies, mistyping analysis, LD removal and single to multipoint linkage analysis, we have chosen the following programs that can handle large pedigrees and many SNPs where required; Alohomora [32], Mega2 [33], MENDEL [30], SNPLINK [29] and SimWalk2 [31].

In advance of the linkage analysis we performed a simulation study to calculate the power using Simlink. The mean LOD score in 1000 simulations in this family was 2.0.

The Alohomora program [32] was used first to combine the genotype data generated with the SNP arrays, and the pedigree and SNP information and secondly, to convert these data into the appropriate format for further analysis. In addition, 1256 of the 10053 SNPs were uninformative and were therefore removed from analysis by Alohomora.

Since errors in genotyping can easily mask linkage, the data were checked for different types of errors. First, we have estimated the genotyping error rate in five duplicate experiments. The mean genotyping error rate between the duplicates was only 0.0051.

Mega2 was then used for several data validation checks, including errors in the pedigree data or Mendelian inconsistent errors. Mega2 was used since it supports 28 different programs, including the programs MENDEL and SimWalk2, which we have used for linkage analysis and haplotype analysis. The genotypes of 18 SNPs (0.21%) were removed from analysis, because of Mendelian inconsistencies. However, with bi-allelic markers not all errors appear as Mendelian inconsistent errors [40]. The data were therefore also checked for Mendelian consistent errors. Because of the computational complexity of these multipoint analyses, this error check was performed only in the regions of interest. The mistyping analysis option in SimWalk2 was used, since this program can handle such a complex analysis in a large pedigree. No Mendelian consistent errors were identified in the ROI.

Affected-only parametric linkage analysis and parametric linkage analysis using liability classes was performed in MENDEL, using two flanking markers. This analysis showed a maximum LOD score of 1.8 in the affected-only analysis and 1.9 using liability classes for a 1.7 Mb region around the *MLH1* gene on chromosome 3 (Figure 2). A second region with a LOD 1.1 was found, both in the affected-only analysis and using liability classes, near the centromere on chromosome 11.

Current linkage analysis programs assume LD between markers and a disease locus and importantly, linkage equilibrium between markers. The presence of linkage disequilibrium between two markers can falsely inflate the LOD score and missing genotypes can increase this effect. Therefore, the r^2 as a measure of LD was computed in Merlin and SNPLINK. Using the threshold $r^2 \geq 0.4$, 5 of the 27 SNPs in the region on chromosome 3 were removed from the analysis. From the region of interest on chromosome 11, 14 of the 30 SNPs with an $r^2 \geq 0.4$ were removed from the analysis. After LD removal, multipoint linkage analysis in the region on chromosome 3 yielded a LOD score of 2.1, whereas on chromosome 11 negative LOD scores were seen by multipoint linkage analysis after LD was removed. This indicates that the strong LD in the region on chromosome 11 was responsible for the peak in the LOD in that region. On both chromosomes, the removal of SNPs with high LD had no significant effect on the information content (not shown).

We inferred the haplotypes of the family members, using SimWalk2 for the linkage region on chromosome 3. All

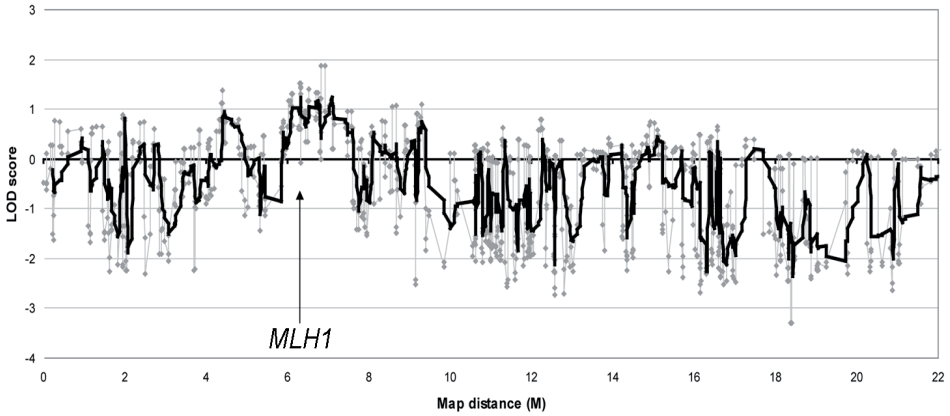

Figure 2
Parametric linkage analysis on chromosome 3, using two flanking markers. The maximum LOD score is 1.9. The gray line represents the raw results of the linkage analysis. The black line is the moving average with a period of ten.

known affected *MLH1*-mutation carriers share the same haplotype, as well as the affected obligate carriers. Therefore, this haplotype perfectly co-segregates with the clinical phenotype of the family members (Figure 1). Case 23, who had developed polyps at age 60, does not share this haplotype. Subsequent mutation analysis showed that this individual indeed did not carry the disease causing mutation in *MLH1*. Therefore, this case showed to be a phenocopy. Another family member, case 39, has to date not developed clinical symptoms of HNPCC, although he did inherit the disease causing allele according to the haplotype analysis. Indeed, sequence analysis showed that this person carries the mutation.

Conclusion

In conclusion, we show that we can perform linkage analysis with high-density 10K SNP arrays in large families for which not all members could be genotyped. We developed a workflow with different publicly available software to perform the analyses: removal of Mendelian consistent and Mendelian inconsistent errors, two and multipoint parametric linkage analysis, removal of linkage disequilibrium and haplotype analysis. The procedure was validated in a large CRC family carrying a known germ line mutation in *MLH1*. Linkage was found with the *MLH1* gene and subsequent haplotype analysis corresponds to the mutation status of the family members. This procedure can now be used for linkage analysis of large families with an inherited condition, such as hereditary colorectal cancer.

List of abbreviations

CRC; colorectal cancer

SNP; single nucleotide polymorphism

LD; linkage disequilibrium

LOD; log of odds

HNPCC; hereditary nonpolyposis colorectal cancer

GDAS; GeneChip DNA Analysis Software

ROI; regions of interest

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

AM performed SNP arrays and mutation analysis, statistical analyses and drafted the manuscript. SJC and QH assisted in the statistical analysis, and QH performed the LD analysis. HMVDK performed SNP arrays. CMJT provided DNA samples and mutation status. HFAV was responsible for family recruitment and surveillance. PD participated in study design. JJHD supervised the statistical analysis, JTW, HM and TVW designed and coordinated the study, TVW helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Grant support: Dutch Cancer Society UL 2005–3247, Nuts Ohra T07-092, NWO 912-03-014.

We thank Katja Philippo for her assistance in identifying the mutation status of several family members.

References

1. Warthin AS: **Heredity with reference to carcinoma.** *Arch Intern Med* 1913, **9:**546-555.
2. Lynch HT, Shaw MW, Magnuson CW, Larsen AL, Krush AJ: **Hereditary factors in cancer. Study of two large midwestern kindreds.** *Arch Intern Med* 1966, **117:**206-212.
3. Vasen HF, Mecklin JP, Khan PM, Lynch HT: **The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC).** *Dis Colon Rectum* 1991, **34:**424-425.
4. Vasen HF, Watson P, Mecklin JP, Lynch HT: **New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC.** *Gastroenterology* 1999, **116:**1453-1456.
5. Peltomaki P, Aaltonen LA, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Green JS, Jass JR, Weber JL, Leach FS, .: **Genetic mapping of a locus predisposing to human colorectal cancer.** *Science* 1993, **260:**810-812.
6. Lindblom A, Tannergard P, Werelius B, Nordenskjold M: **Genetic mapping of a second locus predisposing to hereditary nonpolyposis colon cancer.** *Nat Genet* 1993, **5:**279-282.
7. Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R: **The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer.** *Cell* 1993, **75:**1027-1038.
8. Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, Earabino C, Lipford J, Lindblom A, .: **Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer.** *Nature* 1994, **368:**258-261.
9. Nicolaides NC, Papadopoulos N, Liu B, Wei YF, Carter KC, Ruben SM, Rosen CA, Haseltine WA, Fleischmann RD, Fraser CM: **Mutations of two PMS homologues in hereditary nonpolyposis colon cancer.** *Nature* 1994, **371:**75-80.
10. Akiyama Y, Sato H, Yamada T, Nagasaki H, Tsuchiya A, Abe R, Yuasa Y: **Germ-line mutation of the hMSH6/GTBP gene in an atypical hereditary nonpolyposis colorectal cancer kindred.** *Cancer Res* 1997, **57:**3920-3923.
11. Miyaki M, Konishi M, Tanaka K, Kikuchi-Yanoshita R, Muraoka M, Yasuno M, Igari T, Koike M, Chiba M, Mori T: **Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer.** *Nat Genet* 1997, **17:**271-272.
12. Al Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, Hodges AK, Davies DR, David SS, Sampson JR, Cheadle JP: **Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors.** *Nat Genet* 2002, **30:**227-232.
13. Kinzler KW, Vogelstein B: **Lessons from hereditary colorectal cancer.** *Cell* 1996, **87:**159-170.
14. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, 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.
15. Peto J, Houlston RS: **Genetics and the common cancers.** *Eur J Cancer* 2001, **37 Suppl 8:**S88-S96.
16. Djureinovic T, Skoglund J, Vandrovcova J, Zhou X, Kalushkova A, Iselius L, Lindblom A: **A genome-wide linkage analysis in Swedish families with hereditary non-FAP/non-HNPCC colorectal cancer.** *Gut* 2005.
17. Kemp Z, Carvajal-Carmona L, Spain S, Barclay E, Gorman M, Martin L, Jaeger E, Brooks N, Bishop DT, Thomas H, Tomlinson I, Papaemmanuil E, Webb E, Sellick GS, Wood W, Evans G, Lucassen A, Maher ER, Houlston RS: **Evidence for a colorectal cancer susceptibility locus on chromosome 3q21-q24 from a high-density SNP genome-wide linkage scan.** *Hum Mol Genet* 2006.
18. Kemp ZE, Carvajal-Carmona LG, Barclay E, Gorman M, Martin L, Wood W, Rowan A, Donohue C, Spain S, Jaeger E, Evans DG, Maher ER, Bishop T, Thomas H, Houlston R, Tomlinson I: **Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United kingdom.** *Cancer Res* 2006, **66:**5003-5006.
19. 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, Stratton MR: **Inherited susceptibility to colorectal adenomas and carcinomas: evidence for a new predisposition gene on 15q14-q22.** *Gastroenterology* 1999, **116:**789-795.
20. Wiesner GL, Daley D, Lewis S, Ticknor C, Platzer P, Lutterbaugh J, MacMillen M, Baliner B, Willis J, Elston RC, 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:**12961-12965.
21. John S, Shephard N, Liu G, Zeggini E, Cao M, Chen W, Vasavda N, Mills T, Barton A, Hinks A, Eyre S, Jones KW, Ollier W, Silman A, Gibson N, Worthington J, Kennedy GC: **Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites.** *Am J Hum Genet* 2004, **75:**54-64.
22. Arinami T, Ohtsuki T, Ishiguro H, Ujike H, Tanaka Y, Morita Y, Mineta M, Takeichi M, Yamada S, Imamura A, Ohara K, Shibuya H, Ohara K, Suzuki Y, Muratake T, Kaneko N, Someya T, Inada T, Yoshikawa T, Toyota T, Yamada K, Kojima T, Takahashi S, Osamu O, Shinkai T, Nakamura M, Fukuzako H, Hashiguchi T, Niwa SI, Ueno T, Tachikawa H, Hori T, Asada T, Nanko S, Kunugi H, Hashimoto R, Ozaki N, Iwata N, Harano M, Arai H, Ohnuma T, Kusumi I, Koyama T, Yoneda H, Fukumaki Y, Shibata H, Kaneko S, Higuchi H, Yasui-Furukori N, Numachi Y, Itokawa M, Okazaki Y: **Genomewide high-density SNP linkage analysis of 236 Japanese families supports the existence of schizophrenia susceptibility Loci on chromosomes 1p, 14q, and 20p.** *Am J Hum Genet* 2005, **77:**937-944.
23. Sellick GS, Coleman RJ, Webb EL, Chow J, Bevan S, Rosbotham JL, Houlston RS: **Dominantly inherited cutaneous small-vessel lymphocytic vasculitis maps to chromosome 6q26-q27.** *Hum Genet* 2005, **118:**82-6.
24. Sellick GS, Webb EL, Allinson R, Matutes E, Dyer MJ, Jonsson V, Langerak AW, Mauro FR, Fuller S, Wiley J, Lyttelton M, Callea V, Yuille M, Catovsky D, Houlston RS: **A high-density SNP genomewide linkage scan for chronic lymphocytic leukemia-susceptibility loci.** *Am J Hum Genet* 2005, **77:**420-429.
25. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: **Parametric and nonparametric linkage analysis: a unified multipoint approach.** *Am J Hum Genet* 1996, **58:**1347-1363.
26. Abecasis GR, Cherny SS, Cookson WO, Cardon LR: **Merlin--rapid analysis of dense genetic maps using sparse gene flow trees.** *Nat Genet* 2002, **30:**97-101.
27. Gudbjartsson DF, Jonasson K, Frigge ML, Kong A: **Allegro, a new computer program for multipoint linkage analysis.** *Nat Genet* 2000, **25:**12-13.
28. Gudbjartsson DF, Thorvaldsson T, Kong A, Gunnarsson G, Ingolfsdottir A: **Allegro version 2.** *Nat Genet* 2005, **37:**1015-1016.
29. Webb EL, Sellick GS, Houlston RS: **SNPLINK: multipoint linkage analysis of densely distributed SNP data incorporating automated linkage disequilibrium removal.** *Bioinformatics* 2005, **21:**3060-3061.
30. Lange K, Cantor R, Horvath S, Perola M, Sabatti C, Sinsheimer J, Sobel E: **Mendel version 4.0: a complete package for the exact genetic analysis of discrete traits in pedigree and population data sets.** *Am J Hum Genet* 2001, **69(Supplement):**A1886.
31. Sobel E, Lange K: **Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics.** *Am J Hum Genet* 1996, **58:**1323-1337.
32. Ruschendorf F, Nurnberg P: **ALOHOMORA: a tool for linkage analysis using 10K SNP array data.** *Bioinformatics* 2005, **21:**2123-2125.
33. Mukhopadhyay N, Almasy L, Schroeder M, Mulvihill WP, Weeks DE: **Mega2: data-handling for facilitating genetic linkage and association analyses.** *Bioinformatics* 2005, **21:**2556-2557.
34. Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, Liu G, Su X, Cao M, Chen W, Zhang J, Liu W, Yang G, Di X, Ryder T, He Z, Surti U, Phillips MS, Boyce-Jacino MT, Fodor SP, Jones KW: **Large-scale genotyping of complex DNA.** *Nat Biotechnol* 2003, **21:**1233-1237.
35. Matsuzaki H, Loi H, Dong S, Tsai YY, Fang J, Law J, Di X, Liu WM, Yang G, Liu G, Huang J, Kennedy GC, Ryder TB, Marcus GA, Walsh PS, Shriver MD, Puck JM, Jones KW, Mei R: **Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array.** *Genome Res* 2004, **14:**414-425.

36. Sobel E, Papp JC, Lange K: **Detection and integration of genotyping errors in statistical genetics.** *Am J Hum Genet* 2002, **70:**496-508.
37. Boehnke M, Ploughman LM: **SIMLINK: A Program for Estimating the Power of a Proposed Linkage Study by Computer Simulation.** 1997 [http://csg.sph.umich.edu/boehnke/simlink.php].
38. de Jong AE, Morreau H, Nagengast FM, Mathus-Vliegen EM, Kleibeuker JH, Griffioen G, Cats A, Vasen HF: **Prevalence of adenomas among young individuals at average risk for colorectal cancer.** *Am J Gastroenterol* 2005, **100:**139-143.
39. Thiele H, Nurnberg P: **HaploPainter: a tool for drawing pedigrees with complex haplotypes.** *Bioinformatics* 2005, **21:**1730-1732.
40. Gordon D, Heath SC, Ott J: **True pedigree errors more frequent than apparent errors for single nucleotide polymorphisms.** *Hum Hered* 1999, **49:**65-70.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2407/7/6/prepub

Chapter 3

Comprehensive Genetic Analysis of Seven Large Families with Mismatch Repair Proficient Colorectal Cancer

Genes, Chromosomes and Cancer (2010) 49:539-548

Comprehensive Genetic Analysis of Seven Large Families with Mismatch Repair Proficient Colorectal Cancer

Anneke Middeldorp,[1] Shantie C. Jagmohan-Changur,[2] Heleen M. van der Klift,[2] Marjo van Puijenbroek,[1] Jeanine J. Houwing-Duistermaat,[3] Emily Webb,[4] Richard Houlston,[4] Carli Tops,[5] Hans F. A. Vasen,[6] Peter Devilee,[1,2] Hans Morreau,[1] Tom van Wezel,[1] and Juul Wijnen[2,5]*

[1]Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands
[2]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
[3]Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands
[4]Section of Cancer Genetics, Institute of Cancer Research, Sutton, United Kingdom
[5]Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
[6]The Netherlands Foundation for the Detection of Hereditary Tumours, Leiden, The Netherlands

Approximately 40% of colorectal cancer (CRC) families with a diagnosis of hereditary nonpolyposis CRC on the basis of clinical criteria are not a consequence of mismatch repair (MMR) deficiency. Such families provide supporting evidence for the existence of a hitherto unidentified highly penetrant gene mutation. To gain further understanding of MMR-competent familial colorectal cancer (FCC), we studied seven large families with an unexplained predisposition for CRC to identify genetic regions that could harbor CRC risk factors. First, we conducted a genome-wide linkage scan using 10K single-nucleotide polymorphism (SNP) arrays to search for disease loci. Second, we studied the genomic profiles of the tumors of affected family members to identify commonly altered genomic regions likely to harbor tumor suppressor genes. Finally, we studied the possible role of recently identified low-risk variants in the familial aggregation of CRC in these families. Linkage analysis did not reveal clear regions of linkage to CRC. However, our results provide support linkage to 3q, a region that has previously been linked to CRC susceptibility. Tumor profiling did not reveal any genomic regions commonly targeted in the tumors studied here. Overall, the genomic profiles of the tumors show some resemblance to sporadic CRC, but additional aberrations were also present. Furthermore, the FCC families did not appear to have an enrichment of low-risk CRC susceptibility loci. These data suggest that factors other than a highly penetrant risk factor, such as low or moderate-penetrance risk factors, may explain the increased cancer risk in a subset of familial CRCs.

INTRODUCTION

Colorectal cancer (CRC) is one of the most common malignancies in Western populations (Parkin et al., 2005). As estimated in twin studies, hereditary factors may play a role in up to 35% of CRC cases (Lichtenstein et al., 2000). In the early 1990s, the first gene conferring a high risk of developing CRC was described for familial adenomatous polyposis (Bodmer et al., 1987; Leppert et al., 1987; Groden et al., 1991; Kinzler et al., 1991). The gene defects of several Mendelian disorders have been identified since then, including Lynch syndrome, *MUTYH*-associated polyposis, Juvenile Polyposis, and Peutz–Jeghers syndrome. However, these syndromes account for only ~6% of CRC cases. In the other familial CRC cases, the underlying genetic factors are currently unknown (Jenkins et al., 2002; Aaltonen et al., 2007).

The Amsterdam criteria I (AC-I), based on family history and age at diagnosis of CRC, are used to identify patients with a presumptive diagnosis of Lynch syndrome (Vasen et al., 1999). However, ~40% of patients fulfilling the AC-I do not have tumors with microsatellite instability, which is characteristic of a mismatch repair (MMR) deficiency. These data provide strong evidence that other genetic factors may play a role in the development of CRC in these families

Supported by: Dutch Cancer Society, Grant number: UL2005-3247; NutsOhra Foundation, Grant number: SNO-T-07-092.

*Correspondence to: Juul Wijnen, Leiden University Medical Center, Department of Human Genetics, PO Box 9600, 2300 RC Leiden, The Netherlands. E-mail: j.wijnen@lumc.nl

Received 21 January 2010; Accepted 3 February 2010

DOI 10.1002/gcc.20763

Published online 10 March 2010 in Wiley InterScience (www.interscience.wiley.com).

TABLE 1. Characteristics of the Seven Families Studied

Family	No. of individuals	AC-I	Mean age at diagnosis (in years)	Number of CRCs	Number of CRAs	Other cancers present in family members
na16	18 (13)	+	44	3	3	Breast
na41	31 (20)	–[a]	60	6	9	–
na46	21 (13)	–[a]	58	5	6	Pancreas
na58	19 (15)	+	51	4	2	Lung, endometrium
na61	21 (17)	+	52	3	4	Gastric
na68	26 (13)	+	53	4	8	–
na209	33 (24)	+	51	6	2	Thyroid, breast, endometrium

Number of individuals indicates the number of individuals included in the linkage analysis. The number in the parentheses is the number of individuals genotyped with a 10K SNP array.
AC-I indicates Amsterdam criteria I.
Mean age at diagnosis is the mean age at time of diagnosis of CRC in individuals within the analyzed families.
Number of CRCs indicates the number of family members diagnosed with CRC.
Number of CRAs indicates the number of family members diagnosed with CRA.
[a]Families were characterized by familial aggregation of CRC and early-onset disease (51 and 55 years for na41 and na46, respectively).

(Wijnen et al., 1998; Vasen et al., 1999; de Jong et al., 2004; Mangold et al., 2005).

Several linkage studies of dominantly inherited microsatellite stable (MSS) CRC families have been performed, and various genomic regions, including 3q21-q24, 7q31, 9q22.2-31.2, 11q23.2, 11q13.4, 14q24.2, and 22q12.1, have been linked to CRC predisposition (Wiesner et al., 2003; Djureinovic et al., 2006; Skoglund et al., 2006; Kemp et al., 2006a, b; Neklason et al., 2008; Papaemmanuil et al., 2008; Picelli et al., 2008). To date, none of these studies has, however led to the identification of a novel CRC susceptibility gene.

In addition to highly penetrant mutations, familial clustering could be caused by polygenic susceptibility. Evidence for several low-risk variants for CRC has recently been provided by genome-wide association (GWA) studies (Broderick et al., 2007; Tomlinson et al., 2007; Zanke et al., 2007; Houlston et al., 2008; Jaeger et al., 2008; Tenesa et al., 2008; Tomlinson et al., 2008). We previously have demonstrated that familial CRC cases display a modest enrichment of these low-risk variants (Middeldorp et al., 2009).

Profiles of genomic aberrations in tumors of patients with familial CRC provide a means of obtaining insight into the biological basis of CRC. Distinct characteristic profiles have already been described for breast cancers from patients with germ line mutations in *BRCA1* and to a lesser extent *BRCA2* (Wessels et al., 2002; Jonsson et al., 2005; Joosse et al., 2008). Similarly, the genomic profiles of Lynch syndrome carcinomas, familial adenomatous polyposis adenomas, *MUTYH*-associated polyposis carcinomas, and sporadic CRC are clearly distinct (Cardoso et al., 2006; Diep et al., 2006; Lips et al., 2007; Middeldorp et al., 2008; van Puijenbroek et al., 2008). Hence, tumor profiles of CRC from families with an unexplained CRC predisposition offer the prospect of identifying regions commonly affected by aberrations that are shared by the affected individuals within the families.

To further our understanding of MMR-competent familial CRC, we analyzed seven large Dutch families with a history of CRC. We performed a genome-wide linkage scan and studied the genomic profile of tumors from family members. Finally, we evaluated the contribution of recently identified low-risk loci in the susceptibility to CRC in these families.

MATERIALS AND METHODS

Families

Seven families with a history of CRC that segregates in a dominant fashion were studied. Five of the families (na16, na58, na61, na68, and na209) fulfilled the AC-I (Table 1). The other two families (na41 and na46), while not AC-I positive, were characterized by familial aggregation of CRC and by early-onset disease (51 and 55 years of age). In a number of the families, several other malignancies were also reported in family members, including ovarian, breast, endometrial, thyroid, gastric, and pancreatic cancers. Furthermore, a number of the younger members of the families had been diagnosed with polypoid precursor lesions by endoscopic surveillance.

The pedigree information for the families was collected through the Netherlands Foundation for the Detection of Hereditary Tumors (http://

www.stoet.nl). Pedigree drawings are available upon request. One de novo APC mutation (exon 9: c.1192_1193delAA, p.Lys398GlufsX5), associated with attenuated Familial Adenomatous Polyposis, was detected in a branch of family na61. In the other families, no pathogenic germ line mutations were identified in *APC*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *MUTYH*, *EXO1*, *MLH3*, *TGFBRII*, or *MED1*. Tumors were tested for microsatellite instability using the marker set recommended by the National Cancer Institute Workshop on Microsatellite Instability (Boland et al., 1998), and eight of nine tumors were analyzed with three additional mononucleotide repeat markers (BAT40, MSH3, and MSH6). All the tumors from affected members of the seven families that were available for MSI testing (9 tumors) were MSS, except for one tumor from a member of family na16. In this family, one family member had an MSI-low tumor, and two other family members had MSS tumors. Two additional tumors stained positive using immunohistochemistry for MLH1, MSH2, MSH6, and PMS2.

This study was approved by the Medical Ethical Committee of the Leiden University Medical Center, Leiden, The Netherlands (protocol P01-019), in accordance with the tenets of the declaration of Helsinki.

Genotyping

Peripheral blood samples were collected from 112 family members. DNA was extracted and quantified using standard techniques. All family members were genotyped using Affymetrix GeneChip® Human Mapping 10K 2.0 single-nucleotide polymorphism (SNP) arrays (Affymetrix, Santa Clara, CA). Arrays were processed according to the manufacturer's protocols. We have estimated the genotyping error rate in five duplicate experiments. The mean genotyping error rate between the duplicates was only 0.0051.

Linkage Analysis

We estimated the power of the seven families to identify a disease locus using Simlink, assuming a single-locus dominant trait with a piecewise linear penetrance. The maximum penetrance was set at 80%, and the disease allele frequency was assumed to be 0.001 (Boehnke and Ploughman, 1997). The estimated maximum logarithm of odds (LOD) score for each family was 2.25, 4.44, 3.16, 4.32, 3.61, 2.30, and 4.51 for na16, na41, na46, na58, na61, na68, and na209, respectively.

We applied two different methods to perform the linkage analysis. In the first method, families were analyzed individually using Mendel (Lange et al., 2001) and SimWalk2 (Sobel and Lange, 1996), as described previously (Middeldorp et al., 2007). In brief, uninformative SNPs and Mendelian-inconsistent errors were removed. Parametric linkage analysis using two flanking markers was performed in Mendel; both affected-only analysis and parametric linkage analysis using liability classes were performed. In the affected-only analysis, family members diagnosed with CRC or with adenomas before the age of 50 years were classified as affected, and all other family members were set to unknown. Liability classes were based on the incidences of CRC and adenoma in the members of Lynch syndrome families in the Netherlands that do not carry a disease causing mutation in one of the DNA MMR genes (de Jong et al., 2005). Four age groups were defined (age at diagnosis <30, 30–45, 45–60, and ≥ 60 years), with penetrances set at 0.1000, 0.3000, 0.6000, and 0.8000 with corresponding phenocopy rates of 0.0001, 0.0010, 0.0100, and 0.0500 for CRC, and 0.0200, 0.0600, 0.2000, and 0.6000 for colorectal adenomas (CRAs). We considered LOD scores greater than three as a significant linkage and LOD scores greater than two as a suggestive linkage.

In the second method, we combined the data from all families to calculate nonparametric linkage (NPL) scores and heterogeneity LOD (HLOD) scores. This method has been described previously by Kemp et al. (2006a) and Papaemmanuil et al. (2008). Briefly, Mendelian-inconsistent errors were removed, and SNPs showing evidence of linkage disequilibrium (LD) were excluded. All families were analyzed together with multipoint linkage analysis. Nonparametric linkage analysis and parametric linkage analysis were performed using SNPLINK (Webb et al., 2005). The parametric analyses were performed under both dominant and recessive models of inheritance. Four liability classes were used based on age at diagnosis (<50, 50–59, 60–69, and >70 years). Individuals with CRAs were considered equivalent to individuals with CRC who were 15 years older, that is, someone with CRA at age 45 was counted as having CRC at age 60. Two analyses were performed, one based on CRC and one analysis in which affected individuals were defined by having either CRC or CRA.

The two methods that we used differ in their linkage statistics. Mendel uses either the Lander–Green or the Elston–Stewart algorithm, depending on whichever is more efficient for the pedigree structure. SNPlink uses Allegro and Merlin to perform linkage analysis; these programs both use the Lander-Green Hidden Markov Model algorithm. Moreover, Mendel can handle larger families compared to SNPlink. We used both methods of analysis to minimize the chance of having missed possible linkage regions.

Analysis of Low-Risk Variants

We genotyped all available family members of the seven familial colorectal cancer (FCC) families and 310 unrelated healthy controls to analyze 10 CRC risk loci that were recently identified by GWA studies, including rs12953717 (18q21), rs3802842 (11q23), rs6983267 (8q24), rs16892766 (8q23), rs4779584 (15q13), rs10795668 (10p14), rs4444235 (14q22), rs9929218 (16q22), rs10411210 (19q13), and rs961253 (20p12) (Broderick et al., 2007; Tomlinson et al., 2007, 2008; Zanke et al., 2007; Houlston et al., 2008; Jaeger et al., 2008; Tenesa et al., 2008). Healthy controls were derived from the Laboratory for Diagnostic Genome Analysis at the Leiden University Medical Center (The Netherlands) and included individuals that tested mutation-negative (presymptomatically) for noncancer-related diseases. SNP genotyping was performed by allele-specific PCR KASPar chemistry (KBiosciences, UK) following the manufacturer's protocol for all SNPs (primer details are available upon request), except for rs10795668 (10p14). Genotype calling was done using ABI PRISM 7900HT technology (Applied Biosystems, CA).

Genotyping of rs10795668 (10p14) was performed using high-resolution melting curve analysis implemented on a LightCycler (Roche, Woerden, NL), and genotypes were analyzed using LightCycler software (version 1.5.0; Roche). Primer sequences and assay conditions are available upon request.

We studied the association of the 10 low-risk variants with CRC, taking into account that the individuals are related, and therefore their genotypes are correlated (Thornton and McPeek, 2007; Uh et al., 2009). We compared the allele frequencies in the affected family members to the allele frequencies in the healthy family members and healthy controls.

To study whether carrying a high number of low-risk alleles correlated with disease in the families, we determined the number of risk alleles for all family members, counting heterozygotes and homozygotes for the risk allele as 1 and 2, respectively. Generalized estimation equations (GEEs) with the identity matrix as the working correlation were used to examine the possible relationship between affected status and the number of risk alleles in family members. This method accounts for the fact that family members have correlated genotypes. The Wald test was used to test for associations. Moreover, GEEs were used to examine whether the families carry more risk alleles than would be expected based on population frequencies of the variants.

Tumor Profiling

We analyzed three carcinomas and five adenomas from individuals belonging to seven large families with a history of CRC for genomic aberrations including chromosome gains, chromosome losses, and loss of heterozygosity (LOH). Five of these tumors were derived from the families (na16, na41, na46, na58, and na209) that we studied using linkage analysis. The other three tumors originated from two other families (na11 and na50), both fulfilling the AC-I but having a low a priori power for linkage analysis due to limited sample availability. We used Illumina Beadarrays in combination with the linkage-mapping panel IV_B4b (Illumina, San Diego, CA). GoldenGate assays were performed following the manufacturer's instructions with the following minor modifications: 1 μg of input DNA was used for multi-use activation and resuspended in 60 μl of RS1 (Fan et al., 2003). Genotypes were extracted using GenCall (version 6.0.7, Illumina) and GTS Reports (version 4.0.10.0, Illumina). Copy number and copy-neutral LOH (cnLOH) profiles were generated by analyzing the allelic state of the tumors and the corresponding normal tissue in the "Beadarray SNP" package with the LAIR algorithm (Oosting et al., 2007; Corver et al., 2008). Criteria for the scoring of copy number aberrations were based on previous experiments (Oosting et al., 2007). LOH was defined as regions of three or more consecutive SNPs showing LOH. In practice, regions of LOH always presented as stretches of markers showing LOH. Samples were handled according to the medical ethical guidelines described in the Code Proper Secondary Use of Human Tissue established by

the Dutch Federation of Medical Sciences (http://www.federa.org).

RESULTS

Families

The aim of this study was to further our understanding of the genetics that underlie MMR proficient familial CRC. Therefore, we analyzed seven large Dutch CRC families, and we studied the genomic profile of tumors from family members. Clinical characteristics of the seven families analyzed in this study are shown in Table 1. These families included 28 CRC patients and 30 family members with CRAs. The number of CRC patients per family ranged from three to five in successive generations. The mean age at diagnosis of CRC was 53 years old (range, 28–82 years). The mean age at diagnosis of CRA was 50 years old (range, 34–72 years). In all but one family, individuals from three generations were diagnosed with CRC. In family na209, individuals from two generations were diagnosed with CRC.

Genotyping and Linkage Analysis

Using Affymetrix 10K SNP arrays, we genotyped 112 individuals from the seven families, including both healthy and affected family members. The mean SNP call rate was >95% (84.1–99.4%). The high number of SNPs combined with large informative pedigrees made the data analysis highly computationally complex. We previously established a procedure for such complex linkage analysis using existing programs, and we validated our method in a Lynch syndrome family with a known *MLH1* germ line mutation (Middeldorp et al., 2007). Here, we applied this method to our seven families under study. Families were analyzed individually, because every family by itself had good a priori power to identify linked loci. This approach reduces the impact of locus heterogeneity in the analysis. We performed both affected-only parametric linkage analysis and parametric linkage analysis using liability classes. No clear regions of linkage (LOD $\geq$ 3.0) or suggestive linkage (LOD $\geq$ 2.0) were identified in either analysis in any of the seven families (results not shown). LOD scores greater than one were identified frequently. However, these were peaks in the LOD scores of only single SNPs. No regions, including consecutive markers with LOD scores greater than one, were identified in any of the families.

Subsequently, we analyzed all the families together using a different previously validated linkage analysis method (Kemp et al., 2006a; Papaemmanuil et al., 2008). Using this method, we performed two different analyses. In the first analysis, we restricted the analysis to the family members with CRC, whereas, in the second analysis, we also included the individuals with CRA. The first analysis did not yield any clear regions of linkage or suggestive linkage. However, in the second analysis, in which individuals with carcinomas and adenomas were both included, four chromosomal regions with HLOD scores of ~1.5 were identified (Fig. 1). The HLOD scores are shown in Table 2. The four chromosomal regions with HLOD scores close to 1.5 were 3q21.3 (HLOD = 1.49), 6q21 (HLOD = 1.59), 8q24.2 (HLOD = 1.48), and 14q22.1 (HLOD = 1.30).

Association analysis of 10 low-risk variants revealed significant associations with CRC for rs16892766 (8q23.3; $P = 0.03$) and rs12953717 (18q21.1; $P = 0.03$). The risk allele frequency for rs16892766 was 14.3% in the affected family members, compared to 6.8% in their healthy relatives. The allele frequency for rs12953717, the nonrisk allele, was 68.5% in the affected family members, whereas it was only 50% in the healthy family members. The mean number of risk alleles carried by family members with CRC was 9.4 (range, 4–14). The unaffected family members carried an average of 9.0 (range, 4–13) risk alleles. The average number of risk alleles in unrelated controls was 9.2 (range, 4–13). Hence, no significant correlation was observed between the number of risk alleles and the CRC status of the family members ($P = 0.38$). We did neither observe a general enrichment of risk alleles in the families with a history of CRC compared to the unrelated controls.

Tumor Analysis

Using DNA isolated from formalin-fixed paraffin-embedded tissue from five adenomas and three carcinomas and Illumina 6K SNP arrays, we evaluated eight tumors for genome-wide copy number aberrations and LOH (Table 3). We identified a few aberrations in the CRAs. One adenoma had no chromosomal aberrations, whereas three other adenomas displayed one chromosomal aberration each, and the fifth adenoma exhibited four aberrations. Different chromosomal regions were targeted, including loss of chromosomes 5 and 13, gain of chromosomes 7, 8, and 13, and

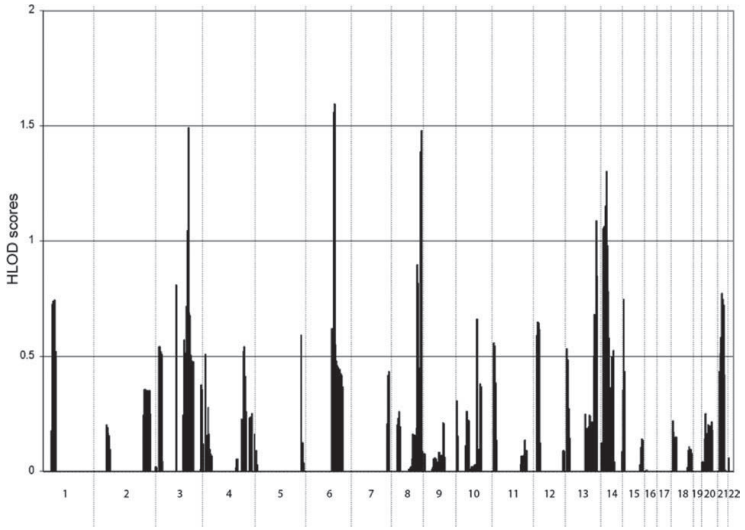

Figure 1. Genome-wide HLOD scores for the combined analysis of all families with both carcinomas and adenomas included in the analysis.

TABLE 2. Maximum HLOD or NPL Scores Determined by the Linkage Analysis

Chromosome	Alpha	HLOD	NPL
Affected = carcinoma			
3p14.1	1.00	1.13	1.82
9q21.33	1.00	1.18	1.78
11q24.2	1.00	1.28	1.68
14q13.3	1.00	1.16	2.26
Affected = carcinomas/adenomas			
3q21.3	0.57	1.49	3.65
6q21	0.54	1.59	3.93
8q24.2	1.00	1.48	2.00
14q22.1	0.51	1.30	3.93

The table shows the results of the linkage analysis in which affected individuals were defined by having CRC (top), and the results of analysis in which affected individuals were defined by having either CRC or CRA (bottom).

cnLOH of chromosome 12. Gain of chromosome 7 and chromosome 13 has been described as prevalent in adenomas (Cardoso et al., 2006; Jones et al., 2007). The aberrations observed at chromosomes 5, 8, and 12 are less common.

The carcinomas we studied displayed many more chromosomal aberrations than the adenomas. The three CRCs displayed 8, 9, and 18 chromosomes with aberrations, respectively (Table 3). The aberrations include genetic changes commonly seen in CRC, such as loss of chromosome 18q, gain of chromosome 13q, gain of chromosome 8q, and gain of chromosome 20q (Diep et al., 2006). However, other aberrations, including gain of chromosome 6p and loss of chromosome 20p, were also observed.

DISCUSSION

Despite the clear familial CRC phenotype and the high estimated a priori power to detect linkage, our analyses did not reveal a novel region of significant linkage in any of the seven large CRC families studied. However, our results do support linkage to a previously reported region on 3q21.3, linked to CRC susceptibility, on the basis of an HLOD score of 1.49 (corresponding to a locus-specific P value of 0.01) (Kemp et al., 2006a; Papaemmanuil et al., 2008). The smallest region of overlap with the different linkage reports in this region is 3q22.1–q22.3, as shown in Figure 2. Collectively, these data provide evidence for a novel CRC susceptibility locus mapping to 3q22. Failure to demonstrate strong evidence of linkage is indicative of the risk conferred by the 3q21.3 locus being modest as opposed to the high-risk profile associated with classical MMR gene mutations. In previous studies, ~40 genes located in this region have been screened for mutations (Papaemmanuil et al., 2008; Picelli et al., 2008). Although no coding mutations have been identified to date, this does not preclude the possibility that the functional basis of the disease locus is mediated through alternative sequence mechanisms, such as regulatory sequences or microRNAs.

TABLE 3. Chromosomal Aberrations Found in Colorectal Adenomas and Carcinomas

Family	Tumor type	Histology	MSI status	Chromosome gain	Chromosome loss	LOH/AI
16–4	Adenoma	Tubular	MSS	13q	–	–[a]
41–3	Adenoma	Tubular	NA	–	–	–
11–2	Adenoma	Tubular, villous	MSS	–	5q14.3–23.3	–
46–1	Adenoma	Villous	MSS	–	13q	–
58–9	Adenoma	Villous	MSS	7pq, 8pq, 12p13.33–q24.1	–	12pq

Family	Tumor type	Tumor grade	MSI status	Chromosome gain	Chromosome loss	LOH/AI
209–15	Carcinoma	Dukes B2	MSS	1p34.3–33, 6p25.3–12.3, 7p22.3–14.1, 8q, 12p13.33–13.31, 13q11–12.3, 13q21.32–32.34, 20q	1p36.33–34.3, 7p14.1, 8p, 11q14.1–25, 12p13.31–11.1, 20p13–12.1	11q12.3–13.5, 13q12.3–21.31, 20p11.23–11.21
11–2	Carcinoma	Dukes C1	MSS	6p, 7pq, 8q, 9q33.3–34.2, 13q21.31–34, 14q11.1–11.2, 16pq, 17q, 18p11.32–q12.1, 19pq, 20pq	1pq, 3p, 5q14.3–23.3, 18q12.2–22.3, 22q	2q21.2–37.3, 3q, 9pter–q33.2, 11q14.3–24.1, 13q11–21.31, 17p, 18q23, 21q
50–1	Carcinoma	Dukes D	MSS	13q, 20q	8p, 14q, 17p, 18q, 20p13–12.2	1q42.13–44, 5q14.3–35.5, 6p, 20p12.1–11.21

MSI indicates the microsatellite stability of the tumor; MSS means that the tumor was microsatellite stable. NA indicates that MSI status could not be assessed. LOH/AI indicates chromosomal regions displaying loss of heterozygosity or allelic imbalance in the absence of copy number alterations. Minus sign (–) indicates that no chromosomal aberrations were present.
[a]Tumor in which LOH could not be analyzed for chromosomes 5, 6, 7, 8, and 9.

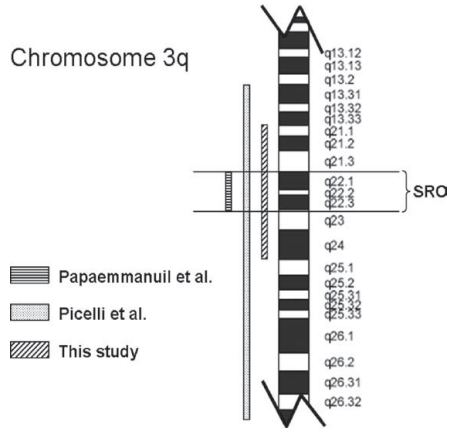

Figure 2. Linkage results for chromosome 3q. The linkage regions determined by Papaemmanuil et al. (2008) Picelli et al. (2008), and our study are shown here. SRO indicates the smallest region of overlap.

In addition to 3q, 14q24.2 has been suggested to be linked to CRC. Although we found some evidence for linkage to 14q22.1, our region of linkage does not overlap with the previously published locus (Djureinovic et al., 2006). Other linkage regions that have been previously reported, albeit nonsignificant on a genome-wide basis include regions on 7q, 9q, 11q, and 22q. No CRC linkages with these loci were supported by our results. The absence of replication of linkage in these regions could be due to differences in family ascertainment, because we restricted our analysis to large pedigrees rather than nuclear CRC families, which were used in previous studies. Furthermore, it is highly unlikely that we failed to demonstrate a significant linkage as we used multiple statistical strategies to maximize the probability of identifying a disease locus.

Our results provide evidence supporting the hypothesis that a single highly penetrant genetic risk factor is unlikely to make a major contribution to the excess familial risk associated with MSS cancers. This establishes that a model based on a combination of moderate-risk or multiple low-risk factors is more likely. Rare alleles conferring moderate risk are very difficult to identify through association-based analyses. To date, most efforts to identify nonhigh-penetrance variants have been directed to common low-risk variants using GWA studies. We studied the impact of the 10 currently known low-risk variants on familial risk in seven large CRC families. The variant rs16892766 (8q23.3) is significantly associated with CRC in these families. Intriguingly, this variant has recently been found to have a modifier effect in Lynch families (Wijnen et al., 2009). Paradoxically, the SNP rs12953717 (18q21.1)

association was counter to that seen in unselected cases. Although this may reflect interaction with another (unknown) risk factor in the CRC families, the observation may simply be reflective of the small number of individuals analyzed in this study. Similarly, no relationship between the number of risk alleles that family members carry and their CRC status was identified. Collectively, these data indicate that the currently identified low-risk variants are insufficient to account for the type of familial clustering of CRC seen in the families we analyzed.

Analysis of genome-wide copy number aberrations and LOH showed that adenomas display only a few chromosomal aberrations, as has been previously described (Jones et al., 2007). Moreover, our results do not suggest the existence of a specific chromosomal target region for tumor initiation that would point to a susceptibility locus responsible for CRC in the families we analyzed. In contrast, the carcinomas we studied displayed many chromosomal aberrations. The profiles of the carcinomas show similarities with the patterns of aberrations observed in sporadic CRC, but additional aberrations were also observed.

When postulating that the regions identified using linkage analysis harbor tumor suppressor genes, it is likely that these regions are targeted early in the tumor by chromosomal aberrations. However, in the adenomas, we did not identify any aberrations at 3q21.3, 6q21, 8q24.2, or 14q22.1, the regions with the highest LOD scores, except for one gain at chromosome arm 8q in tumor 58–9. However, a gain at chromosome 8q is a frequently observed event in CRC. Overall, the profiles identified do not resemble the profile of tumors from Lynch syndrome patients, *MUTYH*-associated polyposis patients, or Familial Polyposis Syndrome patients. Although the profiles do show some resemblance to sporadic CRC profiles, other aberrations were identified, including gain of 6p and loss of 20p (Middeldorp et al., 2008; van Puijenbroek et al., 2008).

In conclusion, we did not find evidence for a high-penetrance genetic factor that can explain the increased CRC risk in these families. However, linkage results for 3q support previous reports that this locus might harbor a moderate- or high-risk CRC allele. However, tumor analysis did not identify a chromosomal loss or LOH at this region on 3q, as would be expected in case of a tumor suppressor function. No enrichment in the number of low-risk alleles was observed in the families we studied. The genomic profiles of the tumors seem distinct from other familial syndromes and show resemblance to sporadic CRC. Overall, these data suggest that factors other than a high-penetrance risk factor, such as low- or moderate-risk factors, may explain the increased cancer risk in a subset of familial CRC.

REFERENCES

Aaltonen L, Johns L, Jarvinen H, Mecklin JP, Houlston R. 2007. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. Clin Cancer Res 13:356–361.

Bodmer WF, Bailey CJ, Bodmer J, Bussey HJ, Ellis A, Gorman P, Lucibello FC, Murday VA, Rider SH, Scambler P. 1987. Localization of the gene for familial adenomatous polyposis on chromosome 5. Nature 328:614–616.

Boehnke M, Ploughman LM. SIMLINK: A program for estimating the power of a proposed linkage study by computer simulation. [Version 4.12] April 2, 1997.

Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, Srivastava S. 1998. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: Development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 58:5248–5257.

Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, Lubbe S, Spain S, Sullivan K, Fielding S, Jaeger E, Vijayakrishnan J, Kemp Z, Gorman M, Chandler I, Papaemmanuil E, Penegar S, Wood W, Sellick G, Qureshi M, Teixeira A, Domingo E, Barclay E, Martin L, Sieber O, Kerr D, Gray R, Peto J, Cazier JB, Tomlinson I, Houlston RS. 2007. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet 39:1315–1317.

Cardoso J, Molenaar L, de Menezes RX, van Leerdam M, Rosenberg C, Moslein G, Sampson J, Morreau H, Boer JM, Fodde R. 2006. Chromosomal instability in MYH- and APC-mutant adenomatous polyps. Cancer Res 66:2514–2519.

Corver WE, Middeldorp A, Ter Haar NT, Jordanova ES, van Puijenbroek M, van Eijk R, Cornelisse CJ, Fleuren GJ, Morreau H, Oosting J, van Wezel T. 2008. Genome-wide allelic state analysis on flow-sorted tumor fractions provides an accurate measure of chromosomal aberrations. Cancer Res 68:10333–10340.

deJong AE, van Puijenbroek M, Hendriks Y, Tops C, Wijnen J, Ausems MG, Meijers-Heijboer H, Wagner A, van Os TA, Brocker-Vriends AH, Vasen HF, Morreau H. 2004. Microsatellite instability, immunohistochemistry, and additional PMS2 staining in suspected hereditary nonpolyposis colorectal cancer. Clin Cancer Res 10:972–980.

deJong AE, Morreau H, Nagengast FM, Mathus-Vliegen EM, Kleibeuker JH, Griffioen G, Cats A, Vasen HF. 2005. Prevalence of adenomas among young individuals at average risk for colorectal cancer. Am J Gastroenterol 100:139–143.

Diep CB, Kleivi K, Ribeiro FR, Teixeira MR, Lindgjaerde OC, Lothe RA. 2006. The order of genetic events associated with colorectal cancer progression inferred from meta-analysis of copy number changes. Genes Chromosomes Cancer 45:31–41.

Djureinovic T, Skoglund J, Vandrovcova J, Zhou XL, Kalushkova A, Iselius L, Lindblom A. 2006. A genome wide linkage analysis in Swedish families with hereditary non-familial adenomatous polyposis/non-hereditary non-polyposis colorectal cancer. Gut 55:362–366.

Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS. 2003. Highly parallel SNP genotyping. Cold Spring Harb Symp Quant Biol 68:69–78.

Groden J, Thliveris A, Samowitz W, Carlson M, Gelbert L, Albertsen H, Joslyn G, Stevens J, Spirio L, Robertson M. 1991.

Identification and characterization of the familial adenomatous polyposis coli gene. Cell 66:589–600.

Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, Chandler I, Vijayakrishnan J, Sullivan K, Penegar S, Carvajal-Carmona L, Howarth K, Jaeger E, Spain SL, Walther A, Barclay E, Martin L, Gorman M, Domingo E, Teixeira AS, Kerr D, Cazier JB, Niittymaki I, Tuupanen S, Karhu A, Aaltonen LA, Tomlinson IP, Farrington SM, Tenesa A, Prendergast JG, Barnetson RA, Cetnarskyj R, Porteous ME, Pharoah PD, Koessler T, Hampe J, Buch S, Schafmayer C, Tepel J, Schreiber S, Volzke H, Chang-Claude J, Hoffmeister M, Brenner H, Zanke BW, Montpetit A, Hudson TJ, Gallinger S, Campbell H, Dunlop MG. 2008. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet 40:1426–1435.

Jaeger E, Webb E, Howarth K, Carvajal-Carmona L, Rowan A, Broderick P, Walther A, Spain S, Pittman A, Kemp Z, Sullivan K, Heinimann K, Lubbe S, Domingo E, Barclay E, Martin L, Gorman M, Chandler I, Vijayakrishnan J, Wood W, Papaemmanuil E, Penegar S, Qureshi M, Farrington S, Tenesa A, Cazier JB, Kerr D, Gray R, Peto J, Dunlop M, Campbell H, Thomas H, Houlston R, Tomlinson I. 2008. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet 40:26–28.

Jenkins MA, Baglietto L, Dite GS, Jolley DJ, Southey MC, Whitty J, Mead LJ, St John DJ, Macrae FA, Bishop DT, Venter DJ, Giles GG, Hopper JL. 2002. After hMSH2 and hMLH1—What next? Analysis of three-generational, population-based, early-onset colorectal cancer families. Int J Cancer 102:166–171.

Jones AM, Thirlwell C, Howarth KM, Graham T, Chambers W, Segditsas S, Page KM, Phillips RK, Thomas HJ, Sieber OM, Sawyer EJ, Tomlinson IP. 2007. Analysis of copy number changes suggests chromosomal instability in a minority of large colorectal adenomas. J Pathol 213:249–256.

Jonsson G, Naylor TL, Vallon-Christersson J, Staaf J, Huang J, Ward MR, Greshock JD, Luts L, Olsson H, Rahman N, Stratton M, Ringner M, Borg A, Weber BL. 2005. Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization. Cancer Res 65:7612–7621.

Joosse SA, van Beers EH, Tielen IH, Horlings H, Peterse JL, Hoogerbrugge N, Ligtenberg MJ, Wessels LF, Axwijk P, Verhoef S, Hogervorst FB, Nederlof PM. 2008. Prediction of BRCA1-association in hereditary non-BRCA1/2 breast carcinomas with array-CGH. Breast Cancer Res Treat 116:479–489.

Kemp Z, Carvajal-Carmona L, Spain S, Barclay E, Gorman M, Martin L, Jaeger E, Brooks N, Bishop DT, Thomas H, Tomlinson I, Papaemmanuil E, Webb E, Sellick GS, Wood W, Evans G, Lucassen A, Maher ER, Houlston RS. 2006a. Evidence for a colorectal cancer susceptibility locus on chromosome 3q21-q24 from a high-density SNP genome-wide linkage scan. Hum Mol Genet 15:2903–2910.

Kemp ZE, Carvajal-Carmona LG, Barclay E, Gorman M, Martin L, Wood W, Rowan A, Donohue C, Spain S, Jaeger E, Evans DG, Maher ER, Bishop T, Thomas H, Houlston R, Tomlinson I. 2006b. Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United kingdom. Cancer Res 66:5003–5006.

Kinzler KW, Nilbert MC, Su LK, Vogelstein B, Bryan TM, Levy DB, Smith KJ, Preisinger AC, Hedge P, McKechnie D. 1991. Identification of FAP locus genes from chromosome 5q21. Science 253:661–665.

Lange K, Cantor R, Horvath S, Perola M, Sabatti C, Sinsheimer J, Sobel E. 2001. Mendel version 4.0: A complete package for the exact genetic analysis of discrete traits in pedigree and population data sets. Am J Hum Genet 69(Suppl):504.

Leppert M, Dobbs M, Scambler P, O'Connell P, Nakamura Y, Stauffer D, Woodward S, Burt R, Hughes J, Gardner E. 1987. The gene for familial polyposis coli maps to the long arm of chromosome 5. Science 238:1411–1413.

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

Lips EH, de Graaf EJ, Tollenaar RA, van Eijk R, Oosting J, Szuhai K, Karsten T, Nanya Y, Ogawa S, van de Velde CJ, Eilers PH, van Wezel T, Morreau H. 2007. Single nucleotide polymorphism array analysis of chromosomal instability patterns discriminates rectal adenomas from carcinomas. J Pathol 212:269–277.

Mangold E, Pagenstecher C, Friedl W, Mathiak M, Buettner R, Engel C, Loeffler M, Holinski-Feder E, Muller-Koch Y, Keller G, Schackert HK, Kruger S, Goecke T, Moeslein G, Kloor M, Gebert J, Kunstmann E, Schulmann K, Ruschoff J, Propping P. 2005. Spectrum and frequencies of mutations in MSH2 and MLH1 identified in 1,721 German families suspected of hereditary nonpolyposis colorectal cancer. Int J Cancer 116:692–702.

Middeldorp A, Jagmohan-Changur S, Helmer Q, van der Klift HM, Tops CM, Vasen HF, Devilee P, Morreau H, Houwing-Duistermaat JJ, Wijnen JT, van Wezel T. 2007. A procedure for the detection of linkage with high density SNP arrays in a large pedigree with colorectal cancer. BMC Cancer 7:6.

Middeldorp A, van Puijenbroek M, Nielsen M, Corver W, Jordanova E, Ter Haar N, Tops C, Vasen H, Lips E, van Eijk R, Hes F, Oosting J, Wijnen J, van Wezel T, Morreau H. 2008. High frequency of copy-neutral LOH in MUTYH-associated polyposis carcinomas. J Pathol 216:25–31.

Middeldorp A, Jagmohan-Changur S, van Eijk R, Tops C, Devilee P, Vasen HF, Hes FJ, Houlston R, Tomlinson I, Houwing-Duistermaat JJ, Wijnen JT, Morreau H, van Wezel T. 2009. Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. Cancer Epidemiol Biomark Prev 18:3062–3067.

Neklason DW, Kerber RA, Nilson DB, Anton-Culver H, Schwartz AG, Griffin CA, Lowery JT, Schildkraut JM, Evans JP, Tomlinson GE, Strong LC, Miller AR, Stopfer JE, Finkelstein DM, Nadkarni PM, Kasten CH, Mineau GP, Burt RW. 2008. Common familial colorectal cancer linked to chromosome 7q31: A genome-wide analysis. Cancer Res 68:8993–8997.

Oosting J, Lips EH, van Eijk R, Eilers PH, Szuhai K, Wijmenga C, Morreau H, van Wezel T. 2007. High-resolution copy number analysis of paraffin-embedded archival tissue using SNP BeadArrays. Genome Res 17:368–376.

Papaemmanuil E, Carvajal-Carmona L, Sellick GS, Kemp Z, Webb E, Spain S, Sullivan K, Barclay E, Lubbe S, Jaeger E, Vijayakrishnan J, Broderick P, Gorman M, Martin L, Lucassen A, Bishop DT, Evans DG, Maher ER, Steinke V, Rahner N, Schackert HK, Goecke TO, Holinski-Feder E, Propping P, van Wezel T, Wijnen J, Cazier JB, Thomas H, Houlston RS, Tomlinson I. 2008. Deciphering the genetics of hereditary non-syndromic colorectal cancer. Eur J Hum Genet 16:1477–1486.

Parkin DM, Bray F, Ferlay J, Pisani P. 2005. Global cancer statistics, 2002. CA Cancer J Clin 55:74–108.

Picelli S, Vandrovcova J, Jones S, Djureinovic T, Skoglund J, Zhou XL, Velculescu VE, Vogelstein B, Lindblom A. 2008. Genome-wide linkage scan for colorectal cancer susceptibility genes supports linkage to chromosome 3q. BMC Cancer 8:87.

Skoglund J, Djureinovic T, Zhou XL, Vandrovcova J, Renkonen E, Iselius L, Bisgaard ML, Peltomaki P, Lindblom A. 2006. Linkage analysis in a large Swedish family supports the presence of a susceptibility locus for adenoma and colorectal cancer on chromosome 9q22.32–31.1. J Med Genet 43:e7.

Sobel E, Lange K. 1996. Descent graphs in pedigree analysis: Applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 58:1323–1337.

Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, Barnetson RA, Theodoratou E, Cetnarskyj R, Cartwright N, Semple C, Clark AJ, Reid FJ, Smith LA, Kavoussanakis K, Koessler T, Pharoah PD, Buch S, Schafmayer C, Tepel J, Schreiber S, Volzke H, Schmidt CO, Hampe J, Chang-Claude J, Hoffmeister M, Brenner H, Wilkening S, Canzian F, Capella G, Moreno V, Deary IJ, Starr JM, Tomlinson IP, Kemp Z, Howarth K, Carvajal-Carmona L, Webb E, Broderick P, Vijayakrishnan J, Houlston RS, Rennert G, Ballinger D, Rozek L, Gruber SB, Matsuda K, Kidokoro T, Nakamura Y, Zanke BW, Greenwood CM, Rangrej J, Kustra R, Montpetit A, Hudson TJ, Gallinger S, Campbell H, Dunlop MG. 2008. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet 40:631–637.

Thornton T, McPeek MS. 2007. Case-control association testing with related individuals: A more powerful quasi-likelihood score test. Am J Hum Genet 81:321–337.

Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, Martin L, Sellick G, Jaeger E, Hubner R, Wild R, Rowan A, Fielding S, Howarth K, Silver A, Atkin W, Muir K, Logan R, Kerr D, Johnstone E, Sieber O, Gray R,

Thomas H, Peto J, Cazier JB, Houlston R. 2007. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 39:984–988.

Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, Spain S, Lubbe S, Walther A, Sullivan K, Jaeger E, Fielding S, Rowan A, Vijayakrishnan J, Domingo E, Chandler I, Kemp Z, Qureshi M, Farrington SM, Tenesa A, Prendergast JG, Barnetson RA, Penegar S, Barclay E, Wood W, Martin L, Gorman M, Thomas H, Peto J, Bishop DT, Gray R, Maher ER, Lucassen A, Kerr D, Evans DG, Schafmayer C, Buch S, Volzke H, Hampe J, Schreiber S, John U, Koessler T, Pharoah P, van Wezel T, Morreau H, Wijnen JT, Hopper JL, Southey MC, Giles GG, Severi G, Castellvi-Bel S, Ruiz-Ponte C, Carracedo A, Castells A, Forsti A, Hemminki K, Vodicka P, Naccarati A, Lipton L, Ho JW, Cheng KK, Sham PC, Luk J, Agundez JA, Ladero JM, de la HM, Caldes T, Niittymaki I, Tuupanen S, Karhu A, Aaltonen L, Cazier JB, Campbell H, Dunlop MG, Houlston RS. 2008. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet 40:623–630.

Uh HW, van der Wijk HJ, Houwing-Duistermaat JJ. 2009. Testing for genetic association taking into account phenotypic information of relatives. BMC Proc 3(Suppl 7):S123.

van Puijenbroek M, Middeldorp A, Tops CM, van Eijk R, van der Klift HM, Vasen HF, Wijnen JT, Hes FJ, Oosting J, van Wezel T, Morreau H. 2008. Genome-wide copy neutral LOH is infrequent in familial and sporadic microsatellite unstable carcinomas. Fam Cancer 7:319–330.

Vasen HF, Watson P, Mecklin JP, Lynch HT. 1999. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology 116:1453–1456.

Webb EL, Sellick GS, Houlston RS. 2005. SNPLINK: Multipoint linkage analysis of densely distributed SNP data incorporating automated linkage disequilibrium removal. Bioinformatics 21: 3060–3061.

Wessels LF, van Welsem T, Hart AA, Van't Veer LJ, Reinders MJ, Nederlof PM. 2002. Molecular classification of breast carcinomas by comparative genomic hybridization: A specific somatic genetic profile for BRCA1 tumors. Cancer Res 62:7110–7117.

Wiesner GL, Daley D, Lewis S, Ticknor C, Platzer P, Lutterbaugh J, MacMillen M, Baliner B, Willis J, Elston RC, Markowitz SD. 2003. A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2–31.2. Proc Natl Acad Sci USA 100:12961–12965.

Wijnen JT, Vasen HF, Khan PM, Zwinderman AH, van der KH, Mulder A, Tops C, Moller P, Fodde R. 1998. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. N Engl J Med 339:511–518.

Wijnen JT, Brohet RM, van Eijk R, Jagmohan-Changur S, Middeldorp A, Tops CM, van Puijenbroek M, Ausems MG, Gomez GE, Hes FJ, Hoogerbrugge N, Menko FH, van Os TA, Sijmons RH, Verhoef S, Wagner A, Nagengast FM, Kleibeuker JH, Devilee P, Morreau H, Goldgar D, Tomlinson IP, Houlston RS, van Wezel T, Vasen HF. 2009. Chromosome 8q23.3 and 11q23.1 variants modify colorectal cancer risk in Lynch syndrome. Gastroenterology 136:131–137.

Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E, Ferretti V, Laflamme P, Sundararajan S, Roumy S, Olivier JF, Robidoux F, Sladek R, Montpetit A, Campbell P, Bezieau S, O'Shea AM, Zogopoulos G, Cotterchio M, Newcomb P, McLaughlin J, Younghusband B, Green R, Green J, Porteous ME, Campbell H, Blanche H, Sahbatou M, Tubacher E, Bonaiti-Pellie C, Buecher B, Riboli E, Kury S, Chanock SJ, Potter J, Thomas G, Gallinger S, Hudson TJ, Dunlop MG. 2007. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet 39:989–994.

Chapter 4

Enrichment of Low Penetrance Susceptibility Loci in a Dutch Familial Colorectal Cancer Cohort

Cancer, Epidemiology, Biomarkers & Prevention (2009)

18(11):3062-3067

Enrichment of Low Penetrance Susceptibility Loci in a Dutch Familial Colorectal Cancer Cohort

Anneke Middeldorp,[1] Shantie Jagmohan-Changur,[2] Ronald van Eijk,[1] Carli Tops,[3] Peter Devilee,[2] Hans F. A. Vasen,[5] Frederik J. Hes,[3] Richard Houlston,[6] Ian Tomlinson,[7] Jeanine J. Houwing-Duistermaat,[4] Juul T. Wijnen,[2,3] Hans Morreau,[1] and Tom van Wezel[1]

Departments of [1]Pathology, [2]Human Genetics, [3]Clinical Genetics, and [4]Medical Statistics and Bioinformatics, Leiden University Medical Centre; and [5]The Netherlands Foundation for the Detection of Hereditary Tumours, Leiden, the Netherlands; [6]Section of Cancer Genetics, Institute of Cancer Research, Sutton, United Kingdom; and [7]Molecular and Population Genetics, Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom

Abstract

Recent genome-wide association studies have identified several loci that confer an increased risk of colorectal cancer (CRC). We studied the role of the 8q24.21 (rs6983267), 18q21.1 (rs12953717), 15q13.3 (rs4779584), 11q23.1 (rs3802842), 8q23.3 (rs16892766), and 10p14 (rs10795668) risk variants in a series of 995 Dutch CRC cases and 1340 controls. The CRC cases were selected on basis of having a family history of CRC and/or early-onset disease. The detailed clinical and molecular data available on the cases allowed us to examine the relationship between risk variants and clinicopathologic characteristics. We replicated the association with an increased risk of CRC cancer for all loci, except 10p14. The association with the variant on chromosome 15q13.3 was confirmed for the first time. The risks associated with variants in our series were higher (not significant) than those previously reported, consistent with our series reflecting genetic enrichment. Moreover, we show that familial CRC cases possess an increased number of risk alleles compared with solitary CRC cases (early-onset; mean age at diagnosis of 48.5 years). We also identified a significant increase in the number of risk alleles in families with early-onset disease (≤50 years) compared with late-onset families (>50 years). In solitary CRC patients, enrichment for risk alleles was not observed, suggesting that other causes of increased CRC risk play a role in these cases. Overall, our results suggest that clustering of low-risk variants may explain part of the excess risk in CRC families. (Cancer Epidemiol Biomarkers Prev 2009;18(11):3062–7)

Introduction

Around a third of all colorectal cancer (CRC) has been shown to be attributable to heritable factors (1). High Penetrance mutations, such as those in the mismatch repair genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) causing the Lynch syndrome, *APC* in Familial Adenomatous Polyposis, and *MUTYH* in *MUTYH*-associated polyposis, only account for ~5% of CRC (2-4). Although the underlying basis of the residual excess familial risk is presently undefined, it is likely that coinheritance of several common alleles each conferring a low CRC risk contribute to this excess familial risk.

Recent genome-wide association studies have vindicated this hypothesis and several loci have been robustly shown to be associated with an increased risk of developing CRC. The first of these was a variant at 8q24.21 defined by the single nucleotide polymorphism (SNP) rs6983267 (5-8). This variant was also shown to be associated with increased risks of both prostate and ovarian cancer (9, 10). Recently, this locus has also been associated with the development of multiple colorectal adenomas (11). Three SNPs on 18q21.1 were reported to be associated with an increased CRC risk. These three variants (rs4939827, rs12953717, and rs4464148) map to an intronic sequence of *SMAD7*, an antagonist of transforming growth factor-β signaling (12, 13). A third locus associated with an increased CRC risk was identified on chromosome 15q13.3 (14). Previously, linkage of this locus to CRC was reported in three Ashkenazi families (15, 16). The strongest association for this locus was reported with rs4779584, close to the genes *SCG5*, *GREM1*, and *FMN1*. A fourth association was identified in a gene-rich region on 11q23 with SNP rs3802842 (13, 17). Furthermore, SNP rs10795668 (located at 10p14) and SNP rs16892766 (located at 8q23.3) were reported to be associated with an increased risk of CRC (18). This latter variant at 8q23.3 is linked to *EIF3H*, a translation initiation factor. Current data suggest that each of these loci act independently in a dose-dependent manner with those individuals possessing multiple risk variants having relatively substantive risks over those carrying few risk alleles (17).

Association of these six risk loci with several clinical or pathologic parameters has been described. For rs3802842 (11q23) and rs4939827 (18q21.1), the risk of developing rectal cancer was found to be greater than the risk of

Received 6/19/09; revised 9/7/09; accepted 9/11/09; published OnlineFirst 10/20/09.

Grant support: Dutch Cancer Society (grant number UL2005-3247) and the NutsOhra Foundation.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Requests for reprints: Tom van Wezel, Department of Pathology, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, the Netherlands. Phone: 31715266813; Fax: 31848331048. E-mail: T.van_Wezel@lumc.nl

doi:10.1158/1055-9965.EPI-09-0601

Table 1. Scoring system for family characteristics

Category	Description
I	Amsterdam Criteria II positive families*
II	At least two FDR affected with CRC, at least one diagnosed ≤50 y in one or two generations
III	Solitary CRC patient
IV	At least three FDR affected with CRC, diagnosed >50 y in one or two generations
V	Two FDR affected with CRC, diagnosed >50 y in one or two generations

Abbreviation: FDR, first-degree relative.
*Including largely non-Lynch or so-called Lindor Type X families (29).

developing colonic cancer (13). Moreover, Tomlinson et al. (18) reported that the association with rs10795668 (10p14) was stronger for rectal cancer than for colonic tumors. The association of the locus at 8q24.21 was described to be stronger for patients under the age of 50 years than for those 50 years of age or older (8). Similarly, the effect of rs16892766 (8q23.3) on CRC risk has also been described to be stronger in younger individuals (<60 years; ref. 18). Finally, Tuupanen et al. (19) identified a tendency for association of rs6983267 (8q24.21) with microsatellite stable (MSS) cancer and a family history of extracolonic cancers.

The effect of the variants at these six loci has also been studied in Lynch syndrome patients, carrying a mutation in one of the mismatch repair genes. In Lynch patients, rs16892766 (8q23.3) is associated with an elevated CRC risk. For rs3802842 (11q23.1), an elevated risk of developing CRC was described for female Lynch patients only (20).

To further inform on the relationship of variants at these six loci with the development of CRC, we have genotyped 8q24.21, 18q21.1, 15q13.3, 8q23.3, 11q23.1, and 10p14 variants in a cohort of Dutch CRC cases, enriched for a positive family history and/or an early onset of disease. Detailed clinical and pathologic data on the cases has allowed us to examine the relationship between clinicopathology and genotype in the setting of familial disease.

Materials and Methods

Study Population. We studied 995 index cases, all of which were diagnosed with CRC (48% male, 52% female), and 1340 controls (47% male, 53% female). These cases fulfill the clinical criteria for microsatellite instability (MSI) testing, installed in the Netherlands since 1997 and formalized in 2008, based on early onset of the disease and/or familial clustering of CRC (Supplementary Table S1). Most cases were sent in through the clinical genetics department. Only one case per family was included in our cohort. Samples were collected mostly in a period from 1997 to 2007. The samples largely originate from the southwestern part of the Netherlands. Eighty-two percent of all tested tumors were MSS, whereas 7% had a low level of MSI (MSI-L), and 12% were MSI-H. Of the MSI-H tumors, 93% were MSI-H on a sporadic basis. Overall, the cohort includes 10 Lynch syndrome patients.

Controls were 894 healthy blood donors from the southwest region of the Netherlands. Additional 446 controls were derived from individuals that presented at the Laboratory for Diagnostic Genome Analysis at the Leiden University Medical Centre for presymptomatic testing for noncancer syndromes and tested mutation-negative for these syndromes. Age was not known for all controls. Therefore, the control cohort might include some young individuals that develop CRC later in life. This would reduce the power of our analysis and therefore could have resulted in an underestimation of the associations.

We genotyped the following SNPs in the CRC cases and the controls: rs16892766, rs6983267, rs10795668, rs3802842, rs4779584, and rs12953717.

The study was approved by the local Medical Ethical Committee (protocol P01.019). Samples were handled according to the medical ethical guidelines described in the Code Proper Secondary Use of Human Tissue established by the Dutch Federation of Medical Sciences.[8]

DNA Isolation. DNA was extracted from either peripheral blood or from formalin-fixed paraffin embedded (FFPE) normal tissue. DNA was extracted from blood samples using an automated procedure (Gentra Systems). DNA from FFPE tissue was isolated using xylol, ethanol, and overnight incubation at 56°C in 120 μL of PK1 lysis buffer and 5 μL Proteinase K. The suspension was then incubated for 10 min at 100°C and centrifuged for 10 min at 13,000 rpm. The supernatant containing the DNA was carefully transferred to a new tube. DNA was quantified using the picogreen method (Invitrogen-Molecular Probes).

MSI Analysis. Eight microsatellite markers were evaluated to determine the MSI status of the tumors. As recommended by the National Cancer Institute Workshop on MSI for Cancer Detection and Familial Predisposition, two mononucleotide repeats (BAT25 and BAT26) and three dinucleotide repeats (D2S123, D5S346, and D17S250) were analyzed, supplemented with three mononucleotide repeat markers (BAT40, MSH3, and MSH6; ref. 21). Tumors were classified as MSS when no instability was seen for any of the eight markers. When instability was seen in <30% of the markers, tumors were classified as MSI-L. When >30% of the markers showed instability, the tumors were classified as MSI-H.

Genotyping. Genotyping of the variants rs12953717, rs3802842, rs6983267, and rs16892766 was done using the KASPar method following the manufacturer's protocol (KBioscience). PCR, with two allele-specific forward primers and one common reverse primer, was done in a GeneAmp PCR system 9700 (Applied Biosystems). Detection was done in an ABI PRISM 7900HT (Applied Biosystems).

Genotyping of rs10795668 was done using a TaqMan SNP Genotyping Assay for samples that were derived from FFPE tissue according to the manufacturer's protocol (Applied Biosystems). The leukocyte DNA samples were genotyped using high-resolution melting curve analysis. Primers for high-resolution melting curve analysis were designed using Primer3. In short, 5 μL PCR reactions were analyzed on a LightCycler (Roche) using the LightCycler software (version 1.5.0, Roche). PCR reactions consisted of 10 ng DNA, iQ Supermix (Bio-Rad), 1 μmol/L SYTO9 (Invitrogen), and 2 pmol primers.

For rs4779584, all leukocyte DNA was genotyped using the KASPar method, whereas the high-resolution melting

[8] http://www.federa.org

Table 2. Characteristics of the study population

	Cases	Controls
Gender		
Male	453 (48%)	633 (47%)
Female	492 (52%)	707 (53%)
Age at diagnosis		
≤50 y	291 (42%)	
>50 y	395 (58%)	
MSI status		
MSI-L/MSI-H	134 (18%)	
MSS	596 (82%)	
Tumor location		
Right sided	218 (33%)	
Left sided	438 (67%)	
Tumor stage		
A	27 (5%)	
B	301 (51%)	
C	204 (35%)	
D	55 (9%)	
Polyps		
No	245 (32%)	
Yes	474 (63%)	
Unknown	38 (5%)	
Family characteristics		
Category I	66 (13%)	
Category II	105 (21%)	
Category III	144 (28%)	
Category IV	102 (20%)	
Category V	66 (13%)	
Other	23 (5%)	

curve analysis method was applied to the samples isolated from FFPE tissue. Samples that failed to produce a genotype were reanalyzed using the TaqMan SNP genotyping assay (Applied Biosystems). Primer details are available upon request.

To check the quality of the genotyping, >5% of the samples were analyzed in duplicate. The concordance was >98%. All SNP genotype frequencies in our cohort fit the Hardy-Weinberg equilibrium as assessed using a χ^2 test.

To facilitate data management and analysis, we developed a relational database that is based on MS Access and structured query language.

Clinicopathologic Characteristics. For all cases, information was available for several clinical genetic and pathologic parameters. This information included age at diagnosis, microsatellite status, tumor grade, location of the tumor (left of the flexura lienalis versus right of the flexura lienalis), additional development of adenomas, and family history characteristics. Information on the presence of adenomas was derived from Pathological Anatomical District Automated Archives, a nationwide network and registry of histopathology and cytopathology in the Netherlands (22). The latter was scored with the scheme displayed in Table 1.

Statistical Analysis. Genotype frequencies in the CRC cases and controls were tested for deviations from expected frequencies under Hardy-Weinberg Equilibrium using a χ^2 test using a significance level of 5% (one degree of freedom). Odds ratios (OR), including their 95% confidence intervals (CI), were calculated using logistic regression analysis. Bonferroni's correction for multiple testing was applied. Associations between CRC risk loci and clinical and pathologic parameters were assessed using a χ^2 test (one degree of freedom) in CRC cases only.

Power calculation was done under the assumption of an additive model and an α of 0.05.

We used a Cochran-Armitage test of trend using ordered categorical data to study the effect of possessing an increased number of risk alleles (counting one for heterozygotes and two for homozygotes). The most common number of risk alleles was used as a reference to calculate ORs. These statistical analyses were done in SPSS 16.0.

Pairwise interactions between the SNPs were studied using Plink (v1.05; ref. 23).

Results and Discussion

We studied the effects of rs16892766 (8q23.3), rs6983267 (8q24.21), rs10795668 (10p14), rs3802842 (11q23.1), rs4779584 (15q13.3), and rs12953717 (18q21.1) in a Dutch familial CRC cohort. This cohort is enriched for a positive family history for CRC and/or early onset of disease. The main characteristics of the study cohort are detailed in Table 2.

Association Analysis. Variant rs6983267 (8q24.21) was significantly associated with an increased risk of CRC in our familial cohort, with an allelic OR of 1.29 (95% CI, 1.14-1.49; $P = 7.2 \times 10^{-5}$; Table 3). We also replicated the association of variant rs12953717 (18q21.1) with CRC risk in our cohort, with an allelic OR of 1.23 (95% CI, 1.09-1.38; $P = 8.6 \times 10^{-4}$). Curtin et al. (24) also described associations

Table 3. Association between CRC risk and low-risk variants

SNP	Chromosome region	Minor allele*	Allele frequencies†		OR (95% CI) Allelic	*P*	GWAS OR (95% CI) Allelic	Power
			Cases	Controls				
rs6983267	8q24.21	G	0.58	0.52	1.29 (1.14-1.46)	7.2×10^{-5}‡	1.24 (1.17-1.33)§	95%
rs12953717	18q21.1	T	0.45	0.40	1.23 (1.09-1.39)	8.6×10^{-4}‡	1.17 (1.12-1.22)‖	75%
rs4779584	15q13.3	T	0.24	0.18	1.45 (1.24-1.69)	3.8×10^{-6}‡	1.23 (1.14-1.34)§	81%
rs3802842	11q23.1	C	0.31	0.27	1.26 (1.11-1.43)	4.4×10^{-4}‡	1.11 (1.08-1.15)¶	38%
rs10795668	10p14	A	0.31	0.31	1.02 (0.88-1.17)	0.84	0.89 (0.86-0.91)§	45%
rs16892766	8q23.3	C	0.10	0.09	1.23 (1.00-1.50)	0.05	1.25 (1.19-1.32)§	53%

NOTE: GWAS OR, OR identified in previous genome-wide association studies. Power was calculated assuming an additive model and an α of 0.05.
*According to the HapMap CEU frequencies.
†In our cohort.
‡Associations remained significant after correction for multiple testing.
§Tomlinson et al., 2008.
‖Jaeger et al., 2008.
¶Tenesa et al., 2008.

Table 4. ORs for cumulative number of risk alleles

No. risk alleles	No. cases (%)	No. controls (%)	OR (95%CI)	*P*
0	13 (1.8)	18 (2.9)	0.65 (0.31-1.36)	0.25
1	47 (6.7)	67 (10.7)	0.63 (0.41-0.96)	0.03
2	135 (19.2)	157 (25.0)	0.77 (0.57-1.04)	0.09
3	197 (28.0)	176 (28.0)	1.00 (reference)	
4	162 (23.0)	132 (21.0)	1.10 (0.81-1.49)	0.56
5	102 (14.5)	60 (9.6)	1.52 (1.04-2.22)	0.03
≥6	47 (6.7)	18 (2.9)	2.33 (1.31-4.17)	4.2×10^{-3}
Total	703 (100)	628 (100)		$P_{Trend} = 1.1 \times 10^{-7}$

NOTE: The results shown in this table are illustrated in Fig. 1.

between the risk alleles on 8q24.21 and 18q21.1 and CRC risk in a cohort that included familial CRC. For rs4779584 (15q13.3), a strong association with CRC risk was identified, with an allelic OR of 1.45 (95% CI, 1.22-1.67; $P = 3.8 \times 10^{-6}$). To the best of our knowledge, these results provide the first replication of the association between the locus at 15q13.3 and CRC risk. A strong association was also identified for rs3802842 on chromosome 11q23.1 (allelic OR, 1.26; 95% CI, 1.11-1.44; $P = 4.4 \times 10^{-4}$). The results for rs3802842 have previously been reported for 783 of 995 samples (17). An allelic OR of 1.23 for rs16892766 (8q23.3) was found in our cohort (95% CI, 1.00-1.51; $P = 0.05$). This association does not remain significant after correction for multiple testing. The results for rs16892766 (8q23.3) have previously been reported for 783 of 953 of the samples (18). Overall, the ORs identified in our cohort tend to be increased compared with the ORs described in the initial genome-wide association studies, consistent with our series reflecting genetic enrichment (Table 3). Although the ORs are consistently increased, the CIs overlap, indicating that the differences are not statistically significant.

Unlike the previous loci, we were unable to replicate the association between rs10795668 (10p14) and CRC risk. In our cohort, the locus on 10p14 was not associated with an increased risk of CRC (allelic OR, 1.02; 95% CI, 0.88-1.17; $P = 0.84$). The absence of an association between this locus and CRC risk may be explained by a lack of power to detect the association in our study, although the power to detect association with a minor allele frequency of 0.33 and an OR of 0.89 was 45% in our study, and association with rs3802842 (11q23.1) was detected with a prior power of 38%. A second explanation could be a difference between the Dutch population compared with the English population. Of note, this locus on chromosome 10p14 was also not captured by meta-analysis of two large genome-wide association studies (25).

Clinicopathologic Characteristics. We studied the association of the six risk loci mentioned above with several clinicopathologic parameters including the following: gender, age at diagnosis, polyp development, tumor stage, family characteristics (Table 1), tumor location, and MSI status. In our cohort, the association between rs12953717 (18q21.1) and CRC risk is stronger in cases with left-sided cancer (OR, 1.43; 95% CI, 1.22-1.68) compared with cases with right-sided cancer, where the association could not be detected (OR, 1.02; 95% CI, 0.82-1.27; $P = 0.03$). Further analysis shows that the association of rs12953717 and CRC is, among the left-sided cancers, stronger for the tumors located at the rectosigmoid junction or in the rectum (OR, 1.58; 95% CI, 1.28-1.95) compared with left-sided tumors proximal of the rectosigmoid junction (OR, 1.30; 95% CI, 1.04-1.61; $P = 0.04$). Similarly, a stronger effect of rs12953717 on CRC risk was seen for familial CRC cases with at least two first-degree relatives affected with CRC (OR, 1.51; 95% CI, 1.25-1.83) compared with solitary CRC cases (early-onset; mean age at diagnosis of 48.5 years; OR, 0.86; 95% CI, 0.65-1.14; $P = 8.6 \times 10^{-4}$). No relationship was identified between tumor location and family characteristics ($P = 0.16$), indicating that these two parameters are independently associated with rs12953717. No statistically significant associations between rs12953717 (18q21.1) and the other clinicopathologic parameters (including gender, age at diagnosis, polyp development, tumor stage, and MSI status) were found. For all other loci, no significant associations were found between the variants and gender, age at diagnosis, tumor location, family characteristics, polyp development, tumor stage, or MSI status. However, ORs for rs12953717 (18q21.1),

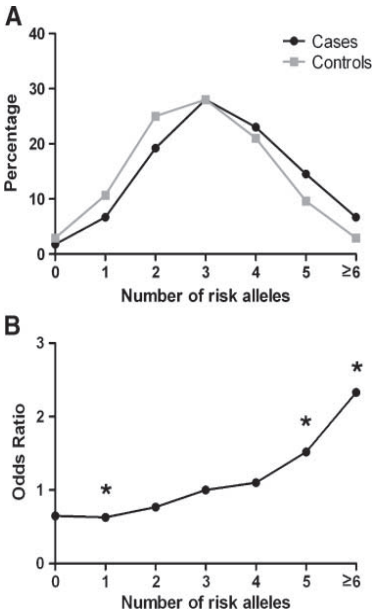

Figure 1. Number of risk alleles in the CRC cases and the controls. **A.** The distribution of the number of risk alleles in the CRC cases and the controls. **B.** The OR for each category. The most common number of risk alleles (three) was set as a reference. *, $P < 0.05$.

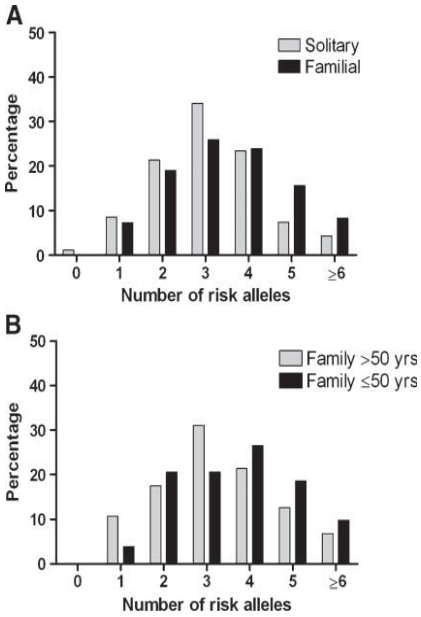

Figure 2. Relationship between number of risk alleles and family characteristics. **A.** The distribution of the number of risk alleles per individual in solitary cases and familial cases (with at least two first-degree relatives affected with CRC). **B.** The distribution of risk alleles in families with an age at diagnosis 50 y or younger and families with an age at diagnosis older than 50 y of age. In both graphs, the distributions differ significantly from each other (**A**, $P_{Trend} = 0.03$; **B**, $P_{Trend} = 0.04$).

rs4779584 (15q13.3), and rs16892766 (8q23.3) were increased in MSS cases compared with MSI-H cases, although these differences were not significant. These results suggest that the role of low-risk variants play in MMR-deficient tumors is reduced compared with MMR-proficient tumors, despite the sporadic nature of the MSI-H cases in our cohort. And although rs16892766 (8q23.3) is significantly associated with an increased CRC risk in MMR mutation carriers (20), its role in MMR-deficient tumors with a sporadic nature seems reduced. Several associations between the risk loci and clinicopathologic parameters have previously been described, including associations between rectal cancer and 11q23.1 and 18q21.1, between early disease onset and both 8q24.21 and 8q23.3, and between MSS tumors and 8q24.21 (8, 13, 18, 19). However, we were unable to replicate any of these associations in our cohort.

Risk Allele Distribution. We investigated the effect of possessing an increased number of risk alleles on CRC risk (Table 4). We counted the number of risk alleles per individual and then compared the distribution in CRC cases versus controls. We identified a significant increase in the number of risk alleles in CRC cases compared with controls ($P_{Trend} = 1.1 \times 10^{-7}$). Moreover, there was a gradual increase in the OR with an increased number of risk alleles, although not all increases in the OR were significant. For individuals possessing six or more risk alleles, an OR of 2.3 (95% CI, 1.31-4.17) was observed. These results are illustrated in Fig. 1. Our findings are in line with previous reports on the cumulative effect of these risk alleles on CRC risk (17). Similar cumulative effects have been described for other cancers, including prostate cancer and head and neck cancer (26-28).

Remarkably, a significant difference in the number of risk alleles was observed between solitary cases (early-onset; mean age at diagnosis of 48.5 years) and familial CRC cases (at least two affected first-degree relatives). The familial CRC cases had significantly more risk alleles compared with solitary CRC cases ($P_{Trend} = 0.03$), suggesting that low-risk variants indeed cluster in families affected with CRC (Fig. 2). In addition, the latter finding suggests that other genetic models (such as a recessive origin) might play a role in solitary CRC cases. Additionally, we found that families with an early onset of the disease (≤50 years of age) had significantly more risk alleles ($P_{Trend} = 0.04$) than families with a late onset of the disease (>50 years of age). These results suggest that possessing an increasing number of risk alleles decreases the age of onset of the disease. However, similar relationship between the number of risk alleles and disease onset was not seen at the individual level.

Analysis of pairwise interactions did not yield evidence for an interaction between any of the risk alleles. These results are consistent with the analyses in a recent meta-analysis, where all comparisons were based on at least 13,000 individuals (25).

In conclusion, we replicated the association between CRC risk and loci on chromosomes 8q23.3, 8q24.21, 11q23.1, 15q13.3, and 18q21.1. This is the first study to replicate the association between CRC risk and the locus on 15q13.3, although the association with a locus on chromosome 10p14 could not be replicated in our cohort. The ORs for rs6983267 (8q24.21), rs12953717 (18q21.1), rs4779584 (15q13.3), and rs3802842 (11q23.1) tend to be higher than those seen in the genome-wide analysis, possibly reflecting enrichment for positive family history of CRC in our cohort. Moreover, we saw enrichment in the number of risk alleles in patients with at least two first-degree family members affected with CRC compared with solitary CRC patients. Similar enrichment was identified in families with early-onset disease (≤50 years of age) compared with families with late-onset of CRC (>50 years of age). Overall, our results suggest clustering of low-risk variants in familial CRC, which is likely to contribute to the observed excess risk in relatives of patients.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Katja Philippo for sample collection and genotyping, the Molecular Diagnostics group (Department of Pathology, Leiden University Medical Centre) and Petra Huijts for sample collection, and Tanaka Moyo for his assistance in data management.

References

1. Lichtenstein P, Holm NV, Verkasalo PK, et al. 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.
2. Jarvinen HJ. Epidemiology of familial adenomatous polyposis in Finland: impact of family screening on the colorectal cancer rate and survival. Gut 1992;33:357–60.
3. Aaltonen LA, Salovaara R, Kristo P, et al. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. N Engl J Med 1998;338:1481–7.
4. Salovaara R, Loukola A, Kristo P, et al. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. J Clin Oncol 2000;18:2193–200.
5. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 2007;39:984–8.
6. Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet 2007;39:989–94.
7. Haiman CA, Le Marchand L, Yamamato J, et al. A common genetic risk factor for colorectal and prostate cancer. Nat Genet 2007;39:954–6.
8. Gruber SB, Moreno V, Rozek LS, et al. Genetic variation in 8q24 associated with risk of colorectal cancer. Cancer Biol Ther 2007;6:1143–7.
9. Haiman CA, Patterson N, Freedman ML, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet 2007; 39:638–44.
10. Ghoussaini M, Song H, Koessler T, et al. Multiple loci with different cancer specificities within the 8q24 gene desert. J Natl Cancer Inst 2008;100:962–6.
11. Berndt SI, Potter JD, Hazra A, et al. Pooled analysis of genetic variation at chromosome 8q24 and colorectal neoplasia risk. Hum Mol Genet 2008;17:2665–72.
12. Broderick P, Carvajal-Carmona L, Pittman AM, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet 2007;39:1315–7.
13. Tenesa A, Farrington SM, Prendergast JG, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet 2008;40:631–7.
14. Jaeger E, Webb E, Howarth K, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet 2008;40:26–8.
15. Jaeger EE, Woodford-Richens KL, Lockett M, et al. An ancestral Ashkenazi haplotype at the HMPS/CRAC1 locus on 15q13-14 is associated with hereditary mixed polyposis syndrome. Am J Hum Genet 2003;72:1261–7.
16. Tomlinson I, Rahman N, Frayling I, et al. Inherited susceptibility to colorectal adenomas and carcinomas: evidence for a new predisposition gene on 15q14-22. Gastroenterology 1999;116:789–95.
17. Pittman AM, Webb E, Carvajal-Carmona L, et al. Refinement of the basis and impact of common 11q23.1 variation to the risk of developing colorectal cancer. Hum Mol Genet 2008;17:3720–7.
18. Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet 2008;40:623–30.
19. Tuupanen S, Niittymaki I, Nousiainen K, et al. Allelic imbalance at rs6983267 suggests selection of the risk allele in somatic colorectal tumor evolution. Cancer Res 2008;68:14–7.
20. Wijnen JT, Brohet RM, van Eijk R, et al. Chromosome 8q23.3 and 11q23.1 variants modify colorectal cancer risk in Lynch syndrome. Gastroenterology 2009;136:131–7.
21. Boland CR, Thibodeau SN, Hamilton SR, et al. 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.
22. Bijlsma F, van der Esch EP. PALGA, of labor pains of the Pathological Anatomical District Automatized Archives. Ned Tijdschr Geneeskd 1972;116:902–3.
23. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.
24. Curtin K, Lin WY, George R, et al. Meta association of colorectal cancer confirms risk alleles at 8q24 and 18q21. Cancer Epidemiol Biomarkers Prev 2009;18:616–21.
25. Houlston RS, Webb E, Broderick P, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet 2008;40:1426–35.
26. Wu X, Spitz MR, Lee JJ, et al. Novel susceptibility loci for second primary tumors/recurrence in head and neck cancer patients: large-scale evaluation of genetic variants. Cancer Prev Res (Phila Pa) 2009;2:617–24.
27. Xu J, Sun J, Kader AK, et al. Estimation of absolute risk for prostate cancer using genetic markers and family history. Prostate 2009.
28. Zheng SL, Sun J, Wiklund F, et al. Cumulative association of five genetic variants with prostate cancer. N Engl J Med 2008;358:910–9.
29. Lindor NM, Rabe K, Petersen GM, et al. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. JAMA 2005;293:1979–85.

High Frequency of Copy-neutral LOH in MUTYH-associated polyposis carcinomas

Journal of Pathology (2008) 216(1):25-31

Chapter 5

High frequency of copy-neutral LOH in *MUTYH*-associated polyposis carcinomas

A Middeldorp,[1] M van Puijenbroek,[1] M Nielsen,[2] WE Corver,[1] ES Jordanova,[1] N ter Haar,[1] CMJ Tops,[2] HFA Vasen,[3] EH Lips,[1] R van Eijk,[1] FJ Hes,[2] J Oosting,[1] J Wijnen,[2,4] T van Wezel[1] and H Morreau[1]*

[1]*Department of Pathology, Leiden University Medical Centre, Leiden, The Netherlands*
[2]*Department of Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands*
[3]*The Netherlands Foundation for the Detection of Hereditary Tumours, Leiden, The Netherlands*
[4]*Department of Human Genetics, Leiden University Medical Centre, Leiden, The Netherlands*

**Correspondence to:*
H Morreau, Department of Pathology, Leiden University Medical Centre, Post Zone L1-Q, PO Box 9600, 2300 RC Leiden, The Netherlands.
E-mail: j.morreau@lumc.nl

No conflicts of interest were declared.

Received: 14 February 2008
Revised: 15 April 2008
Accepted: 15 April 2008

Abstract

Genetic instability is known to drive colorectal carcinogenesis. Generally, a distinction is made between two types of genetic instability: chromosomal instability (CIN) and microsatellite instability (MIN or MSI). Most CIN tumours are aneuploid, whereas MSI tumours are considered near-diploid. However, for *MUTYH*-associated polyposis (MAP) the genetic instability involved in the carcinogenesis remains unclear, as near-diploid adenomas, aneuploid adenomas and near-diploid carcinomas have been reported. Remarkably, our analysis of 26 MAP carcinomas, using SNP arrays and flow sorting, showed that these tumours are often near-diploid (52%) and mainly contain chromosomal regions of copy-neutral loss of heterozygosity (LOH) (71%). This is in contrast to sporadic colon cancer, where physical loss is the main characteristic. The percentage of chromosomal gains (24%) is comparable to sporadic colorectal cancers with CIN. Furthermore, we verified our scoring of copy-neutral LOH versus physical loss in MAP carcinomas by two methods: fluorescence *in situ* hybridization, and LOH analysis using polymorphic markers on carcinoma fractions purified by flow sorting. The results presented in this study suggest that copy-neutral LOH is an important mechanism in the tumorigenesis of MAP.

Keywords: *MUTYH*; tumour profiling; copy-neutral LOH; SNP arrays; MAP carcinomas; colorectal adenomatous polyposis; colorectal cancer

Introduction

MUTYH-associated polyposis (MAP) is the first colorectal cancer syndrome shown to be inherited in an autosomal recessive fashion. Biallelic mutations in the base excision repair (BER) gene *MUTYH* have been shown to cause colorectal adenomatous polyposis, and correlate with a high risk of developing carcinomas [1]. BER is a DNA repair mechanism that guards oxidative DNA damage and other metabolic DNA damage. Upon oxidative DNA damage, MUTYH removes incorrectly incorporated adenines opposite to an 8-oxo-guanine. Consequently, MAP patients show somatic G : C → T : A mutations in crucial genes such as *APC* and *KRAS*. In *APC*, these G : C → T : A transversions seem to occur primarily in GAA sequences [1,2]. In *KRAS*, a specific GGT → TGT mutation (c.34 G → T, p.Gly12Cys) is found in up to 64% of MAP carcinomas [3]. Interestingly few *p53* and *SMAD4* mutations are found in MAP carcinomas, whereas these genes are frequently affected in sporadic colorectal cancer [3]. Although MUTYH deficiency triggers carcinogenesis by G : C → T : A transversions, the exact role of MUTYH deficiency in the tumour progression in MAP patients is still unknown.

For colorectal cancers, different types of genetic instability are known to drive carcinogenesis. The two main types of genetic instability are microsatellite instability (MIN or MSI) and chromosomal instability (CIN). CIN is defined as an accelerated rate of chromosomal missegregation resulting in an aberrant chromosomal content, and is found in the vast majority of sporadic colorectal cancers [4]. On the other hand, ~15% of the sporadic colorectal cancers show MSI, due to *MLH1* promoter hyper-methylation [5]. Moreover, MSI is typically seen in the carcinomas of Lynch syndrome patients. Colon carcinomas that display neither CIN nor MSI have also been described [6]. More recently, abnormal epigenetic modification has been described in colorectal cancer, exhibiting the CpG island methylator phenotype (CIMP) [7,8].

The genomic profile of MAP tumours has been described in three studies to date. Using flow

cytometry, Lipton *et al* found MAP carcinomas to be predominantly near-diploid. Comparative genomic hybridization (aCGH) of two near-diploid MAP carcinomas showed no detectable chromosomal gains or losses. Furthermore, they analysed chromosomes 1p, 2p, 5q, 10p, 15q, 18q and 20q for LOH, using microsatellite markers, and reported a high frequency of LOH for chromosome 18q but low levels of LOH for the other regions [3]. Recently, the same research group identified only a small number of copy number changes in MAP adenomas [9]. These changes were mainly restricted to chromosomes 1p, 13, 17p, 19 and 22. Additionally, in a single MAP adenoma, copy-neutral LOH (cnLOH) of whole chromosome 7 and 12 was reported. On the other hand, Cardoso *et al* identified chromosomal copy number aberrations in MAP adenomas using aCGH analysis. The most prevalent aberrations identified were gains at chromosomes 7 and 13, as well as physical losses on chromosomes 17p, 19p and 22q [10]. However, the ploidy status of these adenomas was not determined.

Although these studies seem to be contradictory, Lipton *et al* studied carcinomas, whereas the other studies analysed adenomas. In addition, different technical platforms were used, i.e. flow cytometry vs. aCGH after amplification of laser capture microdissected DNA.

In order to gain further insight into the genetic instability involved in MAP carcinogenesis, we analysed formalin-fixed paraffin-embedded tumour tissue from 26 carcinomas for patterns of chromosomal losses and gains and copy-neutral LOH using SNP arrays [11,12].

Materials and methods

Samples

From 19 MAP patients, 26 formalin-fixed paraffin-embedded (FFPE) carcinomas and corresponding normal tissue were selected (Table 1). This series of carcinomas included metastases of primary colon carcinomas (t10 and t11). Corresponding normal tissue was either histological normal colon tissue or tissue from unaffected lymph nodes. The carcinomas originated from 11 biallelic Y165C mutation carriers, two biallelic P391L mutation carriers, three Y165C/G382D compound heterozygotes, one 1105delC/G382D, one P391L/G382D and one P391L/R233X compound heterozygote. Clinical details of patients 2, 3, 8, 9, 10, 12, 13, 14, 15, 16 and 17 were previously described

Table 1. Characteristics of the MAP carcinomas

Tumour	Patient	*MUTYH* mutation	Site CRC*	Age at diagnosis	Tumour stage	DNA index
t1	1	Y165C/Y165C	Distal	52	I	0.9 + 1.7†
t2	2	Y165C/Y165C	Distal	49	II	1.1 + 1.4†
t3	3	Y165C/Y165C	Proximal	39	II	1.0
t4	4	Y165C/Y165C	Proximal	49	III	1.0 + 1.5†
t5	5	Y165C/Y165C	Distal	56	I	1.6
t6	6	Y165C/Y165C	Proximal	53	II	1.0
t7.1	7	Y165C/Y165C	Proximal	43	II	1.0 + 1.5†
t7.2	7	Y165C/Y165C	Distal	43	II	1.0 + 1.5†
t8.1	8	Y165C/Y165C	Proximal	41	III	na
t8.2	8	Y165C/Y165C	Proximal	41	III	na
t8.3	8	Y165C/Y165C	Proximal	41	III	1.0
t8.4	8	Y165C/Y165C	Distal	41	III	1.0
t9	9	Y165C/Y165C	Ileum	77	II	1.0
t10	10	Y165C/Y165C	Metastases‡	45	IV	1.5 + 2.7§
t11	11	Y165C/Y165C	Metastasis‡	64	IV	1.5
t12	12	Y165C/G382D	Proximal	67	III	1.0
t13.1	13	Y165C/G382D	Proximal	43	II	1.0 + 1.1†
t13.2	13	Y165C/G382D	Proximal	46	II	1.0
t14	14	Y165C/G382D	Proximal	59	II	1.0
t15.1	15	P391L/P391L	Proximal	37	III	1.1 + 1.4†
t15.2	15	P391L/P391L	Proximal	37	III	Na
t16	16	P391L/P391L	Distal	58	II	1.0
t17.1	17	1105delC/G382D	Distal	42	I	1.1
t17.2	17	1105delC/G382D	Distal	42	I	1.0
t18	18	R233X/P391L	Proximal	48	II	1.4
t19	19	G382D/P391L	Proximal	51	III	1.1

The tumours were located before (proximal) or after (distal) to the splenic flexura of the colon. T11 is a metastasis of an earlier colon carcinoma. T10 consists of two metastases of a colorectal carcinoma from patient 10.The DNA index was measured by multiparameter DNA flow cytometry. When two populations were identified in the keratin-positive fraction, the DNA index of both tumour fractions is shown in the table. Tumour staging was performed according to the TNM classification (http://tnm.uicc.org).

na, could not be analysed for technical reasons.

* All tumours were colorectal with the exception of t9 (ileum).

† Multiple clones.

‡ Exact location of the primary tumour in the colon not known.

§ Two metastases of a primary colorectal carcinoma with DNA index 1.5 and 2.7, respectively.

by Nielsen *et al* [13] (as the respective numbers 13, 4, 11, 12, 14, 20, 18, 16, 35, 34 and 30). Twenty-two previously published sporadic CRCs [14] were included as reference controls.

The study was approved by the local Medical Ethical Committee (protocol P01.019); samples were handled according to the medical ethical guidelines described in the Code Proper Secondary Use of Human Tissue established by the Dutch Federation of Medical Sciences (www.federa.org). Tumour samples were enriched for tumour tissue by taking 0.6 mm tissue punches, using a tissue microarrayer (Beecher Instruments, Sun Prairie, WI, USA) guided by a haematoxylin and eosin (H&E)-stained slide. DNA was isolated by the previously described method, and subsequently cleaned using the Genomic Wizard kit (Promega, Leiden, The Netherlands) [15]. DNA concentrations were measured with the picogreen method (Invitrogen–Molecular Probes, Breda, The Netherlands).

Flow cytometry and cell sorting

For 23 carcinomas, the DNA index was determined by flow cytometry, as described previously with minor modifications [16]. In short, cell suspensions were prepared from FFPE samples and stained for keratin (APC), vimentin (RPE) and DNA (DAPI). Samples were analysed on a LSRII flow cytometer (BD Biosciences, San Jose, CA, USA). From five MAP carcinomas (t2, t4, t10, t12 and t18) and one sporadic carcinoma (sp1), those cell fractions that were vimentin-positive, keratin-negative (V^+K^-) and vimentin-negative, keratin-positive (V^-K^+) were flow-sorted using a FACSAria cell sorter (BD Biosciences).

Single nucleotide polymorphism arrays

Illumina BeadArrays were used in combination with the linkage mapping panel IV_B4b (Illumina, San Diego, CA, USA) [11], which consists of four panels. Panel I covers chromosomes 1, 2, 3 and 22; panel II covers chromosomes 5, 6, 7, 8 and 9; panel III covers chromosomes 10, 11, 12, 13, 14, 15 and 21; and panel IV covers chromosomes 4, 16, 17, 18, 19, 20, X and Y. The GoldenGate assay was performed according to the manufacturer's protocol, with minor adjustments: 1 µg input DNA was used for multi-use activation and resuspended in 60 µl RS1 [17]. Genotypes were extracted using GenCall (version 6.0.7) and GTS Reports (version 4.0.10.0; Illumina). Tumours t3, t9, t10 and t14 could only be analysed for copy number abnormalities, since corresponding normal tissue was unavailable to determine cnLOH. For t1, t17.2 and t18 only three panels could be analysed, due to limited availability of the FFPE tumour DNA. We corrected for this missing information in our calculations.

Analysis of copy numbers and loss of heterozygosity

Copy number and cnLOH profiles were generated by analysing the carcinomas and corresponding normal tissue in 'Beadarray SNP' [12]. Criteria for the scoring of copy number aberrations were based on previous experiments [12]. LOH was determined as follows. The ratio between the GenCall Score (GCS) and the GenTrain Score (GTS) was computed as a relative measure for the quality of the clustering of the SNP. All high-quality heterozygous SNPs (GCS/GTS >0.8) in the normal sample were included in the analysis. For homozygous SNPs and those with a GCS/GTS <0.8 in the tumour, LOH was assigned. LOH at one or two SNPs was ignored. In practice, regions of LOH always presented as stretches of markers showing LOH. When both a copy number change and LOH were detected at a specific region, the detected LOH was considered to be a consequence of the copy number alteration. If no copy number change was detected, LOH was interpreted as cnLOH.

For verification, conventional LOH analysis was performed for chromosomes 17p and 18q, using microsatellite markers (D17S938, D17S921, D18S877, D18S65, D18S460 and D18S1137) in pure tumour DNA of five MAP carcinomas obtained after flow sorting. Normal DNA was used as a reference. As a positive control, one sporadic carcinoma with known physical loss of chromosomes 17p and 18q was included. A standard PCR protocol was used for amplification. Mixtures of 9.5 µl HiDi formamide, 0.5 µl ROX 500 size standard and 2.0 µl PCR product were run on an ABI 3130 Genetic Analyser (Applied Biosystems) and analysed using GeneMapper version 4.0 (Applied Biosystems).

Interphase fluorescence *in situ* hybridization (FISH)

FISH was performed on flow-sorted nuclei that were spotted onto glass slides, as described previously [18]. The nuclei of five carcinomas were hybridized with a BAC on 17p13.1 (RP11-199F11, spanning the *p53* locus), a BAC on 18q21.1 (RP11-748M14, spanning the *SMAD2* locus) and centromere probes for chromosome 17 and 18. For all carcinomas 50 nuclei were scored. For heterogeneous tumours, each cell population that represented at least one-third of the scored nuclei was considered as a separate fraction.

Statistics

The amount of chromosomal aberrations identified in the MAP carcinomas and the sporadic carcinomas was compared using a Mann–Whitney U-test for independent samples. The analyses were performed using SPSS 12.0.1.

Results

We studied a series of 26 Dutch carcinomas from 19 biallelic *MUTYH* mutation carriers. All patients

were diagnosed with >10 colon polyps (median age at diagnosis 49 years, range 37–77 years), ranging from 10–50 polyps to polyposis with >50–100 polyps. The carcinomas were predominantly located proximal to the splenic flexura (15/24 reported) (Table 1). Most carcinomas were stage II (11/26 or 42%) or stage III (9/26 or 35%); 68% of the MAP carcinomas contained a somatic mutation in *KRAS* (16/17 mutations: c.34 G > T, p.Gly12Cys). A low level of mutations (12%) in the mutation cluster region of *APC* was identified and all carcinomas were microsatellite-stable (data not shown).

Using SNP arrays suitable for analysis of FFPE tissue, we were able to study the 26 carcinomas for genome-wide copy number abnormalities and genome-wide copy-neutral loss of heterozygosity (cnLOH) (see Supplementary Table 1, available at: http://www.interscience.wiley.com/jpages/0022-3417/suppmat/path.2375.html). Remarkably, this analysis revealed that 71% of all changes in the MAP carcinomas concerned cnLOH, whereas only 29% comprised copy number abnormalities (mainly chromosomal gains). On average, 5.1 (range 1–14) cnLOH events were identified per carcinoma. The cnLOH involved chromosome arms or complete chromosomes, but cnLOH of smaller chromosomal regions was also frequently identified. The regions most commonly affected by cnLOH in these tumours were chromosome 17p (57%), 18q (52%) and 15q (52%). Copy-neutral LOH was also frequently present at chromosome 6p (36% of the carcinomas). Lower frequencies of cnLOH were found for chromosomes 4p (24%), 4q (29%), 6q (23%), 8p (23%), 10q (24%), 18p (24%), 21q (24%) and 22q (29%) (Figure 1, Supplementary Table 1).

The MAP carcinomas we studied displayed only a few copy number abnormalities (on average 2.5, range 0–9). This is in contrast to sporadic colorectal cancer, where many chromosomal gains and losses are generally seen [4]. In all patients, the tumours showed five or fewer changes, except for patients 10 and 11, who showed eight, nine and six aberrations, respectively. Gain of chromosome 13q was the most prevalent aberration, seen in 9/26 (35%) carcinomas. Chromosome 11q was amplified in 6/26 (23%) carcinomas. Very limited physical chromosomal loss occurred in the MAP carcinomas (Figure 1, Supplementary Table 1).

The absence of gross chromosomal copy number alterations in our series of MAP carcinomas may reflect a near-diploid genome. For 23 carcinomas, we were able to measure ploidy status using flow cytometry. This analysis concluded that 12/23 (52%) MAP carcinomas were, indeed, near-diploid (DNA index, 1.0 ± 0.1). We found three cases with a near-triploid DNA index (1.5 ± 0.1). In addition, flow cytometry revealed that seven carcinomas contained two fractions, each with a different DNA index. In all seven of these carcinomas, one of the fractions was near-diploid, while the other fraction had a DNA index of 1.5 in five of the seven cases. Tumour 10 consisted of two metastases from the same primary tumour, each with a different DNA ploidy (Table 1).

We further compared the 19 MAP carcinomas to the CIN profile of sporadic carcinomas (Figure 1). For accurate comparison, we used a series of 22 microsatellite-stable sporadic carcinomas with CIN that were analysed previously using the same SNP methodology [14] and displayed the typical CIN profile of sporadic carcinomas [4]. Our comparison showed that the amount of cnLOH in MAP carcinomas

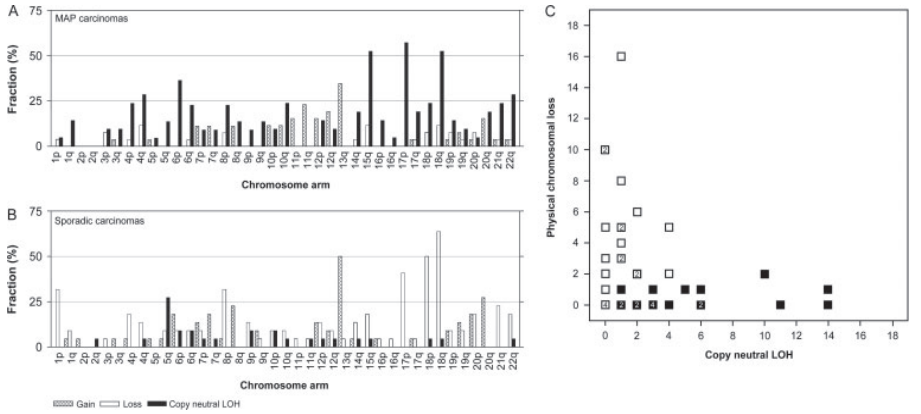

Figure 1. Chromosomal aberrations in MAP carcinomas versus sporadic CRCs. (A, B) The bars indicate the percentage of the 26 MAP carcinomas and 22 sporadic carcinomas, respectively, that exhibit an event of gain, loss or cnLOH of a chromosome. This percentage has been calculated for the respective chromosome arms. White bars, chromosomal gains; checked bars, physical losses of chromosomes; black bars, cnLOH. (C) In this graph the number of cnLOH events versus the number of physical losses is depicted for the 19 MAP carcinomas for which all genomic information was collected (see Materials and methods) versus 22 sporadic carcinomas. White squares, sporadic carcinoma; black squares, MAP carcinoma. The numbered squares represent multiple carcinomas that share the same amount of copy-neutral LOH and physical chromosomal loss

is significantly increased compared to sporadic carcinomas ($p < 0.001$). Moreover, the amount of physical chromosomal losses is significantly ($p < 0.001$) decreased compared to sporadic carcinomas (Figure 1). No differences were seen in the number of chromosomal gains between MAP carcinomas and the sporadic carcinomas. The majority of chromosomal events that are targeted by cnLOH in MAP comprise physical loss instead of cnLOH in sporadic CRC.

The observed pattern of cnLOH versus physical loss was confirmed for five representative MAP carcinomas (t2, t4, t10, t12 and t18) after flow sorting, by FISH for chromosome 17p and 18q on tumour nuclei, in combination with LOH analysis using microsatellite markers. One sporadic carcinoma was included as a control (Table 2). The SNP arrays revealed that four of these five MAP carcinomas exhibited cnLOH on chromosome 17p (t2, t4, t12 and t18) and three exhibited cnLOH on chromosome 18q (t2, t12 and t18). Two MAP cases and the sporadic CRC displayed physical loss of chromosomes 17p and/or 18q. All FISH results that could be obtained were in agreement with our estimation based on the DNA index in combination with the SNP array results. For example, in the tumours with a near-diploid genome content, two copies of chromosome 17p and 18q were identified by FISH in case of cnLOH and in tumours with a near-triploid genome three copies were identified in case of cnLOH (Figure 2). However, within MAP carcinoma t18 (DI = 1.4) only half of the tumour nuclei showed three chromosomal arms of 18q, indicating intratumour heterogeneity. The sporadic carcinoma also harboured two cell populations, with different copy numbers on chromosomal arms 17p and 18q. LOH was unambiguously identified for all informative microsatellite markers in all these cases, also in the cases with cnLOH in the context of a triploid genome content (implying the presence of three copies of a single allele), except for D17S921 in the diploid fraction of MAP carcinoma t4, which showed retention. These results are concordant with the results obtained with the SNP array analysis.

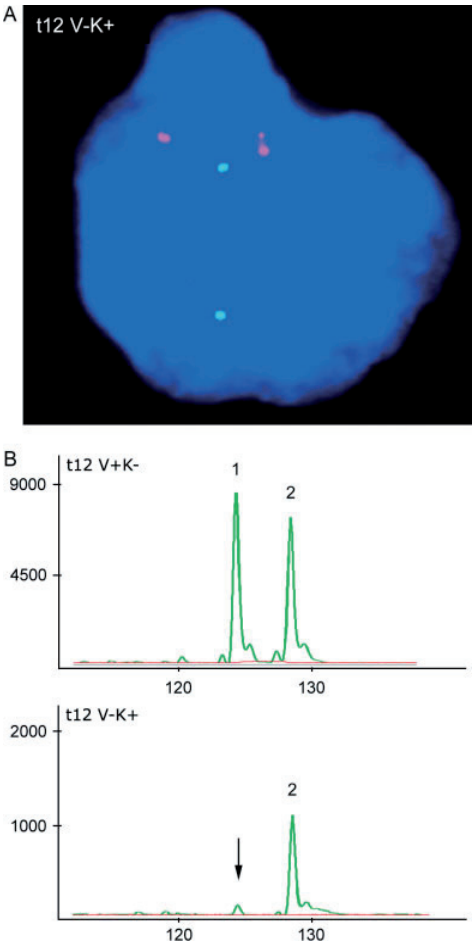

Figure 2. Microsatellite LOH analysis and fluorescent *in situ* hybridization on chromosome 18q21.1 after flow sorting of MAP carcinoma t12 (see also Table 2). (A) FISH showed two centromeric chromosome 18 signals (red) and two signals on 18q21.1 (green) for MAP carcinoma t12 (DNA index = 1.0). (B) Microsatellite LOH analysis (D18S877) on the flow-sorted MAP carcinoma t12 is shown: (upper panel) vimentin-positive, keratin-negative (normal) fraction; (lower panel) the vimentin-negative, keratin-positive (tumour) fraction. Unambiguous LOH is seen of allele 1 in the tumour. In combination with the FISH result shown in (A), copy-neutral LOH for chromosome 18q can be concluded

Discussion

Three studies have reported on the genetic profiles of MAP tumours [3,9,10]. Unfortunately, the results of these studies are seemingly contradictory. Copy number changes in adenomas have been reported, as well as near-diploidy in adenomas and carcinomas. In order to gain more insight into the genetic instability in MAP tumours we studied a series of 26 MAP carcinomas using SNP array analysis in FFPE tissue. In contrast to sporadic colorectal cancer, copy-neutral LOH (cnLOH) appears to be a prevalent characteristic of MAP carcinomas, while only a few copy number abnormalities were identified (4). However, the percentage of chromosomal gains (24%) is comparable to sporadic colorectal cancers with CIN. Such a genomic tumour profile of colon cancer has, to our knowledge, not been described before. With the recent availability of SNP arrays, more detailed information can be obtained on genome-wide cnLOH and several studies now report on cnLOH in cancers [19,20]. However, no study has described cnLOH to the extent seen in our series of MAP carcinomas.

The relative absence of chromosomal loss in our series of MAP carcinomas indeed reflects a

Table 2. Confirmation of copy-neutral LOH by FISH and microsatellite analysis

		p53 locus (17p)				SMAD4 and SMAD2 locus (18q)					
Tumour	**DNA index**	**SNP array Chr. 17**	**D17S938**	**D17S921**	**FISH 17p**	**SNP array Chr. 18**	**D18S877**	**D18S65**	**D18S460**	**D18S1137**	**FISH 18q**
t2 K^+ Dip	1.1	17p cnLOH*	LOH	LOH	na	18pq cnLOH*	LOH	na	na	LOH	na
t2 K^+ An	1.4	17p cnLOH*	LOH	LOH	3/3	18pq cnLOH*	LOH	na	LOH	na	na
t4 K^+ Dip	1	17pq cnLOH*‡	U	R	2/2	18pq phLoss*	na	LOH	LOH	U	2/2†
t4 K^+ An	1.5	17pq cnLOH*‡	U	LOH	3/3	18pq phLoss*	LOH	LOH	LOH	U	
t10 K^+	1.5	17p phLoss	U	U	na	18q phLoss	LOH	LOH	U	na	na
t12 K^+	1	17p cnLOH	LOH	LOH	2/2	18pq cnLOH	LOH	na	LOH	na	2/2
t18 K^+	1.4	17p cnLOH	U	LOH	3/3	18pq cnLOH	LOH	LOH	U	LOH	2/2, 3/3
sp1 K^+	1	17p phLoss	U	LOH	2/1, 1/1	18pq phLoss	LOH	na	U	na	2/2, 1/1

Microsatellite LOH analysis and FISH after flow sorting of five MAP carcinomas and one sporadic carcinoma was concordant with our estimation based on the DNA index and SNP array results. Chr., chromosome, K^+, keratin-positive, vimentin-negative (tumour) fraction after flow sorting; Dip, diploid fraction; An, aneuploid fraction; PhLoss, physical loss; cnLOH, copy-neutral LOH. For the LOH analysis: LOH, loss of heterozygosity; R, retention of both alleles; U, uninformative; na, could not be analysed for technical reasons. For the FISH results, the first number indicates the amount of centromeres and the second number indicates the amount of chromosomal arms 17p and 18q, respectively.

* Assay performed on unsorted tumour material.

† The FISH for t4 on chromosome 18q was, due to technical limitations, not performed on flow-sorted tumour nuclei, but on a tissue slide.

‡ cnLOH of complete chromosome 17.

near-diploid genome. Ploidy analysis using flow cytometry concluded that 12/23 (52%) MAP carcinomas analysed were near-diploid (DNA index, 1.0 ± 0.1). Lipton *et al* [3] found a near-diploid genome in 12/13 MAP carcinomas tested, with one carcinoma showing a polyploid status. We found three cases with a near-triploid DNA index (1.5 ± 0.1). In addition, flow cytometry revealed that seven carcinomas contained two fractions, each with a different DNA index. In all seven of these carcinomas, one of the fractions was near-diploid, while the other fraction had a DNA index of 1.5 in five of the seven cases. Interestingly, the distribution of the DNA ploidy of the MAP carcinomas is very different from sporadic colorectal cancers, which are primarily highly aneuploid. A DNA index of ~1.5 is uncommon in sporadic CRC, although near-triploidy has been described for sporadic CRC [21,22]. We confirmed the scoring of our SNP results in a purified set of tumours by a combination of FISH and LOH analysis, using polymorphic chromosomal microsatellite markers on chromosomes 17p and 18q. In the tumours with a near-diploid genome content, two copies of chromosome 17p and 18q were identified by FISH in case of cnLOH, and in tumours with a near-triploid genome three copies of a single allele were identified in case of cnLOH. Possible limited sensitivity in detecting copy number aberrations, especially in heterogeneous tumours, is unlikely in view of the FACS sorting in combination with FISH and conventional LOH analysis. Moreover, the results we obtained on the sporadic CRC are reassuring in this respect, since these are analysed and scored in exactly the same way as the MAP tumours.

Recently, we studied by SNP analysis of FFPE tissue a series of microsatellite-unstable sporadic and Lynch syndrome colon carcinomas, often with a near-diploid DNA content. All MSI-H carcinomas showed few chromosomal aberrations. CnLOH was infrequent in these tumours and usually confined to the locus harbouring a pathogenic mutation in *MLH1, MSH2* or *PMS2* [23]. These results further underline the uniqueness of the phenotype of the MAP carcinomas.

Interestingly, the cnLOH events identified in the MAP carcinomas frequently involve the same chromosomes affected by physical loss in sporadic colorectal cancer, indicating that the same tumorigenic pathway may be involved in tumour initiation and progression. For example, chromosomes 17p and 18q are commonly affected by physical loss in sporadic colorectal cancer, whereas cnLOH is identified primarily on these chromosome arms in MAP carcinomas. How frequent the genes that are targeted in sporadic colorectal cancer on these respective chromosomes, e.g. *p53* and the *SMAD* genes, are targeted in MAP carcinomas remains elusive. Lipton *et al* found only three *p53* (located on chromosome 17) somatic mutations in 14 MAP carcinomas analysed, although immunohistochemistry for p53 over-expression (indicative for mutation) was positive in four tumours that were negative for mutation testing. *SMAD4* mutations on 18q were not found in the MAP carcinomas analysed by Lipton *et al* [3], although analysis of two chromosome 18q microsatellite markers showed a high frequency of 18q LOH in 7/14 cases analysed.

Our studies also indicate that chromosome 15q is often targeted by cnLOH in the MAP carcinomas. Physical loss of this chromosome has been associated with distant metastasis of sporadic colorectal cancer [4].

Copy-neutral LOH can arise via mitotic recombination, non-disjunction, or deletion and reduplication events. In our series, we identified cnLOH on whole chromosomes and on parts of chromosomes. The high prevalence of cnLOH in MAP carcinomas suggests a relationship between mitotic recombination and the MUTYH deficiency. However, it is difficult to explain why MAP cancers show few copy number aberrations. First, the occurrence of copy-neutral LOH might be

directly linked to BER malfunctioning. Secondly, in parallel to a mismatch repair deficiency, the mutational burden might be relatively high due to the BER defect, favouring mitotic recombination but not physical loss. Therefore, further research into this possible relation is important.

Acknowledgements

This work was supported by the Dutch Cancer Society (Grant Nos UL2003-2807 and UL2005-3247).

Supplementary material

Supplementary material may be found at the web address: http://www.interscience.wiley.com/jpages/0022-3417/suppmat/path.2375.html

References

1. Al Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, *et al*. Inherited variants of MYH associated with somatic G : C → T : A mutations in colorectal tumors. *Nat Genet* 2002;**30:**227–232.
2. Jones S, Emmerson P, Maynard J, Best JM, Jordan S, Williams GT, *et al*. Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G : C → T : A mutations. *Hum Mol Genet* 2002;**11:**2961–2967.
3. Lipton L, Halford SE, Johnson V, Novelli MR, Jones A, Cummings C, *et al*. Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway. *Cancer Res* 2003;**63:**7595–7599.
4. Diep CB, Kleivi K, Ribeiro FR, Teixeira MR, Lindgjaerde OC, Lothe RA. The order of genetic events associated with colorectal cancer progression inferred from meta-analysis of copy number changes. *Genes Chromosomes Cancer* 2006;**45:**31–41.
5. Grady WM, Markowitz S. Genomic instability and colorectal cancer. *Curr Opin Gastroenterol* 2000;**16:**62–67.
6. Georgiades IB, Curtis LJ, Morris RM, Bird CC, Wyllie AH. Heterogeneity studies identify a subset of sporadic colorectal cancers without evidence for chromosomal or microsatellite instability. *Oncogene* 1999;**18:**7933–7940.
7. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, *et al*. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with *BRAF* mutation in colorectal cancer. *Nat Genet* 2006;**38:**787–793.
8. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci USA* 1999;**96:**8681–8686.
9. Jones AM, Thirlwell C, Howarth KM, Graham T, Chambers W, Segditsas S, *et al*. Analysis of copy number changes suggests chromosomal instability in a minority of large colorectal adenomas. *J Pathol* 2007;**213:**249–256.
10. Cardoso J, Molenaar L, de Menezes RX, van Leerdam M, Rosenberg C, Moslein G, *et al*. Chromosomal instability in *MYH*- and *APC*-mutant adenomatous polyps. *Cancer Res* 2006;**66:**2514–2519.
11. Lips EH, Dierssen JW, van Eijk R, Oosting J, Eilers PH, Tollenaar RA, *et al*. Reliable high-throughput genotyping and loss-of-heterozygosity detection in formalin-fixed, paraffin-embedded tumors using single nucleotide polymorphism arrays. *Cancer Res* 2005;**65:**10188–10191.
12. Oosting J, Lips EH, van Eijk R, Eilers PH, Szuhai K, Wijmenga C, *et al*. High-resolution copy number analysis of paraffin-embedded archival tissue using SNP BeadArrays. *Genome Res* 2007;**17:**368–376.
13. Nielsen M, Franken PF, Reinards TH, Weiss MM, Wagner A, van der KH, *et al*. Multiplicity in polyp count and extracolonic manifestations in 40 Dutch patients with MYH-associated polyposis coli (MAP). *J Med Genet* 2005;**42:**e54.
14. Lips EH, van Eijk R, de Graaf EJ, Doornebosch PG, Miranda NF, Oosting J, *et al*. Progression and tumor heterogeneity analysis in early rectal cancer. *Clin Cancer Res* 2008;**14:**772–781.
15. de Jong AE, van Puijenbroek M, Hendriks Y, Tops C, Wijnen J, Ausems MG, *et al*. Microsatellite instability, immunohistochemistry, and additional PMS2 staining in suspected hereditary non-polyposis colorectal cancer. *Clin Cancer Res* 2004;**10:**972–980.
16. Corver WE, Ter Haar NT, Dreef EJ, Miranda NF, Prins FA, Jordanova ES, *et al*. High-resolution multi-parameter DNA flow cytometry enables detection of tumour and stromal cell subpopulations in paraffin-embedded tissues. *J Pathol* 2005;**206:**233–241.
17. Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, *et al*. Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol* 2003;**68:**69–78.
18. Jordanova ES, Corver WE, Vonk MJ, Leers MP, Riemersma SA, Schuuring E, *et al*. Flow cytometric sorting of paraffin-embedded tumor tissues considerably improves molecular genetic analysis. *Am J Clin Pathol* 2003;**120:**327–334.
19. Gaasenbeek M, Howarth K, Rowan AJ, Gorman PA, Jones A, Chaplin T, *et al*. Combined array-comparative genomic hybridization and single-nucleotide polymorphism–loss of heterozygosity analysis reveals complex changes and multiple forms of chromosomal instability in colorectal cancers. *Cancer Res* 2006;**66:**3471–3479.
20. Andersen CL, Wiuf C, Kruhoffer M, Korsgaard M, Laurberg S, Orntoft TF. Frequent occurrence of uniparental disomy in colorectal cancer. *Carcinogenesis* 2007;**28:**38–48.
21. Tollenaar RA, Bonsing BA, Kuipers-Dijkshoorn NJ, Hermans J, van de Velde CJ, Cornelisse CJ, *et al*. Evidence of clonal divergence in colorectal carcinoma. *Cancer* 1997;**79:**1304–1314.
22. Giaretti W, Venesio T, Sciutto A, Prevosto C, Geido E, Risio M. Near-diploid and near-triploid human sporadic colorectal adenocarcinomas differ for *KRAS2* and *TP53* mutational status. *Genes Chromosomes Cancer* 2003;**37:**207–213.
23. van Puijenbroek M, Middeldorp A, Tops CM, van Eijk R, van der Klift HM, Vasen HF, *et al*. Genome-wide copy-neutral LOH is infrequent in familial and sporadic microsatellite unstable carcinomas. *Fam Cancer* 2008;DOI: 10.1007/s10689–008–9194–8.

Increased frequency of 20q gain and copy-neutral LOH in mismatch repair proficient familial colorectal carcinomas

Submitted

Chapter 6

Increased frequency of 20q gain and copy-neutral LOH in mismatch repair proficient familial colorectal carcinomas

A. Middeldorp[1], R. van Eijk[1], J. Oosting[1], M. van Puijenbroek[1], M. van Nieuwenhuizen[1], W.E. Corver[1], D. Ruano[1], T. Caldes[2], J. Wijnen[3], H. Morreau[1], T. van Wezel[1]

[1] Department of Pathology, Leiden University Medical Center, Leiden, the Netherlands

[2] Laboratory of Molecular Oncology, Hospital Clínico San Carlos, Madrid, Spain

[3] Department of Human Genetics and Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands

Abstract

Many hereditary non-polyposis colorectal cancers (CRCs) cannot be explained by Lynch syndrome. Other high penetrance genetic risk factors are likely to play a role in these mismatch repair (MMR)-proficient CRC families. Because genomic profiles of CRC tend to vary with CRC susceptibility syndromes, our aim was to analyze the genomic profile of MMR-proficient familial CRC to obtain insight into the biological basis of MMR-proficient familial CRC.

We studied 30 MMR-proficient familial colorectal carcinomas, from 15 families, for genomic aberrations, including gains, physical losses, and copy-neutral LOH (cnLOH) using SNP array comparative genomic hybridization. In addition, we performed somatic mutation analysis for *KRAS*, *BRAF*, *PIK3CA*, and *GNAS*.

The frequency of 20q gain (77%) is remarkably increased when compared to sporadic CRC, suggesting that 20q gain is involved in tumor progression of familial CRC. There is also a significant increase in the frequency of cnLOH and, as a consequence, a reduced frequency of physical loss compared to sporadic CRC. The most frequent aberrations observed included gains of 7p, 7q, 8q, 13q, 20p, and 20q, and physical losses of 17p, 18p, and 18q. Most of these changes are also observed in sporadic CRC. Mutations in *KRAS* were identified in 26% of the MMR-proficient CRCs and mutations in *BRAF* were identified in 12%. No mutations were identified in *PIK3CA* or the chromosome 20 candidate gene *GNAS*.

In conclusion, while the global patterns of MMR-proficient familial CRC resemble sporadic CRC, the chromosomal instability patterns exhibit a distinct pattern of aberrations with increased levels of cnLOH and 20q gain.

Introduction

Clinical criteria are used for the identification of Lynch syndrome patients. These so-called Amsterdam Criteria and Bethesda criteria include type of cancer, family history, and age at onset of disease.[1,2] Analysis of families that fulfill the strict Amsterdam Criteria I (AC-I) has proved to be successful in identifying germ-line mutations in mismatch repair (MMR) genes to be responsible for the increased CRC susceptibility in Lynch syndrome families. Mutations in MMR genes lead to deficient mismatch repair, which is reflected by microsatellite instability in tumors from Lynch syndrome patients. However, no mutations are identified in the MMR genes *MLH1* and *MSH2* in over half of the patients that meet the AC-I, suggesting that Lynch syndrome cannot explain all AC-I positive families.[3] Some of these AC-I positive families might have mutations in *MSH6*, *PMS2*, or undetected mutations in *MLH1* or *MSH2*; however, this will not explain all AC-1 positive families. Similarly, approximately 40% of families that fulfill the less stringent Amsterdam Criteria II do not display microsatellite instability, a characteristic of MMR deficiency, in their tumors.[4] Moreover, it has been estimated that approximately 32% of the excess CRC risk that is associated with a positive family history of CRC remains unexplained by known genes.[5] In these families, it is likely that other high penetrance genetic risk factors play a role. Analysis of the incidence of cancer in these families showed that they have an increased risk of CRC compared to the general population, albeit to a lesser extent than Lynch syndrome families.[5-7] Lindor and colleagues observed that MMR-proficient AC-I positive families have an increased risk for CRC but not for other cancers. Furthermore, members of MMR-proficient AC-I positive families tend to develop CRC at an older age than individuals in MMR-deficient AC-I positive families.[6] Aaltonen et al. estimated that first-degree relatives (FDRs) of probands with microsatellite stable cancer had a 1.3-fold increase in CRC risk.[5] In addition, a recent study reported that FDRs of CRC patients with microsatellite stable tumors had an increased risk for CRC and observed a strong correlation between the risk of CRC and the number of affected FDRs.[7] Linkage analysis in CRC pedigrees and affected siblings has been performed to identify novel high penetrance risk factors. Several chromosomal regions have been linked to colorectal cancer susceptibility, including 3q21-q24, 7q31, 9q22.2-31.2, 11q23.2, 11q13.4, 14q24.2, and 22q12.1.[8-15] However, none of these studies have led to the identification of a novel CRC susceptibility gene yet.

Target genes for somatic mutations tend to vary by cancer type. In CRC, p.V600E (c.1199T>A) mutations in *BRAF* are predominantly seen in tumors that have sporadic promoter hypermethylation of *MLH1*.[16] *BRAF* mutations are, on the other hand, rare in the tumors of Lynch syndrome patients. Similarly, in *MUTYH*-associated polyposis G>T transversions are observed, with GAA>TAA mutations in *APC* and specific GGT>TGT mutations in codon 12 of *KRAS* (c.34 G>T, p.Gly12Cys). Tumors in Lynch syndrome patients, in contrast, often carry mutations in codon 13 of *KRAS* (c.38 G>A, p.Gly13Asp).[17,18] In addition to somatic mutations, distinct patterns of genetic instability are associated with specific CRC syndromes. Lynch syndrome carcinomas are characterized by microsatellite instability and copy-neutral

loss of heterozygosity (cnLOH) at the locus of the mutated mismatch repair gene, and rarely have large chromosomal aberrations.[19,20] In contrast, at the genomic level, carcinomas of *MUTYH*-associated polyposis patients are characterized by a high level of cnLOH throughout the genome. In addition, the cancer cells are generally near-diploid or near-triploid.[21] Similarly, characteristic genomic aberrations are observed in breast cancers from patients with a germ-line mutation in *BRCA1* or, to a lesser extent, *BRCA2*.[22-24] Therefore, genomic profiling of familial CRCs may provide insight into the biological basis of CRC in these families.
We studied the pattern of genomic aberrations in carcinomas from MMR-proficient familial CRC patients. Our goal was to generate a profile of genomic aberrations in MMR-proficient familial CRC that might provide insight into the biological basis of the increased CRC susceptibility of these families.

Material and Methods

Tumor samples

Thirty MMR-proficient familial colorectal carcinomas and corresponding histologically normal tissues were selected (Table 1). The formalin-fixed paraffin-embedded (FFPE) tumors originate from fifteen families. All families have a positive history of CRC, with either one or two affected generations. Two families fulfill the AC-I and in the other families, at least two family members are affected with CRC. Two tumors from each of the fifteen families were analyzed. Tumor samples were enriched for tumor tissue by taking 0.6 mm tissue punches from the tumor field using a tissue microarrayer (Beecher Instruments, Sun Prairie, WI, USA), guided by a haematoxylin and eosin (H&E)-stained slide. DNA was isolated using a method, described previously; subsequently, DNA was cleaned using the Genomic Wizard kit (Promega, Leiden, the Netherlands).[4] DNA concentrations were measured using the picogreen method (Invitrogen-Molecular Probes, Breda, the Netherlands).
Tumors were tested for microsatellite instability using the marker set recommended by the National Cancer Institute Workshop on Microsatellite Instability, supplemented with three additional mononucleotide repeat markers (BAT40, MSH3, and MSH6), as described previously.[4,25] All of the tumors were microsatellite-stable (MSS).
The study was approved by the local Medical Ethical Committee (protocol P01.019); samples were handled according to the medical ethical guidelines described in the Code Proper Secondary Use of Human Tissue established by the Dutch Federation of Medical Sciences.[1]

SNP array profiling

We analyzed thirty carcinomas from fifteen CRC families for genomic aberrations, including gains, physical losses and cnLOH. We used Illumina Beadarrays in combination with the linkage mapping panel IV_B4b and human linkage V panel (Illumina, San Diego, CA, USA). The

[1] http://www.federa.org

Family	Colorectal cancers[a]	Other cancers[b]	Affected with CRC	Generations[c]	AC-I[d]	Relationship[e]
F1	C46, C55, C56, C69	/	4 of 10 sibs	1	-	FDR (sibs)
F2	C34, C50, C60	/	3 of 14	2	-	TDR
F3	C48, C53	Thyroid55	2 of 4 sibs	1	-	FDR (sibs)
F4	C51, C60, C76, C77, C83	/	5 of 8	2	-	FDR
F5	C40, C56 + unknown	Unknown	unknown	unknown	unknown	unknown
F6	C56, C58, C63, C72, C73, C76[f]	Breast, Ovary67	5 of 11 sibs	1	-	FDR (sibs)
F7	C48, C52, C54, C55, C55, C60, C60, C66, C68	Lung65, Breast52, Breast55, Breast49	9 of 24	2	+	Fourth-degree relatives
F8	C52, C61, C73	/	3 of 10	2	-	FDR (sibs)
F9	C51, C53	/	2 of 7 sibs	1	-	FDR (sibs)
F10	C53, C54, C63, C65	/	4 of 5 sibs	1	-	FDR (sibs)
F11	C49, C51, C76	/	3 of 4	2	+	FDR (sibs)
F12	C50, C71	Pancreas65	2 of 6 sibs	1	-	FDR (sibs)
F13	C42, C42, C60, C64	/	4 of 15	2	-	FDR (sibs)
F14	C28, C36, C50, C56	Thyroid58	4 of 14	2	-	TDR
F15	C52, C65, C66, C78	Leukemia54, Eusop/ Gastric68, Gastric45	4 of 12	2	-	TDR

Table 1. Characteristics of the cohort of MMR-proficient CRC families.
[a] Colorectal cancer cases in the families, including the corresponding age at diagnosis for each CRC; [b] Other cancers diagnosed in the families, including the corresponding age at diagnosis for each tumor; [c] The number of generations affected with CRC in the family; [d] Fulfillment of the Amsterdam Criteria I; [e] The relationship of the two family members from which the tumors were analyzed in this study. FDR, first-degree relative; SDR, second-degree relative; and TDR, third-degree relative; [f] One family member had two colorectal cancers (C56 and C76).

GoldenGate assay was performed at the Leiden Genome Technology Center[2] following the manufacturer's instructions, with minor adjustments: 0.5 µg input DNA was used for multi-use activation and resuspended in 60 µl RS1.[26] Genotypes were extracted using BeadStudio (V3.2, Illumina, San Diego, CA, USA). Genomic profiles were generated from the tumors and their corresponding normal tissue using the "Beadarray SNP" package, as described previously.[19,21,27]
The nature of the tissue, archival FFPE tissue, did not allow us to study germ line copy num-

[2] http://www.lgtc.nl

ber changes. Such small inherited germ line deletions and amplifications might be the underlying cause of the inherited risk for CRC in these patients. SNP arrays with a higher resolution should be used to study these types of copy number changes; however, these arrays are not yet suitable for DNA isolated from FFPE tissue.

Homozygosity mapping

To study the possibility of a recessively inherited risk locus in MMR-proficient familial CRC patients, we analyzed the normal tissue DNA for regions of homozygosity. For each SNP the chance of homozygosity was calculated as 1 minus the frequency of the heterozygote genotype. These calculations used the frequency of the heterozygote genotype found in the Hapmap project for the CEU population.[28] A measure for the extent of homozygosity was derived by multiplying the chances for homozygosity for each SNP in a consecutive stretch of homozygous SNPs. The extent of homozygosity along the chromosomes was visualized as a weighted grey value (darker grey corresponds with an increasing extent of homozygosity), with a cut-off of 10-3, representing stretches of at least ten homozygous SNPs with minor allele frequencies of approximately 0.5.

Somatic mutation analysis

Tumor samples were screened for mutations in *KRAS* exons 1 and 2, *BRAF* exon 15, and *PIK3CA* exons 9 and 20 by Sanger sequence analysis. PCR was performed on DNA extracted from FFPE material using iQ supermix (Bio-Rad Laboratories, Veenendaal, the Netherlands) using standard conditions.[29] Sequence analysis was performed at the Leiden Genome Technology Center.

Tumor samples were analyzed for the hotspot mutation in *GNAS* (c.601C>T, p.Arg201Cys) using the Taqman SNP Genotyping assay, following the manufacturer's protocol (Applied Biosystems), in a LightCycler 480 (Roche Applied Science, Almere, The Netherlands).

Some DNA fragments isolated from the FFPE tissues failed to amplify because of the limited fragment size that can be amplified from FFPE tissue. Details of the reaction conditions are available upon request. Primer sequences are provided in supplementary table 1.

Mutation frequencies were compared with the mutation data obtained from the Sanger Institute COSMIC (Catalogue of Somatic Mutations in Cancer) database[3].[30]

Results and discussion

We analyzed 30 colorectal carcinomas from 15 families with a history of CRC, using Illumina 6K Beadarrays, to study the genomic profile of MMR-proficient familial CRC. Characteristics of the studied cohort are detailed in Table 1. The studied families were predominantly affected with CRC; however, other malignancies were observed in family members, including leuke-

[3] http://www.sanger.ac.uk/cosmic

mia, gastric cancer, thyroid, breast, ovarian, lung, and pancreatic cancers.

The mean age at diagnosis of CRC was 56.4 years (range 28-77). All carcinomas had a microsatellite stable (MSS) phenotype. The majority of CRCs were Dukes B (14/28) or C (8/28) and nearly all were left-sided (25/29). The tumor location varies in different CRC syndromes. Seventy percent of the tumors of Lynch syndrome patients are located in the right side of the colon, whereas almost all familial adenomatous polyposis (FAP) cancers develop in the left side of the colon.[31-33] In addition, carcinomas from patients with *MUTYH*-associated polyposis are predominantly localized in the proximal (right) colon.[34] Familial MSS tumors have been reported to be predominantly located on the left side of the colon, which is consistent with the distribution we observed in our study.[35]

We analyzed the 30 MMR-proficient familial colorectal carcinomas for genome-wide chromosomal gains, physical losses, and cnLOH.[21] To study possible family specific tumor phenotypes, two carcinomas from each family were analyzed. We observed between 0 and 23 genomic aberrations in the MMR-proficient familial CRCs, with a mean of 11.4 aberrations per carcinoma. In addition, we observed a gradual increase in the number of aberrations per tumor from Dukes stage A to stage D. However, we did not detect a significant correlation between the number of aberrations and the Dukes stage of the tumor. This might be explained by the low number of Dukes stages A and D carcinomas in our series.

In a previous study, Rahman et al. reported an average of 5.9 copy number aberrations per MMR-proficient familial CRC, using comparative genomic hybridization. In addition, half of the MMR-proficient familial CRCs displayed less than 5 aberrations.[36] In order to compare our results with this earlier study, we counted the number of copy number aberrations (excluding cnLOH) in our series. In our series of MMR-proficient familial CRCs, there was on average 7.5 copy number aberrations per carcinoma, which is increased compared to the study of Rahman et al. We observed fewer than 5 copy number aberrations in 30% (9/30) of the CRCs. The most frequent aberrations, which we observed in at least 30% of the MMR-proficient familial CRCs, included gains of chromosome 7p (40%), 7q (33%), 8q (30%), 13q (57%), 20p (37%) and 20q (77%); and physical losses of 17p (37%), 18p (37%), and 18q (53%) (Figure 1). The observed aberrations that were most frequent in the MMR-proficient familial CRCs are well-known colorectal cancer aberrations, which are typically observed in sporadic colorectal cancers.[37,38]

We compared the profile of aberrations in our MMR-proficient familial CRCs series to that of sporadic CRC, MAP carcinomas, and Lynch carcinomas series that we analyzed previously, using the same methodology.[19,21,38] The frequencies of gains at 8q and 13q, physical losses at 17p and chromosome 18, and cnLOH at 5q are similar in MMR-proficient familial CRC and sporadic CRC. Gains of chromosomes 7 and 20 were observed more frequently in MMR-proficient familial CRC than in sporadic CRC. On the contrary, physical loss of 8p is observed less frequently in the familial carcinomas. cnLOH of 8p and 17p is observed more frequently in familial CRC than in sporadic CRC. We observed many aberrations that occur in only 10% to 20% of the familial tumors, suggesting heterogeneity among MMR-proficient

familial CRC (Figure 1).

The observed frequency of 20q gain (77%) is remarkably high in the series of MMR-proficient familial CRC compared to sporadic CRC. Based on the literature and our work, a frequency of 30-50% has been reported for 20q gain in sporadic CRC.[37,38] When compared with the 20q gain frequency in a series of sporadic CRCs (36%), which was previously analyzed using the same methodology, we found that the frequency of 20q gain is significantly increased in MMR-proficient familial CRC (p=3.4 x 10-3).[38] In addition, we extended the comparison to a previously reported large series of sporadic CRCs, in which both tumor and paired normal tissue were analyzed by SNP array CGH. [39] While the frequency of 20q gain in this large Japanese series of sporadic CRC is higher (56%), there is still a significant increase of 20q gain in familial CRC (p=4.7 x 10-2).[39] Furthermore, our results are in line with a study of Finnish colorectal cancers, which reported an 85% frequency of 20q gain in MMR-proficient familial CRCs using CGH.[40] In addition, a 70% frequency of 20q gain was reported in a group of early-onset MSS CRC patients.[41] This high frequency of 20q gain suggests that 20q is involved in tumor progression, particularly in MMR-proficient familial CRC. Several genes on 20q, including *AURKA*, *TH1L*, *ADRM1*, and *TCFL1*, have already been described to be differentially expressed in tumors with 20q gain when compared to tumors without 20q gain. [42] Analysis of the coding sequence of eleven CRCs by Sjöblom et al. identified three activating missense mutations in *GNAS*, which is located on 20q13.32.[43] Therefore, we analyzed our cohort of MMR-proficient familial CRC for the most frequent mutation (p.Arg201Cys, c.601C>T) in *GNAS*; however, we did not identify any mutation.

Unlike sporadic CRCs, the MMR-proficient familial CRCs displayed increased levels of cn-LOH genome-wide and, as a consequence, reduced physical loss. In the MMR-proficient familial CRCs, 37% of the aberrations were gains, 31% were physical losses, and 32% were cnLOH. This distribution is different from sporadic CRC, where 31% of the aberrations are gains, 55% are physical losses, and only 14% are cnLOH.[38] The percentage of cnLOH is significantly increased in MMR-proficient familial CRC over that of sporadic CRC (p=6.1 x 10-3). Additionally, the percentage of physical losses is significantly reduced in familial CRC compared to sporadic CRC (p=0.034).

Interestingly, we have previously shown that cnLOH is associated with tumor types that are deficient in DNA repair. cnLOH is the most frequent type of aberration (71%) in MUTYH-associated polyposis (MAP) colorectal carcinomas, which is caused by base excision repair deficiency.[21] Furthermore, in Lynch syndrome carcinomas, which are caused by DNA mismatch repair deficiency, cnLOH is predominantly observed at the locus of the mutated MMR gene. [19] While the MMR-proficient familial CRCs profile does not resemble the MAP carcinoma nor the Lynch syndrome carcinoma profile, the increased level of cnLOH suggests that an unknown DNA repair defect might be involved in the tumorigenesis of MMR-proficient familial CRC.

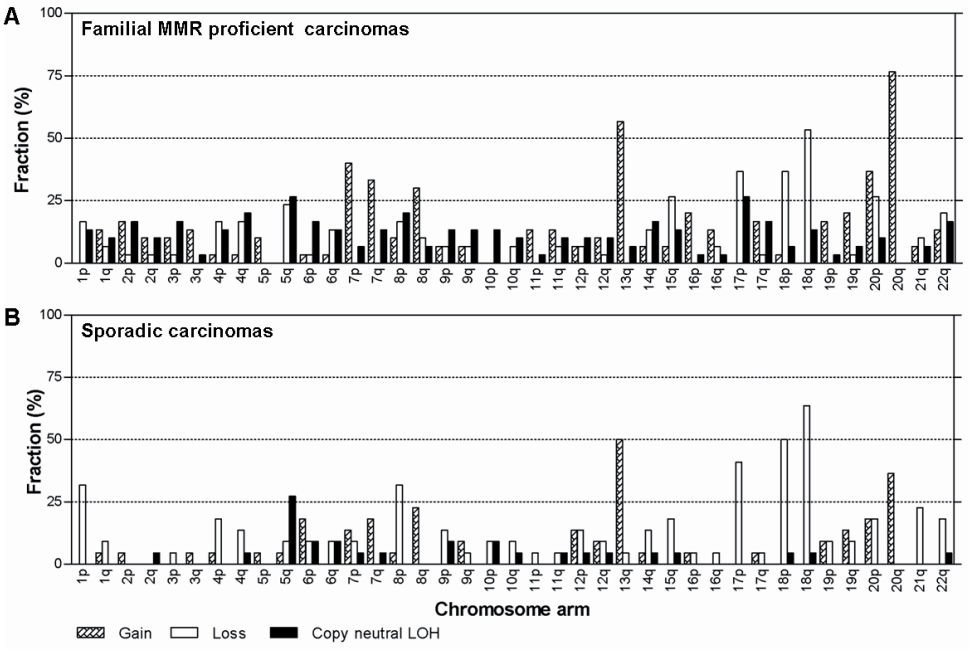

Figure 1. Chromosomal aberrations in MMR-proficient familial colorectal carcinomas
(A, B) The bars indicate the percentage of the 30 mismatch repair proficient familial colorectal carcinomas and 22 sporadic carcinomas that exhibit an event of gain, loss or cnLOH of a chromosome. This percentage has been calculated for the respective chromosome arms. White bars, chromosomal gains; checked bars, physical losses of chromosomes; black bars, cnLOH. Panel B shows the results from a previous study of Lips et al.[38]

In addition to comparing the MMR-proficient familial CRC profile with the sporadic CRC profile, we compared the genomic profiles of tumors within one family to identify possible family-specific genomic profiles. Aberrations that were most frequently shared by tumors from family members included the common aberrations at 13q, 17p, 18q, and 20q. There was a high incidence of 20q gain, which was shared by both tumors in over half of the families. Less frequently shared affected regions include 5q, 8p, 8q, 14q, and 15q. No distinctive profile of aberrations was observed in any of the families. The tumor phenotypes of all families roughly fit the overall genomic profile that was observed for MMR-proficient familial CRCs.
In two families, cnLOH of 8p was shared (F3 and F7) and in one family physical loss of part of 8p was shared (F14) by the tumors. In three families (F3, F7, and F13) aberrations at 8q were shared, however, gains, physical losses, as well as cnLOH were observed in these tumors. Three families (F1, F7, and F12) shared physical loss or cnLOH at the region on 5q that encompasses the APC gene. In addition, both tumors that were analyzed from three families (F6, F14, and F15) exhibited physical loss or cnLOH of 14q. Finally, chromosome 15q

displayed aberrations in both tumors from three families (F6, F11, and F15); however, gains, physical losses, and cnLOH were observed.

Homozygosity analysis

We searched normal tissue for regions of homozygosity that are shared within a family or between multiple individuals across families. These homozygous regions might suggest the presence of a recessively inherited gene, similar to *MUTYH*. In six of our fifteen families only one generation was affected with CRC, which suggests a recessive mode of inheritance. Moreover, in a previous study, we observed a homozygous region in MAP patients that had a shared haplotype that encompassed the *MUTYH* gene (results not shown).[21] In the current study, we did not identify any shared homozygous regions in the MMR-proficient familial CRCs, with the exception of a small region on chromosome 21q22.13. However, we found that this region was often homozygous in tumors from MAP and Lynch syndrome patients, suggesting that this region of homozygosity is not specific for MMR-proficient familial CRC.[19]

Somatic mutation analysis

We also studied the MMR-proficient familial CRCs for somatic mutations in *KRAS*, *BRAF*, and *PIK3CA*. Mutations in *KRAS* were identified in 5 of 19 tumors (26%). Three mutations were identified in codon 12 (one c.34G>T, p.Gly12Cys; and two c.35G>A, p.Gly12Asp), one in codon 13 (c.38G>A, p.Gly13Asp), and one in codon 63 (c.187G>A, p.Glu63Lys) of the *KRAS* gene. In *BRAF*, we detected mutations in codon 600 in 3 of the 26 tumors (12%) analyzed. Two of these mutations were V600E transversions (c.1199T>A, p.Val600Glu) and the third was a V600A transition (c.1199T>C, p.Val600Ala). None of the tumors carried mutations in both *KRAS* and *BRAF*, which has been observed in previous studies also.[16,36,44] The frequency of mutations in *KRAS* is lower and in *BRAF* is slightly increased compared to a previous study of MMR-proficient familial CRC (40% and 4%, respectively).[44] The frequency for both *KRAS* and *BRAF* mutations is similar to the reported mutation frequency in sporadic CRC (COSMIC, 32% and 12%, respectively).[30] We also screened exons 9 and 20 of *PIK3CA* for mutations; however, we did not detect a mutation in any of the 16 tumors that could be analyzed. These results are consistent with a previous report, in which *PIK3CA* mutations were found in only 4% of MSS familial CRCs.[45] The reported *PIK3CA* mutation frequency in sporadic CRC was higher (between 11% (COSMIC) and 32%).[30,46]

In conclusion, we show that the chromosomal instability patterns of MMR-proficient familial CRC are distinct from sporadic CRC, with significantly increased levels of 20q gain and genome-wide cnLOH. However, the overall aberration pattern resembles sporadic CRC.

The increased level of cnLOH in familial MMR-proficient CRC suggests a weak DNA repair defect. The high frequency of 20q gain suggests that there is an important role for this chromosomal region in tumorigenesis, similar to what was found by Laiho and colleagues. Howe-

ver, no gene on 20q has been reported to be mutated in familial CRC.[40] Further evaluation of 20q will be valuable and could include mutation analysis of candidate genes or next-generation sequencing of the entire region.

Acknowledgements

This work was supported by the Dutch Cancer Society (Grant UL2005-3247) and the NutsOhra Foundation (Grant SNO-T-07-092). Dr. T Caldes was supported by FIS 07/0359 and RTICC 06/0020/0021, ISCIII, Ministerio de Sanidad Español.

Table 2. Copy number changes and cnLOH identified in MMR-proficient familial CRC

Family	Tumor	Age at diagnosis	Location	MSI	Dukes	Genomic aberrations	Mutations
F1	1.1	60	Rectum	S	B1	*1p31.3-p22.2*, 2pter-p14, *4p14-q22.1*, *5q11.2-q31.1*, *8pter-p12*, *12pter-p12.3*, 12q11-q21.33, 13q, *15q*, 16pter-p12.1, *17p*, *18pq*, 19pq, **21q11.2-q21.3**, 21q22.11-qter, *22q*	
F1	1.2	46	Rectum	S	A	*1p*, **5q14.3**, 7pq, 12pq, 13q, 20pq, *21q*	*KRAS*, c.35G>A, p.Gly12Asp
F2	2.1	65	Rectum	S	C1	*1pter-p34.3*, **2pter-p12**, *6pter-p21.2*, *6q*, 8q21.11-q24.22, 11q13.2-qter, *14q21.1-q32.12*, *15q14-q25.3*, *17p*, *18q*, *20p*, 20q	*KRAS*, c.187G>A, p.Glu63Lys
F2	2.2	78	Sigmoid	S	B2	4pter-p15.1, **4q22.1-q22.3**, **6p22.2-qter**, **9pq**, 13q, **14q**, **15q**, *17p*, 17q, *18pq*, *20p*, 20q	
F3	3.1	53	Rectum	S	B1	*1pter-p35.1*, **2pq**, *8pter-q13.2*, *9q34.11-qter*, *14q*, 13q, *15q11.2-q25.1*, *17pter-p12*, *18pq*, *20pter-p11.22*, 20p11.22-qter	
F3	3.2	48	Colon asc.	S	C1	1q42.12-qter, *2q21.2-qter*, **3p14.3-p12.2**, **6q13-qter**, 7p, **8p**, 8q, 9q33.3, 14q, 22q11.21-q12.2, *22q12.2-qter*	
F4	4.1	77	Sigmoid	S	unknown	**1pq**, 2pq, **4pq**, *5q*, 9p, 13q, *14q*, 19q, 20pq, **21q**, 22q	
F4	4.2	51	Rectosigmoid	S	unknown	**2pter-p24.2**, 2p24.1-p22.3, **2p22.2-qter**, 8q23.3-qter, *18pq*, 20q	
F5	5.1	40	Rectum	S	B2	**4pter-q31.1**, 4q31.21-qter, *8p*, *10q21.1-qter*, *11q12.1-14.1*, *15q*, 16pq, **17pq**, **18pq**, 19pter-q13.2, **19q13.31-qter**, 20pq, *22q*	
F5	5.2	56	Rectosigmoid	S	B1	7pq, **9pq**, 13q, **17pq**, *20p*, 20q	
F6	6.1	72	Colon asc.	S	B1	**1q**, 7pq, **9pq**, **10pq**, **11pq**, **14q**, **15q**, *18pq*, **19pter-p13.2**, 20pq, **21q**	BRAF, c.1199T>A, p.Val600Glu
F6	6.2	76	Colon desc.	S	B2	*3p*, **5q13.2-qter**, *6q16.1, 8q22.1-qter, 9pter-p13.3, 10q23.31-25.1*, **12pter-12.1**, **12q15-q21.1**, 13q, *14q*, *15q11.2-12*, 15q12-q14, *15q14-qter*, *17p*, *18q*, *19q13.31-qter*, **20p**, 20q, 21q22.11-qter, *22q*	
F7	7.1	48	Sigmoid	S	B1	**1p36.13**, **5q13.2**, **7pq**, **8p**, 8q, **13q**, **17p**, **18q**, 20pq, **22q**	
F7	7.2	52	Rectosigmoid	S	D	*1pter-35.2*, *1q43-qter*, 2p16.1-13.3, 2q24.1-qter, 3q24-qter, *4pter-14*, **5q11.2-q33.3**, *5q34-qter*, **6pter-p21.31**, 7pq, **8p**, 8q, **9pter-21.2**, 11q13.2-22.3, 11q23.1-qter, **12pq**, **14q**, *17p*, 17q, *18q*, *20p*, 20q, *21q*	
F8	8.1	61	Rectum	S	B1	**2pq**, **5q**, **6pq**, **7q21.11-qter**, 13q, *16q21-qter*, *17p*, *18pq*, 20pq	*KRAS*, c.34G>T, p.Gly12Cys

Table 2 continued. Copy number changes and cnLOH identified in MMR-proficient familial CRC

Family	Tumor	Age at diagnosis	Location	MSI	Dukes	Genomic aberrations	Mutations
F8	8.2	53	Sigmoid	S	C2	1q, **3pter-p22.3**, **5q**, **6pter-q22.1**, 7p, **7q**, **8pter-p21.3**, **10pter-p13**, 13q, **15q**, **17p**, **18pq**, 19pter-q13.11, **19q13.11-qter**, 20pq, **22q11.23-qter**	
F9	9.1	51	Colon transv.	S	D	**4pq**, 5p, 6pter-22.1, 6q23.2-23.3, 7pter-q22.1, **8pq**, **10pq**, 11pter-q12.1, **11q22.3-qter**, 12pq, 14q, 15q11.2-25.3, **16pq**, **17pq**, 20q12-qter, **22q11.21-q12.2**, *22q12.3-qter*	
F9	9.2	53	Sigmoid	S	D	No chromosomal aberrations	
F10	10.1	62	Rectum	S	B1	**1pter-p34.1**, *4pq*, *5q13.1-q31.1*, 11p, **13q**, 16pq, **17p**, 17q, *18pq*, 19pq, **20pter-p11.23**, *20p11.22-qter*	
F10	10.2	54	Sigmoid	S	A	1q32.1-qter, 2pq, 3pq, 7pq, 8pq, 13q, 16pq, *20p*, 20q	
F11	11.1	50	Sigmoid	S	B1	3pq, 7pq, 8pq, 9pq, 13q, *15q11.2-q21.2*, **15q21.3-qter**, *17p*, *18pq*, 20pq, 22q	*BRAF*, c.1199T>A, p.Val600Glu
F11	11.2	51	Colon right-side	S	B2	*1q*, *2p24.3-p21*, **3pq**, *6q14.1-16.3*, **10pq**, **12pq**, 13q, *15q*, *17p*, 17q, *18pq*, *20p*, 20q, *21q*, *22q*	
F12	12.1	50	Sigmoid	S	C2	**1pter-35.1**, **1q12-q42.11**, 1q42.12-qter, 3pter-p13, **3p12.3-qter**, *4pq*, *5q14.1-q23.2*, *6q13-qter*, 7pter-14.1, **7p13-qter**, **8p**, 8q, *9pter-p33.1*, **9q33.2-qter**, 13q, **17p**, 17p11.2-q25.1, **17q25.1-qter,** 18p, **18q**, 19pq, **20pter-p12.3**, 22q	
F12	12.2	71	Rectosigmoid	S	B1	**5q**	
F13	13.1	42	Rectum	S	B1	**4pq**, 5p, **5q**, 7pq, **8q**, 13q, 20pq	
F13	13.2	42	Rectum	S	C2	*4pq*, *8q*, *11q*, *12p13.31-qter*, 13q, *15q*, *16q12.1-23.3*, *17pq*, *18pter-q22.3*	*KRAS*, c.35G>A, p.Gly12Asp
F14	14.1	56	Colon unspec	S	Cx	**4q12-q22.2**, *4q22.3-qter*, *5q22.2-q23.3*, 7pq, *8pter-p11.23*, 8p11.23-qter, 11p, 11q13.4-q23.2, **14q**, 16p, *17p*, *18pq*, **20p**, 20q, **22q**	
F14	14.2	28	Rectum	S	D	**2p22.3-p11.2**, **2q14.1-qter**, 5p, **5q11.2-q14.2**, 5q14.3-q21.3, **6p**, *8p*, **9q32-q33.1**, 13q, **14q**, *18q12.3-q22.1*, **20p**, 20q, **22q**	
F15	15.1	50	Rectosigmoid	S	C2	**3pter-p13**, 11pter-p13, 13q, 16p12.1-q13, **17pq**, *18q21.2-qter*, 20q	*BRAF*, c.1199T>C, p.Val600Ala
F15	15.2	34	Sigmoid	S	C2	No chromosomal aberrations	*KRAS*, c.38G>A, p.Gly13Asp

Chromosomal physical losses are depicted in italics, gains are underlined and cnLOH is depicted in bold face.

Reference List

1. Vasen HF, Watson P, Mecklin JP et al. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology (1999), 116: 1453-1456.
2. Umar A, Boland CR, Terdiman JP et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst (2004), 96: 261-268.
3. Wijnen JT, Vasen HF, Khan PM et al. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. N Engl J Med (1998), 339: 511-518.
4. de Jong AE, van Puijenbroek M, Hendriks Y et al. Microsatellite instability, immunohistochemistry, and additional PMS2 staining in suspected hereditary nonpolyposis colorectal cancer. Clin Cancer Res (2004), 10: 972-980.
5. Aaltonen L, Johns L, Jarvinen H et al. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. Clin Cancer Res (2007), 13: 356-361.
6. Lindor NM, Rabe K, Petersen GM et al. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. JAMA (2005), 293: 1979-1985.
7. Lubbe SJ, Webb EL, Chandler IP et al. Implications of familial colorectal cancer risk profiles and microsatellite instability status. J Clin Oncol (2009), 27: 2238-2244.
8. Kemp Z, Carvajal-Carmona L, Spain S et al. Evidence for a colorectal cancer susceptibility locus on chromosome 3q21-q24 from a high-density SNP genome-wide linkage scan. Hum Mol Genet (2006), 15: 2903-2910.
9. Kemp ZE, Carvajal-Carmona LG, Barclay E et al. Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United kingdom. Cancer Res (2006), 66: 5003-5006.
10. Middeldorp A, Jagmohan-Changur SC, van der Klift HM et al. Comprehensive genetic analysis of seven large families with mismatch repair proficient colorectal cancer. Genes Chromosomes Cancer (2010), 49: 539-548.
11. Neklason DW, Kerber RA, Nilson DB et al. Common familial colorectal cancer linked to chromosome 7q31: a genome-wide analysis. Cancer Res (2008), 68: 8993-8997.
12. Papaemmanuil E, Carvajal-Carmona L, Sellick GS et al. Deciphering the genetics of hereditary non-syndromic colorectal cancer. Eur J Hum Genet (2008), 16: 1477-1486.
13. Skoglund J, Djureinovic T, Zhou XL et al. Linkage analysis in a large Swedish family supports the presence of a susceptibility locus for adenoma and colorectal cancer on chromosome 9q22.32-31.1. J Med Genet (2006), 43: e7.
14. Wiesner GL, Daley D, Lewis S et al. A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. Proc Natl Acad Sci U S A (2003), 100: 12961-12965.
15. Djureinovic T, Skoglund J, Vandrovcova J et al. A genome wide linkage analysis in Swedish families with hereditary non-familial adenomatous polyposis/non-hereditary non-polyposis colorectal cancer. Gut (2006), 55: 362-366.
16. Deng G, Bell I, Crawley S et al. BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer. Clin Cancer Res (2004), 10: 191-195.
17. Lipton L, Halford SE, Johnson V et al. Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway. Cancer Res (2003), 63: 7595-7599.
18. Oliveira C, Westra JL, Arango D et al. Distinct patterns of KRAS mutations in colorectal carcinomas according to germline mismatch repair defects and hMLH1 methylation status. Hum Mol Genet (2004), 13: 2303-2311.
19. van Puijenbroek M, Middeldorp A, Tops CM et al. Genome-wide copy neutral LOH is infrequent in familial and sporadic microsatellite unstable carcinomas. Fam Cancer (2008), 7: 319-330.
20. Peltomaki P, Lothe RA, Aaltonen LA et al. Microsatellite instability is associated with tumors that characterize the hereditary non-polyposis colorectal carcinoma syndrome. Cancer Res (1993), 53:

5853-5855.
21. Middeldorp A, van Puijenbroek M, Nielsen M et al. High frequency of copy-neutral LOH in MUTYH-associated polyposis carcinomas. J Pathol (2008), 216: 25-31.
22. Jonsson G, Naylor TL, Vallon-Christersson J et al. Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization. Cancer Res (2005), 65: 7612-7621.
23. Joosse SA, van Beers EH, Tielen IH et al. Prediction of BRCA1-association in hereditary non-BRCA1/2 breast carcinomas with array-CGH. Breast Cancer Res Treat (2008).
24. Wessels LF, van Welsem T, Hart AA et al. Molecular classification of breast carcinomas by comparative genomic hybridization: a specific somatic genetic profile for BRCA1 tumors. Cancer Res (2002), 62: 7110-7117.
25. Boland CR, Thibodeau SN, Hamilton SR et al. 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: 5248-5257.
26. Fan JB, Oliphant A, Shen R et al. Highly parallel SNP genotyping. Cold Spring Harb Symp Quant Biol (2003), 68: 69-78.
27. Oosting J, Lips EH, van Eijk R et al. High-resolution copy number analysis of paraffin-embedded archival tissue using SNP BeadArrays. Genome Res (2007), 17: 368-376.
28. The International HapMap Project. Nature (2003), 426: 789-796.
29. van Eijk R, van Puijenbroek M, Chhatta AR et al. Sensitive and specific KRAS somatic mutation analysis on whole-genome amplified DNA from archival tissues. J Mol Diagn (2010), 12: 27-34.
30. Bamford S, Dawson E, Forbes S et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer (2004), 91: 355-358.
31. Lynch HT, Watson P, Lanspa SJ et al. Natural history of colorectal cancer in hereditary nonpolyposis colorectal cancer (Lynch syndromes I and II). Dis Colon Rectum (1988), 31: 439-444.
32. Bulow S. Clinical features in familial polyposis coli. Results of the Danish Polyposis Register. Dis Colon Rectum (1986), 29: 102-107.
33. Iacopetta B. Are there two sides to colorectal cancer? Int J Cancer (2002), 101: 403-408.
34. Nielsen M, de Miranda NF, van Puijenbroek M et al. Colorectal carcinomas in MUTYH-associated polyposis display histopathological similarities to microsatellite unstable carcinomas. BMC Cancer (2009), 9: 184.
35. Llor X, Pons E, Xicola RM et al. Differential features of colorectal cancers fulfilling Amsterdam criteria without involvement of the mutator pathway. Clin Cancer Res (2005), 11: 7304-7310.
36. Abdel-Rahman WM, Ollikainen M, Kariola R et al. Comprehensive characterization of HNPCC-related colorectal cancers reveals striking molecular features in families with no germline mismatch repair gene mutations. Oncogene (2005), 24: 1542-1551.
37. Diep CB, Kleivi K, Ribeiro FR et al. The order of genetic events associated with colorectal cancer progression inferred from meta-analysis of copy number changes. Genes Chromosomes Cancer (2006), 45: 31-41.
38. Lips EH, van Eijk R, de Graaf EJ et al. Progression and tumor heterogeneity analysis in early rectal cancer. Clin Cancer Res (2008), 14: 772-781.
39. Kurashina K, Yamashita Y, Ueno T et al. Chromosome copy number analysis in screening for prognosis-related genomic regions in colorectal carcinoma. Cancer Sci (2008), 99: 1835-1840.
40. Laiho P, Hienonen T, Karhu A et al. Genome-wide allelotyping of 104 Finnish colorectal cancers reveals an excess of allelic imbalance in chromosome 20q in familial cases. Oncogene (2003), 22: 2206-2214.
41. Berg M, Agesen TH, Thiis-Evensen E et al. Distinct high resolution genome profiles of early onset and late onset colorectal cancer integrated with gene expression data identify candidate susceptibility loci. Mol Cancer (2010), 9: 100.
42. Carvalho B, Postma C, Mongera S et al. Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. Gut (2009), 58: 79-89.

43. Sjoblom T, Jones S, Wood LD et al. The consensus coding sequences of human breast and colorectal cancers. Science (2006), 314: 268-274.
44. Sanchez-de-Abajo A, de la Hoya M, van Puijenbroek M et al. Molecular analysis of colorectal cancer tumors from patients with mismatch repair proficient hereditary nonpolyposis colorectal cancer suggests novel carcinogenic pathways. Clin Cancer Res (2007), 13: 5729-5735.
45. Ollikainen M, Gylling A, Puputti M et al. Patterns of PIK3CA alterations in familial colorectal and endometrial carcinoma. Int J Cancer (2007), 121: 915-920.
46. Samuels Y, Wang Z, Bardelli A et al. High frequency of mutations of the PIK3CA gene in human cancers. Science (2004), 304: 554.

Genome-Wide Allelic State Analysis on Flow-Sorted Tumor Fractions Provides an Accurate Measure of Chromosomal Aberrations

Cancer Research (2008) 68(24):10333-10340

Chapter 7

Genome-wide Allelic State Analysis on Flow-Sorted Tumor Fractions Provides an Accurate Measure of Chromosomal Aberrations

Willem E. Corver, Anneke Middeldorp, Natalja T. ter Haar, Ekaterina S. Jordanova, Marjo van Puijenbroek, Ronald van Eijk, Cees J. Cornelisse, Gert Jan Fleuren, Hans Morreau, Jan Oosting, and Tom van Wezel

Department of Pathology, Leiden University Medical Center, Leiden, the Netherlands

Abstract

Chromosomal aberrations are a common characteristic of cancer and are associated with copy number abnormalities and loss of heterozygosity (LOH). Tumor heterogeneity, low tumor cell percentage, and lack of knowledge of the DNA content impair the identification of these alterations especially in aneuploid tumors. To accurately detect allelic changes in carcinomas, we combined flow-sorting and single nucleotide polymorphism arrays. Cells derived from archival cervical and colon cancers were flow-sorted based on differential vimentin and keratin expression and DNA content and analyzed on single nucleotide polymorphism arrays. A new algorithm, the lesser allele intensity ratio, was used to generate a molecular measure of chromosomal aberrations for each case. Flow-sorting significantly improved the detection of copy number abnormalities; 31.8% showed an increase in amplitude and 23.2% were missed in the unsorted fraction, whereas 15.9% were detected but interpreted differently. Integration of the DNA index in the analysis enabled the identification of the allelic state of chromosomal aberrations, such as LOH ([A]), copy-neutral LOH ([AA]), balanced amplifications ([AABB]), and allelic imbalances ([AAB] or [AAAB], etc.). Chromosomal segments were sharply defined. Fluorescence *in situ* hybridization copy numbers, as well as the high similarity between the DNA index and the allelic state index, which is the average of the allelic states across the genome, validated the method. This new approach provides an individual molecular measure of chromosomal aberrations and will likely have repercussions for preoperative molecular staging, classification, and prognostic profiling of tumors, particularly for heterogeneous aneuploid tumors, and allows the study of the underlying molecular genetic mechanisms and clonal evolution of tumor subpopulations. [Cancer Res 2008;68(24):10333–40]

Introduction

Chromosomal aberrations are common characteristics of human cancer and arise early during tumorigenesis. These aberrations are believed to be one of the driving forces behind tumor progression (1, 2). The process results in aneuploid cancer cells, which can be observed by genomic copy number abnormalities (CNA), allelic imbalances, loss of heterozygosity (LOH), and abnormal DNA content. Recently, a meta-analysis showed that these aberrations are associated with a worse prognosis in colorectal cancer (3). CNAs can be detected by array CGH (aCGH) or single nucleotide polymorphism (SNP) arrays (4–6). SNP arrays are preferred over aCGH because they additionally identify copy-neutral LOH (cnLOH; ref. 7), balanced CNAs (equal multiplication of both alleles), and allelic imbalances in which one allele is duplicated or amplified. For example, Kloth and colleagues showed that 75% of the LOH events found by SNP array were unnoticed using aCGH on the same cohort of cervical cancer cell lines (8). Furthermore, we and others have shown that reliable genotypes and profiles of CNAs and LOH can be generated from the fragmented DNA derived from formalin-fixed, paraffin-embedded (FFPE) cancer tissue using SNP arrays (5, 9–11).

The generated data is usually interpreted relative to the average DNA content of a tumor. For that reason, only relative copy numbers can be detected. Because tumors often show extensive genomic CNAs with almost doubled, near-tetraploid genomes (12), the extent of CNAs could be misinterpreted. For example, without knowledge of the DNA index, the CNA profiles from near-tetraploid tumors can be difficult to distinguish from those of near-diploid tumors. Consequently, only five or more copies will be interpreted as a gain, four copies will be misinterpreted as neutral and two or three copies as a loss. To improve copy number analysis, use of the DNA index has been suggested (13). Yamamoto and colleagues (14) analyzed cell lines and acute leukemia with limited CNAs and a mostly hyperdiploid DNA content (15). A recent study from Lyng and colleagues clearly showed the necessity for measuring the DNA index for calculation of absolute copy numbers from aCGH data (16). Also, varying proportions of normal cells (inflammatory and stromal cells) impair the detection of genomic and genetic alterations in tumor samples. For example, for the detection of LOH, samples should contain at least 50% tumor cells (17–19). Furthermore, solid tumors often contain subpopulations of tumor cells that harbor different chromosomal aberrations and may differ in their DNA index. Microdissection, either manually or by laser capture, is only a partial solution to the sampling problem in solid tumors because it is not possible to select tumor cells based on ploidy. Also, microdissection has a low cell yield and is prohibitively time-consuming when the neoplastic cells are highly intermingled with normal cells.

Fluorescence-activated cell sorting has proven to be an excellent tool for the purification of cell subpopulations from human tumors (20–23). Furthermore, since the development of robust dissociation methods in the early 1980s, flow cytometry has been widely used to determine the DNA content of solid tumors. Although these studies initially required fresh or frozen tumor samples, we developed a technique that markedly improved the yield and resolution of flow

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Willem E. Corver or Tom van Wezel, Department of Pathology, Leiden University Medical Center, P.O. Box 9600 L1-Q, 2300 RC Leiden, the Netherlands. Phone: 31-71526-6813/6580; Fax: 31-84833-1048; E-mail: w.e.corver@lumc.nl or t.van_wezel@lumc.nl.

doi:10.1158/0008-5472.CAN-08-2665

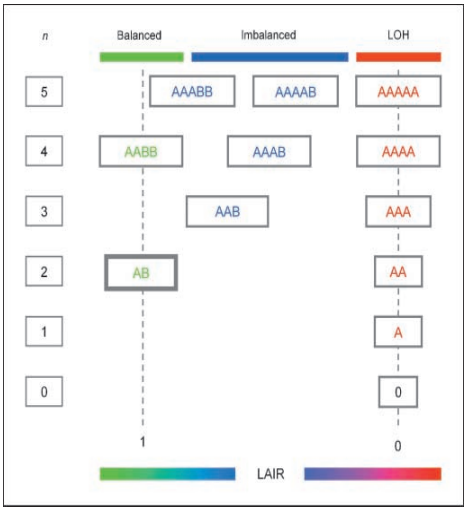

Figure 1. LAIR is a measure of the contribution of two informative alleles. LAIR is 1 when the contribution of both alleles of a certain SNP in the tumor, as compared with the total intensity, is similar to that of paired alleles of the reference sample (balanced, *left dotted line*; two copies [AB], four copies [AABB], etc.). LAIR is 0 when no signal is found for one of the alleles in the tumor (LOH, *right dotted line*; one or more copies [A], [AA], and [AAA], etc.). Allelic imbalances (imbalanced) are indicated by intermediate values depending on the copy number ratio between the two alleles: [AAABB], [AAAAB], [AAAB], and [AAB] are shown equidistantly (*n* = number of copies).

cytometric DNA content measurements of FFPE samples (24). This technique allows simultaneous flow-sorting of tumor and stromal cells based on differential expression of vimentin and keratin, as well as DNA content, and was successfully applied to study cervical, gastric, and colon cancers (25–27).

In the present study, we show that the combined use of multiparameter DNA flow-sorting and SNP array analysis significantly improves the detection of CNAs in archival FFPE cervical and colon cancers. For analysis, we used a novel algorithm, lesser allele intensity ratio (LAIR), which is incorporated in beadarraySNP (5). LAIR integrates the DNA index in the analysis and defines the allelic state of CNAs such as LOH (e.g., [A], cnLOH [AA], amplified LOH [AAA]), balanced amplifications (e.g., [AABB], [AAABBB]), and allelic imbalances (e.g., [AAB], [AAAB]). It provides a molecular measure of chromosomal aberrations which might serve as a clinical marker (28), and can be useful in preoperative molecular staging of rectal cancer (29).

Materials and Methods

Tumor dissociation. Cervical and colorectal tumors were obtained from the FFPE tissue bank of the Department of Pathology, Leiden University Medical Center (LUMC), Leiden, the Netherlands. Samples were handled according to the medical ethical guidelines described in the Code Proper Secondary Use of Human Tissue established by the Dutch Federation of Medical Sciences.[1] Paraffin sections taken from all samples were H&E-stained and reviewed by two pathologists (G.J. Fleuren and H. Morreau). Cell suspensions were prepared as described (24) from either 6 to 10 60-μm sections or 2 4-mm tissue punches from each paraffin block.

Antibodies. Clone MNF116 [anti–keratin 5, 6, 8, and 17, IgG_1 (DAKO)] was used at working concentrations of 2 μg/mL for 1×10^6 cells and 10 μg/mL for 5×10^6 cells. Clones AE1/AE3 [anti–pan-keratin, premixed 20:1, IgG_1 (Chemicon)] were used at working concentrations of 5 μg/mL for 1×10^6 cells and 25 μg/mL for 5×10^6 cells. Clone V9-2b (anti-vimentin, IgG_{2b}), originally developed at our department, was used as a diluted culture supernatant (1:5 or 1:1, depending on the cell concentration). Goat $F(ab_2)'$ anti-mouse IgG_1-FITC and goat $F(ab_2)'$ anti-mouse IgG_{2b}-RPE (Southern Biotechnology Associates) were both diluted 1:100 in PBATw.

Staining. One million cells were incubated with 100 μL of a monoclonal antibody mixture containing clones MNF116, AE1/AE3, and V9-2b overnight at 4°C. The next day, cells were washed twice with ice-cold PBATw and centrifuged at $500 \times g$ for 5 min at 4°C. The cells were then incubated with 100 μL of premixed FITC- or RPE-labeled secondary reagents. After 30 min on ice, cells were washed twice with ice-cold PBATw and incubated with 500 μL of DNA staining solution containing 10 μmol/L of propidium iodide (PI; Calbiochem) and 0.1% DNase-free RNase (Sigma) diluted in PBATw. Cells were kept at room temperature for 30 min to activate the RNase and were then incubated at 4°C overnight to allow for stoichiometric staining of the DNA.

For DNA index validation, two tissue blocks from an archival cervical carcinoma were taken and thick sections were cut at different time intervals and prepared for multiparameter DNA analysis as described. In total, nine independent measurements were performed, of which the DNA index and coefficient of variation (CV) of the G_0G_1 populations was calculated.

Flow cytometry and sorting. For analysis, data from 20,000 single cell events were collected using a standard FACScalibur (BD Biosciences) flow cytometer, equipped with a 15 mW Argon-ion laser (488 nm) and a 12 mW diode laser (635 nm; ref. 30). The FL3-A versus FL3-W pulse-processor was used to enrich for single cell events during acquisition and analysis. For data analysis, DNA index, and CV calculation, the WinList 6.0 and ModFit 3.1 software packages were used (Verity Software House, Inc.). N-color compensation was used for postacquisition spectral cross-talk correction according to the manufacturer's instructions, without the use of hyperlog transformation or log bias.

For flow-sorting, the cell concentration was increased to 5×10^6 cells/mL. The PI concentration was simultaneously increased to 50 μmol/L. G_0G_1 vimentin-negative, keratin-positive tumor cells and G_0G_1 vimentin-positive, keratin-negative stromal cells were flow-sorted using a FACSAria flow-sorter at 40 psi (BD Biosciences) with a 100-μm nozzle at a frequency of ~52 kHz. The 488 laser line was used for excitation. The FACSAria purity mode was used during sorting. These settings allowed us to typically flow-sort 800×10^3 cells in 5 mL Falcon tubes. The following detector and filter settings were used during sorting: FITC fluorescence, detector E, 530/30 nm BP filter; R-PE fluorescence, detector D, 575/26 nm BP filter; PI fluorescence, detector C, 610/20 nm BP filter. A detector C-Area versus detector C-Width dot plot was used to gate out doublet and aggregates during sorting. After sorting, cells were centrifuged at $4{,}000 \times g$ for 10 min before DNA was extracted.

For fluorescence *in situ* hybridization (FISH) analysis of flow-sorted cells (20 psi, 100 μm nozzle), samples were labeled for keratin, vimentin and DNA using APC- and RPE-conjugated antibodies, and 4′,6-diamidino-2-phenylindole as DNA stain. This approach reduced background fluorescence during the examination of the interphase nuclei after hybridization.

DNA isolation. DNA was isolated as described (31) and DNA was further purified using the Promega Protein Precipitation solution (Promega) according to the manufacturer's instructions. DNA concentrations were determined using the Picogreen method (Invitrogen).

SNP array analysis. SNP arrays were performed at the Leiden Genome Technology Center[2] as described (32) with minor modifications: 1 μg of DNA was used as the input in a multi-use activation step and was subsequently dissolved in 60 μL of resuspension buffer. Genotypes and the Gene

[1] http://www.federa.org/

[2] http://www.lgtc.nl/

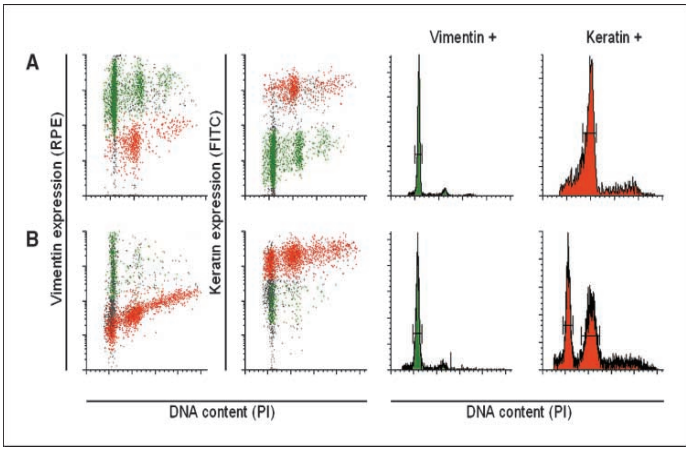

Figure 2. Flow-sorting of tumor cell subpopulations from cervical and colorectal carcinoma FFPE tissue. *A*, cervical cancer sample shows a unimodal DNA histogram after gating of the vimentin-negative, keratin-positive cell fraction. *B*, colorectal sample shows a bimodal DNA histogram after keratin gating, containing two DNA fractions: a near-diploid fraction with a DNA index of 0.97 and an aneuploid fraction with a DNA index of 1.87. Cell suspensions were simultaneously stained for keratin (epithelial cells, FITC), vimentin (stromal cells, R-PE fluorescence), and DNA (PI fluorescence). Sorting was restricted to the G_0G_1 populations (*vertical bars, right*).

Call Score were extracted using GeneCall version 6.0.7 (Illumina). Only genotypes with a GCS <0.5 were used to eliminate low-quality calls from this analysis.

Reference normal sample with the highest average quality score was chosen. LOH within a tumor sample was determined by comparison to the reference sample. Informative SNPs, SNPs at which the reference sample is heterozygous, were checked for homozygosity and LOH was indicated when two neighboring informative SNPs were homozygous in the tumor sample. The beadarraySNP package (5) was adapted to combine copy number profiles, allele-specific intensities, and the DNA index.

First, a segmentation procedure is applied to find genomic regions that have the same copy number (33). Secondly, LAIR, a measure of the contribution of the two original alleles, is calculated for all informative SNPs. This value is close to 1 when the contribution of both alleles of an SNP to the total intensity in the tumor is similar to that of the reference sample. The value is close to 0 when there is no signal for either of the alleles in the tumor (LOH). Allelic imbalances will show intermediate values (Fig. 1). By using the assumption that regions with LAIR close to 1 should have an even copy number and that the total calculated DNA index should be similar to the measured DNA index obtained by flow cytometry, it is possible to determine the allelic copy number in each genomic region. We refer to this as the allelic state. The following allelic states [modified from Nancarrow and colleagues (34)] can be distinguished: (*a*) [AB], normal; (*b*) [A], abnormalities with copy number 1, called LOH, resulting from the loss of either the A or B allele; (*c*) [AA], a diploid abnormality with either the genotype AA or BB, which is referred to as cnLOH; (*d*) [AAA] or [AAAA], etc., or amplified LOH, in which only the A or B allele is present at the locus and is present in three or more copies; (*e*) [AABB] or [AAABBB], etc., or balanced amplifications resulting in a 4n, 6n or higher genomic region with equal amounts of both alleles; and (*f*) allelic imbalances for [AAB], [AAABB], or [AAAABB], etc., for which the copy number at the locus is three or higher and both alleles are present at unequal amounts.

Generally, all SNPs on the array are equidistantly spaced across the genome. Thus, the average copy number of the allelic states of all SNPs, the allelic state index, was calculated and found to be an accurate measure of the copy number of the tumor. The allelic state index closely matched the DNA index measured by flow cytometry.

Interphase FISH analysis of flow-sorted cells. Interphase FISH analysis was performed as previously described (35) on 500 to 2,000 cells. The following probes were used: alphoid satellite centromeric probes for chromosome 4 (PYAM 11.39, kindly provided by Dr. A. von Bergh, Department of Clinical Genetics, Erasmus MC, Rotterdam, the Netherlands), chromosome 6 (p308), chromosome 8 (D8Z2), and chromosome 18 (L1.84; kindly provided by Dr. K. Szuhai, Department of Molecular Cell Biology, LUMC, Leiden, the Netherlands). The probes were biotin-16-dUTP–labeled (Roche Diagnostics) by standard nick translation. The BAC probes 149A7 (4q), 86C11 (6p), 10G10 (8q), 536K17 (8q), 748M14 (18q), and 154H12 (18qter; gift from Dr. K. Szuhai) were similarly labeled with digoxigenin-12-dUTP (Roche Diagnostics).

The centromere CEP17 Alpha SpectrumGreen and Vysis LSI TP53 SO SpectrumOrange (17p) probes were purchased from Abbot Molecular, Inc. Hybridization and immunodetection were performed as advised by the manufacturer with an additional denaturation step for 8 min at 80°C.

Results

Flow cytometry and sorting. Flow cytometric analysis of six solid tumors showed DNA histograms with two G_0G_1 fractions, which shows the presence of an aneuploid population in all cases. Representative DNA histograms and dot plots are shown in Fig. 2. After gating of the vimentin-positive, keratin-negative populations, a single population of tumor stromal cells remained in all cases (CV, 4.48 $\pm$ 0.78%). In the vimentin-negative, keratin-positive cell fraction of five cases, a single DNA aneuploid population was found. The DNA index ranged from 1.30 to 1.91 (Supplementary Table S1). The vimentin-negative, keratin-positive population of sample 5 showed two distinct populations with a near-diploid DNA index of 0.97 and an aneuploid DNA index of 1.86 (Fig. 2*B*; CV, 5.71 $\pm$ 1.30%). All vimentin-positive, keratin-negative cell populations, also those present in the normal samples (lymph nodes, endometrium) showed a single G_0G_1 fraction. The vimentin-positive, keratin-negative and the vimentin-negative, keratin-positive G_0G_1 fractions from all FFPE cancer samples were flow-sorted. From the normal samples, the vimentin-positive, keratin-negative G_0G_1 fractions were flow-sorted. Cell yields ranged from 0.5×10^6 to 2.5×10^6 and from 0.2×10^6 to 2.5×10^6 cells for the keratin-positive and vimentin-positive fractions, respectively (Supplementary Table S1). Independent measurements (n = 4 and n = 5) on two tissue blocks showed the robustness of the method. The DNA index ranged from 1.83 to 1.91 and the CVs from 3.19% to 5.81% for the G_0G_1 fractions.

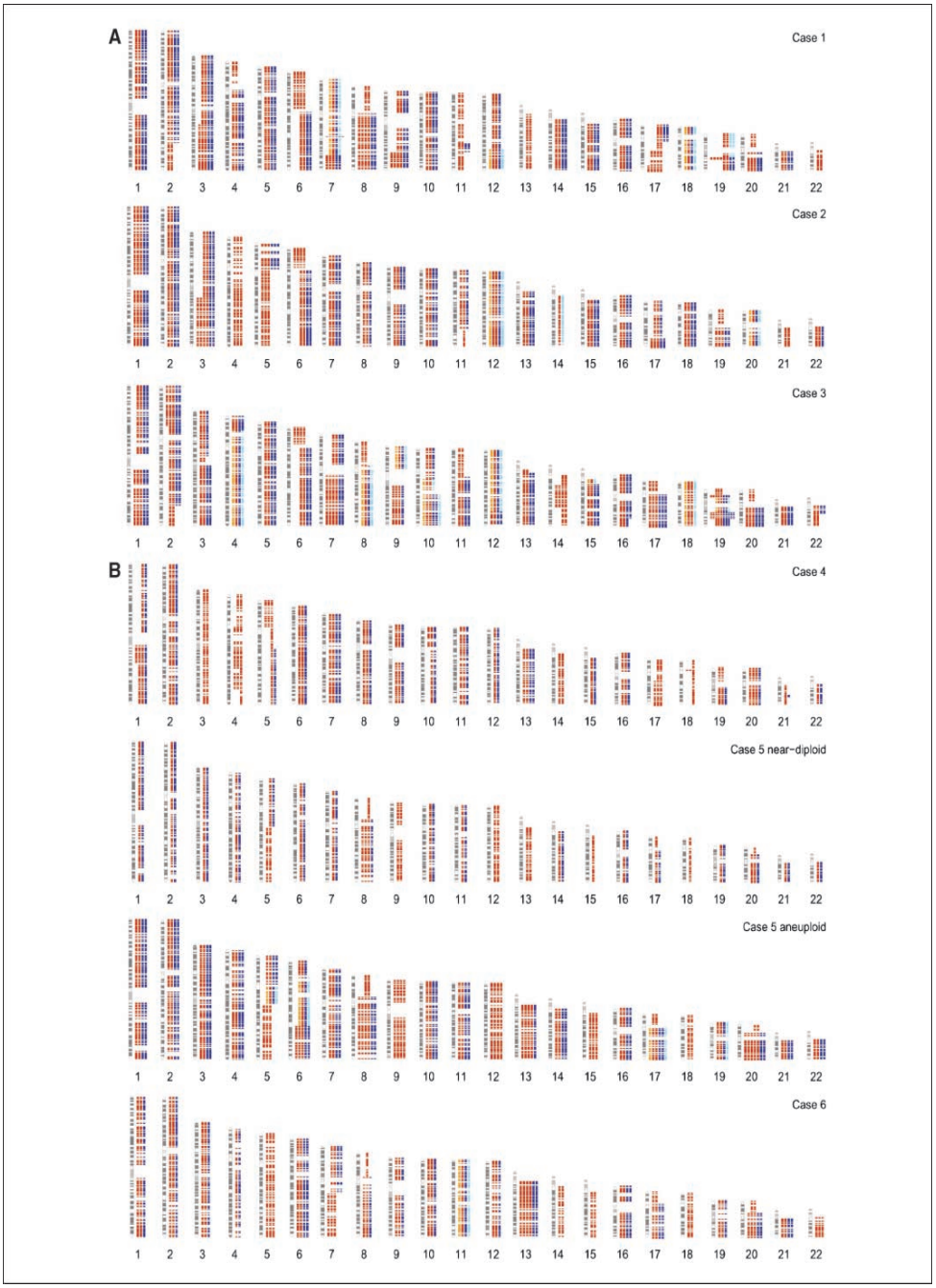

Figure 3. The allelic state of the autosomes for all cases provides a detailed measure of chromosomal aberrations. Cervical carcinomas, cases 1 to 3 (*A*); colon carcinomas, cases 4 to 6 (*B*). For each chromosome, the number of bar(s) to the right of the ideogram indicates the copy number. The bars are only depicted for informative SNPs (heterozygous in normal tissue). *Red*, the A allele; *blue*, B allele. *Orange* and *cyan bars*, clonal heterogeneity of this locus. For example; case 1, chromosome 7—a mixture of [AB] and [AABB] allelic states; see also Supplementary Fig. S2. Note the striking clonal relation between the near-diploid and near-tetraploid fractions of case 5. Besides endoreduplication of most chromosomes, including LOH ([AA]) on chromosome 5, the breakpoints on chromosomes 5, 8, 17, and 20 are identical.

Table 1. Summary of FISH analysis of flow-sorted G_0G_1 tumor cells

Locus	Probe	Case*						
		1	2	3	4	5, near-diploid fraction	5, aneuploid fraction	6
4q28.3	149A7	—	—	—	—	2/2 85%, 4/4 13%	4/4	—
6p25.1	86C11	—	3/3† 31%, 4/4 67%	—	—	2/2	4/4 34%, 4/2 57%	4/4
8q22	10G10	—	2/2 33%, 3/3 63%	—	—	2/2 52%, 4/4 31%, >4/>4 12% (52)	>4/>4‡	4/4 75%, 4/3 15% (20)
8q23.3	536K17	—	2/2 42%, 3/3 52%	—	—	2/2 86%, >4/>4 10%	>4/>4‡	4/4 85%, 4/3 10% (20)
Smad2	748M14	—	3/3 38%, 4/4 50% (24)	—	—	2/2 36%, 1/1 61%	2/2	2/2
18qter	154H12	—	3/3 42%, 4/4 58% (26)	—	—	2/2 50%, 1/1 43%	2/2	2/2
DNA index		1.84	1.91	1.82	1.30	0.97	1.86	1.54
Allelic state index		1.86	1.88	1.88	1.32	0.96	1.95	1.54

NOTE: The DNA index and the allelic state index are given for each fraction (—, not determined).
* One hundred nuclei were counted, unless otherwise noted in parentheses.
† Copy numbers are given as n/n representing centromere signal/probe signal.
‡ >4 = more than four copies.

Signal amplitudes and improved detection of chromosomal alterations. Using the beadarraySNP package, LOH profiles and relative copy numbers were generated for all cancers. Vimentin-positive, keratin-negative stromal cells from the archival tissue were used as a reference because these showed a high genotypic concordance (>98.8%) with paired normal fractions (lymph nodes or endometrium) and contained normal diploid genomes.

To evaluate the effect of flow-sorting for the macrodissected cervical tumors, the unsorted and flow-sorted tumor fractions were compared (Supplementary Table S2). Flow-sorting considerably improved signal amplitudes and identification of chromosomal segments with LOH (Supplementary Fig. S1). In total, 119 segments were identified: 24 with gains, 33 with physical losses; 12 showed cnLOH, and 50 showed retention. Of these segments, 92 were identical in the sorted and unsorted fractions, although 22 of the CNAs (31.8%) showed an increase in amplitude. Of all of the abnormalities, 16 were missed in the unsorted fraction (23.2%), whereas 11 were detected but were differently interpreted (15.9%). Five of these misinterpretations were probably due to tumor heterogeneity, e.g., a change from retention to gain or loss or from gain to cnLOH.

Integration of DNA index in CNA and LOH analyses. In addition to the percentage of tumor cells, the DNA index of the tumor cells will influence the interpretation of the copy number profiles. Therefore, we integrated the DNA index into the analysis. Based on the genotype and allele intensity of the SNP in the paired normal sample, we calculated the LAIR value and estimated the absolute allelic copy number or the allelic state of each chromosomal abnormality (see Materials and Methods) and provided an accurate molecular measure of chromosomal aberrations (Fig. 3) of these tumors. The allelic state index was almost equal to the DNA index measured by flow cytometry in all cases. FISH on flow-sorted tumor cells of cases 2, 5, and 6 validated the calculated copy numbers (Table 1).

The aneuploid cervical cancers (Fig. 3*A*) showed frequent aberrant allelic states, including LOH, on chromosomes 4p, 6p, 8p, and 11q, whereas gains on chromosomes 1q, 3q, 7q, 8q, 11q, 13q, and 17q were found in all three cases. Strikingly, all three cases were found to harbor four copies [AAAA] of a residual region on chromosome 6p (including HLA class I). LAIR analysis showed gains and losses involving chromosomes 5, 8, 13q, 17, 18, and 20 in the three colon cancers (Fig. 3*B*). Detailed analysis of all cases can be found in Supplementary Fig. S2 and Table S2. Chromosomal segments with CNAs were sharply defined. For example, the allelic imbalance [AAAAAABB] on chromosome 19 was restricted to a 3 Mb region.

LAIR analysis remarkably improved the detection and interpretation of CNAs for the aneuploid tumors. Figure 4 shows examples of chromosomes 8 and 18. For aneuploid case 6, through integration of the DNA index (1.54), on chromosome 8q, an allelic imbalance with three copies [AAB] was identified and on 8p LOH [A]. The copy number of chromosome 18 (n = 2, [AA]) was confirmed by interphase FISH (Fig. 4*A*). Similarly, the patterns of CNAs on chromosomes 8 and 18 in case 5 seem to be identical, but their allelic states differ after integration of the DNA index (Fig. 4*B* and *C*).

Intratumor heterogeneity. In case 5 (colon cancer), two vimentin-negative, keratin-positive tumor fractions with a different DNA index were clearly distinguished (0.97 and 1.86 for the near-diploid and aneuploid fraction, respectively; Fig. 2). To study their clonal relation, both fractions were analyzed for genomic aberrations. Fifteen chromosomes (1–5p, 7, 9–12, 14–16, 18, 20–22) were duplicated in the aneuploid fraction compared with the near-diploid fraction (Fig. 3; Supplementary Fig. S2). For example, the allelic states of chromosomes 1, 2, and 3 were duplicated from [AB] to a balanced gain [AABB] in the aneuploid fraction. Similarly, LOH [A] at 8p, 15, and 18 in the near-diploid fraction was duplicated to cnLOH [AA] in the aneuploid fraction. FISH confirmed the chromosome 18 copy numbers (Fig. 4*B* and *C*).

On chromosomes 8q, four copies with [AAAB] in the near-diploid fraction, and with [AAAAABB] in the aneuploid fraction were found, suggesting duplication with an additional loss of an A allele (Fig. 4*B* and *C*). Allelic imbalances [AAB] and [AAAAB] on chromosomes 6p and 13, respectively, in the aneuploid fraction, showed an unchanged [AB] and cnLOH [AA] allelic state in the near-diploid fraction (Supplementary Fig. S2). FISH confirmed the copy number of chromosome 6 in the near-diploid fraction

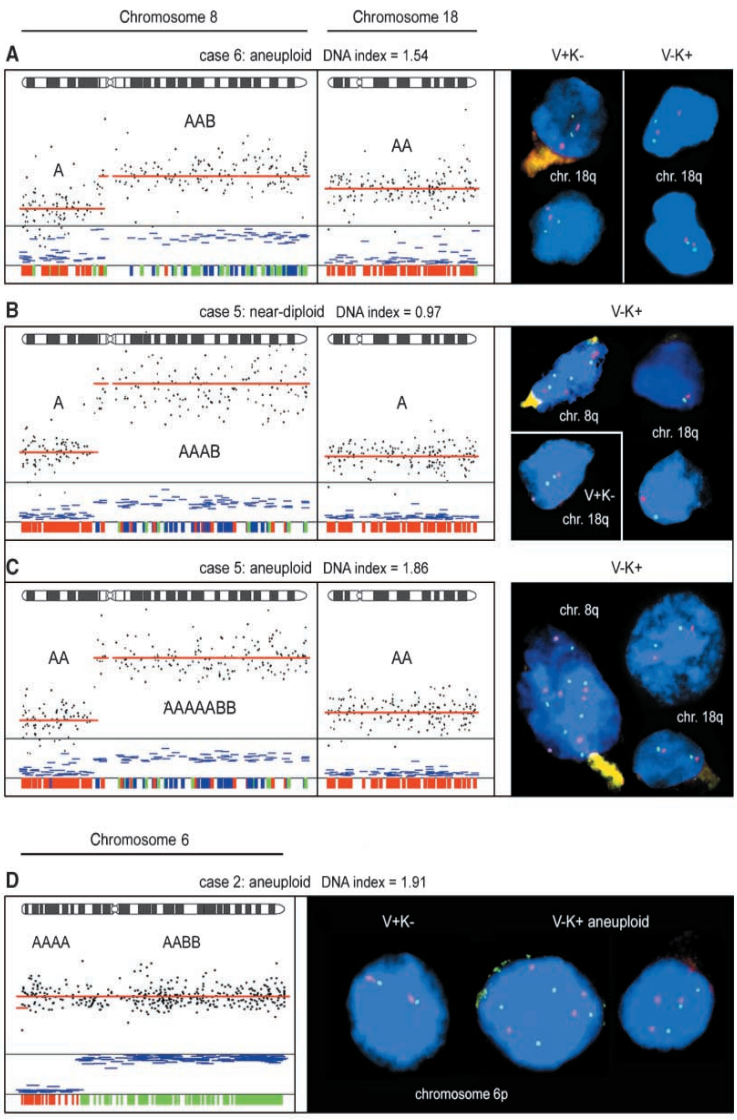

Figure 4. FISH confirmation of allelic state copy number and tumor heterogeneity. *A,* case 6, allelic state analysis of chromosomes 8 and 18 and interphase FISH for chromosome 18q on vimentin-positive, keratin-negative and vimentin-negative, keratin-positive nuclei. Two centromere 18 signals (*red*) and two signals (*green*) in the SMAD2 region are visible. *B,* case 5, near diploid (DNA index, 0.97) fraction. Copy numbers of the allelic states at 8q ([AAAB]) and 18 ([A]) are confirmed by FISH. On chromosome 8, four centromere signals (*red*) and four 8q signals (*green*) are visible, and on chromosome 18, one centromere signal and one 18q signal. *Inset,* vimentin-positive, keratin-negative interphase nucleus showing two centromere signals and two 18q signals. *C,* case 5, aneuploid fraction (DNA index, 1.86) and FISH for chromosomes 8 and 18 of the aneuploid fraction. Chromosome 8 shows seven centromere signals and seven 8q signals. Chromosome 18 shows two centromere signals and two 18q signals. Note the striking similarity between the chromosomal aberrations of different colon tumor fractions. Chromosomes 8 and 18 are from different OPA panels, causing the small difference in the level of the red segmentation line. *D,* intratumor heterogeneity observed by FISH. The allelic state of chromosome 6p of case 2 was [AAAA] according to LAIR analysis (a low LAIR score; *red bars*). FISH analysis of flow-sorted vimentin-negative, keratin-positive tumor cells showed that this population is composed of a mixture of two fractions: one fraction (67%) containing four copies of 6p25/4 centromeric copies and one fraction (33%) containing three copies of 6p25/3 centromeric copies. The vimentin-positive, keratin-negative fraction was shown to be normal (2/2, [AB]). [A], [AAB], etc., indicate the allelic state; *black dots,* normalized copy number with a red segmentation line for all SNPs; *horizontal blue dashes,* LAIR (calculated on informative SNPs) scale from 0 to 1; *vertical bars; green,* LAIR $\approx$ 1 (retention); *red,* LAIR $\approx$ 0 (LOH); *blue,* intermediate LAIR ($\sim$0.2 to $\sim$0.8, allelic imbalance). Probes: centromere 6, p308 (*red*); 6p, 86C11 (*green*); centromere 8, D8Z2 (*red*); 8q, 536K17 (*green*); centromere 18, L1.84 (*red*); 18q, 748M14 (*green*); 18qter, 154H12 (*green*).

(data not shown); however, FISH revealed a nearly equal mixture of 2n and 4n abnormalities on chromosome 6p in the aneuploid fraction (Table 1). These abnormalities may be a mixture of either [AA] and [AABB] or [AB] and [AAAB], leading to the detected [AAB]. Similarly, in case 2 (cervical squamous cell carcinoma), the allelic state estimate of chromosome 6p is [AAAA] (Fig. 3; Supplementary Fig. S2), whereas FISH analysis of the flow-sorted G_0G_1 aneuploid vimentin-negative, keratin-positive tumor fraction clearly revealed a mixture of a 3n and 4n population, with allelic states [AAA] in 31% and [AAAA] in 67% of the nuclei, respectively (Fig. 4*D*).

Discussion

We showed the feasibility of combining high-throughput SNP arrays and flow-sorting of tumor cell subpopulations from different formalin-fixed archival samples. Our approach significantly improved the simultaneous detection of numerical and structural chromosomal aberrations, allowing clear discrimination between allelic imbalances and LOH, and definition of aberrant chromosomal segments. Signal amplitudes of the losses and gains were generally higher when compared with those of the unsorted samples. Furthermore, <20% of the chromosomal abnormalities were missed in the unsorted cervical tumor fractions, whereas 16% were interpreted differently, either due to intratumor heterogeneity or to low tumor percentage. In the sorted samples, absolute LOH ([A]) or cnLOH ([AA]) was frequently identified, creating a sharp definition of a chromosomal segment.

We determined the allelic states of flow-sorted tumor fractions and confirmed copy numbers by FISH for several loci. The concordance between the DNA index and the allelic state index verified the method. In line with Nancarrow and colleagues (34), we propose the use of the following nomenclature: *I,* LOH for the loss of one allele, resulting in the allelic state [A]; *II*, cnLOH for the allelic state [AA]; *III,* amplified LOH for the allelic states [AAA], [AAAA], etc.; *IV*, balanced amplification for 4n [AABB] or 6n [AAABBB], etc.; and *V*, allelic imbalances for [AAB], [AAABB], [AAAABB], etc. We prefer allelic imbalance over amplification because allelic imbalance is more neutral and not suggestive of the molecular mechanism that was responsible for the observed abnormality. For example, an [AAB] status can be caused by a single gain of an A allele within an overall diploid genomic background; however, for a highly aneuploid tumor, this status can arise from endoreduplication of [AB] to an [AABB] status, followed by an additional loss of a B allele during tumor progression.

For accurate estimates of the allelic states of CNAs in FFPE aneuploid tumors, we showed that integration of the DNA index in conjunction with SNP arrays (8, 27) is crucial. Otherwise, most, if not all, CNAs will be misinterpreted, which is also supported by the findings of Lyng and coworkers on aCGH (16). For example, the patterns of CNAs on chromosomes 8 and 18 of the near-diploid and the aneuploid fractions of case 5 seem to be identical, but their allelic states are shown to differ after integration of the DNA index (Fig. 4). For near-diploid tumors, the allelic state of a tumor fraction will likely be correctly interpreted (14). Determination of the DNA index alone is feasible in most laboratories. However, for highly intermingled tumors (infiltrate and stroma) or for tumors containing different tumor subpopulations, flow-sorting seems to be mandatory.

An important potential of flow cytometry is the identification and sorting of multiple clones based on differences in DNA content. This advancement has not yet been achieved by any other enrichment or purification method. This approach allows for the study of intratumor heterogeneity, chromosomal aberrations which develop during tumor progression and clonal relationships between tumor subpopulations (20, 36). A clear example is case 5, in which the near-diploid and aneuploid tumor fractions probably originated from a common hypothetical near-diploid precursor fraction. Most chromosomes were endoreduplicated from [AB] in the near-diploid fraction to [AABB] in the aneuploid fraction (Fig. 3). Subsequent parallel divergence of the two clones resulted in the current fractions. For example, on chromosome 13, physical loss and mitotic recombination could have led from [AB] in the common precursor to cnLOH [AA] in the near-diploid fraction. The [AAAAB] status of chromosome 13 of the aneuploid fraction could have resulted from endoreduplication with subsequent loss of the B allele, and doubling of the A alleles.

In addition, the aneuploid fraction showed a mixture of two subpopulations for chromosome 6p, which were only identified after FISH analysis. At this locus, the near-diploid fraction showed a normal allelic state [AB]. Given the high DNA index (1.86), an average of three copies, an intermediate LAIR-score ([AAB]) and FISH analysis, these aneuploid subpopulations likely contained allelic states [AA] (cnLOH) and [AABB], respectively, clearly demonstrating heterogeneity within the tumor fraction. Similar intratumor heterogeneity within tumor cell fractions was shown for case 2 (Fig. 4*D*). These delicate differences in chromosomal copy numbers might indicate the ongoing generation of tumor subclones due to chromosomal instability (37). These differences also show that intratumor heterogeneity is more extensive than is generally observed by ploidy and LAIR analysis because both techniques register the dominant clone(s). Intratumor heterogeneity has also been identified by aCGH (16). However, LAIR analysis has increased value relative to classical methods. It allows for a more accurate estimate of the true allelic state of a chromosome or regions of chromosomes, which, for example, could result in discrimination between three chromosomal copies [AAB] and a balanced mixture of two allelic states, [AB] and [AABB].

Integration of copy number and ploidy analysis creates a detailed view of chromosomal aberrations during tumor progression. In all three aneuploid cervical carcinoma samples, LAIR analysis revealed amplified LOH, [AAAA], of the HLA region on chromosome 6p. This region is known to be frequently targeted by LOH in cervical cancer (8, 38, 39). From these events, it might be concluded that LOH on chromosome 6p occurred in a near-diploid precursor fraction, followed by mitotic recombination and endoreduplication, leading to aneuploidy. The identification of the chromosomal break points of the amplified LOH on 6p also indicates the possibility of accurately studying the smallest regions of overlap.

The consequences of allelic state analysis for the classification of tumors and profiling, as well as for allelic dosage determination of cancer-related genes, remain to be established in a large cohort of flow-sorted tumors. For example, in a near-diploid tumor with predominantly diploid chromosomes, a locus with the allelic imbalance [AAB] will be interpreted as a gain of the A allele, whereas in a tumor with a predominantly near-tetraploid [AABB] genomic background, [AAB] will be interpreted as an additional loss of a B allele after endoreduplication. One could speculate on the relation between the allelic dosage and RNA expression of the genes on segments with CNAs (40). For the [AAB] segment, the effect might be the opposite in a near-diploid background as compared with an aneuploid background.

In conclusion, we have explicitly shown that a combined approach of flow-sorting and high-throughput SNP array profiling with DNA index integration (*a*) significantly improves the detection of numerical and structural chromosomal aberrations in formalin-fixed tumor samples, (*b*) allows the definition of the allelic states of complex CNAs and, (*c*) provides an individual molecular measure of chromosomal aberrations. Interestingly, a recent meta-analysis showed that chromosomal aberrations are associated with a worse prognosis in colorectal cancer (3) and should be evaluated as a prognostic marker in clinical trials. Our new procedure could be implemented to define chromosomal aberrations and the allelic state of regions that are identified in whole genome association studies and which are associated with a cancer risk (41). Finally, our combined method may provide further insight into the dynamics of genomic instability during clonal evolution in heterogeneous human carcinoma samples and can be implemented in large-scale retrospective studies and studies in which detailed genome-wide information is required.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Received 7/11/2008; revised 9/11/2008; accepted 10/14/2008.

Grant support: Dutch Cancer Society grant UL2005-3247 (T. van Wezel and H. Morreau).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Yavuz Ariyurek for assistance with SNP arrays, Esther Lips for unsorted colorectal tumor samples, and Karoly Szuhai and Anne von Bergh for providing FISH probes.

References

1. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. Nat Genet 2003;33 Suppl:238–44.
2. Rajagopalan H, Lengauer C. Aneuploidy and cancer. Nature 2004;432:338–41.
3. Walther A, Houlston R, Tomlinson I. Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. Gut 2008;57:941–50.
4. Freeman JL, Perry GH, Feuk L, et al. Copy number variation: new insights in genome diversity. Genome Res 2006;16:949–61.
5. Oosting J, Lips EH, van Eijk R, et al. High-resolution copy number analysis of paraffin-embedded archival tissue using SNP BeadArrays. Genome Res 2007;17:368–76.
6. Geigl JB, Obenauf AC, Schwarzbraun T, Speicher MR. Defining "chromosomal instability". Trends Genet 2008; 24:64–9.
7. Nannya Y, Sanada M, Nakazaki K, et al. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. Cancer Res 2005;65:6071–9.
8. Kloth JN, Oosting J, van Wezel T, et al. Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex genetic alterations in cervical cancer. BMC Genomics 2007;8:53.
9. Lips EH, Dierssen JW, van Eijk R, et al. Reliable high-throughput genotyping and loss-of-heterozygosity detection in formalin-fixed, paraffin-embedded tumors using single nucleotide polymorphism arrays. Cancer Res 2005;65:10188–91.
10. Schubert EL, Hsu L, Cousens LA, et al. Single nucleotide polymorphism array analysis of flow-sorted epithelial cells from frozen versus fixed tissues for whole genome analysis of allelic loss in breast cancer. Am J Pathol 2002;160:73–9.
11. Jacobs S, Thompson ER, Nannya Y, et al. Genome-wide, high-resolution detection of copy number, loss of heterozygosity, and genotypes from formalin-fixed, paraffin-embedded tumor tissue using microarrays. Cancer Res 2007;67:2544–51.
12. Devilee P, Cornelisse CJ. Somatic genetic changes in human breast cancer. Biochim Biophys Acta 1994;1198: 113–30.
13. van de Wiel MA, Kim KI, Vosse SJ, et al. CGHcall: calling aberrations for array CGH tumor profiles. Bioinformatics 2007;23:892–4.
14. Yamamoto G, Nannya Y, Kato M, et al. Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. Am J Hum Genet 2007;81: 114–26.
15. Garcia-Sanz R, Orfao A, Gonzalez M, et al. Primary plasma cell leukemia: clinical, immunophenotypic, DNA ploidy, and cytogenetic characteristics. Blood 1999;93: 1032–7.
16. Lyng H, Lando M, Brovig RS, et al. GeneCount: genome-wide calculation of absolute tumor DNA copy numbers from array comparative genomic hybridization data. Genome Biol 2008;9:R86.
17. Lindblad-Toh K, Tanenbaum DM, Daly MJ, et al. Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. Nat Biotechnol 2000;18:1001–5.
18. Zheng HT, Peng ZH, Li S, He L. Loss of heterozygosity analyzed by single nucleotide polymorphism array in cancer. World J Gastroenterol 2005;11:6740–4.
19. Mei R, Galipeau PC, Prass C, et al. Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. Genome Res 2000;10:1126–37.
20. Bonsing BA, Corver WE, Fleuren GJ, et al. Allelotype analysis of flow-sorted breast cancer cells demonstrates genetically related diploid and aneuploid subpopulations in primary tumors and lymph node metastases. Genes Chromosomes Cancer 2000;28:173–83.
21. Geido E, Sciutto A, Rubagotti A, et al. Combined DNA flow cytometry and sorting with k-ras2 mutation spectrum analysis and the prognosis of human sporadic colorectal cancer. Cytometry 2002;50:216–24.
22. Giaretti W, Molinu S, Ceccarelli J, Prevosto C. Chromosomal instability, aneuploidy, and gene mutations in human sporadic colorectal adenomas. Cell Oncol 2004;26:301–5.
23. Paulson TG, Galipeau PC, Reid BJ. Loss of heterozygosity analysis using whole genome amplification, cell sorting, and fluorescence-based PCR. Genome Res 1999;9:482–91.
24. Corver WE, ter Haar NT, Dreef EJ, et al. High-resolution multi-parameter DNA flow cytometry enables detection of tumour and stromal cell subpopulations in paraffin-embedded tissues. J Pathol 2005;206:233–41.
25. de Boer MA, Jordanova ES, Kenter GG, et al. High human papillomavirus oncogene mRNA expression and not viral DNA load is associated with poor prognosis in cervical cancer patients. Clin Cancer Res 2007;13:132–8.
26. Carvalho R, Milne AN, Polak M, et al. Exclusion of RUNX3 as a tumour-suppressor gene in early-onset gastric carcinomas. Oncogene 2005;24:8252–8.
27. Middeldorp A, van Puijenbroek M, Nielsen M, et al. High frequency of copy neutral LOH in MUTYH-associated polyposis carcinomas. J Pathol 2008;216:25–31.
28. Leslie A, Stewart A, Baty DU, et al. Chromosomal changes in colorectal adenomas: relationship to gene mutations and potential for clinical utility. Genes Chromosomes Cancer 2005;45:126–35.
29. Lips EH, van Eijk R, de Graaf EJ, et al. Progression and tumor heterogeneity analysis in early rectal cancer. Clin Cancer Res 2008;14:772–81.
30. Corver WE, Koopman LA, van der AJ, et al. Four-color multiparameter DNA flow cytometric method to study phenotypic intratumor heterogeneity in cervical cancer. Cytometry 2000;39:96–107.
31. Abeln EC, Corver WE, Kuipers-Dijkshoorn NJ, Fleuren GJ, Cornelisse CJ. Molecular genetic analysis of flow-sorted ovarian tumour cells: improved detection of loss of heterozygosity. Br J Cancer 1994;70:255–62.
32. Fan JB, Oliphant A, Shen R, et al. Highly parallel SNP genotyping. Cold Spring Harb Symp Quant Biol 2003;68: 69–78.
33. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 2007;23:657–63.
34. Nancarrow DJ, Handoko HY, Stark MS, Whiteman DC, Hayward NK. SiDCoN: a tool to aid scoring of DNA copy number changes in SNP chip data. PLoS ONE 2007; 2:e1093.
35. Jordanova ES, Corver WE, Vonk MJ, et al. Flow cytometric sorting of paraffin-embedded tumor tissues considerably improves molecular genetic analysis. Am J Clin Pathol 2003;120:327–34.
36. De Angelis PM, Stokke T, Beigi M, et al. Chromosomal 20q gain in the DNA diploid component of aneuploid colorectal carcinomas. Int J Cancer 2007;120: 2734–8.
37. Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. Nature 1997;386: 623–7.
38. Koopman LA, Corver WE, van der Slik AR, Giphart MJ, Fleuren GJ. Multiple genetic alterations cause frequent and heterogeneous human histocompatibility leukocyte antigen class I loss in cervical cancer. J Exp Med 2000;191:961–76.
39. Chatterjee A, Pulido HA, Koul S, et al. Mapping the sites of putative tumor suppressor genes at 6p25 and 6p21.3 in cervical carcinoma: occurrence of allelic deletions in precancerous lesions. Cancer Res 2001;61: 2119–23.
40. Tsafrir D, Bacolod M, Selvanayagam Z, et al. Relationship of gene expression and chromosomal abnormalities in colorectal cancer. Cancer Res 2006;66: 2129–37.
41. Tuupanen S, Niittymaki I, Nousiainen K, et al. Allelic imbalance at rs6983267 suggests selection of the risk allele in somatic colorectal tumor evolution. Cancer Res 2008;68:14–7.

Concluding Remarks and Future Perspectives

Chapter 8

The aim of the work described in this thesis was to identify novel genetic risk factors for colorectal cancer (CRC). We applied several approaches to identify such novel CRC risk factors. Our approaches can broadly be divided into germ-line genetic analyses and somatic genomic analyses.

Using the germ-line approach, linkage analysis in seven large familial CRC families provided supportive evidence for region on 3q, which has previously been linked to CRC susceptibility. However, no novel regions of linkage were identified. Study of low-risk CRC susceptibility loci revealed an enrichment of risk alleles in familial CRC patients as compared to sporadic CRC patients. In solitary patients with an early age at onset of disease no such enrichment of risk alleles was observed. These results suggest that clustering of low-risk factors explains part of the excess risk observed in CRC families.

Somatic genomic analysis showed that carcinomas from *MUTYH*-associated polyposis (MAP) patients have a profile of aberrations that is distinct from sporadic CRC and from Lynch syndrome carcinomas. The most distinguishing factor is the high frequency of LOH in the absence of copy number alterations (copy-neutral LOH; see also page 15) in MAP carcinomas. In analogy to the distinct genomic profiles of Lynch carcinomas and MAP carcinomas, we studied the genomic profile of mismatch repair (MMR) proficient familial CRC. The profiles of these MMR proficient familial carcinomas show an increased frequency of 20q gain and an increased frequency of genome-wide cnLOH compared to sporadic CRC, while the overall profile largely resembles the profile of sporadic CRC. These results suggest an important role for 20q in tumor progression in familial CRC.

Germ-line genetic analyses

Familial CRC of which the underlying genetics are currently unknown is likely to represent a heterogeneous group, including both cases with a strong familial clustering of CRC which are likely to have an inherited basis as well as cases with a more sporadic form of CRC that aggregated in the family as a result of shared environment and lifestyles or simply by chance.[1,2] Using germ-line approaches, we searched for rare high penetrance risk factors and we studied the role of common low penetrance risk variants in CRC families.

Linkage analysis in seven large DNA mismatch repair (MMR) proficient CRC families did not provide a novel region of significant linkage that could harbor a high penetrance risk factor (chapter 3). However, our results support linkage to 3q21-q24, a region that has previously been identified as a CRC susceptibility locus.[3,4] Other regions that were previously reported to be linked to CRC susceptibility using linkage analysis, including 7q31, 9q22.2-31.2, 11q23.2 and 11q13.4, 14q24.2, and 22q12.1, were not supported by our linkage results.[5-9] This might be explained by differences in family ascertainment, since the other studies analyzed nuclear families or sib-pairs.

Three independent studies, including our study, now reported linkage of the 3q region to CRC susceptibility, with a smallest region of overlap encompassing 3q22.1-q22.3.[3,4,10] Together, these studies provide evidence for a novel CRC risk factor on 3q22. However, none of the

studies found strong evidence of linkage, suggesting that the risk of this locus is moderate. Mutation analyses of 46 genes in this region in previous studies did not identify pathogenic mutations.[3,4,10] The genes were screened for pathogenic mutations using a common approach by analyzing all exonic sequences, intron/exon boundaries, 5'- and 3'-untranslated regions (UTRs) and the promoter sequences of genes. The type of mutations that is generally searched for is truncating mutations. However, less obvious types of genetic alterations can be responsible for CRC predisposition. A recent study in CRC families, for example, showed that a heterozygous germ-line deletion of the last exons of *TACSTD1*, upstream of *MSH2*, causes epigenetic inactivation of *MSH2* in *TACSTD1*-expressing tissues and thereby predisposes these families to colorectal cancer.[10] These results demonstrate the potential profit of broader screening approaches for mutations, insertions, and deletions (see also future perspectives section below).

We performed association analyses to study the role of low-risk variants in familial CRC (chapter 4). Six loci that were identified in genome wide association studies (GWAS) were analyzed and the association with CRC risk could be replicated for five of these loci (located on chromosomes 8q23.3, 8q24.1, 11q23.1, 15q13.3, and 18q21.2). The odds ratios for these loci were increased, although not significantly, as compared to initial GWAS. This is likely a result of the familial nature of our cohort. The association between rs10795668 (10p14) and CRC risk was not observed in our Dutch familial CRC cohort, possibly due to either a lack of power or a population difference between our cohort and the English cohort in the initial GWAS.
Interestingly, we observed a significant increase in the number of risk alleles in cases compared with controls and an increase in odds ratio with increasing numbers of risk alleles. These results are in line with previous studies in unselected CRC patients.[12] Moreover, we observed an increased number of risk alleles in the patients with a family history of CRC as compared to solitary cases with an early age of onset of CRC, where no enrichment of risk alleles was observed. This shows that although low-risk alleles initially were thought to play a role in 'sporadic' CRC, they also play a role in FCC families. Therefore, clustering of low-risk variants may explain part of the excess risk in CRC families. Our results were recently confirmed by a study of Finnish familial CRC patients, that also observed an increased number of risk alleles in familial CRC patients compared to a group of sporadic CRC patients.[11] In solitary cases, however, other genetic models are likely to play a role; rare recessive high-risk variants might provide an explanation for their increased CRC risk, as reflected by a lack of affected first-degree relatives and their early onset of disease.
A further example of a role for common low-risk variants in familial CRC was provided by analysis of the seven large CRC families that we also studied with linkage analysis. We detected a significant association between rs16892766 and rs12953717 and CRC within these families (chapter 3). Moreover, two of the risk variants (rs16892766 and rs3802842) appeared to have a modifier role in Lynch syndrome families.[13] In line with these results, a meta-analysis of two large genome-wide association studies estimated that the ten low-risk variants that are

currently known, together account for 6% of the excess familial risk on basis of an additive model.[14]

Overall, our linkage analysis results and those results of others do not support a model in which a single highly penetrant gene explains the excess risk in familial colorectal cancer. All linkage analysis efforts over the last decade did not yet provide a novel high-risk factor, suggesting that the underlying genetics of the remaining familial CRC may be more complex. To further analyze the possibility of a high penetrance factor following a recessive model of inheritance, we analyzed the seven MMR proficient CRC families by homozygosity mapping. Very recently, preliminary results provided a candidate region on chromosome 5 that could harbor a CRC susceptibility locus. In ongoing and future studies, we will analyze this region in further detail.

Our association studies showed that low-risk variants may explain part of the excess risk in CRC families. Moderate risk factors might explain part of the excess risk in the remaining CRC families, but these are very difficult to identify with the current methods because of their assumed relatively low population frequency and their moderate penetrance. With the appearance of novel sequencing methods the identification of rare variants involved in CRC predisposition becomes feasible, for example using exome-sequencing or whole genome sequencing in highly selected cases (further discussed below in the future perspectives section).

However, while it is tempting to explain the excess CRC risk by genetic factors, the observed aggregation of CRC in large families could also be partly explained by environmental factors that are shared by family members. Moreover, the families could represent sampling artifacts and thus show aggregation of CRC largely by chance, although based on the cancer burden in these families, this seems less likely.

Also for other complex diseases, it was observed that much of the estimated heritability is not explained by the low-risk variants identified through genome wide association studies (reviewed by [15]). Several explanations for this so-called “missing heritability” were proposed. The missing heritability might be explained by risk variants with a low minor allele frequency (below 5%) which are not captured in genome-wide association studies and exert risks that are too low to be detected in linkage analyses. Structural variation like inversions or translocations might also partly explain the missing heritability. In addition, the estimated heritability could be an overestimation of the actual heritability of disease.[15] All these aspects could also apply to colorectal cancer.

The spectrum of CRC, including sporadic, hereditary, and familial CRC is summarized in Figure 1.

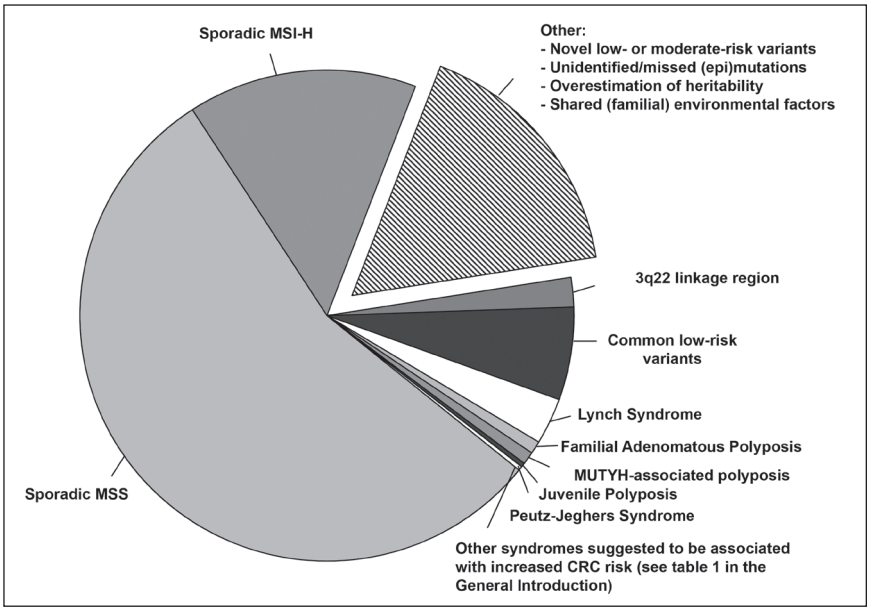

Figure 1. Spectrum of CRC.

In up to 35% of all CRC, hereditary factors play a role. About 5% of all CRC can be explained by known CRC syndromes. Common low risk variants account for 6% of the excess familial risk and linkage analyses identified 3q22 as a candidate region for a CRC susceptibility locus. In the remaining familial CRC cases other factors probably underlie the CRC susceptibility.

Several other approaches were applied by other groups to further study the nature of CRC susceptibility. Two recent studies analyzed homozygosity levels in colorectal cancer patients to study the role of consanguinity in CRC predisposition. Up till now, MUTYH-associated polyposis is the only known recessively inherited colorectal cancer syndrome. The results of both studies are inconclusive, since contradicting results were obtained. On one hand, the study of Bacolod et al. observed more and longer homozygosity regions in 74 colorectal cancer patients as compared to 264 controls, suggesting a role for recessively inherited CRC predisposition.[12] The number of homozygous regions (>4 Mb) per patient was low. The regions differed among the patients and were spread throughout the genome. On the other hand, a large study of Spain et al. did not provide evidence for increased homozygosity in CRC patients in a cohort comprising 921 cases and 929 controls.[13] The latter results show that the larger part of genetic CRC susceptibility likely follows a dominant mode of inheritance and that a minor part is explained by recessively inherited factors. A study of regions of homozygosity in specific subgroups of CRC patients in which a recessive inheritance mode is expected, such as solitary CRC patients with an early age of onset, might, however, be a good strategy to identify novel recessively inherited CRC risk factors.

The loci discovered in the genome-wide association studies inform on novel genes and/or pathways involved in CRC and might point to common molecular mechanisms involved in cancers.[14] Unraveling the biological mechanisms explaining the association between the identified risk loci and CRC will provide important novel insights in the etiology of colorectal cancer. In silico analyses to investigate the causality showed that many associated variants are in linkage disequilibrium (LD) with DNA sequence changes that influence gene expression rather than with nonsynonymous sequence changes that lead to altered proteins.[14] This might relate to the high population frequency and the low-risk these variants confer; different levels of expression likely exert more subtle effects than altered proteins.

Interestingly, five out of the ten identified low-risk factors are SNPs that are in linkage disequilibrium (LD) with genes of the TGF-β superfamily signaling pathway, including the genes *SMAD7*, *GREM1*, *BMP2*, *BMP4*, and *RHPN2*.[15] This further underlines the important role that the TGF-β pathway plays in CRC susceptibility; germ-line mutations in *SMAD4* and *BMPR2* were already known to be involved in Juvenile Polyposis and several members of the TGF-β pathway, including *SMAD4* and *TGFBR2*, are targeted by somatic mutations in colorectal tumors.[16,17]

An alternative approach to perform association studies as compared to using tagging SNPs, was adopted by Webb et al. who studied associations with gene-centric SNPs.[18] These gene-centric SNPs included 7000 genome-wide nonsynonymous SNPs, which alter the encoded amino-acid sequence. However, this study did not yield any significant association between CRC risk and any of the nonsynonymous SNPs, which a priori are more likely to have functional impact than synonymous SNPs. An explanation for the absence of associations could be that natural selection on alleles in coding regions has rendered the risk alleles rather rare.[18]

Somatic genomic analyses

In this thesis, we applied single nucleotide polymorphism (SNP) arrays to study colorectal tumors for genome-wide chromosomal copy number aberrations and loss of heterozygosity (LOH). Our aim was to generate a profile of genomic aberrations in MMR proficient familial CRC that might provide further insight in the biological basis of the increased CRC susceptibility in these families. Furthermore, these profiles might identify a candidate region that could harbor a CRC susceptibility factor.

Although many comparative genomic hybridization (CGH) studies of CRC have been performed over the last years (an overview is provided in chapter 1, table 2), most studies analyzed sporadic CRC, whereas few studies analyzed hereditary or familial CRC. Moreover, the majority of the studies used metaphase-based CGH or arrayCGH which only provides information on copy number aberrations. Genome-wide SNP arrays were used to a much lesser extent, even though these arrays provide information on both copy number and genome-wide LOH in the absence of copy number aberrations (copy-neutral LOH or cnLOH). In addition, we showed that, using SNP arrays, the genomic profiling of tumors can be further improved by analy-

zing flow-sorted tumor cells and incorporation of the DNA index in the analysis. We developed a novel algorithm, the lesser allele intensity ratio (LAIR), which can accurately determine the allelic state of all chromosomes. Upon incorporation of the DNA index of the tumors, LOH, cnLOH, balanced amplifications, and allelic imbalances can be distinguished (chapter 7). In addition to the assessment of allelic states, this method can address tumor heterogeneity.

The great value of genome-wide cnLOH analysis is illustrated by our study of MAP carcinomas (chapter 5). The main characteristic of these tumors was the high frequency of cnLOH; whereas physical loss occurred to a much lesser extent (Figure 2). This is in contrast to sporadic CRC, in which physical loss is frequent and few regions of cnLOH are observed.[19] The tumors from Lynch syndrome patients, which we studied previously, showed a characteristic profile, lacking gross chromosomal aberrations but only exhibiting a small region of copy neutral LOH around the locus of the mutated mismatch repair gene (Figure 2).[20] Also for many other cancers regions of cnLOH have described, including basal cell carcinoma and retinoblastoma (reviewed by [21]).

Compared to the unique and distinct profiles of genomic aberrations that were observed in MAP carcinomas and Lynch syndrome carcinomas, the group of MMR proficient familial colorectal carcinomas that was studied showed resemblance to sporadic CRC but with an increased frequency of 20q gain and genome-wide cnLOH (chapter 6). The most frequent aberrations in MMR proficient familial CRC included gains of chromosome 7, 8q, 13q, 20p and 20q, physical losses of 17p, 18p, and 18q. Remarkably, an increased frequency of 20q gain (77%) was observed as compared to the frequency in sporadic CRC (30-50%), in which it is considered to be an early event during tumorigenesis.[23,26] Moreover, an increased frequency of genome-wide cnLOH was observed at the expense of the frequency of physical losses in MMR proficient familial CRCs as compared to sporadic CRC. The observed high frequency of 20q gain in familial CRC confirmed a previous report on chromosomal aberrations in a Finnish cohort of familial CRC.[22] In this study 99 familial CRCs were compared to 186 sporadic CRCs using genome-wide allelotyping with microsatellite markers and copy number analysis using CGH on a subset of tumors. They observed gain of 20q in 85% of familial CRCs. They could, however, not confirm their results in a series of 67 familial and 96 sporadic CRCs from the UK.

Further analysis of chromosome 20q seems valuable and could include somatic sequence analysis of candidate genes or sequence analysis of the entire region. Gene expression analysis could provide information on differentially expressed genes. In a recent study, gene expression levels of genes located on 20q were compared between sporadic colorectal adenomas and carcinomas to identify oncogenes involved in adenoma to carcinoma progression. Several genes were found to be differentially expressed in carcinomas with gain of 20q as compared to carcinomas without such gain.[23] These genes could be also involved in familial CRCs, although the genetic targets could be very well be different from sporadic CRC. Moreover, profiles of imprinting on 20q could be studied, since for example *GNAS* (located on 20q13.32) is known to be regulated by complex tissue-specific imprinting patterns.[24]

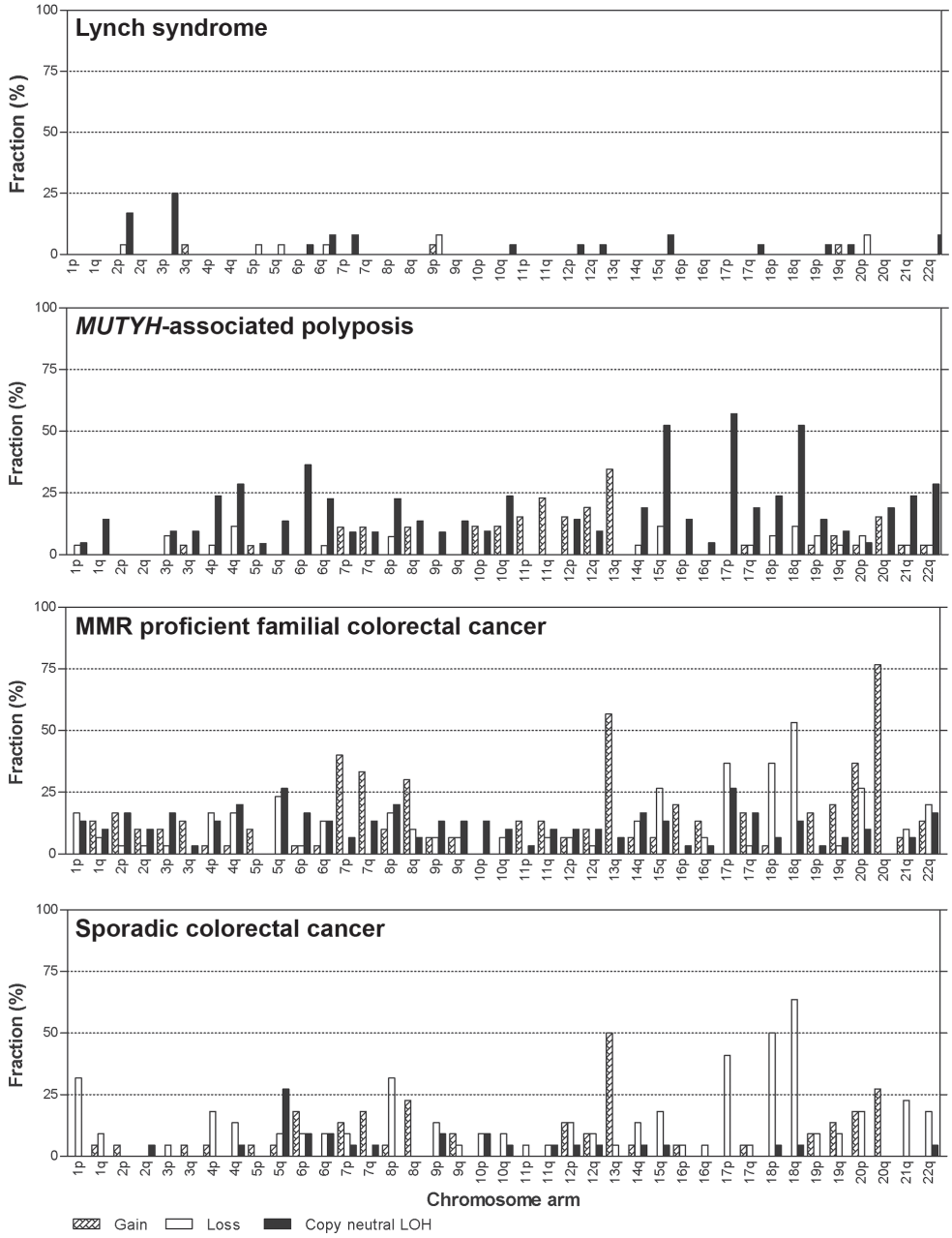

Figure 2. Profile of genomic copy number aberrations and cnLOH in colorectal cancer.

The bars indicate the percentage of carcinomas that exhibit an event of gain, loss or cnLOH of a chromosome. This percentage has been calculated for the respective chromosome arms. White bars, chromosomal gains; checked bars, physical losses of chromosomes; black bars, cnLOH. The profile of Lynch syndrome carcinomas (upper panel) is derived from Van Puijenbroek et al.[20]

Even though the genetic analysis of familial CRC did provide few distinct features for this group of MMR proficient familial CRC, tumor analysis provides a valuable approach to investigate CRC susceptibility. A successful example concerns the identification of mutations in *MUTYH* as the cause of polyposis (now termed *MUTYH*-associated polyposis), which came from the observation that tumors exhibited specific G>T DNA sequence transversions.[25] This provides a clear example of the value of tumor analysis of familial CRC. Additionally, before the identification of *BRCA2* as a breast cancer susceptibility gene, 13q deletions were already observed to be more frequent in familial breast cancer patients.[26]

Future perspectives

Future studies to identify novel genetic factors involved in CRC susceptibility, will require novel approaches because, despite of all efforts that were made over the last decades, a large part of the heredity remains unexplained. First, the search for novel genetic colorectal cancer risk factors has so far been focused mainly on genotypic variants. The study of copy number variants (CNV) may also yield novel CRC risk factors. As already discussed above, deletion of the last exons of *TACSTD1* is associated with CRC predisposition, because this deletion causes epigenetic inactivation of *MSH2* in *TACSTD1*-expressing tissues.[10] In addition, genomic (micro-)deletions showed already to be instrumental for the identification of cancer predisposing genes in for example retinoblastoma, Von Hippel-Lindau disease, and Wilms' tumor associated with WAGR syndrome (Wilms tumor, Aniridia, Genitourinary anomalies, and mental Retardation syndrome).[32-35] More recently, analysis of CNVs by array comparative genomic hybridization has led to the identification of a causative gene for example for CHARGE syndrome (Coloboma, Heart anomaly, choanal Atresia, Retardation, Genital and Ear anomalies syndrome).[36] In addition, association studies could be performed for CNVs that are polymorphic in the general population. Several associations between DNA copy number variants and common complex diseases have already been described (reviewed by [27]).

For example, a significant association was found between a low number of copies of a polymorphism in the human beta-defensin gene *HBD-2* (or *DEFB4*) and Crohn's disease, with a corresponding odds ratio of 3.06 (95% CI 1.46–6.45).[38] Other CNV associations were reported with HIV/AIDS susceptibility, rheumatoid arthritis, systemic autoimmune disease, systemic lupus erythematosus, psoriasis, and asthma, with odds ratios ranging from 1.34-5.27.[27]
Secondly, a useful next strategy could be to sequence the "exome" of colorectal cancer patients, including all protein encoding regions of the genome, or even to sequence the whole genome of CRC patients. Exome sequencing has already been applied successfully by Ng et al. on the exomes of 12 individuals.[28] Ng et al. studied the exomes of eight HapMap individuals and four unrelated individuals affected with Freeman-Sheldon syndrome, a rare autosomal dominant disease caused by mutations in *MYH3*. Filters were applied to identify the possible deleterious variant among all identified variants. Non-causal variants were removed by excluding variants that were not observed in one or more of the affected individuals.

Presumably common variants were removed by removing dbSNP catalogued variants and by removing the variants identified in the eight HapMap individuals. After application of these filters, *MYH3* was the only gene that was left on the candidate list.[28] The genetic homogeneity of the affected individuals and the availability of the HapMap individuals were important factors in the successful identification of *MYH3*. When applying this strategy for the identification of novel CRC susceptibility factors, probable genetic heterogeneity will have a significant impact on the performance and larger sample sizes will be required. Several genetic variants are likely to remain after filtering of the identified variants. To determine the significance of identified mutations, functional analyses could be performed.

Thirdly, in addition to a gene-centric approach, microRNA (miRNA) sequences could be studied for alterations. MiRNAs are small non-coding RNA sequences involved in post-transcriptional regulation of gene expression. Evidence for aberrant expression of miRNAs in human cancers is growing, indicating that they are involved in tumorigenesis (reviewed by [40]). However, both a global increase of miRNA levels in prostate cancer as global inhibition of miRNA processing have been described in cancer, suggesting a complex relation between miRNAs and tumorigenesis. A list of miRNAs involved in colorectal cancer has already been described.[29] Whether miRNAs are involved in CRC susceptibility still needs to be studied.

Fourthly, epigenetic changes or susceptibility to epigenetic changes might be involved in CRC predisposition and analysis of the epigenome could therefore be a fruitful approach. For example, germ-line methylation of the *MLH1* promoter region has been described in colorectal cancer patients.[42,43] However, the mode of inheritance of such epigenetic mutations remains unclear. Few examples of apparent inheritance of epigenetic states exist and it is generally believed that epigenetic modification are reset in germ cells.[30]

Finally, the role of the recently identified low-risk variants in tumor initiation (and progression) should be determined in future studies. This will provide important novel insights in the etiology of colorectal cancer and insight in the biological mechanisms involved in CRC susceptibility. Additional analyses need to be performed to identify the causal variants at the different loci and to unravel the biological mechanisms that cause the increased CRC risk. This will be a challenging task, since several of the identified risk alleles are located in regions that are without known genes (so-called gene deserts). Resequencing of the locus on 18q21.1 has identified Novel 1 to be the causal variant on this locus.[31] Variant rs6983267 is likely to be itself the causal variant, as determined by resequencing and linkage disequilibrium analysis in this region.[45] Further analyses of the region on 18q21.1, already showed that its causal allele (Novel 1) is associated with a reduced expression of *SMAD7* in a Xenopus laevis model.[31] The variant rs6983267 (8q24.1) is located in a transcriptional enhancer region that is bound by TCF7L2 (also referred to as TCF4), a transcriptional effector of the Wnt signaling pathway. The alleles of rs6983267 were found to differentially bind the transcription factor TCF7L2. It was shown that the region around rs6983267 physically interacts with *MYC*, but no robust association could be detected between rs6983267 and *MYC* mRNA expression.[46,47] However, a role for *MYC* is still likely, since *MYC* is a known target of TCF7L2 and is an important

oncogene in colorectal tumorigenesis. A strong indication for a role in somatic tumor evolution for rs6983267 was found by Tuupanen et al., who observed that in case of allelic imbalance at 8q24, the risk allele was favored in about two-thirds of the tumors.[32] Similar analyses for the other low-risk variants that were identified in recent genome-wide association studies will provide more and probably new insight into mechanisms of CRC initiation and progression.

In conclusion, the results described in this thesis suggest that it is unlikely that the excess risk in many of the MMR proficient familial CRC cases is explained by dominant high-risk genetic factors. Single young patients without a family history of CRC might be explained by a recessive origin of disease. In MMR proficient CRC families, one or more moderate risk factors might play a role. Research should therefore be directed more towards identifying novel factors conferring a moderate risk, even though these factors are more difficult to find. Recent advances in sequencing technology as well as novel knowledge of mechanisms involved in CRC development provided by the recently identified low-risk factors might facilitate the identification of novel moderate CRC risk factors.

Reference List

1. Castells A, Castellvi-Bel S, Balaguer F. Concepts in familial colorectal cancer: where do we stand and what is the future? Gastroenterology (2009), 137: 404-409.
2. Lichtenstein P, Holm NV, Verkasalo PK et al. 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.
3. Kemp Z, Carvajal-Carmona L, Spain S et al. Evidence for a colorectal cancer susceptibility locus on chromosome 3q21-q24 from a high-density SNP genome-wide linkage scan. Hum Mol Genet (2006), 15: 2903-2910.
4. Papaemmanuil E, Carvajal-Carmona L, Sellick GS et al. Deciphering the genetics of hereditary non-syndromic colorectal cancer. Eur J Hum Genet (2008), 16: 1477-1486.
5. Djureinovic T, Skoglund J, Vandrovcova J et al. A genome wide linkage analysis in Swedish families with hereditary non-familial adenomatous polyposis/non-hereditary non-polyposis colorectal cancer. Gut (2006), 55: 362-366.
6. Kemp ZE, Carvajal-Carmona LG, Barclay E et al. Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United kingdom. Cancer Res (2006), 66: 5003-5006.
7. Neklason DW, Kerber RA, Nilson DB et al. Common familial colorectal cancer linked to chromosome 7q31: a genome-wide analysis. Cancer Res (2008), 68: 8993-8997.
8. Skoglund J, Djureinovic T, Zhou XL et al. Linkage analysis in a large Swedish family supports the presence of a susceptibility locus for adenoma and colorectal cancer on chromosome 9q22.32-31.1. J Med Genet (2006), 43: e7.
9. Wiesner GL, Daley D, Lewis S et al. A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. Proc Natl Acad Sci U S A (2003), 100: 12961-12965.
10. Ligtenberg MJ, Kuiper RP, Chan TL et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. Nat Genet (2009), 41: 112-117.
11. Niittymaki I, Kaasinen E, Tuupanen S et al. Low-penetrance susceptibility variants in familial colorectal cancer. Cancer Epidemiol Biomarkers Prev (2010), 19: 1478-1483.
12. Bacolod MD, Schemmann GS, Wang S et al. The signatures of autozygosity among patients with colorectal cancer. Cancer Res (2008), 68: 2610-2621.
13. Spain SL, Cazier JB, Houlston R et al. Colorectal cancer risk is not associated with increased levels of homozygosity in a population from the United Kingdom. Cancer Res (2009), 69: 7422-7429.
14. Kingsmore SF, Lindquist IE, Mudge J et al. Genome-wide association studies: progress and potential for drug discovery and development. Nat Rev Drug Discov (2008), 7: 221-230.
15. Tenesa A, Dunlop MG. New insights into the aetiology of colorectal cancer from genome-wide association studies. Nat Rev Genet (2009).
16. Howe JR, Roth S, Ringold JC et al. Mutations in the SMAD4/DPC4 gene in juvenile polyposis. Science (1998), 280: 1086-1088.
17. Howe JR, Bair JL, Sayed MG et al. Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. Nat Genet (2001), 28: 184-187.
18. Webb E, Broderick P, Lubbe S et al. A genome-wide scan of 10 000 gene-centric variants and colorectal cancer risk. Eur J Hum Genet (2009).
19. Lips EH, van Eijk R, de Graaf EJ et al. Progression and tumor heterogeneity analysis in early rectal cancer. Clin Cancer Res (2008), 14: 772-781.
20. van Puijenbroek M, Middeldorp A, Tops CM et al. Genome-wide copy neutral LOH is infrequent in familial and sporadic microsatellite unstable carcinomas. Fam Cancer (2008), 7: 319-330.
21. Tuna M, Knuutila S, Mills GB. Uniparental disomy in cancer. Trends Mol Med (2009), 15: 120-128.
22. Laiho P, Hienonen T, Karhu A et al. Genome-wide allelotyping of 104 Finnish colorectal cancers reveals an excess of allelic imbalance in chromosome 20q in familial cases. Oncogene (2003), 22:

2206-2214.
23. Carvalho B, Postma C, Mongera S et al. Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. Gut (2009), 58: 79-89.
24. Kelsey G. Epigenetics and imprinted genes: insights from the imprinted Gnas locus. Horm Res (2009), 71 Suppl 2: 22-29.
25. Al Tassan N, Chmiel NH, Maynard J et al. Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. Nat Genet (2002), 30: 227-232.
26. Thorlacius S, Jonasdottir O, Eyfjord JE. Loss of heterozygosity at selective sites on chromosomes 13 and 17 in human breast carcinoma. Anticancer Res (1991), 11: 1501-1507.
27. Ionita-Laza I, Rogers AJ, Lange C et al. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. Genomics (2009), 93: 22-26.
28. Ng SB, Turner EH, Robertson PD et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature (2009), 461: 272-276.
29. Cummins JM, He Y, Leary RJ et al. The colorectal microRNAome. Proc Natl Acad Sci U S A (2006), 103: 3687-3692.
30. Cropley JE, Martin DI, Suter CM. Germline epimutation in humans. Pharmacogenomics (2008), 9: 1861-1868.
31. Pittman AM, Naranjo S, Webb E et al. The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. Genome Res (2009), 19: 987-993.
32. Tuupanen S, Niittymaki I, Nousiainen K et al. Allelic imbalance at rs6983267 suggests selection of the risk allele in somatic colorectal tumor evolution. Cancer Res (2008), 68: 14-17.

Summary

Nederlandse samenvatting

Curriculum Vitae

List of publications

Chapter 9

The aim of the work described in this thesis was to identify novel genetic factors that predispose to colorectal cancer (CRC). CRC is one of the most common malignancies in the Western world, currently affecting about 11,000 individuals in the Netherlands each year. The risk of developing CRC is influenced by both genetic and environmental factors. Inherited predisposition plays a role in up to 30% of all CRC, whereas in only about 6% of all cases the genetics underlying the increased cancer risk are known. Identification of novel genetic CRC risk factors will improve our insight in the etiology of the disease. Moreover, it allows identifying individuals at increased risk for CRC. These individuals can then be offered tailor-made colonoscopic surveillance schemes to detect precursor lesions, thereby preventing them to develop into a malignancy.

In chapter 1, a general introduction into the known genetic and environmental risk factors for CRC is provided. The high penetrance genes that give rise to hereditary CRC syndromes like Lynch syndrome and Familial Adenomatous Polyposis are described, as well as low level genetic risk loci for colorectal cancer. Chapter 1 also describes the different paths of tumorigenesis seen in colorectal cancer. The different forms of genetic instability are discussed and new fields in CRC research, prevention, and treatment are briefly reviewed.

Linkage analysis is a suitable method to identify high penetrance susceptibility loci in families affected with a disease. Traditionally, linkage analysis is performed with multi-allelic microsatellite markers. However, with the availability of high density SNP arrays, these arrays have been brought into use for linkage analysis. The information content of these dense arrays is higher as compared to microsatellite markers. Moreover, SNP arrays can be processed with higher throughput. The use of high density arrays in large families, however, yields a computational complex analysis that challenges existing linkage programs. In chapter 2, we developed a procedure for linkage analysis in large pedigrees using high density SNP arrays using existing software. We validated our procedure in a Lynch syndrome family with a known *MLH1* germ line mutation. In chapter 3, we applied this procedure to seven large colorectal cancer families in which known CRC syndromes had been excluded, to identify novel genetic regions linked to CRC predisposition. The linkage scan did not yield novel CRC susceptibility candidate regions, but our results support the previously reported linkage to a region on chromosome 3q. In addition, further analysis using homozygosity mapping recently revealed a candidate region for harboring a CRC susceptibility locus. We will analyze this region in further detail. To further explore genetic factors that could explain the increased CRC risk in the seven mismatch repair proficient CRC families, we also examined the presence of low risk variants in the families. Although no enrichment for low-risk variants could be observed in the families, two loci (8q23.3 and 18q21.1) were associated with CRC risk in the families. Collectively, our data indicate that the currently identified low-risk variants are insufficient to account for the type of familial clustering of CRC seen in the families we analyzed. Finally, analysis of the genomic tumor profiles of the affected family members revealed that these profiles resemble genomic profiles of sporadic colorectal cancer. Overall, these data suggest that factors other than a high-penetrance risk factor, such as low- or moderate-risk factors, may explain

the increased cancer risk in a subset of familial CRC.

In chapter 4, we studied the role of six CRC susceptibility loci (on chromosome 8q24.21, 18q21.1, 15q13.3, 11q23.3, 8q23.3, and 10p14) in a CRC cohort that was enriched for a positive family history of CRC and/or early onset of disease. We found an association with CRC risk for five out of the six susceptibility loci. In addition, we studied the relation between these risk alleles and clinical and pathological parameters, including gender, age at diagnosis, family characteristics, and tumor location. The locus on chromosome 18q21.1 appeared to be stronger associated with left-sided cancer as compared to right-sided cancer. Additionally, a stronger effect of this locus on CRC risk was seen for familial CRC cases with at least two first-degree affected relatives as compared to solitary CRC cases. Analysis of the number of risk alleles per individual revealed that the CRC risk increases with the possession of an increasing number of risk alleles. Furthermore, familial CRC cases carried significantly more risk alleles as compared to solitary CRC cases, suggesting that other causes of increased CRC risk, e.g. recessive factors, play a role in solitary cases. And cases from families with an early onset of disease carried significantly more risk alleles as compared to cases from families with a late onset of disease. Overall, our results in chapter 4 suggest a clustering of low-risk variants exists in familial CRC which is likely to contribute to the observed excess risk in relatives of patients.

Chapter 5 describes our results of the genomic profiling of carcinomas from *MUTYH*-associated polyposis patients. Although MUTYH deficiency triggers carcinogenesis by G:C>T:A transversions, the exact role of MUTYH deficiency in the tumor progression in MAP patients is still unknown. Therefore, we studied 26 MAP carcinomas for genome-wide copy number aberrations and loss of heterozygosity (LOH) using SNP arrays. Our results showed that these tumors mainly show copy-neutral LOH and less chromosomal losses, suggesting a relation between the base excision repair mechanism and mitotic recombination. The number of gains in the MAP carcinomas is similar to sporadic CRCs. Flow cytometry showed that most tumors had a near-diploid or near-triploid DNA content.

In Chapter 6 we applied the same approach to study the genomic tumor profile of patients with familial CRC. We studied 30 microsatellite stable carcinomas from 15 MMR proficient CRC families. Our aim was to generate a familial colorectal cancer profile of genomic aberrations. In addition, we studied the tumor profiles from family members to identify candidate regions that might harbor high or moderate penetrance risk factors. We observed an increased frequency of 20q gain and an increased frequency of genome-wide cnLOH in MMR proficient familial CRC, while the overall pattern of aberrations resembles sporadic CRC.

The detection of copy number aberrations and LOH in tumor samples is generally impaired by tumor heterogeneity, low tumor cell percentage and lack of knowledge of the ploidy status of the tumor. In chapter 7, we set up a novel approach to study chromosomal copy number aberrations and allelic imbalance in tumors. In our study, we combined flow sorting with SNP array analysis, which significantly improved the detection of chromosomal aberrations. Additionally, we developed a new algorithm, the lesser allele intensity ratio (LAIR), to accurately determine

the allelic (im)balances. Further incorporation of the ploidy status of the tumor enabled the identification of the allelic state of all chromosomal aberrations, including LOH, copy-neutral LOH, balanced amplifications, and allelic imbalances.

In chapter 8, concluding remarks and perspectives for future research are given. Collectively, the results presented in this thesis suggest that the increased risk in the remaining familial CRC is not explained by a single dominant high penetrance factor. Solitary young patients without a family history of CRC might be explained by a recessive origin of disease. In MMR proficient CRC families, one or more moderate- or low-risk factors might play a role.

Ieder jaar wordt er in Nederland bij ongeveer 11.000 patiënten dikke darmkanker gediagnosticeerd. Het is samen met borstkanker en longkanker een van de meest voorkomende kankersoorten in de Westerse wereld. Het gemiddelde risico om in de loop van het leven darmkanker te ontwikkelen bedraagt in Nederland ongeveer 6%. En ongeveer 45% van de patiënten overlijdt binnen 5 jaar na het stellen van de diagnose aan de ziekte.
Het risico op het ontwikkelen van kanker van de dikke darm wordt beïnvloed door erfelijke (genetische) factoren en omgevingsfactoren, zoals voeding. In studies met eeneiige en tweeeiige tweelingen is bepaald dat genetische factoren een rol spelen in 10%-30% van alle darmkanker patiënten. Er zijn verschillende syndromen bekend die darmkanker veroorzaken, maar deze syndromen kunnen samen slechts ongeveer 6% van het totaal aantal darmkankergevallen verklaren. Voor de overige patiënten met erfelijke darmkanker is onduidelijk welke genetische factor of factoren bij hen een verhoogd risico op darmkanker veroorzaakt. Deze groep patiënten wordt vaak aangeduid als familiaire darmkankerpatiënten.
Het onderzoek dat is beschreven in dit proefschrift had als doel nieuwe genetische factoren te identificeren die het verhoogde kankerrisico in familiare darmkankerpatiënten verklaren. Zeldzame genetische varianten die een sterk verhoogd risico op darmkanker veroorzaken zouden een rol kunnen spelen in families waarin veel familieleden zijn gediagnosticeerd met dikke darmkanker. In andere patiënten zou een combinatie van minder zeldzame genetische factoren die een klein verhoogd risico veroorzaken een verklaring kunnen bieden.
In dit proefschrift zijn verschillende methodes toegepast om zulke genetische risicofactoren te identificeren. Alle methodes hebben gemeen dat ze het erfelijk materiaal - het DNA - onderzoeken. Het DNA is te vergelijken met een bouwtekening voor het menselijk lichaam. Het DNA is verspreid over 23 verschillende chromosomen en elke cel bevat twee sets van 23 chromosomen (een set van vader en een set van moeder). In totaal bevat elke cel dus 46 chromosomen. Alle eigenschappen die een mens heeft, bijvoorbeeld de kleur van de ogen, maar ook het risico op bepaalde ziekten, staan “beschreven” in het DNA. Het DNA is een hele lange keten opgebouwd uit vier verschillende moleculen, die worden aangeduid met de letters A, C, G en T. De volgorde van de “letters“ van het DNA is voor alle mensen bijna identiek, maar er bestaan kleine – veelal onschadelijke – variaties (‘single nucleotide polymorfismen’, SNPs). Van deze polymorfismen is gebruik gemaakt bij het onderzoek. In hoofdstuk 3 werden circa 10.000 polymorfismen geanalyseerd, met behulp van SNP arrays, in zeven grote families die belast zijn met dikke darmkanker. In deze families werd de SNPs van de gezonde familieleden vergeleken met die van de aangedane familieleden met behulp van een zogenaamd koppelings-onderzoek of linkage analysis. Aangezien de analyse van 10.000 polymorfismen in grote families statistisch zeer complex is en veel computer capaciteit vergt, werd eerst een methode opgezet om deze analyses uit te voeren. Dit is beschreven in hoofdstuk 2. Een procedure werd opgezet om zulke analyses met gebruik van bestaande programmatuur die vrij te verkrijgen is uit te voeren. De ontwikkelde procedure werd gevalideerd in een familie met een bekende afwijking in het gen MLH1 gelegen op chromosoom 3. Deze afwijking veroorzaakt het Lynch Syndroom, waarbij individuen een sterk verhoogd risico op dikke darmkanker

hebben. De ontwikkelde procedure was in staat dit gen te identificeren als risicofactor in deze darmkanker familie. Toepassing van de procedure op zeven grote darmkanker families met onbekende onderliggende erfelijkheid leverde echter geen nieuw kandidaatgebied in het DNA op, dat een risicofactor voor dikke darmkanker zou kunnen herbergen. Het onderzoek leverde wel een bevestiging op van een eerder beschreven kandidaatgebied voor darmkankergevoeligheid op chromosoom 3q (hoofdstuk 3).

Om het verhoogde darmkankerrisico in deze families verder te verklaren, werd de mogelijke rol van zes laagrisicofactoren onderzocht. Deze zes risicofactoren zijn recent geïdentificeerd in grote genoombrede associatie studies (GWAS). Het zijn factoren die veel voorkomen in de algemene populatie en die slechts een klein verhoogd risico met zich meebrengen. Twee van deze factoren (op chromosoom 8q23.3 en chromosoom 18q21.1) waren significant geassocieerd met darmkanker in de families. Wanneer het aantal risicofactoren dat in de families voorkomt wordt vergeleken met controles, werd geen verrijking voor laagrisicofactoren gevonden in deze families. Tot slot werden de tumoren van de patiënten onderzocht op genetische afwijkingen. De genetische afwijkingen die in de tumoren werden waargenomen, komen sterk overeen met de afwijkingen die in sporadische tumoren ontstaan. Er werden echter ook andere afwijkingen geïdentificeerd.

Samenvattend, werd er geen bewijs gevonden dat één genetische hoogrisicofactor het verhoogde darmkankerrisico in deze families verklaard. De resultaten suggereren dat andere factoren, zoals risicofactoren die een laag tot matig risico met zich mee brengen, het verhoogde kankerrisico in deze families verklaren.

In hoofdstuk 4 werd de rol van zes laagrisicofactoren, gelegen op chromosoom 8q24.21, 18q21.1, 15q13.3, 11q23.3, 8q23.3 en 10p14, in een groep van 995 familiaire darmkankerpatiënten en 1340 gezonde controles bestudeerd. Al deze factoren, met uitzondering van de risicofactor gelegen op 10p14, waren significant geassocieerd met een verhoogd darmkankerrisico in de bestudeerde groep van familiaire darmkankerpatiënten. Er werd ook onderzocht of deze risicofactoren geassocieerd waren met klinische parameters als geslacht, leeftijd bij diagnose, locatie van de tumor in de darm en de familieanamnese. De risicofactor op chromosoom 18q21.1 was significant geassocieerd met tumoren aan de linkerzijde van de darm. Bovendien werd er een associatie gevonden tussen deze factor en patiënten met ten minste twee eerstegraads familieleden met darmkanker, afgezet tegen patiënten zonder eerstegraads familieleden met darmkanker (solitaire patiënten).

Ook werd vastgesteld dat het totaal aantal risicofactoren (allelen) dat familiaire darmkankerpatiënten hebben hoger is vergeleken met gezonde controles. Bovendien hadden patiënten met twee aangedane eerstegraads familieleden meer risico-allelen dan solitaire patiënten. Tot slot, hadden families waarin darmkanker werd gediagnosticeerd onder de leeftijd van 50 jaar meer risico-allelen dan families waarin de diagnose boven de 50 jaar werd gesteld. Al deze resultaten duiden erop dat een cluster van laagrisicofactoren een deel van het verhoogde darmkankerrisico in families kan verklaren, hoewel dit effect in zeven grote darmkankerfamilies (hoofdstuk 3) niet werd waargenomen.

Naast analyse van DNA uit de kiembaan, werd ook het DNA van darmtumoren onderzocht in dit proefschrift. Tumoren ontstaan door een opeenstapeling van foutjes in het DNA. Enerzijds ontstaan er kleine veranderingen van één “letter” op plaatsten in het DNA die de belangrijk zijn voor celgroei en celdeling. Anderzijds ontstaan er veranderingen waarbij grote stukken DNA, vaak zelfs hele chromosomen, verloren gaan of waarbij grote stukken DNA verdubbeld worden. Dit proces wordt chromosomale instabiliteit genoemd. Met behulp van de analyse van 10.000 SNPs (polymorfismen) werden in dit proefschrift het DNA van verschillende darmtumoren onderzocht. In hoofdstuk 5 werd het DNA van carcinomen van patiënten met *MUTYH*-geassocieerde polyposis (MAP) onderzocht. MUTYH is een eiwit dat een rol speelt bij het herstellen van fouten die ontstaan in het DNA bij celdeling. Bij MAP patiënten werkt dit eiwit niet meer naar behoren. Uit ons onderzoek is gebleken dat de tumoren van MAP patienten een uniek patroon van chromosomale instabiliteit hebben. De chromosomen in MAP carcinomen worden met name getroffen door verlies van heterozygotie (LOH) zonder dat er een verschil ontstaat in het aantal kopieën van dat chromosoom (kopie-neutrale LOH of cnLOH). Van het totaal aantal gevonden chromosomale afwijkingen betrof 71% cnLOH. Het aantal chromosomen waarvan een kopie verloren gaat was daarentegen laag. Dit profiel van afwijkingen is anders dan het profiel van sporadische darmtumoren, waarin juist veel verlies van chromosomen wordt waargenomen, echter maar weinig kopie-neutrale LOH. Het percentage afwijkingen waarbij een extra kopie van een chromosoom aanwezig is in de tumoren is ongeveer gelijk voor de MAP carcinomen en de sporadische carcinomen. Kopie-neutrale LOH werd in de MAP carcinomen voornamelijk waargenomen op chromosoom 17p (57%), 18q (52%) en 15q (52%). In sporadische tumoren gaat vaak een kopie van chromosoom 17p en 18q verloren.

Naast de analyse van chromosomale instabiliteit, werd ook de totale DNA inhoud van de tumor cellen geanalyseerd met behulp van ‘flow cytometrie’. Gezonde cellen bevatten van elk chromosoom twee kopieën en zijn daarmee diploïd. Van de MAP tumoren had circa de helft (52%) een diploïde DNA index, hetgeen betekent dat in elke cel van elk chromosoom gemiddeld twee kopieën aanwezig zijn. Bovendien werd er in acht MAP tumoren (35%) een bijna triploïde DNA index gemeten, dus van elk chromosoom gemiddeld drie kopieën. In sporadische darmtumoren is een triploïde DNA index ongebruikelijk.

Naar analogie van het unieke chromosomale patroon van MAP carcinomen en Lynch syndroom carcinomen (die eerder werden onderzocht in de context van het proefschrift van M. van Puijenbroek) werden de chromosomale profielen van familiaire darmkankerpatiënten met onbekende oorzaak onderzocht (hoofdstuk 6). Dertig tumoren afkomstig uit vijftien darmkanker families werden onderzocht. Het profiel van afwijkingen toonde veel overeenkomsten met de afwijkingen die in sporadische tumoren worden waargenomen, zoals een extra kopie van chromosoom 13 en verlies van chromosoom 17p en chromosoom 18p en 18q. Echter, een zeer hoog percentage (77%) familiaire tumoren had een extra kopie van chromosoom 20q, hetgeen ze onderscheid van sporadische tumoren waarin dit percentage lager is (30-50%). Dit hoge percentage tumoren met een extra kopie van chromosoom 20q suggereert dat deze

afwijking belangrijk is voor de progressie van familiaire darmtumoren. Daarnaast betrof in de familiare tumoren een verhoogd aantal afwijkingen kopie-neutrale LOH en daarmee minder verlies van chromosomen vergeleken met sporadische darmtumoren.
Naast chromosomale instabiliteit werden ook drie gebieden in het DNA (genen) die coderen voor eiwitten die belangrijk zijn voor celgroei en celdeling, *KRAS*, *BRAF* en *PIK3CA*, onderzocht op mutaties. De frequenties van de mutaties in *KRAS* (26%), *BRAF* (12%) waren vergelijkbaar met eerdere studies in familiaire tumoren. Er werden geen mutaties gevonden in *PIK3CA*. Samenvattend, wijzen de resultaten van de analyse van familiaire darmtumoren er op dat de chromosomale afwijkingen lijken of die van sporadische darmtumoren, maar met een verhoogde frequentie van een extra kopie van chromosoom 20q en een verhoogde frequentie van kopie-neutrale LOH.
Tot slot werd in hoofdstuk 7 de methode om chromosomale instabiliteit in tumoren te analyseren verfijnd. In tumorweefsel bevinden zich tussen de tumorcellen ook gezonde cellen, zoals cellen van het immuunsysteem. Wanneer er DNA wordt geïsoleerd uit tumorweefsel, wordt er dus ook vaak wat DNA uit gezonde cellen geïsoleerd. Bovendien bevinden er zich in een tumor soms twee groepen tumorcellen die ieder een andere genetische samenstelling hebben. Deze twee factoren vertroebelen het beeld, wanneer de chromosomale afwijkingen van tumoren worden geanalyseerd. In hoofdstuk 7 werd een methode opgezet om dit probleem te omzeilen. Met behulp van een ‘flow cytometrie’ is het mogelijk om de tumorcellen van de gezonde cellen te scheiden. Bovendien kan deze techniek verschillende groepen tumorcellen met een verschil in DNA index (DNA inhoud van een cel) van elkaar scheiden. Wanneer deze groepen tumorcellen vervolgens afzonderlijk worden geanalyseerd ontstaat een zuiver beeld van de chromosomale afwijkingen. Een additioneel voordeel van de zuivering van de cellen met behulp van ‘flow cytometrie’, is dat het de gemiddelde DNA inhoud van de cellen (DNA index) wordt gemeten met deze techniek. De DNA index maakt het mogelijk om het aantal kopieën van elk chromosoom nauwkeurig te schatten, terwijl zonder het betrekken van de DNA index in de analyse alleen relatieve winst en verlies van chromosomen is waar te nemen. Deze verfijning van de methode om chromosomale instabiliteit in tumoren te analyseren leidt dus tot een verbeterde analyse van chromosomale afwijkingen en maakt het mogelijk genetische mechanismen in de klonale evolutie van subpopulaties te bestuderen.
In hoofdstuk 8 worden een aantal concluderende opmerkingen gemaakt aangaande het onderzoek dat werd beschreven in dit proefschrift. Bovendien worden een aantal implicaties voor toekomstig onderzoek gegeven. Over het geheel genomen wijzen de resultaten van het onderzoek beschreven in dit proefschrift er op dat er niet één enkele hoogrisicofactor is die het verhoogde darmkankerrisico in familiaire darmkanker veroorzaakt. Een combinatie van een aantal laagrisicofactoren kan een deel van het verhoogde risico verklaren. Het overige deel van het verhoogde kankerrisico in familiaire darmkankerpatiënten is misschien te verklaren door nog te ontdekken genetische factoren die een laag of matig risico met zich mee brengen.

Curriculum Vitae

Anneke Middeldorp werd geboren op 17 mei 1980 te Deventer. In 1998 behaalde zij haar VWO diploma op Scholengemeenschap Alexander Hegius in Deventer. Daarna is ze Biomedische Wetenschappen aan Universiteit Leiden gaan studeren. Ze deed de master Communicatie van de studie Biomedische Wetenschappen, waar de praktijkstudie *Journalistiek en Nieuwe Media* een onderdeel van was.
Tijdens haar stage op de afdeling Pathologie van het Leids Universitair Medisch Centrum deed ze onderzoek naar methylering van *E-cadherine* in borsttumoren. Daarnaast heeft ze een stage gedaan in het Nationaal Natuurhistorisch Museum Naturalis, alwaar ze onderzoek deed naar de ontvangst van een tweetal tentoonstellingen door de bezoekers van het museum. In november 2004 heeft ze haar studie afgerond.
Na haar afstuderen werkte ze van december 2004 tot mei 2005 als projectmedewerker bij ZonMw in Den Haag. Ze heeft daar onderzocht hoe wetenschappers omgaan met de opmerkingen die ze tijdens de subsidieaanvraag over de opzet van hun onderzoeksproject hebben gekregen. In mei 2005 is Anneke Middeldorp als promovendus begonnen op de afdeling Pathologie van het Leids Universitair Medisch Centrum aan het project *Identification novel genes predisposing to colorectal cancer*. De resultaten van dit werk staan beschreven in dit proefschrift. Van oktober 2009 tot oktober 2010 was ze werkzaam als “medical writer” op de afdeling Pathologie van het Leids Universitair Medisch Centrum. Per 15 oktober 2010 werkt zij als (bio)medisch subsidieadviseur bij KCGroup te Amsterdam.

List of publications

van Roon EH, van Puijenbroek M, Middeldorp A, van Eijk R, de Meijer EJ, Erasmus D, Wouters KA, van Engeland M, Oosting J, Hes FJ, Tops CM, van Wezel T, Boer JM, Morreau H. Early onset MSI-H colon cancer with MLH1 promoter methylation, is there a genetic predisposition? BMC Cancer 2010, 5;10:180.

Middeldorp A, Jagmohan-Changur SC, Van der Klift HM, Van Puijenbroek M, Houwing-Duistermaat JJ, Webb E, Houlston R, Tops C, Vasen HFA, Devilee P, Morreau H, Van Wezel T, Wijnen T. Comprehensive Genetic Analysis of Seven Large Families with Mismatch Repair Proficient Colorectal Cancer. Genes Chromosomes Cancer 2010, 49(6):539-48.**

Middeldorp A, Jagmohan-Changur S, van Eijk R, Tops C, Devilee P, Vasen HF, Hes FJ, Houlston R, Tomlinson I, Houwing-Duistermaat JJ, Wijnen JT, Morreau H, van Wezel T. Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. Cancer Epidemiol Biomarkers Prev. 2009, 18(11):3062-7.**

Nielsen M, de Miranda NF, van Puijenbroek M, Jordanova ES, Middeldorp A, van Wezel T, van Eijk R, Tops CM, Vasen HF, Hes FJ, Morreau H. Colorectal carcinomas in MUTYH-associated polyposis display histopathological similarities to microsatellite unstable carcinomas. BMC Cancer 2009, 9:184.

Corver WE, Middeldorp A, ter Haar NT, Jordanova ES, van Puijenbroek M, van Eijk R, Cornelisse CJ, Fleuren GJ, Morreau H, Oosting J, van Wezel T. Genome-wide allelic state analysis on flow-sorted tumor fractions provides an accurate measure of chromosomal aberrations. Cancer Res. 2008, 68(24):10333-40.**

Wijnen JT, Brohet RM, van Eijk R, Jagmohan-Changur S, Middeldorp A, Tops CM, van Puijenbroek M, Ausems MG, Gomez GE, Hes FJ, Hoogerbrugge N, Menko FH, van Os TA, Sijmons RH, Verhoef S, Wagner A, Nagengast FM, Kleibeuker JH, Devilee P, Morreau H, Goldgar D, Tomlinson IP, Houlston RS, van Wezel T, Vasen HF. Chromosome 8q23.3 and 11q23.1 variants modify colorectal cancer risk in Lynch syndrome. Gastroenterology 2009, 136:131-137.

Middeldorp A, van Puijenbroek M, Nielsen M, Corver WE, Jordanova ES, ter Haar N, Tops CM, Vasen HF, Lips EH, van Eijk R, Hes FJ, Oosting J, Wijnen J, van Wezel T, Morreau H. High frequency of copy-neutral LOH in MUTYH-associated polyposis carcinomas. J Pathol. 2008, 216(1):25-31.**

van Puijenbroek M, Middeldorp A, Tops CM, van Eijk R, van der Klift HM, Vasen HF, Wijnen JT, Hes FJ, Oosting J, van Wezel T et al.: Genome-wide copy neutral LOH is infrequent in familial and sporadic microsatellite unstable carcinomas. Fam Cancer 2008, 7:319-330.

Middeldorp A, Jagmohan-Changur S, Helmer Q, van der Klift HM, Tops CM, Vasen HF, Devilee P, Morreau H, Houwing-Duistermaat JJ, Wijnen JT, van Wezel T. A procedure for the detection of linkage with high density SNP arrays in a large pedigree with colorectal cancer. BMC Cancer. 2007, 12;7:6.**

Lombaerts M, Middeldorp JW, van der Weide E, Philippo K, van Wezel T, Smit VT, Cornelisse CJ, Cleton-Jansen AM. Infiltrating leukocytes confound the detection of E-cadherin promoter methylation in tumors. Biochem Biophys Res Commun 2004, 319:697-704.

** This thesis