Department of Oncology-Pathology
Cancer Center Karolinska
Karolinska Institutet, Stockholm, Sweden

COLORECTAL CANCER

Genome, Transcriptome, and Proteome Dynamics

by

Jens Karsten Habermann

Stockholm 2005

Abstract

Colorectal carcinomas are the third most common malignancies in the Western World. Despite screening programs about 70% of these tumors are detected at advanced stages. Late diagnosis, however, results in a significant reduction of average survival times. Thus, alternative tools or biomarkers for early detection, prognosis and therapeutic intervention are needed.
Patients with ulcerative colitis (UC) have an increased risk to develop an ulcerative colitis-associated colorectal carcinoma (UCC). To evaluate independent cellular features with possible predictive value, a total of 683 UC biopsies from two patient groups were analyzed by image cytometry and immunohistochemistry: group A consisted of 8 patients with UCC, group B of 16 ulcerative colitis patients without any malignancy. Whereas no difference in inflammatory activity and dysplasia between patient groups could be observed, genomic aneuploidy, laminin-5 immunopositivity and increased Cyclin A expression could be confirmed as significant predictors for malignant transformation. We then analyzed 23 UC-associated colorectal carcinomas by image cytometry, comparative genomic hybridization (CGH), and immunohistochemistry. We found that DNA aneuploidy in UCC correlates with specific chromosomal imbalances and increased cyclin A expression levels. Distinct amplifications that were not known to be associated with colorectal cancer will likely reveal novel genes involved in colorectal tumorigenesis. However, the overall pattern of specific chromosomal aberrations in UCC is similar to that seen in sporadic colorectal carcinomas. To further elucidate the importance of these highly conserved chromosomal aberrations, we modeled specific chromosomal aneuploidies in cancer cells and dissected the immediate consequences of genomic imbalances on the transcriptome: we generated artificial trisomies in a diploid colorectal cancer cell line and in normal human breast epithelial cells using microcell-mediated chromosome transfer. Gene expression levels were analyzed using cDNA arrays. Regardless of chromosome or cell type, chromosomal trisomies resulted in a significant increase in the average transcriptional activity of the trisomic chromosome and affected the expression of numerous genes on other chromosomes as well, revealing a more complex global transcriptional dysregulation. Analogous results could be obtained in primary colorectal tumors: we found that specific and recurrent chromosomal aneuploidies exert strong and direct influence on gene expression levels of resident genes on the affected chromosomes by analyzing tissue samples from 36 patients with sporadic colorectal carcinoma. In addition, increasing genomic instability, aneuploidy and a recurrent pattern of chromosomal aberrations are accompanied by distinct gene- and protein expression patterns that correlate with subsequent stages of colorectal cancer progression. Our analysis identified 58 genes differentially expressed between normal mucosa and adenoma, 116 genes between adenoma and carcinoma, and 158 genes between primary carcinoma and liver metastasis. Proteomics analysis revealed 42 differentially expressed proteins that were related to similar canonical pathways as the differentially expressed genes. The identified proteins underwent extensive posttranslational modifications, thus multiplying the transcriptional dysregulation. A malignancy related protein profile could also be identified in serum samples: comprehensive serum protein profiling in 147 patients with colorectal malignancy and healthy individuals using SELDI-TOF mass spectrometry revealed a set of 13 differentially expressed features. Using 16 different classifiers allowed class prediction with 96.7% sensitivity and 100% specificity in a blinded validation set, including patient sera with early stage disease. Large scale prospective multicenter studies are now warranted to establish the clinical value of SELDI-TOF based protein profiling for colorectal cancer screening.

Cover: Composite CGH image of metaphase spread and the according karyogram of a lymph node metastasis from ulcerative colitis-associated colorectal carcinoma.

Published and printed by Larserics Digital Print AB, Sundbyberg.

ISBN 91-7140-411-2

To my Teachers and Mentors:

Karolinska Institutet and Hospital, Stockholm, Sweden:

Prof. G. Auer, MD, PhD

Genetics Branch, National Cancer Institute, NIH, Bethesda, USA:

Prof. T. Ried, MD

Dept. of Surgery, University of Schleswig-Holstein, Campus Lübeck, Germany:

Prof. H.-P. Bruch, MD
Prof. H. Schimmelpenning, MD, PhD
Prof. R. Broll, MD

To my Parents and Sisters:

Ronald and Gertrud Habermann
with Inga and Kristina

LIST OF PUBLICATIONS

This thesis is based on the following articles and manuscripts:

1. **Habermann J**, Lenander C, Roblick UJ, Krüger S, Ludwig D, Alaiya A, Freitag S, Dümbgen L, Bruch HP, Stange E, Salo S, Tryggvason K, Auer G, Schimmelpenning H. Ulcerative colitis and colorectal carcinoma: DNA-profile, laminin-5 γ2 chain and cyclin A expression as early markers for risk assessment. Scand J Gastroenterol. 2001 Jul;36(7):751-8.

2. **Habermann JK**, Upender MB, Roblick UJ, Krüger S, Freitag S, Blegen H, Bruch HP, Schimmelpenning H, Auer G, Ried T. Pronounced chromosomal instability and multiple gene amplifications characterize ulcerative colitis-associated colorectal carcinomas. Cancer Genet Cytogenet. 2003 Nov;147(1):9-17.

3. Upender MB, **Habermann JK**, McShane LM, Korn EL, Barrett JC, Difilippantonio MJ, Ried T. Chromosome transfer induced aneuploidy results in complex dysregulation of the cellular transcriptome in immortalized and cancer cells. Cancer Res. 2004 Oct 1;64(19):6941-9.

4. **Habermann JK**, Paulsen U, Roblick UJ, Upender MB, McShane L, Korn E, Krüger S, Bruch HP, Auer G, Ried T. The pattern of aneuploidy and genomic imbalances correlates with specific gene and protein expression changes along the adenoma-carcinoma-sequence of colorectal carcinogenesis. Manuscript.

5. Roblick UJ, Hirschberg D, **Habermann JK**, Palmberg C, Becker S, Krüger S, Gustafsson M, Bruch HP, Franzen B, Ried T, Bergmann T, Auer G, Jörnvall H. Sequential proteome alterations during genesis and progression of colon cancer. Cell Mol Life Sci. 2004 May;61(10):1246-55.

6. **Habermann JK**, Roblick UJ, Luke B, Prieto DA, Oevermann E, Schiedeck T, Veenstra TD, Burt SK, Bruch HP, Auer G, Ried T. SELDI-TOF mass spectrometry identifies specific m/z values that allow detection of colorectal malignancy in human serum. Manuscript.

Related Articles not included in this thesis:

1. Lenander C**, Habermann JK**, Ost A, Nilsson B, Schimmelpenning H, Tryggvason K, Auer G. Laminin-5 γ2 chain expression correlates with unfavorable prognosis in colon carcinomas. Anal Cell Pathol. 2001;22(4):201-9.

2. Lenander C, Roblick UJ, **Habermann JK**, Ost A, Tryggvason K, Auer G. Laminin 5 gamma 2 chain expression: a marker of early invasiveness in colorectal adenomas. Mol Pathol. 2003 Dec;56(6):342-6.

3. **Habermann JK**, Hellman K, Freitag S, Heselmeyer-Haddad K, Hellstrom AC, Shah K, Auer G, Ried T. A recurrent gain of chromosome arm 3q in primary squamous carcinoma of the vagina. Cancer Genet Cytogenet. 2004 Jan 1;148(1):7-13.

4. Korn EL, **Habermann JK**, Upender MB, Ried T, McShane LM. Objective method of comparing DNA microarray image analysis systems. Biotechniques. 2004 Jun;36(6):960-7.

ABBREVIATIONS

ABC = avidin-biotin peroxidase complex
AJCC = American Joint Committee on Cancer
ANCA = average number of copy number alterations
BSA = bovine serum albumine
Bp = base pair
cDNA = complementary DNA
CEA = carcinoembryonic antigen
CGH = comparative genomic hybridization
CID = collision-induced dissociation
CIN = chromosomal instability
CRP = C-reactive protein
Da = dalton
DAB = diaminobenzidine tetrahydro-chloride
DD-KNN = Distance-Dependent K-Nearest Neighbors
2-DE = two-dimensional gel electrophoresis
DNA = desoxyribonucleic acid
ES = electrospray ionization
FAP = familial adenomatous polyposis
FISH = fluorescence in situ hybridization
Gal file = gene annotation file
HNPCC = hereditary non-polyposis colorectal cancer
IAP = Ingenuity Analysis Pathways
KEGG = Kyoto Encyclopedia of Genes and Genomes
LDI = laser desorption ionization
LOH = loss of heterozygosity
MALDI = matrix-assisted laser desorption ionisation
MMCT = microcell-mediated chromosome transfer
MMR = mismatch repair deficiency
mRNA = messenger RNA
MSI = microsatellite instability
NSAID = nonsteroidal anti-inflammatory drugs
NW = network
PCA = principal component analysis
PCR = polymerase chain reaction
PSC = primary sclerosis cholangitis
Q-PCR = quantitative polymerase chain reaction
RNA = ribonucleic acid
SD = standard deviation
SDS-PAGE = sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis
SCC = sporadic colorectal carcinoma
SELDI-TOF = surface-enhanced laser desorption ionization time-of-flight
SKY = spectral karyotyping
TBS = Tris buffered saline
TNM = tumor–node–metastasis classification
TOF-MS = time-of-flight mass spectrometer
UC = ulcerative colitis
UCC = ulcerative colitis associated colorectal carcinoma
UICC = International Union Against Cancer
WHO = World Health Organization

INTRODUCTION

BACKGROUND

Colorectal cancer is one of the most common malignancies in the world. [1] While the five-year disease free survival rate for early stage tumors (UICC stage I) exceeds 90%, this percentage is reduced to 63% in advanced stage carcinomas (UICC stage III). [2] Therefore, detection of cancer at an early stage is critical for curative treatment interventions and utilization or application of tools and methodologies for early cancer detection can directly result in improving patient survival rates. In current clinical practice, screening for cancer and pre-invasive polyps of the colorectum is based on clinical examination, the detection of fecal occult blood (FOBT), and on sigmoidoscopy or colonoscopy. [3] The successful implementation of these screening procedures has contributed to a reduction in the mortality of colorectal carcinomas. [4] However, despite these screening programs, about 70% of carcinomas are detected at advanced tumor stages (UICC III/IV) presenting poor patient prognosis.

The lifetime risk for the development of colorectal carcinomas is considerably increased in patients with ulcerative colitis (UC). [5] Ulcerative colitis can therefore be considered a bona fide premalignant condition leading to the recommendation that patients with UC should participate in surveillance programs to screen for early signs of malignancy. [6] However, reliable surveillance is difficult and 50% of the detected malignancies are already at an advanced tumor stage. [7-9] Against this background, it was the aim of this thesis to further elucidate colorectal carcinogenesis in order to discover novel biomarkers relevant for improved diagnostics and therapeutics.

EPIDEMIOLOGY

Colorectal carcinoma is the fourth most common malignancy in the Western World. It ranks third as a cause of death from carcinoma, surpassed only by lung and prostate neoplasms in men, and lung and breast cancers in women. [10] Worldwide, almost 800,000 new patients are diagnosed with colorectal cancer per year with almost 450,000 dying of this malignancy. [1] The incidence per 100,000 amounts to 19.4 in men and 15.3 in women. [1] In asymptomatic individuals, the prevalence of any neoplastic lesions was found to be 37.5 %, e.g., 1% for invasive cancers and 7.9% for adenomas with a diameter of at least 10 mm or a villous adenoma. [11] The incidence of colorectal cancer shows a geographic variation with a 30- to 40-fold difference between the highest and lowest incidence rates. [12] For a comprehensive overview of detailed statistics please visit the National Cancer

Institute's website (http://www3.cancer.gov/atlasplus/). Interestingly, migrants from low-incidence areas succumb to the higher incidence rate of the host country within one generation. [13] This emphasizes the importance of environmental factors for colorectal cancer development, such as diet and lifestyle, which might explain the overall higher cancer incidence and mortality rates in Western industrialized countries. [14] The predominant location of colon cancer was thought to be the distal colon. However, the incidence of proximal (right-sided) colon cancer has been increasing. [15,16]

HISTOPATHOLOGY

More than 90% of all colorectal carcinomas are adenocarcinomas and can present, e.g., as mucinous or non-mucinous cell subtypes. [17] Other histological types of colorectal cancer are rare and include carcinoid tumors, lymphomas, and sarcomas. A different prognostic significance for the various histological types has been discussed but remains controversial. [18,19]

STAGING AND GRADING

The tumor stage has a high prognostic significance in regards to the outcome of the patient. [20] The first staging system was introduced by Dukes who classified the tumors into stage A, B, C, and D. [21] Hereby, stage A tumors are restricted to the bowel wall, stage B tumors penetrate the wall, stage C tumors exhibit lymph node metastases, and stage D tumors present with distant metastasis. Even though the original Dukes' classification was further modified and improved by Dukes and others, it is still hampered by the insufficient consideration of nodal sampling, extent of lymph node involvement, and tumor grade. [22-24] In contrast, the TNM classification of the American Joint Committee on Cancer (AJCC) and International Union Against Cancer (UICC) has proven its relevance in terms of routine practice and prognostic significance. [20] Tumors are thus classified according to the invasiveness of the primary tumor (T stage), the number of affected lymph nodes (N stage), and the occurrence of distant metastases (M stage). Carcinoma in situ (Tis) defines neoplasms confined to the lamina propria. T1 tumors reach the submucosa, T2 tumors reach the muscularis propria, T3 tumors infiltrate the subserosa, nonperitonealized pericolic or perirectal tissue, and T4 tumors invade other organs. The prefix "p" denotes the histopathological assessment (in contrast to the clinical/surgical one), and the prefix "y" defines tumors staged after neoadjuvant (presurgical) therapy. N0 denotes all examined lymph nodes to be negative. N1 defines tumors with one to three metastatic regional lymph nodes, N2 tumors with more than three affected nodes. Distant

metastasis is not apparent in stage M0, but becomes evident in stage M1. In addition, the residual tumor stage indicates if a tumor has been completely gross resected either with histologically negative margins (R0) or with histologically positive margins (R1). Tumors with incomplete gross resection fall into stage R2.
The histological grade has been divided into well differentiated (G1), moderately (G2), poorly differentiated (G3), or undifferentiated (G4). However, the distinction into two groups, low grade (G0 and G1) and high grade (G3 and G4) differentiated, seems to be of higher prognostic significance. [25]

Table: TNM classification of the American Joint Committee on Cancer (AJCC) and International Union Against Cancer (UICC)

UICC Stage	TNM classificAtion		
	T (tumor)	**N (lymph node)**	**M (metastasis)**
I	T1	N0	M0
I	T2	N0	M0
IIa	T3	N0	M0
IIb	T4	N0	M0
IIIa	T1 - T2	N1	M0
IIIb	T3 – T4	N1	M0
IIIc	Any T	N2	M0
IV	Any T	Any N	M1

DIAGNOSIS, PROGNOSIS, and SCREENING

Diagnosis

Various symptoms have been described as being associated with colorectal cancer. However, symptoms do not necessarily indicate the malignancy and can be subjective or nonspecific, such as changes in bowel movement and appetite, weakness, abdominal pain, or weight loss. [26] Gastrointestinal bleeding and especially acute obstructive symptoms or perforation are more severe signs of colorectal malignancy. [27] Unfortunately, symptoms develop when the malignancy is already rather advanced. Thus, patients having symptoms present more often with advanced stage tumors (in 53% UICC III or IV), while patients without symptoms harbor mainly early stage malignancies (in 85% UICC I or II). [27]

Prognosis

The most significant prognostic factor for patient survival is the tumor stage, in particular the invasiveness of the primary tumor (T stage), the number of affected lymph nodes (N stage), the completeness of surgical resection (R stage), and the occurrence of distant metastases (M stage). For a comprehensive overview and rank of prognostic factors please see the "College of American Pathologists Consensus Statement 1999". [20] While the five-year disease free survival for early stage tumors (UICC stage I) exceeds 90%, this percentage is reduced to 63% in advanced stage carcinomas (UICC stage III). [2] Therefore, detection of cancers at an early stage is critical for curative treatment interventions and it should be obvious that the tools and methodologies for early cancer detection directly impact on survival times.

Screening

Clinical practice currently employes five screening techniques for colorectal cancer: rectal digital exam, fecal occult blood testing (FOBT), sigmoidoscopy, colonoscopy, and barium enema as radiology diagnostics. All techniques show certain advantages and disadvantages. [3] The ***rectal digital exam*** can detect most of the rectal cancers, however it does not cover malignancies of the colon. ***Fecal occult blood testing*** (FOBT) has been shown to decrease the mortality from colorectal cancer by 33%. [28-32] However, the low specificity is a major limitation for this test. ***Flexible sigmoidoscopy*** and ***colonoscopy*** provide an advantage in the potential removal of precancerous lesions, such as adenomas, and in the subsequent histopathological assessment of biopsies taken from macroscopically suspicious regions. Thus, screening and treatment can be performed at the same time. [33] The use of sigmoidoscopy and colonoscopy have contributed to decreased mortalilty rates. [34] In addition, colonoscopy might need to be performed only in five year or higher intervals since the development from normal mucosa via adenoma to invasive carcinoma has been estimated to take approximately 7 to 10 years. [35] However, the detection of adenomas and even early stage carcinomas often depends on the bowel preparation and the experience of the examiner. ***Double-contrast barium enema*** allows screening of the entire colon and rectum. The technique itself, however, does not allow direct treatment interventions and requires appropriate equipment and expertise. [36]

The successful implementation of the above mentioned screening procedures has contributed to a reduction of the mortality of colorectal carcinomas. [4] However, colonic tumors still rank among the most common malignant cancers in the Western World and present a major health care problem today. The high mortality is attributable to a low compliance to some screening tests (e.g., colonoscopy) or to the low specificity of others (e.g., FOBT). [10] It is therefore essential to explore the

value of emerging technologies for early cancer detection for further reduction of mortality rates.
The efficiency of innovative technologies such as *virtual colonoscopy* or *magnetic resonance colonography* still needs to be confirmed. [37,38] However, these emerging techniques will probably not overcome all of the disadvantages known in the current screening techniques. The most desirable screening approach would need to be highly sensitive and specific, cost effective, fast, simple, and non-invasive, hence allowing high patient compliance. The assessment of carcinoembryonic antigen (CEA) in serum as a prognostic marker was hopeful but it has been unable to fulfill its expectations as a simple test for early diagnosis of colorectal cancer. [39,40] Therefore, it is promising that the employment of functional genomics and proteomics will reveal malignancy specific genes, proteins or specific profiles which can be identified in stool, blood, or serum to detect colorectal cancer at its earliest stage.

ETIOLOGY

The etiology of colorectal carcinomas is complex and related to the interaction of environmental and genetic factors. [41] The most prominent environmental factors known today are diet, lifestyle and the use of nonsteroidal anti-inflammatory drugs (NSAIDs). Obesity, total caloric intake, consumption of red meat, and cigarette smoking have been found to be associated with increased colorectal cancer risk. [42-45] Colonic microbiota have been recently implicated with colorectal cancer development. [46,47] A high-fiber diet was believed to be protective against colorectal cancer development but recent studies were not able to verify this inverse correlation. [42,48] In contrast, vegetable and fruit consumption, the use of NSAIDs, and physical activity have been found to decrease the risk for colorectal cancer. [43,45,49-51] Genetic factors play an important role in the pathogenesis of colorectal cancer and can be inherited or acquired.

Pathogenesis of Colorectal Cancer

The normal colonic epithelium (mucosa) is a highly dynamic system: Stem cells are located at the basis of epithelial crypts. [52] They are a source of constantly proliferating cell population that – while differentiating - migrate to the surface of the colonic crypts from where they shed into the lumen. The intestinal epithelium is thus renewed every 5 to 6 days. Mucosal cells are prone to genetic damage due to the highly toxic and mechanically stressful intra-luminal environment. The rapid clearance of mucosal cells however prevents these cells from being a source of malignant transformation. However, a high proliferative rate in a toxic environment

could also easily accelerate malignant transformation once regulatory mechanisms for cell homeostasis are bypassed. Most colorectal tumors are caused by acquired genetic lesions of single mucosal cells that harbor growth advantage and – through clonal expansion – rise to invasive carcinomas. Whether these genetic aberrations occur predominantly in stem cells, migrating cells, or mucosal cells at the crypt surface has not been conclusively clarified. [52-55] Genetic aberrations can become evident either on the subchromosomal or chromosomal level and target regulatory mechanisms required for the genetic equilibrium such as cell cycle regulation, cellular signaling pathways, proliferation, differentiation, growth inhibition, and apoptosis signaling. Vogelstein and colleagues defined a model of colorectal carcinogenesis in which a non-random accumulation of genetic aberrations can be correlated with morphologic changes of the colon epithelium: the transition from normal mucosa via adenomatous polyp to colorectal cancer and eventually distant metastasis. [56]

The Chromosomal Instability (CIN) Pathway

The "adenoma-carcinoma-sequence" describes one of two identified genetic pathways leading to colorectal cancer. The fundamental genetic aberration in this pathway is a mutation of the *adenomatous polyposis coli* (APC) gene, which is located on chromosome 5q21. The APC gene normally initiates degradation of the β-catenin oncoprotein and is thus considered to be a tumor-suppressor gene. [57] Mutation in the APC gene leads to loss of its gatekeeper function and subsequent chromosomal instability (CIN). [58] CIN can already be detected in premalignant lesions, e.g., adenomas, as trisomy of chromosome 7. [59,60] The APC mutation is present in 80% of adenomas and is accompanied by mutations of the Ras oncogene in 40% to 50% of adenomas that progress to invasive disease. [57] The Ras mediated pathways trigger cell proliferation, growth, and transformation. In addition, invasive carcinomas lose the tumor-suppressor-function of the p53 gene that guards cell-cycle progression by initiating cell cycle arrest or apoptosis once DNA damage has occurred. [61,62] Additional tumor-suppressor genes acquire loss of function during the progression from adenoma to invasive and metastatic disease (e.g., DCC, SMAD4). [57]

The mutation of the APC gene can be inherited autosomal dominant with almost 100% penetrance and causes the ***familial adenomatous polyposis*** (FAP) syndrome. [63] FAP patients characteristically develop hundreds to thousands of adenomas and – if the colon is not surgically removed – develop colorectal cancer. [64] The affected patients inherit a mutated copy of the APC gene that causes early-onset polyposis. The additional inactivation of the remaining APC allele then leads inevitably to colorectal cancer development. The chromosomal instability pathway

via APC gene mutation also accounts for approximately 80% of all ***sporadic colorectal carcinomas.*** [65]

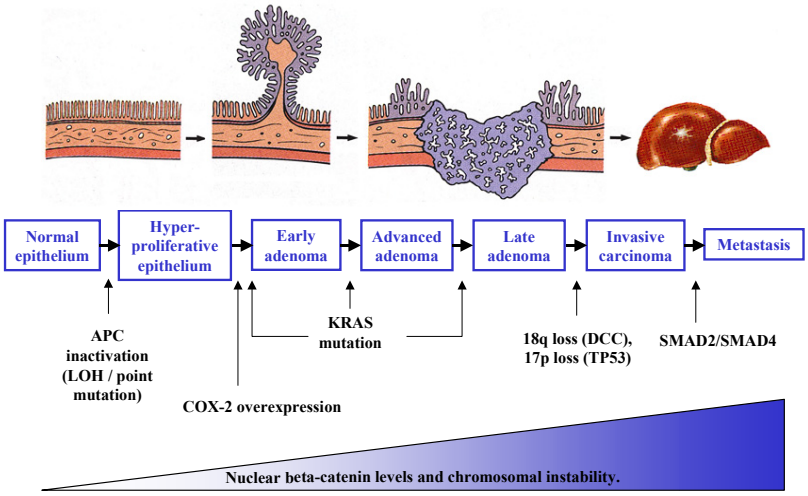


Figure: Genomic alterations during sequential carcinogenesis of the colorectum, the "adenoma-carcinoma-sequence".

The Microsatellite Instability (MSI) Pathway

In contrast to the chromosomal instability pathway, genetic aberrations of genes within the microsatellite instability pathway result in DNA replication errors, known as microsatellite instability. Microsatellites are repetitive DNA sequences that are the target of mutations caused by deficiency in the ***mismatch repair pathway***. [66,67] Most mutations occur in the hMLH1 and hMLH2 gene but mutations of other members of this family have been described (hMSH6, hPMS1, hPMS2). [68,69] Such mutations are inherited autosomal dominant and occur in patients with the ***hereditary non-polyposis colorectal cancer*** (HNPCC) syndrome. [70] HNPCC accounts for approximately 3% of all colorectal cancers. The syndrome is clinically defined by the Amsterdam criteria. [71] About 10% of all sporadic colorectal carcinomas are related to the MSI pathway. [72]

Inherited, Familial and Non-familial Causes of Colorectal Cancer

In addition to FAP and HNPCC, other additional rare syndromes exist with an increased risk for colorectal malignancy, such as Juvenile Polyposis, Peutz-Jeghers syndrome, and Cowden disease. [63] Even though a few germline mutations have been described for these syndromes (PTEN, SMAD4) their etiology is still less defined. In addition, up to 30% of all colorectal carcinomas seem to have an inherited predisposition but are not related to any known syndromes. [73] It has been

shown, that, e.g., persons that have a first-degree relative with colorectal cancer harbor a 2-fold risk of colorectal cancer development. [48,74] Even though susceptibility factors are still lacking, genetic polymorphisms may play an important role in predisposing certain families and different racial or ethnic groups to colorectal cancer. [45,51] The Cronkhite-Canada syndrome, Crohn's colitis and ulcerative colitis are non-familial diseases with an increased risk for colorectal cancer. Whereas the lifetime risk for colorectal cancer development is relatively low in patients with Cronkhite-Canada syndrome and Crohn's colitis, it is much higher in ulcerative colitis patients. [75]

Ulcerative colitis-associated colorectal cancer (UCC)

Three in 10 ulcerative colitis patients will eventually develop cancer after a longstanding colitis. [76] Ulcerative colitis can therefore be considered a bona fide premalignant condition and it is recommended that patients with UC participate in surveillance programs in order to screen for early signs of malignancy. [6] No genetic basis has been identified yet to explain the predisposition to colorectal cancer. Ulcerative colitis-associated colorectal carcinomas (UCC) do not develop through the adenoma-carcinoma-sequence. [75] Instead, epithelial dysplasias have been defined as precursor lesions and are meant to be the most predictive feature of intensive and expensive surveillance programs today. [77] However, reliable endoscopic sampling and histopathological evaluation is difficult. [7,9] Additionally, a review of 12 surveillance studies with 92 detected carcinomas in 1916 patients revealed that about half of them were advanced Dukes´ C and D malignancies, and only 12% were early carcinomas. [8] For these reasons, it should be obvious that additional markers with high prognostic impact for individual risk assessment are needed. [78]

CLONAL EXPANSION AND PROLIFERATION

In 1914, Theodor Boveri proposed the model that the loss or gain of chromosomes might cause uncontrolled cell proliferation. [79] This hypothesis could only be validated after technical progress allowed a more detailed analysis of the human genome. In 1956 it was discovered, that the DNA of normal human mitotic cells is packed into 44 autosomes and two sex chromosomes. Each chromosome is constituted by two chromatids that are joined at the centromere. The centromeric region is important for the attachment of kinetochores that are responsible for the segregation of the chromatids during meiotic cell devision (see below). Chromosomes can be classified by their centromere position, size and banding

pattern of heterochromatic and euchromatic DNA. Telomeres are located at the ends of chromosomes to protect the integrity of the chromosomal DNA. They harbor proteins that protect the ends of chromosomes from, e.g., recombination, nuclease attacks, and end-to-end fusions. The DNA polymerase *telomerase* is responsible for the maintenance of the telomere length that physiologically becomes shorter with each cell division. Telomerase reactivation has been identified as an important mechanism for malignant transformation. [80]

Boveri's hypothesis that chromosomal abberations might cause uncontrolled cell proliferation was first supported by the detection of the Philadelphia chromosome in 1960. The Philadelphia chromosome shows a translocation (9;22), characteristic for chronic myelogenous leukemia. [81] This translocation causes synthesis of a tyrosine kinase fusion protein (p210) that increases cell proliferation. Translocations belong to those mutations that affect the chromosomal structure. Inversions, point mutations, deletions and duplications of DNA sequences also belong to the same group of mutations. Such cancer promoting mutations can arise in different genes causing either a gain of function (protooncogenes) or loss of function (tumor suppressor genes). Whereas mutations in oncogenes are mainly dominant, mutations in tumor suppressor genes are typically recessive and follow Knudson's two-hit model: The first mutation "hits" one allele of a tumor suppressor gene which still retains its function due to the remaining wild-type allele. The second "hit" mutates the remaining wild-type allele, which results in the complete loss of gene function. [82] In contrast to mutations that affect the chromosomal structure, mutations of the genome can cause numeric chromosomal alterations that are visible as either a loss or gain of whole chromosomes or chromosome regions. [83] Even though the etiology for the occurrence of genomic mutations has not been conclusively clarified, there is substantial evidence that numeric chromosome imbalances are caused by chromatid segregation errors during mitotic cell division. [84-87] In particular overexpression of cyclin E, a cell cycle stimulator, has been observed in a variety of malignancies causing chromosome instability and aneuploidy. [88,89] The completion of the cell and DNA replication cycle requires the coordination of a variety of macromolecular syntheses, assemblies and movements. These complex processes are normally tightly controlled by regulatory pathways or checkpoints in order to protect genetic integrity and genomic stability. However, cancer cells often exhibit mutations that allow bypassing those regulatory mechanisms leading to growth advantage and clonal expansion. In order to enhance cell proliferation, mutations can affect growth factors, growth factor receptors (that initiate the cell cycle machinery) or cell cycle regulators. There are multiple growth factors (e.g, EGF, VEGF, PDGF,

Insulin) and transmembrane growth factor receptors that have been found to be significantly up-regulated in malignancies. [90-92]

Cell Cycle Regulation

Cell cycle regulators, namely cyclin-dependent kinases (cdks) and their regulatory subunits (cyclins), are the driving forces of the cell cycle and act at different cellular checkpoints: in 1951, Howard and Pelc divided the cell cycle into four phases (GAP1, synthetic phase, GAP2, and mitosis). [93] Later on, abbreviations were used that described the preceding phase as G1, the synthetic phase as S phase, the phase before cell division as G2, and the mitosis phase as M phase. Since the genetic material is duplicated in the S phase and divided in the M phase, the transition of a cell into these two phases is crucial and regulated by the *G1/S checkpoint* and the *G2/M DNA damage checkpoint*. The cascade of interacting cyclins and cdks during the cell cycle can be briefly summarized as follows: The activation of cdk4 and cdk6 by cyclin D leads the cell from the middle of G1 to the G1/S checkpoint. Active cyclin E/cdk2 complexes then trigger the transition from G1 to S phase. The cyclin A/cdk2 complex promotes the cell cycle progress from the G1/S checkpoint into G2. [94] Cyclin A can therefore serve as a proliferation marker for committed cells that will pass through the S and G_2 phase. [95] Cyclin A also binds cdk1 from the end of S to the beginning of the M phase. The function has not been conclusively elucidated but aberrant expression of cyclin A/cdk1 complexes has been associated with tumorigenesis. [96] In addition, cyclin A overexpression itself significantly reflects poor prognosis of colorectal carcinoma patients. [97] For the transition from G2 into M phase, cyclin B activates cdk1. In addition to the cell cycle regulation by cdks and their cyclins, other regulatory factors have been described such as the transcription factor p53 which is responsible for leading the cell into G1 and G2 arrest. [98] Another checkpoint has been described for the M phase which has been subdivided into five phases that harbor specific stages of the meiotic cell division: prophase, prometaphase, metaphase, anaphase, and telophase. The appropriate transition from prometaphase and metaphase to anaphase is highly important to guarantee genomic stability. The cellular mechanisms that could be used to delay prometaphase or metaphase in response to spindle defects or impaired chromosome segregation has been termed the spindle integrity checkpoint. [99]

TUMOR DISSEMINATION

Metastases are one of the hallmarks of solid tumor malignancy. Once a primary tumor has been detected and surgically removed, the survival of the patient greatly depends on the occurrence of local or distant metastases. Rather than the primary

carcinoma itself, it is mainly the metastatic disease that leads to death. The ability of tumor cells to metastasize depends on the acquisition of certain characteristics that allow local or distant spread via the lymphatic or venous system. Early detection of metastasis is important for treatment interventions, however, has proved to be difficult. [100]

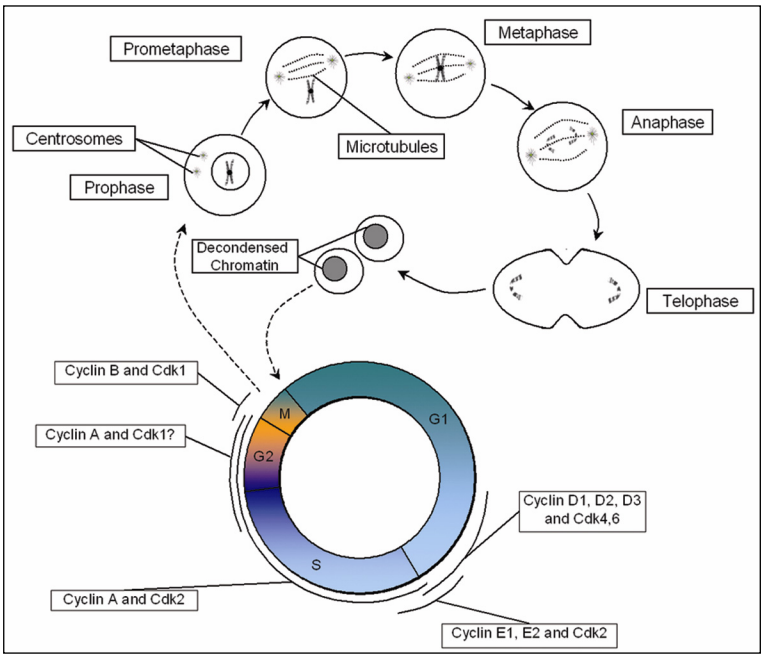


Figure: Cyclins and cdks around the cell cylce.

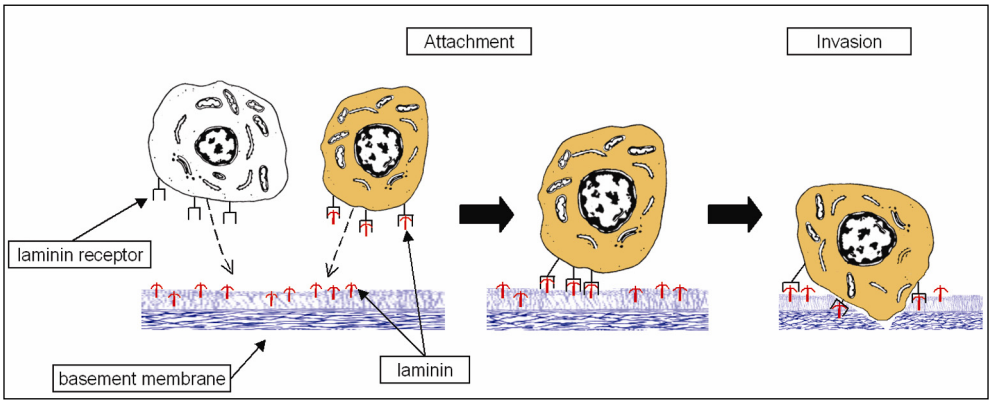


Figure: Invasion of a tumor cell through the basement membrane. Normal cells (white) express laminin receptors whereas neoplastic cells (ocher) are additionally capable to excrete laminin. Laminin builds the bridge between the tumor cell and the collagen in the basement membrane allowing the neoplastic cell to penetrate the basement membrane barrier.

Local invasion

One of the necessary characteristics of metastasizing tumor cells is the ability to invade the basement membrane. Laminins are major components of the basement membranes that belong to a family of heterotrimeric glycoproteins. They are composed of at least α, β and γ subunits that can form 12 or more isoforms. [101] The various isoforms have different tissue specific biological functions, such as cell adhesion, migration, proliferation, as well as growth and differentiation. [102] The laminin-5 isoform (α3: β3: γ2), also known as kalinin, nicein, epiligrin, and ladsin, plays an important role for epithelial cell adhesion to the basement membrane. [103] In order to invade surrounding tissue, tumor cells attach to the basement membrane by binding to laminin receptors from laminin implemented in the basement membrane. This mimics a physiological process: for example, non-neoplastic cells such as inflammatory and endothelial cells regularly cross the basement membrane. These processes are controlled by regulatory mechanisms and it remains unclear how tumor cells can bypass those mechanisms. One possible mechanism is the ability of tumor cells to express laminin themselves. That would enable the attachment to the basement membrane independent from available receptors of the basement membrane laminin. Interestingly, there is much evidence that shows increased expression of the laminin-5 γ2 gene has been found in invasively growing malignant cells at the epithelial-stromal junction, i. e., at the invasion front of different tumors and colorectal cancers. [104-106]

COLORECTAL CANCER GENOMICS, TRANSCRIPTOMICS, and PROTEOMICS

Genomics

Aneuploidy is a consistent genetic alteration of the cancer genome. [107-109] When the first quantitative measurements of the DNA content of cancer cells were performed, aneuploidy was defined as a variation in nuclear DNA content in the population of cancer cells within a tumor. [110] With increased resolution of cytogenetic techniques, such as chromosome banding, comparative genomic hybridization (CGH), spectral karyotyping (SKY), and multicolor fluorescence *in situ* hybridization, it has become clear that in addition to nuclear aneuploidy, specific nonrandom chromosomal imbalances (heretofore referred to as chromosomal aneuploidy) exist. [111-114] Indeed, despite genetic instability in cancer genomes, cancer cell populations as a whole display a surprisingly conserved, tumor-specific pattern of genomic imbalances. [109,115,116] At early steps in the sequence of malignant transformation during human tumorigenesis, *e.g.,* in

preinvasive dysplastic lesions, chromosomal aneuploidies can be the first detectable genetic aberration found. [117-120] This suggests that there is both an initial requirement for the acquisition of specific chromosomal aneuploidies and a requirement for the maintenance of these imbalances despite genomic and chromosomal instability. This would be consistent with continuous selective pressure to retain a specific pattern of chromosomal copy number changes in the majority of tumor cells. [59,109,121,122] Additionally, in cell culture model systems in which cells are exposed to different carcinogens, chromosomal aneuploidy is the earliest detectable genomic aberration. [123,124] The conservation of these tumor specific patterns of chromosomal aneuploidies suggests that they play a fundamental biological role in tumorigenesis.

The progression of colorectal cancer is defined by the sequential acquisition of genetic alterations. [56] At the cytogenetic level, many of these aberrations can be visualized as specific chromosomal gains and losses. These aneuploidies result in a recurrent pattern of genomic imbalances, which is specific and conserved for these tumors. [60] For instance, one of the earliest acquired genetic abnormalities during colorectal tumorigenesis are copy number gains of chromosome 7. [59] These trisomies can already be observed in benign polyps, and can emerge in otherwise stable, diploid genomes. At later stages, e.g., in high-grade adenomas or in invasive carcinomas, additional specific cytogenetic abnormalities become common, such as gains of chromosome and chromosome arms 8q, 13, and 20q, and losses that map to 8p, 17p, and 18q. For a comprehensive summary see the "Mitelman Database of Chromosome Aberrations in Cancer" at http://cgap.nci.nih.gov/Chromosomes/Mitelman. These chromosomal aneuploidies are accompanied by specific mutations in oncogenes and tumor suppressor genes, including ras, APC, and p53. [90] It is therefore well established that both, chromosomal aneuploidies and specific gene mutations, are required for tumorigenesis.

Transcriptomics

A rather well-defined correlation of tumor phenotype and genotype has been established, mainly through the application of molecular and molecular cytogenetic techniques to study sequential changes during tumorigenesis. However, it remains less clear how genomic aneuploidy and chromosomal imbalances impact on the transcriptome. One could postulate that expression levels of all transcriptional active genes on trisomic chromosomes would increase in accordance with the chromosome copy number. Alternatively, changing the expression level of only one or a few genes residing on that chromosome through tumor-specific chromosomal aneuploidies may be the selective advantage necessary for tumorigenesis. This would require the permanent transcriptional silencing of most

of the resident genes. Another formal possibility that must be entertained is that chromosomal copy number changes are either neutral or inversely correlated with respect to gene expression levels. This would mean that gains or losses of chromosomes are a byproduct of specific gene mutations and may not offer any selective advantage. Because of the many chromosomal aberrations usually found in cancer cells, it is difficult, if not impossible, to identify the consequences of specific trisomies, independent from other coexisting genomic imbalances, gene mutations, or epigenetic alterations. [125]

Methodology to analyze the consequences of chromosomal imbalances in tumor genomes has become available through the development of microarray based gene expression profiling. This method has been first described by Schena et al. and enables one the simultaneous analysis of thousands of genes for their gene expression. [126] Despite the exponential increase in the number of publications describing microarray experiments, only a few reports have attempted to specifically address the question regarding the immediate consequences of chromosomal aneuploidies vis-a-vis the dysregulation of the cellular transcriptome. These reports came to quite different conclusions and none of them attempt to address this question in the clinical setting of colorectal carcinogenesis. [127-131] In addition, a comprehensive exploration of how global alterations of the cellular transcriptome might correlate with sequential steps of cellular transformation from normal mucosa via adenoma and carcinoma up to distant metastases has not been described. Such analyses, however, could reveal potential candidate genes for improved prognostics, diagnostics and therapeutics.

Proteomics

The term proteome was first defined in 1994 and denotes the entirety of proteins expressed by the genome. Proteomics is thus understood as the consecutive step following genomics. Proteomics techniques have rapidly evolved and are now widely applied to monitor disease specific alterations. [132] For instance, two-dimensional gel electrophoresis (2-DE) can be used to identify differentially expressed peptides and mass spectrometry can then be subsequently used for protein identification. [133] The application of these techniques could already show that tumor specific quantitative or qualitative changes of protein patterns are indeed discernable. One particularly intriguing possibility develops if tumor specific changes could be detected with non-invasive, cost efficient formats, for instance, by detection of disease specific markers in the peripheral blood. However, the use of single serum markers, e.g., carcinoembryonic antigen, has so far failed to deliver markers of high sensitivity and specificity for colon cancer and most other tumors. [39,40] Comprehensive serum proteome profiling for such tumor specific markers has therefore become a field of intensive research. [134] A particular promising technique

for serum proteome screening is based on surface-enhanced laser desorption ionization time-of-flight (SELDI-TOF) mass spectrometry. SELDI-TOF utilizes chromatographic surfaces that retain proteins from a complex sample mixture according to their specific properties (e.g., hydrophobicity and charge), with the molecular weights of the retained proteins then being measured by TOF mass spectrometry. [134] The identification of SELDI based protein profiles and the subsequent protein identification of features that allow the distinction between malignancy related and normal sera would be highly beneficial.

AIMS OF THIS THESIS

The general aim of this thesis is to further elucidate molecular changes during colorectal carcinogenesis in order to discover biomarkers relevant for improved diagnostic and therapeutic approaches. More specific aims were as follows:

Paper I
To investigate three independent cellular features, i.e. genetic instability, invasive potential, and proliferative commitment, and their possible role as predictors of malignant transformation in the individual risk assessment of ulcerative colitis patients.

Paper II
To identify specific genomic imbalances in ulcerative colitis-associated colorectal carcinomas that present with gross aneuploidy and to establish a comprehensive map of DNA gains and losses of such tumors.

Paper III
To elucidate how genomic imbalances affect chromosome-specific gene expression patterns in particular and how chromosomal aneuploidy dysregulates the genetic equilibrium of cells in general.

Paper IV and V
To identify sequential alterations of the genome, transcriptome, and proteome that define the transformation of normal epithelium and the progression from adenoma to invasive disease; in particular to identify gene and protein expression patterns and genetic pathways associated with malignant transformation, to analyze on how precisely chromosomal aneuploidies affect resident gene expression levels, with the final goal being discovery of novel diagnostically and therapeutically relevant biomarkers of colon cancer progression.

Paper VI
To identify colorectal cancer specific protein signatures in serum samples using surface-enhanced laser desorption ionization (SELDI) methodology and study the potential of such signatures for disease class prediction.

MATERIAL AND METHODS

MATERIALS AND SAMPLE PREPARATIONS

Clinical Material

All clinical material was collected at the Department of Surgery in collaboration with the Department of Gastroenterology and Institute of Pathology, all University Hospital Schleswig-Holstein, Campus Lübeck, Germany. All samples were acquired in adherence with protocols approved by the local Institutional Ethical Review Board.

Paper I

The study was based on data of 629 ulcerative colitis (UC) patients diagnosed between 1986 and 1999. Two cohorts were defined: Group A, comprising eight patients who underwent surgery for an ulcerative colitis-associated colorectal carcinoma (UCC), and Group B, comprising 16 colitis patients with no clinical or morphological signs of malignancy during the period of observation. All 24 patients had been determined as follows: According to the database at the Endoscopy Unit, only 96 out of all 629 UC patients presented at least one of the following four criteria: long-standing UC with a minimum duration of seven years, moderate- to high-grade inflammatory activity, disease extent of at least half of the colon, and occurrence of epithelial dysplasia. According to the 96 patient records, only 24 individuals fulfilled all four parameters together. These patients were selected and divided into groups A and B on the basis of the criteria of the carcinoma diagnosis. All patients underwent at least three colonoscopies, during which mucosal biopsies were sampled in a standardized manner at eight different locations throughout the colon: caecum, ascending colon, right flexure, transverse colon, splenic flexure, descending colon, sigmoid colon and rectum. A total of 683 mucosal biopsies were therefore analyzed.

Paper II

The study was based on data from 23 patients with ulcerative colitis-associated colorectal carcinoma (UCC) that were diagnosed between 1986 and 2001. Synchronous UCC could be observed in five of the 23 cases. Clinical material was collected from surgically removed tumors, diagnosed on hematoxylin–eosin-stained tissue sections, and classified using the tumor–node–metastasis (TNM) classification. Eight sections were prepared from each tumor and used for histopathological diagnosis, immunohistochemistry (thickness 4 µm), DNA ploidy measurements (8 µm), and microdissection prior to DNA extraction (50 µm). A second hematoxylin–eosin-stained section was prepared subsequent to the sections

for CGH analysis, and the histopathological diagnosis was confirmed. All data were obtained from the dissected areas.

Paper IV

The study was based on data of 36 patients (20 male, 16 female) diagnosed with colorectal adenocarcinoma. The tumors were diagnosed at the University of Schleswig-Holstein, Campus Lübeck, Germany. Tissue samples were collected during surgery from different stages of the adenoma-carcinoma-sequence: 17 normal mucosa specimen, 17 adenomas, 19 primary sporadic colorectal carcinomas, and 13 liver metastases. We were successful in collecting the complete sequence of normal mucosa-adenoma-carcinoma of eight individual patients, paired normal mucosa and carcinoma samples from additional eight patients, and paired carcinoma and metastasis from two patients. All samples were stained with hematoxylin & eosin and re-evaluated by one experienced pathologist (S.K.). Before further processing, all normal mucosa samples and adenomas were microdissected. All tumor samples and metastases revealed at least 80% tumor cells, as assessed by Giemsa staining. Each sample was divided into two parts: one for simultaneous DNA and RNA extraction, and one for protein extraction.

Paper V

Colon samples were obtained immediately after resection from 20 patients. Thus, 14 normal mucosa samples, eight adenomas, 16 carcinomas, and seven metastases could be analyzed. Cells were collected from the surface of nonnecrotic tumor or metastatic tissue, normal mucosa and adenomas by scraping with a scalpel, and were transferred into 2–5 ml ice-cold RPMI-1640 medium containing 5% fetal calf serum and 0.2 mM phenylmethylsulfonyl fluoride/ 0.83 mM benzamidine. This was followed by aspiration and squirting steps with a syringe and a 29-gauge needle to preferentially release the tumor cells, which are less attached to each other than the connective-tissue cells. A two-phase nylon filter (pore sizes of 250 and 100 μm) was used to catch remaining stromal components and to allow the passage of the tumor cells. These enriched cell suspensions, underlaid with 2 ml ice-cold Percoll (54.7% in PBS), were centrifuged at 1000 g for 10 min at +4°C. Cells at the interface were collected and washed twice with PBS. Each sample was checked for its tumor cell representativity by comparison of Giemsa-stained smears with histological slides. Only samples with more than 95% tumor cells were used. Histopathological characterization was carried out using hematoxylin-eosin-stained sections of formalin-fixed and paraffin-embedded specimens. Polyps, tumors and liver metastases were characterized according to size, lymph node status and site of metastasis. Histopathological classification followed WHO criteria (TNM classification).

Paper VI

A total of 147 serum samples were collected under standardized conditions and analyzed by surface-enhanced laser desorption ionization time-of-flight (SELDI-TOF) mass spectrometry. After outlier detection, a total of 139 serum samples remained for the final analysis: The training set consisted of 32 healthy controls and 52 patients with colorectal malignancy. The independently collected, non-overlapping, blinded validation set comprised 55 samples (19 healthy controls and 36 patients with colorectal malignancy) and was used to test the usefulness of identified m/z values for prediction of malignant disease.

Cell Lines

Paper III

For Microcell-Mediated Chromosome Transfer (MMCT), mouse/human hybrid cell lines were purchased from the Coriell Repository (http://locus.umdnj.edu/nigms/) and the Japan Health Sciences Foundation (http://cellbank.nihs.go.jp/). All hybrids were cultured according to manufacturers' recommendations. The diploid colorectal cancer cell line, DLD1, was purchased from American Type Culture Collection (http://www.atcc.org). The telomerase immortalized mammary epithelial cell line (hTERT-HME) was purchased from Clontech (Palo Alto, CA) and cultured in the recommended medium. Both recipient cell lines were first tested for the optimal concentration of G418 (Geneticin; Invitrogen, Carlsbad, CA).

METHODS

Histopathology

The original histopathological evaluation was performed by different pathologists. To avoid interobserver variation, all samples were reclassified by one pathologist for inflammatory activity, grade of dysplasia, TNM staging, grading, and the validation of tumor cell representativity. Histological examination was performed without knowledge of clinical data or of patient affiliation to the experimental groups. Inflammatory activity was assessed semiquantitatively (0 = none, 1 = low, 2 = moderate, 3 = high grade). The degree of dysplasia was graded as suggested by Ridell et al. as follows: 0 = none, 1 = indefinite, 2 = low grade, 3 = high grade. [135]

Microcell-mediated chromosome transfer (MMCT)

MMCT methodology was performed as described previously. [136] Briefly, donor A9 cells were grown in six Nunclon T-25 flasks at 1 x 10^6 cells/flask in media containing 500 µg/mL Geneticin. Cells were incubated with 0.05 µg/mL Colcemid in media (plus 20% serum) for 48 hours to induce micronuclei formation. Cells were centrifuged in the presence of 10 µg/mL cytochalasin B at 8000 rpm for 1 hour at 34°C to isolate micronuclei. Micronuclei were purified by sequential filtration through sterile 8-, 5-, and 3-µm filters (Millipore, Billerica, MA). Purified micronuclei were incubated with the recipient cells for 15 to 20 minutes in phytohemagglutinin P containing medium (100 µg/mL). The medium was removed, and cells were coated with 1 mL of PEG 1500 (Roche, Indianapolis, IN) for 1 minute followed by three washes with serum-free medium and incubated overnight in serum containing medium. Cells were plated onto 100 mm^2 plates at 1 x 10^6 cells/plate in medium plus Geneticin (200 µg/mL for DLD1 clones and 50 µg/mL for hTERT-HME clones) for 2 to 3 weeks, until clones appeared. Clones were expanded and tested for incorporation of neomycin-tagged chromosome by fluorescence *in situ* hybridization (FISH) using whole chromosome-specific paint probes and a neomycin-specific DNA probe. We generated four derivative cell lines: DLD1 containing an extra copy of chromosome 3 (DLD1 + 3); an extra copy of chromosome 7 (DLD1 + 7); and chromosome 13 (DLD1 + 13). In addition, chromosome 3 was introduced into the karyotypically normal immortalized mammary epithelial cell line (hTERT-HME + 3).

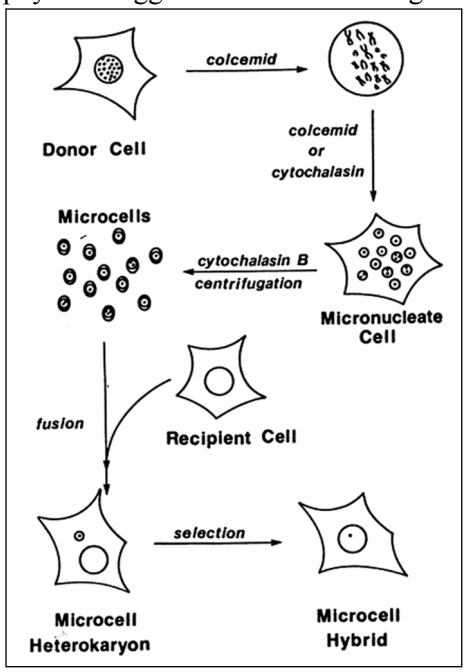


Figure: Theory of microcell-mediated chromosome transfer.

Immunohistochemistry

Immunohistochemistry was performed using the standard peroxidase avidin-biotin-complex (ABC) technique (Vector, Elite Standard Kit, cat. PK-6100). The sections were deparaffinized, rehydrated and microwave treated in 0.01 M sodium citrate buffer (pH 6) for 10 min at 500 W. After rinsing in Tris buffered saline solution (TBS; pH 7.6), the endogenous peroxidase activity was blocked by immersion of

the slides in 0.5% hydrogen peroxidase in distilled water for 30 min. Unspecific staining was prevented by application of 1% bovine serum albumin (BSA) for 20 min. For laminin-5 immunohistochemistry, a polyclonal antibody was raised in rabbits against a fusion protein containing the C-terminus of the laminin-5 γ2 chain (containing amino acid residues 1017–1178). This antibody was used for overnight incubation at 4 °C at a dilution of 1:200 in 1% BSA, followed by a biotinylated antirabbit IgG (diluted 1:200) for 30 min. Cyclin A analysis was performed using a monoclonal mouse antibody against human cyclin A protein (Novocastra Laboratories, Newcastle-upon-Tyne, UK) diluted 1:100 in 1% BSA. This was followed by incubation with the avidin-biotin peroxidase complex for an additional 30 min. The peroxidase reaction was visualized using DAB (diaminobenzidine tetrahydro-chloride, 0.6 mg/ml with 0.03% H2O2 for 6 min). After counterstaining with haematoxylin, the slides were dehydrated and mounted with a xylene-soluble mounting medium. As negative controls, the primary antibodies were replaced by BSA. Cells were regarded as laminin-5 immunoreactive when a distinct staining of the cytoplasm was visible. The cyclin A immunoreactivity was confined to the cell nuclei. The percentage of stained cells was calculated in each specimen. Based on a semiquantitative scoring system, the immunoreactivity for both laminin-5 γ2 chain and cyclin A was recorded as follows: category 0 = no specific antibody expression, category 1 = <20%, category 2 = 20%–50%, and category 3 = >50% immunoreactivity of all mucosal cells. This evaluation was done by three independent investigators being unaware of either the clinical or the histopathological data of the patients. Inter-observer variability was calculated and thus, strong and almost perfect agreement concerning negative and positive laminin-5 expression could be observed (0.78, 0.81, 0.81). Kappa values also showed a strong agreement to distinguish between low (category 0 or 1) and increased (category 2 or 3) cyclin A immunoreactivity (0.89, 0.90, 0.92).

DNA Image Cytometry

Nuclear DNA assessments were performed by means of image cytometry using Feulgen stained histological sections of 4 μm thickness. The staining procedure, internal standardization and cell selection criteria were based on methods described by Auer et al. [137] DNA distribution profiles (histograms) were sampled from 100 interphase nuclei for each specimen. All DNA values were expressed in relation to the corresponding staining controls, which were given the value 2c, denoting the normal diploid DNA content. The DNA profiles were classified according to Auer. Histograms characterized by a single peak in the diploid or near-diploid region (1.5c–2.5c) were classified as type I. The total number of cells with DNA values exceeding the diploid region (>2.5c) was <10%. Type II histograms showed a single peak in the tetraploid region (3.5c–4.5c) or peaks in both the diploid and

tetraploid regions (>90% of the total cell population). The number of cells with DNA values between the diploid and tetraploid region and those exceeding the tetraploid region (>4.5c) was <10%. Type III histograms represented highly proliferating near-diploid cell populations and were characterized by DNA values ranging between the diploid and the tetraploid region. Only a few cells (<5%) showed more than 4.5c. The DNA histograms of types I, II and III thus characterize euploid cell populations. Type IV histograms showed increased (>5%) and/or distinctly scattered DNA values exceeding the tetraploid region (>4.5c). These histograms were suggested to reflect aneuploid populations of interphase nuclei with decreased genomic stability.

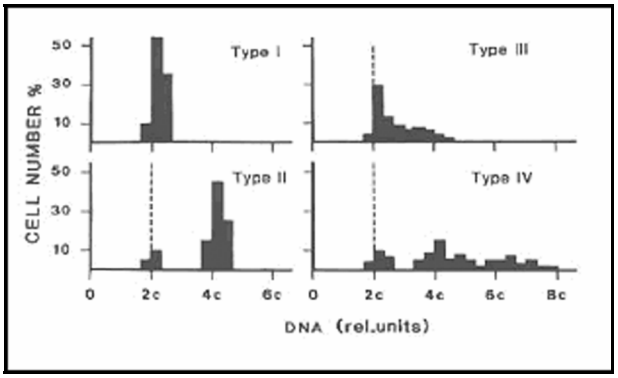


Figure: The DNA histogram types I, II, III, and IV.

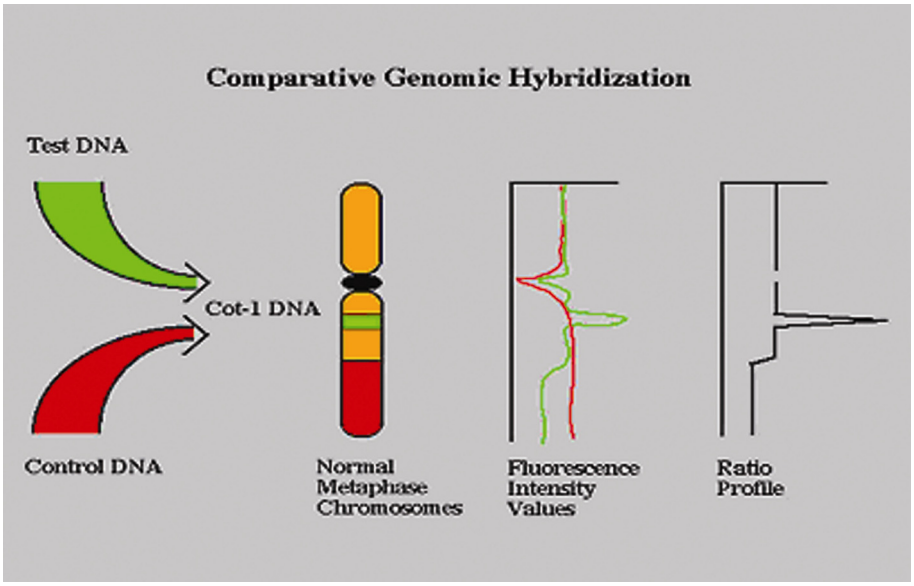


Figure: Theoretical background of Comparative Genomic Hybridization (CGH).

Comparative Genomic Hybridization (CGH)

CGH is based on a quantitative two-color fluorescence in situ hybridization. [112,138] It allows the identification and mapping of DNA copy number changes in tumor genomes. Karyograms visualize gains and losses of tumor-specific DNA as a result of CGH analyses. The sensitivity for detection of copy number alterations is limited mainly by the degree of chromosome condensation on the reference metaphase chromosomes and the frequency of representative copy number changes in the tumor cell population. [139] Thus, the resolution limit for low copy number alterations (losses or gains) is estimated to be in the range of 10 Mbp. However, high copy number changes (amplicons) can be detected being as small as 50 Kbp. CGH is therefore especially useful for detecting recurrent patterns of chromosomal copy number changes in tumor genomes and has been used for solid tumor cytogenetics.

Reference or control DNA isolated from an individual with a normal karyotype and test DNA e.g. from a solid tumor are differentially labeled with reporter molecules (usually green for the tumor genome and red for the reference genome), hybridized to reference human metaphase spreads and subsequently visualized with different fluorochromes. Differences in fluorescence intensities along the reference chromosomes reflect the copy number of corresponding sequences in the tumor DNA. If chromosomes or chromosomal subregions are present in identical copy numbers in both the control and the tumor genome, the observed fluorescence is a blend of an equal contribution of red and green fluorescence. If chromosomes are lost or chromosomal subregions deleted in the tumor genome, the resulting color is shifted to red. A gain of a certain chromosome in the tumor would be reflected by a more intense green staining on the respective chromosome in the reference metaphase preparation. Using computer software, the painted chromosomes will be segmented and the fluorescence values determined on a pixel to pixel basis. The final step in a quantitative fluorescence measurement includes the calculation of average ratio profiles along the chromosomal axis.

The CGH experiments for the here presented studies were conducted as follows. Formalin-fixed and paraffin-embedded samples were provided in 50-µm-thick tissue sections. The tissue was incubated in xylene (3 × 5 minutes), followed by washes in 95% ethanol. According to subsequent hematoxylin–eosin sections, the deparaffinized tissue sections were microdissected to obtain representative tissue containing at least 80% of epithelial or cancer cells respectively. The microdissected samples were placed into Eppendorf tubes with 95% ethanol. After centrifugation, the samples were dried and resuspended in 1 ml sodium isothiocyanate (1 mol/l) and incubated overnight at 37 ºC. DNA was prepared using high-salt extraction and phenol purification. DNA from fresh frozen tissue

was extracted subsequently to RNA extraction using Trizol reagent (Invitrogen, Carlsbad, CA) according to the manufacturers protocol. The extracted DNA was labeled by nick-translation using biotin-11-dUTP (Boehringer Mannheim, Indianapolis, IN). Genomic DNA from cytogenetically normal individuals was labeled with digoxigenin-12-dUTP (Boehringer Mannheim) as a control. Hybridization was performed on karyotypically normal metaphase chromosomes using an excess of Cot1-DNA (GIBCO BRL, Gaithersburg, MD). The biotin-labeled sequences were visualized with avidin–fluorescein isothiocyanate (Vector Laboratories, Burlingame, CA) and the digoxigenin-labeled sequences were detected with a mouse-derived antibody against digoxigenin followed by a secondary rhodamine-conjugated anti-mouse antibody (Sigma-Aldrich, Milwaukee, WI). Detailed protocols can be retrieved from http://www.riedlab.nci.nih.gov. Quantitative fluorescence imaging and CGH analysis were performed using Leica Q-CGH software (Leica Imaging Systems, Cambridge, UK). Interpretation of changes at 1pter and chromosomes 16, 19, and 22 required careful examination, because these loci are prone to artifacts due to the high proportion of repetitive sequences. CGH results of individual cases and CGH comparison tools can be found at http://www.ncbi.nlm.nih.gov/sky/skyweb.cgi.

Fluorescence in situ Hybridization (FISH)

The technique of fluorescence in situ hybridization allows the visualization and localization of specific DNA sequences on the chromosomal and cellular level by utilizing locus-specific gene probes or chromosome-specific DNA libraries, respectively. [139] A particular advantage of this method is the possibility to analyze chromosomal aberrations, such as gene amplifications, also in non-dividing cells (interphase nuclei). [140]

Chromosome-specific painting probes were hybridized to confirm the incorporation of a given chromosome in the microcell-mediated chromosome transfer clones. Centromere-specific probes (Vysis, Downers Grove, IL) were used for quantitation of chromosome incorporation rate. The FISH technique was performed as described previously. [141] Briefly, slides were pretreated with RNase, fixed, and denatured in 70% formamide / 2x SSC for 1.5 minutes at 80°C. Centromere/telomere probes were denatured at 74°C for 5 minutes and placed on the denatured slides, coverslipped, and incubated at 37°C overnight (16 to 20 hours). Slides were washed, counterstained with 4',6-diamidino-2- phenylindole and mounted with antifade solution. Images were acquired on a DMRXA epifluorescence microscope (Leica, Wetzlar, Germany) using Qfluoro software (Leica, Cambridge, United Kingdom).

Spectral Karyotyping (SKY)

The SKY method allows color karyotyping of all human chromosomes simultaneously. [113] Chromosome homologues are displayed with a unique color due to a simultaneous hybridization of a 24-chromosome-specific probe pool. Thus, SKY has been in particular useful in identifying chromosomal rearrangements in tumor cytogenetics. [142] We used SKY analysis to determine whether the chromosome transfer process induced secondary karyotypic changes. [143] Slides were pretreated with RNase, followed by pepsin to remove cytoplasm and denatured. Slides were then hybridized with a SKY probe mixture for 72 hours at 37°C. Images were acquired and processed as described previously. [143]

Gene expression micro-arrays

The microarray technique is based on a quantitative two-colour fluorescence in situ hybridisation and allows simultaneous gene expression analysis of thousands of genes. [126] This technique has been successfully utilized to identify differential gene expression patterns in clinical cancer research. [144] We used cDNA microarray expression profiling to elucidate consequences of chromosomal aneuploidies on the transcriptional level and to monitor gene expression patterns during the sequential steps of colorectal carcinogenesis.

Total RNA was extracted using Trizol (Invitrogen, Carlsbad, CA) followed by Qiagen RNeasy column purification (Qiagen, Valencia, CA). Each RNA preparation was hybridized using a slightly modified protocol from Hedge et al. [145] Extraction and hybridization protocols used can be viewed in detail online at http://www.riedlab.nci.nih.gov. In brief, 20 μg of total RNA were reverse transcribed using random primers and converted into cDNA using reverse transcriptase. After incorporation of aminoallyl-conjugated nucleotides, the RNA was indirectly labeled with Cy3 (cell line/tissue RNA) and Cy5 (reference RNA; Amersham, Piscataway, NJ). Each sample was hybridized against universal human reference RNA (Stratagene, La Jolla, CA) in a humid chamber (Arraylt Hybridization Cassette, TeleChem Intl., Sunnyvale, CA) for 16 hours at 42°C, washed and scanned by the Axon GenePix 4000B Scanner (Axon Instruments, Union City, CA). The amount of labelled cDNA that hybridizes with its target cDNA clone on the array is proportional to its abundance in the original sample. Thus, the signal of each spot is proportional to the concentration of the corresponding mRNA in the sample: if a gene is equally expressed in both the reference and the tumor genome, the observed fluorescence is a blend of an equal contribution of red and green fluorescence. A higher gene expression in the tumor is thus be reflected by a more intense green staining. Signal intensities will be normalized and expression levels are calculated in comparison to the reference

RNA. Customized arrays were obtained from the National Cancer Institute's Advanced Technology Center. All arrays were composed of 9128 cDNAs that were denatured and immobilized on a poly-L-lysine–coated glass surface. The gene annotation files (GAL file) were obtained from the Advanced Technology Center's web site at http://nciarray.nci.nih.gov. GenePix software version 4.0.1.17 was used to apply the GAL file through an interactive gridding process. All images of the scanned microarray slides were meticulously inspected for artifacts. Empty spots and aberrant spots and slide regions were flagged for exclusion from analyses. [146]

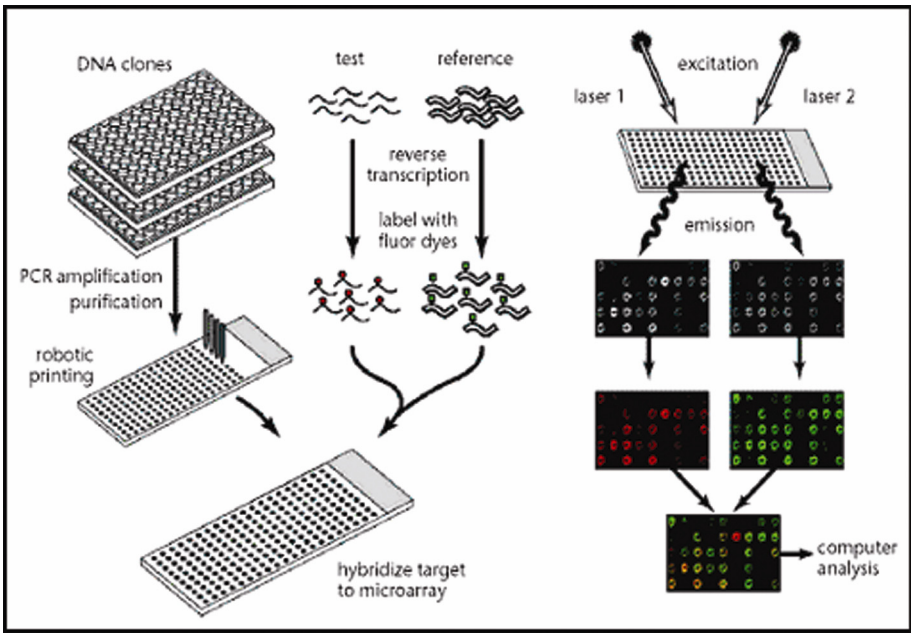


Figure: Theoretical background of cDNA microarray experiments.

Quantitative-PCR (TaqMan)

Complementary DNAs (cDNA) were prepared by reverse transcribing 25 ng of total cellular RNA using the Single-Strand cDNA Synthesis Kit (Roche, Inc.) according to manufacturer's protocol. Quantative-PCR (Q-PCR) analysis was performed using TaqMan probes (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions, in 10μL final volumes, in 384-well microtiter plates. [147] Thermocycling conditions using an Applied Biosystems ABI-7900 SDS were as follows; 50°C for 2 minutes, 95°C for 10 minutes and 40 cycles of 95°C for 10seconds and 60°C for 1 minute. Specific primers for Q-PCR of GAPDH, GUS, and 18S as housekeeping genes and APC, MYC, EAF1, EGR1,

EPHA7, HRG, MFAP4, SERPINA1, OPCML, and CLCA1 were designed using Applied Biosystems Assay-by-Design primer design software, and their sequences are proprietary in nature. Accurate quantification of each mRNA was achieved using the normalization of the sample ΔCT values to one reference. This value, referred to as the $\Delta\Delta C_{T\text{-}Sample}$ value ($\Delta\Delta C_{T\text{-}Sample} = \Delta C_{T\text{-}Sample} - \Delta C_{T\text{-}Reference}$), is derived by taking the result of the expression: If $2^{(-\Delta\Delta CT)} - 1 > 0$ then the result = $2^{(-\Delta\Delta CT)} - 1$ or else the result = $-1 / 2^{(-\Delta\Delta CT)}$. This equation changes the range for down regulation from 0 - 1 to -∞ - 0, and up regulation from 1 - ∞ to 0 - ∞. The target mRNA expression was normalized separately to GAPDH, GUS, and 18S expression, and the relative expression was calculated back to the untreated controls for each cell type. Raw data from each Q-PCR run was exported into a comparative C_T analysis workbook. C_T represents the threshold cycle or the PCR cycle at which an increase in reporter fluorescence above baseline signal can be detected. The comparative C_T workbook allows for normalization with different endogenous controls on a number of samples and genes. Each graph displays the analyzed results in a format (both numerically and graphically) showing their expression relative to not only an endogenous control but also a reference sample.

Two-Dimensional Gel Electrophoresis (2-DE)

The 2-DE approach allows the separation of complex protein mixtures. [148] Proteins mixtures are applied to a two-dimensional SDS gel and single polypeptides will be separated depending on their individual size and charge. Proteins of interest can be extracted from the gel and subsequently identified by mass spectrometry. [133] 2-DE was performed as described previously. [149,150] Protein concentrations of samples were determined by addition of 25 µl concentrated assay reagent (Bio-Rad) to 1 µl solubilized sample diluted in 100 µl Milli-Q water using 96-well microplates. A standard curve was constructed using different concentrations of bovine serum albumin. The plate was read using a Multiscan reader (Labsystems). Before 2-DE application, all samples were diluted to 500 µl containing 7 M urea, 2 M thiourea, 1% 3-(3-chloramidopropyl) dimethylammonio-1-propanesulfonate (CHAPS), 0.4% immobilized pH gradient (IPG) buffer, 0.3% dithiothreitol (DTT) and a trace of bromophenol blue. Total cell extracts were applied via active rehydration on precast IPG strips (pH 4–7, 17 cm; Bio-Rad) in sample solution for isoelectric focusing. Protein (75 µg/IPG strip) was loaded and focused in PROTEAN IEF Cells (Bio-Rad) for ~52,900 Vh overnight. After isoelectric focusing (IEF), the strips were immediately equilibrated for 2 x 15 min with 50 mM Tris-HCl, pH 8.8, in 6 M urea, 30% glycerol and 2% SDS. DTT (2%) was included in the first and iodoacetamide (2.5%) in the second equilibration step to reduce and alkylate free thiols. SDS gels with a 10–13% linear acrylamide gradient (1.5 x 200 x 230 mm)

were used for the second dimension. The IEF strips were carefully applied on top of the gels and fixed in position using warm agarose (0.5%) dissolved in SDS/ PAGE running buffer. Electrophoresis was carried out overnight at 42,000 Vh. For image analysis, gels were stained with silver nitrate. Preparative gels utilized more material (two- to tenfold more) and spots were then visualized using Coomassie Brillant Blue, Sypro Ruby (Bio-Rad) or Silver Plus staining (Bio-Rad). 2-D gels were scanned at 84.7 x 84.7 µm resolution using a GS-710 imaging densitometer (Bio-Rad).

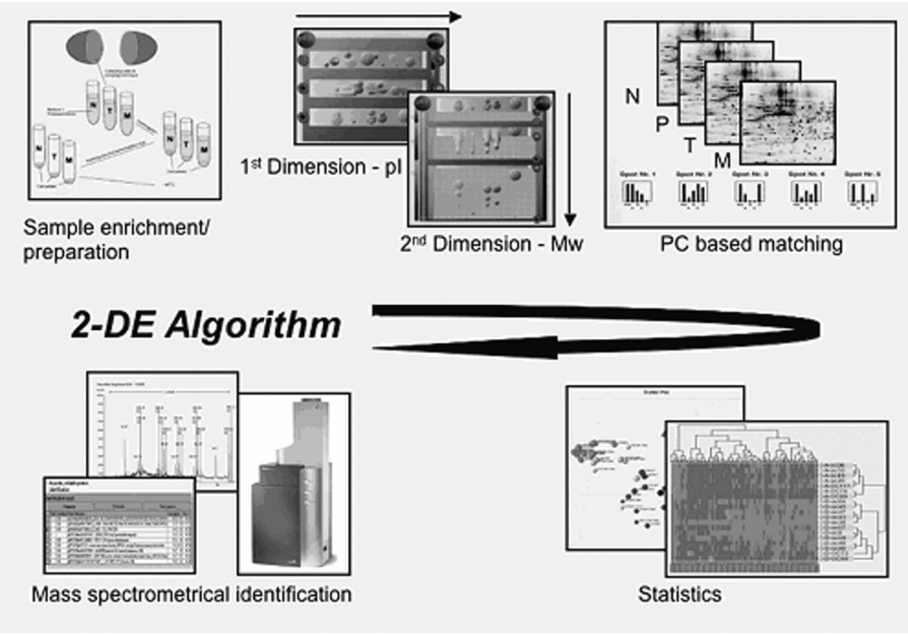


Figure: Theoretical background of two-dimensional gel electrophoresis.

Matrix-assisted Laser Desorption Ionisation (MALDI) Mass Spectrometry

Within the MALDI technique, matrix and sample are co-crystallized on the MALDI plate and irradiated with a laser pulse. [151] The matrix absorbs the energy and acts as an intermediary for the co-desorption and ionization of sample and matrix. The ions are accelerated in an electrical field and enter a field free drift tube. The mass-related time of flight is detected and the analogue signal converted and digitalized. The experimentally generated masses are compared to a set of mass profiles in a protein database, e.g., SwissProt (http://au.expasy.org/sprot/), ExPASy (http://au.expasy.org/), or UniProt (http://www.ebi.uniprot.org/index.shtml). The most similar pattern determines the

protein-"hit". The tighter the mass tolerance, the more stringent is the identification.

Protein spots were excised manually from Coomassie stained gels and in-gel digested using a MassPREP robotic protein-handling system (Micromass). Gel pieces were destained twice with 100 µl 50 mM ammonium bicarbonate (Ambic)/50% (v/v) acetonitrile at 40°C for 10 min. Pieces containing protein were reduced by 10 mM DTT in 100 mM Ambic for 30 min, shrunk in acetonitrile, and the proteins were then alkylated with 55 mM iodoacetamide in 100 mM Ambic for 20 min. Trypsin (25 µl of a 12 ng/µl solution in 50 mM Ambic) was added and incubation was carried out for 4.5 h at 40°C. Peptides were extracted with 30 µl 5% formic acid/2% acetonitrile followed by extraction with 24 µl 2.5% formic acid/50% acetonitrile. The acetonitrile was evaporated under atmospheric pressure overnight at 10°C. For electrospray (ES) ionization MS/MS, the peptide extracts were desalted with C18 ZipTips (Millipore), activated and equilibrated using 10 µl 70% acetonitrile/0.1% trifluoroacetic acid (TFA) twice, 10 µl 50 % acetonitrile/0.1% TFA twice, and finally 10 µl 0.1% TFA twice. The sample was loaded onto the ZipTip by pipetting 20 times and washed using 10 µl 0.1% TFA twice. The tryptic fragments were eluted with 60% acetonitrile/1% acetic acid. Samples with proteins in low yield were analyzed using a Gyrolab MALDI SP1 Workstation (Gyros AB). In this approach, 96 microcolumns (packed with a C18 resin to a volume of 10 nl) are incorporated into a CD platform and used for desalting by reverse-phase chromatography. The columns are conditioned with 50% acetonitrile in water. The samples are loaded onto the columns, and solvents passed through by spinning of the disc. The wash solution (200 nl 0.1% TFA) is directed to a waste exit. Peptides are eluted from the columns using 200 nl 50% acetonitrile containing 1 mg/ml *a*-cyano-4-hydroxycinnamic acid matrix and 0.1% TFA. The eluate is captured in an open MALDI target area of 200 x 400 µm for solvent evaporation and the peptide/matrix crystallization. For on-CD MALDI analysis, the cut CD was accommodated in the target compartment of the MALDI instrument.

The tryptic fragments were mass analyzed by matrix-assisted laser desorption ionization (MALDI) mass spectrometry (Voyager DE-PRO; Applied Biosystems) and, where relevant, also with ES ionization quadrupole time-of flight (Q-TOF) tandem MS (Micromass) for sequence information. Samples for MALDI analysis were mixed at a 1:1 (v/v) ratio with a saturated α-cyano-4-hydroxycinnamic acid solution in 50% acetonitrile/0.1% TFA. Database searches were carried out using the MS-Fit search program (http://prospector.ucsf.edu/). Only protein hits with three or more matching peptide masses were considered, and only if the assigned protein was human, and the theoretical pI and M_r values did not deviate excessively from the values observed in the 2-D gel separation. Samples for the ES analysis

were introduced via goldcoated nano-ES needles (Protana). A capillary voltage of 800–1000 V was applied together with a cone voltage of 40–45 V and collision energy of 4.2 eV. The sample aerosol was desolvated in a stream of nitrogen. During the collision-induced dissociation (CID), the collision energy was in the range of 15–30 eV with argon as collision gas.

Surface-enhanced Laser Desorption Ionisation (SELDI)

A major advantage of Surface-enhanced Laser Desorption Ionization (SELDI) is that complex protein mixtures can be directly analyzed by mass spectrometry without any prior separation and purification. SELDI employs a panel of different chips with binding characteristics for different proteins. Microliter quantities of serum are directly applied to chips and the bound proteins are treated and analyzed by mass spectrometry. The mass spectra patterns obtained for different samples reflect the protein and peptide contents of these samples. Protein identification itself needs to be performed in an additional analysis step. [152-154] The reliability and reproducibility have been proven even if variation coefficients of 8%-10% indicate the need for technical repeats. [153,154] SELDI – TOF mass spectrometry is particularly well suited to evaluate low-molecular proteins (0.5-25kDa) and is, as such, complementary to the 2-DE approach.

Sample preparations were processed according to protocols provided by the manufacturer (Ciphergen Biosystems, Freemont, CA, USA) of the two protein chips utilized (IMAC3: immobilized metal affinity capture array with a nitriloacetic acid surface; WCX2: weak cation exchange array with carboxylate functionality). Serum samples were thawed on ice and subsequently diluted 1:1 in 9M Urea/2% CHAPS. Following 30 minutes incubation on ice, samples for IMAC3 chip use were diluted 1:5 in PBS, pH 7.4 + 0.1% Triton X-100, whereas samples for WCX2 application were diluted 1:5 in 50mM sodium acetate, pH 4.5 + 0.1% Triton X-100. All samples were stored on ice until ready to use. After bioprocessor set up, both chip platforms were pretreated: all solutions were incubated at room temperature with gentle agitation and decanted before a new solution was applied. The IMAC3 chip was incubated with 100 µl/well of 100 mM CuSO4 for 10 minutes with one repeat followed by a brief wash with 200 µl/well of HPLC-grade water for 30 seconds. Then, 100 µl/well of 50mM sodium acetate pH 4.0 were incubated for 10 minutes before 200 µl/well of HPLC-grade water were applied for 30 seconds. The WCX2 chip was incubated with 100 µl/well of 10mM HCl for 10 minutes followed by application of 200 µl/well of HPLC-grade water for 5 minutes. After pretreatment of the chip surfaces each well of the bioprocessor was incubated with 100 µl of binding/wash buffer according to the chip to be used (IMAC3: PBS, pH 7.4 + 0.1% Triton X-100; WCX2: 50mM sodium acetate, pH 4.5 + 0.1% Triton X-100). After 15 minutes incubation, 100 µl

sample were added to each well of the bioprocessor. The bioprocessor was covered and incubated for 1.5 hours at room temperature while shaking. Samples were then removed and each well of the bioprocessor was washed three times with 100 µl of the binding/wash buffer as specified above for 5 minutes each. Additionally, 200 µl of final wash buffer were added to each well of the bioprocessor and incubated for 30 seconds with shaking (*IMAC3*: 5mM Tris, pH 8.0; *WCX2*: HPLC grade water). Incubation with final wash buffer was repeated once. Protein chips were removed from the bioprocessor and air-dried. One µl of a saturated sinapinic acid solution in 50% acetonitrile, 0.5% trifluoroacetic acid was added to each spot of the proteinchip array.

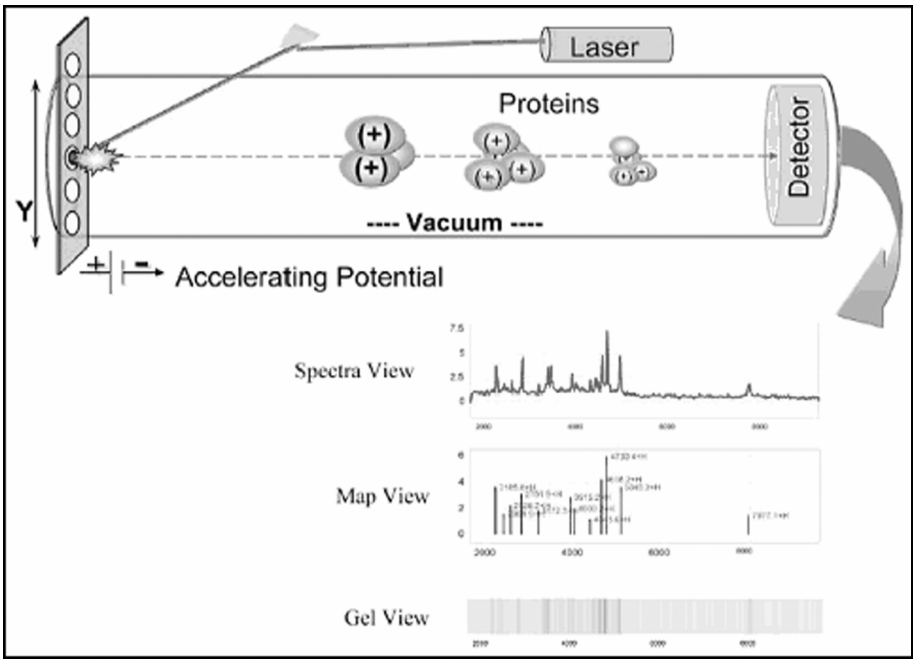


Figure: Schematic diagram of the SELDI Ciphergen mass spectrometer. Samples are loaded onto ProteinChip arrays that are analyzed by a laser desorption ionization (LDI) time-of-flight mass spectrometer (TOF-MS). The TOF MS measures the molecular weights of the various proteins that are retained on the array. For comparison purposes, the software associated with the SELDI Ciphergen instrument is capable of displaying the resultant data as either a spectra, map, or gel view. [152]

Both protein chips were analyzed on the Protein Biology System 2 SELDI-TOF mass spectrometer (Ciphergen Biosystems, Freemont, CA, USA). Mass accuracy was assessed daily through external calibration utilizing a standard mixture obtained from Ciphergen. The arrays were analyzed using the following PBS-II

SELDI-TOF MS automated settings: laser intensities 215 (IMAC3) and 220 (WCX2), detector sensitivity 8, focus mass 5000, *m/z* range 0-200,000, 130 averaged laser shots per sample spectrum. Data were collected using Ciphergen ProteinChip software version 3.0.2.

Statistical Analysis

Paper I

Statistical analysis was performed using SAS Version 6.0 for Windows NT exclusively. To investigate whether differences between the two patient groups exist, the two-sided Wilcoxon rank sum test at a level of a = 0.05 was used. According to this test, a mean value was calculated for each patient based on all biopsies taken throughout the individual observation time. This can only be done if one assumes that different biopsies regarding time and location in the colon can be treated as coherent observations when they originate from the same patient or rather the same genome. In addition, for the patients with ulcerative colitis-associated carcinoma, the progress of parameters was tested (Wilcoxon signed rank test) in order to detect a trend over time. The correlation analysis was performed by means of Spearman correlation coefficients. Kappa values were calculated to analyse inter- and intra-observer agreement between histopathological and immunohistochemical evaluations. Finally, a logistic regression was applied to calculate the prediction of cancer development by the combined approach of all markers.

Paper II

To detect relationships among cytogenetic imbalances, expression levels of immunohistochemical markers, and clinical course, the data were dichotomized and analyzed by Fisher's exact test at a level of $\alpha = 0.05$ using exclusively SPSS version 10.0 (SPSS, Inc., Chicago, IL). Further, Kaplan–Meier survival curves were estimated and compared by a log rank test.

Paper III

Log ratios for each spot were calculated as follows: in each channel, signal was calculated as foreground mean minus background median. If the signal was <100 in a single channel, the signal value in that channel was set to 100. If the signal was <100 in both channels, the spot was flagged as unreliable and not used in further analyses. Ratios for non-flagged spots were calculated by dividing the green (Cy3) signal by the red (Cy5) signal and then a log base 2 transformation was applied. Log ratios were normalized within each array by subtracting from each the median log ratio value across the spots on the array. These median-normalized log ratios

(removing all flagged spots) were used in all subsequent analyses. Intensity-based normalization was considered; however, based on the diagnostic plots, a global normalization approach (with the truncation of signal at 100) was considered superior for this data set. To determine whether the hybridizations were of sufficient quality to extract meaningful and statistically relevant conclusions, we examined the correlation among the four replicate arrays within each experimental condition. These pairwise correlations were calculated to assess the variability between hybridizations (technical repeats), cells at the time of harvest (biological repeats), and due to the physical properties of the individual fluorochromes (reverse-fluorochrome labeling). This was done for each of the six different experimental groups: DLD1, DLD1 + 3, DLD1 + 7, DLD1 + 13, hTERT-HME, and hTERT-HME + 3. DLD1 was analyzed on two print batches (I and II). For the three purely technical replicates, the pairwise correlations were 0.83, 0.87, and 0.87. In subsequent analyses, all replicates within experimental conditions were treated equally, regardless of whether they were between-extraction, between-hybridization, or reverse fluorochrome replicates because the correlation analysis indicated that in this experimental system, between-extraction variability, or reverse fluorochrome variability (ratios inverted) contributed little additional variability over and above that due to hybridization. The second phase of the analysis was to produce summary statistics of the expression levels within each group at the gene level, chromosome level, and chromosome arm level. All calculations were done on the log base 2 scale. To produce gene level summaries, the mean and standard deviation (SD) of the log ratios for each gene across the four arrays within each group were calculated. Diagnostic plots suggested that the SD was constant as a function of mean log ratio across genes within a group, so an overall average SD was computed for each group as the median SD overall genes. For each array and each chromosome, chromosome level summaries were calculated by averaging log ratios from genes known to be located on that chromosome. When average expression ratios are reported, they have been obtained by taking antilog of the mean log ratios. Comparisons between experimental groups were conducted at each of the chromosome, chromosome arm, and gene levels.

Paper IV

Log ratios for each spot were calculated as follows: in each channel, the signal was calculated as foreground mean minus background median. If the signal was less than 100 in any single channel, the signal value in that channel was set to 100. If the signal was less than 100 in both channels, the spot was flagged as unreliable and not used in any further analyses. For all remaining (non-flagged) spots, a log ratio was calculated as $\log_2$((green signal)/(red signal)). M vs A plots were

constructed, and an intensity-based normalization curve was estimated using locally weighted regression (lowess). The lowess curve was calculated using the statistical software Splus (version 6.0, Release 2, Insightful Corporation). Normalization was accomplished by subtracting the lowess curve-determined correction factor from each log ratio. To avoid unduly influencing results by a few very extreme expression ratios, any normalized log ratios > 5 or < -5 (i.e., ratios > 32 or < 1/32) were truncated at 5 or -5 prior to subsequent statistical analyses. Comparisons of gene expression between stage groups (normal mucosa, adenoma, carcinoma, metastasis) were performed using two-sample pooled-variance t-statistics. Unless otherwise noted, a gene was considered differentially expressed between groups if the two-sample t-statistic calculated for that gene reached statistical significance with $p < 0.001$.

In order to examine the association between gene expression and gene copy number changes (measured by CGH), a linear mixed model was fit to data for each chromosome arm. For this purpose, CGH and gene expression data were used only from those tissue samples for which both analyses could be performed in parallel. For each array, a chromosome arm-level average log expression ratio was calculated as the average of the log expression ratios observed for that array for all genes known to reside on that chromosome arm. Copy numbers determined from CGH analysis were quantified for each chromosome arm for each specimen by an ordinal scoring system taking possible values into account (-1 = loss of chromosomal arm / -0.5 = partial loss / 0 = no loss or gain / 0.5 = partial gain / 1 = gain / 2 = high amplification). Separately for each chromosome arm, a linear mixed model was fit. The dependent variable in each linear mixed model was the chromosome-arm level average log expression ratio. Independent variables in each model were the CGH score treated as a continuous variable (fixed effect), disease stage (normal mucosa, adenoma, carcinoma, metastasis) as categorical fixed effect, and random subject effects. P-values reported for the expression ratio-CGH score associations correspond to p-values obtained from tests presuming that the regression slope coefficients for CGH score in the above-described linear mixed models are different form zero.

Paper V

Data were analyzed using PDQuest software from Bio-Rad. A Principal Component Analysis (PCA) was performed with 60 gels to screen for outliers and clusters. Spot data from the PDQuest gel analysis package were subsequently transferred to the Spotfire statistical software (DecisionSite for Functional Genomics; www.spotfire.com). PCA is effective for identifying discriminating features in a data set by finding two or three linear combinations of the original features that best summarize the variation in the data. If much of the variation is

captured by these two or three most significant principal components, group membership of many data points can be observed. Using the PDQuest 7.1 software, polypeptide spots from the different stages of the disease were matched to spots in the reference pattern. Spot intensities were normalized, background was subtracted, peaks were located and the relative staining intensities were determined. Several known proteins served as landmarks to facilitate the gel matching. Spots were identified using boolean analysis and the Mann-Whitney test ($p < 0.05$). A high-level match set comparing the clinical cancer samples with two colorectal cancer cell lines (HCT 116, diploid / Lovo, aneuploid) was also constructed, as were gels from two normal liver samples, recovered from the liver metastases resection specimens. Only those protein spots that indicated up- or down-regulation (> twofold) in the metastases samples were considered for identification.

Paper VI

The Ciphergen chip data were treated as described briefly: Background subtraction using Ciphergen software was followed by truncation of the spectra to eliminate m/z values below 1500 Da. Values below that range can be due to noise inherent to SELDI-TOF based protein profiling. After scaling each spectrum to a constant total ion current, the spectra were combined to identify peak regions with sufficient intensity: 648 peaks were identified on the IMAC3 platform, and 761 peaks on the WCX2 chip. The spectra of the two chip surfaces were then combined, such that each spectrum presented 1445 features. It was then determined whether the duplicate spectra (two technical repeats per serum sample) should be averaged or kept as duplicates. Outlier detection identified eight samples that were excluded from subsequent analysis. Only training set spectra were then used to identify features that distinguish malignant sample sera from control sera: through model building based on evolutionary programming, a total of 16 top models selected a set of 13 features that were chosen in different combinations by Distance-Dependent K-Nearest Neighbors (DD-KNN). The value of these features for the detection of colorectal malignancy was then tested with an independently collected, blinded validation set consisting of 55 samples. For this purpose, Maximum Likelihood was used for class prediction in all 16 models that were based on K-Nearest-Neighbors. Only samples that revealed identical results with all 16 models were considered for further classification into the malignant or healthy group. All analytical procedures were completed before patient diagnoses were decoded.

RESULTS AND DISCUSSION

Paper I

Patients with ulcerative colitis have a significantly increased lifetime risk for the development of colorectal carcinoma. [155] Despite extensive surveillance programs, clinical practice reflects an uncertainty in individual risk assessment. The efficiency of such programs, especially endoscopic surveillance with the purpose of finding dysplastic lesions as markers for an impending malignant transformation, has been questioned for several reasons. [9] In order to evaluate independent cellular features with possible predictive value in patient risk assessment, two groups were selected: *group A* comprised eight patients with ulcerative colitis-associated colorectal carcinomas (UCC), and *group B* consisted of 16 ulcerative colitis (UC) patients with risk factors (duration of disease, extent of inflammation, epithelial dysplasias). A total of 683 paraffin-embedded mucosal biopsies were retrospectively evaluated for inflammatory activity, grade of dysplasia, ploidy status, laminin-5 γ2 chain and cyclin A expression.

In all biopsies, mild or moderate inflammatory activity was present in 78% while low-grade or high-grade dysplasia was found in 5.5%. There was, however, no difference in inflammatory activity and dysplasia between patient groups (**Table II, Paper I**). It is a known fact that dysplasia is absent in 20% - 30% of colectomy specimens containing UCC. [156] Additionally, a review of 12 surveillance studies with 92 detected carcinomas in 1916 patients revealed that about half of them were advanced Dukes' C and D malignancies. Only 12% were early carcinomas detected by surveillance. [8] In our study, only two cancer patients had a distinct high-grade dysplastic lesion prior to the final diagnosis. One of them was underestimated in the original routine histopathological diagnosis. The tumor stages of the eight UCCs were as follows: one Dukes' A, three Dukes' B, two Dukes' C and two Dukes' D.

One of the most important findings of this study was the detection of highly aneuploid epithelial cell populations scattered over the colon and rectum in premalignant biopsies of all eight UCC patients (**Figure 1, Paper I**). These lesions could be observed up to 11 years prior to the final cancer diagnosis (average 7.8 years). They were found in macro- and microscopically unsuspicious mucosa, could even be detected in regenerative epithelium, and were not related to dysplasia. This DNA aneuploidy occurred more frequently in group A biopsies (75%) than in those of group B (14%, $p = 0.006$). The carcinoma samples of the eight UCC patients also exhibited highly aneuploid DNA distribution patterns.

Löfberg et al. reported aneuploid biopsies in 25% of high-risk patients at least once during 10 years of observation. [157] In other studies, aneuploidy has been repeatedly observed by flow-cytometry in non-dysplastic mucosa of high-risk patients. [158] The results of the present study support the above-mentioned observations. Genomic instability, represented by DNA aneuploidy, could initiate the process of malignant transformation in colitis as an early event.

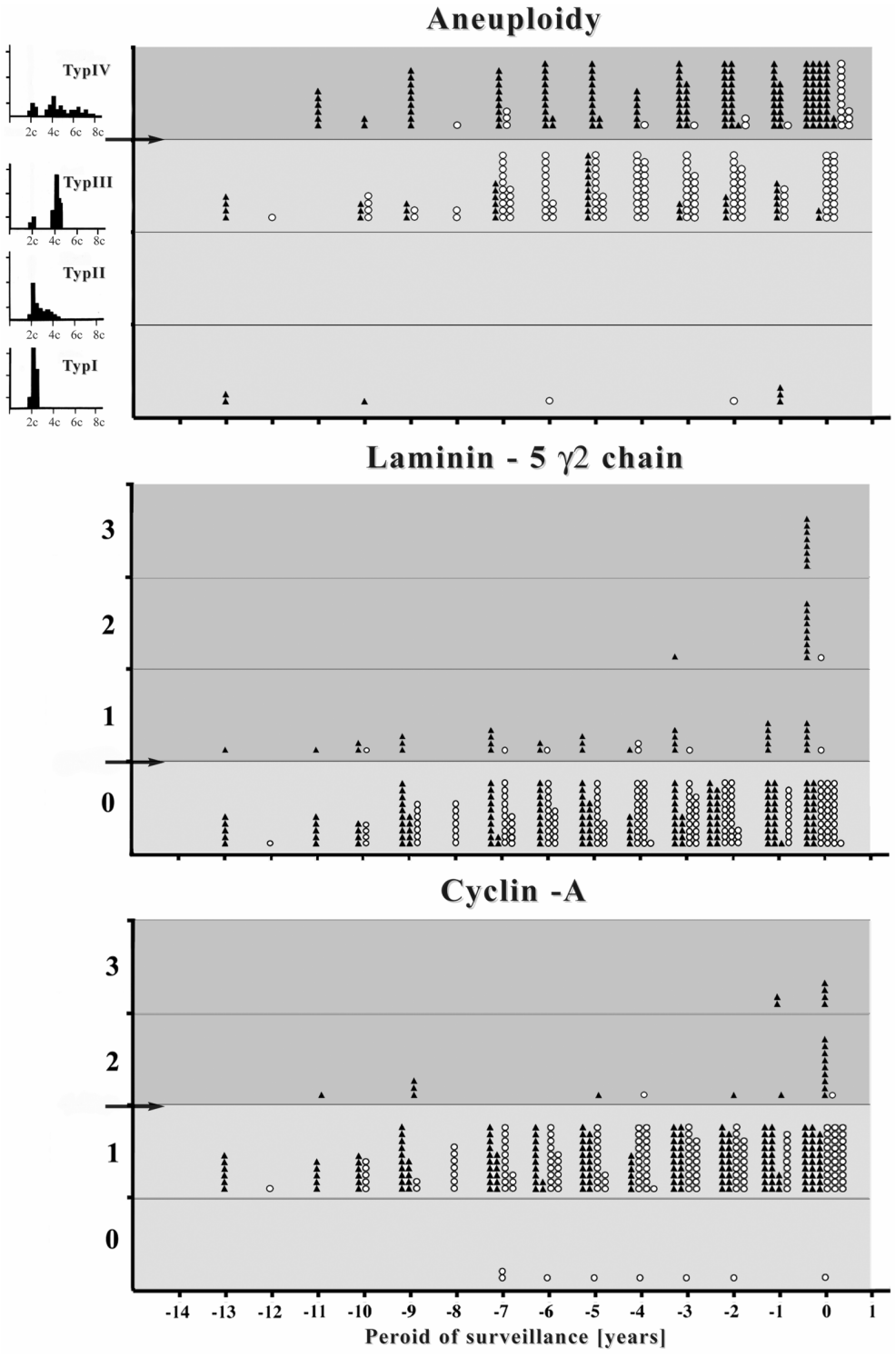
Little is known about laminin-5 γ2 chain immunoreactivity in precancerous lesions of malignant human tumors. Previous data from our group had shown various degrees of cytoplasmic laminin-5 γ2 chain immunoreactivity in 96% of primary colon carcinomas, whereas staining was absent in stromal cells and adjacent normal colonic mucosa. There was also an interrelationship between a strong staining pattern and a worse clinical outcome. [159] In the present study, seven of the eight UCC specimens had a moderate or strong γ2 chain expression, which was observed predominantly in malignant cells at the invasive front (**Figure 2, Paper 1**). An interesting finding in patient group A was the detection of immunoreactivity in biopsy specimens up to 13 years prior to the subsequent carcinoma (average 8 years). The overall expression of laminin-5 was significantly more frequent throughout the entire observation period in group A biopsies (20%) than in those of group B (5%, $p = 0.002$). Laminin-5 γ2 chain positive cells were distributed over the whole colon and rectum and were not correlated with dysplasia ($r = 0.357$) or inflammatory activity ($r = 0.142$), but interestingly, related to aneuploidy: in group A as many as 26 of 37 of the laminin-5 immunopositive specimens were aneuploid. In group B, only 2 of 8 immunopositive biopsies showed aneuploidy and the rarely occurring laminin-5 positive cells were generally localized close to the basement membrane, mainly in flat, regenerative epithelium. Thus, the observation of laminin-5 immunoreactivity years prior to a UCC might not just represent the ability of cells to invade but also to wound healing. In fact, strong laminin-5 expression has also been observed in migrating keratinocytes in healing skin wounds. [160] The observed phenomenon could thus be effected by the underlying inflammatory disease. [161,162] Laminin-5 γ2 chain overexpression in repeated biopsies might be related to upregulated regenerative processes in ulcerative colitis. However, since normal regenerative processes and wound healing generally occur in diploid cell populations, the combined analysis of ploidy and laminin-5 γ2 chain expression may allow identification of premalignant populations with invasive capacity. The present data strongly indicate an increased risk of progression to invasive properties in genetically unstable cells. However, aneuploidy may be reversible over time once cells are not longer exposed to the inducing agent or carcinogen. [163,164] Thus, it is reasonable to suggest that the genomic instability reflected by aneuploidy has to be followed by multiple cellular alterations in order

to reach malignant properties. One of the decisive steps in this transformational process is the ability of genomically altered cells to proliferate, which is compulsory for clonal expansion. [165]

In the present study, six out of eight UCC specimens exhibited an increased cyclin A expression pattern. In addition, cyclin A expression was found in 98% of all biopsies, with a higher number of immunopositive cells in group A biopsies ($p = 0.014$), as well as being mainly observed in aneuploid populations: in group A as many as 12 of 13 biopsies with increased cyclin A staining were aneuploid, whereas in group B only one of four biopsies with similar staining intensity showed aneuploidy. Thus in group A, increased cyclin A expression was significantly correlated to aneuploidy ($r = 0.791$). However, there was no significant correlation with inflammatory activity ($r = 0.178$), grade of dysplasia ($r = 0.485$) or laminin-5 immunopositivity ($r = 0.140$). Since cyclin A expression indicates whether a cell is committed to pass through the cell cycle and divide, i.e. participate in clonal expansion, the fraction of cyclin A positive cells may be used to estimate the risk of aneuploid populations to progress to malignancy. At the study's conclusion and establishment of a risk profile, based on the three parameters discussed above, it was discovered that six out of the 16 patients in group B could be identified as high-risk patients (**Figure 1, Paper I**). Of these six patients, one patient developed an invasive carcinoma after the endpoint of the study. By fitting a logistic regression model, DNA-cytometry, laminin-5 positivity and increased cyclin A expression were confirmed as significant predictors for malignant transformation. Among these, laminin-5 immunopositivity showed the strongest prognostic correlation.

Following page:

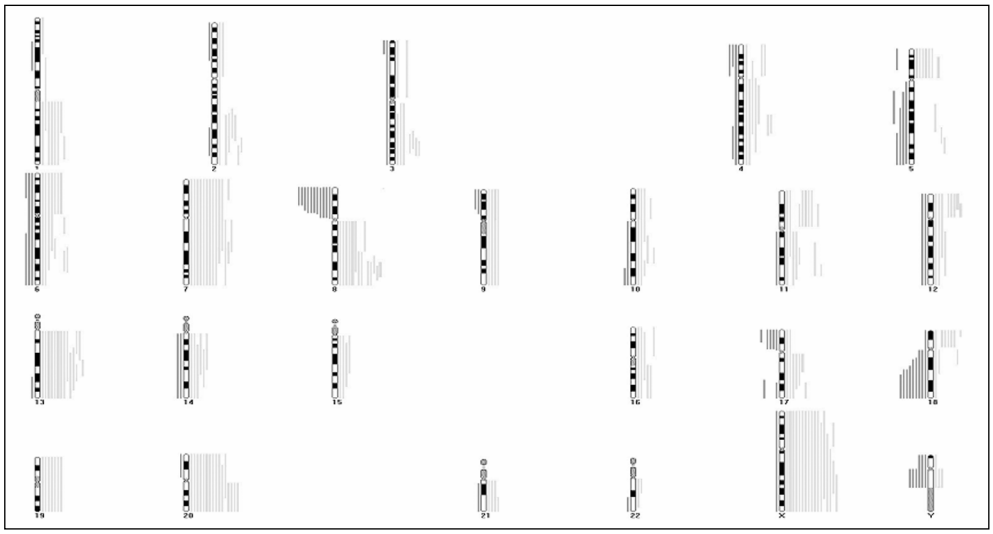
Figure 2, paper I. DNA histogram type, degree of laminin-5 γ2 chain and cyclin A expression (0 = 0%, 1 = <20%, 2 = 20 – 50% and 3 = >50% immunopositive mucosal cells) in 8 patients with (group A = ▲) and 16 patients without UCC (group B = ○). Cutoff points to high risk levels are marked by arrows, which allow the definition of 6 patients out of group B with a high UCC risk. The endpoint of the study is represented by year 0 on the x-axis.


Aneuploidy
TypIV
TypIII
TypII
TypI
2c 4c 6c 8c
Laminin - 5 γ2 chain
3
2
1
0
Cyclin -A
3
2
1
0
-14 -13 -12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1
Peroid of surveillance [years]

Paper II

Detection of aneuploid lesions in ulcerative colitis patients seems to indicate imminent carcinogenesis with faithful progression to ulcerative colitis-associated colorectal carcinoma (UCC). Recent reports have provided evidence that genomic aneuploidy or imbalances observed in UCC, cluster on the same chromosomes. [166-169] Unlike sporadic colorectal tumors, UCCs do not follow the adenoma–carcinoma sequence, and the sequential acquisition of chromosomal aneuploidy and gene mutations is less well established. It was therefore interesting to investigate whether the pattern of chromosomal gains and losses in UCC are similar to that described in sporadic carcinomas. This would indicate that the final distribution of genomic imbalances is the product of continuous selection, and that this distribution is independent of whether a carcinoma occurs spontaneously or as a result of, for example, chronic inflammation. The aim of this study was therefore to identify specific genomic imbalances in ulcerative colitis–associated colorectal carcinomas and to establish a comprehensive map of DNA gains and losses by investigating carcinoma specimens from 23 patients. The degree of genomic instability was determined by DNA cytometry and comparative genomic hybridization (CGH). Markers for proliferative activity (cyclin A) and invasive potential (laminin-5) were determined with immunocytochemistry. All parameters were compared with tumor stage and grade, clinical features, and patient survival.

Figure: Summary of genomic imbalances in 19 ulcerative colitis-associated colorectal cancer specimen defined by CGH. Bars on the left side of the ideogram denote a loss of sequence in the tumor genome, while bars on the right side designate a gain. Bold-face squares or bars indicate high-level copy number increases (amplifications).



All 23 UCC specimens revealed highly aneuploid DNA distribution patterns of the nuclear DNA content, independent of the tumor stage. CGH analysis could be performed on 19 UCC specimens, with all of them showing chromosomal imbalances as follows: the most common DNA gains were mapped to chromosomes or chromosome arms 20q (84% of all cases), 7 (74%), 8q (74%), 13q (74%), 11p and 12 (both 42%), 5p and 18p (both 37%), and 17q (31%). Recurrent losses occurred on 8p (58%), 18q (47%), and 5q (26%) (**Figure 2, Paper II**). These results show that chromosomal imbalances observed in UCC mainly cluster on the same chromosomes as described for sporadic colorectal cancer. For instance, Ried et al reported DNA gains that frequently mapped to chromosomes or chromosome arms 7, 8q, 13q, and 20 in sporadic colorectal carcinomas. [60] However, it also becomes clear that sporadic colorectal carcinomas have fewer genomic imbalances than UCCs (**Figure 3, Paper II**). Additional significant differences exist that characterize ulcerative colitis-associated colorectal carcinomas in contrast to sporadic carcinomas. Our previous analyses of sporadic colorectal carcinomas revealed an average number of DNA copy alterations (ANCA, calculated as the number of chromosomal copy number changes divided by the number of cases) of 5.6, which was elevated to 13.3 in UCC. This number exceeds that observed in primary liver metastases from colorectal carcinomas, for which the ANCA had been determined to be 11.7. [129] This high degree of genomic instability is also supported by measurements of the nuclear DNA content, which invariably revealed gross aneuploidy. We also observed a large number of localized high-level copy number increases (amplifications). Amplifications have been described as a reflection of advanced disease and poor prognosis in other malignancies. [170] Some of the amplifications occurred in regions known to be affected in colorectal carcinomas, such as chromosome arms 6p, 8q, 13q, 17q, and 20q, and for which the target genes are either known or likely candidates have been identified (http://www.helsinki.fi/cmg/cgh_data.html). For instance, the frequent gain of chromosome 8 and amplifications that map to band 8q24 target the *MYC* oncogene. Candidates on chromosome 20 include the nuclear co-receptor activator gene *NCOA3* and a member of the aurora kinase family; however, amplicons that map to chromosomes 5p, 12p, 12q, 9p, 10p, 11p, and 18p were not known to be associated with colorectal cancers prior to this study. Interestingly, most of these amplifications appear in very distinct locations that allow one to identify possible candidate genes (**Figure 1, Paper II**). The amplification on 12q could indicate copy number increases of *MDM2* and could therefore denote an alternative pathway to *TP53* inactivation. [171,172] Also, the 12p amplification could point to an as yet rarely described mechanism of *KRAS* upregulation, whose activation by way of point mutations, but not genomic amplification, has been described in colorectal carcinomas. [173] Another correlation is the coincidental overexpression of laminin-5

and gain of chromosome band 1q25~q31, the map position of the *LAMC2* gene (laminin-5). According to our previous results, laminin-5 seems to serve as an independent predictor of cancer risk in UCC, but it has not been elucidated how increased expression levels are produced. Genomic amplification could be one molecular mechanism leading to laminin-5 overexpression.

The CGH profile for UCC as presented here, dominated by overall gains and numerous amplifications, is in concordance with the relatively high ANCA value and severe aneuploidy observed in the majority of all 23 UCCs. In comparison, sporadic colon carcinomas show aneuploidy in only 70%–80% of the cases, combined with an overall lower ANCA value. The surprisingly high level of ANCA values in UCC could be a reflection of a generally increased genetic instability in UCC, due to the long latency of inflammatory disease before overt tumors develop; however, the data presented here and in the literature clearly indicate that the tumor cell population as an entity selects for a distribution of genomic imbalances that is similar to sporadic carcinomas. Therefore, the tissue origin of the tumor cell, and not the mode of tumor induction, seems to define the similarity between sporadic colorectal cancers and UCC. This is in striking contrast to hereditary colorectal carcinomas arising in the background of mismatch repair deficiency, where neither aneuploidy nor specific chromosomal imbalances are observed. [174,175]

The correlation between survival and grading ($p = 0.008$), staging ($p = 0.03$) and presence of metastases at the time of surgery ($p = 0.001$) was significant. In contrast, neither high-risk factors for cancer development, such as the diagnosis of a primary sclerosing cholangitis in UC, nor participation in surveillance programs seemed to significantly influence a patient's outcome. These findings clearly suggest that additional and more reliable markers for early cancer detection are needed, especially within this subtype of colorectal carcinomas. In this respect, the positive correlation of high ANCA values and the occurrence of amplifications with elevated cyclin A expression ($p = 0.04$) indicate a high proliferative activity combined with genomic instability, necessary features for tumor growth and clonal expansion. This finding thus increases the value of cyclin A as an independent prognostic marker for carcinoma development in ulcerative colitis. The correlation between increased laminin-5 expression and the presence of metastases at the time of surgery ($p = 0.04$) could also point to those patients that – presenting elevated laminin-5 expression – are at higher risk for the development of micro- or macrometastases.

Paper III

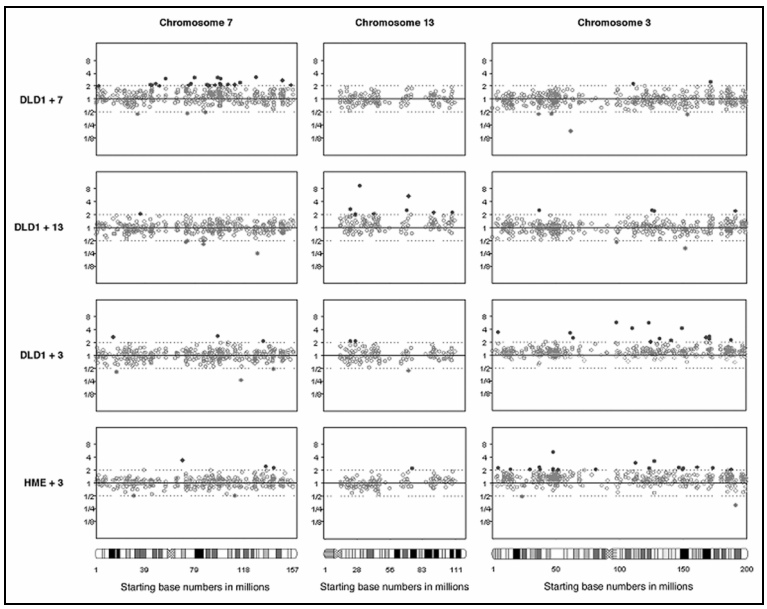
Chromosomal aneuploidies are not only observed in sporadic and ulcerative colitis-associated colorectal carcinomas but in essentially all sporadic carcinomas. These aneuploidies result in tumor-specific patterns of genomic imbalances that are acquired early during tumorigenesis. [109,115,176] For instance, one of the earliest genetic abnormalities observed in the development of sporadic colorectal tumors is trisomy of chromosome 7. [59] Usually, once acquired, these specific imbalances are maintained despite ongoing chromosomal instability. [177] It is therefore reasonable to assume that continuous selective pressure for the maintenance of established genomic imbalances exists in cancer genomes. It is not known how genomic imbalances affect chromosome-specific gene expression patterns in particular and how chromosomal aneuploidy dysregulates the genetic equilibrium of cells in general. To model specific chromosomal aneuploidies in cancer cells and dissect the immediate consequences of genomic imbalances on the transcriptome, we set up a unique experimental model system in which the only genetic alteration between parental and derived cell lines is an extra copy of a single chromosome: we generated derivatives of the diploid yet mismatch repair-deficient colorectal cancer cell line, DLD1, and immortalized cytogenetically normal human mammary epithelial cell line (hTERT-HME) using microcell-mediated chromosome transfer to introduce extra copies of neomycin-tagged chromosomes 3, 7, and 13. Fluorescence in-situ hybridization (FISH) with chromosome-specific probes confirmed the maintenance of extra copies of these chromosomes under neomycin selective cell culture conditions (**Figure 1, Paper III**). In addition, spectral karyotyping (SKY) was performed to determine whether the chromosome transfer process induced secondary karyotypic changes. With the exception of loss of the Y chromosome in all DLD1 + 3 cells analyzed and in a small percentage of DLD1 + 7 cells, all four derivative lines maintained the diploid karyotype of the parental cell line and contained the additional copy of the introduced chromosome (**Figure 1, Paper III**). These results were also confirmed by CGH. The global consequences on gene expression levels were analyzed using cDNA arrays. Our results show that regardless of chromosome or cell type, chromosomal trisomies resulted in a significant increase in the average transcriptional activity of the trisomic chromosome ($p < 0.001$) (**Table 1 and Figure 2, Paper III**).

Thus, several important conclusions regarding the impact of chromosomal aneuploidy on cellular transcription levels can be drawn from our analysis. First, alterations in the copy number of whole chromosomes resulted in an increase in average gene expression for that chromosome. The average increase in gene expression (1.21) however, was lower than the average increase of genomic copy number (1.44). These results were consistent with results from similar analyses of

aneuploid colorectal, pancreatic and renal cancer derived cell lines in which we observed a trend, indicating that indeed chromosomal aneuploidies correlate with global transcription levels. [116,128,129] Second, chromosomes not observed to be aneuploid in particular tumor types (*i.e.,* chromosome 3 in colorectal tumors) also had increased transcriptional activity when placed into that cellular environment. Thus, their presence is not neutral with respect to the transcriptome. Third, aneuploidy not only affects gene expression levels on the chromosomes present in increased copy numbers but a substantial number of genes residing on other chromosomes significantly increased or decreased, apparently in a stochastic manner: The influence of chromosomal aneuploidy on the expression level of individual genes was examined by considering only those genes whose expression ratios were >2.0 (up-regulated) or <0.5 (down-regulated) when compared with the parental cell line. For clone DLD1 + 7, 155 genes were up-regulated and 47 down-regulated (23%) beyond these thresholds. Of those genes with known map location, 18% (35 of 194) map to chromosome 7 and 32 of these were up-regulated (**Table 2, Paper III**). Regarding chromosome 13, 92 genes were up-regulated and 72 were down-regulated (44%). Of those with known map location, 6% (10 of 162) mapped to chromosome 13, of which all were up-regulated (**Table 3, Paper III**). The introduction of chromosome 3 in DLD1 resulted in up-regulation of 81 genes and down-regulation of 67 genes (45%). Here, those with known map location, 12% (17/144) mapped to chromosome 3 and all were up regulated (**Table 4, Paper III**). The introduction of this same chromosome 3 into the hTERT-HME cells resulted in the up-regulation of 91 genes and down-regulation of 49 genes (35%). Of those with known map location, 17% (23 of 135) mapped to chromosome 3 and 21 were up-regulated (**Table 5, Paper III**). Strikingly, none of the genes were affected in common among any of the four cell lines. This observation is of course consistent with known mechanisms of gene regulation (*e.g.,* activator and suppressor proteins, signaling pathways) and the fact that genes residing in a given pathway, are for the most part distributed throughout the genome on different chromosomes. Three groups have analyzed the consequences of constitutional chromosomal trisomies on transcriptional activity in noncancerous fetal cells and in a mouse model of human trisomy 21, and attained similar conclusions as us. [178-180] These studies concluded that the average gene expression of trisomic chromosomes is clearly increased, although this was not due to high-level up-regulation of only a few specific genes. However, expression levels of multiple genes throughout the genome were dysregulated. Normal cells with constitutional chromosomal aneuploidies (or segmental duplications) cannot tolerate trisomy of >4.3% of the genome. [181,182] However, this limit is not merely a reflection of the DNA content because multiple copies of an inactivated X-chromosome can be tolerated. Therefore, this limit is more likely to be imposed by global disturbance of the

transcriptome as a consequence of genomic imbalances. This hypothesis is supported by the recent identification of differential average expression levels of specific chromosomes. For instance, the average gene expression levels and gene density of chromosomes 13, 18, and 21 are lower than those of smaller chromosomes, yet trisomy of only these chromosomes is compatible with life in noncancerous cells. [182] It is interesting to speculate that one of the specific features of emerging cancer cells, which would differ from nontransformed cells that carry constitutional trisomies, would be the ability to exceed this transcriptional threshold during the multiple steps required for tumorigenesis. This global dysregulation of the transcriptome of cancers of epithelial origin may also reflect on our ability for therapeutic intervention: although the consequences of a simple chromosomal translocation, such as the *BCR/ABL* fusion in chronic myelogenic leukemia, can be successfully targeted with an inhibitor of the resulting tyrosine kinase activity such as Gleevec, the normalization of the complex dysregulation of transcriptional activity in carcinomas requires a more general, less specific, and hence more complex interference. [183]

Figure: Global gene expression profiles. Each scatterplot displays all genes and their corresponding normalized ratio values along the length of each chromosome. Values in open circles represent ratio values between 0.5 and 2.0. Dark dots represent expression ratios ≥ 2.0 or ≤ 0.5.



PAPER IV and V

A correlation of tumor phenotype to genotype has been firmly established, yet significant advances in correlative gene and protein expression profiling are lacking. We present a detailed analysis to identify sequential alterations of the genome, transcriptome, and proteome that define the transformation of normal epithelium and the progression from adenomas to invasive disease. We have analyzed tissue samples from 36 patients, including the mucosa-adenoma-carcinoma sequence from eight individual patients, paired normal mucosa and carcinoma samples from an additional eight patients, and paired carcinoma and metastasis from two other patients. We determined the degree of genomic instability during carcinogenesis by measuring DNA contents, performing CGH experiments, and analyzing global gene expression profiles using cDNA arrays. Microarray-based gene expression profiling was complemented by identification of protein expression levels.

Our CGH analyses confirmed the importance of specific chromosomal aneuploidies during colorectal tumorigenesis, such as gains of chromosomes and chromosome arms 7, 8q, 13q, and 20q, as well as losses that map to 8p, 17p, and 18q. This strict conservation of genomic imbalances that so clearly defines colorectal carcinomas prompted us to identify the immediate consequences of chromosomal aneuploidies on gene expression levels. Our results indicate that there is, in general, a significant correlation between chromosomal copy number and average transcriptional activity of resident genes. This correlation applies to almost all chromosomes that are frequently affected in colorectal tumors; however, a few exceptions exist. For instance, gene expression levels on the long arm of chromosome 8 did not seem to be up-regulated by genomic copy number increases. It remains to be established whether this particular effect on chromosome arm 8q is a reflection of local transcriptional down-regulation, e.g., through chromatin remodeling events, or whether down-regulation of expression levels of genes on chromosome arm 8q is a result of gene expression changes that occur elsewhere in the genome.

Unsupervised principal component analysis allowed separation of adenomas, carcinomas, and metastases based on gene and protein expression profiles. Interestingly, two dysplastic polyp samples did not conformingly cluster in their cohort and were closer located to the malignant samples. Both polyps revealed aneuploid DNA distribution patterns, indicating an increased malignancy potential.

A total of 368 genes were differentially expressed between the four stages of colorectal cancer progression ($p<0.001$). In detail, 58 genes were differentially expressed between normal mucosa and adenoma (20 genes were upregulated, 38 genes had lower expression levels); 116 genes were differentially expressed

between adenoma and carcinoma (80 elevated, 36 decreased), and 158 genes were differentially expressed between primary carcinoma and liver metastasis (138 elevated, 20 decreased). The differentially expressed genes are involved in various cellular pathways and networks as identified with Ingenuity Pathways Analysis (IPA) software (Ingenuity, Mountain View, CA 94043). The term "pathway" is used to refer to canonical pathways that annotate genes of interest to known metabolic or signaling cascades as described in the Kyoto Encyclopedia of Genes and Genomes (KEGG, www.genome.jp/kegg). Networks however, are determined by IPA as groups of genes that directly interact with a defined gene of interest, i.e., in our analysis a gene that is significantly differentially expressed among certain stages of colorectal tumorigenesis. "Focus genes" are defined as genes that were differentially expressed ($p<0.001$) in our group comparisons and that additionally were part of the IPA database genes. IPA networks can include several genes or proteins, and also allow identification of indirect interactions between focus genes. Thus, focus genes are used to map molecular networks that indicate how these genes of interest may influence each other above and beyond already known interaction maps, derived from, e.g., KEGG pathways. The IPA generated networks (NW) are listed in a certain order with the top NWs having a lower likelihood that the generation of networks was serendipitous. The application of IPA-based algorithms to our data set identified 50 individual pathways that were either directly or indirectly affected by gene expression changes of at least one focus gene. We then selected further for pathways that were affected in more than one stage comparison and identified 20 pathways. Of these 20 pathways, 14 were directly affected by four focus genes, which reflects the multiple functions of any one of these focus genes (NFKB2, YWHAE, NP, PRKCG). While the number of dysregulated genes increased sequentially during progression from one stage to another, we observed redundancy in terms of involvement of gene networks and affected pathways. For instance, for all three comparisons, the genes that were sequentially up- or down-regulated during malignant transformation belonged to networks that influence signaling pathways involved in cell cycle control, apoptosis, angiogenesis, and proliferation. Consistent to three prior independently conducted studies, we observed a specific up-regulation of SLC12A2 (Solute carrier family 12, member 2), and SORD (Sorbitol dehydrogenase), and the pronounced down-regulation of SEPP1 (Selenoprotein P, plasma, 1), providing supporting evidence that all three genes might play a yet underestimated role in the progression of colorectal cancers. [184-187]

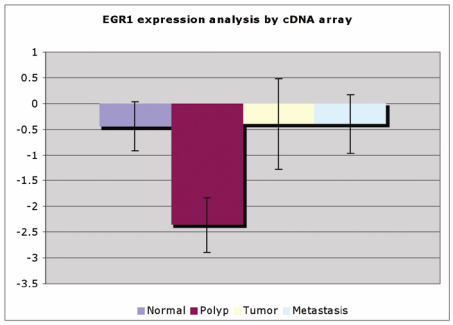
We also found that expression levels of 616 genes increased constantly through the progression from normal mucosa to adenoma, carcinoma, and metastasis, while they decreased for 1100 genes. We also extracted from our dataset expression levels of genes known to be involved in colorectal tumorigenesis. Very few

reached the significance levels required in our analysis, however, most changed in intuitive directions. Few exceptions exist, most notably EGFR, which was down-regulated during sequential steps of carcinogenesis. The reason for this discrepancy remains unresolved at this time.

A total of 10 genes were analyzed by quantitative reverse transcriptase-PCR for all different stages using three different housekeeping genes (GAPDH, GUS, 18S) as internal controls. These experiments confirmed the array expression values for eight of the 10 genes (APC, MYC, EGR1, EPHA7, HRG, MFAP4, SERPINA1, CLCA1), thus showing the reliability of our array expression data. For two of the genes (OPCML and EAF1), the array derived expression values could not be confirmed for all individual stages.

Some of the samples included in this study were also analyzed for protein expression changes using two-dimensional gel electrophoresis. A total of 112 polypeptide spots showed an at least two-fold differential expression between the four stages of carcinogenesis. A total of 72 of these polypeptides could be characterized by mass spectrometry and 46 of those were exclusively over-expressed in tumors and metastases. Interestingly, these 72 polypeptides belonged to only 42 individual proteins, suggesting a high frequency of posttranslational modifications, such as phosphorylation. For 27 of these 42 proteins, the corresponding cDNA was included on our microarray platform, allowing comparison of the results with gene expression changes. While there was no direct match of the 42 differentially expressed proteins and genes, the majority of differentially expressed genes and proteins were related to the same canonical pathways. For 26 differentially expressed proteins with corresponding cDNA expression analysis, a total of 12 corresponding genes showed a similar trend in transcriptional expression as observed for translational changes.

Figure: The average gene expression levels of EGR1 during sequential steps of colorectal carcinogenesis assessed by cDNA microarray (left) and quantitative-PCR (right).



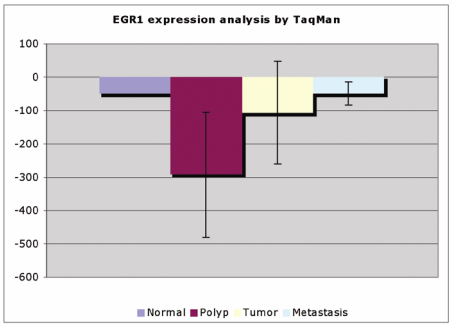
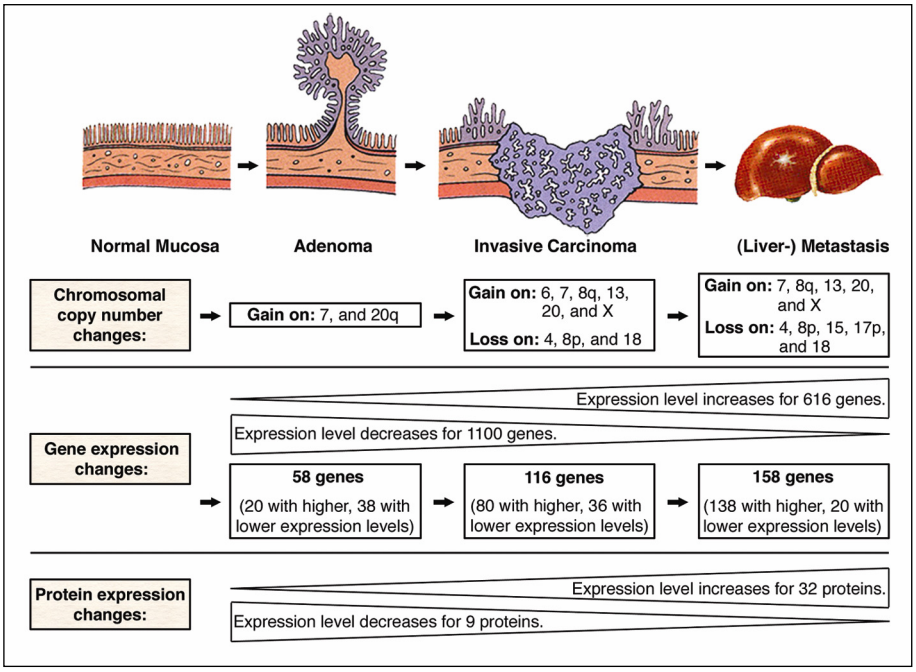


Figure: Summary figure of genomic, transcriptional and proteomic changes along the "adenoma-carcinoma-sequence" of colorectal carcinogenesis.



Paper VI

Against the background that distinct gene and protein expression patterns correlate with subsequent stages of colorectal cancer progression, we were then eager to investigate if a malignancy specific protein pattern could potentially be detected not only in clinical tissues but also in patient serum samples. We therefore conducted a comprehensive serum protein profiling in 88 patients with sporadic colorectal malignancy and in 51 healthy individuals using SELDI-TOF mass spectrometry. The first step of the analytical procedure focused on the characterization of 32 healthy controls and 52 samples with colorectal malignancy. This training set was used to identify differentially expressed m/z values, or features between the two groups. These features were then used to classify an independently collected series of serum samples (n=55). This validation set was analyzed in a double-blinded fashion.

The comparison of 52 patients with colorectal malignancy with 32 healthy controls revealed a set of 13 features that were chosen in different combinations by K-Nearest Neighbors to distinguish the malignant sera from healthy control sera. The 13 discriminative features were located at the following m/z values: 2172.8, 2591.1, 2646.9, 5715.9, 7005.2, 7568.9, 7683.6, 7722.6, 7905.8, 9148.7, 9556.6, 14653.8, and 14698.6 Da. One of the m/z values, at 9148.7 Da, was selected by all 16 models and therefore appeared to be the strongest single discriminative feature. Since the control sera were collected from significantly younger individuals as compared to the malignant sera (**Table 1, Paper VI**), we analyzed each selected feature for the possibility that the observed differences might simply be a reflection of age. We could not detect any age-dependent expression of certain features; for instance, the most dominant feature at 9148.7 m/z revealed a Pearson's correlation coefficient of expression levels and age of r = -0.04221, indicating that there is no correlation between expression levels and age. The analysis of the training set therefore suggested that serum profiling using SELDI-TOF identifies protein peaks that allow the discernment of patients with colorectal malignancy from control individuals. The predictive value of these 13 features was then tested with an independently collected, blinded validation set consisting of 55 samples. For this purpose we used 16 models based on K-Nearest Neighbors comprising of two, three, four or five selected features and six to nine neighbors. Only samples that revealed identical results with all 16 models were considered for further classification into the malignant or control group (**Supplementary Table 2, Paper VI**). Applying this criterion, 48 of 55 samples were classifiable. Of those 48 samples, 47 were classified correctly (98%). The sera from the seven individuals that revealed ambiguous class predictions - i.e., not all 16 classifiers agreed with one another - belonged to two healthy controls (a 30-year-old female and a 35-

year-old male) and five tumor patients with carcinomas classified as T3N0MX, T3N1MX (twice), T2N0MX, and T1N0MX. One of these patients had a synchronous gastrointestinal stromal tumor. Three of the five tumors were localized at the right colon, a location that was underrepresented in the training set (**Table 1, Paper VI**). Four of the five non classifiable tumors had been operated with curative attempt (R0 resection), whereas one tumor could be resected only incompletely (R2 resection). The latter tumor was one of three inoperable tumors that were included in the validation set. However, only one advanced, inoperable tumor was included in the training set, and hence its characteristics were underrepresented for classifier training. This might also explain that the only misclassified sample represented an inoperable tumor (a 63-year-old male patient with an advanced staged rectal carcinoma with synchronous metastasis (UICC IV)). The technical repeats for this sample were analyzed separately and assigned to the healthy control group with very high probability (55% to 97% likelihood). This rather surprising lack of sensitivity for the detection of advanced tumors could be due not only to the under-representation of such advanced malignancies in the training set but potentially also due to the following factors: (A) Highly advanced tumors with early metastasis potential could have a different protein profile than primary tumors and/or their metastasis alone. Certain proteins could be expressed exclusively in these advanced tumors and - compared to features that are normally expressed from less advanced cancers - could reach higher peak intensities. Due to normalization, the dominant features used for classification would be scaled down and thus the advanced tumor would not be classifiable or even determined to be normal. (B) Another possible explanation is that dominant features used for classification might be expressed only temporarily and become less prominent or disappear during progression towards advanced tumor stages. A repeated SELDI-TOF analysis of the misclassified sample, together with an additional serum sample from the same patient showed that the spectra of this particular patient were then defined as outliers; confirming our above discussed hypotheses. It is to emphasize that the classifiers that we applied were designed more for the identification of potential biomarkers, rather than for the creation of the best possible classifier. This is different from several previous studies that employed SELDI-TOF mass spectrometry for detection of, e.g., ovarian carcinomas. In these studies, an overall scaling of each spectrum was performed once instead of scaling each feature over the set of samples or scaling over a selected set of features. In addition, our classifiers were only allowed to use a small number of features (2 to 5) and a relatively large number of neighbors (6 to 9); these stringent conditions rendered it very difficult to produce a fortuitous separation of samples into regions containing members of the same class.

The observation that two control and five malignant serum samples were not classifiable does not impair the usefulness of SELDI based KNN-classification as a powerful screening tool. Individuals whose sera cannot be classified with certainty would have to undergo conservative medical examination. None of the normal samples were assigned to the malignant group. The SELDI based KNN-classification presented here, identified 13 features that permitted the discrimination of colorectal cancer-associated sera from healthy controls in an independent, blinded validation with 96.7% sensitivity, 100% specificity, 100% positive predictive value, and 94.4% negative predictive value (**Table 1, Paper VI**). In comparison, Cordero et al. determined a sensitivity and specificity of 90% for colorectal cancer diagnosis by using preoperative serum CD26 levels. [188,189] Any other diagnostic approach, such as detection of CEA or VEGF, TIMP1, circulating D-dimer, IL-6, or CRP concentrations, either alone or in combination, reaches only lower diagnostic efficiencies. [92,190-195] Real time quantification of human telomerase reverse transcriptase mRNA in plasma obtained a sensitivity of 98%, but a specificity of only 64%. [196] None of these markers showed high enough discriminative power to become implemented in clinical tumor screening. One similar approach to detect colorectal malignancy by SELDI-TOF based protein profiling was performed by Yu et al. [197] Their results suggest that colorectal cancer patients can be discriminated from healthy controls with 89% sensitivity and 92% specificity based on four features. These four features showed different m/z values than the 13 features identified in our analysis. Yu and colleagues did not verify their findings with an independent validation set. [197]

Early detection of cancer is a main-important clinical goal. A total of 16 tumor samples in the training set tested here were UICC stage I and II, i.e., early stage tumors. The independent validation set contained nine such tumors. Our data did not show a correlation between peak intensity and tumor stage, thus allowing a sensitive and specific class prediction also for early stage colorectal carcinomas.

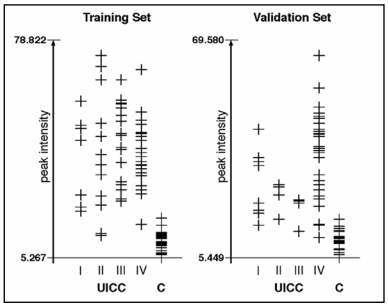


Figure: Ratio plot of peak intensities for the feature 9148.7 m/z used in all 16 prediction models for the training (left) and validation (right) set. The peak intensities (one cross per spectrum/sample) are grouped according to the UICC stage of the malignancies (first four columns to the left, respectively) and are compared to peak intensities of the control (C) spectra (right column). Note that there is no correlation between peak intensity and the UICC stage in either the training or the validation set.

CONCLUSIONS

Genomic aneuploidy occurs early and is commonly found in precancerous biopsies of ulcerative colitis patients who subsequently develop an ulcerative colitis-associated colorectal carcinoma (UCC). The assessment of DNA ploidy could therefore become a basic element in future surveillance programs in ulcerative colitis. The complementary detection of laminin-5 positivity and increased cyclin A expression in aneuploid lesions - indicating invasive potential and clonal expansion - seems to be the most powerful combination to predict imminent malignant transformation for an individual patient. Moreover, genomic aneuploidy in UCC tumors correlates with specific chromosomal gains and losses, which, in turn, are associated with increased cyclin A levels. The molecular characterization of distinct amplifications that were not known to be associated with colorectal cancer will likely reveal novel genes involved in colorectal tumorigenesis. However, the overall pattern of specific chromosomal aberrations in UCC tumors is similar to that seen in sporadic colorectal carcinomas. Given the predominance of specific chromosomal aneuploidies in colorectal cancers, we were now interested to explore on how these aneuploidies affect the transcriptome of cancer cells. Therefore, we developed a well defined model system in which we could show that the introduction of an extra copy of a given chromosome increases very specifically the overall average expression of genes on the trisomic chromosomes. Additionally, a large number of genes on diploid chromosomes were also significantly increased, revealing a more complex global transcriptional dysregulation. Analogous results could be obtained while deciphering how chromosomal aneuploidies affect the transcriptional equilibrium of emerging cancer cells in primary tumor samples: we found that specific and recurrent chromosomal aneuploidies exert strong and direct influence on gene expression levels of resident genes on the affected chromosomes, therefore enhancing the complexity of transcriptional dysregulation in colorectal carcinomas and associated metastases. In addition, increasing genomic instability and a recurrent pattern of chromosomal aberrations are accompanied by distinct gene- and protein expression patterns that correlate with subsequent stages of colorectal cancer progression. Our analysis also identified novel, cancer-associated transcripts. Proteomic analysis revealed differentially expressed proteins that were related to similar canonical pathways as the differentially expressed genes. The identified proteins underwent extensive posttranslational modifications, thus multiplying the transcriptional dysregulation. In addition, a malignancy related protein profile could be identified also in serum samples: comprehensive serum protein profiling in patients with colorectal malignancy and healthy individuals using SELDI-TOF mass spectrometry could reveal 13 differentially expressed features. Using 16 classifiers based on these features allowed a highly sensitive and specific class prediction in

an independently collected, blinded validation set. SELDI based serum protein profiling thus enables the distinction of sera from colorectal cancer patients, including those with early stage disease, from healthy controls.

FUTURE PERSPECTIVES

The evolving technique of array CGH allows the identification of DNA copy number changes of, ideally, individual genes and thus enables an increasing resolution compared to conventional CGH, which is performed on more or less condensed chromosomes. One particular application could be the analysis of amplicons identified in sporadic and ulcerative colitis-associated colorectal carcinomas. Custom designed arrays that contain genes located on distinct amplified chromosome segments would enable to select for individual genes rather then chromosomal segments that are highly amplified. Such genes could be used as marker genes for colorectal malignancy. Their diagnostic and prognostic potential could be tested with gene specific probes by means of multi-color fluorescence in situ hybridization (FISH) in premalignant lesions (e.g., ulcerative colitis biopsies, adenomas). Thus, the combined analysis of DNA ploidy measurements and colorectal cancer specific Multi-FISH probes in premalignant colorectal lesions could profoundly improve individual risk assessment for imminent colorectal cancer development. However, the design of customized array CGH chips that contain the amplified genes might provide a more rapid and high throughput screening approach compared to Multi-FISH. The application of array CGH would also allow a direct correlation how single gene copy number changes influence the transcriptional equilibrium. The employment of comprehensive gene and protein expression profling in subsequent stages of colorectal cancer progression allowed the identification of genes and proteins that now warrant further validation by, e.g., RNA interference (RNAi) experiments, in order to prove their potential for gene and protein expression tailored individualized diagnostic, prognostic, and therapeutic approaches. The identification of malignancy related serum features by means of SELDI protein profiling and their high accuracy for class prediction in a blinded valdiation set now warrants protein identification by subsequent mass spectrometry and the validation of these identities by, e.g., ELISA technique in even larger sample sets collected from different clinics. The combined evaluation of ploidy status, amplification of disease and stage specific gene probes, and gene and protein expression patterns in clinical tissue samples should be utilized in prospective studies to corroborate their value for improved diagnostics, prognostication, and identification of therapeutic targets. The detection of malignancy related proteins in the serum might provide a rapid, sensitive and specific screening method for colorectal malignancies even for early disease stages.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to all my friends and colleagues who have supported me throughout the work of this thesis. I am in particular grateful to

Professor Gert Auer, my head supervisor, for welcoming me to the Karolinska Institute and allowing me to conduct my projects at different laboratories. I am especially grateful for all our highly interesting conversations and your enormous enthusiasm, encouragement, and support.

Professor Thomas Ried, my supervisor at the NCI, Bethesda, for welcoming me in his lab. I am especially thankful for introducing me to the world of fluorescence in situ hybridization techniques, our most stimulating discussions and your friendship.

Professor Hendrik Schimmelpenning, my supervisor of my german doctoral thesis, for introducing me into the world of research. I am especially thankful for your never ending guidance and friendship.

Professor Hans-Peter Bruch, head of the Department of Surgery at Lübeck University, for your outstanding support. This work would not have been possible without your generosity.

All the members of the Auer group, Ayodele Alaiya, Ulla Aspenblad, Susanne Becker, Juan Castro, Kaoru Fujioka, Barbro Larsson, Helena Lexander, Inga Maurin, Carin Östring, Görel Söderblom, Birgitta Sundelin, and Kicki Svanholm for always being helpful.

All current and past Riedlab members, especially Turid, Hesed and Bill, Kerstin and Bassem, Madhvi and Barg, Nicole, Linda, Cristina, Raluca, Diana, Danny, Mike and Simone, Buddy and Chin-Law, Tricia, John and Kate, Kundan and Richa, Joseph, Tom, Kuolin, Chris, and Andy. Thanks so much for all your support and friendship.

Lisa McShane and Ed Korn for their extraordinary professional help and guidance in gene expression data analysis. I am looking forward to our next collaborations.

DaRue Prieto, Tim Veenstra, and Brian Luke for our exciting collaboration and the introduction into serum protein profiling. Your expertise was the basis for our results.

My colleagues and friends in Lübeck at the Department of Surgery and Laboratory for Surgical research, especially Professor Rainer Broll, Michael Duchrow, Vera Grobleben, Dagmar Gebhardt, Regina Kaatz, Gisela Grosser-Pape, Elke Gheribi, Annemarie Aumüller, Barabara Scheurer, Carsten Pichler, Claudia Killaitis, Elisabeth Oevermann, Andreas Unger, Jan Nolde, Daniela Bauer, Gabor Horvath, Matthias Birth, Thomas Schiedeck, and Martin Strik for always being helpful and supporting.

Claes Lenander and Harald Blegen, my coauthors, for their support and successful cooperation.

All the people from the 4th floor at CCK from Professor Olle Larson's, Anders Zetterberg's and Klas Wiman's group. Especially, Ada, Leo, Cristina, Helene, Frederik, Bertha, Vladimir, Pädraig, Hanna-Stina, Margareta W., Margareta R., Lisa, and Tao for their friendship.

Professor Britta Norström, Caroline Lundgren, and Kristina Hellman for our interesting discussions and collaborations.

Professor Hans Jörnvall and his group at MBB, Tomas and Daniel, for being such generous collaborators.

Professor Thilo Wedel and Professor Wolfgang Jelkmann for stimulating discussions and sharing interesting projects.

Nils Homan, Thomas Witthoeft, Professor Dieter Ludwig, and Professor Eduard Stange, from the Gastroenterology Department at Lübeck University for our fruitful collaboration and interesting discussions.

My doctorands in Lübeck, Jana Doering, Nana Buendgen, Ulrike Paulsen, Constanze Brucker, Kathrin Overbeck, Stephanie Zillmer, and Marco Gerling for your help and all the time we spend together.

Sandra Freitag and Professor Lutz Dümbgen for introducing me into research statistics. Sandra, thank you for all your support, patience, and friendship through all these years.

Stefan Krüger and Professor A. Feller, from the Institute of Pathology at Lübeck University, for generous cooperating. Thank you, Stefan, for all your enthusiastic support, profound knowledge and friendship.

To Ulrike Kronenwett for interesting discussions, collaborative projects and your support to reach common goals.

To the NIH community of postdocs and PhD students for all the nice time we spend and all your support, especially but not exclusive Nicola and Roshni, Till, Runa, Yenan, Rosie, Gernot, Andy, Satoshi, Chun, Elke, Pascaline and Francois, Michal and Michael.

To Paul Savelkoul for his very professional and generous help and his friendship.

To Patricia Ried and Dara Wangsa for always being there with suggestions and help.

Sampsa Hautaniemi and Jonny Finlay for their friendship, our exciting conversations and fruitful collaborations.

Michel and Nora for making my stays in Stockholm always most comfortable. Thanks for your support and friendship throughout all years.

To my good friends Oliver and Dheepa, Hauke and Moni, Axel, Jens, and Andreas and Margit, and the entire Celle group. Thank you for being so supportive and understanding when I didn't have time to see you.

To Madhvi, for your never ending support and friendship. I am thankful for letting me benefit from your extraordinary skills. I am looking forward to our future cooperations.

To Uwe and Ruby, for their extraordinary support and friendship during all years.

To my dearest Martika, for always supporting and understanding me. Thank you for always believing in me.

To all my relatives and especially my parents, Gertrud and Ronald Habermann, my sisters Kristina and Inga and my brother in law Andrè. Without your support I could not have done this work.

References

1. Parkin, D.M., Pisani, P. & Ferlay, J. Global cancer statistics. *CA Cancer J Clin* **49**, 33-64, 1 (1999).
2. O'Connell, J.B., Maggard, M.A. & Ko, C.Y. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J Natl Cancer Inst* **96**, 1420-5 (2004).
3. Mak, T., Lalloo, F., Evans, D.G. & Hill, J. Molecular stool screening for colorectal cancer. *Br J Surg* **91**, 790-800 (2004).
4. Fleischer, D.E. et al. Detection and surveillance of colorectal cancer. *Jama* **261**, 580-5 (1989).
5. Bernstein, C.N., Blanchard, J.F., Rawsthorne, P. & Collins, M.T. Population-based case control study of seroprevalence of Mycobacterium paratuberculosis in patients with Crohn's disease and ulcerative colitis. *J Clin Microbiol* **42**, 1129-35 (2004).
6. Daperno, M. et al. The role of endoscopy in inflammatory bowel disease. *Eur Rev Med Pharmacol Sci* **8**, 209-14 (2004).
7. Eaden, J., Abrams, K., McKay, H., Denley, H. & Mayberry, J. Inter-observer variation between general and specialist gastrointestinal pathologists when grading dysplasia in ulcerative colitis. *J Pathol* **194**, 152-7 (2001).
8. Lynch, D.A., Lobo, A.J., Sobala, G.M., Dixon, M.F. & Axon, A.T. Failure of colonoscopic surveillance in ulcerative colitis. *Gut* **34**, 1075-80 (1993).
9. Riddell, R.H. How reliable/valid is dysplasia in identifying at-risk patients with ulcerative colitis? *J Gastrointest Surg* **2**, 314-7 (1998).
10. Schulmann, K., Reiser, M. & Schmiegel, W. Colonic cancer and polyps. *Best Pract Res Clin Gastroenterol* **16**, 91-114 (2002).
11. Lieberman, D.A. et al. Use of colonoscopy to screen asymptomatic adults for colorectal cancer. Veterans Affairs Cooperative Study Group 380. *N Engl J Med* **343**, 162-8 (2000).
12. Brown, M.O., Lanier, A.P. & Becker, T.M. Colorectal cancer incidence and survival among Alaska Natives, 1969-1993. *Int J Epidemiol* **27**, 388-96 (1998).
13. Whittemore, A.S. et al. Diet, physical activity, and colorectal cancer among Chinese in North America and China. *J Natl Cancer Inst* **82**, 915-26 (1990).
14. Landis, S.H., Murray, T., Bolden, S. & Wingo, P.A. Cancer statistics, 1998. *CA Cancer J Clin* **48**, 6-29 (1998).
15. Obrand, D.I. & Gordon, P.H. Continued change in the distribution of colorectal carcinoma. *Br J Surg* **85**, 246-8 (1998).
16. Thorn, M. et al. Trends in colorectal cancer incidence in Sweden 1959-93 by gender, localization, time period, and birth cohort. *Cancer Causes Control* **9**, 145-52 (1998).

17. Cusack, J.C. et al. Survival factors in 186 patients younger than 40 years old with colorectal adenocarcinoma. *J Am Coll Surg* **183**, 105-12 (1996).
18. de Bruine, A.P. et al. Endocrine cells in colorectal adenocarcinomas: incidence, hormone profile and prognostic relevance. *Int J Cancer* **54**, 765-71 (1993).
19. Messerini, L., Palomba, A. & Zampi, G. Primary signet-ring cell carcinoma of the colon and rectum. *Dis Colon Rectum* **38**, 1189-92 (1995).
20. Compton, C.C. et al. Prognostic factors in colorectal cancer. College of American Pathologists Consensus Statement 1999. *Arch Pathol Lab Med* **124**, 979-94 (2000).
21. Dukes, C.E. The classification of cancer of the rectum. *The Journal of Pathology and Bacteriology* **35**, 323-332 (1932).
22. Deans, G.T. et al. Jass' classification revisited. *J Am Coll Surg* **179**, 11-7 (1994).
23. Jass, J.R. Lymphocytic infiltration and survival in rectal cancer. *J Clin Pathol* **39**, 585-9 (1986).
24. Wolmark, N., Fisher, B. & Wieand, H.S. The prognostic value of the modifications of the Dukes' C class of colorectal cancer. An analysis of the NSABP clinical trials. *Ann Surg* **203**, 115-22 (1986).
25. Greene, F.L., Stewart, A.K. & Norton, H.J. A new TNM staging strategy for node-positive (stage III) colon cancer: an analysis of 50,042 patients. *Ann Surg* **236**, 416-21; discussion 421 (2002).
26. Majumdar, S.R., Fletcher, R.H. & Evans, A.T. How does colorectal cancer present? Symptoms, duration, and clues to location. *Am J Gastroenterol* **94**, 3039-45 (1999).
27. Stein, W. et al. Characteristics of colon cancer at time of presentation. *Fam Pract Res J* **13**, 355-63 (1993).
28. Mandel, J.S., Church, T.R., Ederer, F. & Bond, J.H. Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. *J Natl Cancer Inst* **91**, 434-7 (1999).
29. Ederer, F., Church, T.R. & Mandel, J.S. Fecal occult blood screening in the Minnesota study: role of chance detection of lesions. *J Natl Cancer Inst* **89**, 1423-8 (1997).
30. Lazovich, D. et al. A case-control study to evaluate efficacy of screening for faecal occult blood. *J Med Screen* **2**, 84-9 (1995).
31. Saito, H. et al. Reduction in risk of mortality from colorectal cancer by fecal occult blood screening with immunochemical hemagglutination test. A case-control study. *Int J Cancer* **61**, 465-9 (1995).
32. Selby, J.V., Friedman, G.D., Quesenberry, C.P., Jr. & Weiss, N.S. Effect of fecal occult blood testing on mortality from colorectal cancer. A case-control study. *Ann Intern Med* **118**, 1-6 (1993).
33. Winawer, S.J., Flehinger, B.J., Schottenfeld, D. & Miller, D.G. Screening for colorectal cancer with fecal occult blood testing and sigmoidoscopy. *J Natl Cancer Inst* **85**, 1311-8 (1993).

34. Muller, A.D. & Sonnenberg, A. Protection by endoscopy against death from colorectal cancer. A case-control study among veterans. *Arch Intern Med* **155**, 1741-8 (1995).
35. Selby, J.V., Friedman, G.D., Quesenberry, C.P., Jr. & Weiss, N.S. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *N Engl J Med* **326**, 653-7 (1992).
36. Winawer, S. et al. Colorectal cancer screening and surveillance: clinical guidelines and rationale-Update based on new evidence. *Gastroenterology* **124**, 544-60 (2003).
37. Johnson, C.D. & Ahlquist, D.A. Computed tomography colonography (virtual colonoscopy): a new method for colorectal screening. *Gut* **44**, 301-5 (1999).
38. Morrin, M.M. & LaMont, J.T. Screening virtual colonoscopy--ready for prime time? *N Engl J Med* **349**, 2261-4 (2003).
39. Macdonald, J.S. Carcinoembryonic antigen screening: pros and cons. *Semin Oncol* **26**, 556-60 (1999).
40. Srivastava, S., Verma, M. & Henson, D.E. Biomarkers for early detection of colon cancer. *Clin Cancer Res* **7**, 1118-26 (2001).
41. Weitz, J. et al. Colorectal cancer. *Lancet* **365**, 153-65 (2005).
42. Willett, W.C., Stampfer, M.J., Colditz, G.A., Rosner, B.A. & Speizer, F.E. Relation of meat, fat, and fiber intake to the risk of colon cancer in a prospective study among women. *N Engl J Med* **323**, 1664-72 (1990).
43. Wilmink, A.B. Overview of the epidemiology of colorectal cancer. *Dis Colon Rectum* **40**, 483-93 (1997).
44. Slattery, M.L. et al. Energy balance and colon cancer--beyond physical activity. *Cancer Res* **57**, 75-80 (1997).
45. Potter, J.D. Colorectal cancer: molecules and populations. *J Natl Cancer Inst* **91**, 916-32 (1999).
46. Hope, M.E., Hold, G.L., Kain, R. & El-Omar, E.M. Sporadic colorectal cancer--role of the commensal microbiota. *FEMS Microbiol Lett* **244**, 1-7 (2005).
47. McGarr, S.E., Ridlon, J.M. & Hylemon, P.B. Diet, anaerobic bacterial metabolism, and colon cancer: a review of the literature. *J Clin Gastroenterol* **39**, 98-109 (2005).
48. Fuchs, C.S. et al. A prospective study of family history and the risk of colorectal cancer. *N Engl J Med* **331**, 1669-74 (1994).
49. Rosenberg, L., Louik, C. & Shapiro, S. Nonsteroidal antiinflammatory drug use and reduced risk of large bowel carcinoma. *Cancer* **82**, 2326-33 (1998).
50. Smalley, W., Ray, W.A., Daugherty, J. & Griffin, M.R. Use of nonsteroidal anti-inflammatory drugs and incidence of colorectal cancer: a population-based study. *Arch Intern Med* **159**, 161-6 (1999).
51. Chen, J. et al. A prospective study of N-acetyltransferase genotype, red meat intake, and risk of colorectal cancer. *Cancer Res* **58**, 3307-11 (1998).

52. Wright, N.A. Epithelial stem cell repertoire in the gut: clues to the origin of cell lineages, proliferative units and cancer. *Int J Exp Pathol* **81**, 117-43 (2000).
53. Shih, I.M. et al. Top-down morphogenesis of colorectal tumors. *Proc Natl Acad Sci U S A* **98**, 2640-5 (2001).
54. Lamprecht, S.A. & Lipkin, M. Migrating colonic crypt epithelial cells: primary targets for transformation. *Carcinogenesis* **23**, 1777-80 (2002).
55. Bach, S.P., Renehan, A.G. & Potten, C.S. Stem cells: the intestinal stem cell as a paradigm. *Carcinogenesis* **21**, 469-76 (2000).
56. Fearon, E.R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-67 (1990).
57. Fodde, R. The APC gene in colorectal cancer. *Eur J Cancer* **38**, 867-71 (2002).
58. Fodde, R. et al. Mutations in the APC tumour suppressor gene cause chromosomal instability. *Nat Cell Biol* **3**, 433-8 (2001).
59. Bomme, L. et al. Clonal karyotypic abnormalities in colorectal adenomas: clues to the early genetic events in the adenoma-carcinoma sequence. *Genes Chromosomes Cancer* **10**, 190-6 (1994).
60. Ried, T. et al. Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Genes Chromosomes Cancer* **15**, 234-45 (1996).
61. Lane, D.P. Cancer. p53, guardian of the genome. *Nature* **358**, 15-6 (1992).
62. Mercer, W.E. Cell cycle regulation and the p53 tumor suppressor protein. *Crit Rev Eukaryot Gene Expr* **2**, 251-63 (1992).
63. Tejpar, S. & Van Cutsem, E. Molecular and genetic defects in colorectal tumorigenesis. *Best Pract Res Clin Gastroenterol* **16**, 171-85 (2002).
64. Gordon, H. Familial adenomatous polyposis. Definition, description and genetics. *Minerva Chir* **44**, 1845-54 (1989).
65. Lengauer, C., Kinzler, K.W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386**, 623-7 (1997).
66. Liu, B. et al. Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. *Nat Genet* **9**, 48-55 (1995).
67. Parsons, R. et al. Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. *Cancer Res* **55**, 5548-50 (1995).
68. Aaltonen, L.A. et al. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N Engl J Med* **338**, 1481-7 (1998).
69. Salovaara, R. et al. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. *J Clin Oncol* **18**, 2193-200 (2000).
70. Muller, H.H., Heinimann, K. & Dobbie, Z. Genetics of hereditary colon cancer - a basis for prevention? *Eur J Cancer* **36**, 1215-23 (2000).
71. Vasen, H.F., Watson, P., Mecklin, J.P. & Lynch, H.T. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome)

proposed by the International Collaborative group on HNPCC. *Gastroenterology* **116**, 1453-6 (1999).
72. Wheeler, J.M. Epigenetics, mismatch repair genes and colorectal cancer. *Ann R Coll Surg Engl* **87**, 15-20 (2005).
73. Burt, R.W. Familial risk and colorectal cancer. *Gastroenterol Clin North Am* **25**, 793-803 (1996).
74. Rozen, P. et al. Family history of colorectal cancer as a marker of potential malignancy within a screening program. *Cancer* **60**, 248-54 (1987).
75. Willenbucher, R.F. Inflammatory bowel disease. *Semin Gastrointest Dis* **7**, 94-104 (1996).
76. Bernstein, C.N., Blanchard, J.F., Kliewer, E. & Wajda, A. Cancer risk in patients with inflammatory bowel disease: a population-based study. *Cancer* **91**, 854-62 (2001).
77. Collins, R.H., Jr., Feldman, M. & Fordtran, J.S. Colon cancer, dysplasia, and surveillance in patients with ulcerative colitis. A critical review. *N Engl J Med* **316**, 1654-8 (1987).
78. Itzkowitz, S. Colon carcinogenesis in inflammatory bowel disease: applying molecular genetics to clinical practice. *J Clin Gastroenterol* **36**, S70-4; discussion S94-6 (2003).
79. Boveri, T. Zur Frage der Entstehung maligner Tumoren. *Jena: Gustav Fischer* (1914).
80. Shay, J.W. & Wright, W.E. Telomerase: a target for cancer therapeutics. *Cancer Cell* **2**, 257-65 (2002).
81. Rowley, J.D. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290-3 (1973).
82. Knudson, A.G., Jr. Hereditary cancer. *Jama* **241**, 279 (1979).
83. Giaretti, W. et al. Chromosomal instability and APC gene mutations in human sporadic colorectal adenomas. *J Pathol* **204**, 193-9 (2004).
84. D'Amours, D. & Jackson, S.P. The Mre11 complex: at the crossroads of dna repair and checkpoint signalling. *Nat Rev Mol Cell Biol* **3**, 317-27 (2002).
85. Jallepalli, P.V. & Lengauer, C. Chromosome segregation and cancer: cutting through the mystery. *Nat Rev Cancer* **1**, 109-17 (2001).
86. Loeb, K.R. & Loeb, L.A. Significance of multiple mutations in cancer. *Carcinogenesis* **21**, 379-85 (2000).
87. Vessey, C.J., Norbury, C.J. & Hickson, I.D. Genetic disorders associated with cancer predisposition and genomic instability. *Prog Nucleic Acid Res Mol Biol* **63**, 189-221 (1999).
88. Spruck, C.H., Won, K.A. & Reed, S.I. Deregulated cyclin E induces chromosome instability. *Nature* **401**, 297-300 (1999).
89. Donnellan, R. & Chetty, R. Cyclin E in human cancers. *Faseb J* **13**, 773-80 (1999).

90. Vogelstein, B. & Kinzler, K.W. Cancer genes and the pathways they control. *Nat Med* **10**, 789-99 (2004).
91. Roberts, R.B. et al. Importance of epidermal growth factor receptor signaling in establishment of adenomas and maintenance of carcinomas during intestinal tumorigenesis. *Proc Natl Acad Sci U S A* **99**, 1521-6 (2002).
92. Celen, O., Kahraman, I., Yildirim, E. & Berberoglu, U. Correlation of vascular endothelial growth factor (VEGF) and CEA with clinicopathological variables in colorectal cancer patients. *Neoplasma* **51**, 293-9 (2004).
93. Howard, A. & Pelc, S.R. Nuclear incorporation of p32 as demonstrated by autoradiographs. *Exp Cell Res* **2**, 178 (1951).
94. Sherr, C.J. Mammalian G1 cyclins. *Cell* **73**, 1059-65 (1993).
95. Zindy, F. et al. Cyclin A is required in S phase in normal epithelial cells. *Biochem Biophys Res Commun* **182**, 1144-54 (1992).
96. Liao, C. et al. Elevated levels and distinct patterns of expression of A-type cyclins and their associated cyclin-dependent kinases in male germ cell tumors. *Int J Cancer* **108**, 654-64 (2004).
97. Handa, K., Yamakawa, M., Takeda, H., Kimura, S. & Takahashi, T. Expression of cell cycle markers in colorectal carcinoma: superiority of cyclin A as an indicator of poor prognosis. *Int J Cancer* **84**, 225-33 (1999).
98. Vousden, K.H. Activation of the p53 tumor suppressor protein. *Biochim Biophys Acta* **1602**, 47-59 (2002).
99. Allshire, R.C. Centromeres, checkpoints and chromatid cohesion. *Curr Opin Genet Dev* **7**, 264-73 (1997).
100. Calaluce, R., Miedema, B.W. & Yesus, Y.W. Micrometastasis in colorectal carcinoma: a review. *J Surg Oncol* **67**, 194-202 (1998).
101. Iivanainen, A., Morita, T. & Tryggvason, K. Molecular cloning and tissue-specific expression of a novel murine laminin gamma3 chain. *J Biol Chem* **274**, 14107-11 (1999).
102. Tryggvason, K. The laminin family. *Curr Opin Cell Biol* **5**, 877-82 (1993).
103. Carter, W.G., Ryan, M.C. & Gahr, P.J. Epiligrin, a new cell adhesion ligand for integrin alpha 3 beta 1 in epithelial basement membranes. *Cell* **65**, 599-610 (1991).
104. Pyke, C. et al. The gamma 2 chain of kalinin/laminin 5 is preferentially expressed in invading malignant cells in human cancers. *Am J Pathol* **145**, 782-91 (1994).
105. Pyke, C. et al. Laminin-5 is a marker of invading cancer cells in some human carcinomas and is coexpressed with the receptor for urokinase plasminogen activator in budding cancer cells in colon adenocarcinomas. *Cancer Res* **55**, 4132-9 (1995).
106. Sordat, I. et al. Differential expression of laminin-5 subunits and integrin receptors in human colorectal neoplasia. *J Pathol* **185**, 44-52 (1998).

107. Duesberg, P., Rausch, C., Rasnick, D. & Hehlmann, R. Genetic instability of cancer cells is proportional to their degree of aneuploidy. *Proc Natl Acad Sci U S A* **95**, 13692-7 (1998).
108. Lengauer, C., Kinzler, K.W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643-9 (1998).
109. Ried, T., Heselmeyer-Haddad, K., Blegen, H., Schrock, E. & Auer, G. Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: a phenotype/genotype correlation. *Genes Chromosomes Cancer* **25**, 195-204 (1999).
110. Caspersson, T.O. Quantitative tumor cytochemistry--G.H.A. Clowes Memorial Lecture. *Cancer Res* **39**, 2341-5 (1979).
111. Caspersson, T., Zech, L., Johansson, C. & Modest, E.J. Identification of human chromosomes by DNA-binding fluorescent agents. *Chromosoma* **30**, 215-27 (1970).
112. Kallioniemi, A. et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**, 818-21 (1992).
113. Schrock, E. et al. Multicolor spectral karyotyping of human chromosomes. *Science* **273**, 494-7 (1996).
114. Speicher, M.R., Gwyn Ballard, S. & Ward, D.C. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat Genet* **12**, 368-75 (1996).
115. Knuutila, S. et al. DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. *Am J Pathol* **152**, 1107-23 (1998).
116. Forozan, F., Karhu, R., Kononen, J., Kallioniemi, A. & Kallioniemi, O.P. Genome screening by comparative genomic hybridization. *Trends Genet* **13**, 405-9 (1997).
117. Hittelman, W.N. Genetic instability in epithelial tissues at risk for cancer. *Ann N Y Acad Sci* **952**, 1-12 (2001).
118. Hopman, A.H. et al. In situ hybridization as a tool to study numerical chromosome aberrations in solid bladder tumors. *Histochemistry* **89**, 307-16 (1988).
119. Heselmeyer, K. et al. Gain of chromosome 3q defines the transition from severe dysplasia to invasive carcinoma of the uterine cervix. *Proc Natl Acad Sci U S A* **93**, 479-84 (1996).
120. Solinas-Toldo, S. et al. Mapping of chromosomal imbalances in pancreatic carcinoma by comparative genomic hybridization. *Cancer Res* **56**, 3803-7 (1996).
121. Nowak, M.A. et al. The role of chromosomal instability in tumor initiation. *Proc Natl Acad Sci U S A* **99**, 16226-31 (2002).
122. Desper, R. et al. Distance-based reconstruction of tree models for oncogenesis. *J Comput Biol* **7**, 789-803 (2000).

123. Barrett, J.C., Oshimura, M., Tanaka, N. & Tsutsui, T. Role of aneuploidy in early and late stages of neoplastic progression of Syrian hamster embryo cells in culture. *Basic Life Sci* **36**, 523-38 (1985).
124. Oshimura, M. & Barrett, J.C. Chemically induced aneuploidy in mammalian cells: mechanisms and biological significance in cancer. *Environ Mutagen* **8**, 129-59 (1986).
125. Matzke, M.A., Mette, M.F., Kanno, T. & Matzke, A.J. Does the intrinsic instability of aneuploid genomes have a causal role in cancer? *Trends Genet* **19**, 253-6 (2003).
126. Schena, M. et al. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci U S A* **93**, 10614-9 (1996).
127. Virtaneva, K. et al. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. *Proc Natl Acad Sci U S A* **98**, 1124-9 (2001).
128. Phillips, J.L. et al. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. *Cancer Res* **61**, 8143-9 (2001).
129. Platzer, P. et al. Silence of chromosomal amplifications in colon cancer. *Cancer Res* **62**, 1134-8 (2002).
130. Pollack, J.R. et al. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci U S A* **99**, 12963-8 (2002).
131. Hyman, E. et al. Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res* **62**, 6240-5 (2002).
132. Misek, D.E., Imafuku, Y. & Hanash, S.M. Application of proteomic technologies to tumor analysis. *Pharmacogenomics* **5**, 1129-37 (2004).
133. Oppermann, M. et al. Identification of foetal brain proteins by two-dimensional gel electrophoresis and mass spectrometry comparison of samples from individuals with or without chromosome 21 trisomy. *Eur J Biochem* **267**, 4713-9 (2000).
134. Conrads, T.P., Hood, B.L., Issaq, H.J. & Veenstra, T.D. Proteomic patterns as a diagnostic tool for early-stage cancer: a review of its progress to a clinically relevant tool. *Mol Diagn* **8**, 77-85 (2004).
135. Riddell, R.H. et al. Dysplasia in inflammatory bowel disease: standardized classification with provisional clinical applications. *Hum Pathol* **14**, 931-68 (1983).
136. Hunt, J.D. Evaluation of phenotypic alteration by microcell-mediated chromosome transfer. *Anal Biochem* **238**, 107-16 (1996).
137. Auer, G.U., Caspersson, T.O. & Wallgren, A.S. DNA content and survival in mammary carcinoma. *Anal Quant Cytol* **2**, 161-5 (1980).
138. du Manoir, S. et al. Detection of complete and partial chromosome gains and losses by comparative genomic in situ hybridization. *Hum Genet* **90**, 590-610 (1993).

139. Ried, T. et al. Tumor cytogenetics revisited: comparative genomic hybridization and spectral karyotyping. *J Mol Med* **75**, 801-14 (1997).
140. Cremer, T. et al. Detection of chromosome aberrations in the human interphase nucleus by visualization of specific target DNAs with radioactive and non-radioactive in situ hybridization techniques: diagnosis of trisomy 18 with probe L1.84. *Hum Genet* **74**, 346-52 (1986).
141. Ried, T. et al. Specific metaphase and interphase detection of the breakpoint region in 8q24 of Burkitt lymphoma cells by triple-color fluorescence in situ hybridization. *Genes Chromosomes Cancer* **4**, 69-74 (1992).
142. Veldman, T., Vignon, C., Schrock, E., Rowley, J.D. & Ried, T. Hidden chromosome abnormalities in haematological malignancies detected by multicolour spectral karyotyping. *Nat Genet* **15**, 406-10 (1997).
143. Padilla-Nash, H.M. et al. Jumping translocations are common in solid tumor cell lines and result in recurrent fusions of whole chromosome arms. *Genes Chromosomes Cancer* **30**, 349-63 (2001).
144. Wadlow, R. & Ramaswamy, S. DNA microarrays in clinical cancer research. *Curr Mol Med* **5**, 111-20 (2005).
145. Hegde, P. et al. A concise guide to cDNA microarray analysis. *Biotechniques* **29**, 548-50, 552-4, 556 passim (2000).
146. Korn, E.L., Habermann, J.K., Upender, M.B., Ried, T. & McShane, L.M. Objective method of comparing DNA microarray image analysis systems. *Biotechniques* **36**, 960-7 (2004).
147. Haslett, J.N. et al. Gene expression comparison of biopsies from Duchenne muscular dystrophy (DMD) and normal skeletal muscle. *Proc Natl Acad Sci U S A* **99**, 15000-5 (2002).
148. Gorg, A., Weiss, W. & Dunn, M.J. Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4**, 3665-85 (2004).
149. Franzen, B. et al. Analysis of polypeptide expression in benign and malignant human breast lesions: down-regulation of cytokeratins. *Br J Cancer* **74**, 1632-8 (1996).
150. Franzen, B., Linder, S., Okuzawa, K., Kato, H. & Auer, G. Nonenzymatic extraction of cells from clinical tumor material for analysis of gene expression by two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis* **14**, 1045-53 (1993).
151. Veenstra, T.D., Prieto, D.A. & Conrads, T.P. Proteomic patterns for early cancer detection. *Drug Discov Today* **9**, 889-97 (2004).
152. Issaq, H.J., Conrads, T.P., Prieto, D.A., Tirumalai, R. & Veenstra, T.D. SELDI-TOF MS for diagnostic proteomics. *Anal Chem* **75**, 148A-155A (2003).
153. Cordingley, H.C. et al. Multifactorial screening design and analysis of SELDI-TOF ProteinChip array optimization experiments. *Biotechniques* **34**, 364-5, 368-73 (2003).

154. Semmes, O.J. et al. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin Chem* **51**, 102-12 (2005).
155. Ekbom, A., Helmick, C., Zack, M. & Adami, H.O. Ulcerative colitis and colorectal cancer. A population-based study. *N Engl J Med* **323**, 1228-33 (1990).
156. Dobbins, W.O., 3rd. Current status of the precancer lesion in ulcerative colitis. *Gastroenterology* **73**, 1431-3 (1977).
157. Lofberg, R., Brostrom, O., Karlen, P., Ost, A. & Tribukait, B. DNA aneuploidy in ulcerative colitis: reproducibility, topographic distribution, and relation to dysplasia. *Gastroenterology* **102**, 1149-54 (1992).
158. Rubin, C.E. et al. DNA aneuploidy in colonic biopsies predicts future development of dysplasia in ulcerative colitis. *Gastroenterology* **103**, 1611-20 (1992).
159. Lenander, C. et al. Laminin-5 gamma 2 chain expression correlates with unfavorable prognosis in colon carcinomas. *Anal Cell Pathol* **22**, 201-9 (2001).
160. Larjava, H., Salo, T., Haapasalmi, K., Kramer, R.H. & Heino, J. Expression of integrins and basement membrane components by wound keratinocytes. *J Clin Invest* **92**, 1425-35 (1993).
161. Haapasalmi, K. et al. Expression of epithelial adhesion proteins and integrins in chronic inflammation. *Am J Pathol* **147**, 193-206 (1995).
162. Thorup, A.K. et al. Can alterations in integrin and laminin-5 expression be used as markers of malignancy? *Apmis* **106**, 1170-80 (1998).
163. Auer, G. et al. Reversibility of bronchial cell atypia. *Cancer Res* **42**, 4241-7 (1982).
164. Ono, J. et al. Reversibility of 20-methylcholanthrene-induced bronchial cell atypia in dogs. *Cancer* **54**, 1030-7 (1984).
165. Wang, Y. et al. Large scale identification of human hepatocellular carcinoma-associated antigens by autoantibodies. *J Immunol* **169**, 1102-9 (2002).
166. Aust, D.E. et al. Chromosomal alterations in ulcerative colitis-related and sporadic colorectal cancers by comparative genomic hybridization. *Hum Pathol* **31**, 109-14 (2000).
167. Holzmann, K. et al. Comparison of flow cytometry and histology with mutational screening for p53 and Ki-ras mutations in surveillance of patients with long-standing ulcerative colitis. *Scand J Gastroenterol* **36**, 1320-6 (2001).
168. Kern, S.E. et al. Molecular genetic profiles of colitis-associated neoplasms. *Gastroenterology* **107**, 420-8 (1994).
169. Willenbucher, R.F., Zelman, S.J., Ferrell, L.D., Moore, D.H., 2nd & Waldman, F.M. Chromosomal alterations in ulcerative colitis-related neoplastic progression. *Gastroenterology* **113**, 791-801 (1997).

170. Blegen, H. et al. Genetic instability promotes the acquisition of chromosomal imbalances in T1b and T1c breast adenocarcinomas. *Anal Cell Pathol* **22**, 123-31 (2001).
171. Tamborini, E. et al. Analysis of the molecular species generated by MDM2 gene amplification in liposarcomas. *Int J Cancer* **92**, 790-6 (2001).
172. Osman, I. et al. Inactivation of the p53 pathway in prostate cancer: impact on tumor progression. *Clin Cancer Res* **5**, 2082-8 (1999).
173. Benhattar, J., Losi, L., Chaubert, P., Givel, J.C. & Costa, J. Prognostic significance of K-ras mutations in colorectal carcinoma. *Gastroenterology* **104**, 1044-8 (1993).
174. Ghadimi, B.M. et al. Centrosome amplification and instability occurs exclusively in aneuploid, but not in diploid colorectal cancer cell lines, and correlates with numerical chromosomal aberrations. *Genes Chromosomes Cancer* **27**, 183-90 (2000).
175. Schlegel, J. et al. Comparative genomic in situ hybridization of colon carcinomas with replication error. *Cancer Res* **55**, 6002-5 (1995).
176. Tirkkonen, M. et al. Molecular cytogenetics of primary breast cancer by CGH. *Genes Chromosomes Cancer* **21**, 177-84 (1998).
177. Roschke, A.V., Stover, K., Tonon, G., Schaffer, A.A. & Kirsch, I.R. Stable karyotypes in epithelial cancer cell lines despite high rates of ongoing structural and numerical chromosomal instability. *Neoplasia* **4**, 19-31 (2002).
178. FitzPatrick, D.R. et al. Transcriptome analysis of human autosomal trisomy. *Hum Mol Genet* **11**, 3249-56 (2002).
179. Mao, R., Zielke, C.L., Zielke, H.R. & Pevsner, J. Global up-regulation of chromosome 21 gene expression in the developing Down syndrome brain. *Genomics* **81**, 457-67 (2003).
180. Saran, N.G., Pletcher, M.T., Natale, J.E., Cheng, Y. & Reeves, R.H. Global disruption of the cerebellar transcriptome in a Down syndrome mouse model. *Hum Mol Genet* **12**, 2013-9 (2003).
181. Brewer, C., Holloway, S., Zawalnyski, P., Schinzel, A. & FitzPatrick, D. A chromosomal duplication map of malformations: regions of suspected haplo- and triplolethality--and tolerance of segmental aneuploidy--in humans. *Am J Hum Genet* **64**, 1702-8 (1999).
182. Caron, H. et al. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291**, 1289-92 (2001).
183. Druker, B.J. Imatinib alone and in combination for chronic myeloid leukemia. *Semin Hematol* **40**, 50-8 (2003).
184. Bertucci, F. et al. Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. *Oncogene* **23**, 1377-91 (2004).
185. Notterman, D.A., Alon, U., Sierk, A.J. & Levine, A.J. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res* **61**, 3124-30 (2001).

186. Kim, H. et al. Different gene expression profiles between microsatellite instability-high and microsatellite stable colorectal carcinomas. *Oncogene* **23**, 6218-25 (2004).
187. Shih, W., Chetty, R. & Tsao, M.S. Expression profiling by microarrays in colorectal cancer (Review). *Oncol Rep* **13**, 517-24 (2005).
188. Ayude, D. et al. Clinical interest of the combined use of serum CD26 and alpha-L-fucosidase in the early diagnosis of colorectal cancer. *Dis Markers* **19**, 267-72 (2003).
189. Cordero, O.J., Ayude, D., Nogueira, M., Rodriguez-Berrocal, F.J. & de la Cadena, M.P. Preoperative serum CD26 levels: diagnostic efficiency and predictive value for colorectal cancer. *Br J Cancer* **83**, 1139-46 (2000).
190. Broll, R. et al. Vascular endothelial growth factor (VEGF)--a valuable serum tumour marker in patients with colorectal cancer? *Eur J Surg Oncol* **27**, 37-42 (2001).
191. Carpelan-Holmstrom, M. et al. Estimating the probability of cancer with several tumor markers in patients with colorectal disease. *Oncology* **66**, 296-302 (2004).
192. Blackwell, K. et al. Circulating D-dimer levels are better predictors of overall survival and disease progression than carcinoembryonic antigen levels in patients with metastatic colorectal carcinoma. *Cancer* **101**, 77-82 (2004).
193. Belluco, C. et al. Interleukin-6 blood level is associated with circulating carcinoembryonic antigen and prognosis in patients with colorectal cancer. *Ann Surg Oncol* **7**, 133-8 (2000).
194. Erlinger, T.P., Platz, E.A., Rifai, N. & Helzlsouer, K.J. C-reactive protein and the risk of incident colorectal cancer. *Jama* **291**, 585-90 (2004).
195. Yukawa, N. et al. Plasma concentration of tissue inhibitor of matrix metalloproteinase 1 in patients with colorectal carcinoma. *Br J Surg* **88**, 1596-601 (2001).
196. Lledo, S.M. et al. Real time quantification in plasma of human telomerase reverse transcriptase (hTERT) mRNA in patients with colorectal cancer. *Colorectal Dis* **6**, 236-42 (2004).
197. Yu, J.K., Chen, Y.D. & Zheng, S. An integrated approach to the detection of colorectal cancer utilizing proteomics and bioinformatics. *World J Gastroenterol* **10**, 3127-31 (2004).