

Aus der Klinik für Chirurgie
der Universität zu Lübeck
Direktor: Prof. Dr. med. H.-P. Bruch

**The Impact of Aneuploidy
on Malignant Transformation in Ulcerative Colitis**

Inauguraldissertation

zur Erlangung der Doktorwürde
der Universität zu Lübeck
- Aus der Medizinischen Fakultät -

vorgelegt von
Marco Gerling
aus Rahden

1. Berichtstatter: Priv.-Doz. Dr. med. Jens K. Habermann PhD
 2. Berichtstatter: Prof. Dr. med. Christoph Thorns
 3. Berichtstatter: Prof. Dr. rer. nat. Ferdinand von Eggeling
- Tag der mündlichen Prüfung: 30.11.2011
- Zum Druck genehmigt. Lübeck, den 30.11.2011

We fall into deep error, not just a harmful oversimplification, when we speak of genes „for“ particular items of anatomy or behavior.

(Stephen Jay Gould, 1941-2002)

Für meine Eltern

TABLE OF CONTENTS

1	INTRODUCTION.....	1
1.1	Ulcerative Colitis	1
1.1.1	Epidemiology and Clinical Presentation.....	1
1.1.2	Diagnosis	2
1.1.3	Treatment	3
1.1.4	Etiology	4
1.2	The Ulcerative Colitis-associated Colorectal Carcinoma	4
1.3	The Role of Aneuploidy in Malignant Lesions	6
1.3.1	Aneuploidy – Cause or Effect?	6
1.3.2	The Role of Aneuploidy in Colorectal Cancers	7
1.3.3	The Role of Aneuploidy in UC and UCC.....	7
2	SYNOPSIS AND RESEARCH GOALS.....	9
3	MATERIAL AND METHODS.....	10
3.1	General Study Design and Setting	10
3.2	Part 1: Patients and Specimens for Ploidy Assessment.....	11
3.3	Part 2: Patients and Specimens for Gene Expression Analyses	12
3.4	Methods	13
3.4.1	DNA Image Cytometry	13
3.4.2	Oligonucleotide Microarrays	15
3.4.3	Real Time-quantitative PCR (RT-qPCR).....	19
3.4.4	Statistical Analyses	21
3.5	Synopsis of Workflow for the Gene Expression Analyses	24
4	RESULTS	25
4.1	Results from Ploidy Assessment in Neoplastic Lesions	25
4.1.1	Ploidy Assessment by DNA Image Cytometry	25
4.1.2	Inter-Observer Reliability of Ploidy Assessment	25
4.1.3	Clinical Differences between Patient Groups.....	25
4.1.4	Logistic Regression Analysis	30
4.1.5	Survival Analyses (Kaplan-Meier-plots)	31
4.2	Results from the Gene Expression Analyses in Colitis Mucosa	32
4.2.1	Ploidy Assessment in Premalignant Colitis Mucosa.....	32
4.2.2	Inter-Observer Reliability of Ploidy Assessment in Mucosa Samples	33
4.2.3	Microarray Gene Expression Analyses	33

5	DISCUSSION	44
5.1	Overview of Experimental Results	44
5.1.1	Frequency of Aneuploidy in UCCs and SCCs.....	44
5.1.2	Inter-observer Variability of Ploidy Assessment.....	45
5.2	Comparative Analyses of UCCs and SCCs.....	45
5.2.1	Clinical Correlations (Univariate Analyses).....	45
5.2.2	Survival Analyses.....	47
5.2.3	Logistic Regression (Multivariate Analysis).....	47
5.3	Gene Expression Analyses of Malignant Transformation in UC	48
5.3.1	Transcriptomic Changes during Aneuploidization.....	48
5.4	Genes Validated with RT-qPCR.....	51
5.4.1	RT-qPCR Validation.....	51
5.4.2	Genes Differentially Regulated between Diploid and Aneuploid Mucosa.....	52
5.4.3	Genes Differentially Regulated between Aneuploid Mucosa and UCC.....	57
5.4.4	Genes Constantly Differentially Regulated	58
5.5	Transition of Microarray Results into Clinical Application	60
6	SUMMARY	62
7	APPENDIX	74
7.1	Ethical Permits	74
7.2	Reagents and Solutions.....	74
7.3	Protocols.....	76
7.3.1	Haematoxylin – Eosin – Staining	76
7.3.2	Feulgen – Staining.....	76
7.3.3	RNA Extraction	76
7.3.4	RNA Amplification.....	77
7.3.5	Microarray Hybridization with aRNA	80
7.3.6	RT-qPCR.....	81
7.4	Supplementary Tables and Figure	83
7.4.1	Group comparisons.....	84
8	ACKNOWLEDGEMENTS	93
9	CURRICULUM VITAE	95
10	PUBLICATIONS	97

ABBREVIATIONS

5-ASA:	5-aminosalicylic acid	IHC:	immunohistochemistry
ANOVA:	analysis of variance	IPA:	Ingenuity Pathways Analysis®
aRNA:	amplified ribonucleic acid	lowess:	locally weighted scatterplot smoothing
CD:	Crohn's disease	m:	mean
cDNA:	complementary DNA	MEF:	mouse embryonic fibroblast
CEA:	carcinoembryonic antigen	mRNA:	messenger ribonucleic acid
CENP:	centromere-associated protein	MSI:	microsatellite instability
CGH:	comparative genomic hybridization	NCI:	National Cancer Institute, USA
CI:	confidence interval	oligo:	oligonucleotide
CIN:	chromosomal instability	OR:	odds ratio
CRC:	colorectal carcinoma	PMT:	photon-multiplier-tube
CT:	cycle threshold (of RT-qPCR)	PSC:	primary sclerosing cholangitis
CV:	coefficient of variation	RNA:	ribonucleic acid
DALM:	dysplasia associated lesion or mass	RNAi:	RNA-interference
DEG:	differentially expressed gene	RR:	relative risk
DFS:	disease free survival	RT-qPCR:	real time quantitative-polymerase chain reaction
DNA:	desoxyribonucleic acid	SAC:	spindle assembly checkpoint
DSB:	double strand break (of DNA)	SCC:	sporadic colorectal carcinoma
fc:	fold change	SD:	standard deviation
fdr:	false discovery rate	SNP:	single nucleotide polymorphism
GAL:	gene allocation table	TLR:	toll like receptor
GIT:	gastrointestinal tract	UC:	ulcerative colitis
GO:	Gene-Ontology	UCC:	ulcerative colitis-associated colorectal carcinoma
HC:	hierarchical clustering	UICC:	union internationale contre le cancer
HE:	Haematoxylin & Eosin staining		
HNPCC:	hereditary non-polyposis colorectal cancer		
IBD:	inflammatory bowel disease		
IF:	immunofluorescence		

TABLES AND FIGURES

Tables

Table 1: Examples of extraintestinal manifestations in UC.	2
Table 2: Clinical findings in ulcerative colitis and Crohn's disease	2
Table 3: Clinical data obtained for UCC and SCC patients	11
Table 4: Patients for ploidy analyses, part one	12
Table 5: Patients for gene expression analyses.....	13
Table 6: Forward and reverse primers for RT-qPCR validation.....	21
Table 7: Significant differences in group comparisons	29
Table 8: Logistic regression analysis	30
Table 9: Validated genes from the comparison of diploid and aneuploid colitis mucosa	36
Table 10: DEGs between aneuploid mucosa and UCC.. ..	38
Table 11: Number of DEGs in group comparisons.....	41
Table 12: Genes constantly down or up regulated.....	42

Figures

Figure 1: Illustrated treatment algorithm for UC.....	3
Figure 2: General study design of both separate parts of this thesis	10
Figure 3: Selection of nuclei with the ACAS software	14
Figure 4: Ploidy types according to Auer's classification	15
Figure 5: Synopsis of microarray workflow	15
Figure 6: Coupling of Cy-5 dye to amino-allyl residues of dUTP.....	17
Figure 7: Microarray gridding process with GenePix Software	18
Figure 8: Example of spatial setting of primers for RT-PCR validation	20
Figure 9: Study design for gene expression analyses	24
Figure 10: Percentage of synchronous carcinomas in UCC, aSCC, and dSCC patients	26
Figure 11: Percentage of distant metastases in UCC, aSCC and dSCC groups	27
Figure 12: Distribution of post-operative tumor grading among UCCs, aSCCs and dSCCs	28
Figure 13: Kaplan-Meier plots	31
Figure 14: Distribution of ploidy types in colitis mucosa.....	32
Figure 15: DEGs among the sequence from uninfamed colon to UCC.....	34
Figure 16: Highest ranked IPA networks of DEG between normal controls vs UC mucosa (A) and UC mucosa vs UCC (B).....	39
Figure 17: Top IPA networks of constantly changed DEGs and DEGs between diploid and aneuploid colitis.....	40

Figure 18: Linear array expression plots and PCR expression plots	43
Figure 19: The cell cycle and associated validated genes	52
Figure 20: Laminin expression on tumor cells	56

Supplementary Tables

Supplementary table 1: Patients for part two, overview of mucosa biopsies	84
Supplementary table 2: Top 30 down regulated DEGs between diploid mucosa and aneuploid mucosa	85
Supplementary table 3: Top 30 up regulated DEGs between diploid mucosa and aneuploid mucosa	86
Supplementary table 4: Top 30 up regulated DEGs between diploid mucosa and UCCs ..	87
Supplementary table 5: Top 30 down regulated DEGs between diploid mucosa and UCC	88
Supplementary table 6: Top 30 up regulated DEGs between UC mucosa and UCCs	89
Supplementary table 7: Top 30 down regulated DEGs between UC mucosa and UCCs ...	90
Supplementary table 8: Top 30 up regulated DEGs between normal controls and UC mucosa	91
Supplementary table 9: Top 30 down regulated DEGs between normal controls and UC mucosa	92

1 INTRODUCTION

1.1 *Ulcerative Colitis*

Ulcerative colitis (UC) is a chronic inflammatory disorder affecting mucosa and submucosa of the colon and rectum ¹. UC and Crohn's disease (CD) comprise the major entities of inflammatory bowel diseases (IBDs). UC is mainly confined to the colorectum, but can also impinge on the terminal ileum, then referred to as "backwash ileitis". Contrarily, CD can affect the whole gastrointestinal tract (GIT) and is not constricted to the innermost histological layers ¹. The first description of UC was published by Sir Samuel Wilks in 1859: "*The morbid appearance of the intestine of Miss Banks*" ². IBDs are nowadays recognized as relatively common disorders: In Germany alone, more than 65,000 citizens suffer from IBDs, underscoring their socioeconomic importance ³.

UC is associated with an increased risk for colorectal carcinoma (CRC) development ⁴. While this complication is not common, it severely aggravates patients' prognoses ⁵. Reliable, objective markers for early detection of malignant transformation in UC are still missing. Furthermore, biological differences between UC associated carcinomas (UCCs) and sporadic colorectal carcinomas (SCCs) have not yet been thoroughly addressed. Identifying markers that can predict carcinoma development are thus highly desirable. Moreover, a more detailed characterization of the molecular pathways associated with UC associated carcinogenesis is warranted to further characterize UCCs and develop novel targeted therapeutics.

1.1.1 **Epidemiology and Clinical Presentation**

Epidemiological studies have described an incidence for UC of about 10/100000 ⁶. Within Europe, a north-south divide can be observed with a higher incidence in northern European countries ⁷. Men are slightly more frequently affected than women ⁶. UC usually develops in the second or third decade of life ⁸. Initially, patients commonly present with diarrhea, bloody stools, abdominal pain or tenderness and unspecific symptoms such as anemia and general malaise ⁹. The course of the disorder is highly variable and to a large extent not predictable: In a Danish study comprising 1,186 patients, disease development varied from remission to relapse without significant predictors, except for severe clinical

presentation at onset being associated with an unfavorable course ¹⁰. Apart from symptoms directly related to the inflamed colon, extraintestinal manifestations occur in up to 40% of all patients ¹¹ (table 1).

Hepatobiliary	primary sclerosing cholangitis (PSC), autoimmune hepatitis, cholelithiasis
Ocular	uveitis, iritis, episcleritis
Rheumatologic	ankylosing spondylitis, sacroiliitis
Mucocutaneous	cheilitis, glossitis, erythema nodosum, pyoderma gangrenosum

Table 1: Examples of extraintestinal manifestations in UC
Modified from Timani and Mutasim "Skin manifestations of inflammatory bowel diseases", 2008, Clinics in Dermatology ¹²

The life expectancy of patients with UC is comparable to that of the average population ¹³. However, a meta-analysis revealed that the causes of death are differently distributed: CRCs were found more frequently as a cause for mortality in UC patients than in the average population, accounting for 42% of UC-related deaths ¹³. Typical complications such as toxic megacolon and PSC leading to chronic liver disease are rare among mortality causes, accounting to 8% of UC-related deaths ¹³.

1.1.2 Diagnosis

Diagnosis of UC is established through a multidisciplinary approach involving clinical, endoscopical, and histopathological findings. In about 10% of all cases, the differential diagnosis between CD and UC remains unclear, then referred to as "undetermined colitis" ¹⁴. Table 2 contrasts the diagnostic hallmarks of UC and CD. Established serum or fecal markers for IBDs do not exist, yet a diversity of genomic, transcriptomic and proteomic markers currently undergo extensive investigation ^{15, 16}.

	<i>Ulcerative colitis</i>	<i>Crohn's disease</i>
Endoscopic	Lesions in colon and rectum	Whole GIT can be affected
	Rectum involvement	Rectal sparing (frequently)
	Continuous disease	"Skipping lesions"
	Pseudopolyps	"Cobblestoning"
Radiographic	Occasionally strictures	Frequently strictures
		Small bowel abnormalities
Histologic	(Sub-)mucosal inflammation	Transmural inflammation
	Crypt abscesses	Granulomas

Table 2: Clinical findings in ulcerative colitis and Crohn's disease
Modified from Harrison's Principals of Internal Medicine, McGraw Hill, 16th Edition, 2006

1.1.3 Treatment

While no specific causal treatment for UC is available, several therapeutic options exist, which are based on the general principle of modulating the immune system. Patients are stratified for treatment according to disease localization and severity of inflammation ¹⁷. Figure 1 gives an overview of treatment options: In general, induction therapy in an acute flare consists of 5-aminosalicylic acid (5-ASA) derivatives or corticosteroids ¹⁷. In addition to established immunosuppressive drugs such as Ciclosporin and Tacrolimus, so-called “biological” drugs became available for severe cases: Infliximab and Adalimumab are TNF- α antagonists that can lead to remission. Abatacept, Basiliximab, and other monoclonal antibodies directed against different epitopes involved in immune response were more recently introduced ^{18, 19}. Fulminant disease can nevertheless necessitate surgery. Pancolectomy with ileal pouch anastomosis is by many authors considered the treatment of choice for severe treatment-refractory disease ²⁰.

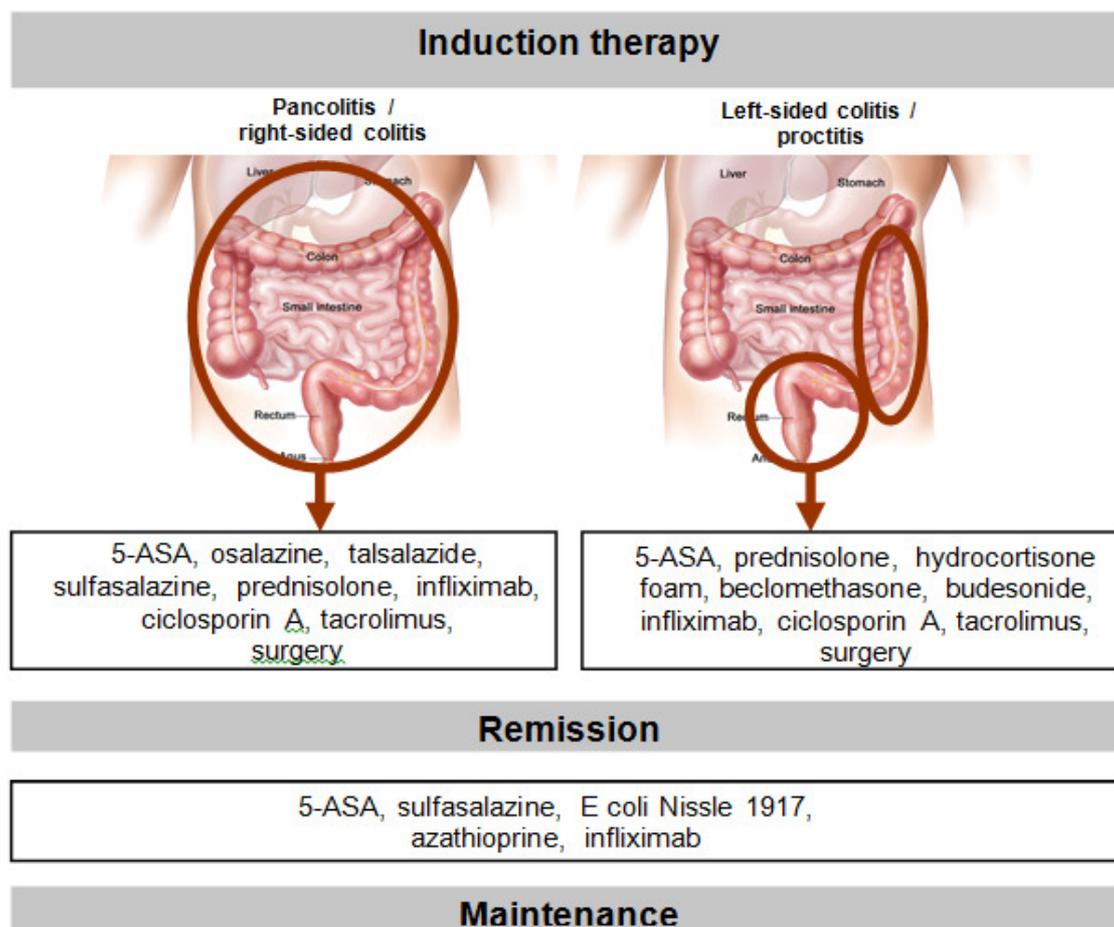


Figure 1: Illustrated treatment algorithm for UC

Modified from Baumgart and Sandborn “*Inflammatory Bowel diseases: clinical aspects and evolving therapies*” in *Lancet*, 2007 ¹⁷. Images from National Cancer Institute - Visuals online (public images); drawing by Terese Winslow

1.1.4 Etiology

Countless putative pathogenetic factors for IBDs have been investigated: The spectrum includes bacterial agents with special attention given to *Mycobacterium avium paratuberculosis*²¹, as well as helminths-colonisation of the gut²², dietary factors²³, and psychosomatic theories²⁴. No single microbial organism or extrinsic risk factor could be identified as the sole causative agent. Two reproducible findings are of interest: Cigarette smoking increases the risk for CD, while it seems to protect from developing UC²⁵. Moreover, appendectomy has a protective effect for UC⁶.

An improved understanding of the immune system facilitated the investigation of intrinsic immunopathogenic factors that play a major role for UC development. In particular, Toll-like receptor (TLR) and Nod-like-receptor (NLR) variants have been identified to be associated with IBDs²⁶. Both receptor types mediate pattern recognition for microbial molecules²⁷. Moreover, murine studies are giving evidence that a bacterial flora is indispensable for colitis development²⁸.

Using positional cloning and candidate gene approaches, *NOD2* (also known as *CARD15*) could be identified as susceptibility gene in CD²⁹. Whole genome association studies led to the discovery that *IL23R* (interleukin-23 receptor) is associated with CD³⁰. In 2008, a cohort of 1,167 UC patients was tested in a genome-wide study for single nucleotide polymorphisms (SNPs). The study identified regions in the *IL10*- and *ARPC2*-genes to be associated with UC³¹. Genetic associations in UC are generally weaker than those in CD. Accordingly, comparative studies in monozygotic twins demonstrated a stronger concordance for CD (35%) than for UC (16%)³².

In summary, the available evidence points to a multifactorial pathogenesis of both major IBD entities. The development of colitis seems to rely on the colonization of the gut with microbial flora in a genetically susceptible host, while environmental factors also contribute to colitis initiation and disease progression.

1.2 The Ulcerative Colitis-associated Colorectal Carcinoma

Patients suffering from UC face an increased lifetime risk for the development of CRCs: The cumulative risk is estimated to reach 20% after 35 years of disease duration³³. Studies on risk factors for cancer development in UC could identify presence of PSC, as well as duration and extent of colitis as risk factors for UCC

³⁴. Contrarily, smoking and the use of aspirin or corticosteroids seems to lower the risk ³⁵. UCCs show distinct clinical differences to sporadic carcinomas: Patients with UCCs are younger, more often males than females, and synchronous carcinomas are found more frequently ^{5, 36, 37}. Conflicting results have been reported for the overall disease free survival (DFS) of UCC patients: Delaunoy *et al.* described the prognosis of UCC and SCC to be similar with an overall DFS of 54% and 53% after five years, respectively ³⁸. In contrast, Aarnio and co-workers found UCC patients to have a significantly inferior survival than SCC patients, with 38.6% five-year DFS for UCC and 58.8% for SCC patients ³⁹. Cancers in UC do not seem to follow the adenoma-carcinoma sequence, which is a widely accepted concept for CRCs arising *de novo* ⁴⁰. Instead, they develop via flat dysplasia or so-called “dysplasia associated lesions and masses” (DALMs) that are particularly hard to detect ⁴¹.

Surveillance programs have been installed for UC patients aiming at early detection of dysplastic areas ⁴². However, a meta-analysis by the Cochrane Collaboration could not find evidence that surveillance colonoscopy prolongs survival in patients with UC: The authors concluded that “the slight apparent benefit [of surveillance programs in UC] is likely to be attributable to lead-time bias” ⁴³. I.e. they suspected that surveillance leads to early recognition of UCC and a relative increase in survival, yet the absolute survival and the overall mortality is not altered by the surveillance program. One reason for this could be that the assessment of dysplastic lesions in the inflamed colon is highly subjective: In a study by Odze *et al.* it was shown that after reviewing the pathologic dysplasia assessment of 38 biopsies of chronic UC by seven pathologists, the histopathological diagnosis “was changed in 51% of the observations” ⁴⁴.

The limited possibilities for early detection of UCC warrant more objective markers to detect premalignant lesions. One additional marker that has been studied by several groups in this context is DNA aneuploidy ^{4, 45-49}. Aneuploidy is defined as an abnormal DNA content in the nucleus of a cell, reflecting chromosomal instability (CIN). Independent investigations have shown that aneuploid colonic epithelial cells can be detected years prior to malignant transformation in UC ^{4, 46}. Moreover, aneuploidy precedes dysplastic lesions in UC patients ⁴⁶. As compared to dysplasia, aneuploidy is an objective marker and easy to assess routinely.

1.3 The Role of Aneuploidy in Malignant Lesions

In 1914, Theodor Boveri proposed his theory that aneuploidy causes cancer⁵⁰. Today, it is possible to scan the whole genome on a subchromosomal level for gains and losses of genetic information. However, the causative role of aneuploidy for malignancy development, as deduced by Boveri from the mere correlation, still remains questionable⁵¹. Aneuploidy is a dominating feature of a variety of cancers, occurring in about 90% of all solid and 75% of all hematological malignancies⁵². As for solid tumors, aneuploidy is probably best characterized in breast cancer and CRCs. In breast cancer, its presence is associated with an inferior prognosis and a higher frequency of distant metastases at time of diagnosis as compared to diploid carcinomas⁵³. In CRC, aneuploidy is likewise correlated with an unfavorable prognosis⁵⁴.

1.3.1 Aneuploidy – Cause or Effect?

Many theories exist concerning the pathogenesis of aneuploidy. Special attention has been given to the spindle assembly checkpoint (SAC, also known as mitotic checkpoint): Genes involved in the SAC act during mitosis to prevent chromosomal missegregation⁵⁵. Yet, these genes are expressed throughout the whole cell cycle and fulfill a variety of functions within the cell⁵⁶. Therefore, it has not been feasible to elucidate whether aneuploidy resulting from impaired SAC protein function is the sole driving force of malignant transformation. Further possible causes for aneuploidy are missegregation events, which occur when the kinetochore of a replicated chromosome attaches to microtubules from the two spindle poles, rather than from merely one (merotelic attachment). Such an event remains undetected by the SAC proteins and is therefore independent of a dysfunctional mitotic checkpoint⁵⁷. A further theory that is being explored focuses on telomere shortening⁵⁸.

Previously, it was outlined that not all cancers are aneuploid. Contrarily, not all aneuploidies lead to cancer: Williams *et al.* reported aneuploidy-introduced growth arrest of murine embryonic fibroblasts (MEFs)⁵⁹. Similar observations are reported by Weaver *et al.*, who describe aneuploid cells due to a lack of *CENP-E*, a gene encoding for a protein associated with the mitotic spindle checkpoint⁶⁰. A subset of cells with specific aneuploidies emerging from the lack of *CENP-E* shows decreased growth, while other cell populations with different patterns of

chromosomal aberrations show unlimited proliferation⁶⁰. For several cancer entities it could be shown that indeed specific rather than stochastic aberrations propagate malignant growth⁶¹. Studies on the effect of aneuploidy on the cellular level have clarified that an altered gene content of the cells is not without consequences: Aneuploidy disrupts global transcription patterns. By using microsomal mediated chromosome transfer, Upender *et al.* could show that aneuploidy due to the gain of one single chromosome results in the deregulation of hundreds of genes. It was in particular remarkable that merely 5–20% of deregulated genes mapped to the introduced trisomic chromosome indicating that aneuploidy causes global gene expression changes⁶².

1.3.2 The Role of Aneuploidy in Colorectal Cancers

In SCCs, at least two different pathways of malignant transformation exist: About 75% of all SCCs show aneuploidy, following the chromosomal instability (CIN) pathway⁵⁴. A smaller subset of SCCs is considered to exhibit “microsatellite instability” (MSI), in most cases due to promoter hypermethylation of the *hMLH1* gene⁶³. The MSI pathway is the hallmark of carcinomas arising in patients with Hereditary-Non-Polyposis-Colorectal-Cancer syndrome (HNPCC)⁶⁴.

A meta-analysis comprising 32 studies and 5,478 patients found different ploidy types in SCC to be associated with an increase in the five-year mortality rate from 29.2% for diploid tumors to 43.2% for aneuploid tumors⁵⁴. This finding is consistent with the fact that HNPCC carcinomas show a favorable prognosis as compared to SCCs⁶⁵.

1.3.3 The Role of Aneuploidy in UC and UCC

Support for the causal role of aneuploidy for cancer development lays in the fact that aneuploid cell populations can frequently be detected in premalignant lesions such as UC: Aneuploid biopsies have been found up to twelve years prior to the diagnosis of UCCs and can serve as prognostic marker for cancer development^{4, 46}. In contrast to dysplasia, aneuploidy seems to be widely spread throughout the whole colon of colitis patients who subsequently develop UCC⁴. Although it is clear that aneuploidy is a valuable marker for prediction of malignant transformation in UC, its frequency in UCCs themselves is insufficiently known: Most studies report on very small samples of UCC cases. In addition, a few case reports exist describing aneuploid UCCs^{66, 67}. In most of these studies, flow-

cytometry was used, which might miss subtle aneuploid cell populations⁶⁷⁻⁷⁰. Furthermore, definitions of the terms “diploidy” and “aneuploidy” varied significantly among the studies, corrupting comparability. The largest cohort of UCC patients investigated by cytometry in a single study consists of 17 patients reported by Burmer *et al.*, who found a frequency of 88% for aneuploidy in UCC⁷¹. However, the 17 neoplastic lesions included in Burmer’s study were comprised of carcinomas and non-malignant dysplastic lesions. Moreover, the effect of aneuploidy in UC on gene expression *in vivo* is not known.

2 SYNOPSIS AND RESEARCH GOALS

Patients with UC bear an increased lifetime risk for the development of CRCs ⁷². For these patients, aneuploidy has been identified as an independent risk factor of malignant transformation ^{4, 46}. Aneuploid CRCs are associated with an inferior prognosis ⁵⁴. Likewise, UCCs could be associated with an inferior prognosis as compared to SCCs ³⁹, but reports are controversial and previous comparative studies did not distinguish between diploid and aneuploid carcinomas ³⁸. The main objective of the first part of this thesis was thus

1. to elucidate whether differences in clinical features and patients' prognoses exist between UCCs and SCCs and,
2. to explore to what extent such differences can be explained by different ploidy status of the tumors.

Specific chromosomal aberrations have been elaborated for UCCs ⁷³. However, little is known of the consequences of aneuploidization on the transcriptome during carcinogenesis in UC. Identifying genes that are differentially regulated during carcinogenesis and related to different ploidy patterns could improve our insight into the molecular changes that characterize CIN and cancer development. Therefore, in the second part of this thesis, it was aimed at

3. elaborating gene expression changes in colonic tissue from patients with non-inflamed colon, diploid UC, aneuploid UC, and UCCs, and
4. establishing a gene expression signature of chromosomal instability in UC.

3 MATERIAL AND METHODS

3.1 General Study Design and Setting

This study consists of two major parts: First, 291 patients with CRC were analyzed with regard to aneuploidy (261 SCCs, 30 UCCs). Correlations between ploidy status and clinical features were elaborated.

Secondly, colonic mucosa was used to analyze gene expression patterns during malignant transformation from non-inflamed mucosa via diploid and aneuploid premalignant stages to UCC. Figure 2 gives an overview of the general study design and the number of patients for each group.

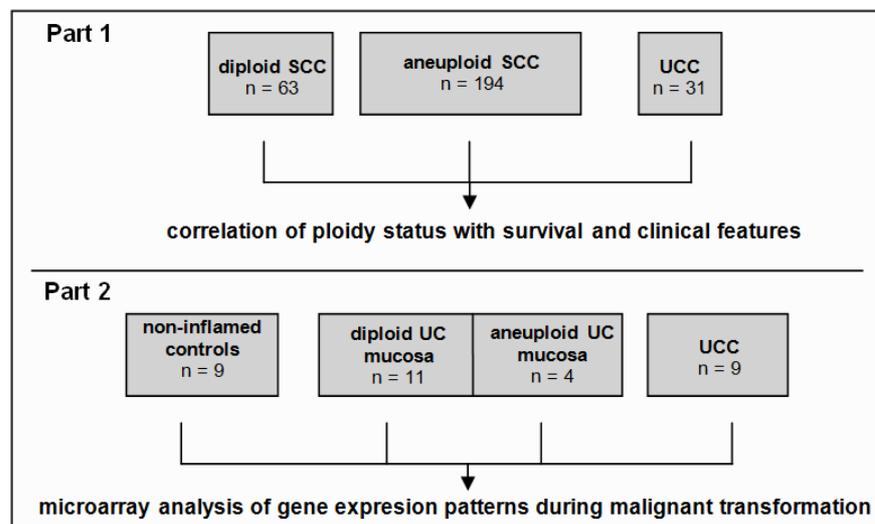


Figure 2: General study design of both separate parts of this thesis; for SCC samples in part one, in 3 cases agreement could not be reached between multiple raters concerning ploidy assessment. Therefore, only 289 out of 291 samples could be used for downstream analyses.

Experiments were performed in close collaboration with different institutes: Feulgen staining and cytometric measurements were conducted at the Laboratory for Surgical Research, University Hospital Schleswig-Holstein, Campus Lübeck. DNA histograms were evaluated in cooperation with the Department of Oncology / Pathology, Cancer Center Karolinska, Stockholm, Sweden (Prof. Gert Auer). Statistical analyses of ploidy data was done in collaboration with the Institute for Medical Biometry and Statistics, University of Lübeck.

Microarray experiments were performed at the Section of Cancer Genomics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA (Dr. Thomas Ried). Statistics for microarray experiments were done by the Computational Systems Biology Laboratory, Genome-Scale Biology Research

Program, Biomedicum Helsinki and Institute of Biomedicine, University of Helsinki, Finland (Dr. Sampsa Hautaniemi). RT-PCR validation of gene targets was done at the Laboratory for Surgical Research in Lübeck.

3.2 Part 1: Patients and Specimens for Ploidy Assessment

Patients with CRC were identified within the *Tumorbank of Colorectal Cancer* of the Department of Surgery, University Hospital Schleswig-Holstein, Campus Lübeck (ethical permits no. 99-121 and 07-124). Criteria for inclusion into the study were histological diagnosis of CRC, operation at the Department of Surgery, University Hospital Schleswig-Holstein, Campus Lübeck, and available imprints of tumor material. Subsequent to their operation, patients included in the database are seen on a regular basis in the department's outpatient clinic for post-operative surveillance (overall mean time of surveillance was $m = 4.74$ years [SD = 0.6 years], mean surveillance for patients included in survival analysis was $m = 6.82$ years [SD = 0.8 years]). In adherence to the Amsterdam II criteria for HNPCC, SCC patients younger than 50 years of age at time of carcinoma diagnosis were excluded from this study ⁷⁴. In total, 5.2% of all carcinomas within the tumor database were affecting patients under the age of 50. A list of clinical features extracted from the tumor database is presented in table 3. For an overview of patients included in part one please refer to table 4.

Clinical data extracted from tumor database for part one of this study
<ul style="list-style-type: none"> ▪ age of patient ▪ sex of patient ▪ type of carcinoma (sporadic or ulcerative colitis-associated) ▪ ploidy status of tumor (diploid or aneuploid) ▪ tumor localization ▪ presence of synchronous carcinoma at time of diagnosis ▪ carcinoembryonic antigen (CEA) levels at time of diagnosis ▪ postoperative resection status (R0, R1, R2) ▪ postoperative tumor histology ▪ postoperative T-, N-, M-stage ▪ postoperative histopathological grading (G1-G3) ▪ postoperative UICC stage (I - IV) ▪ curative / palliative primary operation ▪ survival within observation time ▪ death within 30 days after surgery ▪ months of survival

Table 3: Clinical data obtained for UCC and SCC patients

	<i>Sporadic</i>		<i>UCC</i>
	<i>diploid SCC</i>	<i>aneuploid SCC</i>	
sex (male/female)	31/32	107/87	23/8
Average age at diagnosis (years)	70.3	69.5	49.3
range (years)	50-92	50-88	32-79
average observation time (months)	57.9	57.5	51.6

Table 4: Patients for ploidy analyses, part one

3.3 Part 2: Patients and Specimens for Gene Expression Analyses

Overall, specimens of 41 UC patients were collected. Onset of colitis varied between 1970 and 2001, sample collection was performed between 2003 and 2005 for non-malignant tissue and between 1995 and 2003 for UCCs. Median time between UC diagnosis and sample collection was $m = 18.3$ years ($SD = 6.5$ years). For an overview, please refer to table 5. For premalignant tissue, collection was performed during surveillance colonoscopy at the Institute of Gastroenterology, University Hospital Schleswig-Holstein, Campus Lübeck by one experienced gastroenterologist. Biopsies were taken in a standardized manner, one each from the caecum, ascending, transverse, descending, and sigmoid colon, and one from the rectum. Biopsies were paraffin-embedded and subjected to *Haematoxylin-Eosin* (HE) staining as well as Feulgen staining for image cytometry. Adjacent to the biopsies for image cytometry, samples were obtained for RNA extraction. Immediately after harvesting, samples were snap frozen in liquid nitrogen and kept at -130°C instantaneously before RNA extraction.

Additionally, UCC specimens were obtained operatively. All carcinoma patients presented with locally advanced tumor stages according to the TNM classification (T3 and T4). According to UICC classification, three carcinomas were stage II, five were stage III, and one was stage IV. These nine patients were also comprised in the UCC group of 31 patients from part one (paragraph 3.2, page 11). Only those patients, of whom snap frozen biopsies were available could be included in microarray analyses, limiting the number of carcinomas analyzed for gene expression. A solid block of each tumor was obtained and immediately processed as described above.

Additionally, non-malignant tissue of nine patients was obtained in the same manner. These patients who served as normal controls underwent colectomy for SCC. Specimens were obtained at least 30 cm distant from tumor margins. The absence of malignancy and inflammation was confirmed histopathologically prior to processing as described.

	<i>normal controls</i>	<i>UC biopsies</i>			<i>UCCs</i>
		<i>diploid</i>	<i>mixed</i> ³	<i>aneuploid</i>	
number of patients	n = 9	n = 11	n = 18	n = 4	n = 9
mean age (ys)² (range)	71.7 (59.7 - 83.7)	52.5 (31.8 - 68.0)	48.5 (25.0 - 72.6)	36.5 (19.9 - 52.1)	52.7 (33.4 - 86.7)
sex (m/f)	6/3	7/4	13/5	2/2	9/2
average duration of colitis before biopsy / operation (ys)²	n/a	17 (10 - 25)	20 (7 - 29)	15 (5 - 34)	25.9 (7 - 66)
total number of biopsies for DNA image cytometry	9	45	71	19	9
number of arrays¹	9	18	0	13	7

Table 5: Patients for gene expression analyses

¹representing only those arrays that passed the quality check; ²range given in brackets; ³mixed refers to the presence of both, diploid and aneuploid biopsies as assessed from one patient in biopsies obtained during one surveillance colonoscopy

3.4 Methods

3.4.1 DNA Image Cytometry

The DNA-image cytometry is based on a nuclear staining procedure developed by Feulgen and Rossenbeck in 1924 that utilizes Schiff's reagent to quantitatively stain DNA⁷⁵. Through hydrolysis, purine bases are removed from the DNA, exposing aldehyde groups from the desoxyribose. Schiff's reagent subsequently reacts with the aldehyde groups, forming Schiff's base. The staining procedure is stoichiometric and highly specific for DNA⁷⁶. In this study, a modified protocol was used as described previously (please refer to appendix for the detailed protocol)⁷⁷. Following the staining procedure, cell nuclei were measured using the Ahrens ACAS DNA-image cytometry system, which consists of a Zeiss Axioplan microscope, a CCD camera and a computer with software that allows the selection of single nuclei in an interactive process (Figure 3).

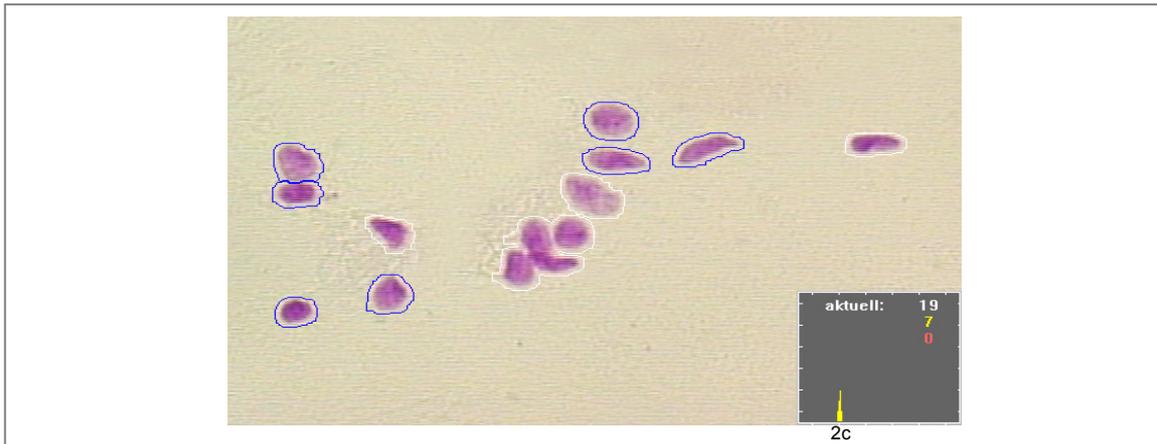


Figure 3: Selection of nuclei with the ACAS software

Feulgen-stained nuclei (here from an SCC imprint) are identified automatically by the software, borders are adjusted manually. Here, seven nuclei are selected showing a single peak at 2c.

The system was calibrated using cerebellar neuronal nuclei, which were given the value 2c. In addition, lymphocytes were measured in each specimen as internal controls as described previously⁷⁸. Paraffin embedded sections of 8 μ m thickness (UC, UCC) or imprints (SCC) were subjected to image cytometric measurements. At least 100 cell nuclei were measured for each specimen.

The DNA profiles were classified according to Auer (Figure 4)⁷⁷. 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, commonly referred to as “diploid”. Type IV histograms showed increased (>5%) and distinctly scattered DNA values exceeding the tetraploid region (>4.5c). These histograms reflect aneuploid populations of colon mucosa nuclei. All DNA histograms were evaluated by three independent investigators being unaware of either the clinical or histopathological data of according patients. Figure 4 shows examples of ploidy histograms as assessed according to Auer’s classification.

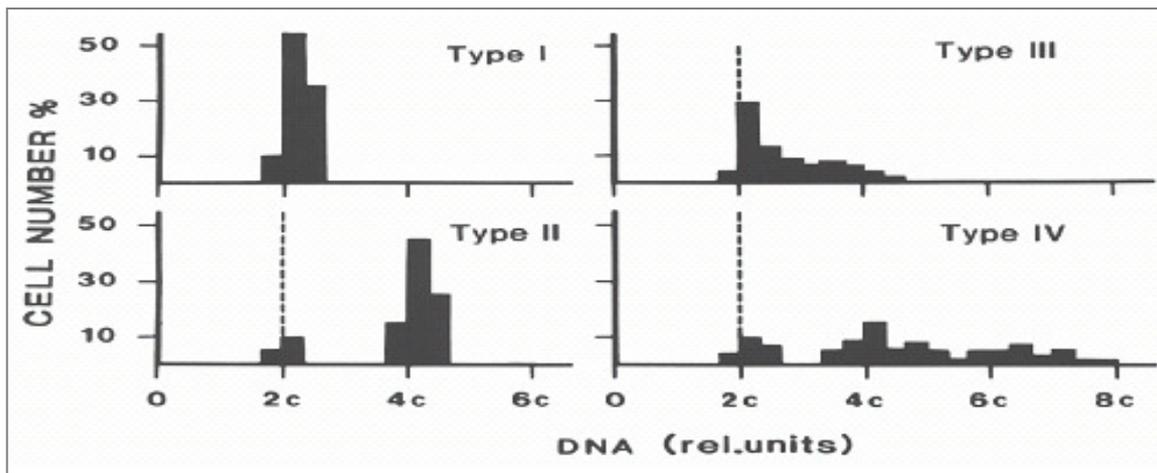


Figure 4 Ploidy types according to Auer's classification

Types I, II, and III represent diploid cell populations, type IV characterizes aneuploid cell populations.

3.4.2 Oligonucleotide Microarrays

3.4.2.1 Basic Principle

Microarray technology was first introduced in 1995 by Schena *et al.*⁷⁹. Despite major technical improvements, the basic principal remained unchanged: RNA is extracted from cells and stained quantitatively with fluorescent dyes. Thereafter, it is hybridized to glass slides bearing complementary nucleotide probes at defined positions. After incubation of the target-RNA to its probes, the arrays are scanned by energizing the fluorochromes using dye-specific laser wavelengths and detecting the ignited signal with a γ -camera⁷⁹. In order to account for inter-array heterogeneities, a reference RNA is hybridized concurrently. Figure 5 gives an overview of the basic microarray workflow as it was pursued in this thesis.

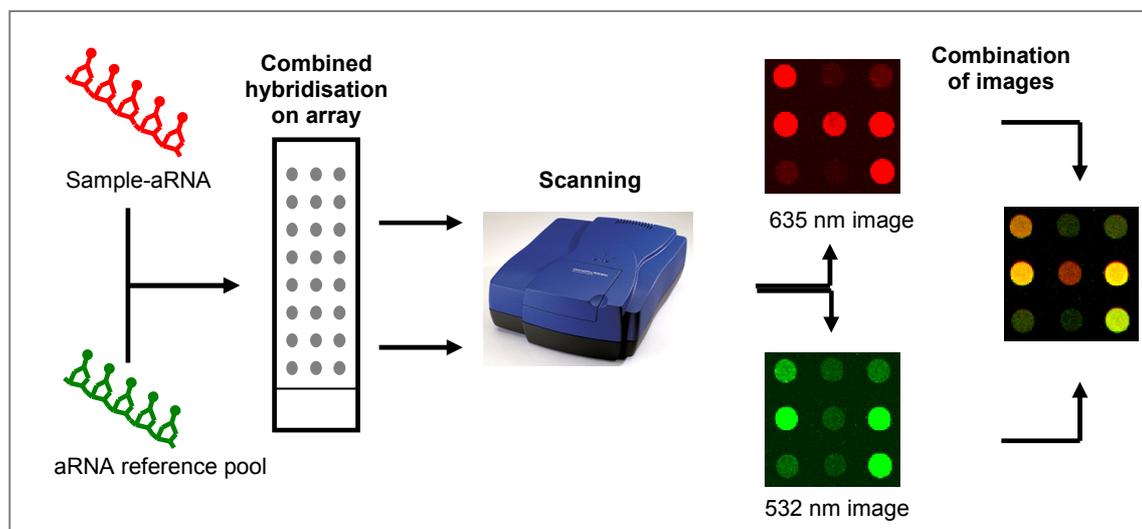


Figure 5: Synopsis of microarray workflow

Amplified RNA is labeled with fluorescent dyes (Cy3 and Cy5). Images are obtained at $5\mu\text{m}/\text{pixel}$ resolution for each dye and combined for post-processing.

3.4.2.2 *Arrays used in this thesis*

Custom-designed microarrays were obtained from the microarray core facility of the National Cancer Institute in Frederick, MD, USA. Arrays were printed with sequences of 34,580 gene products based on version 3.0 of the Hs-Operon OligoSet (for details please visit <http://nciarray.nci.nih.gov/index.shtml>, last access March 2011). Pre-synthesized oligonucleotides containing a 5'-terminal aminolinker were printed to a 3-amino-triethoxysilan coated glass slide. Probes were chosen based on the Ensembl gene database (<http://www.ensembl.org>) and are commercially available (Operon, Huntsville, Alabama, USA, https://www.operon.com/array/OpArray_human.php; last access June 2010).

3.4.2.3 *RNA Preparation and Microarray Hybridization*

Based on the standard protocol provided by the core facility, an adjusted protocol was developed to allow hybridization of aRNA. For the detailed protocols please refer to the appendix (pages 74ff). In the following, a brief outline of the steps for array hybridization is given.

3.4.2.4 *RNA Extraction*

Tissue was transferred to phenol/guanidine-isothiocyanate buffer (TRIzol reagent ®). As for the colonoscopically acquired material and the tumor samples, the whole sample was used for RNA extraction. The non-malignant tissue (normal controls) was macrodissected immediately after thawing while removing all layers except for mucosa and submucosa to assure comparability. The RNA/buffer solution was washed through silica filters, thereby adsorbing RNA that was resuspended in RNase-free water. RNA concentration was measured using a micro-photometer (Nanodrop, Version 2.4.5, Thermo Scientific, Wilmington, USA).

3.4.2.5 *RNA Amplification*

RNA was amplified with a commercially available kit from Ambion (Ambion MessageAmp Kit, Ambion, Texas, USA)⁸⁰. The extracted total RNA was used to create complementary DNA (cDNA) using T7 Oligo(dT) primers and a reverse transcriptase. Primers bind to the 3'poly(A) tail of mRNA and reverse transcriptase synthesizes a complementary DNA strand⁸¹. Thereafter, a DNA polymerase synthesizes double stranded DNA from that template (dsDNA). This template is used for linear transcription to generate amplified RNA (aRNA), whereby

aminoallyl-UTP was added in a 1 to 1 ratio to TTP to synthesize RNA with aminoallyl residues for the subsequent coupling of fluorescent Cy3 and Cy5 dyes.

3.4.2.6 Coupling of Cy-Dyes

Following amplification, 50% of all UTP nucleotides bear amino-allyl residues available for ester-coupling of fluorescent dyes (Figure 6). Cy3 and Cy5 (Amershan Pharmacia, Piscataway, NJ, USA), respectively, were added to the aRNA solution. Dye incorporation was measured with the micro-photometer (wavelength of 532 nm for Cy3 and 635 nm for Cy5). Only aRNA with highly similar concentrations and dye incorporation ratios was used.

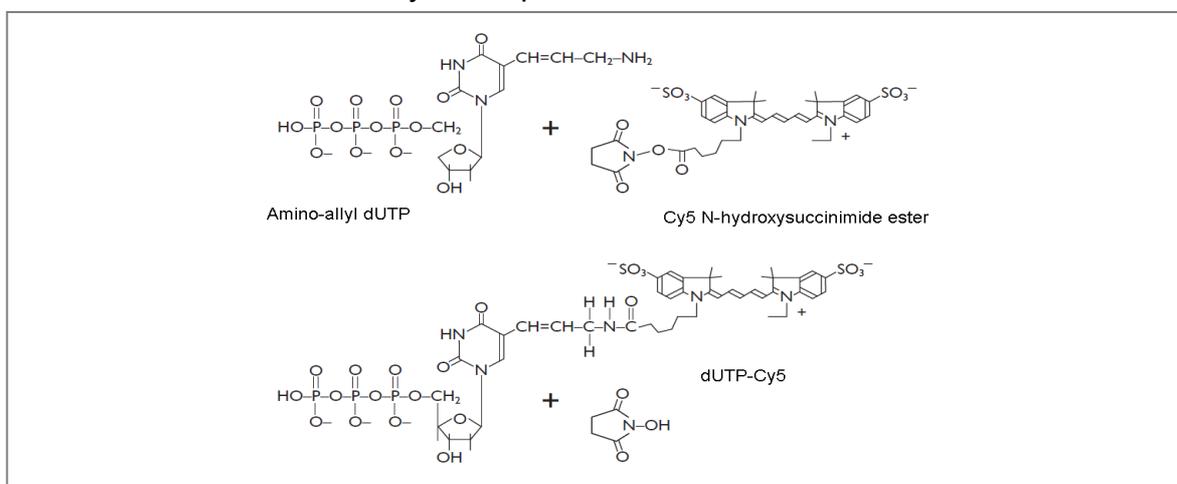


Figure 6: Coupling of Cy-5 dye to amino-allyl residues of dUTP

Cy5 (and Cy3 likewise) gets covalently attached in a C-N bound. Image modified according to Ambion Amino Allyl MessageAmp™ aRNA Kit protocol, version 0302

3.4.2.7 aRNA Fragmentation

RNA fragmentation facilitates binding of RNA to the 70-80mer oligonucleotides on the array and enhances signal strength of the spots on the array⁸². RNase (RNA Fragmentation Reagents, Ambion, Texas, USA) was given to the aRNA solution, fragmenting RNA to 60 to 200 base pairs.

3.4.2.8 Microarray Hybridization

Arrays were pre-hybridized with buffer containing bovine serum albumin to avoid unspecific binding of target RNA. Reference aRNA and sample aRNA were used for incubation of the array in a sealed humid chamber at 48°C for 16 hours. Subsequently, arrays were washed, spinned dry, and scanned within 24 hours after hybridization to avoid bleaching artifacts.

3.4.2.9 Microarray Scanning

Scanning was performed at a resolution of 5 $\mu\text{m}/\text{pixel}$ using GenePix Software version 4.0.1.17 and an Axon Microarray Scanner 4000B (both MDS Analytical Technologies, California, USA). Cy-dyes were energized at 532 nm for Cy3 and 635 nm for Cy5. The signal was amplified using photon-multiplier-tubes (PMTs) to create a digital image on a 16-bit scale (maximum intensity 65536). PMTs were manually adjusted to reduce the number of saturated spots (intensity > 65535) and to match the distribution of the intensities of both wavelengths, adjusting for dissimilarities in incorporation of the two dyes arising from chemical differences between the fluorescent molecules.

3.4.2.10 Microarray Gridding

The two-color images were overlaid with a grid derived from the Gene Allocation File (GAL file) containing spatial information for all 34,580 gene products and assigning every spot on the array to its known oligonucleotide sequence. Gridding was performed in an interactive process (figure 7)⁸³. First, the grid was overlaid manually and single spots were then automatically detected by the GenePix Software. Thereafter, a visual control was performed for all spots, correcting spot detection where necessary. In parallel, a visual quality check was performed to exclude spots with evident artifacts. For all spots, mean intensity values for both channels (reference and sample RNA) were calculated.

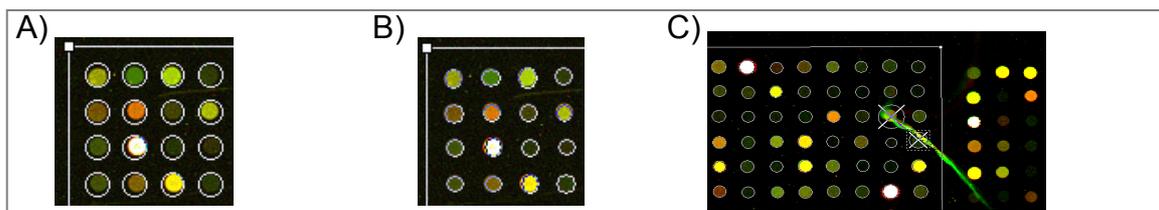


Figure 7: Microarray gridding process with GenePix © Software

First, a grid containing the information of gene features for each spot is manually overlaid over each block (A). Then, the software detects each spot automatically and assigns the correct pixel area to each feature (B). Automatic assignment has to be controlled visually for each spot. C) shows an example of minor scratches on the array surface, leading to aberrant fluorophore signals in a confined area. Spots within the damaged area have to be selected manually and flagged “bad” (illustrated by an “X” through the spot circle). These spots are excluded from downstream analyses.

3.4.3 Real Time-quantitative PCR (RT-qPCR)

3.4.3.1 General Principle

Real time reverse transcription quantitative polymerase chain reaction (RT-qPCR) was used with SybrGreen 1 as fluorescent dye for amplicon quantification. Gene expression of ten targets in the gene lists generated by microarray experiments was measured in 33 randomly selected samples. PCR was based on non-amplified mRNA to include testing for amplification related biases. 1 µg of total sample RNA was used as described previously⁸⁴. Please refer to the appendix, page 81f, for the detailed protocols, reagents, and cycler specifications used.

For downstream analysis, $\Delta\Delta CT$ values and relative expression were calculated under consideration of the respective reference group (normal controls, diploid colitis, and aneuploid colitis, respectively) as described previously⁸⁵, with

$$E_r = 2^{-\Delta\Delta CT}$$

where

E_r : relative expression to reference group

CT: PCR cycle in which signal reaches detection threshold

$$(\Delta CT = CT_{\text{target}} - CT_{\text{house keeping gene}}; \Delta\Delta CT = \Delta CT_{\text{reference group}} - \Delta CT_{\text{comparative group}})$$

Subsequently, expression was compared to the analogous ratios of the array expression values on linear scale.

3.4.3.2 RT-qPCR Reference Gene Selection Based on the own Dataset

In order to identify a suitable reference gene, we chose 14 commonly used “house keeping genes” (HKGs) (*RPLP0*, *GAPDH*, *GUS*, *β -actin*, *HPRT*, *β 2-microglobulin*, *RS18*, *POLR2A*, *PGK1*, *TBP*, *YWHAZ*, *UBC* (Ubiquitin), *RPL13A*, *HMBS*⁸⁶⁻⁸⁸) and calculated the coefficient of variation (CV) of their non-log transformed array expression values in all arrays for each gene. The gene with the smallest coefficient of variation (CV = 0.16; phosphoglycerate kinase-1 [*PGK1*]) was chosen for normalization of PCR data⁸⁹.

3.4.3.3 RT-qPCR Primer Design

Forward and reverse primers for PCR were designed using CLC DNA workbench software version 5.0 (<http://clcbio.com>; last accessed September 2010; CLC bio Aarhus, Denmark). In case a gene consisted of more than one exon, primers were

designed to cover exon-exon intersections, minimizing interference by contaminating genomic DNA. Primers were chosen to ideally cover the oligonucleotide sequence printed on the array. Figure 8 illustrates the criteria used for the spatial setting of primers and PCR amplicons.

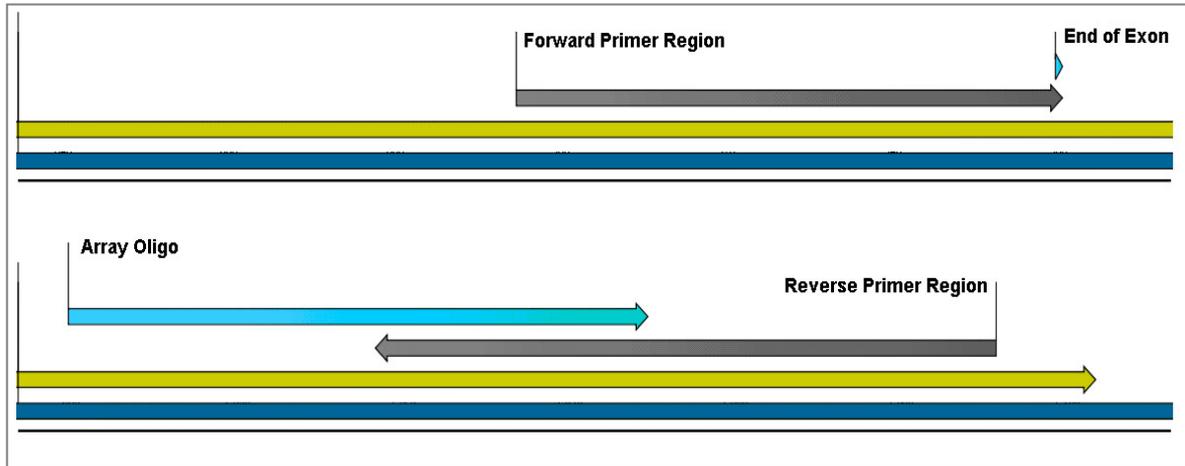


Figure 8: Example of spatial setting of primers for RT-PCR validation

Yellow strand represents translated sequences of the gene to be amplified; dark blue strand illustrates hmRNA (including non-translated regions). Light blue arrows illustrate the location of the oligonucleotide printed on the array for this respective gene ("ArrayOligo") and the end of an exon in the gene sequence ("End of Exon"). Please note that the regions for forward and reverse primers (grey arrows) are chosen to a) include the sequence of the oligonucleotide on the array and b) include an exon-exon intersection. Image derived from CLC workbench 5.0 software (CLC bio, Aarhus, Denmark).

After specifying the spatial locations of the primers, additional limitations were set considering the following parameters as described previously⁸⁴:

- Melting temperature (T_m): 56 – 61 °C
- GC-content. 45 - 60%
- Primer length: 18 - 22 base pairs
- Amplicon length: 50 – 150 base pairs

Scores for primer self-annealing, secondary structure, primer end-annealing, and primer dimers were calculated using a patented algorithm implemented in the CLC bio software. Primers with the best scores matching all other criteria described above were chosen for RT-qPCR.

3.4.3.4 RT-qPCR Efficiency Testing

Each primer was tested in a series of five dilution steps in pooled cDNA derived from colonic mucosa and SCCs (1.5 µg; 0.75 µg; 0.375 µg; 0.1875 µg; 0.09375 µg of pooled cDNA). PCR efficiency was assessed by plotting Δ CT values for the dilution steps on logarithmic scale. Slopes of linear regression graphs were calculated, with a slope of 0 representing perfect efficiency. Primers with slopes < 0.1 or > 0.1 were discarded and new primers were designed. Additionally, RT-qPCR melting curves were assessed as described previously⁸⁴: Double peaks were suggested to represent formation of primer dimers. Thus primers repeatedly showing double peaks in melting curves were discarded and new primers were designed. Please refer to table 6 for a summary of primers used.

Gene	forward primer	reverse primer
CSPG6	AGAGCAGCAACAGGAAAG	GACGGAAGTGGTCTAGCA
KIF20B	CAACCAAACGAGCCAAA	GATCACTCTCCTTCATTTTC
CENPH	GCACAGACAAAACAACAAC	GCTTCGATTTGCTTTTCTTG
TNFAIP3	CCCCATTGTTCTCGGCTA	TCTTCCCCGGTCTCTGTT
LAMC2	GAATGGAAAAAGTGGGAGAGAG	AAGATTGGCACGGGAAAG
CYR61	ACAGCCAGTGTACAGCAGCCTGA	GGGCCGGTATTTCTTCACACTCA
TRIB3	CCAGAAGGGAGAAAGGCAGAAGC	CACCCAAGCAGGAACTGCATGT
SMARCA1	GCGAGCCAAAGATTCCAAAGGC	CCAGGAGCTCAAATAAGCGTGGTG
NUF2	CAAAAGCAAACGACAAAGTCGG	TCCTGGTGTGCGGCGTTTAAC
TFPI2	GGGCAACGCCAACAATTTCTACA	TGGTCGTCCACACTCACTTGCA

Table 6: Forward and reverse primers for RT-qPCR validation

3.4.4 Statistical Analyses

3.4.4.1 Part 1: Statistical Analyses of Ploidy Data

Data from ploidy measurements and clinical data were combined and statistical analyses were performed using the software package SAS version 9.1 (SAS, Heidelberg). Data were presented using descriptive statistics like mean values (m), standard deviations (SD) or frequencies. For inductive inference, non-parametric rank-sum tests were used to compare location parameters of data distributions. More precisely, two independent samples e.g. having different ploidy status were compared using Wilcoxon's test. In addition, corresponding frequencies were analyzed by computing Fisher's exact test.

For multivariate analysis, logistic regression analysis was performed to determine the effect of eight variables (age, UICC status, underlying inflammation, ploidy

status, T-, N-status, histological tumor grading) on patients' prognosis. Survival 60 months after surgery was defined as the dependent variable. Patients with palliative surgery, non-R0 resection, distant metastases at time of diagnosis, and patients who died within 30 days after operation were excluded from logistic regression, leaving 216 patients for this analysis (190 SCC and 26 UCC patients). For the independent variables that met the criteria of significance, a reduced model was generated. To test for inter-independence of the significant variables, further logistic regression analyses were performed, each with one single significant variable left out of the calculation. Independence was assessed with regard to the alteration of odds ratios (ORs) of the remaining variables. The same exclusion criteria were chosen for survival analysis, except for patients with tumor stage UICC IV being included, leaving 222 patients for survival plot analyses (196 SCC and 26 UCC), which were estimated using Kaplan-Meier estimates and compared using a log-rank test.

In addition, inter-observer reliability was analyzed using Fleiss' kappa index (K)⁹⁰. Observations were illustrated by histograms and estimated Kaplan-Meier plots. The type 1 error rate for testing statistical hypotheses was set to 5% (0.05).

3.4.4.2 Part 2: Statistical Methods used for Microarray Experiments

Normalization of microarray data was done using locally weighted scatterplot smoothing (lowess). Group comparisons were performed using the false discovery rate (fdr) to account for multiple testing error⁹¹.

First, for both channels at a single spot, the segmented pixel values were summarized with median. After background subtraction, median intensities at each spot were divided by the median intensity of the corresponding spot in the reference hybridization. The ratios were then log-transformed natural base. Within-slide normalization was done with lowess in MA-space, where

$$M = \log_{10}(R/G), \text{ and}$$

$$A = \log_{10}(\sqrt{RG})$$

and where R and G correspond to the median intensities of the sample RNA and the reference RNA. After within-slide normalization, quality control was performed based on the MA-plots and the cumulative distribution estimates. Samples with clear aberrations were excluded from downstream analysis. Student's t-test (two-sided) was used in order to identify DEGs. The false discovery rate (fdr) according

to Benjamini and Hochberg was computed for each gene tested in every group comparison⁹¹.

In order to discover constantly up and down regulated genes, a linear model was developed, which identified genes that were constantly up or constantly down regulated over the sequence from normal mucosa via diploid and aneuploid colitis to cancer. Medians of expression values were ordered according to the sequence from normal mucosa, diploid colitis, aneuploid colitis, and UCCs. Using the least squares method, a curve was fitted to the data points. The curves with the highest positive slopes were considered as being candidates for constantly up regulated genes, while those genes with the lowest negative slopes were considered to be the candidates for constantly down regulated genes. We computed a list of the 25 highest ranked genes for up- and down regulation, respectively. Therein, we chose those genes, that had a mean constantly increasing or decreasing over the whole sequence and that were significantly differentially expressed ($p \leq 0.0001$) in the comparison of normal mucosa vs. UCC.

For visualization and network analysis Ingenuity Pathways Analyses (Ingenuity Systems, Mountain View, CA, www.ingenuity.com) was applied to identify canonical pathways of interest from the obtained gene lists. IPA is a comprehensive database and software based on the Ingenuity Pathways Knowledge Base^{92, 93}. After querying the IPA data base with the gene lists created, genes were overlaid onto a global molecular network developed from information in the Ingenuity Pathways Knowledge Base. Networks of these genes were then algorithmically generated based on their connectivity. The IPA software creates networks rated by scores, which represent the negative exponent of a p-value calculation and indicate the number of network eligible genes within a network; the higher the number of network eligible genes in a network, the higher the score. Networks with a score ≥ 5 were considered significant (please refer to www.ingenuity.com for details, last access January 2011).

3.5 Synopsis of Workflow for the Gene Expression Analyses

Figure 9 gives an overview of the workflow pursued to elucidate gene expression changes during malignant transformation in UC.

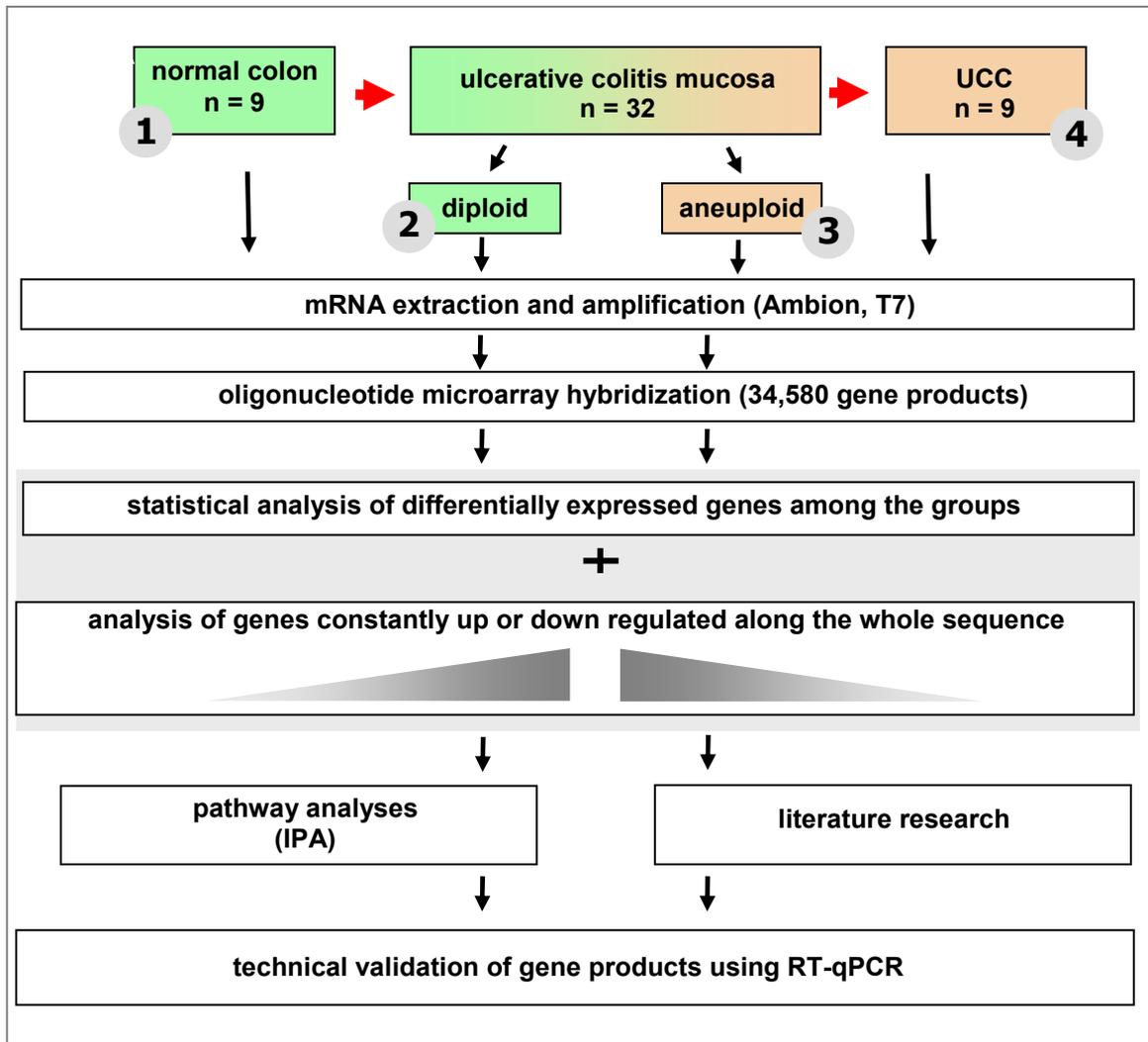


Figure 9: Study design for gene expression analyses

First row represents tissue specimens and gives number of patients available for downstream analysis after quality check. After ploidy measurements (second row) four main groups were defined: 1. Normal controls, 2. diploid biopsy samples from patients with ulcerative colitis, 3. aneuploid biopsies from patients with ulcerative colitis, 4. UCCs.

4 RESULTS

4.1 Results from Ploidy Assessment in Neoplastic Lesions

4.1.1 Ploidy Assessment by DNA Image Cytometry

For all 291 carcinoma specimens, an average of $m = 216$ cell nuclei were measured (range 100 to 586, $SD = 186.9$ nuclei) depending on the availability of non-overlapping nuclei. Within the group of SCCs, 63 out of 260 tumors (24.2%) showed a diploid DNA distribution pattern. The remaining 194 SCCs showed aneuploid patterns, while for three tumors no agreement in ploidy-classification could be reached. Within the group of UCCs ($n = 31$) all specimens showed aneuploidy (type IV of Auer's classification as shown in figure 4).

4.1.2 Inter-Observer Reliability of Ploidy Assessment

Ploidy was assessed independently by three investigators. When all four subtypes of Auer's classification were considered on all 291 carcinoma samples, strong agreement was obtained (Fleiss' $K = 0.962$). When merely distinguishing euploid (subtype I, II, III combined) from aneuploid (IV) cell populations, disagreement was observed for three out of 291 samples, accounting for an overall strong agreement ($K = 0.980$). The three samples, in which concordance in the gross ploidy assessment could not be reached (1% of the total of all tumors), were left out of further analyses.

4.1.3 Clinical Differences between Patient Groups

4.1.3.1 Age and Sex

Patients in the UCC group were younger ($m = 49.3$ years, $SD = 11.8$ years) than in the SCC group ($m = 69.7$ years, $s = 10.0$ years; $p < 0.0001$). This observation remained valid when UCCs were compared separately with diploid SCCs ($m = 70.3$ years, $SD = 10.6$ years; $p < 0.0001$) and aneuploid SCCs (mean 69.5 years, $SD = 9.8$ years; $p < 0.0001$), respectively. 25.8% of the UCC patients were female (8 out of 31) compared to 46.3% in the SCC group (119 out of 257) ($p = 0.035$). This difference was more pronounced when comparing UCCs with diploid SCCs (50.8% females; $p = 0.027$). There was no statistically significant difference in sex comparing the UCC with the aneuploid SCC patients (44.2% female, $p = 0.052$). When comparing aneuploid and diploid carcinomas regardless of sporadic or UC-

associated origin, there was no significant difference in patients' sex. However, patients with aneuploid tumors were significantly younger than patients with diploid tumors (average of $m = 66.7$ years for the aneuploid group, $SD = 12.3$ years; and $m = 70.3$ years for the diploid group, $SD = 10.6$ years; $p = 0.036$). Within the SCC group, after stratification into diploid and aneuploid carcinomas, there was no significant difference in age or sex.

4.1.3.2 Synchronous Carcinomas and Histology

While in the UCC group ten out of 31 patients (32.3%) had a synchronous carcinoma, this was the case for 34 out of 257 SCC patients (13.2%; $p < 0.0001$, figure 10). The significant difference remained when comparing UCCs separately to diploid ($p = 0.004$) and aneuploid SCCs ($p = 0.0002$), respectively. In addition, UCCs showed differences in the distribution of histological subtypes when compared to SCC tumors ($p < 0.0001$). These differences were still detectable when comparing UCCs separately to the diploid ($p = 0.034$) and aneuploid SCCs ($p < 0.0001$). The sporadic malignancies were mainly comprised of adenocarcinomas, whereas mucinous carcinomas were more frequently observed in the UCC group. There was no significant correlation between histological subtype and the presence of synchronous malignancy. When comparing aneuploid and diploid carcinomas regardless of sporadic or UC-associated origin, there were neither significant differences in the frequency of synchronous carcinomas nor were there differences in histological subtypes.

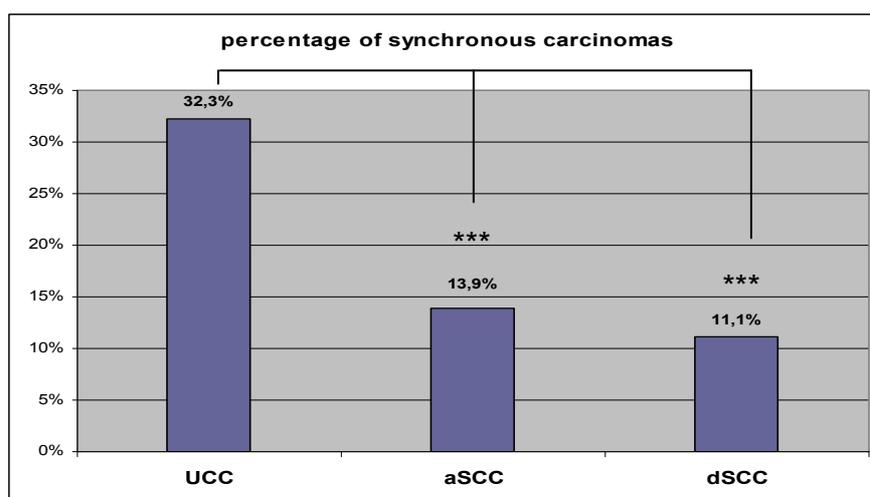


Figure 10: Percentage of synchronous carcinomas in UCC, aSCC, and dSCC patients. Please note the statistically significant differences of UCCs compared to aSCCs ($p = 0.0002$) and dSCCs ($p = 0.004$).

4.1.3.3 Staging

When comparing aneuploid and diploid carcinomas regardless of sporadic or UC-associated origin, cancer was diagnosed at advanced stages (UICC III and IV) in 50.7% of all aneuploid entities, while 48.9% were discovered in UICC stages I and II (no data for this variable was available for the remaining 0.4% of aneuploid SCCs). As for diploid entities, UICC III and IV tumors could be found in 42.9%, while 57.1% were found in UICC stages I and II. These differences did not reach the level of significance ($p = 0.25$). Significant differences could neither be found for any other TNM criteria, nor for UICC criteria in this group comparison. Furthermore, in the comparison of SCCs with UCCs, there was no significant difference in tumor stage as defined by TNM and UICC criteria. However, when comparing aneuploid and diploid SCCs alone, a significantly higher frequency of metastases at time of diagnosis could be observed in the aneuploid group ($p = 0.046$), while for T- and N-, as well as for UICC-status no differences could be shown. Figure 11 illustrates the frequencies of metastases among UCC, dSCC and aSCC groups.

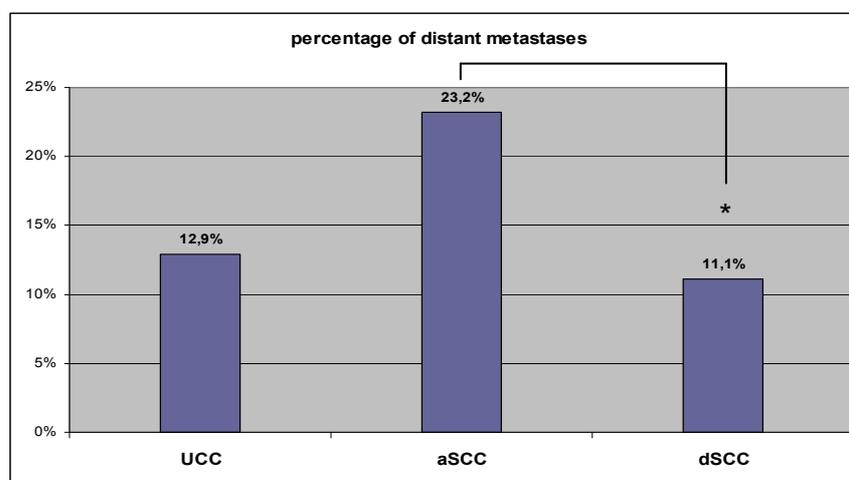


Figure 11: Percentage of distant metastases in UCC, aSCC and dSCC groups. Between aSCC and dSCC patients the difference in metastases frequency was significant ($p = 0.046$), while there was no significant difference between UCCs and neither aSCCs or dSCCs.

4.1.3.4 Grading

Overall, UCC tumors showed a higher degree of dedifferentiation (G3) than SCC tumors ($p < 0.0001$, figure 12). This finding remained significant when UCCs were compared separately to diploid ($p = 0.0005$) and aneuploid ($p < 0.0001$) SCCs. There was a correlation between the histological subtype and the degree of dedifferentiation with uncommon histological subtypes such as mucinous

carcinomas showing higher tumor grading ($p = 0.001$). When comparing aneuploid and diploid carcinomas overall, there was no significant difference regarding tumor grading. Within the group of SCCs there was no difference in tumor grading between diploid and aneuploid SCCs.

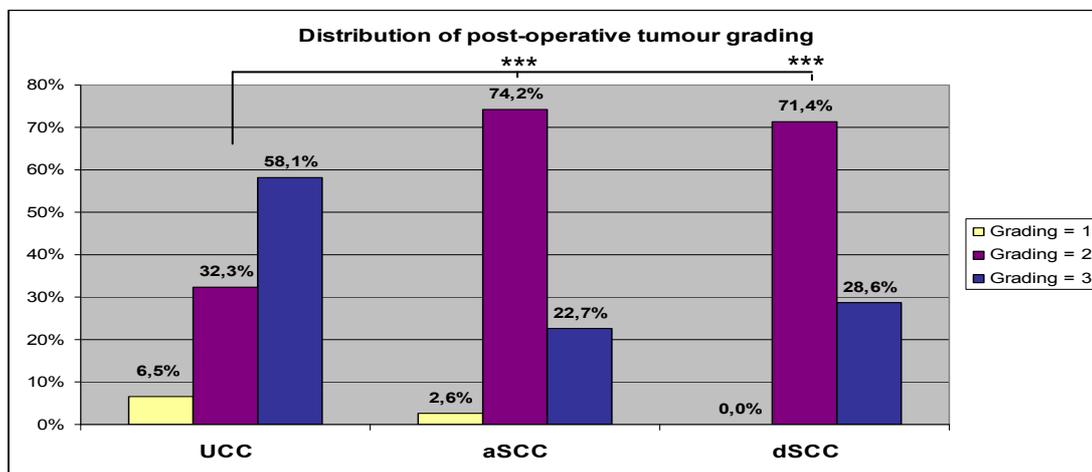


Figure 12: Distribution of post-operative tumor grading among UCCs, aSCCs and dSCCs. Please note the significantly higher degree of dedifferentiation in the UCC group when compared to aSCC and dSCC groups ($p < 0.0001$ and $p = 0.0005$). In case total percentages do not add up to 100% (UCCs and aSCC), single observations for this variable were missing.

4.1.3.5 Resection Status

For nine SCC patients (3.5%) R0-resection could not be achieved, while this was the case for four UCC patients (12.9%; $p = 0.047$). This difference remained significant when comparing UCCs with aneuploid SCCs ($p = 0.048$). There was a significant correlation between UICC staging and resection status, with R1- and R2-resection being more frequently found in advanced carcinomas (UICC III and IV; $p = 0.040$). When comparing aneuploid and diploid carcinomas regardless of sporadic or colitis-associated origin, there was no significant difference regarding the resection status ($p = 0.555$).

significant parameter	groups used in comparison and descriptive statistics		p-value ¹
	first group	second group	
Age	aneuploid CAs (mean: 65,1 yrs)	diploid CAs (mean: 67,6 yrs)	0.0363
	aneuploid SCC (mean: 69.5 yrs)	UCC (mean: 49.3 yrs)	0.0001
	diploid SCC (mean 70.3 yrs)	UCC (mean: 49.3 yrs)	0.0001
	all SCC (mean 69.7 yrs)	UCC (mean: 49.3 yrs)	0.0001
adenocarcinomas versus rare histologic types²	aneuploid SCC (n = 193) (adeno: 95.3% rare: 4.7%)	UCC (n = 30) (adeno: 73.3% rare: 26.7%)	0.0002
	diploid SCC (n = 63) (adeno: 92.1% rare: 7.9%)	UCC (n = 30) (adeno: 73.3% rare: 26.7%)	0.0336
	all SCC (n = 256) (adeno: 95% rare: 5%)	UCC (n = 30) (adeno: 73.3% rare: 26.7%)	0.0001
postoperative resection status³	all SCC (n = 257) (R0: 96.5% R1: 3.1% R2: 0.4%)	UCC (n = 31) (R0: 87.1% R1: 9.7% R2: 3.2%)	0.0472
	aneuploid SCC (n = 194) (R0: 96.9% R1: 2.6 R2: 0.5%)	UCC (n = 31) (R0: 87.1% R1: 9.7% R2: 3.2%)	0.0477
presence of synchronous carcinoma	all SCC (n = 257) (synchr.: 13.2%)	UCC (n = 31) (synchr.: 32.3%)	0.0001
	diploid SCC (n = 63) (synchr.: 11.1%)	UCC (n = 31) (synchr.: 32.3%)	0.0044
	aneuploid SCC (n = 194) (synchr.: 13.9%)	UCC (n = 31) (synchr.: 32.3%)	0.0002
tumor grading³	diploid SCC (n = 63) (I: 0%; II: 71,4%; III: 28,6%)	UCC (n = 30) (I: 6,7%; II: 33,3%; III: 60%)	0.0005
	all SCC (n = 256) (I: 2%; II: 74%; III: 24%)	UCC (n = 30) (I: 6,7%; II: 33,3%; III: 60%)	< 0.0001
	aneuploid SCC (n = 193) (I: 2.6%; II: 74.6%; III: 22.8%)	UCC (n = 30) (I: 6,7%; II: 33,3%; III: 60%)	< 0.0001
UICC stage	all SCC (n = 257) I: 17.5% III: 29.6% II: 32.7% IV: 20.2%	UCC (n = 30) I: 20% III: 30% II: 36.7% IV: 13.3%	0.0047
	aneuploid SCC (n = 194) I: 16.5% III: 28.9% II: 31.4% IV: 23.2%	UCC (n = 30) I: 20% III: 30% II: 36.7% IV: 13.3%	0.0083
Sex	all SCC (n=257) (m: 53.7%; f = 47.3%)	UCC (n = 31) (m: 74.2%; f = 25.8%)	0.0351
	diploid SCC (n = 63) (m : 49.2%; f = 50.8%)	UCC (n = 31) (m: 74.2%; f = 25.8%)	0.0269
Aneuploidy	all SCC (n=257)	UCC (n=31)	0.00058
M-status³	diploid SCC (n = 63) (M1: 11.1%)	aneuploid SCC (n = 194) (M1: 23.2%)	0.0464

Table 7: Significant differences in group comparisons

Left column represents differentially distributed variables. Only variables with p-values < 0.05 are shown. They are followed by the respective groups, in which analyses yielded significant differences. Rightmost column shows corresponding p-values. ¹p-values from Fisher's exact test, student's t-test and Wilcoxon rank sum test, respectively (please refer to methods section for details). ²comprising rare histopathological types such as mucinous carcinomas and signet-ring cell carcinomas; ³ according to TNM classification. yrs = years; m = male; f = female; CA = carcinoma

4.1.4 Logistic Regression Analysis

For logistic regression analysis, eight different parameters were selected based on routine clinical application (age, sex, UICC stage, T-, and N-status, histological tumor grading) and hypothesized features of prognostic impact (underlying inflammation, DNA ploidy status). These parameters were suggested to reflect the most promising candidates of prognostic significance^{54, 94-97}.

Logistic regression yielded two parameters of significant prognostic value for five-year survival subsequent to operation for CRC (table 8). Those two significant parameters were age and DNA ploidy status, indicating that patients of higher age at diagnosis or with aneuploid tumor cell populations have an inferior survival. Additional logistic regression analysis in the reduced model confirmed age (OR 1.05, 95% CI 1.02 – 1.09; $p = 0.003$) and DNA aneuploidy (OR 4.07, 95% CI 1.46 – 11.36; $p = 0.007$) to be independent prognostic parameters (see “Methods”).

<i>parameter</i>	<i>point estimate</i>	<i>95% CI</i>	<i>p-value</i> ¹
age	1.05	1.02-1.09	0.0031
sex	0.64	0.32-1.26	0.1972
UICC status	1.04	0.49-2.24	0.9163
underlying colitis	1.12	0.88-1.42	0.372
DNA aneuploidy	4.07	1.46-11.36	0.0072
T – status ²	1.31	0.74-2.30	0.3537
N – status ²	1.15	0.51-2.56	0.7355
grading ²	1.38	0.66-2.92	0.3947

Table 8: Logistic regression analysis

Logistic regression analysis was performed for eight variables indicated above with endpoint defined as survival after five years subsequent to operation for colorectal malignancy. Patients who died within 30 days after operation, patients who were not operated on in a curative approach, patients for whom R0 resection could not be achieved, and patients with metastases at time of diagnosis were excluded (please refer to “Materials and Methods” for details); thus, regression analysis could be performed on altogether 216 patients. ¹p-values from Chi²-test; ²T- and N-status and tumor grading were determined postoperatively by histopathological examination; significant p-values are printed in bold letters

4.1.5 Survival Analyses (Kaplan-Meier-plots)

Kaplan-Meier-Plots are presented in figure 13. No difference in survival became evident when comparing UCCs and SCCs (plot A). When stratifying SCCs into diploid and aneuploid carcinomas, there was a trend for longer survival for patients with diploid SCCs (five-year survival: 88.2%) when compared to UCCs (73.1%) and aneuploid SCCs (69.0%) ($p = 0.074$; plot B). Stratification of UCCs and SCCs into early UICC stages [I, II] and advanced UICC stages [III]) showed most unfavorable survival for late stage UCCs (plot C). Stratification according to ploidy type of SCCs yielded most favorable outcome for diploid SCCs in early stages, followed by advanced diploid SCCs, UCCs in early stages, early and advanced aneuploid SCCs, and worst outcome for UCCs (all aneuploid) in advanced stages (plot D).

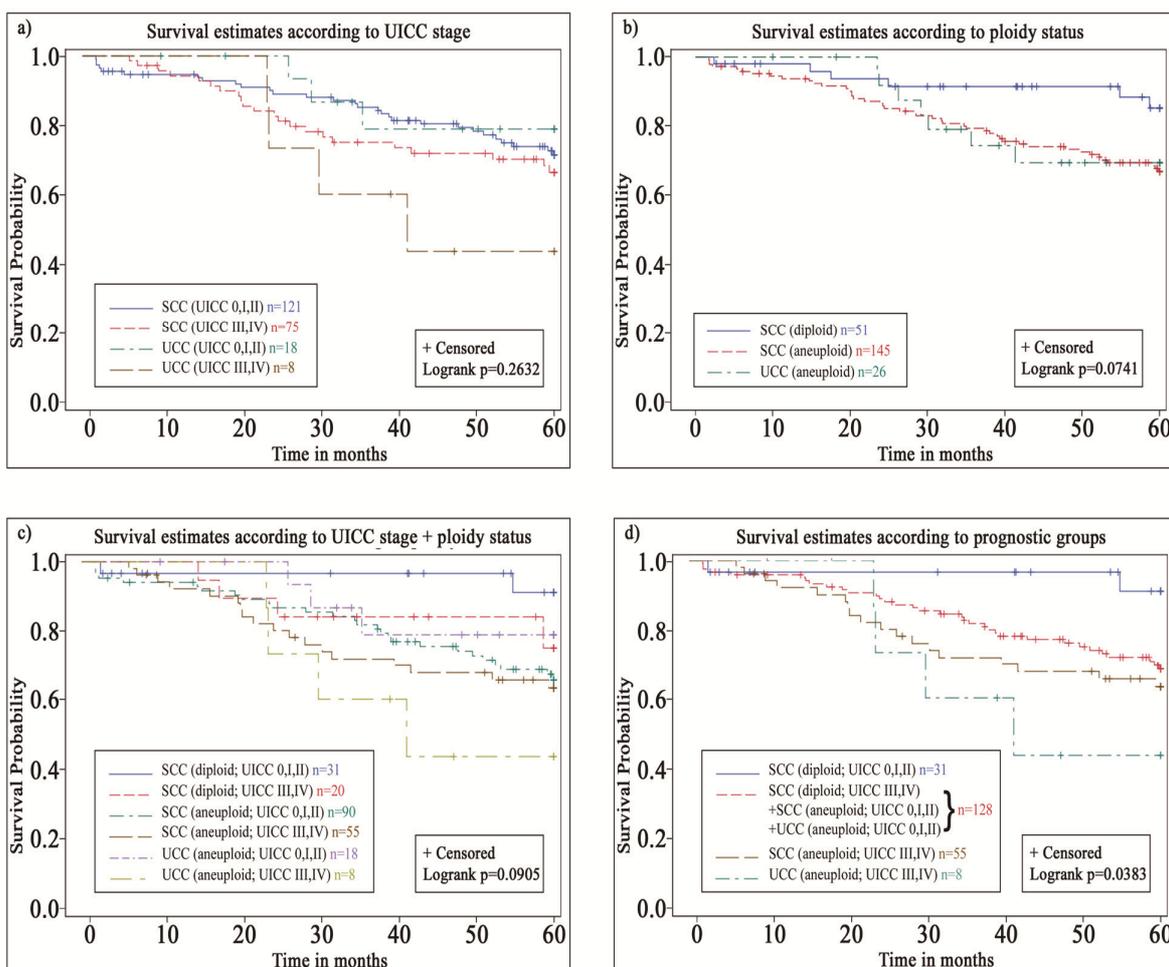


Figure 13: Kaplan-Meier plots

Plot **A**: Survival estimates of UCCs and SCCs

Plot **B**: Survival estimates of aneuploid and diploid CRCs irrespective of underlying colitis.

Plot **C**: Survival estimates of UCCs and aneuploid and diploid SCCs.

Plot **D**: Survival estimates of 1: SCC in early stages (UICC I and II); 2: SCC in advanced stages (UICC III); 3: UCC in early stages (UICC I and II); 4: UCC in advanced stages (UICC III).

4.2 Results from the Gene Expression Analyses in Colitis Mucosa

4.2.1 Ploidy Assessment in Premalignant Colitis Mucosa

DNA ploidy measurements were conducted on altogether 139 UC mucosa samples and nine carcinomas. For four out of 139 mucosa samples (2.88%) an insufficient amount of non-overlapping cell nuclei ($n < 100$) was available and measurements were considered not representative. Of the remaining 135 mucosa samples, in twelve samples (8.9%) ploidy status was rated type I according to Auer's classification, no sample was found to be type II, 68 samples (50.4%) were type III, and 55 samples (40.7%) were type IV. All carcinomas showed aneuploid cell populations (type IV).

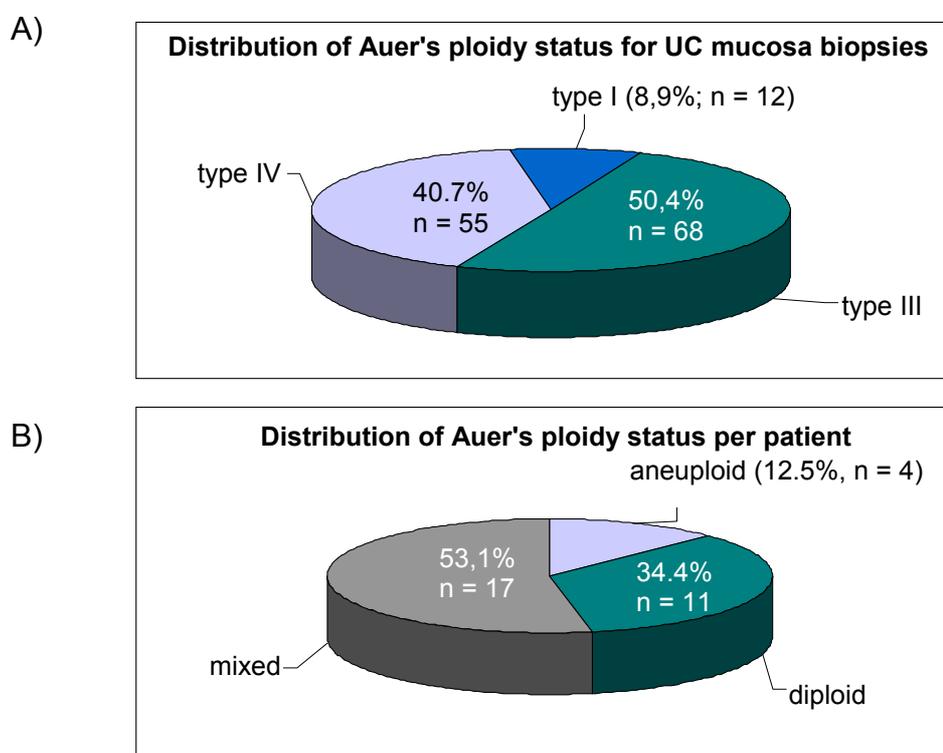


Figure 14: Distribution of ploidy types in colitis mucosa
Ploidy status in 135 mucosa biopsies (A) and 32 corresponding UC patients (B).

In eleven patients (34.4%), only diploid ploidy types (I, II, III) could be observed throughout all biopsies measured for each patient (corresponding to 26 mucosa biopsies), while in four patients (12.5%; 16 mucosa biopsies) all biopsies measured were aneuploid. The remaining 17 patients showed both aneuploid and diploid biopsies. Figure 14 illustrates the distribution of ploidy status for biopsies and patients.

4.2.2 Inter-Observer Reliability of Ploidy Assessment in Mucosa Samples

For ploidy assessment according to all four Auer categories, strong agreement could be observed ($K = 0.934$). For distinguishing between diploid (Auer's I, II, and III) and aneuploid (IV) biopsies $K = 0.944$. All nine UCCs were found aneuploid by all investigators ($K = 1$). All nine normal controls were diploid ($K = 1$).

4.2.3 Microarray Gene Expression Analyses

In the following, results from group comparisons of gene expression patterns are presented. Due to the large amount of data generated, only excerpts can be given in the appendix of this thesis, comprising the most interesting candidate genes (see Supplementary Tables and Figures, pp. 83ff). Figure 15 illustrates the most important group comparisons and presents the number of genes differentially expressed throughout the sequence of malignant transformation.

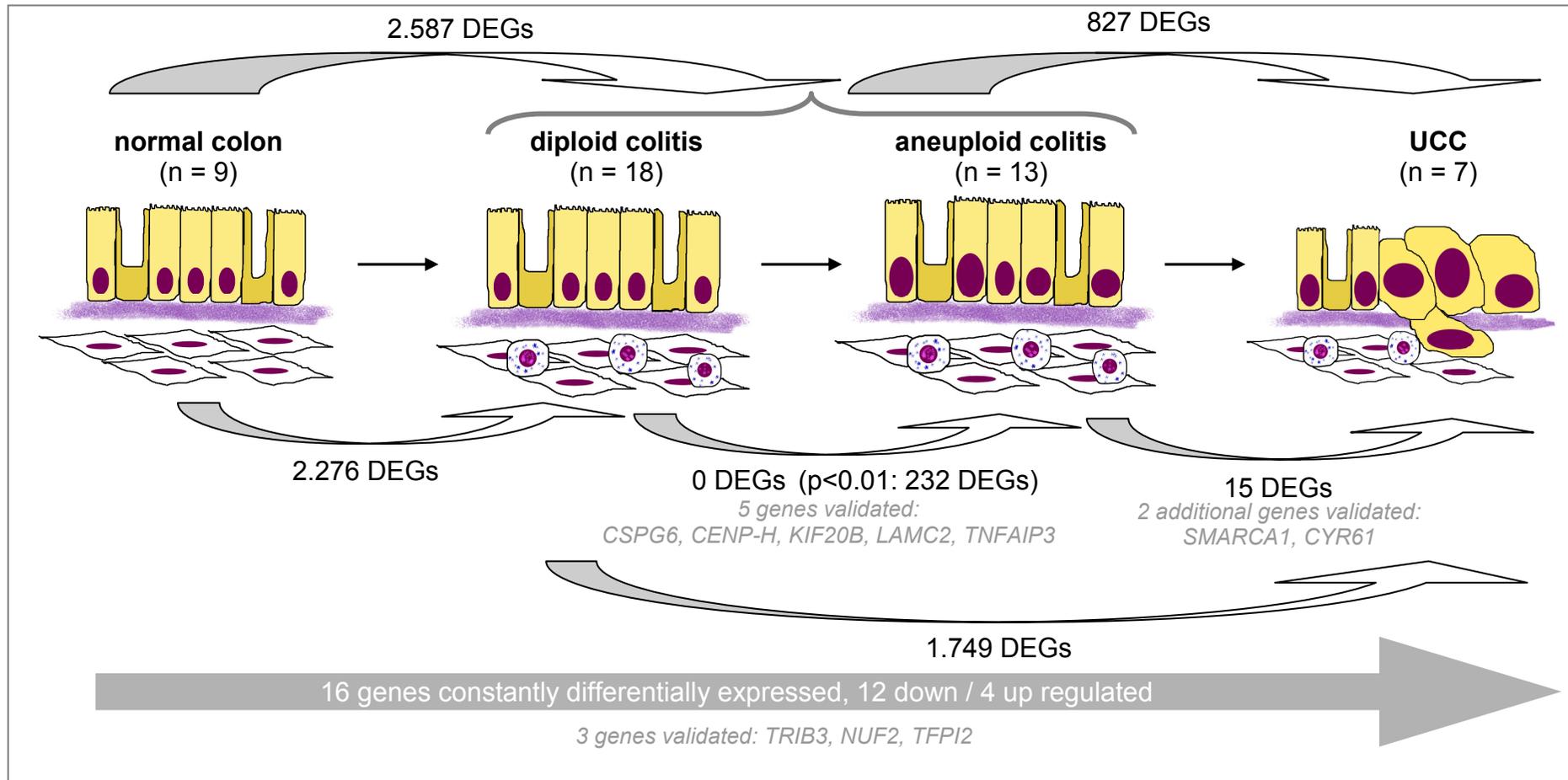


Figure 15: Illustration of DEGs among the sequence from uninfamed colon via diploid and aneuploid colitis to UCC. Only genes with a fold change > +/- 1.2 are presented. Below, genes constantly differentially expressed are presented (for details please refer to text). Genes for which RT-PCR was performed are given in italic where applicable

4.2.3.1 Differentially Expressed Genes between Normal Controls and UC Mucosa

Gene expression changes were compared between normal controls (n = 9) and patients with longstanding UC (n = 31) independent of the respective ploidy status. A total of 2,587 genes were significantly differentially expressed ($p < 0.01$ & $\text{fdr} < 0.1$ & ratios < 0.8 or > 1.2). Herein, 1,122 DEGs were lower expressed in UC-affected tissue, while 1,465 were higher expressed. This comparison yielded the highest number of DEGs of all group comparisons. IPA software recognized 1,441 genes (55.7%). Based on these IPA eligible genes, 25 networks with scores between 18 and 40 were found. The two highest ranked networks (scores 40 and 38) comprised 32 and 31 DEGs, respectively, that were associated with *Cell-To-Cell Signaling*, *DNA Replication*, *Recombination & Repair*, and *Cellular Assembly and Organization*. Associated canonical pathways were *p53 Signaling*, *DNA Methylation & Transcriptional Repression Signaling*, and *Mammalian Embryonic Stem Cell Pluripotency* ($p < 0.002$). Figure 16a presents the highest ranked network.

4.2.3.2 Differential Gene Expression of Ulcerative Colitis Mucosa and Colitis-associated Cancer

When comparing all UC biopsies (comprising aneuploid and diploid samples, n = 31) against UCCs (n = 7), 827 genes were found to be significantly differentially expressed ($p < 0.01$ & $\text{fdr} < 0.1$ & ratios < 0.8 or > 1.2). Herein, 399 genes were higher expressed in carcinomas and 428 genes were lower expressed. Pathway analysis detected 536 DEGs (64.8%) projecting 25 high-ranked networks (scores 13 to 35). The top two networks were associated with *Cell Cycle*, *Embryonic Development*, and *Cell Morphology*. DEGs were involved in the canonical pathways *Rac Signaling* and *Molecular Mechanisms of Cancer* ($p < 0.002$). Figure 16b illustrates the highest ranked network of this comparison.

4.2.3.3 Differentially Expressed Genes between Aneuploid and Diploid Mucosa

Oligonucleotide array analyses were performed on patients with only aneuploid or only diploid biopsies throughout the colorectum. Biopsies of patients presenting diploid and aneuploid biopsies simultaneously were excluded. When comparing clinical features of patients and samples included in this analysis, no significant differences in patients' sex ($p = 0.634$), age ($p = 0.096$), duration of colitis at time

of sampling ($p = 0.771$), degrees of dysplasia ($p = 0.179$) or inflammation ($p = 0.813$) or biopsy location ($p = 0.634$) became evident. However, DNA stemline ($p = 6 \times 10^{-5}$), 2.5c exceeding rate ($p = 5 \times 10^{-5}$), and 5c exceeding rate ($p = 0.006$) differed significantly among both groups. Nevertheless, gene expression analysis between the 13 aneuploid and 18 diploid colitis samples did not yield any DEGs according to thresholds applied previously ($p < 0.01$ & $fdr < 0.1$ & ratios < 0.8 or > 1.2). However, a total of 232 genes were differentially expressed using the uncorrected $p < 0.01$. Herein, 77 genes were lower expressed in aneuploid samples and 155 genes were higher expressed. Five of the DEGs of this comparison were validated with RT-qPCR in a subset of samples, for which mRNA was available after array hybridization (*SMC3*, *CENP-H*, *KIF20B*, *LAMC2*, *TNFAIP3*). PCR confirmed the trend of differential regulation for all five genes (figure 18a). In addition, *CENP-H* was significantly differentially expressed for both, array analysis using the uncorrected p -value < 0.01 and PCR validation ($p = 0.001$ and $p = 0.016$, respectively). Table 9 shows the five validated genes of this analysis.

Symbol	Description	fold change	p-value	fdr	GB_accession
CSPG6/SMC3	chondroitin sulfate proteoglycan 6 (bamacan)	2,11	0,00015	0,548	AF020043
CENP-H	centromere protein H	1,48	0,00015	0,548	BC015355
MPHOSPH1 /KIF20B	M-phase phosphoprotein 1	2,08	0,0094	0,548	AB033337
LAMC2	laminin, gamma 2	0,65	0,00228	0,548	-
TNFAIP3	tumor necrosis factor, alpha-induced protein 3	0,72	0,00928	0,548	M59465

Table 9: Validated genes derived from the comparison of diploid and aneuploid colitis mucosa; fold change refers to aneuploid over diploid calculations (< 1 indicates up regulation in aneuploid samples); „GB_accession“ refers to GenBank accession number of the respective gene.

When querying the IPA database with the set of DEGs derived from the comparison of diploid and aneuploid colitis mucosa, 179 out of 232 genes (77.2%) were recognized. Network analyses revealed nine significant networks (scores 8 to 36). The three most significant networks were associated with *Cell Mediated Immune Response*, *Cellular Movement*, *Immune Cell Trafficking*, *Infection Mechanism*, *Cell Morphology*, and *Cellular Compromise*. DEGs were involved in the canonical pathways *Lymphotoxin beta-Receptor Signaling*, and *CD40 Signaling* ($p < 0.002$). Figure 17b presents network 1 with overlaid canonical pathways.

4.2.3.4 Comparison of Diploid UC vs. UCC and Aneuploid UC vs. UCC

The comparison of diploid colitis samples with carcinomas detected 1,749 DEGs ($p < 0.01$ & $fdr < 0.1$ & ratios < 0.8 or > 1.2). Herein, 894 genes were higher expressed in carcinomas and 855 genes were lower expressed. A total of 786 genes were found overlapping between the comparisons *diploid mucosa vs. UCC* and *overall colitis mucosa vs. UCC*.

Differences in gene expression between aneuploid UC biopsies and UCCs revealed a total of 15 genes to be differentially expressed ($p < 0.01$ & $fdr < 0.1$ & ratios < 0.8 or > 1.2). These genes were *ABCC13*, *TIEG2*, *CASP9*, *FLJ31668*, *MBC2*, *SLC17A7*, *TRIM45*, *ODZ3*, *GBGT1*, *NXN*, *DKFZp761D221*, *Oculorhombin*, *DRD4*, *HRASLS3*, and one EST. Herein, 10 genes were higher expressed in UCCs and five genes were lower expressed. Of these genes, 14 were concomitantly found to be differentially expressed between UC mucosa overall and UCCs (except *GBGT1*). Additionally, based on the highest fold changes in the dataset, *SMARCA1* and *CYR61* were validated with RT-qPCR in a subset of samples. PCR confirmed the trend of differential regulation for both genes. In addition, *CYR61* was found to be significantly differentially expressed with both techniques showing decreased expression in carcinomas ($p < 0.05$; figure 18b). Table 10 presents the total of these 15 and two ($n = 17$) genes relevant for this analysis.

Symbol	Description	fold change	p-value	fdr	GB_accession
ABCC13	ATP-binding cassette, sub-family C (CFTR/MRP), member 13	0.42	1.86E-07	0.004	AY063515
TIEG2	Kruppel-like factor 11	0.59	3.87E-06	0.049	AF028008
CASP9	caspase 9, apoptosis-related cysteine protease	0.38	1.05E-05	0.054	U60521
Q96MZ3	-	1.51	9.44E-06	0.054	BC035889
NP_056107	likely ortholog of mouse membrane bound C2 domain containing protein	1.99	8.18E-06	0.054	AB018290
SLC17A7	solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member 7	0.46	1.62E-05	0.069	AB032436
-	-	0.45	2.43E-05	0.088	-
TRIM45	tripartite motif-containing 45	1.88	2.99E-05	0.095	AY669488
ODZ3	odz, odd Oz/ten-m homolog 3 (Drosophila)	0.28	4.94E-05	0.096	AK001336
GBGT1	globoside alpha-1,3-N-acetylgalactosaminyltransferase 1	0.55	3.46E-05	0.096	-
NXN	nucleoredoxin	0.59	3.90E-05	0.096	AK027451
NP_115667	hypothetical protein DKFZp761D221	0.66	5.31E-05	0.096	AL136561
-	Unknown EST	0.76	4.87E-05	0.096	-
DRD4	dopamine receptor D4	1.41	5.67E-05	0.096	L12398
HRASLS3	HRAS-like suppressor 3	3.27	5.60E-05	0.096	X92814
CYR61*	cysteine-rich, angiogenic inducer, 61	6.56	0.001819	0.281	Y11307
SMARCA1*	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1	5.31	0.062858	0.392	

Table 10: DEGs between aneuploid mucosa and UCC. Validated genes marked with asterisk (*). fold change refers to UCC over aneuploid mucosa calculations (<1 indicates up regulation in UCC); „GB_accession“ refers to GenBank accession number of the respective gene.

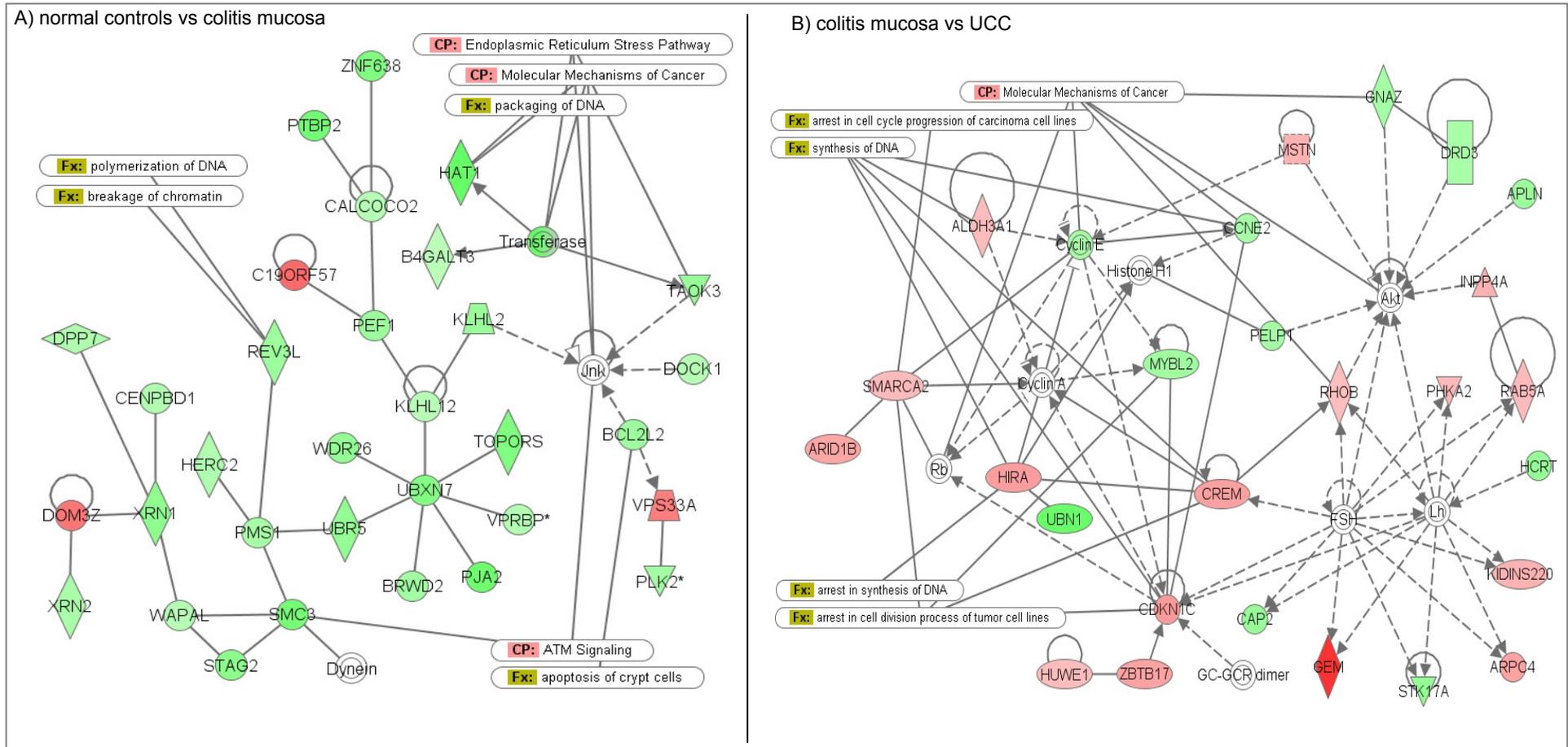


Figure 16: Highest ranked IPA networks of DEG between normal controls vs UC mucosa (A) and UC mucosa vs UCC (B)

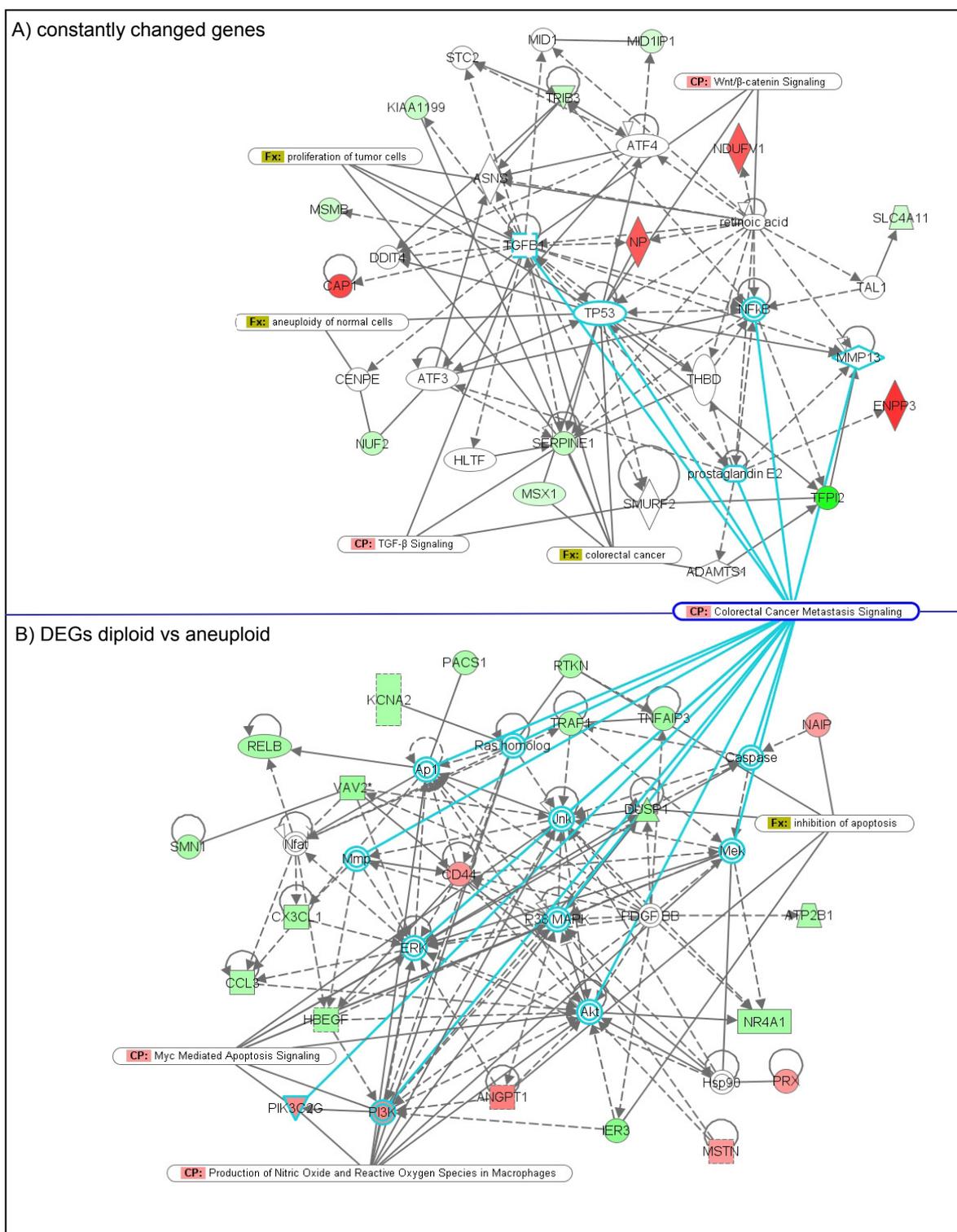


Figure 17: Top IPA networks of constantly changed DEGs and DEGs of diploid vs. aneuploid colitis

A: IPA network 1 (Score 35) of constantly up and down regulated genes across malignant transformation from normal colitis to UCC. Red highlighted genes were constantly up regulated over the sequence, green genes constantly down regulated. Please note *P53*, *TGFβ1* and *NFκB* as central nodes of this network, as well as canonical pathways (CPs) and common functions (Fx) of genes involved in wnt/ β -catenin signaling, CRC-metastasis signaling, cell proliferation, and aneuploidy.

B: Top rated network of IPA analysis diploid vs aneuploid. Please note genes involved in canonical pathways such as CRC-metastasis, apoptosis signaling, and production of reactive oxygen species. Remarkably, genes known to function in metastatic spread of colorectal carcinomas are found as central nodes in both analyses. Genes participating in this pathway are highlighted in light blue.

4.2.3.5 Comparison of Normal Colon vs. Diploid and vs. Aneuploid UC

Comparison between normal controls (n = 9) and diploid UC mucosa (n = 18) revealed 2,276 DEGs ($p < 0.01$ & $fdr < 0.1$ & ratios < 0.8 or > 1.2). Of these, 1,362 genes were higher expressed in diploid UC mucosa, 914 genes were lower expressed. A total of 1,867 genes were simultaneously found as DEGs in the overall comparison of *normal controls vs. colitis mucosa*.

When comparing normal controls with aneuploid UC mucosa, 2,326 DEGs showed differential expression ($p < 0.01$ & $fdr < 0.1$ & ratios < 0.8 or > 1.2). Herein, 1,216 genes were higher expressed in aneuploid UC mucosa and 1,110 genes were lower expressed. Ratios ranged from 0.13 to 3.88. A total of 1,830 genes were overlapping with the comparison *normal controls vs. all colitis mucosa* samples. Table 11 presents an overview of the number of significant genes for the aforementioned group comparisons as well as for additional group comparisons, which are not explicitly referred to in the text.

no	group 1	group 2	p-value			false discovery rate		
			significant genes	up*	down*	significant genes	up*	down*
1	normal	D&A	2587	1122	1465	3165	1294	1871
2	D	A	232	155	77	0	0	0
3	D&A	UCC	1503	744	759	827	399	428
4	A	UCC	676	356	320	15	5	10
5	normal	UCC	559	381	178	18	18	0
6	normal	D	2276	914	1362	2551	1005	1546
7	normal	A	2326	1110	1216	2377	1129	1248
8	D	UCC	2054	1000	1054	1749	855	894

Table 11: Number of DEGs in group comparisons

Second and third columns specify groups compared (normal: normal controls, D: diploid mucosa, A: aneuploid mucosa, D&A: all colitis mucosa specimens, UCC: UC-associated CRC). Thereafter the number of DEGs according to the uncorrected p-value is given, threshold $p < 0.01$. Subsequently, DEGs according to the fdr is presented ($fdr < 0.1$). *) For both, uncorrected p-value and fdr , the threshold for the fold change was set to ± 1.2 . "Up" regulated genes show higher expression in "group 1" over "group 2" and vice versa.

4.2.3.6 Constantly Up and Down Regulated Genes

Based on the linear model, altogether 16 genes were found: four genes with constant up regulation and 12 genes with constant down regulation during malignant progression (table 12). All of the 16 DEGs were recognized by the IPA knowledge base. Network analysis revealed three networks, the first one with a score of 35 involving 13 IPA eligible genes of interest (figure 17a). This network

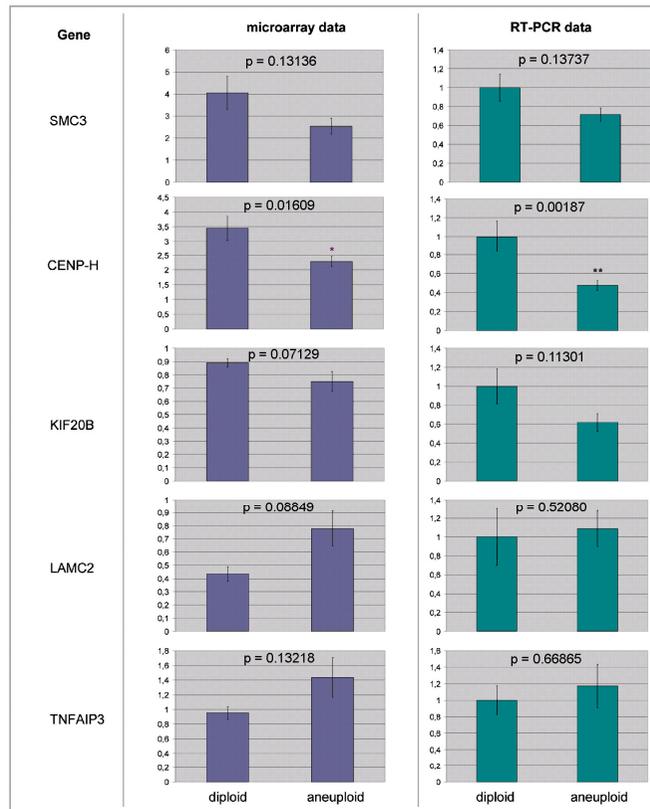
was associated with *Cellular Development* and *Cellular Growth and Proliferation*. Central nodes of this network were *TGF β* , *NF κ B* and *TP53*. Three of the 16 genes (*TRIB3*, *NUF2*, *TFPI2*) were validated with RT-qPCR showing similar trends of expression. In addition, *NUF2* was significantly differentially expressed with both techniques showing constant down regulation ($p = 0.021$ and $p = 0.030$; Figure 18c). Constantly changed genes and DEGs found between diploid and aneuploid samples shared *Colorectal Cancer Metastasis Signaling* as a common pathway.

Nr	Gene symbol	Name	Normal controls	Diploid colitis	Aneuploid colitis	UCC	IsSlope ¹
constantly down regulated							
1	<i>MSX1</i>	msh homeo box homolog 1	5.1293	4.1601	3.1764	1.4888	-1.1905
2	<i>TRIB3*</i>	tribbles homolog 3	9.4710	7.1342	6.6624	3.7760	-1.7557
3	<i>KIAA1199</i>	KIAA1199	6.8218	4.6015	3.7614	1.8066	-1.5886
4	<i>SERPINE1</i>	serine proteinase inhibitor, clade E, 1	9.1889	8.0568	7.1037	3.3610	-1.8437
5	<i>TFPI2*</i>	tissue factor pathway inhibitor 2	35.0618	25.5804	19.3078	14.999	-6.6459
6	<i>KIF18A</i>	kinesin family member 18A	7.2580	6.6963	4.7609	3.2719	-1.3894
7	<i>MSMB</i>	Microseminoprotein	6.7859	5.3939	4.0477	2.1870	-1.5143
8	<i>NUF2*</i>	cell division cycle associated 1	8.5800	7.2291	4.1056	3.3738	-1.8742
9	<i>MID1IP1</i>	MID1 interacting protein 1	5.4880	2.4226	2.2959	1.4611	-1.2207
10	<i>FBXL21</i>	F-box and leucine-rich repeat protein 21	11.6939	8.8064	3.6571	3.5668	-2.9531
11	<i>SLC4A11</i>	solute carrier 4	5.7088	3.2619	2.7208	1.2976	-1.3775
12	<i>MGC27005</i>	Hypothetical protein MGC27005	12.0696	8.7209	8.0800	3.2498	-2.7100
constantly up regulated genes							
13	<i>NP</i>	Purine nucleoside phosphorylase	1.5298	2.0541	2.1734	3.4713	0.5944
14	<i>CAP1</i>	Adenylyl cyclase-associated prot. 1	0.9776	1.0039	1.7089	2.9251	0.6547
15	<i>ENPP3</i>	Ectonucleotide pyrophosphatase/phosphodiester. 3	0.4824	0.5792	0.7493	2.8956	0.7410
16	<i>NDUFV1</i>	NADH-ubiquinone oxidoreductase	1.3737	1.8821	2.5857	3.1317	0.5978

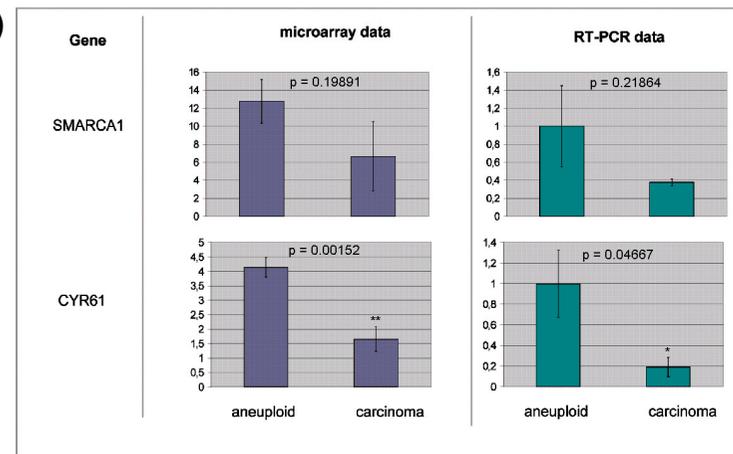
Table 12: Genes constantly down or up regulated

Constantly differentially expressed genes over the whole sequence from normal mucosa via diploid and aneuploid colitis biopsies to UC CRC; ¹IsSlope = Slope of curves fitted through data points by least squares method; asterisk after gene symbol indicates gene has been validated by RT-PCR

A)



B)



C)

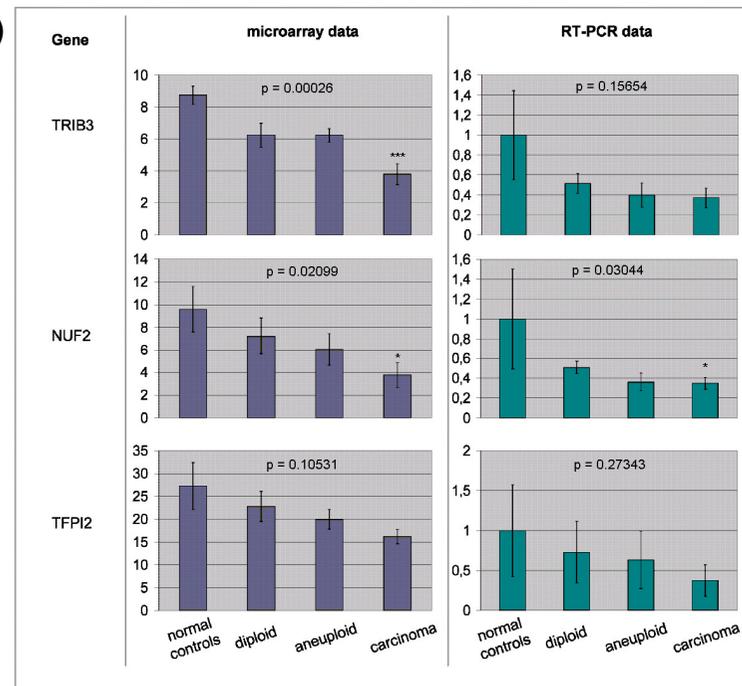


Figure 18: Linear array expression plots and PCR expression plots. PCR expression was normalized to HKG (*PGK1*). Error bars represent standard error of the means (SEM). **A)** DEGs between diploid (n = 7) and aneuploid (n = 7) mucosa biopsies; **B)** DEGs between aneuploid (n = 10) mucosa biopsies and carcinomas (n = 5); **C)** genes constantly differentially regulated among the sequence from normal controls (n = 9), via diploid (n = 8) and aneuploid (n = 11) mucosa to UCCs (n = 5); * / ** / *** indicate significant differences between reference group (the furthestmost left bars represent reference group) and the so-marked group with p < 0.05 / p < 0.01 / p < 0.001, respectively. To facilitate comparison of array and PCR results and inference from analysis, p-values are presented from both PCR and array data.

5 DISCUSSION

5.1 Overview of Experimental Results

This thesis sought to elucidate the role of CIN in UC-associated colorectal carcinogenesis. Using DNA-image cytometry, a high frequency of CIN reflected by aneuploidy could be elaborated, reaching 100% in our cohort of UCCs examined. Additionally, aneuploidy could be related to unfavorable outcome of patients with CRCs regardless of underlying UC. These results underscore the importance of aneuploidy for colitis-based carcinogenesis and prompted us to further investigate the molecular changes accompanying aneuploidization.

Gene expression patterns were examined in premalignant and malignant colitis mucosa based on presence of aneuploidy. Genes differentially expressed were identified and network analyses revealed numerous canonical pathways to be involved in colitis-associated carcinogenesis. Finally, a subset of ten genes could be validated with RT-qPCR.

5.1.1 Frequency of Aneuploidy in UCCs and SCCs

All 31 UCCs analyzed were observed to be aneuploid. It would have been desirable to detect also diploid UCCs in order to clearly distinguish between the impact of inflammation and aneuploidy for prognosis in UCC. However, the approach used allowed investigating the role of aneuploidy for the prognosis of patients that undergo surgical resection of CRCs among both, SCCs and UCCs, based on the hypothesis that aneuploidy might account for different clinical behavior of distinct tumor entities of the same organ. As for SCCs, ploidy measurements in this study revealed 74.6% to be aneuploid. This finding is in line with the literature as shown in the meta-analysis of 31 studies by Araujo *et al.*, in which the authors report percentages ranging from 36% to 89% of aneuploidy in CRCs⁵⁴. The difference in aneuploidy frequency between UCCs and SCCs was statistically significant (table 7). One advantage of the comparison performed in this study is based on the fact that all patients were operated on in one single centre and ploidy measurements were performed using one standardized technical approach. However, for the critical assessment of cytometry results slight differences in the cytometry technique have to be noted: in contrast to SCCs, in which ploidy was measured using imprints, in UCCs ploidy measurements were

performed on histological sections. Yet, comparative studies of DNA cytometry based on paraffin embedded specimens and imprints have shown that both techniques yield comparable results⁹⁸.

In summary, our results provide evidence that the CIN pathway of malignant transformation is dominant in UC carcinogenesis. However, due to the general sparsity of UCCs, multicentric studies based on a single and standardized method of ploidy assessment are necessary to detect whether differences in ploidy status are indeed as pronounced as suggested by the results presented herein.

5.1.2 Inter-observer Variability of Ploidy Assessment

Aneuploidy was assessed by DNA-image cytometry for 291 CRCs and 139 UC mucosa biopsies. In all instances, Kappa statistics showed a strong agreement for ploidy assessment ($K > 0.90$). In contrast, dysplasia assessment in UC biopsy samples is impeded by a high degree of inter-observer variability: Eaden *et al.* reported the Kappa statistics for inter-observer variability in the assessment of dysplasia in UC to be $K = 0.30$. In only 7.8% of 51 cases, all 13 pathologists participating in the study came to exactly the same dysplasia rating⁹⁹.

Our results suggest that objectivity of ploidy status assessment is superior to that of dysplasia assessment. It must be taken into account that all investigators in this study were trained in one center and measurements were performed using one cytometry device. Therefore, it remains uncertain to what extent using different cytometry systems by differently trained raters would decrease performance. Nevertheless, criteria for Auer's categories are well defined and it can thus be expected that inter-observer variability of ploidy assessment will remain superior to that of dysplasia.

5.2 Comparative Analyses of UCCs and SCCs

5.2.1 Clinical Correlations (Univariate Analyses)

5.2.1.1 Age and Sex

UCCs occur at younger age and are more often found in male patients than in females³⁶. These observations could be confirmed: Mean age at diagnosis of UCCs was 49.3 years, while SCCs were detected at a mean age of 69.7 years. It has to be taken into account that for the specific aims of this study, patients younger than 50 years were excluded in the SCC group to avoid including HNPCC

patients. However, previous studies have reported the mean age at diagnosis of SCCs to range between 61 and 74 years¹⁰⁰⁻¹⁰². We therefore consider our cohort representative.

The reasons for the male predominance among the UCCs and the association of male sex and aneuploidy remain ill-understood. One could speculate that dietary as well as genetic factors and hormonal influences might play a role³⁶.

5.2.1.2 Synchronous Carcinomas

A higher prevalence of synchronous carcinomas in UC patients compared to patients with SCCs as shown in our study has been reported previously¹⁰³. One possible explanation for this finding might be that in UCC patients not only the primary carcinoma itself, but also the non-cancerous mucosa throughout the entire colorectum presents CIN⁴. This stands in contrast to SCCs, in which the adjacent normal mucosa seldom shows aneuploid cells¹⁰⁴. Special attention should be given to the detection of synchronous carcinomas in UC patients. Pancolonoscopy screening prior to operation planning seems pivotal for UC patients in particular, moreover, pancolectomy, as routinely performed for UCC patients, could be highly recommended.

5.2.1.3 Tumor Staging

The frequency of distant metastases was comparable for UCCs and SCCs, while it was significantly higher in aneuploid SCCs as compared to their diploid counterpart. T- and N-status were equally distributed among all groups.

Our results do not legitimate the assertion that aneuploidy *per se* leads to a higher frequency of distant metastases, since UCCs and diploid SCCs show a similar frequency of M1 stages. Yet, it has been described for several malignant tumors that aneuploidy is associated with a higher prevalence of distant metastases^{53, 105}. UC patients in our study underwent an intensive surveillance program and therefore, cancers are likely to be discovered at early stages. This hypothesis is in line with a meta-analysis of the Cochrane Collaboration, which concluded that lead-time bias could contribute substantially to the apparent benefit of cancer surveillance in UC⁴³. Metastases at early stages may be difficult to diagnose and might be indiscernible as micro-metastases. However, based on this hypothesis, one would also expect to detect more frequently early stage tumors in UC patients, which could not be shown in our patient cohort.

5.2.1.4 Tumor Grading

UCCs are less differentiated than SCCs, regardless of ploidy status. Interestingly, a higher degree of dedifferentiation was associated with the mucinous subtype. Since these subtypes comprise only a small subset of CRCs, large sample cohorts will be needed to validate this finding.

5.2.2 Survival Analyses

Despite the overall significant differences between SCCs and UCCs regarding sex, synchronous malignancy, histology, and grading, patients with aneuploid carcinomas, being sporadic or UC-associated, show highly similar survival rates, while patients with diploid SCCs are distinct from both groups presenting better prognosis (figure 13b). Interestingly, diploid tumors at advanced stages (UICC stage III/IV) present similar survival rates as compared to aneuploid carcinomas at early stages (figure 13c&d). This finding might point to the conclusion that the presence of aneuploid tumor cell populations influences patients' prognoses more dominantly than does tumor stage. This is in line with the finding that the most pronounced difference in prognosis can be observed between diploid SCCs at early stages and aneuploid CRCs at advanced stages. Furthermore, UCCs at advanced stages show a prognosis inferior to that of their sporadic counterpart at the same tumor stage (figure 13d). This could lead to the assumption that UC is associated with inferior survival in these patients. Colitis could therefore be considered a poor prognostic factor in addition to aneuploidy and advanced tumor stage. In contrast, regression analysis did not yield inflammatory disease as a significant influencing prognostic factor. However, the group of advanced stage UCCs in survival analyses was comprised of only eight patients and is thus underrepresented in this particular analysis.

5.2.3 Logistic Regression (Multivariate Analysis)

Age at time of diagnosis and tumor ploidy status proved to be of significance for survival five years subsequent to operation. In our study, it was not distinguished between cancer-associated deaths and deaths from other causes. Thus, it could be expected that death among elderly patients is more likely to occur within the defined observation period. Consequently, inferior survival of UCC patients being younger than SCC patients must be appraised even more highly. The regression model applied to our data set suggests that aneuploidy *per se*, but not an

underlying colitis determines the five-year survival prognosis of patients. With a point estimate of 4.07 for aneuploid tumors, aneuploidy seems to be the dominating variable in our logistic regression model. One of the most interesting and challenging tasks will be to elucidate the role of inflammatory activity in the development of CIN ¹⁰⁶.

5.3 Gene Expression Analyses of Malignant Transformation in UC

The high frequency of aneuploidy in UCCs and the fact that aneuploid colonic epithelial cells can be detected years prior to malignant transformation render UC a suitable model for the study of aneuploidization in human carcinogenesis *in vivo*. UC specifically has several advantages for the study of aneuploidization:

- 1.) the distribution of aneuploidy in UC has been well described in the literature
- 2.) aneuploidy can be easily assessed based on biopsy specimens that are routinely available from surveillance colonoscopy
- 3.) UCCs present a high frequency of aneuploidy as demonstrated in this thesis, underscoring the importance of CIN specifically in this disease.

5.3.1 Transcriptomic Changes during Aneuploidization

5.3.1.1 Deregulation of Gene Expression after UC Initiation

In the comparison of normal controls and UC mucosa, a massive deregulation of gene expression could be observed (table 11). This result could be expected from the strong background of inflammatory activity and has been described previously ¹⁰⁷. DEGs participate in a variety of functions: IPA analysis yielded diverse significant networks, among others involved in *DNA Replication, Recombination & Repair*. Interestingly, the highest ranked network of the DEGs indicated replicative and endoplasmic reticulum stress to be dominant in colitis afflicted mucosa (figure 16a). This finding puts further emphasis on the hypothesis that genomic stress is imposed by chronic inflammation ¹⁰⁸ and might facilitate the ability of unlimited growth and proliferation by rearrangements of chromosomal material ¹⁰⁹.

5.3.1.2 Subtle Gene Expression Changes between Diploid and Aneuploid Mucosa

The problem of multiple testing when determining subtle gene expression changes in a small proportion of all hypotheses tested is increasingly recognized: Low statistical power to detect genes that are truly differentially expressed is a result of

multiple test adjustment ¹¹⁰. Omitting multiple test statistics leads to a high number of false positive findings, i.e. genes that are found to be differentially regulated, but indeed are not changed. Yet, weighing the options of not finding any interesting candidate genes or taking the risk of validating false positives, the latter approach seems more promising ¹¹¹. Finally, a subset of five DEGs derived from the comparison of diploid and aneuploid mucosa could be validated using RT-PCR as a complementary technique, acknowledging the trend of differential regulation for all validated genes and reaching statistical significance for one gene (*CENP-H*). Successful validation thus yielded additional confidence in our array results.

IPA analysis resulted in nine significant networks for the gene signature of this comparison. It goes beyond the scope of this study to discuss all canonical pathways found. Yet, the variety of significant pathways and networks involved suggests that aneuploidization causes a magnitude of gene expression changes. The top ranked network (figure 17) contained members of apoptosis signaling (Caspase, *TNFAIP3*, *TRAF1*, *NAIP*, *IER3*) and other genes associated with CRC-metastasis (*Ap1*, *MMP*, *p38*, *Akt*, *CASP*, *Jnk*, *RASH*, *ERK*, *Mek*), and production of RONS in macrophages (*PI3K*, *Akt*, *MekAp1*, *ERK*). We were able to identify the connection of two focus genes involved in RONS- as well as cancer metastases-pathways: In our dataset *PIK3C2G* (also referred to as *PIK3*) and *PI3K*, both members of the phosphoinositide 3-kinase family of genes, proved to be associated with both important pathways, giving further evidence that RONS, which are released abundantly in inflamed tissues might contribute to gross structural DNA aberrations, as described previously ¹¹². Interestingly the third highest ranked network contains *CENP-I*, a gene involved in kinetochore assembly, which is currently of major interest for researchers focusing on aneuploidization ⁶⁰. The kinetochore assembly pathway will be discussed in detail below (pages 51ff).

While groups for this comparison were defined by ploidy status, it cannot be excluded that other factors contribute to differential gene expression. However, neither inflammatory activity, dysplasia, age, gender nor biopsy localization differed significantly among both groups. Factors that were not controlled in our dataset were, e.g., diet and medication at time of harvest. Especially the latter factor might contribute to DEGs. While medications at time of harvest were recorded, numerous different drugs were taken by patients, thereby impeding

statistical analysis of putative influences. Especially corticosteroids, which are given regularly to UC patients are known to alter gene expression ¹¹³. However, one study conducted using microarray technique was not able to detect DEGs between Crohn's patients with and without corticosteroid treatment ¹¹⁴.

5.3.1.3 *Deregulation of Gene Expression in Malignant Tissue*

While between diploid mucosa and UCCs 1,749 DEGs were found, this was the case for only 15 genes between aneuploid mucosa and UCCs. IPA analysis of the 1,749 genes revealed a high number of networks associated with molecular mechanisms of cell division and DNA synthesis (figure 17), underscoring the importance of genes involved in DNA replication and mitosis for cancer development in UC. Conceivably, DEGs between aneuploid mucosa and UCCs could be interesting candidate genes for the determining step in carcinogenesis and thus reveal pivotal pathways for malignant transformation. On the other hand, as suggested by our results, crucial events for aneuploidization might occur early during carcinogenesis and accumulate with increasing CIN. Therefore, the last steps from aneuploid mucosa to UCCs might be diverse and random, while the main damage that paves the ground for malignant transformation takes place earlier.

5.3.1.4 *Constantly Changed Genes during Malignant Transformation*

In total, 16 DEGs could be identified over the sequence from normal via diploid and aneuploid mucosa to UCC. 13 genes were found to be part of the most significant IPA network. Interestingly, the top networks of DEGs derived from the comparison of diploid and aneuploid mucosa and the network drawn by genes constantly differentially regulated share a common pathway with *CRC metastasis signaling* (figure 17). Affected canonical pathways by genes of this network were *proliferation of tumor cells*, *TGF- β* , and *Wnt/ β -catenin*-signaling, all involved in colorectal carcinogenesis. For the wnt-pathway, early activation has been reported in UCCs, which is supported by our results ¹¹⁵.

The approach using the linear model could identify new genes of interest that had not been previously associated with UC carcinogenesis (*MSX1*, *TRIB3*, *KIAA1199*, *KIF18A*, *MSMB*, *MID1IP1*, *FBXL21*, *SLC4A11*, *MGC27005*, *NP*, *CAP1*, *ENPP3*, *NDUFV1*). *SERPINE1*, a gene encoding for a molecule involved in the plasminogen activator system, which was constantly down regulated during the

sequence, is known to be differentially regulated in IBDs ¹¹⁶. Furthermore, its role in cancer development is being intensively investigated, but no reports exist that associate *SERPINE1* with CIN. Interestingly, in the majority of published data, *SERPINE1* up regulation is associated with tumor development and positively correlated to tumor aggressiveness, which is in contrast to our finding ¹¹⁷. However, RT-PCR validation showed comparable expression values confirming the validity of our gene expression data.

The involvement of *NUF*, *TFPI2*, and *SMARCA1*, which were all constantly down regulated, is discussed in the following section.

5.4 Genes Validated with RT-qPCR

5.4.1 RT-qPCR Validation

PCR could be performed on only a limited number of samples owing to small biopsy size. For all ten genes tested, the general trend of differential expression could be confirmed (figure 18). For three genes, differences of expression were significant in PCR analyses. For nine genes, PCR and array p-values were congruent when assessed for equivalent samples. Thus, as the comparison of t-test statistics from PCR and array data suggests, with the exception of *TRIB3*, PCR and array data are congruent ($p < 0.05$) confirming array results. Figure 19 illustrates the cell cycle associated genes used for technical validation with PCR and puts them into the context of proliferation, DNA replication, and cell division. Please note that in the following gene and protein nomenclature will follow the Human Gene Nomenclature (HGNC, www.genenames.org, last accessed February 2011). Therefore, gene and mRNA symbols are written in italics, protein symbols are not italicised.

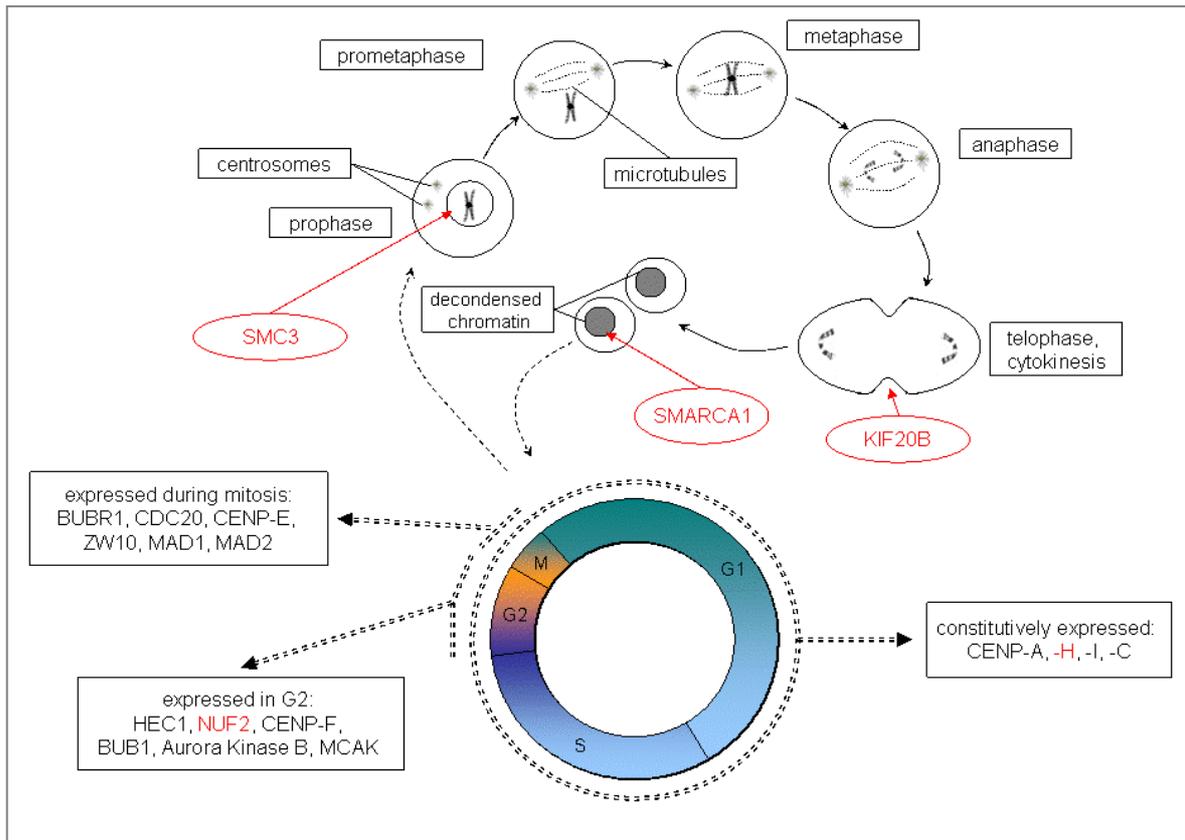


Figure 19: The cell cycle and associated validated genes

Upper part magnifies steps during mitosis and highlights validated genes associated with mitotic dysfunction in red. *SMC3* is essential for chromatid cohesion and aberrant expression might cause chromosome segregation errors. *KIF20B* is pivotal for the last step, cytokinesis. *SMARCA1* influences transcription in decondensed chromatin through nucleosome repositioning. Lower part of the image shows cell cycle progression from mitosis through G1/G2 phases onwards to G2 (G = “gap”). Kinetochores are given in boxes indicating their expression during respective phases of the cycle. *NUF2* and *CENP-H* are found to be differentially expressed in our data set and highlighted in red. (Kinetochores expression alongside cell cycle adopted from Liu *et al.*: “Mapping the assembly pathways that specify formation of the trilaminar kinetochore plates in human cells”; The Journal of Cell Biology, Vol. 175, No. 1, October 9, 2006 41–53¹¹⁸).

5.4.2 Genes Differentially Regulated between Diploid and Aneuploid Mucosa

5.4.2.1 *SMC3* / *CSPG6*

In our dataset, *SMC3* (Structural Maintenance of Chromosome, No. 3) was significantly down regulated between diploid and aneuploid UC mucosa and associated with one of the lowest p-values in this comparison overall. RT-qPCR confirmed down regulation in a subset of samples.

SMC3 encodes a protein, which is pivotal for sister chromatid cohesion. During mitosis, DNA first condensates into distinct chromosomes and subsequently, sister chromatids are to be divided in two halves during anaphase¹¹⁹. To make these

specific steps possible, condensation and cohesion of sister chromatids are facilitated by protein complexes, the so called condensins and cohesins¹²⁰. SMC3 belongs to a family of chromosomal ATPases involved in this process¹²⁰. Within conserved domains, nucleotide-binding motifs have been identified, connected by a “hinge” region that allows formation of homo- and heterodimers¹²⁰. SMC3 and the interacting proteins SMC1 and SCC1 have been demonstrated to form a triangular “ring” that traps sister chromatids. Enzymatic cleavage of SMC3 causes the release of the cohesin complex from sister chromatids and suffices to destroy chromatid cohesion completely¹²¹. Interestingly, the cohesin complex is also involved in DNA repair: SMC1 and SMC3 have been shown to be recruited to DNA double strand breaks (DSBs) and it has been suggested that the cohesin complex facilitates DNA repair by holding sister chromatids together locally at DSBs, thereby allowing exchange with the sister chromatid repair template during homologous recombination¹²². From the above, it could be speculated that impaired function of cohesin complex proteins might lead to impaired distribution of sister chromatids during mitosis and cause aneuploidy. Indeed, *SMC3* knockdown using RNA interference (RNAi) was performed previously and high degrees of CIN in the so altered cells were observed¹²³. Specific mutations in the *SMC3* gene were described in a small cohort of human CRCs¹²⁴. Contrarily, in other studies increased levels of *SMC3* were observed in CRCs and overexpression of *SMC3* was found to transform human cell lines and mouse fibroblasts *in vitro*¹²⁵.

Our results provide evidence derived from an *in vivo* model of aneuploidization that *SMC3* might play an important role in malignant transformation and CIN development. They further emphasize the involvement of *SMC3* in premalignant lesions with genetic instability, thereby suggesting that impaired *SMC3* expression occurs early during malignant transformation. Interestingly, gene expression levels were almost similar for aneuploid mucosa biopsies and UCCs in our patient cohort (fold change 0.904 [aneuploid over UC CRC]), indicating that possible genetic damage caused by impaired *SMC3* function occurs in the premalignant stage and is not aggravated in carcinomas.

5.4.2.2 *CENP-H*

In our analyses, *CENP-H* expression was down regulated in aneuploid mucosa as compared to diploid mucosa. RT-qPCR confirmed the trend of differential regulation. Similar to *SMC3*, expression between aneuploid mucosa and UCCs was not significantly different for *CENP-H* (fold change 1.041).

The gene product encoded by the *CENP-H* gene belongs to the family of centromere associated proteins¹²⁶. CENPs (*CEN*tromere *P*roteins) are evolutionary conserved molecules that are essential for kinetochore formation and chromosome segregation¹²⁷. Centromeric DNA is primarily characterized by CENP-A accumulation¹²⁷. Distally from CENP-A, numerous CENPs accumulate, each involved in kinetochore formation and function¹²⁸. CENP-H has been found to be important for the localization of newly synthesized *CENP-A* to centromeric regions¹²⁹. RNAi knockdown of *CENP-H* in human HEP-2 cells showed an abundance of misaligned chromosomes¹³⁰. In contrast, *CENP-H* over expression is likewise associated with CIN, suggesting that stoichiometric expression of the gene is essential for its function¹³¹. Over expression might sterically prevent cofactors, which yet have to be elucidated, to recruit CENP-H to the kinetochore. Therefore, up and down regulation, respectively, might lead to kinetochore dysfunction and cause aneuploidy. Recently, a new mouse model of aneuploidization involving a CENP gene was introduced by Weaver *et al*: Mice bearing depletion of one *CENP-E* allele showed aneuploid cell populations and were more likely to develop lymphomas of the spleen and lung adenomas⁶⁰. Our results indicate that down regulation of *CENP-H* possibly plays a role in colitis-associated carcinogenesis and warrants further functional analysis.

5.4.2.3 *KIF20B*

The present study is the first to show differential expression of *KIF20B* in aneuploid premalignant UC mucosa, while no significantly different expression could be observed between aneuploid samples and UCCs (fold change 1.047). These results indicate that *KIF20B* might be involved in initiation of aneuploidy. It is conceivable that *KIF20B* down regulation leads to aneuploid cells as shown in cell culture experiments^{132, 133}. *KIF20B* (also known as MPHOSPH1; *M*-phase *phosphoprotein 1*) is a kinesin-like protein involved in mitosis¹³⁴. Overexpression of *KIF20B* causes cells to arrest at G2-M-phase *in vitro*¹³⁴. Interestingly,

knockdown of *KIF20B* induces apoptosis and defects in cytokinesis, rendering *KIF20B* an important protein for completion of mitosis, but not for its initiation (see also figure 19) ¹³². RNAi knockdown of *KIF20B* leads to “a significant increase of multinuclear cells and subsequent cell death of bladder cancer cells” ¹³³. As described in human bladder cancer, most of these aneuploid cells are not likely to be viable and subsequently undergo apoptosis. Concurrent deficiency to initiate apoptosis (please refer to *TRIB3* for one possible example) might, however, give some cells the ability to perform further cell divisions and eventually lead to the evolution of a viable aneuploid cell clone.

5.4.2.4 *LAMC2*

LAMC2 expression was increased in aneuploid samples. No significant difference could be observed between aneuploid samples and UCCs or in the comparison of normal controls and UC tissue. The *LAMC2* gene encodes for laminin-5- γ -2, which belongs to a family of glycoproteins that are major constituents of the basement membrane. By different composition of five α , four β , and three γ chains, 16 known isoforms of laminins exist ¹³⁵. Laminins provide a molecular anchor for cells to attach to the basement membrane and thereby facilitate cellular migration. Laminin-5- γ -2, in particular, has been associated with cancer progression and inferior prognosis: it is over expressed at the invasive front of CRCs and its expression levels at the infiltrating tumor edge correlate with tumor aggressiveness ¹³⁶. Our workgroup has previously shown that *LAMC2* expression is increased on the protein level in aneuploid UC mucosa ⁴. Figure 20 illustrates the finding that tumor cells expressing laminin facilitate tumor invasion.

The herein presented results indicate that *LAMC2* up regulation occurs relatively early during malignancy development in UC. Interestingly, it has recently been shown that *LAMC2* expression correlates with genomic instability: In a small cohort of MSI CRCs, a significantly lower *LAMC2* expression was observed as compared to carcinomas characterized by CIN ¹³⁷. Thus, *LAMC2* overexpression might indeed be attributable to CIN, which should be evaluated in further studies.

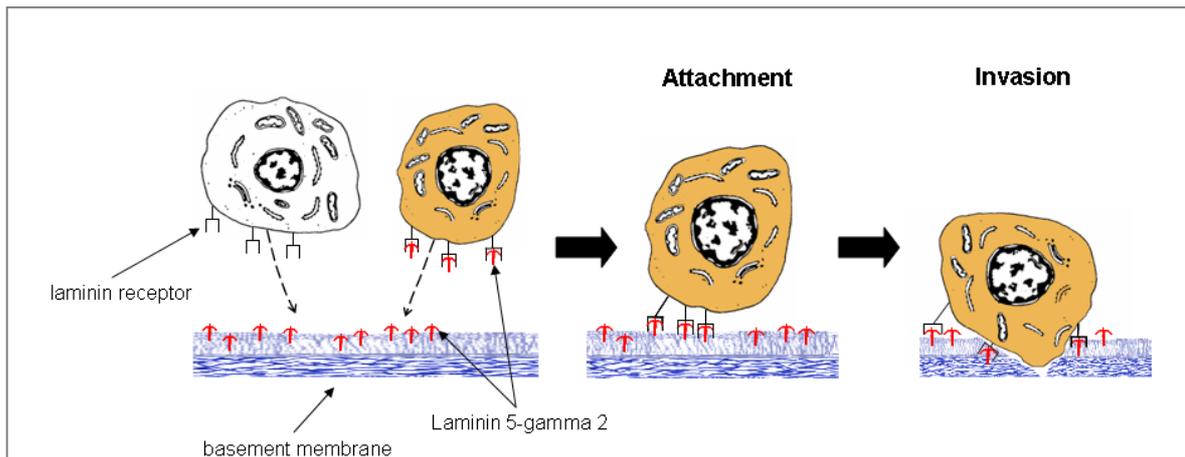


Figure 20: Laminin expression on tumor cells

Normal cells (white) express laminin receptors. Malignant cells (orange) additionally excrete laminins at the invasive tumor front^{138, 139}. Laminins build the bridge between tumor cells and the basement membrane, mediating penetration through the basement membrane and dissemination.

5.4.2.5 *TNFAIP3* / A20

TNFAIP3 was found up regulated in aneuploid mucosa. *TNFAIP3* (synonym A20) is a zinc finger protein and a potent inhibitor of NF κ B signaling¹⁴⁰. It functions as a negative regulatory feedback inhibitor: Its up regulation upon NF κ B activation terminates NF κ B signaling¹⁴¹. Lack of *TNFAIP3* leads to prolonged NF κ B activation and causes sustained inflammatory responses and cachexia in mice¹⁴¹. *TNFAIP3* has ubiquitinating and de-ubiquitinating activity, thereby adding or removing ubiquitin chains from RIP (receptor interacting protein)¹⁴². RIP is a pivotal mediator of the TNFR1 (TNF receptor 1) complex, a canonical pathway involved in NF κ B activation¹⁴². Thus, by influencing proteasome degradation of RIP, *TNFAIP3* has the potential to regulate NF κ B activity. *TNFAIP3*^{-/-} mice were found to develop colitis at a young age due to the inability to terminate NF κ B driven inflammatory response¹⁴¹. Interestingly, *TNFAIP3*^{-/-} cells die after exposure to tumor necrosis factor (TNF), while *TNFAIP3*^{+/+} cells can survive the same treatment, demonstrating that *TNFAIP3* can protect cells from TNF-induced cell death¹⁴¹. Clinically, *TNFAIP3* polymorphisms have recently been associated with type I diabetes, rheumatoid arthritis and systemic lupus erythematoses^{143, 144}. With regard to its important role in inflammation, up regulation of *TNFAIP3* in aneuploid mucosa as shown in this thesis could be interpreted as an alteration in inflammatory activity. However, the degree of inflammation was not significantly different between aneuploid and diploid specimens, suggesting a different

explanation. Interestingly, *NFκB* expression was not significantly different between the two groups (fold change 0.947; data presented for *NFKB1*). Contrarily, expression of *TRAF1* (*TNF receptor-associated factor 1*), encoding for a protein that interacts with TNFAIP3 was significantly increased in aneuploid mucosa (fold change 0.829) providing further evidence for the involvement of the TNFAIP3/TRAF pathway in aneuploid mucosa. *RIP1* expression could not be evaluated in array expression data, since it was not included in the genes that passed the quality check.

Notably, significant differences in *TNFAIP3* expression were found neither between normal controls and UC mucosa, nor between aneuploid mucosa and UCCs. To the knowledge of the author, no evidence exists that links *TNFAIP3* to aneuploidization or that suggests its overexpression to be a characteristic feature of aneuploid cell populations.

5.4.3 Genes Differentially Regulated between Aneuploid Mucosa and UCC

5.4.3.1 *CYR61 / CCN1*

In our study we demonstrate down regulation of *CYR61* in UCC as compared to aneuploid mucosa. Expression of *CYR61* was not significantly different between diploid and aneuploid mucosa biopsies, indicating that alteration of *CYR61* expression occurs relatively late during carcinogenesis.

CYR61 (*Cysteine-rich protein 61*) belongs to the CCN gene family (*connective tissue growth factor / cysteine-rich 61 / neuroblastoma overexpressed*), which comprises genes that are involved in numerous cell functions such as mitosis, cell-adhesion, apoptosis and growth arrest, as well as extracellular matrix production¹⁴⁵. *CYR61* (or *CCN1*) also contributes to angiogenesis in malignant tumors¹⁴⁶. It has an integrin binding domain and is involved in intra-cellular signaling of several cancer-associated pathways (e.g. *TGFβ*, *IGF*, and *VEGF*)¹⁴⁵. Altered *CYR61* expression could be correlated to outcome in different malignant diseases: *CYR61* mRNA expression was decreased in 74 out of 94 lung tumors (78.7%) and decreased expression was associated with advanced disease¹⁴⁷. In gastric carcinomas, *CCN1* protein expression was detected in 43 out of 49 early gastric carcinomas (87.7%), while only 19 out of 79 advanced gastric carcinomas (24.4%) showed detectable *CYR61* levels¹⁴⁸. On the contrary, increased expression was observed in prostate cancer and breast cancer^{149, 150}. The contribution of CCN

proteins to malignant transformation might be achieved through several of their various functions, i.e. anti-apoptotic, growth stimulating and angiogenesis-inducing features of *CYR61* and warrants further functional studies.

5.4.3.2 *SMARCA1*

SMARCA1 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1) encodes for a gene involved in the SWI/SNF signaling pathway. *SMARCA1* has helicase and ATPase activity and regulates gene transcription by altering chromatin structures¹⁵¹. However, only about 5% of all genes of *Saccharomyces cerevisiae*, in which SWI/SNF function is well-characterized, are dependent on the presence of SWI/SNF¹⁵². *SMARCA1* seems to be a target gene of *MT-MMP1* (matrix metalloproteinase 1), indicating a role in cancer progression¹⁵³. *SMARCA1* expression was significantly down regulated in UCCs (fold change 5.895). Similar to *CYR61*, differential expression of *SMARCA1* failed multiple test statistics. However, this gene was chosen for PCR validation based on the high fold change observed. In our data set down regulation of *SMARCA1* did not occur from diploid to aneuploid mucosa but during the transition from aneuploid mucosa to cancer. As suggested by the general role of SWI/SNF genes, depletion of *SMARCA1* expression might lead to altered expression of multiple genes, which could involve oncogenes and tumor suppressor genes. These findings make *SMARCA1* a suitable target for functional validation studies.

5.4.4 Genes Constantly Differentially Regulated

5.4.4.1 *NUF2*

NUF2, alongside with *CENP-H*, is the second validated gene associated with the kinetochore. In contrast to *CENP-H*, *NUF2* failed to meet the level of significance in the comparison of diploid and aneuploid mucosa by a small margin (fold change 1.816, p-value 0.0319). However, *NUF2* was identified as one of the genes matching the criteria for constant down regulation over the whole sequence analyzed.

NUF2 was first identified in 1994 as a myosin-like protein associated with the mitotic spindle¹⁵⁴. The kinetochore undergoes constant assembly and disassembly¹⁵⁵. Some of its associated proteins are constitutively expressed (i.e. *CENP-A*, *CENP-H*, *CENP-I*), while others are cell-cycle dependent and only found

in cells around mitosis (i.e. CENP-E). There is a substantial interdependency of kinetochore proteins, i.e. the temporal and spatial recruitment to the kinetochore of some proteins has been shown to greatly depend on other proteins¹¹⁸. For NUF2, it has been described that its localization to the kinetochore depends on CENP-I expression and Liu *et al.* suggest a “CENP-I assembly pathway”¹¹⁸. Once localized to the kinetochore, NUF2 has a crucial function for attachment of microtubules: a heterodimer of NUF2 and NDC80 binds to microtubules with binding sites localized distally from the kinetochore-binding region¹⁵⁶. Cells lacking NUF2 do not form functional trilaminar kinetochores, but instead a “fuzzy ball”-shaped structure” develops with greatly reduced ability to bind microtubules¹¹⁸. Interestingly, there is evidence that NUF2 and CENP-H interact to form stable structural components of the centromeres during mitosis¹⁵⁷.

Little is known on the involvement of NUF2 and associated proteins during aneuploidization in clinical samples. The results presented herein demonstrate that *NUF2* down regulation can be observed constantly from normal controls to UCCs. Considering the significant down regulation of CENP-H in aneuploid cells - a protein that interacts with NUF2 to orchestrate mitotic division - the importance of kinetochore assembly for malignant transformation is emphasized by our data. It would be highly interesting to investigate spatial and temporal distributions of CENP-H and NUF2 in clinical specimens (figure 19).

5.4.4.2 *TRIB3*

Our data indicate that *TRIB3* down regulation begins alongside with colitis initiation and is aggravated through aneuploidization and carcinogenesis. Thereby, we identify *TRIB3* down regulation as a novel mechanism putatively related to aneuploidy in UC.

TRIB3 belongs to the tribbles family of genes, which have been described in *Drosophila melanogaster* to regulate embryonic development by influencing cell cycle progression¹⁵⁸. Human *TRIB3* is involved in pathways such as lipolysis, insulin signal transduction, and muscle differentiation¹⁵⁹. Ohoka *et al.* have shown that knockdown of *TRIB3* in a cell line system decreases endoplasmic reticulum stress-dependent cell death mediated by *ATF4* (activating transcription factor 4) and *CHOP* (*C/EBP homologous protein*), an inducer of cell cycle arrest and apoptosis due to endoplasmic reticulum stress. Sparse data are available for

tribbles expression levels in human cancers. One study reported overexpression of *TRIB3* mRNA in a small cohort of colon, lung, and esophagus cancers, while the authors found down regulation in kidney tumors ¹⁶⁰.

5.4.4.3 *TFPI2*

Our results demonstrate that expression of *TFPI2* is diminished along malignant transformation in UC. *TFPI2* (*Tissue factor pathway inhibitor-2*) is a serine proteinase inhibitor, originally isolated from placental tissue as “placental protein 5” ¹⁶¹. Most of the protein produced is secreted into the extracellular department and plays a role in maintaining the integrity of the extracellular matrix for cell attachment ¹⁶². *TFPI2* inhibits matrix metalloproteinases ¹⁶³. In malignant tumors, decreased expression of *TFPI2* is associated with more aggressive behavior and advanced tumor stages, which has been most clearly demonstrated in human gliomas ¹⁶⁴. In other cancer entities such as pancreatic carcinomas and esophagus carcinomas, *TFPI2* down regulation could be observed due to hypermethylation of its promoter region ^{165, 166}. As *TFPI2* exerts inhibitory effects on extracellular proteinases, loss of *TFPI2* leads to increased proteolytic activity in the extracellular matrix, a mechanism that is believed to facilitate tumor expansion and metastatic spreading ¹⁶⁷. In general, *TFPI2* holds promise as a molecular target of new cancer therapies, as its expression is altered in numerous malignant tumors.

5.5 *Transition of Microarray Results into Clinical Application*

There are two principle ways to utilize microarray data for inference of valuable clinical instruments: First, gene expression signatures can be used directly to scan tissue samples in order to gain prognostic information or identify a biological status that defies discovery through conventional techniques: In the context of UC this could e.g. mean distinguishing patients who are likely to develop CRCs from those who are not. Other hopes lie in the development of classifiers for treatment prediction or prognosis. Until now, no such approach has been introduced to routine clinical practice.

Our sample numbers do not allow for a reasonable cross-validation of a gene signature elaborated: Previous studies on gene expression-based classifiers of breast cancer used 307 patients to cross-validate a 70-gene signature with

promising results ¹⁶⁸. Such a high number of patients can only be achieved in a concerted effort and under tremendous financial burdens. Nevertheless, studies such as the MINDACT trial for breast cancer (“*Microarray for Node Negative Disease may Avoid Chemotherapy*”) are on the way to generate level I evidence for the clinical usefulness of microarray-built classifiers ¹⁶⁹.

Secondly, Microarrays can be understood as “hypotheses generators” that allow holistic scanning of transcriptomic alterations. Subsequent to array experiments, single genes or groups of DEGs can be validated in clinical material by complementary techniques such as PCR, IHC, and others. We validated ten genes with RT-qPCR and could thereby acknowledge their differential expression, which was significant for three genes in the reduced subset of samples used for RT-qPCR. Thereby, the first step of validation is done for those three genes and in future projects, functional experiments can elucidate the role of these genes in aneuploidization or cancer development specifically.

6 SUMMARY

Ulcerative colitis is a premalignant lesion that imposes an increased risk for colorectal carcinomas on afflicted patients. The search for diagnostic markers had previously led to the discovery that aneuploidy precedes malignancy development in colitis almost invariably and is of prognostic value for impending carcinoma. Aneuploidy or “chromosomal instability” had moreover been recognized as one major genetic pathway of colorectal carcinogenesis in sporadic carcinomas where its presence is associated with inferior outcome. Likewise, for colitis-associated carcinomas, a poor outcome has been reported as compared to their sporadic counterpart. Therefore, one could hypothesize that aneuploidy *per se* might influence prognosis for ulcerative colitis patients. Yet, the extent of aneuploidy in colitis-associated carcinomas has not been systematically analyzed.

Thus, in the present study, 260 sporadic and 31 ulcerative-colitis associated carcinomas were analyzed with DNA-image cytometry for aneuploidy. While 74.6% of all sporadic cancers were aneuploid, all colitis-associated carcinomas showed chromosomal instability. Clinical parameters and survival characteristics of patients from both groups were investigated with respect to the tumor ploidy status. Hereby, known differences between sporadic and colitis-associated carcinomas could be acknowledged, e.g., colitis-associated carcinomas tending to occur more often in male patients than in females. Furthermore, it could be shown that colitis-associated carcinomas are characterized by a higher frequency of rare histopathological subtypes such as mucinous carcinomas, as well as an increased rate of synchronous malignancies. Using multivariate analyses, it could be demonstrated that aneuploidy confers a bad prognosis upon the patient, whereas the presence of ulcerative colitis as such does not. This finding was supported by survival analyses demonstrating that aneuploid carcinomas show inferior overall outcome albeit inflammation.

In summary, chromosomal instability can be found significantly more often in colitis-associated colorectal carcinomas and impairs patients' outcome, yet prognosis depends on aneuploidy and not on the presence of colitis.

To elaborate the impact of aneuploidy during malignant transformation, global gene expression profiling was utilized for transcriptomic profiling of normal colonic

controls (n = 9), diploid (n = 18) and aneuploid (n = 13) colitis mucosa in premalignant stages and colitis associated carcinomas (n = 7). While a massive deregulation of the transcriptome was observed after initiation of colitis, subsequent gene expression changes were relatively subtle. Yet, 16 genes that were constantly changed over the whole sequence could be identified and network analysis of these genes revealed close interactions in pathways involving *p53*, *TGF β 1* and *NF κ B*, thereby linking these canonical pathways at the crossroads of inflammation and cancer.

From the gene lists created, ten gene products were validated with RT-qPCR, which confirmed the trend of differential regulation in all cases. Among the so-validated differentially expressed genes, *CENP-H* and *NUF2* provide two examples of interacting partners in a putative chromosomal instability pathway: Both genes are involved in kinetochore assembly and orchestrate mitotic division. Lack of the genes or impaired function might drive aneuploidy and thereby malignant transformation. Thus, altered gene expression, as demonstrated in this study for both genes, might contribute to cancer development on the basis of chronic inflammation.

In conclusion, the results of the present thesis suggest a link between chronic colonic inflammation and chromosomal instability. Whole transcriptomic analyses reveal gene expression differences along the sequence of colitis initiation to cancer development, which warrant further investigations to elucidate cause and effect of aneuploidization and to clarify their potential as novel diagnostic and therapeutic targets.

7 REFERENCES

1. Kumar V, Abbas AK, Fausto N. Pathologic Basis of Diseases. 7 ed: W.B. Saunders (Elsevier), 2005.
2. Wilks S. The morbid appearance of the intestine of Miss Banks. *The Medical Times & Gazette*. 1859; 2(1):264.
3. Daiss W, Scheurlen M, Malchow H. Epidemiology of inflammatory bowel disease in the county of Tübingen (West Germany). *Scand J Gastroenterol Suppl* 1989; 170:39-43; discussion 50-5.
4. Habermann J, Lenander C, Roblick UJ, et al. Ulcerative colitis and colorectal carcinoma: DNA-profile, laminin-5 gamma2 chain and cyclin A expression as early markers for risk assessment. *Scand J Gastroenterol* 2001; 36(7):751-8.
5. Greenstein AJ. Cancer in inflammatory bowel disease. *Mt Sinai J Med* 2000; 67(3):227-40.
6. Loftus EV, Jr. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* 2004; 126(6):1504-17.
7. Shivananda S, Lennard-Jones J, Logan R, et al. Incidence of inflammatory bowel disease across Europe: is there a difference between north and south? Results of the European Collaborative Study on Inflammatory Bowel Disease (EC-IBD). *Gut* 1996; 39(5):690-7.
8. Xavier RJ, Podolsky DK. Unravelling the pathogenesis of inflammatory bowel disease. *Nature* 2007; 448(7152):427-34.
9. Cottone M, Scimeca D, Mocciaro F, et al. Clinical course of ulcerative colitis. *Dig Liver Dis* 2008; 40 Suppl 2:S247-52.
10. Langholz E, Munkholm P, Nielsen OH, et al. Incidence and prevalence of ulcerative colitis in Copenhagen county from 1962 to 1987. *Scand J Gastroenterol* 1991; 26(12):1247-56.
11. Williams H, Walker D, Orchard TR. Extraintestinal manifestations of inflammatory bowel disease. *Curr Gastroenterol Rep* 2008; 10(6):597-605.
12. Timani S, Mutasim DF. Skin manifestations of inflammatory bowel disease. *Clin Dermatol* 2008; 26(3):265-73.
13. Jess T, Gamborg M, Munkholm P, Sorensen TI. Overall and cause-specific mortality in ulcerative colitis: meta-analysis of population-based inception cohort studies. *Am J Gastroenterol* 2007; 102(3):609-17.
14. Podolsky DK. Inflammatory bowel disease. *N Engl J Med* 2002; 347(6):417-29.
15. Zisman TL, Rubin DT. Novel diagnostic and prognostic modalities in inflammatory bowel disease. *Med Clin North Am* 2010; 94(1):155-78.
16. Roda G, Caponi A, Benevento M, et al. New proteomic approaches for biomarker discovery in inflammatory bowel disease. *Inflamm Bowel Dis*. 2010; 16(7):1239-46
17. Baumgart DC, Sandborn WJ. Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet* 2007; 369(9573):1641-57.
18. Moreland L, Bate G, Kirkpatrick P. Abatacept. *Nat Rev Drug Discov* 2006; 5(3):185-6.

19. Creed TJ, Norman MR, Probert CS, et al. Basiliximab (anti-CD25) in combination with steroids may be an effective new treatment for steroid-resistant ulcerative colitis. *Aliment Pharmacol Ther* 2003; 18(1):65-75.
20. Becker JM, Stucchi AF. Treatment of choice for acute severe steroid-refractory ulcerative colitis is colectomy. *Inflamm Bowel Dis* 2009; 15(1):146-9.
21. Hermon-Taylor J. Protagonist. *Mycobacterium avium* subspecies paratuberculosis is a cause of Crohn's disease. *Gut* 2001; 49(6):755-6.
22. Weinstock JV, Summers RW, Elliott DE, et al. The possible link between de-worming and the emergence of immunological disease. *J Lab Clin Med* 2002; 139(6):334-8.
23. Jowett SL, Seal CJ, Pearce MS, et al. Influence of dietary factors on the clinical course of ulcerative colitis: a prospective cohort study. *Gut* 2004; 53(10):1479-84.
24. Aronowitz R, Spiro HM. The rise and fall of the psychosomatic hypothesis in ulcerative colitis. *J Clin Gastroenterol* 1988; 10(3):298-305.
25. Hanauer S. Update on the etiology, pathogenesis and diagnosis of ulcerative colitis. *Nature Clinical Practice* 2004; 1:26-31.
26. Franchimont D, Vermeire S, El Housni H, et al. Deficient host-bacteria interactions in inflammatory bowel disease? The toll-like receptor (TLR)-4 Asp299gly polymorphism is associated with Crohn's disease and ulcerative colitis. *Gut* 2004; 53(7):987-92.
27. Yeretssian G, Labbe K, Saleh M. Molecular regulation of inflammation and cell death. *Cytokine* 2008; 43(3):380-90.
28. Elson CO, Cong Y, McCracken VJ, et al. Experimental models of inflammatory bowel disease reveal innate, adaptive, and regulatory mechanisms of host dialogue with the microbiota. *Immunol Rev* 2005; 206:260-76.
29. Ogura Y, Bonen DK, Inohara N, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001; 411(6837):603-6.
30. Duerr RH, Taylor KD, Brant SR, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006; 314(5804):1461-3.
31. Franke A, Balschun T, Karlsen TH, et al. Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. *Nat Genet* 2008; 40(11):1319-23.
32. Spehlmann ME, Begun AZ, Burghardt J, et al. Epidemiology of inflammatory bowel disease in a German twin cohort: results of a nationwide study. *Inflamm Bowel Dis* 2008; 14(7):968-76.
33. Harpaz N, Talbot IC. Colorectal cancer in idiopathic inflammatory bowel disease. *Semin Diagn Pathol* 1996; 13(4):339-57.
34. Jess T, Loftus EV, Jr., Velayos FS, et al. Risk factors for colorectal neoplasia in inflammatory bowel disease: a nested case-control study from Copenhagen county, Denmark and Olmsted county, Minnesota. *Am J Gastroenterol* 2007; 102(4):829-36.
35. Velayos FS, Loftus EV, Jr., Jess T, et al. Predictive and protective factors associated with colorectal cancer in ulcerative colitis: A case-control study. *Gastroenterology* 2006; 130(7):1941-9.

36. Soderlund S, Granath F, Brostrom O, et al. Inflammatory bowel disease confers a lower risk of colorectal cancer to females than to males. *Gastroenterology* 2010; 138(5):1697-703.
37. Pohl C, Hombach A, Kruis W. Chronic inflammatory bowel disease and cancer. *Hepatogastroenterology* 2000; 47(31):57-70.
38. Delaunoit T, Limburg PJ, Goldberg RM, et al. Colorectal cancer prognosis among patients with inflammatory bowel disease. *Clin Gastroenterol Hepatol* 2006; 4(3):335-42.
39. Aarnio M, Mustonen H, Mecklin JP, Jarvinen HJ. Prognosis of colorectal cancer varies in different high-risk conditions. *Ann Med* 1998; 30(1):75-80.
40. Vogelstein B, Fearon ER, Hamilton SR, et al. Genetic alterations during colorectal-tumor development. *N Engl J Med* 1988; 319(9):525-32.
41. Sjoqvist U. Dysplasia in ulcerative colitis--clinical consequences? *Langenbecks Arch Surg* 2004; 389(5):354-60.
42. Schmidt C, Bielecki C, Felber J, Stallmach A. Surveillance strategies in inflammatory bowel disease. *Minerva Gastroenterol Dietol* 2010; 56(2):189-201.
43. Collins PD, Mpofo C, Watson AJ, Rhodes JM. Strategies for detecting colon cancer and/or dysplasia in patients with inflammatory bowel disease. *Cochrane Database Syst Rev* 2006(2):CD000279.
44. Odze RD, Tomaszewski JE, Furth EE, et al. Variability in the diagnosis of dysplasia in ulcerative colitis by dynamic telepathology. *Oncol Rep* 2006; 16(5):1123-9.
45. Holzmann K, Klump B, Borchard F, et al. Flow cytometric and histologic evaluation in a large cohort of patients with ulcerative colitis: correlation with clinical characteristics and impact on surveillance. *Dis Colon Rectum* 2001; 44(10):1446-55.
46. Befrits R, Hammarberg C, Rubio C, et al. DNA aneuploidy and histologic dysplasia in long-standing ulcerative colitis. A 10-year follow-up study. *Dis Colon Rectum* 1994; 37(4):313-9; discussion 319-20.
47. Rosman-Urbach M, Niv Y, Birk Y, et al. A high degree of aneuploidy, loss of p53 gene, and low soluble p53 protein serum levels are detected in ulcerative colitis patients. *Dis Colon Rectum* 2004; 47(3):304-13.
48. Stenling R, Jonsson BO, Palmqvist R, Rutegard JN. DNA aneuploidy in ulcerative colitis and in colorectal carcinoma--a comparative study. *Anal Cell Pathol* 1999; 18(2):69-72.
49. Lindberg JO, Stenling RB, Rutegard JN. DNA aneuploidy as a marker of premalignancy in surveillance of patients with ulcerative colitis. *Br J Surg* 1999; 86(7):947-50.
50. Boveri T. Zur Frage der Entstehung maligner Tumoren. Fischer, Jena 1914.
51. Holland AJ, Cleveland DW. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nat Rev Mol Cell Biol* 2009; 10(7):478-87.
52. Weaver BA, Silk AD, Cleveland DW. Cell biology: nondisjunction, aneuploidy and tetraploidy. *Nature* 2006; 442(7104):E9-10; discussion E10.
53. Li L, Mu K, Zhou G, et al. Genomic instability and proliferative activity as risk factors for distant metastases in breast cancer. *Br J Cancer* 2008; 99(3):513-9.
54. Araujo SE, Bernardo WM, Habr-Gama A, et al. DNA ploidy status and prognosis in colorectal cancer: a meta-analysis of published data. *Dis Colon Rectum* 2007; 50(11):1800-10.

55. Schliekelman M, Cowley DO, O'Quinn R, et al. Impaired Bub1 function in vivo compromises tension-dependent checkpoint function leading to aneuploidy and tumorigenesis. *Cancer Res* 2009; 69(1):45-54.
56. Campbell MS, Chan GK, Yen TJ. Mitotic checkpoint proteins HsMAD1 and HsMAD2 are associated with nuclear pore complexes in interphase. *J Cell Sci* 2001; 114(Pt 5):953-63.
57. Cimini D, Howell B, Maddox P, et al. Merotelic kinetochore orientation is a major mechanism of aneuploidy in mitotic mammalian tissue cells. *J Cell Biol* 2001; 153(3):517-27.
58. O'Sullivan JN, Bronner MP, Brentnall TA, et al. Chromosomal instability in ulcerative colitis is related to telomere shortening. *Nat Genet* 2002; 32(2):280-4.
59. Weaver BA, Cleveland DW. The aneuploidy paradox in cell growth and tumorigenesis. *Cancer Cell* 2008; 14(6):431-3.
60. Weaver BA, Silk AD, Montagna C, et al. Aneuploidy acts both oncogenically and as a tumor suppressor. *Cancer Cell* 2007; 11(1):25-36.
61. Helou K, Padilla-Nash H, Wangsa D, et al. Comparative genome hybridization reveals specific genomic imbalances during the genesis from benign through borderline to malignant ovarian tumors. *Cancer Genet Cytogenet* 2006; 170(1):1-8.
62. Upender MB, Habermann JK, McShane LM, et al. Chromosome transfer induced aneuploidy results in complex dysregulation of the cellular transcriptome in immortalized and cancer cells. *Cancer Res* 2004; 64(19):6941-9.
63. Kuismanen SA, Holmberg MT, Salovaara R, et al. Epigenetic phenotypes distinguish microsatellite-stable and -unstable colorectal cancers. *Proc Natl Acad Sci U S A* 1999; 96(22):12661-6.
64. Mueller J, Gazzoli I, Bandipalliam P, et al. Comprehensive molecular analysis of mismatch repair gene defects in suspected Lynch syndrome (hereditary nonpolyposis colorectal cancer) cases. *Cancer Res* 2009; 69(17):7053-61.
65. Stigliano V, Assisi D, Cosimelli M, et al. Survival of hereditary non-polyposis colorectal cancer patients compared with sporadic colorectal cancer patients. *J Exp Clin Cancer Res* 2008; 27:39.
66. Makiyama K, Tokunaga M, Itsuno M, et al. DNA aneuploidy in a case of rectosigmoid adenocarcinoma complicated by ulcerative colitis. *J Gastroenterol* 1995; 30(2):258-63.
67. Clausen OP, Andersen SN, Stroomkjaer H, et al. A strategy combining flow sorting and comparative genomic hybridization for studying genetic aberrations at different stages of colorectal tumorigenesis in ulcerative colitis. *Cytometry* 2001; 43(1):46-54.
68. Klump B, Holzmann K, Kuhn A, et al. Distribution of cell populations with DNA aneuploidy and p53 protein expression in ulcerative colitis. *Eur J Gastroenterol Hepatol* 1997; 9(8):789-94.
69. Holzmann K, Klump B, Borchard F, et al. Comparative analysis of histology, DNA content, p53 and Ki-ras mutations in colectomy specimens with long-standing ulcerative colitis. *Int J Cancer* 1998; 76(1):1-6.
70. Levine DS, Rabinovitch PS, Haggitt RC, et al. Distribution of aneuploid cell populations in ulcerative colitis with dysplasia or cancer. *Gastroenterology* 1991; 101(5):1198-210.

71. Burmer GC, Rabinovitch PS, Loeb LA. Frequency and spectrum of c-Ki-ras mutations in human sporadic colon carcinoma, carcinomas arising in ulcerative colitis, and pancreatic adenocarcinoma. *Environ Health Perspect* 1991; 93:27-31.
72. Bernstein CN, Blanchard JF, Kliewer E, Wajda A. Cancer risk in patients with inflammatory bowel disease: a population-based study. *Cancer* 2001; 91(4):854-62.
73. Habermann JK, Upender MB, Roblick UJ, et al. Pronounced chromosomal instability and multiple gene amplifications characterize ulcerative colitis-associated colorectal carcinomas. *Cancer Genet Cytogenet* 2003; 147(1):9-17.
74. Half EE, Bresalier RS. Clinical management of hereditary colorectal cancer syndromes. *Curr Opin Gastroenterol* 2004; 20(1):32-42.
75. Feulgen R, Rossenbeck H. Mikroskopisch-chemischer Nachweis einer Nucleinsäure vom Typus Thymusnucleinsäure und die darauf beruhende elektive Färbung von Zellkernen in mikroskopischen Präparaten. *Hoppe Seyler's Z Physiol Chem* 1924; 135:203-248.
76. Kjellstrand P. Mechanisms of the Feulgen acid hydrolysis. *J Microsc* 1980; 119(3):391-6.
77. Auer GU, Caspersson TO, Wallgren AS. DNA content and survival in mammary carcinoma. *Anal Quant Cytol* 1980; 2(3):161-5.
78. Thunnissen FB, Ellis IO, Jutting U. Quality assurance in DNA image analysis on diploid cells. *Cytometry* 1997; 27(1):21-5.
79. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; 270(5235):467-70.
80. Gomes LI, Silva RL, Stolf BS, et al. Comparative analysis of amplified and nonamplified RNA for hybridization in cDNA microarray. *Anal Biochem* 2003; 321(2):244-51.
81. Pabon C, Modrusan Z, Ruvolo MV, et al. Optimized T7 amplification system for microarray analysis. *Biotechniques* 2001; 31(4):874-9.
82. Relogio A, Schwager C, Richter A, et al. Optimization of oligonucleotide-based DNA microarrays. *Nucleic Acids Res* 2002; 30(11):e51.
83. Korn EL, Habermann JK, Upender MB, et al. Objective method of comparing DNA microarray image analysis systems. *Biotechniques* 2004; 36(6):960-7.
84. Nolan T, Hands RE, Bustin SA. Quantification of mRNA using real-time RT-PCR. *Nat Protoc* 2006; 1(3):1559-82.
85. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001; 25(4):402-8.
86. Cicinnati VR, Shen Q, Sotiropoulos GC, et al. Validation of putative reference genes for gene expression studies in human hepatocellular carcinoma using real-time quantitative RT-PCR. *BMC Cancer* 2008; 8:350.
87. Steinau M, Rajeevan MS, Unger ER. DNA and RNA references for qRT-PCR assays in exfoliated cervical cells. *J Mol Diagn* 2006; 8(1):113-8.
88. Abruzzo LV, Lee KY, Fuller A, et al. Validation of oligonucleotide microarray data using microfluidic low-density arrays: a new statistical method to normalize real-time RT-PCR data. *Biotechniques* 2005; 38(5):785-92.

89. Radonic A, Thulke S, Mackay IM, et al. Guideline to reference gene selection for quantitative real-time PCR. *Biochem Biophys Res Commun* 2004; 313(4):856-62.
90. Fleiss JL. Measuring nominal scale agreement among many raters. *Psych. Bulletin* 1971; 76(5):4.
91. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* 1995; 57(1):289-300.
92. Ficenec D, Osborne M, Pradines J, et al. Computational knowledge integration in biopharmaceutical research. *Brief Bioinform* 2003; 4(3):260-78.
93. Calvano SE, Xiao W, Richards DR, et al. A network-based analysis of systemic inflammation in humans. *Nature* 2005; 437(7061):1032-7.
94. Rosenberg R, Friederichs J, Schuster T, et al. Prognosis of patients with colorectal cancer is associated with lymph node ratio: a single-center analysis of 3,026 patients over a 25-year time period. *Ann Surg* 2008; 248(6):968-78.
95. Zlobec I, Lugli A. Prognostic and predictive factors in colorectal cancer. *J Clin Pathol* 2008; 61(5):561-9.
96. Denlinger CS, Cohen SJ. Progress in the development of prognostic and predictive markers for gastrointestinal malignancies. *Curr Treat Options Oncol* 2007; 8(5):339-51.
97. Coleman MP, Quaresma M, Berrino F, et al. Cancer survival in five continents: a worldwide population-based study (CONCORD). *Lancet Oncol* 2008; 9(8):730-56.
98. Danque PO, Chen HB, Patil J, et al. Image analysis versus flow cytometry for DNA ploidy quantitation of solid tumors: a comparison of six methods of sample preparation. *Mod Pathol* 1993; 6(3):270-5.
99. Eaden J, Abrams K, McKay H, et al. Inter-observer variation between general and specialist gastrointestinal pathologists when grading dysplasia in ulcerative colitis. *J Pathol* 2001; 194(2):152-7.
100. Tsuchiya A, Ando Y, Ishii Y, et al. Flow cytometric DNA analysis in Japanese colorectal cancer. A multivariate analysis. *Eur J Surg Oncol* 1992; 18(6):585-90.
101. Nori D, Merimsky O, Samala E, et al. Tumor ploidy as a risk factor for disease recurrence and short survival in surgically-treated Dukes' B2 colon cancer patients. *J Surg Oncol* 1995; 59(4):239-42.
102. Geido E, Sciutto A, Rubagotti A, et al. Combined DNA flow cytometry and sorting with k-ras2 mutation spectrum analysis and the prognosis of human sporadic colorectal cancer. *Cytometry* 2002; 50(4):216-24.
103. Itzkowitz SH, Harpaz N. Diagnosis and management of dysplasia in patients with inflammatory bowel diseases. *Gastroenterology* 2004; 126(6):1634-48.
104. Sacconi Jotti G, Fontanesi M, Orsi N, et al. DNA content in human colon cancer and non-neoplastic adjacent mucosa. *Int J Biol Markers* 1995; 10(1):11-6.
105. Chen WS, Chen JY, Liu JM, et al. Microsatellite instability in sporadic-colon-cancer patients with and without liver metastases. *Int J Cancer* 1997; 74(4):470-4.

106. Colotta F, Allavena P, Sica A, et al. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis* 2009; 30(7):1073-81.
107. Noble CL, Abbas AR, Cornelius J, et al. Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis. *Gut* 2008; 57(10):1398-405.
108. Kaser A, Martinez-Naves E, Blumberg RS. Endoplasmic reticulum stress: implications for inflammatory bowel disease pathogenesis. *Curr Opin Gastroenterol* 2010; 26(4):318-26.
109. Mantovani A. Molecular pathways linking inflammation and cancer. *Curr Mol Med* 2010; 10(4):369-73.
110. Hackstadt AJ, Hess AM. Filtering for Increased Power for Microarray Data Analysis. *BMC Bioinformatics* 2009; 10(1):11.
111. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003; 100(16):9440-5.
112. Clevers H. At the crossroads of inflammation and cancer. *Cell* 2004; 118(6):671-4.
113. Almon RR, DuBois DC, Jusko WJ. A microarray analysis of the temporal response of liver to methylprednisolone: a comparative analysis of two dosing regimens. *Endocrinology* 2007; 148(5):2209-25.
114. Csillag C, Borup R, Olsen J, et al. Treatment response and colonic gene expression in patients with Crohn's disease. *Scand J Gastroenterol* 2007; 42(7):834-40.
115. van Dekken H, Wink JC, Vissers KJ, et al. Wnt pathway-related gene expression during malignant progression in ulcerative colitis. *Acta Histochem* 2007; 109(4):266-72.
116. Magro F, Dinis-Ribeiro M, Araujo FM, et al. High prevalence of combined thrombophilic abnormalities in patients with inflammatory bowel disease. *Eur J Gastroenterol Hepatol* 2003; 15(11):1157-63.
117. Samarakoon R, Higgins CE, Higgins SP, Higgins PJ. TGF-beta1-Induced Expression of the Poor Prognosis SERPINE1/PAI-1 Gene Requires EGFR Signaling: A New Target for Anti-EGFR Therapy. *J Oncol* 2009; 2009:342391.
118. Liu ST, Rattner JB, Jablonski SA, Yen TJ. Mapping the assembly pathways that specify formation of the trilaminar kinetochore plates in human cells. *J Cell Biol* 2006; 175(1):41-53.
119. Hirano T. Chromosome cohesion, condensation, and separation. *Annu Rev Biochem* 2000; 69:115-44.
120. Hirano T. The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair. *Genes Dev* 2002; 16(4):399-414.
121. Gruber S, Haering CH, Nasmyth K. Chromosomal cohesin forms a ring. *Cell* 2003; 112(6):765-77.
122. Kim ST, Xu B, Kastan MB. Involvement of the cohesin protein, Smc1, in Atm-dependent and independent responses to DNA damage. *Genes Dev* 2002; 16(5):560-70.
123. Ghiselli G. SMC3 knockdown triggers genomic instability and p53-dependent apoptosis in human and zebrafish cells. *Mol Cancer* 2006; 5:52.
124. Barber TD, McManus K, Yuen KW, et al. Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. *Proc Natl Acad Sci U S A* 2008; 105(9):3443-8.

125. Ghiselli G, Iozzo RV. Overexpression of bamacan/SMC3 causes transformation. *J Biol Chem* 2000; 275(27):20235-8.
126. Mellone B, Erhardt S, Karpen GH. The ABCs of centromeres. *Nat Cell Biol* 2006; 8(5):427-9.
127. Sullivan KF. A solid foundation: functional specialization of centromeric chromatin. *Curr Opin Genet Dev* 2001; 11(2):182-8.
128. Foltz DR, Jansen LE, Black BE, et al. The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol* 2006; 8(5):458-69.
129. Okada M, Cheeseman IM, Hori T, et al. The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. *Nat Cell Biol* 2006; 8(5):446-57.
130. Orthaus S, Ohndorf S, Diekmann S. RNAi knockdown of human kinetochore protein CENP-H. *Biochem Biophys Res Commun* 2006; 348(1):36-46.
131. Tomonaga T, Matsushita K, Ishibashi M, et al. Centromere protein H is up-regulated in primary human colorectal cancer and its overexpression induces aneuploidy. *Cancer Res* 2005; 65(11):4683-9.
132. Abaza A, Soleilhac JM, Westendorf J, et al. M phase phosphoprotein 1 is a human plus-end-directed kinesin-related protein required for cytokinesis. *J Biol Chem* 2003; 278(30):27844-52.
133. Kanehira M, Katagiri T, Shimo A, et al. Oncogenic role of MPHOSPH1, a cancer-testis antigen specific to human bladder cancer. *Cancer Res* 2007; 67(7):3276-85.
134. Kamimoto T, Zama T, Aoki R, et al. Identification of a novel kinesin-related protein, KRMP1, as a target for mitotic peptidyl-prolyl isomerase Pin1. *J Biol Chem* 2001; 276(40):37520-8.
135. Tzu J, Marinkovich MP. Bridging structure with function: structural, regulatory, and developmental role of laminins. *Int J Biochem Cell Biol* 2008; 40(2):199-214.
136. Lenander C, Habermann JK, Ost A, et al. Laminin-5 gamma 2 chain expression correlates with unfavorable prognosis in colon carcinomas. *Anal Cell Pathol* 2001; 22(4):201-9.
137. Shinto E, Baker K, Tsuda H, et al. Tumor buds show reduced expression of laminin-5 gamma 2 chain in DNA mismatch repair deficient colorectal cancer. *Dis Colon Rectum* 2006; 49(8):1193-202.
138. Shinto E, Tsuda H, Ueno H, et al. Prognostic implication of laminin-5 gamma 2 chain expression in the invasive front of colorectal cancers, disclosed by area-specific four-point tissue microarrays. *Lab Invest* 2005; 85(2):257-66.
139. Pyke C, Romer J, Kallunki P, et al. The gamma 2 chain of laminin 5 is preferentially expressed in invading malignant cells in human cancers. *Am J Pathol* 1994; 145(4):782-91.
140. Beyaert R, Heyninck K, Van Huffel S. A20 and A20-binding proteins as cellular inhibitors of nuclear factor-kappa B-dependent gene expression and apoptosis. *Biochem Pharmacol* 2000; 60(8):1143-51.
141. Lee EG, Boone DL, Chai S, et al. Failure to regulate TNF-induced NF-kappaB and cell death responses in A20-deficient mice. *Science* 2000; 289(5488):2350-4.
142. Wertz IE, O'Rourke KM, Zhou H, et al. De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-kappaB signalling. *Nature* 2004; 430(7000):694-9.

143. Dieguez-Gonzalez R, Calaza M, Perez-Pampin E, et al. Analysis of TNFAIP3, a feedback inhibitor of NFkappaB, and the neighbor intergenic 6q23 region in rheumatoid arthritis susceptibility. *Arthritis Res Ther* 2009; 11(2):R42.
144. Fung EY, Smyth DJ, Howson JM, et al. Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. *Genes Immun* 2009; 10(2):188-91.
145. Yeger H, Perbal B. The CCN family of genes: a perspective on CCN biology and therapeutic potential. *J Cell Commun Signal* 2007; 1(3-4):159-64.
146. Meyuhas R, Pikarsky E, Tavor E, et al. A Key role for cyclic AMP-responsive element binding protein in hypoxia-mediated activation of the angiogenesis factor CCN1 (CYR61) in Tumor cells. *Mol Cancer Res* 2008; 6(9):1397-409.
147. Mori A, Desmond JC, Komatsu N, et al. CYR61: a new measure of lung cancer outcome. *Cancer Invest* 2007; 25(8):738-41.
148. Maeta N, Osaki M, Shomori K, et al. CYR61 downregulation correlates with tumor progression by promoting MMP-7 expression in human gastric carcinoma. *Oncology* 2007; 73(1-2):118-26.
149. O'Kelly J, Chung A, Lemp N, et al. Functional domains of CCN1 (Cyr61) regulate breast cancer progression. *Int J Oncol* 2008; 33(1):59-67.
150. Lv H, Fan E, Sun S, et al. Cyr61 is up-regulated in prostate cancer and associated with the p53 gene status. *J Cell Biochem* 2009; 106(4):738-44.
151. Muchardt C, Yaniv M. When the SWI/SNF complex remodels...the cell cycle. *Oncogene* 2001; 20(24):3067-75.
152. Sudarsanam P, Iyer VR, Brown PO, Winston F. Whole-genome expression analysis of snf/swi mutants of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 2000; 97(7):3364-9.
153. Rozanov DV, Savinov AY, Williams R, et al. Molecular signature of MT1-MMP: transactivation of the downstream universal gene network in cancer. *Cancer Res* 2008; 68(11):4086-96.
154. Osborne MA, Schlenstedt G, Jinks T, Silver PA. Nuf2, a spindle pole body-associated protein required for nuclear division in yeast. *J Cell Biol* 1994; 125(4):853-66.
155. Cleveland DW, Mao Y, Sullivan KF. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell* 2003; 112(4):407-21.
156. Wilson-Kubalek EM, Cheeseman IM, Yoshioka C, et al. Orientation and structure of the Ndc80 complex on the microtubule lattice. *J Cell Biol* 2008; 182(6):1055-61.
157. Mikami Y, Hori T, Kimura H, Fukagawa T. The functional region of CENP-H interacts with the Nuf2 complex that localizes to centromere during mitosis. *Mol Cell Biol* 2005; 25(5):1958-70.
158. Grosshans J, Wieschaus E. A genetic link between morphogenesis and cell division during formation of the ventral furrow in *Drosophila*. *Cell* 2000; 101(5):523-31.
159. Ohoka N, Yoshii S, Hattori T, et al. TRB3, a novel ER stress-inducible gene, is induced via ATF4-CHOP pathway and is involved in cell death. *Embo J* 2005; 24(6):1243-55.
160. Xu J, Lv S, Qin Y, et al. TRB3 interacts with CtIP and is overexpressed in certain cancers. *Biochim Biophys Acta* 2007; 1770(2):273-8.
161. Seppala M, Wahlstrom T, Bohn H. Circulating levels and tissue localization of placental protein five (PP5) in pregnancy and trophoblastic disease:

- absence of PP5 expression in the malignant trophoblast. *Int J Cancer* 1979; 24(1):6-10.
162. Chand HS, Foster DC, Kisiel W. Structure, function and biology of tissue factor pathway inhibitor-2. *Thromb Haemost* 2005; 94(6):1122-30.
 163. Herman MP, Sukhova GK, Kisiel W, et al. Tissue factor pathway inhibitor-2 is a novel inhibitor of matrix metalloproteinases with implications for atherosclerosis. *J Clin Invest* 2001; 107(9):1117-26.
 164. Sierko E, Wojtukiewicz MZ, Kisiel W. The role of tissue factor pathway inhibitor-2 in cancer biology. *Semin Thromb Hemost* 2007; 33(7):653-9.
 165. Tsunoda S, Smith E, De Young NJ, et al. Methylation of CLDN6, FBN2, RBP1, RBP4, TFPI2, and TMEFF2 in esophageal squamous cell carcinoma. *Oncol Rep* 2009; 21(4):1067-73.
 166. Sato N, Parker AR, Fukushima N, et al. Epigenetic inactivation of TFPI-2 as a common mechanism associated with growth and invasion of pancreatic ductal adenocarcinoma. *Oncogene* 2005; 24(5):850-8.
 167. Rao CN, Cook B, Liu Y, et al. HT-1080 fibrosarcoma cell matrix degradation and invasion are inhibited by the matrix-associated serine protease inhibitor TFPI-2/33 kDa MSPI. *Int J Cancer* 1998; 76(5):749-56.
 168. Buyse M, Loi S, van't Veer L, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 2006; 98(17):1183-92.
 169. Mook S, Bonnefoi H, Pruneri G, et al. Daily clinical practice of fresh tumour tissue freezing and gene expression profiling; logistics pilot study preceding the MINDACT trial. *Eur J Cancer* 2009; 45(7):1201-8.

8 APPENDIX

8.1 Ethical Permits

These studies were approved by the local ethical review board (Universität zu Lübeck) under the following running heads and numbers:

1. Nr. 99-121 vom 30.11.1999: "Untersuchung von klinischen, makroskopischen und immunhistochemischen Merkmalen sowie genetischen Veränderungen als mögliche Erkennungsfaktoren der Tumorentstehung bei Colitis ulcerosa im Vergleich zu sporadischen kolorektalen Karzinomen"
2. Nr. 07-124 vom 21.09.2007: "Norddeutsche Tumorbank zur Erforschung von Darmkrebs für verbesserte (Früh-)Diagnose, Therapie, Nachsorge und Prognose"

8.2 Reagents and Solutions

- *Di-Potassiumhydrogenphosphat*; MERCK 4871.1000
- *Di-Natriumhydrogenphosphat*; MERCK 3095.1000
- *Entellan*; MERCK 1.07961.0100
- *Ethanol 100%*; JT Baker Laboratory Chemicals
- *Natriumchloride* (NaCl); M = 58,44g/mol, MERCK 1.06404.
- *Natriumnitrit* (NaNO₂); M = 69,0g/mol, MERCK, 6549
- *Neufuchsin*; Cat Nr. 4040, MERCK
- *XyloI*; M = 106,17g/mol, Merck 1.08685.
- *15 ml centrifuge tubes with screw caps*; Cat.Nr. 2610R22, Thomas Scientific, Swedesboro, NJ, USA
- *50 ml centrifuge tubes with screw caps*; Cat.Nr. 2610R54, Thomas Scientific, Swedesboro, NJ, USA
- *RNeasy® Midi Kit (50)*; Cat.Nr. 75144, Qiagen, Valencia, CA, USA
- *RNeasy® Maxi Kit (12)*; Cat.Nr. 75162, Qiagen, Valencia, CA, USA
- *3M Sodium Acetate pH 5,2*; Cat. Nr. 351-035-060, Quality Biological, Inc., Gaithersburg, MD, USA
- *Chloroform (CHCL₃)*; Cat. Nr. 351-035-060, Mallinckrodt, Hazelwood, MO, USA
- *Ethyl Alcohol*, U.S.P.; Cat Nr. 64-17-5, The Warner-Graham-Company, Cockeysville, MD, USA

- *Amino Allyl MessageAmp™ aRNA*; Cat. Nr. 1752, Lot Nr. 025K23, Ambion Inc., Austin, TX, USA
- *RNase AWAY*; Cat Nr. 7000, Molecular BioProducts, San Diego, CA, USA
- *TRIzol® Reagent*; Cat. Nr. 15596-018, invitrogen, Carlsbad, CA, USA
- *RNA Fragmentation Reagents*; Cat. Nr. 8740, Ambion Inc., Austin, TX, USA
- *Bovines Serum Albumin*, Cat. Nr. A-9478, Sigma, St. Louis, MD, USA
- *Cy5 Mono-Reactive Dye Pack*, Cat. Nr. PA25001, Amersham Biosciences Group, NJ, USA
- *Cy3 Mono-Reactive Dye Pack*, Cat. Nr. PA23001, Amersham Biosciences Group, NJ, USA
- *Hybridization Cassette for Microarrays*; Cat. Nr. AHC-1 ArrayIt™ DNA Microarray Products, Sunnyvale, CA, USA
- *Microcon Centrifugal Filter Units YM-30*; Cat.Nr. 42410 Millipore, Billerica, MA, USA
- *Poly A-DNA (40-60)*; Cat.Nr. 27-7988-01, apbiotech, Piscataway, NJ, USA
- *DNA Cot1-Human*; Cat. Nr. 1581074, Boehringer-Mannheim/Roche-Applied Science, Indianapolis, IN, USA
- *DEPC treated water*, Cat.Nr. 750024, Research Genetics, Huntsville, AL, USA
- *20X SSC, 1 Liter*, Cat.Nr. 750020, Research Genetics, Huntsville, AL, USA
- *SYBR Green PCR Master Mix*; Cat Nr. SKU# S-7563, Invitrogen
- *PCR primer*, individually designed using CLC bio software DNA bench 5.0 (Aarhus, Denmark), ordered from biomers.net
- *MicroAmp Optical 96-Well reaction plates*; Cat Nr. N801-0560, Applied Biosystems
- *Optical Adhesive Covers for PCR plates*; Cat Nr. 4311971, Applied Biosystems
- *DNAse I*; Cat Nr.18047019, Invitrogen

8.3 Protocols

8.3.1 Haematoxylin – Eosin – Staining

1. deparaffinate for 3*5' in xyluol
2. rehydrate in ethanol (100%, 75%, 95%, 95%, 70%) for 2' each
3. rinse in aqua dest. briefly, then rinse in tap water
4. stain for 3-5' in Meyer's Haematoxylin
5. rinse for 10' in tap water
6. stain for 30-45'' in Eosin
7. rinse in tap water
8. dehydrate in Ethanol (70%, 95%, 95%) briefly and in 2*100% for 2' each
9. 2*xyluol, cover with Entellan

8.3.2 Feulgen – Staining

1st day:

1. add 5g Pararosnilin (Basic Fuchsin, Aldrich, 85, 734-3) to 150ml 1M HCl at RT (use a 2L bottle), shake gently
2. add 5g K₂S₂O₅ and 850ml Aqua dest., shake gently
3. cover bottle in aluminium foil and store at RT over night

2nd day:

1. add 3g active charcoal and shake thoroughly for 4'
2. filter twice (2nd time into a clear-glass bottle)
3. store pink solution at 4°C (can be used for up to 3 days)
4. deparaffinate embedded samples (as in H. & E. protocol) and refixate in 4% formaldehyde over night

3rd day:

1. rinse paraffin specimens in tap water until formaldehyde odour disappears
2. incubate in 5M HCl for 60'
3. rinse carefully in Aqua dest. three times
4. incubate for 120' in Feulgen's reagent (in the dark at RT)
5. rinse carefully to remove excessive staining
6. rinse three times for 10' each in new Na₂S₂O₅ solution
7. rinse for 5' in tap water
8. rehydrate and cover in Entellan (as described in H&E protocol)

8.3.3 RNA Extraction

Extraction kit: RNeasy® Midi Kit (50); Cat.Nr. 75144, Qiagen, Valencia, USA,
for abundant material:
RNeasy® Maxi Kit (12); Cat.Nr. 75162, Qiagen, Valencia, USA

1. Add 6ml TriZol Reagent into 15ml tube, add frozen tissue and immediately homogenize to avoid RNA degradation

2. Fill up with 4ml TriZol, shake for 15'' and let sit for at least 5'
3. Add 2ml Chloroform, shake for 15'', let sit for 10' or until phase separation becomes clearly visible (may take longer for large specimens)
4. centrifuge for 30' at 4°C and 8000 g
5. pipette transparent, upper phase into new 15ml tube
6. add 2ml 100% ethanol drop wise while shaking
7. for Qiagen RNAeasy-purification pipette half of the liquid in each tube to Qiagen-centrifuge column (pink lid).
8. centrifuge at 5500 g briefly until rotor is a full speed
9. remove lid, place over column and add second half over the liquid
10. centrifuge as in 8.) and repeat steps 7.-9. to complete four rounds of centrifugation
11. discard elute (RNA should be retained in column)
12. replace column upon the tube and add 4ml RW1 buffer
13. centrifuge at 5000 g for 5'
14. repeat steps 12 and 13 once
15. repeat steps 7 and 8 twice using RPE buffer, centrifuge at 10' in the second round
16. place column in a new, clean tube
17. add 120 µl DEPC treated water and let sit for 1' at RT
18. centrifuge at 5000 g for 5', do not discard the elute – it contains the RNA!
19. repeat steps 17 and 18 three more times, centrifuge for 10' at last spin
20. pipette elute in 1.5ml Eppendorf tubes
21. add 960µl 100% ethanol and shake well
22. Add 48µl Natrium Acetate (pH 5.5), vortex
23. let precipitate at -80°C for 30' or at -30°C over night
24. centrifuge at 4°C and 14000 g
25. remove supernatant carefully
26. wash in 75% ethanol and vortex for 10'
27. centrifuge for 30' at 4°C and 14000 g and remove supernatant
28. let the RNA pellet air dry for not more than 30' in RNase-free environment
29. resuspend pellet in DEPC-water
30. let sit at RT for 15', then heat for 10' at 65°C in waterbath
31. store at -80°C or quantify immediately

8.3.4 RNA Amplification

Amplification kit: Amino Allyl MessageAmp™ aRNA; Cat. Nr. 1752, Ambion Inc., Austin, TX, USA

Prior to first-time use:

- add 11.2 ml ethanol to cDNA wash-buffer
- add 22.4ml 100% ethanol to aRNA wash-buffer

8.3.4.1 cDNA synthesis and purification:

1. place RNA in sterile, RNase-free Eppendorf tube and add 1µl T7 oligo(dT)-primer
2. Add nuclease-free water to a total volume of 12 µl
3. incubate for 10' at 70°C in thermal cycler

4. centrifuge briefly, place on ice during preparation of next steps
5. add 8µl Master-Mix, pipette up and down, centrifuge briefly and incubate for 2h in thermal cycler (Master-Mix: 2µl 10X first strand buffer; 1 µl ribonuclease inhibitor, 4 µl dNTP-Mix, 1µl reverse transcriptase)
6. centrifuge briefly, place on ice, proceed immediately with 2nd strand synthesis
7. add 80 µl Master-Mix-II, pipette up and down, centrifuge briefly, incubate in thermal cycler at 4°C for 2h (pause possible at -20°C after incubation step) (Master-Mix-II: 67µl nuclease free water, 10µl 10X second strand buffer, 4µl dNTP mix, 2 µl DNA polymerase, 1µl RNase H)
8. before purification, warm up bottle of nuclease free water to 50°C
9. equilibrate cDNA-filter-cartridge with 50µl cDNA binding buffer for 5' at RT
10. add 250µl cDNA binding buffer and mix thoroughly
11. pipette mixture to column and centrifuge for 1' at 10000 g
12. discard eluate and place cartridge back in tube
13. add 500µl cDNA wash-buffer and centrifuge for 1' at 10000 g, discard eluate and centrifuge again for 1' at 10000 g to remove abundant liquids
14. add 9µl nuclease free water (50°C) to centre of cartridge, let sit for 2' at RT and centrifuge at 10000 g for 1.5'
15. repeat step 14, eluate should amount to 14µl. If it does not, fill up with nuclease free water

8.3.4.2 *In vitro* transcription and purification of aRNA

1. Prepare aRNA-Master-Mix: 3µl aaUTP solution (50mM), 12 µl NTP mix (25mM), 3µl UTP solution, 4µl T7 10X reaction buffer, 4µl T7 enzyme mix
2. add aRNA-Master-Mix to 14µl double strand cDNA and incubate for 14h at 37°C
3. add 2µl DNase I, mix and centrifuge briefly (pause possible at -20°C after this step)
4. add 58µl nuclease free water (50°C) and mix thoroughly
5. add 350µl aRNA binding buffer, vortex
6. add 250µl ethanol 100%, vortex
7. place aRNA filter cartridge on an aRNA collection tube and pipette specimen to the centre of the cartridge
8. centrifuge for 1' at 10000 g, discard liquid
9. add 650µl aRNA-washing buffer onto cartridge, centrifuge at 10000 g for 1', discard liquid and repeat centrifugation for 1' at 10000 g
10. place cartridge on new collection tube
11. add 50 µl nuclease free water (50°C), let sit for 2' and centrifuge at 10000g for 1.5'
12. repeat step 11 once
13. aRNA is contained in the elute, can be stored at -80°C

8.3.4.3 *Dye coupling to aRNA*

1. place 5µg amino-allyl-aRNA in nuclease-free microtube and air dry (approximately 5-10', do not overdry!)

2. resuspend in 9 μ l coupling buffer
3. add 4.5 μ l dye (Cy3 or Cy5, respectively) and incubate at RT for 1h in the dark
4. stop reaction with 4.5 μ l 4M hydroxylamine and mix thoroughly
5. incubate for 15' in the dark
6. repeat aRNA purification as described above (steps 5 onwards), RNA should be used immediately

8.3.4.4 aRNA Fragmentation

fragmentation kit: RNA Fragmentation Reagents; Cat. Nr. 8740, Ambion Inc.,
Austin, TX, USA

1. add nuclease free water to a volume of 9 μ l aRNA
2. add 1 μ l fragmentation buffer 10X
3. centrifuge briefly, incubate at 70°C for 15'
4. add 1 μ l stop solution, place on ice or store at -80°C

8.3.5 Microarray Hybridization with aRNA

1. Clean arrays with 190 proof ethanol or 2-propanol. Do not use 200 proof ethanol! Prehybridize arrays in hyb chamber at 42°C for at least 30'; wash for 2' each in molecular grade water and 2-propanol and spin dry for 3' at 650 rpm after prehybridization

prehybridization buffer:

- a. 600 ul BSA (prefiltered)
 - b. 750 ul 20X SSC
 - c. 30 ul 10% SDS
 - d. 1620 µl H₂O
2. combine samples with similar dye concentrations and pipette to microfuge-column on a microfuge tube; spin at 8000 g for 6'
 3. flip column and place it in new tube; spin at 1000 g for 3'
 4. if volume > 9 µl, vacuum dry and add nuclease free water to a final volume of 9 µl
 5. add 1 µl 10X Fragmentation Buffer (Ambion Fragmentation Reagents), vortex, and incubate at 70°C for 15'
 6. Add 1 µl Stop Solution (Ambion Fragmentation Reagents), vortex, and place samples on ice
 7. prepare hybridization buffer:
 - a. 100 ul Formamide (store at RT in dark bottle)
 - b. 100 ul 20X SSC
 - c. 4 ul 10% SDS

vortex thoroughly, spin down briefly and warm up to 48°C

8. prepare arrays and lifterslips. Clean arrays with 190 proof ethanol or 2-propanol. Do not use 200 proof ethanol!
9. Add nuclease free water to each sample to a final volume of 40 µl
10. Denature sample at 90°C for 2' and snap cool on ice
11. Add 40 µl hyb-buffer and mix thoroughly
12. hybridize array with 80 µl buffer/sample mix and incubate for 16h
13. wash arrays:
 1. step (2'): 179 ml water, 20 ml 20X SSC, 1 ml SDS (10%)
 2. step (2'): 190 ml water, 10 ml 20X SSC,
 3. step (30''): 198 ml water, 2ml 20X SSC
14. spin dry at 650 rpm for 3' and scan within 24h

8.3.6 RT-qPCR

8.3.6.1 Reverse Transcription

DNase-I digestion (all reagents from Invitrogen):

1. Add 1µl DNase-I buffer, 1µl DNase-I and 1µg of total RNA and fill up with DEPC-treated water to a final volume of 10µl
2. let sit for 15' at RT
3. add 1µl EDTA to stop reaction
4. incubate for 10' at 65°C

Preprecipitation:

1. add 1.1µl Na-acetate pH 7.3
2. add 30µl ethanol 100%, vortex
3. incubate for 30' at -80°C
4. centrifuge at 15000g for 20' at RT, remove supernatant carefully
5. wash pellet with 200µl ethanol 70% (4°C)
6. centrifuge at RT at 15000g for 5', remove supernatant, let air dry

RT-reaction (all reagents from Invitrogen):

1. redissolve pellet in 8µl DEPC-treated water
2. add 1µl random hexamere (Invitrogen) and 1µl dNTPMix (10mM, Invitrogen)
3. linearize briefly at 65°C (5-10'), then place on ice (>1'), thereafter place in thermal cycler
4. Add cDNA-synthesis-master-mix (volumes given for single reaction):
 - a. 10X RT-buffer: 2µl
 - b. 25mM MgCl₂: 4µl
 - c. 0.1 DTT 2µl
 - d. RNase out 1µl
 - e. SSIII 1µl
5. cycling steps: 10' at 25°C, 50' at 50°C, 5' at 85°C
6. cool down to 4°C (place on ice), centrifuge briefly, add 1µl RNase H, mix, spin down and incubate for 20' at 37°C
7. can store cDNA at -20°C

8.3.6.2 RT-qPCR Amplification and Detection

Hardware: ABI Sequence Detection System 5700

PCR efficiency assessment:

1. prepare 5 dilution steps of template-cDNA (log₂).
2. prepare PCR-Master Mix: 225µl 2X SybrGreen PCR Master Mix (Invitrogen), 9µl Primer-Mix (10µM), 36µl DEPC-water
3. add 15µl PCR-Master-Mix to template

4. perform at least double technical repeats, temperature settings should be used as in quantification experiments (see below)
5. set baseline and threshold according to manufacturer's protocol and calculate ΔCT -values ($CT_{\text{target}} - CT_{\text{house keeping gene}}$).
6. calculate linear regression slope with $y = ax + b$; PCR efficiencies with $a < 0.1$ should yield valid results.

Relative quantification of gene expression:

1. dilute cDNA from reverse transcription using nuclease-free water or DEPC-water according to efficiency testing results (see above) to a volume of 10 μ l for each single PCR reaction (take HGK reactions into account)
2. prepare PCR-Master-Mix: 12.5 μ l SybrGreen PCR Master Mix, 0,75 μ l Primer-Mix (forward and reverse, 1 μ M), 1,75 μ l DEPC-water
3. load 96-well PCR plate with 15 μ l PCR-Master-Mix and 10 μ l diluted cDNA template
4. gently pipette up and down
5. cover with cover foil, centrifuge briefly and place in cycler
6. set reaction steps:
 - a. Start temperature: 50°C
 - b. Stage 1: 95°C for 10'
 - c. Stage 2: 95°C for 15'' and 60 °C for 1', repeat 40 times
7. after end of reaction perform dissociation curves:
 - a. Stage 1: 15'' at 95°C
 - b. Stage 2: 20'' at 60°C
 - c. Ramp Time: 19'59''
 - d. Stage 3: 15'' at 95°C
8. Export data for downstream analyses

8.4 Supplementary Tables and Figure

C No	sex	age	duration of colitis	localisation	inflammation	dysplasia	ploidy	array
diploid patients (n=11)								
D1	f	59	8	ascending	1	0	3	X
				transverse	3	0	3	X
				descending	1	0	3	
				sigmoid	1	0	3	X
D2	m	54	21	ascending	2	0	3	
				transverse	2	i	3	
				descending	1	0	3	X
				sigmoid	2	0	3	X
				rectum	2	i	3	X
D3	m	64	21	transverse	2	0	3	
				descending	1	0	3	X
				sigmoid	2	0	3	X
D4	m	61	21	ascending	1	0	1	X
				transverse	1	0	3	
				descending	1	0	3	X
				rectum	3	i	1	X
D5	f	67	24	ascending	1	0	3	
				descending	1	0	3	X
				sigmoid	1	0	3	X
				rectum	2	0	3	X
D6	m	56	> 20	caecum	1	0	3	X
				ascending	1	0	3	
				descending	1	0	3	X
				sigmoid	0	0	3	
				rectum	3	0	3	X
D7	m	39	16	transverse	1	0	3	X
				descending	2	0	3	
				sigmoid	0	0	3	X
				rectum	0	0	3	
D8	m	31	9	ascending	2	0	3	
				transverse	2	0	1	
				descending	1	0	3	X
				sigmoid	1	0	3	X
				rectum	2	0	3	X
D9	m	43	12	ascending	1	i	3	
				transverse	2	0	3	X
				descending	1	0	3	X
				sigmoid	1	i	3	X
D10	f	57	17	transverse	0	0	1	
				descending	1	0	1	
				sigmoid	1	0	1	X
				rectum	0	0	1	X
D11	f	37	14	transverse	1	0	3	
				descending	1	0	3	X
				sigmoid	2	0	3	X
				rectum	1	0	3	X

aneuploid patients (n=4); supplementary table 1 ctd.								
A1	f	52	33	caecum	1	0	4	X
				ascending	0	0	4	X
				transverse	0	0	4	X
				descending	0	0	4	X
A2	m	20	2	ascending	2	0	4	X
				transverse	2	0	4	X
				descending	2	i	4	X
				sigmoid	3	i	4	X
				rectum	1	i	4	X
A3	f	39	13	ascending	1	0	4	X
				transverse	1	0	4	X
				descending	0	0	4	X
				sigmoid	0	0	4	X
				rectum	1	0	4	
A4	m	34	10	caecum	2	i	4	X
				ascending	2	0	4	X
				descending	2	0	4	X
				sigmoid	2	0	4	
				rectum	2	i	4	X

Supplementary table 1: Patients for part two, overview of mucosa biopsies; in the first column, “Dn” refers to diploid samples, “An” refers to aneuploid samples. Inflammation is rated semiquantitatively as described in the text (see “Methods”), the degree of dysplasia is rated according to Riddell *et al.* as described in the text; ploidy is presented in categories of Auer’s classification, “array” indicates whether for this specific sample microarray analysis was performed. “i” = indefinite for dysplasia.

8.4.1 Group comparisons

In the following, the top 30 genes for microarray based group comparisons are presented. “Down regulated” refers to a gene with increased expression in the second mentioned group, and *vice versa* for “up regulated”. Genes were chosen among significant genes according to fold change. In the following tables “symbol” refers to the abbreviation assigned to the respective gene by the “HUGO Gene Nomenclature Committee (HGNC)”; „Name“ is given as the first row of the gene description assigned to in the Operon® dataset, clipping may occur; „fc“ = fold change, herein expressed as negative or positive expression ratio; p-value refers to the uncorrected p-value prior to multiple test correctation; „fdr“ = false discovery rate after correction with Benjamini and Hochberg’s method; „RefSeq“ presents assigned number for the gene in the NCBI (National Center for Biotechnology Information) reference sequence database; „GB_accesssion“ refers to GenBank accession number of the respective gene (multiple numbers might appear for redundant entrys or protein modifications); „LocusLink“ presents associated code for querying NCBI’s „entrez gene“ database. This table legend can be applied to the following supplementary tables 2 - 9.

Nr.	symbol	Name	fc	p-value	fdr	RefSeq	GB_accession	LocusLink
1	UGT2B7	UGT2B7--UDP glycosyltransferase 2 family	0,39	1,04E-03	0,55	NM_001074	J05428	7364
2	CPB2	CPB2--Carboxypeptidase B2	0,41	5,30E-03	0,55	NM_016413	M75106,AB011969	1361
3	-	Unknown	0,47	3,26E-03	0,55	-	-	-
4	CSPG6	CSPG6--Chondroitin sulfate proteoglycan	0,47	1,46E-04	0,55	NM_005445	AF020043	9126
5	LRRC17	LRRC17--Leucine rich repeat	0,48	1,49E-03	0,55	NM_005824	U32907	10234
6	MAGEB2	MAGEB2--Melanoma antigen, family B, 2	0,48	8,77E-03	0,55	NM_002364	BC026071	4113
7	MPHOSPH1	MPHOSPH1--M-phase phosphoprotein 1	0,48	9,40E-03	0,55	NM_016195	AB033337,AL117496	9585
8	-	Unknown	0,50	7,51E-03	0,55	-	-	-
9	ITGBL1	ITGBL1--Integrin, beta-like 1	0,50	6,97E-03	0,55	NM_004791	AF072752	9358
10	NP_057157	HDGFRP3--Hepatoma-derived growth factor,	0,52	6,07E-03	0,55	NM_016073	AL133102	50810
11	SLC6A15	SLC6A15--Solute carrier family 6	0,52	8,79E-03	0,55	NM_182767	AK001178,BC070040	55117
12	ZIC1	ZIC1--Zic family member	0,52	5,17E-03	0,55	NM_003412	-	7545
13	-	CLGN--Calmequin	0,52	3,30E-03	0,55	NM_004362	D86322	1047
14	-	Unknown	0,54	8,58E-03	0,55	-	-	-
15	NP_612143	LOC148213--Hypothetical protein FLJ31526	0,54	8,79E-04	0,55	NM_138286	AK122869,AK056088	148213
16	NP_659485	MGC29898--Hypothetical protein MGC29898	0,54	6,88E-03	0,55	-	BX647215	-
17	POLR3F	POLR3F--Polymerase (RNA) III	0,55	7,69E-03	0,55	NM_006466	BC016761,U93869	10621
18	-	Unknown	0,55	7,38E-03	0,55	XR_000211	-	158177
19	-	Unknown	0,56	1,49E-03	0,55	-	-	-
20	C6orf139	C6orf139--Chromosome 6 open reading fram	0,57	8,94E-03	0,55	NM_018132	BC072444	55166
21	-	GLYCOPHORIN B PRECURSOR (PAS-3)	0,57	8,35E-03	0,55	-	-	-
22	-	Unknown	0,57	8,98E-03	0,55	-	-	-
23	NMU	NMU--Neuromedin U	0,57	8,56E-03	0,55	NM_006681	X76029,BF034907	10874
24	USP25	USP25--Ubiquitin specific protease 25	0,58	9,85E-03	0,55	-	AF170562	29761
25	EEF1A1	EEF1A1--Eukaryotic transl elong factor1A1	0,59	7,18E-04	0,55	-	-	-
26	AFG3L1	AFG3L1--AFG3 ATPase family gene 3-like 1	0,59	3,47E-03	0,55	NM_001132	AK056488,AF329691	172
27	NP_775928	FLJ31139--Hypothetical protein FLJ31139	0,60	1,64E-03	0,55	NM_173657	AK055701	285315
28	MS4A3	MS4A3--Membrane-spanning 4-domains	0,60	1,07E-03	0,55	NM_006138	L35848	932
29	-	Unknown	0,60	2,35E-03	0,55	-	-	-
30	-	Unknown	0,60	7,45E-03	0,55	-	-	-

Supplementary table 2: Top 30 down regulated DEGs between diploid mucosa and aneuploid mucosa; for legend please refer to page 84

Nr.	symbol	name	fc	p-value	fdr	RefSeq	GB_accession	LocusLink
1	HBB	HBD--Hemoglobin, delta	3,35	2,49E-03	0,55	NM_000518	BM811415	3043
2	-	C7--Complement component 7	1,89	5,29E-03	0,55	-	J03507	730
3	-	Unknown	1,87	1,05E-04	0,55	-	BC028245	-
4	IER3	IER3--Immediate early response 3	1,76	5,56E-03	0,55	NM_003897	-	8870
5	C20orf31	C20orf31--Chromosome 20 open reading fr.31	1,75	8,42E-03	0,55	NM_018217	AK001645,BC001371	55741
6	LAMC2	LAMC2--Laminin 5, gamma 2	1,55	2,27E-03	0,55	NM_005562	-	3918
7	VAV2	VAV2--Vav 2 oncogene	1,49	3,75E-03	0,55	NM_003371	BX640754	7410
8	SLC28A2	SLC28A2--Solute carrier family 28	1,49	5,86E-03	0,55	NM_004212	U84392	9153
9	DUSP1	DUSP1--Dual specificity phosphatase 1	1,46	9,31E-04	0,55	NM_004417	AK127679	1843
10	VAV2	VAV2--Vav 2 oncogene	1,42	1,15E-04	0,55	NM_003371	BX640754	7410
11	IDH3B	IDH3B--Isocitrate dehydrogenase 3 (NAD+)	1,40	9,86E-04	0,55	NM_174856	AK001905,BQ051868	3420
12	DTR	DTR--Diphtheria toxin receptor	1,40	8,93E-03	0,55	NM_001945	M60278,BC033097	1839
13	CPEB3	CPEB3--Cytoplasmic polyadenylation elem.	1,40	5,67E-03	0,55	NM_014912	AB023157	22849
14	UBE2H	UBE2H--Ubiquitin-conjugating enzyme E2H	1,40	4,53E-04	0,55	NM_182697	Z29331	7328
15	TNFAIP3	TNFAIP3--Tumor necrosis factor, alpha-ind.	1,40	9,28E-03	0,55	NM_006290	M59465,BC064689	7128
16	TGOLN2	TGOLN2--Trans-golgi network protein 2	1,37	2,81E-03	0,55	NM_006464	BX640868,AF027516	10618
17	NR4A1	NR4A1--Nuclear receptor subfamily 4	1,35	3,07E-03	0,55	NM_002135	AK131566	3164
18	RAI14	RAI14--Retinoic acid induced 14	1,35	6,79E-03	0,55	NM_015577	AB037755	26064
19	LGALS9	LGALS9--Lectin, galactoside-binding	1,35	4,96E-03	0,55	NM_002308	AK126017	3965
20	TUFT1	TUFT1--Tuftelin 1	1,34	7,35E-03	0,55	NM_020127	AF254260	7286
21	UPF2	UPF2--UPF2 regulator of nonsense transcr.	1,34	3,12E-03	0,55	-	-	-
22	GPR108	GPR108--G protein-coupled receptor 108	1,33	5,94E-03	0,55	XM_290854	AL365404	56927
23	RELB	RELB--V-rel reticuloendotheliosis viral.	1,33	6,70E-03	0,55	NM_006509	M83221	5971
24	-	Unknown	1,31	2,30E-03	0,55	-	BC008757,BC007916	-
25	CASKIN2	CASKIN2--CASK interacting protein 2	1,31	7,30E-03	0,55	NM_020753	AB032965,BC066643	57513
26	CX3CL1	CX3CL1--Chemokine (C-X3-C motif) ligand	1,31	6,62E-03	0,55	NM_002996	U84487	6376
27	USP33	USP33--Ubiquitin specific protease 33	1,30	8,58E-03	0,55	NM_201626	AB029020,AF383172	23032
28	-	Unknown	1,30	1,58E-03	0,55	-	AK093055	-
29	KCNA2	KCNA2--Potassium voltage-gated channel	1,29	5,12E-03	0,55	NM_004974	L02752	3737
30	CENTA1	CENTA1--Centaurin, alpha 1	1,29	1,10E-03	0,55	NM_006869	AF082324	11033

Supplementary table 3: Top 30 up regulated DEGs between diploid mucosa and aneuploid mucosa; for legend please refer to page 84

Nr.	symbol	name	fc	p-value	fdr	RefSeq	GB_accession	LocusLink
1	SI	SI--Sucrase-isomaltase (alpha-glucosidas	7,14	4,47E-03	0,082	NM_001041	X63597	6476
2	ABCG2	ABCG2--ATP-binding cassette, sub-family	4,76	5,74E-04	0,039	NM_004827	AF103796	9429
3	NP_733746	CSE-C--Cytosolic sialic acid 9-O-acety	4,76	1,07E-03	0,050	NM_170601	AL137496	54414
4	NP_996663	Unknown	4,55	2,93E-03	0,071	NM_206832	AY358153	388364
5	ENPP3	ENPP3--Ectonucleotide pyrophosphatase/ph	4,35	5,65E-03	0,089	NM_005021	AK024899	5169
6	NP_694576	FLJ32063--Hypothetical protein FLJ32063	4,35	1,16E-03	0,051	NM_153031	AK123639	150538
7	UGP2	UGP2--UDP-glucose pyrophosphorylase 2	4,35	1,67E-03	0,057	NM_006759	-	7360
8	SGK2	SGK2--Serum/glucocorticoid regulated kin	4,17	3,20E-03	0,072	NM_170693	BC006523	10110
9	TFCP2L1	TFCP2L1--Transcription factor CP2-like 1	3,57	5,84E-03	0,091	NM_014553	BC064698	29842
10	CASP9	CASP9--Caspase 9, apoptosis-related cyst	2,86	5,38E-07	0,003	NM_032996	U60521	842
11	NP_061183	SLC30A10--Solute carrier family 30 (zinc	2,70	1,35E-03	0,053	NM_018713	BC036078	55532
12	-	ODZ3--Odz, odd Oz/ten-m homolog 3 (Droso	2,70	6,12E-04	0,040	XM_371717	AK001336	55714
13	NP_775761	LOC134285--Hypothetical protein LOC13428	2,63	1,97E-03	0,061	NM_173490	BC018083	134285
14	HOOK1	HOOK1--Hook homolog 1 (Drosophila)	2,44	8,87E-04	0,046	NM_015888	BC011621	51361
15	SLC4A4	SLC4A4--Solute carrier family 4, sodium	2,44	1,39E-04	0,023	NM_003759	AF007216	8671
16	LDHD	LDHD--Lactate dehydrogenase D	2,38	1,84E-03	0,059	NM_194436	BC040279	197257
17	ESPN	ESPIN	2,38	3,80E-03	0,077	NM_031475	AL136880	83715
18	AMACR	AMACR--Alpha-methylacyl-CoA racemase	2,38	4,15E-03	0,080	NM_203382	CR616479	23600
19	AQP11	AQP11--Aquaporin 11	2,33	6,03E-03	0,092	NM_173039	BC040443	282679
20	NP_940884	Unknown	2,33	8,63E-06	0,008	NM_201594	AY517502	284948
21	FRMD1	Unknown	2,33	5,02E-04	0,037	-	AL133077	79981
22	NP_689835	FLJ35954--Hypothetical protein FLJ35954	2,27	7,35E-04	0,043	NM_152622	BX537798	166968
23	O94873	ProSAPiP2--ProSAPiP2 protein	2,27	1,01E-03	0,049	NM_014726	AB018318	9755
24	SLC17A7	SLC17A7--Solute carrier family 17 (sodiu	2,27	2,58E-07	0,003	NM_020309	AB032436	57030
25	HRASLS2	HRASLS2--HRAS-like suppressor 2	2,27	5,75E-03	0,090	NM_017878	AK000563	54979
26	NP_061140	MESP1--Mesoderm posterior 1	2,22	4,92E-03	0,085	NM_018670	AL357535	55897
27	-	Unknown	2,22	5,47E-05	0,017	-	-	-
28	NP_899063	PR1--Voltage-dependent calcium channel g	2,17	4,18E-03	0,080	NM_183240	BC046362	140738
29	NP_079490	SE57-1--CTCL tumor antigen se57-1	2,17	1,42E-03	0,054	NM_025214	AF273051	80323
30	NP_075384	FLJ12949--Hypothetical protein FLJ12949	2,13	2,87E-03	0,071	-	-	-

Supplementary table 4: Top 30 up regulated DEGs between diploid mucosa and UCCs; for legend please refer to page 84

Nr.	symbol	name	fc	p-value	fdr	RefSeq	GB_accession	LocusLink
1	GEM	GEM--GTP binding protein overexpressed i	0,14	1,06E-03	0,050	NM_005261	CR603445	2669
2	RAB31	RAB31--RAB31, member RAS oncogene family	0,15	4,45E-03	0,082	NM_006868	AF183421	11031
3	-	IL8--Interleukin 8	0,17	8,24E-04	0,045	-	M17017	3576
4	SMARCA1	SMARCA1--SWI/SNF related	0,17	6,84E-03	0,095	NM_003069	-	6594
5	NP_955806	URB--Steroid sensitive gene 1	0,19	6,41E-03	0,094	NM_199511	AY548106	151887
6	CYR61	CYR61--Cysteine-rich, angiogenic inducer	0,19	4,04E-03	0,079	NM_001554	Y11307	3491
7	NP_057157	HDGFRP3--Hepatoma-derived growth factor	0,22	6,03E-03	0,092	NM_016073	AL133102	50810
8	NP_689986	LOC116211--Hypothetical protein BC013113	0,23	1,27E-03	0,052	NM_152773	-	255758
9	-	HETEROG. NUCLEAR RIBONUCLEOPROT.	0,25	6,28E-04	0,040	XM_496344	CR599515	440563
10	JAZ1	JAZF1--Juxtaposed with another zinc finger1	0,25	7,38E-03	0,099	NM_175061	AK091311	221895
11	PLS3	PLS3--Plastin 3 (T isoform)	0,26	7,12E-03	0,097	NM_005032	BC056898	5358
12	DFNA5	DFNA5--Deafness, autosomal dominant 5	0,27	7,04E-03	0,096	NM_004403	BX647389	1687
13	-	NUCLEOSOME ASSEMBLY PROTEIN 1-LIKE	0,28	3,05E-03	0,071	-	-	-
14	-	CXXC5--CXXC finger 5	0,28	2,14E-04	0,026	-	-	-
15	-	C14orf78--Chromosome 14 open reading frame	0,28	2,22E-03	0,064	XM_290629	BC011859	113146
16	C20orf53	C20orf53--BA353C18.4 (NOVEL PROTEIN)	0,28	1,28E-04	0,022	-	-	-
17	-	Unknown	0,28	5,91E-03	0,091	-	-	-
18	LPL	LPL--Lipoprotein lipase	0,29	5,83E-03	0,091	NM_000237	M15856	4023
19	WASPIP	WASPIP--Wiskott-Aldrich syndrome protein	0,29	5,63E-03	0,089	NM_003387	-	7456
20	MEIS2	MEIS2--Meis1, myeloid ecotropic	0,30	3,94E-03	0,078	NM_170674	AK001298	4212
21	ITGBL1	ITGBL1--Integrin, beta-like 1 (with EGF-	0,30	1,98E-03	0,061	NM_004791	AF072752	9358
22	ROPN1	ROPN1--Ropporin	0,30	3,91E-03	0,078	XM_042178	-	152015
23	MAGEB2	MAGEB2--Melanoma antigen, family B, 2	0,30	4,55E-03	0,083	NM_002364	BC026071	4113
24	DDX53	DDX43--DEAD (Asp-Glu-Ala-Asp) box polype	0,30	6,05E-03	0,092	NM_018665	AJ278110	55510
25	CLIC4	CLIC4--Chloride intracellular channel 4	0,31	3,18E-03	0,072	NM_013943	AL117424	25932
26	NP_061060	GALNACT-2--Chondroitin sulfate GalNAcT-2	0,31	5,96E-04	0,040	NM_018590	BX647369	55454
27	-	Unknown	0,31	4,88E-03	0,085	-	-	-
28	-	CAMP-RESPONSIVE ELEMENT MODULATOR	0,31	4,92E-03	0,085	-	-	-
29	CI21	Unknown	0,31	7,99E-04	0,044	XM_114685	-	195827
30	SAA4	SAA4--Serum amyloid A4, constitutive	0,31	1,03E-03	0,049	NM_006512	M81349	6291

Supplementary table 5: Top 30 down regulated DEGs between diploid mucosa and UCCs; for legend please refer to page 84

<i>Nr.</i>	<i>symbol</i>	<i>name</i>	<i>fc</i>	<i>p-value</i>	<i>fdr</i>	<i>RefSeq</i>	<i>GB_accession</i>	<i>LocusLink</i>
1	NP_733746	CSE-C--Cytosolic sialic acid 9-O-acetyle	4,46	2,33E-03	0,08	NM_170601	AL137496	54414
2	NP_694576	FLJ32063--Hypothetical protein FLJ32063	4,31	1,60E-03	0,07	NM_153031	AK123639	150538
3	UGP2	UGP2--UDP-glucose pyrophosphorylase 2	3,37	3,59E-03	0,10	NM_006759	-	7360
4	-	ODZ3--Odz, odd Oz/ten-m homolog 3	3,10	1,33E-04	0,03	XM_371717	AK001336	55714
5	ABCG2	ABCG2--ATP-binding cassette, sub-family	3,09	6,64E-04	0,05	NM_004827	AF103796	9429
6	CASP9	CASP9--Caspase 9, apoptosis-related cyst	2,77	5,18E-07	0,00	NM_032996	U60521	842
7	SLC4A4	SLC4A4--Solute carrier family 4, sodium	2,33	8,57E-05	0,02	NM_003759	AF007216	8671
8	SLC17A7	SLC17A7--Solute carrier family 17	2,25	4,37E-08	0,00	NM_020309	AB032436	57030
9	-	Unknown	2,21	2,61E-06	0,003	-	-	-
10	LDHD	LDHD--Lactate dehydrogenase D	2,21	2,48E-03	0,09	NM_194436	BC040279	197257
11	NP_940884	Unknown	2,16	6,01E-06	0,01	NM_201594	AY517502	284948
12	HOOK1	HOOK1--Hook homolog 1 (Drosophila)	2,15	1,27E-03	0,07	NM_015888	BC011621	51361
13	FSIP1	FSIP1--Fibrous sheath interacting protein	2,13	3,64E-04	0,04	NM_152597	BC045191	161835
14	NP_079490	SE57-1--CTCL tumor antigen se57-1	2,09	1,34E-03	0,07	NM_025214	AF273051	80323
15	AMACR	AMACR--Alpha-methylacyl-CoA racemase	2,06	3,71E-04	0,04	NM_203382	CR616479	23600
16	UBN1	UBN1--Ubinuclein 1	2,05	4,65E-04	0,05	NM_016936	AF108461	29855
17	NP_689835	FLJ35954--Hypothetical protein FLJ35954	2,05	1,56E-03	0,07	NM_152622	BX537798	166968
18	HIST1H4E	HIST4H4--Histone 4, H4	2,00	2,66E-06	0,00	NM_003542	-	8364
19	FRMD1	FRMD1--FERM domain containing 1	1,96	1,64E-04	0,03	-	AK074110	79981
20	ABCC13	ABCC13--ATP-binding cassette, sub-family	1,92	6,60E-07	0,001	NM_172024	AY063515	150000
21	Q8NHW1	ENVERIN-2	1,85	5,85E-05	0,02	-	-	-
22	GH1	GH1--Growth hormone 1	1,82	1,02E-03	0,06	NM_022559	CD513545	2688
23	ZNF575	Unknown	1,81	6,81E-04	0,05	NM_174945	AK057129	284346
24	AP3S2	AP3S2--Adaptor-related protein complex 3	1,79	1,13E-04	0,03	NM_005829	BC002785	10239
25	OLFM4	OLFM4--Olfactomedin 4	1,79	3,68E-03	0,10	NM_006418	AY358567	10562
26	NXN	NXN--Nucleoredoxin	1,78	1,29E-05	0,01	NM_022463	AK027451	64359
27	-	Unknown	1,76	2,63E-03	0,09	-	-	-
28	ENTPD5	ENTPD5--Ectonucleoside triphosphate	1,75	1,21E-03	0,07	NM_001249	AF039918	957
29	MMS19L	MMS19L--MMS19-like (MET18 homolog)	1,74	1,73E-06	0,00	-	AF319947	64210
30	-	Unknown	1,71	2,91E-04	0,04	-	AK097622	-

Supplementary table 6: Top 30 up regulated DEGs between UC mucosa and UCCs; for legend please refer to page 84

<i>Nr.</i>	<i>symbol</i>	<i>name</i>	<i>fc</i>	<i>p-value</i>	<i>fdr</i>	<i>RefSeq</i>	<i>GB_accession</i>	<i>LocusLink</i>
1	CYR61	CYR61--Cysteine-rich, angiogenic inducer	0,16	2,92E-03	0,09	NM_001554	Y11307	3491
2	GEM	GEM--GTP binding protein	0,18	1,43E-03	0,07	NM_005261	CR603445	2669
3	NP_689986	LOC116211--Hypothetical protein	0,26	3,38E-03	0,10	NM_152773	-	255758
4	-	IL8--Interleukin 8	0,27	1,86E-03	0,08	-	M17017	3576
5	-	CXXC5--CXXC finger 5	0,28	3,08E-04	0,04	-	-	-
6	-	C14orf78—Chr. 14 open reading frame	0,29	2,75E-03	0,09	XM_290629	BC011859	113146
7	NP_061060	GALNACT-2--Chondroitin sulfate GalNAcT2	0,31	1,38E-03	0,07	NM_018590	BX647369	55454
8	C20orf53	C20orf53--BA353C18.4	0,33	5,44E-04	0,05	-	-	-
9	CBR3	CBR3--Carbonyl reductase 3	0,33	3,10E-03	0,09	NM_001236	AB041012	874
10	NP_598407	ZAK--Sterile alpha motif and leucine zipper	0,34	3,16E-03	0,09	NM_133646	AF238255	51776
11	-	Unknown	0,34	1,68E-03	0,07	-	-	-
12	SAA4	SAA4--Serum amyloid A4, constitutive	0,34	2,59E-03	0,09	NM_006512	M81349	6291
13	CI21_HUMAN	Unknown	0,35	1,85E-03	0,08	XM_114685	-	195827
14	-	Unknown	0,36	3,06E-03	0,09	-	-	-
15	-	Unknown	0,36	1,59E-03	0,07	XM_210365	-	284288
16	RANBP5	RANBP5--RAN binding protein 5	0,37	2,40E-03	0,09	NM_002271	-	3843
17	USP36	USP36--Ubiquitin specific protease 36	0,38	4,93E-05	0,02	NM_025090	AB040886	57602
18	-	MGC12458--Hypothetical protein	0,38	2,15E-04	0,03	NM_032328	AK090927	84288
19	SUPT3H	SUPT3H--Suppressor of Ty 3 homolo	0,39	3,16E-03	0,09	NM_181356	BC050384	8464
20	FAIM	FAIM--Fas apoptotic inhibitory molecule	0,40	2,20E-03	0,08	NM_018147	AK001444	55179
21	PTPRC	PTPRC--Protein tyrosine phosphatase	0,40	2,58E-03	0,09	NM_002838	Y00062	5788
22	SELM_HUMAN	Unknown	0,40	1,35E-03	0,07	NM_080430	AY043487	140606
23	LOXL1	LOXL1--Lysyl oxidase-like 1	0,41	3,82E-03	0,10	NM_005576	BC068542	4016
24	SGCE	SGCE--Sarcoglycan, epsilon	0,41	3,68E-03	0,10	NM_003919	CR622102	8910
25	-	OAS2—2'-5'-oligoadenylate synthetase 2	0,41	1,01E-03	0,06	-	-	-
26	CPVL	CPVL--Carboxypeptidase, vitellogenic-like	0,42	3,66E-03	0,10	NM_031311	AK124472	54504
27	RLF	RLF--Rearranged L-myc fusion sequence	0,42	4,44E-04	0,04	NM_012421	U22377	6018
28	MPP6	MPP6--Membrane protein, palmitoylated 6	0,43	1,28E-04	0,03	NM_016447	AL136836	51678
29	-	HNR	0,43	1,18E-03	0,07	XM_496344	CR599515	440563
30	PSPC1	PSPC1--Paraspeckle component 1	0,43	9,19E-04	0,06	-	-	-

Supplementary table 7: Top 30 down regulated DEGs between UC mucosa and UCCs; for legend please refer to page 84; please note that CYR61 appears on top of this list and was validated for the comparison of aneuploid mucosa and UCC, while the other gene validated according to high fold changes (SMARCA1) was not significant for the comparison presented in this table.

Nr.	symbol	name	fc	p-value	fdr	RefSeq	GB_accession	LocusLink
1	NP_056198	MGC8902--Hypothetical protein MGC8902	6,9	0,0003502	0,01	NM_173638	AK126734	284565
2	NP_056198	Unknown	6,53	0,000186088	0,01	NM_173638	AK126734	284565
3	DCN	DCN--Decorin	6,47	0,000134477	0,01	NM_133504	-	1634
4	VIP	VIP--Vasoactive intestinal peptide	6,47	0,013971307	0,09	NM_003381	M36634	7432
5	NP_899228	Unknown	5,74	0,000807589	0,02	XM_496399	AF131738	440675
6	NP_899228	Unknown	5,6	5,42728E-06	0,00	NM_183372	AL832622	200030
7	-	Unknown	5,5	0,000681885	0,02	XM_496394	-	440673
8	GNA13	GNA13--Guanine nucleotide binding protein	5,14	0,003482352	0,05	NM_006572	BC036756	10672
9	GPNMB	GPNMB--Glycoprotein (transmembrane)	4,62	0,000733307	0,02	NM_002510	X76534	10457
10	FOS	FOS--V-fos FBJ murine osteosarcoma viral	4,62	0,011006031	0,08	NM_005252	BX647104	2353
11	EIF4A2	EIF4A2--Eukaryotic translation init. fact.	4,61	0,003558931	0,05	NM_001967	AL117412	1974
12	SGK	SGK--Serum/glucocorticoid regulated kinase	4,55	0,004392283	0,05	NM_005627	BX649005	6446
13	HSPA1L	HSPA8--Heat shock 70kDa protein 8	4,47	0,002641295	0,04	NM_006597	BC016179	3312
14	-	C7--Complement component 7	4,29	0,009668333	0,07	-	J03507	730
15	TCERG1	TCERG1--Transcription elongation regulator	4,25	0,01438019	0,09	NM_006706	AF017789	10915
16	AHNAK	AHNAK (desmoyokin)	4,1	0,001227006	0,03	-	M80899	195
17	SCGN	SCGN--Secretagogin	4,09	0,002933551	0,04	NM_006998	Y16752	10590
18	ANKHD1	ANKHD1	4,05	0,000570139	0,02	NM_017747	AF521883	54882
19	ZFP36L2	ZFP36L2--Zinc finger protein 36, C3H	4,02	0,003584928	0,05	NM_006887	-	678
20	ACAT1	ACAT1	3,97	0,014421809	0,09	NM_000019	BC063853	38
21	COPEB	COPEB--Core promoter elem. binding prot.	3,94	0,001101745	0,03	NM_001008490	U51869	1316
22	ASAH1	ASAH1--N-acylsphingosine amidohydrolase	3,89	0,001417662	0,03	NM_004315	AK025732	427
23	GPM6B	GPM6B--Glycoprotein M6B	3,81	8,63336E-07	0,001	NM_005278	-	2824
24	-	Unknown	3,75	4,58685E-05	0,01	-	-	-
25	PCOLCE2	PCOLCE2--Procollagen C-endopeptidase	3,74	0,000338394	0,01	NM_013363	AY358557	26577
26	SFRP2	SFRP2--Secreted frizzled-related protein	3,7	7,48557E-05	0,01	NM_003013	AF311912	6423
27	PDE9A	PDE9A--Phosphodiesterase 9A	3,7	0,002171974	0,04	NM_001001583	AK127770	5152
28	-	Unknown	3,69	1,8362E-06	0,00	XM_496394	BM911099	440673
29	SFRP1	SFRP1--Secreted frizzled-related protein	3,69	5,22989E-05	0,01	NM_003012	AF056087	6422
30	PDLIM3	PDLIM3--PDZ and LIM domain 3	3,66	0,004421042	0,05	NM_014476	BX648290	27295

Supplementary table 8: Top 30 up regulated DEGs between normal controls and UC mucosa; for legend please refer to page 84; genes 1 and 2, and 5 and 6 were printed with two different clones yielded highly similar results.

Nr.	symbol	name	fc	p-value	fdr	RefSeq	GB_accession	LocusLink
1	TRIM7	TRIM7--Tripartite motif-containing 7	0,31	8,88E-08	0,001	NM_033342	-	81786
2	C6orf117	C6orf117--Chromosome 6 open reading frame	0,36	2,12E-04	0,01	NM_138409	AK090775	112609
3	HYAL1	HYAL1--Hyaluronoglucosaminidase 1	0,38	2,02E-03	0,03	NM_007312	U03056	3373
4	TRIM29	TRIM29--Tripartite motif-containing 29	0,44	2,57E-06	0,001	NM_058193	L24203	23650
5	GPT2	GPT2--Glutamic pyruvate transaminase	0,47	3,88E-03	0,05	NM_133443	AY029173	84706
6	ABCG5	ABCG5--ATP-binding cassette, sub-family	0,47	6,09E-04	0,02	NM_022436	AF312715	64240
7	SOD2	SOD2--Superoxide dismutase 2	0,49	2,82E-03	0,04	NM_000636	-	6648
8	PF4	PF4--Platelet factor 4	0,50	1,34E-03	0,03	NM_002619	M25897	5196
9	OLFM4	OLFM4--Olfactomedin 4	0,51	1,04E-02	0,08	NM_006418	AY358567	10562
10	-	COP--CARD only protein	0,51	5,78E-03	0,06	NM_052889	-	114769
11	TRIM29	TRIM29--Tripartite motif-containing 29	0,51	1,94E-03	0,03	NM_012101	BX648072	23650
12	MS4A3	MS4A3--Membrane-spanning 4-domains	0,51	1,54E-02	0,10	NM_006138	L35848	932
13	SLC4A11	SLC4A11--Solute carrier family 4, sodium	0,51	5,56E-05	0,01	NM_032034	AF336127	83959
14	FABP6	FABP6--Fatty acid binding protein 6	0,51	3,09E-05	0,004	NM_001445	X90908	2172
15	BRUNOL4	BRUNOL4--Bruno-like 4, RNA binding prot.	0,52	2,29E-03	0,04	NM_020180	-	56853
16	SLC6A14	SLC6A14--Solute carrier family 6	0,53	5,78E-03	0,06	NM_007231	AF151978	11254
17	AQP11	AQP11--Aquaporin 11	0,53	2,44E-03	0,04	NM_173039	BC040443	282679
18	CARD6	CARD6--Caspase recruitment domain family	0,53	1,39E-02	0,09	NM_032587	AF356193	84674
19	SMARCB1	SMARCB1--SWI/SNF related	0,54	6,86E-04	0,02	NM_003073	AK024025	6598
20	MRPL52	Unknown	0,54	9,84E-03	0,08	NM_181306	-	122704
21	-	F12--Coagulation factor XII	0,55	5,34E-03	0,06	NM_000505	AB095845	2161
22	KLK11	KLK11--Kallikrein 11	0,56	9,04E-04	0,02	NM_144947	AB041036	11012
23	-	MGC42951--MGC42951 gene	0,56	4,61E-07	0,001	-	BC031958	414926
24	NP_872440	Unknown	0,56	3,33E-03	0,04	NM_182634	-	349152
25	-	Unknown	0,57	8,43E-05	0,01	XM_378738	AK095347	400643
26	-	KYNURENINASE	0,57	7,15E-03	0,06	-	-	-
27	KIAA1199	KIAA1199--KIAA1199	0,58	6,69E-03	0,06	NM_018689	AB033025	57214
28	-	Unknown	0,58	8,59E-03	0,07	-	-	-
29	ASS	ASS--Argininosuccinate synthetase	0,58	6,17E-03	0,06	NM_000050	-	445
30	C6orf57	C6orf57--Chromosome 6 open reading	0,58	2,93E-03	0,04	NM_145267	BU598152	135154

Supplementary table 9: Top 30 down regulated DEGs between normal controls and UC mucosa; for legend please refer to page 84

9 ACKNOWLEDGEMENTS

Without the support of many people, this thesis would not have been possible.

I am especially grateful to:

PD Dr. Dr. Jens Habermann for his support and friendship throughout the years this thesis was created. Jens, thank you for all the work you have done to organize the exchange with the Riedlab, for establishing the collaborations with Stockholm and Helsinki, and for your helpful advice on how to proceed with ploidy measurements and life in general. You introduced me to the world of science and, even more importantly, to fascinating people from virtually all over the world and I am very grateful for that.

Dr. Thomas Ried for welcoming me wholeheartedly to his laboratory in Bethesda, and for stimulating and humorous discussions not only about genes and German soccer. Without your generosity this thesis would not have been possible.

Prof. Dr. Hans-Peter Bruch for letting me utilize the laboratory facilities in the Laboratory in Lübeck and for generously supporting academic research and innovative ideas.

Dr. Britta Fritzsche, Stefanie Bünger and Timo Gemoll for their persistent help in completing and analyzing PCR experiments and, in particular, for many exhilarating moments at the lab in Lübeck.

Prof. Gert Auer for advice and assistance with the evaluation of DNA histograms.

Dr. Sampsa Hautaniemi and Kari Nousianen for sophisticated knowledge in testing 34580 hypotheses at once and tireless support in the interpretation of array data.

Dr. Bernd Wolfgang Igl for the statistical analyses of ploidy results.

The wonderful people of the Riedlab for their incredible warm welcome and their support and friendship throughout my time in the US. Thank you, Hersed, Nicole, Turid, Linda, Kerstin, Buddy, Tom, Kundan, Amanda, Jordi, Mike, Danny, and all the others for making my stay an unforgettable experience.

The technicians of the lab in Lübeck, Vera Grobleben, Elke Gheribi, Regina Kaatz, Katja Klempt-Gießing, and Carsten Pilcher for a warm welcome to the group and their continuous efforts to support scientific work. I am especially grateful for Gisela Grosser-Pape's tireless help in performing DNA cytometry and Feulgen staining. Even more importantly, I would like to thank you for the virulently optimistic atmosphere you create in your laboratory.

My fellow students in Lübeck, especially Constanze Brucker, Sarah Louise Nause, Henrike Preis, and Arnika Wagner for their friendship and for mutual support towards our common goal. I am particularly grateful for Karl-Frederick Meyer's help with performing Feulgen staining and with making sense out of all the data created.

The Medical Faculty of the University of Lübeck and the German Academic Exchange Service for the financial support to complete this thesis.

I have countless reasons to be grateful to Nanni Schnell. At the end of this thesis I would like to thank you for showing me how to float on the Krumme Lanke without going in circles.

9 CURRICULUM VITAE

November 2011

Marco Gerling

Born November 10th, 1980 in Rahden, Germany

German citizenship

Medical Career:

Since 05/2009: Residency („Assistenzarzt“), Medical Clinic I,
Gastroenterology / Infectious Diseases / Rheumatology,
Charité – Universitätsmedizin Berlin,
Campus Benjamin Franklin (director: Prof. Dr. Martin Zeitz)

University and High School Education:

12/2008: License to Practice Medicine
10/2003 to 12/2008: University of Lübeck, clinical studies, examination
 („Zweite Ärztliche Prüfung“) 12/2008, average grade of 1.0
10/2001 to 09/2003: University of Lübeck, preclinical studies, examination
 („Ärztliche Vorprüfung“) 09/2003, average grade of 1.0
1991 to 2000: Söderblom-Gymnasium Espelkamp, examination („Abitur“)
 06/2000, average grade of 1.8

Abroad Education during Scientific Training:

03/2005 to 03/2007: Several research visits (total of nine months) to the National
 Cancer Institute / NIH, Bethesda, USA (Dr. Thomas Ried):
 Learning and applying RNA-extraction and oligonucleotide-
 microarray-hybridization techniques
Since 01/2005: Several visits to the Department of Oncology & Pathology,
 Cancer Center Karolinska, Stockholm, Sweden (Prof. Gert
 Auer): Learning and applying DNA image cytometry

Clinical Education („Praktisches Jahr“):

- 12/2007 to 02/2008: Interdisciplinary Emergency Department, University Clinic Schleswig-Holstein, Campus Lübeck, Germany
- 10/2007 to 11/2007: Gastroenterology, University of Sydney, Australia
- 06/2007 to 09/2007: Neurology, Pretoria Academic Hospital, South Africa
- 02/2007 to 06/2007: Surgery, Asklepios Klinikum Bad Oldesloe, Germany

Scholarships:

- 05/2010 Travel grant from the „GlaxoSmithKline Stiftung“ for a poster presentation at the *Digestive Disease Week*, New Orleans, USA
- 06/2005 to 12/2006: Scholarship „Experimentelle Medizin“ from the University of Lübeck
- 03/2005 to 05/2005: Scholarship from the German Academic Exchange Service (DAAD) for a research visit at the National Institutes of Health, Bethesda, USA
- 2006 and 2007: Travel grants from the Allianz AG for clinical training in Nigeria and South Africa

10 PUBLICATIONS

Congress Contributions / Abstracts:

“Elucidating Aneuploidy as Prognostic Marker for Ulcerative Colitis-associated Colorectal Carcinogenesis“

Gerling M, Meyer KF, Freitag-Wolf S, Hautaniemi S, Nousiainen K, Gemoll T, Fritzsche B, Bünger S, Schimmelpenning H, Ried T, Auer G, Bruch HP, Roblick UJ, Habermann JK, poster at the 12th Annual Meeting on Surgical Resarch, Freiburg [Langenbecks Arch Surg (2008) 393:759–815]

“Laser capture micro-dissection – an optimized protocol for cytogenetic characterization of colon cancer specimens”

Fritzsche B, Roblick UJ, Bünger S, **Gerling M**, Gemoll T, Gheribi E, Kalies K, Bruch H-P, Habermann JK; poster at the 12th Annual Meeting on Surgical Resarch, Freiburg [Langenbecks Arch Surg (2008) 393:759–815]

„Critical comparison of genomic instability in inflammation-associated colorectal cancer of man and mice“

Gerling M, Habermann JK, Glauen R, Lehr HA, Meyer KF, Bruch HP, Zeitz M, Siegmund B; Poster at the Digestive Disease Week 2010, May 3rd, 2010, New Orleans, USA

„Inflamations-assoziierte Karzinogenese und Chromosomale Instabilität: Validierung eines murinen Modells Colitis-assoziiierter Karzinome“

Gerling M, Habermann JK, Glauen R, Lehr H-A, Bruch H-P, Zeitz M, Siegmund B
Poster at the DGVS meeting, Viszeralmedizin 2010, September 14th, 2010

„Aneuploidie als prädiktiver und prognostischer Faktor für Vorsorge und Therapie von sporadischen und Colitis ulcerosa-assoziierten kolorektalen Karzinomen“

Habermann JK, **Gerling M**, Meyer KF, Fritzsche B, Freitag-Wolf S, Hautaniemi S, Nousiainen K, Krüger S, Schimmelpenning H, Bruch HP, Auer G, Ried T, Roblick UJ
Poster at the DGVS meeting, Viszeralmedizin 2010, September 14th, 2010

Paper:

Parts of this thesis have been published:

“High frequency of aneuploidy defines ulcerative colitis-associated carcinomas – a comparative prognostic study to sporadic colorectal carcinomas”

Gerling M, Meyer KF, Fuchs K, Igl B.W., Fritzsche B, Ziegler A, Bader F, Kujath P, Schimmelpenning H, Bruch H-P, Roblick UJ, Habermann JK

[*Annals of Surgery*; 2010 Jun 4, 252(1):84-89, PMID 20531006]

Based on the findings of this thesis, a related study that does not comprise data presented herein has been published:

“Characterization of chromosomal instability in murine colitis-associated colorectal cancer”

Gerling M, Glauben R, Habermann JK, Kühn AA, Loddenkemper C, Lehr HA, Zeitz M, Siegmund B

[*PLoS ONE*; 2011, 6(7): e22114. doi:10.1371/journal.pone.0022114]