Aus dem Institut für Pathologie
der Medizinischen Fakultät Mannheim
(Direktor: Prof. Dr. med. Alexander Marx)

Genomic Instability in Colorectal Adenomas

with and without Recurrence

Inauguraldissertation
zur Erlangung des Doctor scientiarum humanarum (Dr. sc. hum.)
der
Medizinischen Fakultät Mannheim
der Ruprecht-Karls-Universität
zu
Heidelberg

vorgelegt von
David Fiedler

aus
Berlin-Friedrichshain

2018

Dekan: Prof. Dr. med. Sergij Goerdt
Referent: Prof. Dr. med. Timo Gaiser

Preliminary remark:
Partial results (comprising chapter 3.2-3.5) of the presented thesis have been accepted for publication (first authorship): *International Journal of Cancer*; August 13th, 2018.

ABBREVIATIONS

2N	diploid
3N	near-triploid
4N	near-tetraploid
μ	micro (10^{-6})
aCGH	array-based comparative genomic hybridization
ANCA	average number of copy number alterations
APC	adenomatous polyposis coli
ASN	average signal number
BAC	bacterial artificial chromosome
bisDNA	bisulfite-converted DNA
bp	base pair
BSA	bovine serum albumin
CDK	cyclin-dependent kinase
CGI	CpG island
Chr	chromosome
CI	confidence interval
CIMP	CpG island methylator phenotype
CIN	chromosomal instability
CMS	consensus molecular subtype
CN	copy number
CNA	copy number alteration
CNV	copy number variation
CpG	cytosine-phosphate-guanine oligodeoxynucleotide
CRC	colorectal cancer
Cy3/-5	cyanine fluorescent dye
DAPI	4′,6-diamidino-2-phenylindole
DCIS	ductal carcinoma *in situ*
ddH$_2$O	*aqua bidestillata*
dH$_2$O	*aqua destillata*
DMP	differentially methylated (CpG) position
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DNMT	DNA methyltransferase
dNTP	deoxynucleoside triphosphate (dATP, dCTP, dGTP, dTTP, dUTP)
dsDNA	double-stranded DNA
DSB	double-strand break
e.g.	for example (*exempli gratia*)
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetic acid
EMR	endoscopic mucosal resection
EtBr	ethidium bromide
EtOH	ethanol
FAP	familial adenomatous polyposis
FC	fold change
FDR	false discovery rate
FFPE	formalin-fixed and paraffin-embedded
FISH	fluorescence *in situ* hybridization

For	forward-primer (5'→ 3'-primer)
g	grams
g	acceleration of gravity
gDNA	genomic DNA
GSEA	gene set enrichment analysis
h	hours
H&E	hematoxylin and eosin
HGD	high-grade dysplasia
HM450K	Human Methylation 450K Bead Chip Methylation array
HNPCC	hereditary nonpolyposis colorectal cancer (Lynch syndrome)
i.a.	among others (*inter alia*)
i.e.	that is; in other words (*id est*)
JAK	Janus kinase
l	liter
IEN	intraepithelial neoplasia
IQR	interquartile range (of the median)
kb	kilo-bases
LB	lysogeny broth
LGD	low-grade dysplasia
LINE	long interspersed nuclear element
LOH	loss of heterozygosity
M	molar
Mb	mega-bases
MHC	major histocompatibility complex
min	minutes
miRNA	microRNA
mM	millimolar
MMR	mismatch repair
MSI	microsatellite instability
MSS	microsatellite stability
n	nano (10^{-9})
NaOH	sodium-hydroxide
ND	not determined
NT	nick translation
NTC	non-template control
PAGE	polyacrylamide gel electrophoresis
p	small arm of chromosome (*petit*)
P	p-value or probability value
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PCA	principal component analysis
pH	negative of the base 10 logarithm of the molar concentration of hydrogen ions (*potentia hydrogenii*)
q	long arm of chromosomes
QC	quality control
Rev	reverse-primer (3'→ 5'-primer)
RNA	ribonucleic acid
ROI	region of interest
rpm	revolutions per minute
RQ	relative quantification; fold change
RT	room temperature (20-23°C)

SCNA   somatic copy number alteration
SD   standard deviation (of the mean)
SE   Shannon entropy
sec   seconds
SI   Simpson diversity index
SINE   short interspersed nuclear element
SKY   spectral karyotyping
SSA   sessile serrated adenoma
SSC   saline sodium citrate (buffer)
ssDNA   single-stranded DNA
STAT   signal transducer and activator of transcription
SU   subunit
TMD   transmembrane domain
TSA   traditional serrated adenoma
TSG   tumor suppressor gene
U   units
UICC   Union internationale contre le cancer
ULS   universal linkage system™
VEGF   vascular endothelial growth factor
vs   versus
w/   with
w/o   without
WGA   whole genome amplification
WGD   whole genome duplication
WHO   world health organization

1 INTRODUCTION

1.1 Anatomy and histology of the large intestine

1.1.1 Anatomy of the colon and rectum

The large intestine (*intestinum crassum;* colorectum) is a 100-150 cm long abdominal organ and the final part of the gastrointestinal (GI) tract (Ponz de Leon and Di Gregorio, 2001). As such, it contributes to the digestive process by absorbing water and nutrients (*e.g.*, salts) from the remnants of the predigested digested foods of the upper GI tract. In the colorectal passage, unabsorbed material is fermented by the aid of bacteria prior to the expelling of stool. These tasks are primarily performed from cecum to sigmoid colon (Figure 1), whereas the rectum instead stores fecal matter (Ponz de Leon and Di Gregorio, 2001; Väyrynen, 2016).
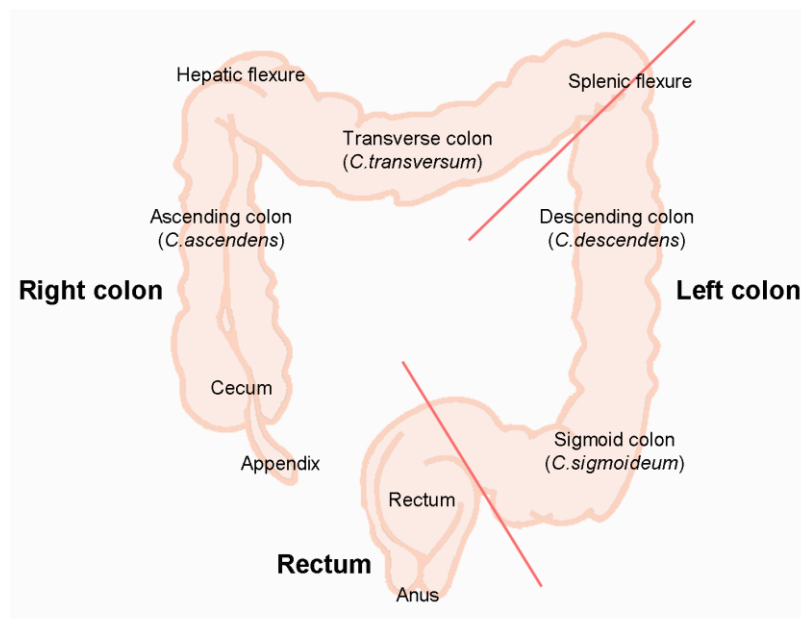
Figure 1. Anatomy of the colorectum. Schematic representation of the anatomical terminology and structure of the large intestine which is subdivided by sections (regular) and sides (**bold**). *Image recreated according to Ponz de Leon and Di Gregorio (2001).*

The large intestine, apart from the rectum, is covered by the serous membrane of the peritoneum anchoring the colon to the abdominal wall. Two artery branches, *i.e.*, superior mesenteric and inferior mesenteric, maintain the blood supply to the colon. This subdivides the organ into two vascular regions: (*i*) from the cecum to the splenic flexure and (*ii*) from the descending colon to the upper rectum (Väyrynen, 2016). However, blood for the lower rectum is derived from the middle and inferior rectal arteries (Ponz de Leon and Di Gregorio, 2001). Thus, the vascularization of the colorectum supports the concept of differentiation between right- and left-sided colon and the

rectum, respectively. Venous blood leaves the colon through the portal vein system into the liver which is the principal organ of hematogenous metastases from colorectal cancers (Ponz de Leon and Di Gregorio, 2001). The unique position occupied by the rectum is underlined by venous blood which leaves this section through the hemorrhoidal plexus anastomosing into the caval vein system, leading to lung metastasis for rectal cancer. It is assumed that three different types of colorectal malignancies may occur varying by localization within the large intestine (Weisburger and Wynder, 1987), namely, cancer of the right colon, cancer of the left colon and cancer of the rectum.

1.1.2 Histology of the colon and rectum

The walls of the colon and rectum are composed of distinct histological layers (Ponz de Leon and Di Gregorio, 2001) (Figure 2). The mucosa constitutes the inner-most (luminal) layer of the colon. It accommodates the colonic crypts and the *lamina propria*, which lies beneath the crypts (Väyrynen, 2016). Moreover, columnar cells shape the surface epithelium which functions as the protective barrier between the host and the gut lumen. Mucosa and subjacent submucosa are separated by the smooth muscle layer of the *muscularis mucosae*. The submucosa contains fibroblasts, vasculature, nerves and a variety of white blood cells which are located beneath strands of collagen tissue (Ponz de Leon and Di Gregorio, 2001). This layer houses a neuronal plexus, *i.e.*, the Meissner plexus, which innervates the *muscularis mucosae*. The inner colon layers are surrounded by a thick muscular layer (*muscularis externa*) which consists of circular and longitudinal muscle tissue (Ponz de Leon and Di Gregorio, 2001). These are innervated by the Auerbach plexus (located in the *tunica muscularis*). The outer-most histological components of the colon are formed by the subserosa and the serosa (intraperitoneal). However, the rectum is surrounded by the adventitia (retroperitoneal).
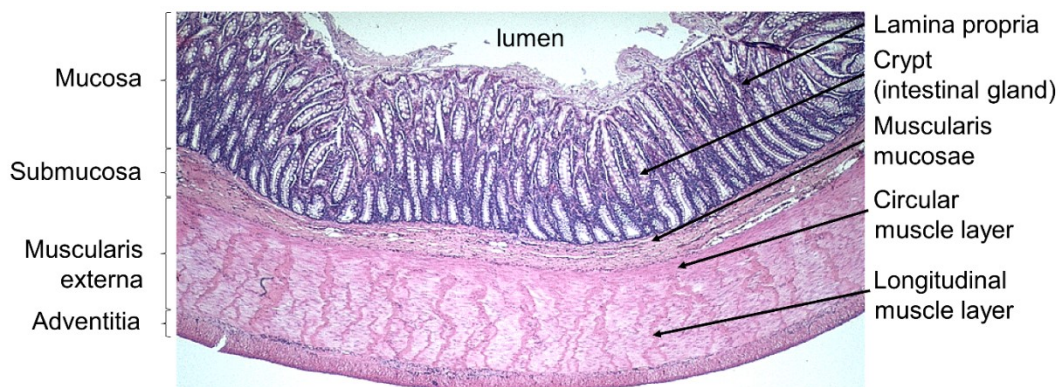
Figure 2. Histopathology of the colorectum. Microscopic image of an H&E stained section from the rectum showing the characteristic histology in the large intestine. *Image modified from Gulwani (2016) with courtesy of PathologyOutlines.com.*

1.1.3 Microanatomy of the colon: epithelial crypts

The anatomy of the colon consists of crypts which are formed by monolayered epithelial cells (Figure 3). The crypts increase the surface of the colorectum by the invaginated architecture. Moreover, the crypts have a polarized topographical organization where the location of a specific cell is closely related to its function (Radtke and Clevers, 2005). At the bottom of the crypt, stem cells give rise to cell types which shape the crypt. Enteroendocrine cells are located at the lower third of the crypt, whereas the well-differentiated goblet cells (mucus-secretion) and absorptive cells (colonocytes) are found in the upper two-thirds of the crypt (Fodde et al., 2001). The cells migrate from the crypt's bottom towards the surface to replenish the colonocyte amassment. Essential to maintain the homeostasis of an efficiently and correctly functioning colon (Barrasa et al., 2013), the senescent colonocytes are exfoliated into the lumen by undergoing a unique form of apoptosis, *i.e.*, anoikis (Aust and Barretton, 2010; Potten et al., 1997). Cell renewal is believed to take three to six days (Cotran et al., 1999). Accordingly, the frequency of mitotic cells equals the rate of apoptotic cells. However, neoplastic lesions of the colorectum are the result of an imbalance of these frequencies towards an accumulation of proliferative cells (Fodde et al., 2001).
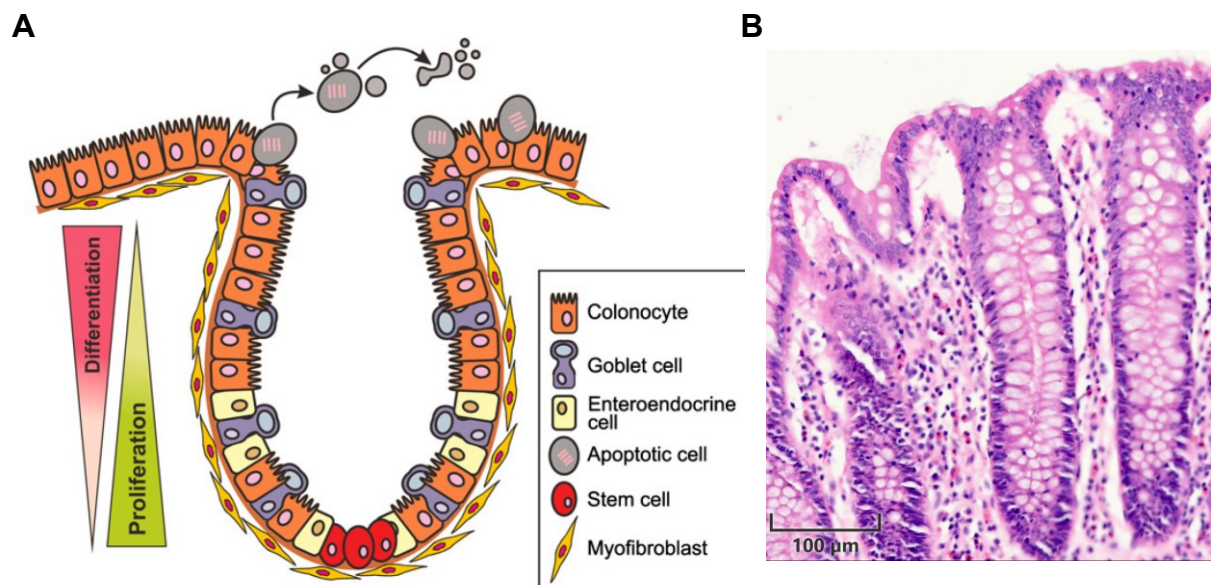
Figure 3. Structure of the colonic crypt. (**A**) Schematic structural architecture of a colonic crypt, which is the basic functional unit of the colon. It consists of different cells varying in proliferation activity and localization. There is an inverse correlation between proliferation and differentiation. Senescent colonocytes undergo apoptosis and are secreted into the lumen of the colon. *Image modified from Barrasa et al. (2013).* (**B**) Histopathology of a regular colonic crypt within normal colon mucosa stained by hematoxylin and eosin. Scale bar, 100 µm.

1.2 Colorectal cancer: epidemiology and etiology

The colorectum is a common cancer site reflected by a lifetime probability of cancer diagnosis of 5.0% in men and 4.3% in women (Hornick and Odze, 2015; Jemal et al., 2008). In Germany, colorectal cancer (CRC) is the third most common cancer in men and the second most common cancer in women with 34,050 and 28,360, respectively, new cases in 2013. Consequently, men have a higher CRC age-standardized incidence rate compared to women (56.6 vs 36.3 cases per 100,000 inhabitants) (Zentrum für Krebsregisterdaten, 2016). Epidemiologic data have shown that CRC incidence is not equally distributed around the world (Figure 4). The (historically) highest incidence rates were reported in long-standing economically developed countries (Center et al., 2009). Albeit, incidence rates in these countries have stabilized and started to decrease due to screening practices, colonoscopic polypectomy and improved treatments (Center et al., 2009; Welch and Robertson, 2016). In contrast, economically transitioning countries still report an incline of CRC incidences, most likely due to the 'westernization' of lifestyles, *e.g.*, obesity, physical inactivity and high-calorie food (Center et al., 2009; Giovannucci, 2002; Popkin, 2015).
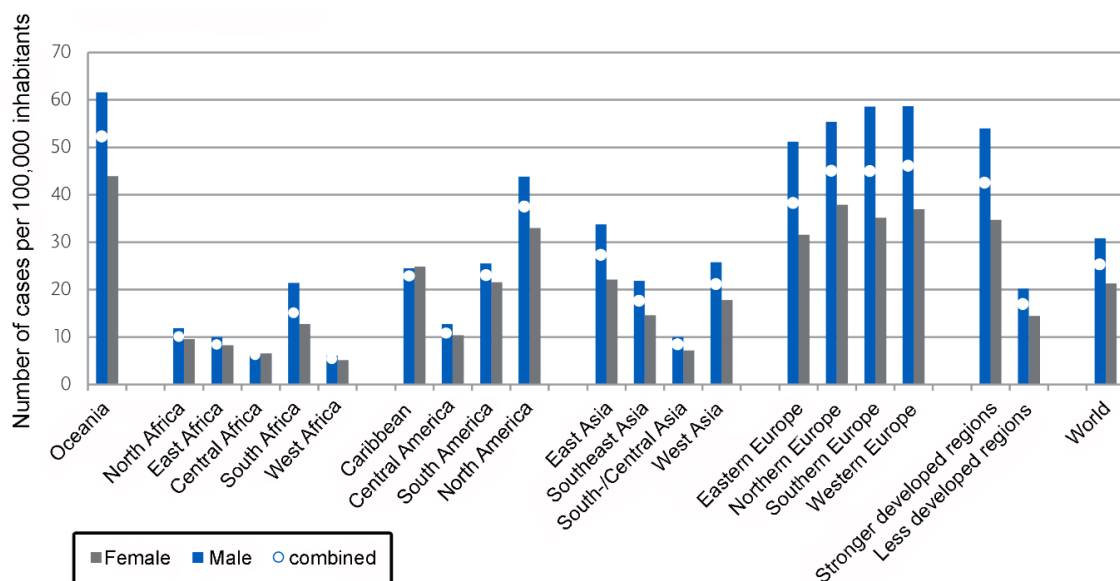
Figure 4. CRC incidences worldwide. Estimated age-standardized incidence rates per 100,000 inhabitants of colorectal cancers distributed by regions of the world and sex. *Extracted and modified from Zentrum für Krebsregisterdaten (2016).* Data was collected by GLOBOCAN 2012 (International Agency for Research on Cancer).

Mortality rates of CRC decreased in the past decades, arguably due to declining incidence rates in developed countries (American Cancer Society, 2017). Disease-related deaths affected 13,609 men and 12,085 women in Germany in 2013. The relative survival rate for CRC is 63% at 5-years following diagnosis, likewise in men and women (Zentrum für Krebsregisterdaten, 2016). Nevertheless, the prognosis of patients

suffering from CRC strongly depends on the tumor stage at the time of diagnosis. This circumstance is emphasized by a 5-year relative survival rate of CRC declining from over 90% in early stages towards 60% in advanced stages and under 10% in metastatic CRC (O'Connell et al., 2004).

Moreover, CRC risk increases with age, with a median age at diagnosis of 70.3 years in men and 72.9 years in women in Germany (Zentrum für Krebsregisterdaten, 2016). Studies have reported that a racial disparity for CRC risk exists, *i.e.*, some ethnicities (*e.g.*, Afroamericans and American natives) are confronted with higher CRC incidence and mortality rates than average (American Cancer Society, 2017; Irby et al., 2006). Aside from age and ethnicity, several risk factors influence CRC formation: (*i*) hereditary and family history, (*ii*) personal medical history and (*iii*) behavioral risk factors (American Cancer Society, 2017).

While approximately 70% of CRC arise sporadically, estimations state that up to 30% of CRC patients have a family history of the disease, approximately 5% of which are caused by known hereditary genetic alterations (Patel and Ahnen, 2012). People with a family history (first-degree relative) have a two-to-four time increased risk for developing the disease compared to people without a family history (Butterworth et al., 2006). Although familial CRC has no apparent association with an identifiable hereditary syndrome so far increasingly more susceptibility genes were and will be identified (Jasperson et al., 2010). Conversely, hereditary CRCs are composed of well-defined inherited syndromes (Figure 5): (*i*) the nonpolyposis phenotype Lynch syndrome (LS); (*ii*) the adenomatous polyposis phenotype including the familial adenomatous polyposis (FAP), the attenuated familial adenomatous polyposis (AFAP) and *MUTYH*-associated polyposis (MAP); (*iii*) and the hamartomatous polyposis phenotype comprising Cowden syndrome (CS), juvenile adenomatous syndrome (JAS), oligopolyposis (OP) and Peutz-Jeghers syndrome (PJS) (Burt, 2000; Jasperson et al., 2010).
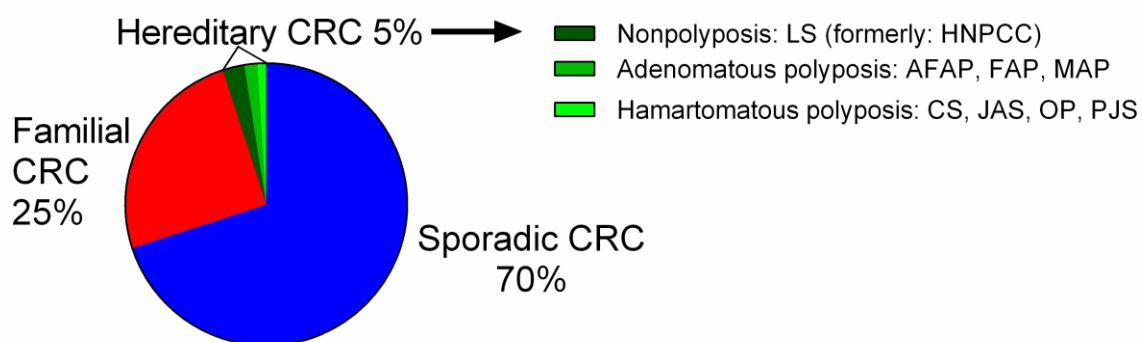
Figure 5. Distribution of CRC types. Fractions and frequencies of CRC arising in various family risk settings. HNPCC, hereditary nonpolyposis colorectal cancer. *Image recreated according to Burt (2000).*

Another important risk factor is the personal medical history: the individual risk for CRC increases in patients which had a history of CRC and/or adenomatous polyps (Mysliwiec et al., 2006; Ren et al., 2016). Moreover, patients suffering from chronic inflammatory bowel disease, *e.g.*, ulcerative colitis and Crohn's disease, have a doubled risk of developing CRC compared to the people of the general population (Lutgens et al., 2013). A meta-analysis revealed a strong association between CRC and type-2 diabetes patients, suggesting this disease is a risk factor for CRC (Tsilidis et al., 2015).

Behavioral risk factors for CRC include (modifiable) parameters traditionally associated with a sedentary lifestyle and Western diet (American Cancer Society, 2017). Consequently, physical inactivity (Schmid and Leitzmann, 2014), overweight and obesity (Ma et al., 2013), consumption of red and/or processed meat (Chan et al., 2011), moderate and heavy alcohol intake (Cho et al., 2004) and smoking (Secretan et al., 2009) negatively affect the individual risk of CRC (Boyle et al., 2012). Nevertheless, several behavioral factors beneficial to CRC prevention have been identified. Dietary consumption of calcium from dairy, fiber, fruits and vegetables may all help to decrease CRC risk (Song et al., 2015; Kushi et al., 2012). Additionally, physical activity and the intake of vitamin D or aspirin were shown to reduce CRC risk (Boyle et al., 2012; Rothwell et al., 2010; Song et al., 2015). Most importantly, endoscopic polypectomy during colonoscopy is believed to reduce CRC risk most efficiently (Elmunzer et al., 2015).

1.3 Precursor lesions of sporadic colorectal cancer

1.3.1 Adenoma-carcinoma sequence: a pathway to cancer

CRC develops slowly in a multi-step tumorigenesis of 10-20 years (Fearon and Vogelstein, 1990). First described by Morson et al. (1975), the carcinogenesis of the large bowel follows the gradual model of the adenoma-carcinoma sequence. This model has developed into a textbook paradigm according to which normal colonic epithelium is genetically transformed and progresses via premalignant lesions, *i.e.*, small and large adenomas, into a malignant adenocarcinoma (Figure 6). Fearon and Vogelstein (1990) have provided the molecular basis for the adenoma-carcinoma-sequence. They identified specific genetic alterations which not only initiated the carcinogenesis but also progressively accumulated during the transformation from adenoma into a solid carcinoma. The advanced model of the adenoma-carcinoma sequence has gained outstanding reward and is often referred to as "Vogelgram" (Haan et al., 2014; Krimpenfort et al., 2012). Despite this reputation, the "model has started to live a life of its own"

(Sillars-Hardebol et al., 2012). Thus, people stopped looking for alternatives. Intensified research on the colorectal carcinogenesis has accumulated striking evidence of multiple gene pathways frequently affected by alterations. Notwithstanding, the model reflects an oversimplification of a slightly more complex and diverse process. Jass (2007) describes the dogmatic linking of CRC to the mono-directional model as detrimental for the research in the field. Hence, Issa (2008) postulates the concept of adenoma-carcinoma sequence needs revision. This fact is underpinned by improved molecular analyzes of the colorectal carcinogenesis which unveiled several overlapping ways of CRC development (Snover, 2011) (see 1.4.1 Molecular pathogenesis).
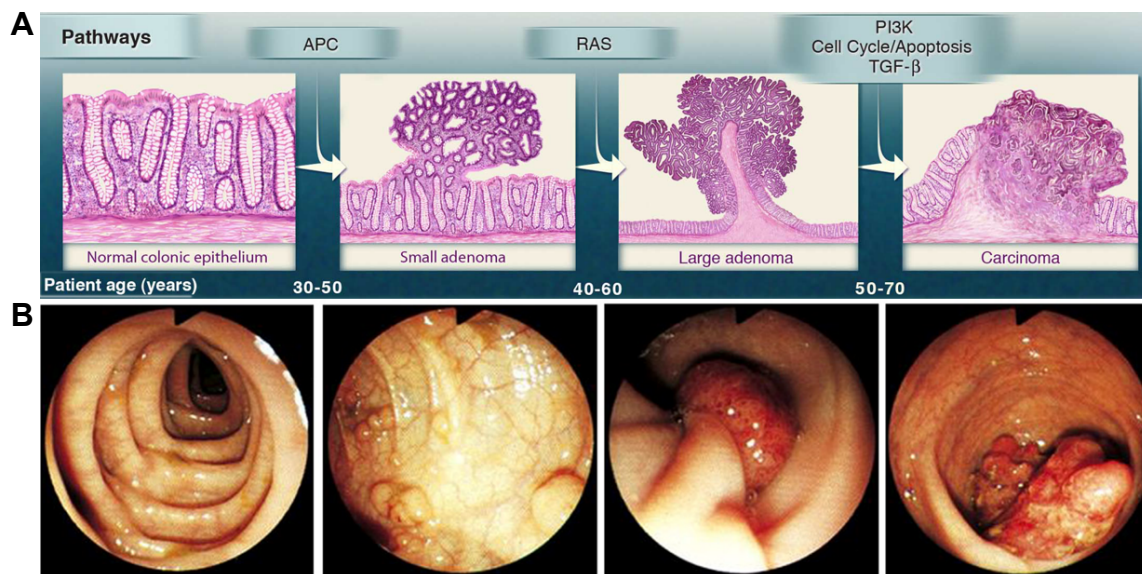
Figure 6. Adenoma-carcinoma sequence. (**A**) Histopathology of the evolution in colorectal cancer. From left to right, normal colorectal mucosa develops stepwise to adenoma and carcinoma by accumulating genomic events (top row). *Image extracted from Vogelstein et al. (2013).* (**B**) Endoscopic images taken at a colonoscopy. From left to right: normal epithelium, small adenoma, large adenoma and adenocarcinoma of the colon. *Image extracted from Cardoso et al. (2007).*

1.3.2 Aberrant crypt foci (ACF)

ACF have emerged as earliest morphological precursors to both adenoma and CRC (Gupta et al., 2009; Hamilton et al., 2000). ACF revealed a prevalence of 50% in patients of 40-49 years, sharply increasing with age and occurrence of adenomas or CRC (73% and 100%, respectively) (Takayama et al., 1998). Defined as microscopical preneoplastic clusters of aberrant crypts, ACF vary in size ranging from single altered glands to plaques greater than 30 crypts (Pretlow et al., 1991). Several genomic and epigenomic alterations, among which mutations of *KRAS* were most consistently described (13-82% of cases), were identified as shared by ACF, adenomas, and CRC (Gupta et al., 2007; Jen et al., 1994; Pretlow et al., 1993). Despite intensive research, the clinical relevance of ACF remains controversial regarding frequencies and impact.

1.3.3 Epithelial polyps

The largest group of precursor lesions of sporadic CRC is composed of epithelial polyps (Figure 7). Historically, these polyps were distinguished by two major types, *i.e.*, the hyperplastic polyp and the adenomatous polyp. Characterized by neoplastic cytological and histological atypia adenomas have been exclusively recognized as precursor lesions of carcinomas, while hyperplastic polyps seemed not capable to malignantly progress into CRC (Morson, 1962). Albeit, evidence has accumulated that the group of polyps with hyperplastic changes is, in fact, rather complicated and heterogeneous concerning both pathological and molecular features (East et al., 2008; Longacre and Fenoglio-Preiser, 1990; Snover et al., 2005; Torlakovic et al., 2003). In general, it was found that the group of hyperplastic polyps (in the sense of "serrated polyps") can be subdivided into four groups: (i) classical hyperplastic polyps (HP); (*ii*) sessile serrated adenomas (SSA); (*iii*) traditionally serrated adenomas (TSA) and (*iv*) mixed polyps (Aust and Baretton, 2010; Bettington et al., 2013; Rüschoff et al., 2007). All serrated polyps share the same pattern of characteristic saw-tooth, luminal serrations with branching and distortion in both crypts and epithelium (Strum, 2016). However, the molecular landscape of these lesions is more diverse with mutations predominantly affecting *KRAS* or *BRAF,* and methylation of CpG islands (Bettington et al., 2013; Snover, 2011). Although the classical HP is not associated with a risk for cancer progression, all other serrated polyps are perceived, meanwhile, as precursors of the serrated pathway carcinoma (Iion et al., 1999; Jass et al., 1999; Snover, 2011).
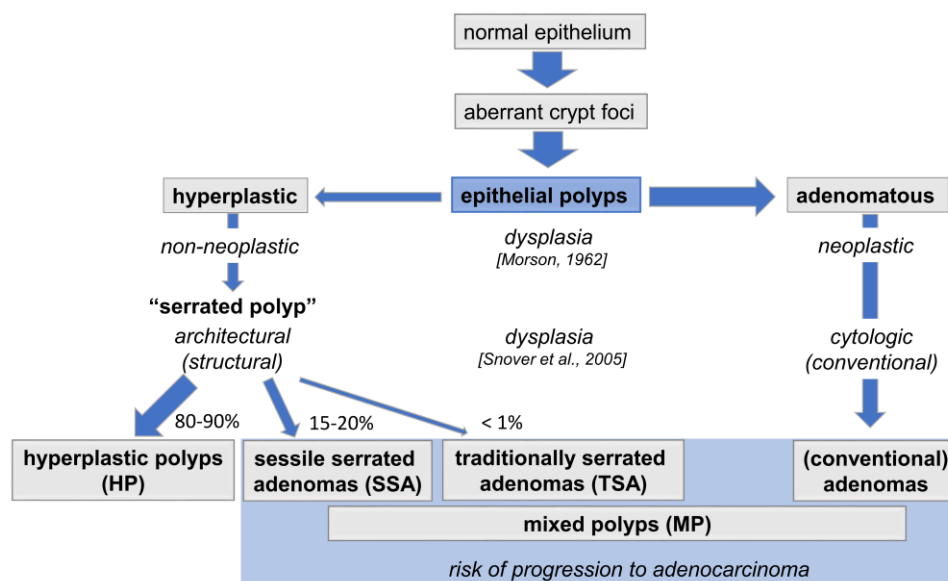
Figure 7. Terminology and systematic scheme of sporadic epithelial polyps of the colorectum. The Scheme displays the hyperplastic and adenomatous subgroups of polyps. Polyps displaying diagnostic criteria which match both adenomatous and hyperplastic cytological dysplasia are termed mixed polyps (MP). Although HPs are the most common hyperplatic polyps they seem not to harbor malignant potential for progression into CRC. *Image recreated according to Rüschoff et al. (2007).*

1.3.4 Adenomas

Adenomatous polyps are the most common premalignant neoplasia of the large intestine with an incidence of about 30-40% in individuals over 60 years of age (Eide, 1991; Levine and Ahnen, 2006; Paspatis et al., 2001; Vatn and Stalsberg, 1982). The incidence has been shown to increase dramatically with age emphasized by half the Western population who will have developed an adenoma by the age of 70 years (Fodde et al., 2001). Histologically, adenomas are defined by their grade of intraepithelial neoplasia (IEN) which is estimated by consideration of histological and cytological findings, *i.e.*, dysplasia of low-grade (LGD) or high-grade (HGD) character (Fodde et al., 2001; Hamilton et al., 2000). Hence, epithelial cells grow in multiple layers while the alignment to the basal membrane is lost (Fodde et al., 2001). The cells display characteristic hypercellularity with enlarged and hyperchromatic nuclei which may appear spindle-shaped (Hamilton et al., 2000). Histopathology of adenomas differentiates between tubular (less than 25% villous structures), tubulo-villous (25-75% villous structures) or villous (>75% villous structures) gland architecture, respectively (Hamilton et al., 2000; Strum, 2016). Adenomas may macroscopically emerge polypoid (pedunculated or sessile) or non-polypoid (flat, superficially elevated or depressed) (Hornick and Odze, 2015; Soetniko et al., 2008).

Adenomatous polyps are perceived as benign but potential precancerous lesions in the colorectal carcinogenesis representing an essential link from normal epithelium towards cancer. Molecularly, there is broad consensus that the first, *i.e.*, "gatekeeping," mutation often affects *APC* conferring a selective growth advantage to the cells to outgrow adjacent normal cells (Kinzler and Vogelstein, 1997; Vogelstein et al., 2013). Adenomas grow slowly, however, the clonal expansion may be promoted by another mutation, *e.g.*, *KRAS*, accelerating the exponential increase of mutated cells (Fearon and Vogelstein, 1990). Despite the paradigm of linear progression, it is estimated that only 3-5% of detected adenomas progress to malignancy which takes many years (Eide, 1991; Shinya and Wolff, 1979; Sillars-Hardebol et al., 2012). Several histological risk factors are believed to be associated with progression of adenomas: *e.g.*, size, villous features and HGD (DiSario et al., 1991; Hornick and Odze, 2015; Muto et al., 1975).

Polyps with malignancies (early adenocarcinomas with adenomas) may comprise two percent of all endoscopically resected polyps (Morson et al., 1984). In addition to adenomatous tissue, these lesions also expose carcinoma cells which invaded the submucosa throughout the *muscularis mucosa* (Cooper, 1983).

1.3.5 Resection of adenomas and surveillance intervals

CRC arises from adenomatous polyps according to the adenoma-carcinoma sequence. Detection and removal of these lesions are routinely performed by endoscopic colonoscopy. This technique provides the opportunity to most accurately detect cancer (at early and curable stages) and adenomas, whose removals reduce the risk of cancer by 76-90% (Winawer et al., 1993; Zauber et al., 2012). However, colonoscopy is also an invasive and cost-intensive procedure which is associated with morbidity and complications (Bond, 2000). As CRC risk increases in an age-dependent manner, colonoscopy is recommended for cancer prevention in patients aged ≥55 years (Stock et al., 2011). Colorectal adenomas are usually detected asymptomatically in colonoscopies and are subsequently endoscopically resected via electrocautery excision techniques (Bond, 2000). The endoscopist removes polyps preferentially by en-bloc excision (*in toto*) or, if not possible, by piecemeal snare excision. However, both techniques aim to retrieve tissue for pathological evaluation of malignancy, and if en-bloc resected to confirm complete resection (*in sano*) (Bond, 2000).

Patients who have undergone colonoscopy (with adenoma excision) are at an increased risk of two-to-four time of developing cancer compared to the average population (Levine and Ahnen, 2006). Accordingly, current guidelines recommend re-colonoscopy in defined time intervals based on clinical and pathological factors (Table 1). Firstly, post-polypectomy surveillance includes to determine the frequency of colonoscopies for re-examination and to clarify complete resection. Secondly, patients need to be defined who benefit from an intensified follow-up which does not invariably apply to all patients (Bond, 2000; Leitlinienprogramm Onkologie, 2014). Today, histology, size, and multiplicity of polyps are the significant factors to define screening intervals (Avidan et al., 2002; Neugut et al., 1995; Noshirwani et al., 2000; van Heijningen et al., 2013). However, none of these factors was prospectively evaluated, and retrospective studies often have methodical limitations. Noteworthy, the utilization of individual pathological characteristics as criteria for surveillance intervals and recurrence prediction will come under scrutiny (Avidan et al., 2002; Bond, 2000).

Since 20-40% of adenomas recur after endoscopic polypectomy (Avidan et al., 2002; Bonithon-Kopp et al., 2004; Neugut et al., 1995; Winawer et al., 1993), studies on adenoma recurrence may represent a unique way of examining the phenomenon of adenoma development. Given that the removal of polyps provides the opportunity for cancer prevention, a failure to detect and remove adenomas during colonoscopy

conversely increases the patient's risk for both adenoma recurrence and cancer (Corley et al., 2014). Adenoma recurrence seems likely due to lesions missed by colonoscopy, incomplete resection and, arguably, the phenomenon of field cancerization (De Maio et al., 2017; Levine and Ahnen, 2006). Hotta and colleagues (2008) also showed that local recurrence of adenomas occurred more frequently after piecemeal resection compared to en-bloc excision. Nevertheless, both procedures were named an adequate treatment achieving cure in most patients (Hotta et al., 2008; Walsh et al., 1992).

In summary, the lack of favorable pathological characteristics for the assessment of an individual risk of adenoma recurrence emphasizes the need for molecular markers which could predict recurrence more accurately in both sensitivity and specificity.

Table 1. Guidelines for post-polypectomy surveillance. Excerpt of guidelines for colonoscopic surveillance after polypectomy defined by the European Society of Gastrointestinal Endoscopy (ESGE) (Hassan et al., 2013). The recommendations apply for patients only who underwent high quality colonoscopy with complete removal (*in toto*) of all neoplastic lesions. Inadequately removed adenomas are not subject to the hereby suggested surveillance intervals.

	Colonoscopic findings	Recommended interval between colonoscopies
Low-risk group	• 1-2 tubular adenomas with low-grade dysplasia and size <10mm • Serrated polyps without dysplasia, size <10mm	10-year surveillance interval
High-risk group	• ≥3 adenomas • Adenomas with villous histology • Adenomas with high-grade dysplasia • Adenomas with size ≥10mm • Serrated polyps with dysplasia and/or size ≥10mm	3-year surveillance interval *(If no high-risk adenoma was detected at the first surveillance examination a 5-year surveillance interval is recommended.)*
	• >10 adenomas	Genetic counseling

1.4 Molecular pathogenesis in colorectal lesions

1.4.1 Molecular carcinogenesis pathways

The model of the adenoma-carcinoma sequence contributed fundamentally to the understanding of the colorectal carcinogenesis. To explain the clinical progression of colorectal tumors, understanding the molecular basis became an important objective in the last decades. Evidence accumulated that three major molecular pathways may drive colorectal carcinogenesis (Issa, 2008):

(*i*) The first pathway to be identified is the *chromosomal instability (CIN) pathway*, also referred to as the suppressor pathway. It accounts for approximately 60% of sporadic CRC including the conventional adenoma-carcinoma sequence (Fearon and Vogelstein, 1990; Jass, 2007). CIN denotes the characteristic aberrations of gains and losses of whole or substantial portions of chromosomes with high frequencies (Lengauer et al., 1998), and is accepted to increase as the tumor progresses (Fearon and Vogelstein, 1990). Consequently, CIN leads to an imbalance in chromosome number (termed aneuploidy) and structure, genetic mutations in oncogenes and tumor suppressor genes (*i.a.*, *APC*), sub-chromosomal genomic amplifications and high frequencies of loss of heterozygosity (LOH) (Pino and Chung, 2010).

(*ii*) The second pathway is termed *microsatellite instability (MSI) pathway*, also referred to as the mutator pathway. It affects up to 18% of sporadic CRC (Ionov et al., 1993; Popat et al., 2005), most of which were following the serrated-precursor pathway (Issa, 2008; Snover, 2011). The basis for this pathway is DNA tandem repeats (microsatellites) of one to six base pairs which are distributed throughout the genome. These repetitive sequences are prone to errors occurring during the process of DNA replication (Ionov et al., 1993). However, coding errors are normally repaired by mismatch repair (MMR) proteins (namely MLH1, MSH2, MSH6, and PMS2) which interact together to recognize and excise mismatches (Vilar and Gruber, 2010). A loss of function in one of the MMR system proteins induces the accumulation of frameshift mutations by inactivation of mismatch-repair (Buecher et al., 2013). This accumulation affects not only microsatellite sequences but also gene coding regions of key genes involved in several cellular processes (Duval and Hamelin, 2002). Consequential, MSI leads to genetic hypermutability. While sporadic CRC with MSI$^+$ is caused by somatic hypermethylation of the *MLH1* promoter, hereditary CRC with MSI$^+$ (LS) is induced by germline mutations of MMR protein genes (Buecher et al., 2013). MSI$^+$ cancers are subdivided into MSI-high or MSI-low, whereas non-affected tumors are termed microsatellite stable (MSS) (Buecher et al., 2013).

(*iii*) The third pathway towards CRC is the *CpG island methylator phenotype (CIMP)* (Toyota et al., 1999), also termed the epigenetic instability pathway. It was demonstrated that this pathway accounts for 20-30% of sporadic CRC (Jass, 2007; van Rijnsoever et al., 2002), which are predominately associated with serrated neoplasia precursor lesions (Snover, 2011). The name is derived from the characteristic epigenetic changes on the DNA, *i.a.*, methylation of CpG sites (Jones and Baylin, 2007). Based on the degree of hypermethylation, CIMP can be either of low-level (CIMP-L) or

high-level (CIMP-H) (Bettington et al., 2013). Tumors affected by CIMP⁺ exhibit hyper-methylation of CpG islands in gene promoter regions of tumor suppressor genes (TSGs) whose expression becomes silenced (Tycko, 2000). Accordingly, the progression of the tumor may be promoted accompanied by an increased tumoral growth.

The concept of linear pathways responsible for the progression towards CRC is yet an oversimplification (Issa, 2008). The multiplicity of pathways increases in complexity due to an overlap of the underlying features (Snover, 2011). That implies that the pathways are not mutually exclusive; however, their relevance is not yet fully defined and somewhat inconclusive (Pino and Chung, 2010). Hence, the inferred model of colorectal carcinogenesis pathways takes the shape of a complex disease reigned by multiple genotypes and phenotypes (Figure 8). The inconsistency of the current classifications emphasizes the need for a more stratified classification system. Recently, the CRC Subtyping Consortium assessed the distinct subtypes by sharing large-scale data and analytics among experts to proclaim new taxonomic patterns: the consensus molecular subtypes (CMS) of CRC (Guinney et al., 2015).
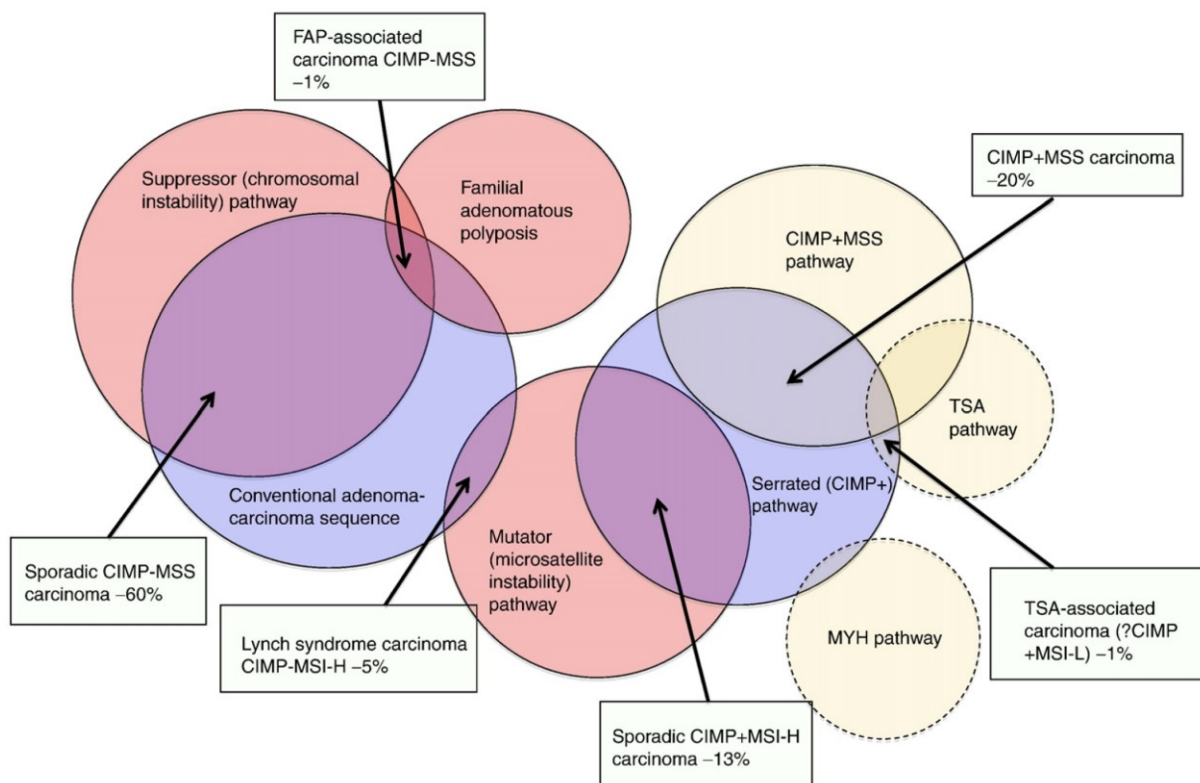
Figure 8. Venn-diagram depicting the overlap of molecular pathways in CRC development. Shown are pathways which are either based on putative initiation factors (*i.a.,* suppressor *versus* mutator pathway; *red circles*), based on the precursor lesion towards CRC (conventional adenoma-carcinoma sequence *versus* serrated pathway; *blue circles*) or which are hitherto poorly understood (*yellow circles*). Annotations in boxes reflect the intersections in the respective pathways. CIMP+/-, CIMP positive/negative; MSI-H/-L, degree of MSI high/low. *Image extracted from Snover (2011).*

1.4.2 Key genes in the colorectal carcinogenesis

It is reported that >70% of solid neoplasms show aneuploidy (Bakhoum and Compton, 2012). These genomic alterations affect specific sets of proto-oncogenes and tumor suppressor genes (Pino and Chung, 2010). Aberrations may lead to activation of oncogenic pathways or inactivation of tumor suppressive pathways. Aneuploidy-induced loss of heterozygosity (LOH) contributes to tumor progression (Weaver and Cleveland, 2006). Frequently targeted genes in the colorectal tumorigenesis are presented below.

Adenomatous polyposis coli (*APC*) and *β-catenin* (*CTNNB1*)

Mutations of *APC* on 5q22 are widely accepted as the earliest genetic event in the tumorigenesis of the colorectum, activating the WNT signaling pathway (Groden et al., 1991; Powell et al., 1992). While germline inactivating mutations in this gene show responsibility for hereditary CRC associated with FAP (Kinzler and Vogelstein, 1996), somatic mutations in *APC* are present in a small number of dysplastic aberrant crypt foci and strongly increase to 30-70% in sporadic adenomas and carcinomas, respectively (De Filippo et al., 2002; Miyaki et al., 1994; Otori et al., 1998). The disruption of *APC* is a "gatekeeping" (*i.e.*, tumor-initiating) event providing a selective growth advantage for the affected cells (Kinzler and Vogelstein, 1997). Mutations truncating the encoded protein within its N-terminal 1,600 amino acids are considered as driver gene mutations, whereas missense mutations in *APC* and protein-truncating mutations of the C-terminal 1,200 amino acids are likely passenger gene mutations (Vogelstein et al., 2013). Hypermethylation of the *APC* promoter was found in 18% of CRCs and adenomas suggesting an alternative process of *APC* inactivation (Esteller et al., 2000).

Expression of *APC* is involved in cellular differentiation, adhesion, polarity, migration, development, apoptosis and chromosomal segregation (Pino and Chung, 2010). Molecularly, intact APC interacts in the WNT signaling pathway in the absence of WNT by forming a "destruction complex" with β-catenin, glycogen synthase kinase-3β (GSK-3β) and casein kinase 1α/ε (CK1) on an axin-conductin scaffold. Accordingly, the phosphorylation of β-catenin by GSK-3β initiates the ubiquitin-driven proteasome-dependent degradation (Rubinfeld et al., 1996). Aberrant APC blocks the formation of the "destruction complex" leading to cytoplasmic accumulation of β-catenin. The protein is translocated into the nucleus to trigger the transcription of downstream genes which promote tumor growth and invasion (Mann et al., 1999). The APC pathway, hence, indirectly targets colorectal oncogenes, *i.a.*, *MYC* and *CCND1* (He et al., 1998). The downstream-target gene β-catenin (*CTNNB1*), *per se,* is renowned as commonly

affected by gain-of-function mutations in colonic tumors with MSI-H (Jass et al., 2003; Sparks et al., 1998). Nevertheless, mutations in *CTNNB1* are mainly limited to LS (Johnson et al., 2005; Miyaki et al., 1999). However, neither of these genes plays a role in clinical practice due to hitherto insufficient validation (Walther et al., 2009).

Proto-oncogene K-Ras (*KRAS*) and *proto-oncogene B-Raf* (*BRAF*)

The alteration of *KRAS* (located on 12q12.1) is an early event in the colorectal tumor-igenesis underlined by a prevalence of 30-50% in sporadic CRC (Andreyev et al., 1998; Santini et al., 2008). *KRAS* encodes the GTPase-enzyme RAS which is involved in multiple signal transduction pathways (Figure 9). As such, RAS is an effector of cellular growth, differentiation, apoptosis, survival, motility, transformation, vesicle trafficking and proliferation (Pino and Chung, 2010). RAS affects the major cellular proliferation MAPK/ERK pathway by triggering the cascade of MEK and ERK. ERK is translocated and phosphorylates cytosolic and nuclear substrates regulating the transcription of downstream targets, *e.g.*, CCND1, p21 and p27 (Pruitt and Der, 2001). Aberrant *KRAS* leads to constitutive activation of the MAPK/ERK cascade (Armaghany et al., 2012). Fearon and Vogelstein (1990) proclaim in the adenoma-carcinoma sequence that *KRAS* mutations occur after *APC* alterations, marking the tumor's transition from the breakthrough phase into the expansion phase (Vogelstein and Kinzler, 2015).

An important downstream target of RAS in the MAPK/ERK pathway is the enzyme RAF (encoded by *BRAF* located on 7q34) (Figure 9). Importantly, *BRAF* mutations are not only observed within serrated neoplasia and CIMP[+] CRC but also in MSI[+] CRC (Iacopetta et al., 2006; Torlakovic et al., 2003). Lesions presenting a mutation of *BRAF* show an inhibition of apoptosis in colonic epithelial cells (Snover, 2011).
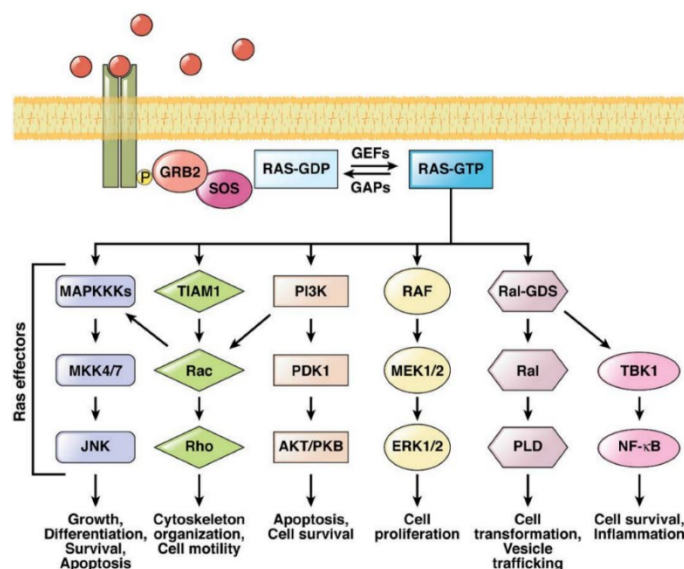
Figure 9. Schematic representation of the RAS signaling pathway. Growth factors (*red*), *e.g.*, EGF and TGF-α, bind to the respective growth factor receptor (*green*). This activates a cascade including the growth-factor-receptor bound protein 2 (GBR2) and son of sevenless (SOS) which stimulate the exchange of guanosine di-phosphate (GDP) for guanosine triphos-phate (GTP) in RAS GTPase by guanosine exchange factors (GEFs). Activated RAS-GTP activates several downstream target proteins (*e.g.*, PI3K and RAF) which are involved in the regulation of a multiplicity of cellular processes. *Image extracted from Pino and Chung (2010).*

Epidermal growth factor receptor (EGFR)

The proto-oncogene *EGFR* (located on 7p11.2) encodes a transmembrane receptor tyrosine kinase of the ERBB family of receptors which binds protein ligands of the EGF family. If a ligand binds to the extracellular domain, a dimerization of EGFR is induced which leads to autophosphorylation of the intracellular domain (Herbst, 2004). Subsequently, signal transduction pathways, *e.g.*, MAPK/ERK pathway and the PI3K/AKT pathway, are activated in a downstream cascade starting with RAS activation (Figure 9). EGFR is involved in cellular differentiation, proliferation, and survival (Herbst, 2004). Hence, EGFR activity was linked to the expansion of intestinal tumors (Roberts et al., 2002). Phenotypically in line with mutations of *KRAS* and *BRAF*, *EGFR* alterations lead to constitutive activation of the downstream signaling cascades empowering cells to growth aggressively (Ethier, 2002). Interestingly, chromosomal aberrations affecting chromosome 7 are believed to be an early and frequent event in the colorectal carcinogenesis (Meijer et al., 1998). Therefore, *EGFR* is a prominent target of copy number gain influencing tumor propagation by gene overexpression. The presence of *EGFR* copy number gains in 30-36% of colorectal adenomas emphasizes the pivotal role of *EGFR* in colorectal tumorigenesis (Habermann et al., 2011; Ried et al., 1996).

Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic SU alpha (PIK3CA)

PIK3CA (located on 3q26.32) encodes the catalytic subunit of class I phosphatidylinositol 3-kinase (PI3K), a downstream regulator of the EGFR pathway targeted by activated RAS, *i.e.*, RAS-GTP (Figure 9). PI3K is a lipid kinase regulating signaling cascades which influence cellular growth, proliferation, motility, and survival (Samuels and Velculescu, 2004). When activated by RAS, PI3K translocates to the plasma membrane and converts phosphatidylinositol-4,5-bisphosphate (PIP$_2$) to phosphatidylinositol-3,4,5-triphosphate (PIP$_3$) by phosphorylation. PIP$_3$ binds to either 3-phosphoinositide-dependent protein kinase-1 (PDPK1) or AKT serine/threonine kinase (protein kinase B, PKB) (Samuels and Velculescu, 2004). Mutations in hotspots of *PIK3CA* lead to constitutive activation of the PI3K-AKT pathway in >30% of CRCs (Samuels et al., 2004). DNA alterations of *PIK3CA* lead to overexpression and were revealed to be frequent in CRC patients with both amplifications (38%) and gains (19%), while 25% of colorectal adenomas exhibited amplification of *PIK3CA* (Jehan et al., 2009). Of note, *PTEN* (located on 10q23.1) is another prominent dysregulated target gene in the PI3K-AKT pathway exhibiting predictive value for metastatic CRC (Frattini et al., 2007; Jhawer et al., 2008).

Caudal type homeobox protein 2 (CDX2)

CDX2 (located on 13q12.2) encodes a transcription factor of the caudal-related home-obox family. Typically expressed in nuclei of intestinal epithelial cells (German et al., 1994), CDX2 is involved in gastrointestinal lineage development and differentiation which is emphasized by the utilization as a biomarker in diagnostic pathology (Liu et al., 2007). While multiple studies demonstrated tumor suppressive abilities for *CDX2* expression (Bai et al., 2003; Choi et al., 2006; Hinoi et al., 2001), oncogenic amplification of *CDX2* was, contrarily, shown to promote cellular proliferation and provide survival advantages for colorectal cancer cells compared to normal mucosa cells (Dang et al., 2006; Salari et al., 2012). Recently, a loss of CDX2 was named a negative biomarker in stage II/III cancer (Dalerba et al., 2016). Anyhow, recurrent copy number gain of chr13 in colorectal tumorigenesis were revealed in numerous cytogenetic studies (Bomme et al., 1994; Grade et al., 2006; Meijer et al., 1998; Ried et al., 1996).

Tumor protein P53 (TP53)

TP53 (located on 17p13.1) represents a frequently altered tumor suppressor gene which lost activity in a variety of cancers (Lane, 1992). The prevalence of mutated *TP53* in CRC is denoted with 35-55% (Markowitz and Bertagnolli, 2009). Despite this, *TP53* alterations are a late event in the adenoma-carcinoma sequence with rare occurrence in adenomatous lesions compared to carcinomas (Fearon and Vogelstein, 1990; Leslie et al., 2002). *TP53* encodes a transcription factor which coordinates the cellular responses to different types of cellular stress, *e.g.*, nutrient deprivation, hypoxia and ribosomal stress (Vogelstein et al., 2000). The protein triggers DNA repair as a response to DNA damage and arrests the cell cycle in G1 phase (Vousden and Lane, 2007). TP53 is capable of inducing senescence and apoptosis of the cell when DNA repair failed (Vousden and Lane, 2007). TP53 dysfunctions deregulate cellular proliferation and apoptotic machinery, resulting in tumor progression and genomic instability (Leslie et al., 2002). Thus, TP53 is recognized as "the guardian of the genome" (Lane, 1992), and frequently affected by allelic losses in aneuploid CRCs (Meling et al., 1993).

Mothers against decapentaplegic homolog-4 (SMAD4) and -7 (SMAD7)

Fearon and Vogelstein (1990) identified the allelic loss of chromosome-arm 18q as a widespread event affecting >70% of CRCs. While chr18 houses several TSGs the most important are *DCC*, *SMAD2*, and *SMAD4.* Last mentioned genes encode intracellular proteins which transduce the extracellular signal of ligands binding to the TGFβ-

receptor. SMAD2 belongs to the receptor-regulated class of R-SMADs. After activation by TGFβ-receptor, R-SMADs form a heteromeric complex with SMAD4 (class of co-SMAD). The complexes translocate into the nucleus to regulate gene expression (Roelen et al., 2003). Therefore, *SMAD4* regulates cellular growth, cellular differentiation and apoptotic processes (Heldin et al., 1997). Aneuploidy-associated heterozygotic loss of *SMAD4* leads to cellular deregulation and has been observed in 16% of CRCs (Takagi et al., 1996). In contrast, SMAD7 is a member of the class of inhibitory SMADs (i-SMAD), which are capable to antagonistically block proteins of other classes of SMADs (Itoh et al., 2001). Hence, SMAD7 regulates TGFβ signaling and provides negative feedback of the pathways of activin and BMP, respectively (Heldin et al., 1997; Ishisaki et al., 1999). Alterations of *SMAD7* have been observed in ~10% of both colorectal adenomas and carcinomas (Habermann et al., 2011).

1.5 Epigenetics

1.5.1 Epigenetic modifications

The definition of epigenetics comprises heritable and acquired abnormalities in gene expression which are not due to changes in the DNA sequence (Holliday, 1987). Although multiple epigenetic processes have been identified in transcription regulation, the covalent modifications of DNA, histone proteins or chromatin remodeling factors are hitherto most intensively studied (Esteller, 2011). Post-translational modifications on N-terminal amino acids ("tails") of histones form a "histone code" which involves different chemical groups (*i.a.,* methylation, acetylation, phosphorylation, ubiquitylation) (Esteller, 2008). The effects vary depending on the degree (*e.g.*, mono-, di-, and trimethylation), histone protein (*e.g.*, H3 or H4), histone residue (*e.g.*, lysine, arginine, serine) and their position (H3K4 and H3K9). However, the most-researched epigenetic modification occurs on the DNA. Methylation of DNA is exclusively observed on cytosines which are followed by guanines in 5'→3' direction, namely CpG dinucleotides. A methyl group is covalently added onto a cytosine which converts the pyrimidine derivative into 5-methylcytosine. Estimations state that about three out of four CpG sites are methylated in mammals (Jabbari and Bernardi, 2004). Moreover, CpG sites are not randomly distributed within the genome but cluster with high density in regions known as CpG islands (CGIs; length 0.5-2 kb). 70% of gene promoters located near the transcription start sites comprise CGIs (Saxonov et al., 2006). Analogous to CGIs, CpG sites in "shore" (up to 2 kb) and "shelf" regions (2-4 kb away from CGI) may also

be considered as regulatory regions for gene expression (Irizarry et al., 2009; Schneider et al., 2016). Methylation of CGIs in development or disease progression was shown to be associated with post-translational modifications of histones. This methylation leads to an inactivating condensation of chromatin structures and, consequently, to gene silencing (Jaenisch and Bird, 2003; Weber et al., 2007). Conversely, unmethylated CGIs and CpG methylation within the gene body is associated with gene transcription and silencing of transposable elements (*e.g.*, *LINE* and *SINE*) (Esteller, 2008; Jones, 2012; Yoder et al., 1997). The methylome, which comprises the sum of the methylated CpG sites, is maintained by DNA methyltransferase 1 (DNMT1). This enzyme adds methyl-groups to replicated DNA strands. However, DNMT3A and DNMT3B are capable of catalyzing the addition of methyl-group to CpG sites *de novo*, and are, thus, associated with aberrant CpG methylation in CRC (Ibrahim et al., 2011).

Most procedures which determine the methylation status of CpGs within the DNA utilize bisulfite conversion. In this procedure, sodium bisulfite converts unmethylated cytosine residues to uracil. However, 5-methylcytosines are protected from hydrolytic deamination by the catalyst. Although methylation analysis can be conducted by restriction enzyme sites, *e.g.*, combined bisulfite restriction analysis (Fraga and Esteller, 2002), most methods use PCR-amplification of converted DNA to characterize differentially methylated positions (DMPs): qualitatively, *e.g.*, by methylation-specific PCR (Herman et al., 1996), or quantitatively by methylation-sensitive qPCR MethyLight (Eads et al., 2000), DNA pyrosequencing (Tost et al., 2003) or DNA methylation microarray (Bibikova et al., 2011). The HumanMethylation450K BeadChip array permits the profiling of CGIs within a genome-wide scale with a single-nucleotide resolution. The array covers more than 485,000 CpG sites and, therefore, is a powerful tool as the analysis allows the detection of many thousands of DMPs (De Meyer et al., 2015).

1.5.2 DNA methylation in colorectal tumorigenesis

One of the first epigenetic alterations discovered was the low level of DNA methylation in tumors compared to corresponding normal tissues (Feinberg and Vogelstein, 1983). The state of hypomethylation contributes to CIN by promoting chromosomal recombination and rearrangement, reactivating transposable elements and causing the loss of imprinting (Bestor, 2005; Eden et al., 2003; Esteller, 2008). In the development of neoplasms DNA methylation of 5-methylcytosines globally decreases as the lesion progresses, *i.e.*, global hypomethylation (Fraga et al., 2004). Conversely, both the

frequency of hypermethylated CGIs and the imbalance of histone modifications increase as a tumor progresses (Esteller, 2008). Hypermethylation of CpG-dense promoters plays a central role in the initiation and the progression of neoplasms since it causes silencing of TSGs which are involved in cellular processes (Esteller, 2007; Weber et al., 2007). Interestingly, patterns of hypermethylated CGIs in promoters of TSGs appear to be specific to the cancer type (Costello et al., 2000; Esteller et al., 2001). The profile of frequently hypermethylated TSGs in colorectal tumorigenesis may comprise (Esteller, 2008): (*i*) transcription factors, *e.g.*, *RARβ2*, *GATA-4* and *GATA-5*; (*ii*) cell-cycle modulators, *e.g.*, *RASSF1A* and *p16^{INK4a}*; (*iii*) DNA repair proteins, *e.g.*, *MLH1* and *MGMT*; (*iv*) WNT-signaling modulators, *e.g.*, *APC*, *WIF1* and *SFRP1*.

Colorectal tumors, which are affected in no small degree by CGI hypermethylation, are classified as CIMP^{+} tumors (Tycko, 2000). The high frequency is evaluated by assessing the methylation status of marker loci, *e.g.*, *CDKN2A*, *MINT1*, *MINT2*, *MINT31*, and *MLH1* (Weisenberger et al., 2006). Nevertheless, CIMP^{+} tumors are predominantly associated with serrated neoplasia precursor lesions (Snover, 2011). This is underpinned by the presence of DMPs in serrated compared to conventional CRC (Conesa-Zamora et al., 2015) and by hyper-methylated loci in sessile polyps and serrated cancers (Andrew et al., 2017). Genome-wide studies on aberrant DNA methylation revealed the presence of four methylation clusters associated with CRC (Hinoue et al., 2012; The Cancer Genome Atlas Network, 2012). Interestingly, the methylation status of *CDKN2A, MLH1*, and *MGMT* in CRC was previously linked to cancer recurrence (Chang et al., 2016). Multiple authors also claim that aberrant DNA methylation appeared already in sporadic colorectal adenomas which are CIMP$^{-/low}$ (Galamb et al., 2016; Patai et al., 2015; Silva et al., 2014; Voorham et al., 2013). Most of these studies focused solely on promoter methylation whereas the methylation of gene bodies is also believed to be crucial for the regulation of gene expression (Jones, 2012; Yang et al., 2014). Galamb et al. (2016) showed that DMPs in colorectal tumors compared to normal tissue were most commonly found in gene body regions of WNT pathway genes, *i.a.*, *APC*, *AXIN2*, *CTNNB1* or *SFRP1.* DNA methylome analysis mapped aberrant CGI methylation to genes involved in the TGF-β pathway (including *BMP3*) (Ashktorab et al., 2014). Luo et al. (2014) found evidence for differences in methylation patterns across the progression from normal mucosa to adenoma and CRC.

Despite these advances in profiling the methylome of colorectal adenomas and cancers, the putative role of CpG methylation in sporadic colorectal adenoma recurrence and tumor evolution is yet elusive and requires intense research.

1.6 Aneuploidy and chromosomal instability

1.6.1 Abnormal karyotypes

Tumor formation was firstly associated with the alteration of chromosomal material almost 130 years ago (von Hansemann, 1890). Boveri (1914) suggested that cancer originates from "imbalances in the chromosomes," and hence, should be recognized as a disease of the chromosomes (Ried, 2009). Nowadays, the presence of an abnormal number of chromosomes in a cell is termed "aneuploidy." The gain and loss of chromosomes due to missegregation at high rates are defined as chromosomal instability (CIN) (Lengauer et al., 1997; Lengauer et al., 1998). Although aneuploidy is a common feature in tumor cells, aneuploidy and CIN are not synonymous (McGranahan et al., 2012). An aneuploid cell may originate from a single event of chromosome missegregation accompanied by clonal expansion. Ultimately, aneuploid cells irrespective of CIN may present rather "homogeneous" clonal patterns of chromosomal aberrations, while CIN causes "heterogeneous" patterns of chromosomal aberrations (Bakhoum and Compton, 2012). However, CIN contributes as a form of genomic instability to the widely accepted "hallmarks of cancer" (Hanahan and Weinberg, 2011). CIN is classified by the gain or loss of whole chromosomes (*i.e.*, numerical CIN) or chromosome fragments and rearrangements (*i.e.*, structural CIN) (Lengauer et al., 1998; McGranahan et al., 2012). Structurally abnormal chromosomes and rearrangements are the results of chromosomal deletions, duplications, translocations, inversions, insertions, rings and isochromosome formation (Huret et al., 2013). CIN cancer genomes often exhibit alterations of the whole- and the sub-chromosomal level (McGranahan et al., 2012).

There is a controversy whether aneuploidy is the origin, or *vice versa*, principle consequence of CIN. While many authors name aneuploidy as the direct consequence of CIN (Duesberg et al., 1998; Reish et al., 2006; Yuen, 2010), more recent studies claim that, in fact, aneuploidy leads to CIN (Lengauer et al., 1997; Nicholson et al., 2015; Passerini et al., 2016; Thompson and Compton, 2008; Sheltzer et al., 2011).

1.6.2 Mechanisms leading to chromosomal instability

CIN is a direct result of defects in the dynamic signaling cascade which conducts the mechanism of segregation of the chromosomes during mitosis. Briefly, the process of chromosome separation begins in prometaphase of mitosis. The nuclear envelope is broken down, and chromosomes attach to the spindle by binding of microtubules at

kinetochores (Compton, 2000), *i.e.*, protein structures located at the chromosomal centromeres (Maiato et al., 2004). Kinetochore microtubules emerge from the spindle poles and reach for the kinetochores. During metaphase, chromosomes are aligned at the metaphase plate in the center of the cell. Nevertheless, a single chromosome unattached or falsely oriented to the spindle is capable of delaying the onset to anaphase by emitting a "wait anaphase" signal (Rieder et al., 1994; Rieder et al., 1995; Thompson et al., 2010). This "signal" is transduced by spindle-checkpoint proteins inducing the inhibition of the anaphase-promoting complex/cyclosome (APC/C; ubiquitin-protein ligase) via formation of the spindle assembly checkpoint (SAC) (Herzog et al., 2009). Both SAC and cyclin-dependent kinases preserve the cohesion of sister chromatids (Thompson et al., 2010). When no errors occur in the transition towards anaphase, APC/C will not be inhibited and ubiquitinylates securin. The protease separase is activated and cleaves the multiprotein complexes of cohesins (Herzog et al., 2009).

Due to the multiplicity of genes involved in the process of accurate chromosome segregation, there are multiple error-prone targets whose alteration eventually leads to cancer formation and CIN (Thompson et al., 2010). Yuen (2010) summarized the mitotic errors possibly affecting chromosome segregation: (*i*) sister chromatid cohesion defects (Zhang et al., 2008), (*ii*) defective centrosome duplications (*i.e.*, supernumerary centrosomes) (Ganem et al., 2009), (*iii*) weakened (*i.e.,* nondisjunction errors) or hyperactivated spindle assembly checkpoint (Cahill et al., 1998; Weaver and Cleveland, 2006), and (*iv*) extremely stable kinetochore-microtubule attachments to chromosomes (Bakhoum et al., 2009). Structural CIN is also linked to being driven by telomere dysfunction caused by chromosomal ends which enter breakage-fusion-bridge-cycles (O'Hagan et al., 2002; Pino and Chung, 2010). Alterations in genes encoding cell cycle proteins are perceived to allow cells with DNA double-strand breaks to divide and accumulate structural CIN (Lengauer et al., 1998). Thus, alterations of DNA repair genes of the DNA damage response machinery (*i.a.*, fixing DNA double-strand breaks) are involved in carcinogenesis (Gorgoulis et al., 2005; Khanna and Jackson, 2001).

The tumorigenesis of epithelial cancers is commonly (incidences exceeding 50% [Carter et al., 2012]) affected by genome duplication, *i.e.*, polyploidization. This catalyzes CIN (Holland and Cleveland, 2009). Polyploidy depicts the condition of a cell which possesses more than two haploid chromosome sets (*i.e.*, diploid; 2N), with near-triploid (3N) and near-tetraploid (4N) being the most common forms observed (Weaver and Cleveland, 2006). Tetraploidy can be caused by cell-cell fusion, cytokinesis failure, mitotic slippage or endoreduplication (Holland and Cleveland, 2009; Quintanilla, 2017).

1.6.3 Consequences of aneuploidy and chromosomal instability

Aneuploidy has been recognized for decades as severely detrimental for the patients. Thus, chromosome abnormalities induce miscarriages of unborn, lead to congenital mental retardation and are associated with cancer (Nagaoka et al., 2012). However, cellular effects caused by aneuploidy remained elusive in the past. Recently, a direct correlation between mRNA expression and chromosome copy number was demonstrated: mRNAs of corresponding genes, located on aneuploid chromosomes, were present in more considerable extent compared to euploid controls (Ried et al., 2012; Rutledge and Cimini, 2016; Stingele et al., 2012). Interestingly, observations in yeast and human cells suggest that aneuploidy not only affects the corresponding genes but also has genome-wide effects on gene expression (Dephoure et al., 2014; Sheltzer et al., 2017). Speculatively, genes located on aneuploid chromosomes may encode transcriptional regulators of genes on other chromosomes (Rutledge and Cimini, 2016). CNAs were also associated with changes in the abundance of proteins, *e.g.*, protein complexes and kinases (Stingele et al., 2012; Torres et al., 2007). Imbalances in protein levels interfere with the stoichiometry of specific protein complexes and lead to an excess of uncomplexed proteins generating proteotoxic stress (Gordon et al., 2012).

Consequently, aneuploid cells may exhibit an overwhelmed protein degradation pathway resulting in protein misfolding and aggregation (Gordon et al., 2012). That in turn negatively affects the cellular fitness and proliferation, although aneuploidy appears to be beneficial for cancer cells under certain circumstances (Rutledge and Cimini, 2016). Storchová (2012) conjectured that aneuploidy protects the cells from other stress factors and that increased CIN accelerates the evolution of clones with advantageous karyotypes. Hence, aneuploidy may be an effective mechanism to confer a selective advantage to the cells by providing tolerance to stress (Gordon et al., 2012).

1.6.4 Aneuploidy in colorectal tumorigenesis

Beroukhim and colleagues (2010) revealed that 25% of a typical cancer sample's genome is affected by arm-level somatic CNAs. This fact corroborates the relevance of aneuploidy for tumor initiation and progression. Hence, chromosomal imbalances were also investigated among colorectal tumors. Proposed by the adenoma-carcinoma sequence, chromosomal abnormalities occur early and increase in the evolution from normal mucosa to adenoma and adenocarcinoma (Fearon and Vogelstein, 1990). Cytogenetic analysis of colorectal adenomas displayed chromosomal gains of 3, 7, 13

and 20 and the loss of 18 (Bomme et al., 1994). Other researchers identified a similar pattern with chromosomal gains of 9, 12, 13 and 20, and losses of 18q (Muleris et al., 1994). The first investigation by comparative genomic hybridization (CGH) showed no chromosomal alterations in MSI+ CRC, whereas sporadically occurring CIN+ CRC revealed characteristic CNAs, namely: gains of the chromosome (arms) 7, 13 and 20q and losses of 9p, 17 and 18 (Schlegel et al., 1995). However, Ried and collaborators (1996) were the first to analyze the progression from adenoma to carcinoma by CGH which unveiled a stepwise accumulation of chromosomal aberrations including gains of 1, 7, 8q, 13 and 20 and losses of 4, 8p and 18q. A rash of similar studies utilized CGH to confirm these earlier results in colorectal adenomas and CRC (Camps et al., 2006; De Angelis et al., 1999; Douglas et al., 2004; Hermsen et al., 2002; Ried et al., 1999). Noteworthy, these characteristic CNAs were not only presented in Western patients but also consistently found in Asians (He et al., 2003; Poeaim et al., 2005).

Accordingly, this led to the updated view (Figure 10) that the accumulation of characteristic chromosomal aberrations drives the colorectal tumorigenesis of sporadically occurring lesions. These include gains of 7 and 20q in low-grade adenomas and is accompanied by gains of 8q and 13 and losses of 4q, 8p, 17p and 18 in the progression from high-grade adenomas towards CRC (Grade et al., 2006; Ried et al., 2012). A few aneuploid adenomas were identified to be affected by CIN (Jones et al., 2007), while studies on CNA profiles of polypoid and nonpolypoid adenomas revealed controversial results (Postma et al., 2005; Richter et al., 2003). Many studies focused on colorectal tumor development, patient survival and metastasis (Grade et al., 2006), hitherto none investigated the relevance of CNAs on local adenoma recurrence.
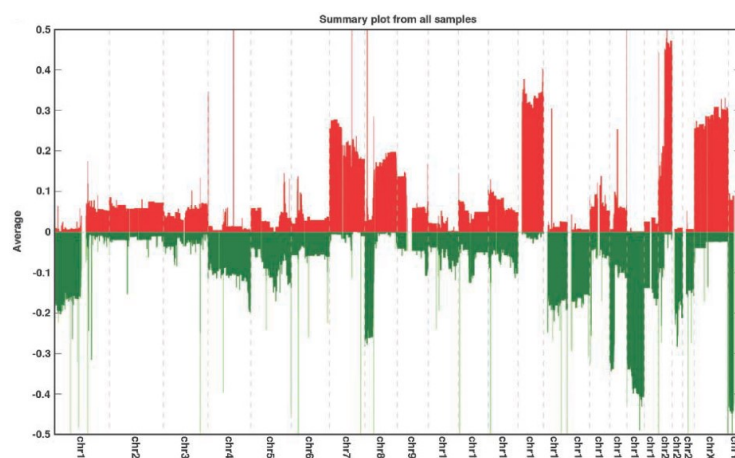
Figure 10. Summary plot of chromosomal aberrations in CRC detected by aCGH. Chromosomal gains (*red*) are predominantly seen on chromosomes 7, 8q, 13 and 20q, while chromosomal losses (*green*) show the highest frequencies on 8p and 18. Y-axis depicts the average values expressed as log$_2$ ratio. *Image extracted from Camps et al. (2008).*

1.7 Tumor heterogeneity

1.7.1 The concept of tumor heterogeneity

The theory of the evolutionary development of cancer subclones dates back four decades ago. Peter Nowell (1976) proposed in his groundbreaking article that a progenitor cell may give rise to mutant clonal subpopulations within cancer and that the progression of neoplasms is based on the acquisition of genetic alterations which lead to the Darwinian development of clones. Since then, intensified research on cancer development utilized a broad spectrum of genomic approaches which substantiated the widely accepted view of tumors bearing inter- and intratumor heterogeneity (Figure 11). Intertumor heterogeneity describes the genetic and phenotypical divergence among tumors from individual patients. That is further substantiated by variations between tumors of different tissues and cell types, which vary in genomic landscapes and, consequently, in prognosis and treatment response (Burrell et al., 2013).

However, intratumor heterogeneity (ITH) denominates the observation that even an individual tumor may be reigned by multiple subclonal populations which acquired different genetic alterations (Burrell et al., 2013). An ongoing generation of (genetically anomalous) subclones leads to an increase of the intratumor heterogeneity and influences the degree of genomic instability (*i.e.*, CIN) which confers the basis for multiple features defining the tumor (Camps et al., 2013).
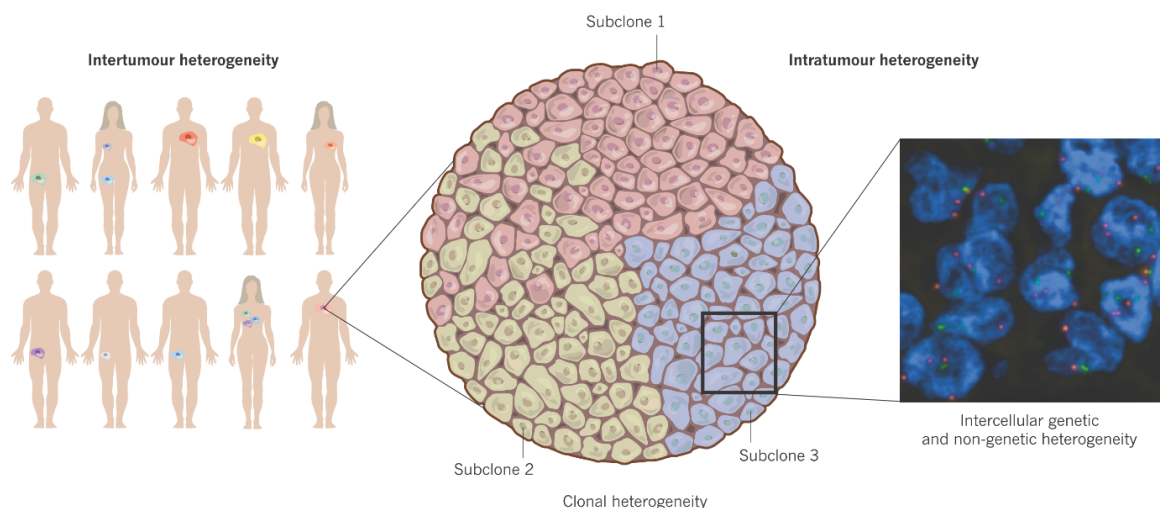
Figure 11. Inter- and intratumor heterogeneity. Among individual cancer patients there is a genetic and phenotypical variation between cancers of different tissue and cell type. Additionally, variation among patients with the same cancer type is frequently observed. A phenomenon called intertumor heterogeneity (*left image*). Moreover, the fraction of cancer cells might display a set of genetically divergent subclones which contribute to intratumor heterogeneity (*middle image*). Subclonal populations can be revealed by fluorescence *in situ* hybridization (FISH; *square*). *Image extracted from Burrell et al. (2013).*

The underlying causations for inter- and intratumor heterogeneity are a constant matter of debate. Visvader (2011) proposed two models causing intertumoral heterogeneity: (*i*) the "(epi-)genetic mutation model" and (*ii*) the "cell-of-origin model." Contrary, ITH is generated in the model of clonal evolution by tumor cells emerging from a single or multiple subclones (Nowell, 1976; Russnes et al., 2011). However, according to the model of cancer stem cells, a single progenitor stem cell expands and generates a monogenomic clone population while multiple rare stem cell progenitors give rise to polygenomic populations. Nevertheless, evidence also suggests a mutator phenotype model gradually accumulating genetic alterations which result in a highly genetically diverse subclonal tumor (Russnes et al., 2011). Burrell and Swanton (2014) proposed that ITH is generated by the clonal evolution of the cancer genome which occurs in a linear or branched fashion (Figure 12A). Linear evolution of the tumor is characterized by the acquisition of beneficial genetic alterations which allow both clonal expansion and clonal succession in competing with other clones. The elimination of competing subclones via outgrowing by a novel dominant clone is described as clonal sweep (Burrell et al., 2013). Branched evolution, in contrast, could be the result of a failed clonal sweep which is accompanied by the appearance of sister clones which grow in parallel and compete in Darwinian evolution (Burrell and Swanton, 2014). Given that tumors are subject to constant evolutionary variation, the composition of the subclonal pattern within a particular lesion is rather temporal than static which is expressed in temporal heterogeneity (Burrell and Swanton, 2016). Cross et al. (2016) suggested that the model of gradualism in the clonal evolution of cancers demands adjustment and is better described by a punctuated equilibrium of variable evolutionary tempos. The clonal composition of a tumor may vary between surgical excision and clinical recurrence as minor and major clones take part in clonal selection (Figure 12B). Understanding the relevance of clones with different frequencies (*i.e.*, major and minor clones) "over space and time" attracts notice (Burrell and Swanton, 2016).

Utilization of single-cell next-generation sequencing (NGS) approaches and the single-cell variant of multiplex-interphase fluorescence *in situ* hybridization (miFISH) (see 1.7.2 FISH) confirmed the presence of branched intratumor heterogeneity in both hematological and solid tumors, *e.g.*, the pancreas and the breast (Anderson et al., 2011; Campbell et al., 2010; Heselmeyer-Haddad et al., 2012). As early as in the 1990s, colonic adenomas were shown to be of polyclonal origin contrary to the scientific perspective at that time which favored monoclonal development (Novelli et al., 1996). Thirlwell and coworkers (2010) accumulated evidence for polyclonality present in

colorectal neoplasms by analyzing the loss of heterozygosity in *APC*, *KRAS*, and *TP53* via Sanger sequencing of multiple crypts within a single tumor. Although the role of ITH was elucidated within colorectal lesions, the putative variation of subclonal patterns between primary and recurrent colorectal adenomas is yet elusive.
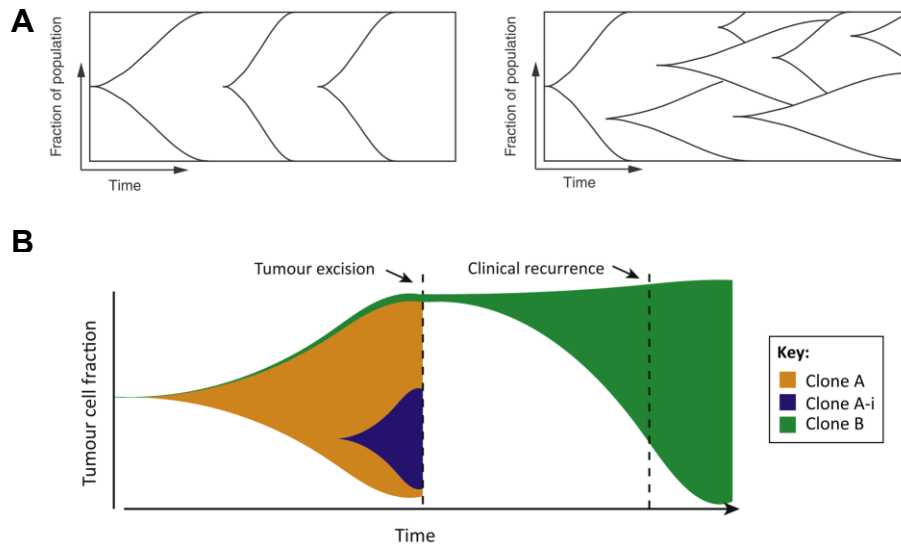
Figure 12. Trajectories reflect the evolution of cancers. (**A**) The evolution of cancer follows either the linear (*left box*) or the branched (*right box*) trajectory. Linear evolution is characterized by clonal sweeps which eliminate antecedent clones by outgrowing the population. In contrast, branched evolution is distinguished by genetically different clones arising independently and expanding in parallel. *Image extracted from Burrell and Swanton (2014).* (**B**) While a tumor may be dominated by several clones at the time of (surgical) excision, early divergence of clones might, in fact, account for clinical recurrence. *Image extracted from Burrell and Swanton (2016).*

1.7.2 Fluorescence *in situ* hybridization (FISH)

The approach of molecular cytogenetics offers the opportunity to utilize chromosomes and their respective architecture as prognostic, predictive and diagnostic biomarkers (Wippold and Perry, 2007). Additionally, the cytogenetic detection of chromosomal structures may support the identification of subclonal populations which contribute to tumor heterogeneity. Therefore, molecular cytogenetics provides several powerful techniques among which FISH is of outstanding importance in research and routine diagnostics. While the first *in situ* hybridization for chromosome analysis was performed by radioactively labeled DNA probes (Pardue and Gall, 1969), utilization of fluorescent dye-labeled probes demonstrated superiority since their discovery (Rudkin and Stollar, 1977). Although nuclei in metaphase preparations were exclusively accessible by FISH back in the days, improvements of the technique then provided hybridization procedures of nuclei in interphase which paved the way for analyses of both fresh and archived clinical samples (Cremer et al., 1986). Moreover, advances in the technique contributed to the widespread and everyday use. FISH is based on the

complementary binding of labeled DNA probe strand fragments onto a target DNA locus (Figure 13A). The probe has to be of a certain length to precisely bind the target DNA and to emit a sufficient signal intensity allowing the detection by fluorescence DNA microscopy (Wippold and Perry, 2007). Combining multiple FISH probes with multi-colored fluorochromes makes the simultaneous hybridization of interphase nuclei assessable (Figure 13B). Nuclei are specifically counter-stained with DAPI. In this study, the whole technique is termed miFISH in the abbreviated notation. Location, presence, and size of the probe signals are of particular relevance for the evaluation of the probe and the targeted DNA. FISH probes can be targeted against (*i*) specific genes and small chromosomal regions, (*ii*) repetitive sequences within the centromeric region of chromosomes, (*iii*) repetitive sequences at the telomeric regions of chromosomes and (*iv*) sequences painting whole-chromosomes for karyotyping (Kontogeorgos, 2004). Specific gene probes were designed using bacterial artificial chromosomes (BAC) which cover the entire human genome. Hence, gene-specific and centromere probes, respectively, proved useful to study chromosomal alterations within tumors. Consequently, numerous studies demonstrated that FISH and single-cell miFISH are powerful cytogenetic techniques to investigate tumorigenesis, disease progression and tumor heterogeneity within a variety of cancers (Habermann et al., 2013; Heselmeyer-Haddad et al., 2012; Heselmeyer-Haddad et al., 2014; Oltmann et al., 2018; Wangsa et al., 2009; Wangsa et al., 2016).
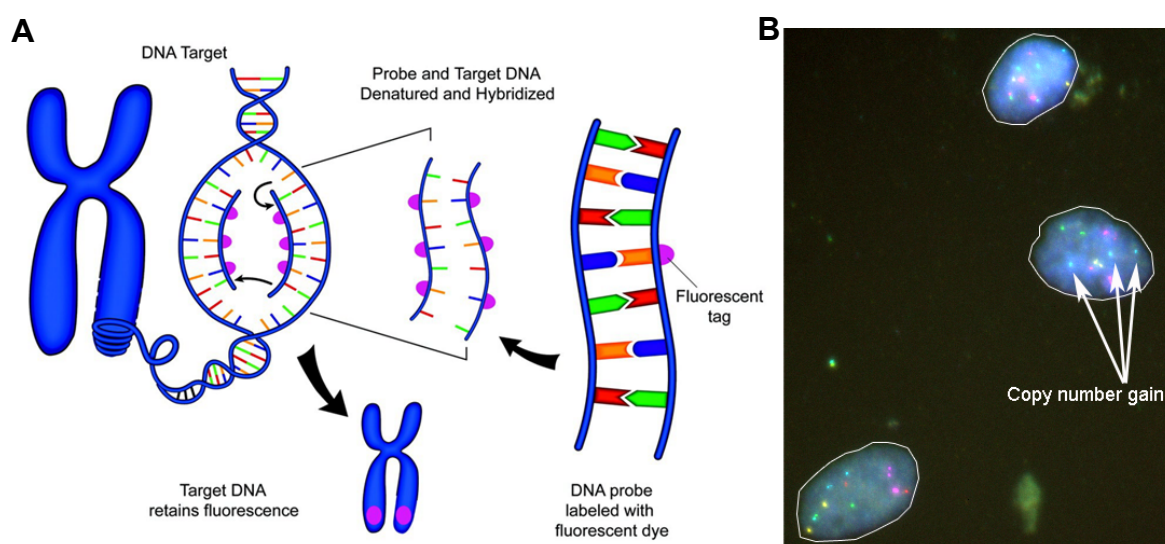
Figure 13. Schematic representation of FISH and miFISH on adenoma nuclei. (**A**) Schematic representation of fluorescence *in situ* hybridization. The fluorescent dye-labeled DNA probe is denatured together with the target DNA. DNAs hybridize with each other according to the complementary bases. The fluorescence tag is detected with a fluorescence microscope allowing to assess the copy number status of genes. *Image extracted from Wippold and Perry (2007)*. (**B**) Representative merged fluorescence image of single-cell miFISH performed on adenoma nuclei fixed on cytospin-slides. Nuclei were counter-stained with DAPI and digitally marked by a white frame. FISH revealed neutral copy numbers (*i.e.*, two signal counts) for all probes hybridized except for the *blue* probe (three signals, *i.e.*, copy number gain).

1.8 Objectives

Knowledge of the epigenetic and genetic signatures underlying local recurrences of colorectal adenomas are hitherto limited, which confers a health burden for patients. Overall, this study aimed to improve the understanding of the development of recurrence of sporadic colorectal adenomas. The thesis does not only provide an integrative analysis of the epigenetic and genetic landscapes of primary colorectal adenomas without and with recurrence but also an examination of primary adenomas with the corresponding recurrent adenomas, so-called matched pairs.

The first objective was to determine whether epigenetic and genetic markers could augment the prediction of adenoma recurrence based on the signature of the primary tumor. With our epigenetic approach, we sought to highlight signatures of putative differentially methylated CpG positions (DMPs). DMPs were evaluated to uncover specific associations with emphasis on the genomic distribution and pathway enrichment. Our genetic approaches assessed copy number alterations (CNA) in the tumor bulk and at the level of single cells. CNAs were utilized to quantify clonal diversity within the samples and to identify predictive biomarkers for local recurrence of adenomas. This should elucidate whether adenoma recurrence occurs as a consequence of an acquisition of distinct genetic alterations.

Second, this study was focused on the clonal evolution from the primary adenoma to the corresponding recurrent adenoma. Particularly, patterns of genomic imbalances within the adenoma samples could aid to clarify whether genomic instability and intratumor heterogeneity change throughout the process of adenoma recurrence. Quantitative diversity measures were applied to assess inter- and intratumor heterogeneity across the adenomas. Moreover, CNAs of single cells were used to infer patterns of clonal evolution in adenoma recurrence.

2 MATERIALS AND METHODS

2.1 Materials

Table 2. List of kits.

Kit	Supplier
DNA Clean & Concentrator™-25 (# D4033)	Zymo Research, Irvine, CA, USA
EnVision Detection System, Peroxidase/DAB, Rabbit/Mouse, Dako (# K5007)	Agilent, Santa Clara, CA, USA
EpiTect® Plus DNA Bisulfite Kit (# 59124)	Qiagen, Hilden, Germany
Genomic ULS Labeling Kit	Agilent, Santa Clara, CA, USA
Genomic ULS Purification Module, KREApure	Agilent, Santa Clara, CA, USA
Gentra Puregene Tissue Kit (# 158667)	Qiagen, Hilden, Germany
Infinium FFPE QC Kit (#WG-321-1001)	Illumina, San Diego, CA, USA
Infinium HD FFPE DNA Restore Kit (#WG-321-1002)	Illumina, San Diego, CA, USA
Infinium HumanMethylation 450K BeadChip Array	Illumina, San Diego, CA, USA
Oligo aCGH/ChiP-on-chip Hybridization Kit	Agilent, Santa Clara, CA, USA
Oligo aCGH/ChiP-on-chip Wash Buffer Kit	Agilent, Santa Clara, CA, USA
Plasmid Maxi Kit 25 (# 12163)	Qiagen, Hilden, Germany
PyroGold SQA Reagent Kit	Qiagen, Hilden, Germany
SurePrint G3 Human CGH Microarray Kit 8x60K	Agilent, Santa Clara, CA, USA

Table 3. List of consumables.

Consumable	Supplier
Abgene™ 96-well 0.8ml, Deep-well plate MSA5	Thermo Fisher Sc., Waltham, MA, USA
Aluminum foil, Reynolds Wrap®	Reynolds Cons. Pro., Lake Forest, IL, USA
Corning™ 30 ml Glass Tube	Corning, Corning, NY, USA
Corning™ Mini Bioreactor 50 ml	Corning, Corning, NY, USA
Coverslips, 22x22 mm², Glass	VWR International, West Chester, PA, USA
Cytology Funnel, Fisherbrand™, Single (# 10-335)	Thermo Fisher Sc., Waltham, MA, USA
Fixogum Rubber Cement	Marabu, Tamm, Germany
Lens cleaning tissue, Whatman®	GE Healthcare, Chicago, IL, USA
Microscope Adhesion Slides, SuperFrost™ Plus	Langenbrinck, Emmendingen, Germany
Microscope Slides, Duran® frosted	DWK Life Sciences, Wertheim, Germany
Magnetic stir bar, Spinbar®	VWR International, West Chester, PA, USA
Nalgene™ Oak Ridge Glass Centrifuge Tube 30 ml	Thermo Fisher Sc., Waltham, MA, USA
Nalgene™ Oak Ridge PPCO Centrifuge Tube 50 ml	Thermo Fisher Sc., Waltham, MA, USA
PCR tube, 0.5 ml, multiply® strip	Sarstedt, Nümbrecht, Germany
Parafilm® M	Bemis Company, Neenah, WI, USA
Pasteur pipettes, Brand®, plastic	Sigma-Aldrich, St.Louis, MO, USA
Pasteur pipettes, glass	Carl Roth, Karlsruhe, Germany
Pipette Tips, epT.I.P.S.® (red, grey, yellow, blue)	Eppendorf, Hamburg, Germany

Consumable	Supplier
Polypropylene Centrifuge Tube 15 ml	Corning, Corning, NY, USA
Precision wipes, KIMTECH	Kimberly-Clark Corp., Neenah, WI, USA
Q-PCR plate MicroAMP Fast 96-well	Applied Biosystems, Foster City, USA
Reaction tube, 1.5 ml	Eppendorf, Hamburg, Germany
Reaction tube, 2.0 ml	Eppendorf, Hamburg, Germany
Serological pipettes (2 ml, 5 ml, 10 ml, 25 ml)	Sarstedt, Nümbrecht, Germany
Surgical Scalpel FEATHER®	Sigma-Aldrich, St. Louis, MO, USA
Syringe and Hypodermic Needle "Sterican"	B. Braun AG, Melsungen, Germany

Table 4. List of reagents and chemicals.

Reagent	Supplier
Acetone	Carl Roth, Karlsruhe, Germany
Agarose, UltraPure™, 100 g (# 16500-100)	Invitrogen, Carlsbad, CA, USA
Antibody, CDX2, clone DAK-CDX2, Dako (# M3636)	Agilent, Santa Clara, CA, USA
Antibody, MLH1, clone ES05, Dako (# M3640)	Agilent, Santa Clara, CA, USA
Antibody, MSH2, clone FE11, Dako (# IR085)	Agilent, Santa Clara, CA, USA
Antibody, MSH6, clone EP49, Dako (# IR086)	Agilent, Santa Clara, CA, USA
Antibody, PMS2, clone EP51, Dako (# M3647)	Agilent, Santa Clara, CA, USA
Antibody diluent (# ZUC025)	Zytomed Systems, Berlin, Germany
Aqua Ad Injectabilia (ddH$_2$O) 10 ml (# 3113087)	B. Braun AG, Melsungen, Germany
β-Mercaptoethanol (2-Mercaptoethanol) (# M6250)	Sigma-Aldrich, St. Louis, MO, USA
BAC clones, stab agar cultures	BACPAC Resources Center, CA, USA
Bovine Serum Albumin (BSA), lyophilized (# A2153)	Sigma-Aldrich, St. Louis, MO, USA
CCP10 FISH Probe (# CT-CCP010)	CytoTest, Rockville, MD, USA
Chloramphenicol, 20 mg/ml (# C0378)	Sigma-Aldrich, St. Louis, MO, USA
DAPI (4′,6-diamidino-2-phenylindole) (# D9542)	Sigma-Aldrich, St. Louis, MO, USA
Dextran Sulfate, 50% Solution, 100 ml (# S4030)	Sigma-Aldrich, St. Louis, MO, USA
DNA Gel Loading Solution (# 351-028-661)	Quality Biological, Gaithersburg, MD, USA
DNA Molecular Weight Marker II (# 10236250001)	Roche-Boehringer, Mannheim, Germany
DNA Polymerase I, Kornberg, 5 units/µl (# 104 485)	Roche-Boehringer, Mannheim, Germany
DNase I, from bovine pancreas, 100 mg (# 104 159)	Roche-Boehringer, Mannheim, Germany
dNTPs: dATP (# 105 1440), dCTP (# 105 1482), dGTP (# 105 1466), dTTP (# 105 1458)	Roche-Boehringer, Mannheim, Germany
EDTA, 0.5M, pH 8.0, 100 ml (# 351-027-721)	Quality Biological, Gaithersburg, MD, USA
Ethidium Bromide, UltraPure™, 10 ml (# 15585011)	Invitrogen, Carlsbad, CA, USA
Ethyl Alcohol, Pure, Proof 200	Warner Graham, Cockeysville, MD, USA
Fluorophore DY-415-dUTP, 100 nmol (# 415-34)	Dyomics, Jena, Germany
Fluorophore DY-505-dUTP, 100 nmol (# 505-34)	Dyomics, Jena, Germany
Fluorophore DY-547P1-dUTP, 100 nmol (# 547-34)	Dyomics, Jena, Germany
Fluorophore DY-590-dUTP, 100 nmol (# 590-34)	Dyomics, Jena, Germany

Reagent	Supplier
Fluorophore DY-651-dUTP, 100 nmol (# 651-34)	Dyomics, Jena, Germany
Formamide, deionized, 500 g (# AM9342)	Invitrogen, Carlsbad, CA, USA
GelRed™ Nucleic Acid Gel Stain (# 41001-41003-T)	Biotium, Fremont, CA, USA
GeneRuler 100 bp DNA ladder (# SM0243)	Thermo Fisher Sc., Waltham, MA, USA
Glycerol, 100 ml (# G5516)	Sigma-Aldrich, St. Louis, MO, USA
Isopropanol (2-Propanol), 500 ml (# I9516)	Sigma-Aldrich, St. Louis, MO, USA
Hematoxylin, Mayer's (# 1092490500)	Merck, Darmstadt, Germany
Human Cot-1 DNA™, 1 mg/ml (# 15279011)	Invitrogen, Carlsbad, CA, USA
Human Genomic DNA, female (# G1521)	Promega, Fitchburg, WI, USA
Human Genomic DNA, male (# G1471)	Promega, Fitchburg, WI, USA
Lysogeny broth, Miller (# L3152-1KG)	Sigma-Aldrich, St. Louis, MO, USA
Magnesium Chloride, 2 M, 100 ml (# 340-034-721)	Quality Biological, Gaithersburg, MD, USA
Magnesium Chloride Solution, 25 mM (# M8787)	Sigma-Aldrich, St. Louis, MO, USA
NP-40 Lysis Buffer (# A611-J619-10)	Quality Biological, Gaithersburg, MD, USA
PBS Buffer, 1x, pH 7.2, 1 L (# RGE-3190)	KD Medical, Columbia, MD, USA
Pepsin from porcine gastric mucosa (# P6887)	Sigma-Aldrich, St. Louis, MO, USA
Peroxidase Blocking Solution, Dako (# S202386)	Agilent, Santa Clara, CA, USA
Pertex Mounting Medium (# 41-4011-00)	Medite, Nussloch, Germany
Primer, desalted, variable DNA sequences	Sigma-Aldrich, St. Louis, MO, USA
Proteinase, Bacterial Type XXIV (# P8038)	Sigma-Aldrich, St. Louis, MO, USA
Proteinase K, Puregene (# 158918)	Qiagen, Hilden, Germany
REDTaq® ReadyMix™ PCR Reaction Mix	Sigma-Aldrich, St. Louis, MO, USA
ROTISOLV® HPLC Gradient, ddH$_2$O (# A5112)	Carl Roth, Karlsruhe, Germany
Salmon Testes Sonicated DNA (# GE27-4565-01)	GE Healthcare, Chicago, IL, USA
Sodium Acetate, 3M, pH 5.2, 100 ml (# 351-035-721)	Quality Biological, Gaithersburg, MD, USA
Sodium Chloride, 5M, 100 ml (# 351-036-721)	Quality Biological, Gaithersburg, MD, USA
SSC Buffer, 20x premixed	Roche, Basel, Switzerland
Stabilization and Drying Solution for aCGH	Agilent, Santa Clara, CA, USA
Sulforhodamine 101 (# S7635)	Sigma-Aldrich, St. Louis, MO, USA
SYBR® Green Premix Ex Taq™ II (# RR820A)	Takara Bio Europe, Saint-Germain-en-Laye, France
Target Retrieval Solution, pH 9.0, Dako (# S2367)	Agilent, Santa Clara, CA, USA
TBS buffer (# WL587C2500)	DCS Diagnostics, Hamburg, Germany
Tris-HCl, 1M, pH 8.0, 100 ml (# 351-007-101)	Quality Biological, Gaithersburg, MD, USA
Tween 20, 50% Solution, 20 ml (# 003005)	Thermo Fisher Sc., Waltham, MA, USA
Vectashield Antifade Mounting Medium w/o DAPI	Vector Laboratories, Burlingame, CA, USA
Xylene	Carl Roth, Karlsruhe, Germany

Table 5. List of technical equipment.

Equipment	Supplier
Agilent Microarray Scanner Bundle 8x60K	Agilent, Santa Clara, CA, USA
Autostainer Plus Staining System, Dako	Agilent, Santa Clara, CA, USA
Centrifuge, BeckmanCoulter© AvantiJ25	BeckmanCoulter, Fullerton, CA, USA
Centrifuge, Eppendorf 5417 R	Eppendorf, Hamburg, Germany
Centrifuge, Minifuge GL	Heraeus, Hanau, Germany
Centrifuge, Shandon Cytospin® 3	Thermo Fisher Sc., Waltham, MA, USA
Dri-Block® heater, Techne	Cole-Parmer, Staffordshire, UK
DUET scanning imaging workstation	BioView, Rehovot, Israel
Electrophoresis Chamber, Bio-Rad 200	Bio-Rad Laboratories, Hercules, CA, USA
Fluorescence microscope BX63 with 8 slide-stage	Olympus, Tokyo, Japan
Glass pen with tungsten tip (# 2-5030)	neoLab Migge, Heidelberg, Germany
Glassware	Schott, Mainz, Germany
Hybridization chamber (SureHyb) and gasket slides	Agilent, Santa Clara, CA, USA
Hybridization chamber and gasket slides (HM450K)	Illumina, San Diego, CA, USA
Hybridization oven, temperature set at 65°C	SciGene, Sunnyvale, CA, USA
Hybridization oven rotisserie	SciGene, Sunnyvale, CA, USA
Incubator (20% O2; 5% CO2; 37°C)	Heraeus / Kendro, Düsseldorf, Germany
Incubator and Shaker, Series 25	New Brunswick Scientific, Edison, NJ, USA
iScan Microarray Scanner	Illumina, San Diego, CA, USA
Laboratory scale	Sartorius, Göttingen, Germany
Magnetic hot plate stirrer, Brinkmann® Heidolph®	Heidolph, Schwabach, Germany
Microscope and Scanner M8	PreciPoint, Freising, Germany
Microtome HM560	Microm International, Walldorf, Germany
Ozone-barrier slide cover	Agilent, Santa Clara, CA, USA
Ozone-controlled enclosure	Agilent, Santa Clara, CA, USA
Qubit 4 Fluorometer	Thermo Fisher Sc., Waltham, MA, USA
Real-time Analyzer, ABI StepOne Plus PCR System	Applied Biosystems, Foster City, USA
pH-Meter, digital, WTW pH320	Xylem Analytics, Weilheim, Germany
Phase contrast microscope, Axiovert 25	Zeiss, Jena, Germany
Pipetboy acu 2	INTEGRA Biosciences, Biebertal, Germany
Pipetus®	Hirschmann, Eberstadt, Germany
Pipettes, Eppendorf Research® plus (0.1-2.5 µl, 0.5-10 µl, 2-20 µl, 10-100 µl, 20-200 µl, 100-1000 µl)	Eppendorf, Hamburg, Germany
PyroMark Q24 cartridge	Qiagen, Hilden, Germany
PyroMark Q24 device	Qiagen, Hilden, Germany
PyroMark Vacuum Prep Station	Qiagen, Hilden, Germany
Scale, BP3015	Sartorius, Göttingen, Germany
Shaker MTS4	IKA-Werke, Staufen, Germany
Slide-staining dish with rack, 250 ml and 1.5 l	DWK Life Sciences, Wertheim, Germany

Equipment	Supplier
SOLO satellite workstation	BioView, Rehovot, Israel
Spectrophotometer, NanoDrop™ ND-1000	Thermo Fisher Sc., Waltham, MA, USA
Speed Vacuum Concentrator, SpeedVac	Savant, Ramney, MN, USA
SureScan mircoarray scanner (# G2565BA)	Agilent, Santa Clara, CA, USA
Thermal Cycler with heated lid, peqSTAR®	PEQLAB, Erlangen, Germany
Thermomixer, Eppendorf 5436	Eppendorf AG, Hamburg, Germany
ThermoBrite StatSpin System	Abbott Molecular, Des Plaines, IL, USA
UV-transilluminator	Intas Science Imaging, Göttingen, German
Vortex MS2 Minishaker	IKA-Werke, Staufen, Germany
Vortex-Genie 2	Scientific Industries, Woburn, USA
Water bath (37°C, 45°C, 60°C), VWR 1245	VWR International, West Chester, PA, USA

Table 6. List of software and databases.

Software	Supplier
Bioconductor software package *minfi* (*R*-based)	http://www.bioconductor.com
BioView, Version 3.6.0.16	BioView, Rehovot, Israel
CpG Island Finder database	http://www.dbcat.cgm.ntu.edu.tw
DotEditor, Version 0.3.1	http://www.vincenthee.github.io/DotEditor/
ENSEMBL database	http://www.ensembl.org
Excel Spreadsheet 2016	Microsoft, Redmond, WA, USA
Feature Extraction Software, Version 10.7.1.1	Agilent, Santa Clara, CA, USA
FISHtrees	ftp://ftp.ncbi.nlm.nih.gov/pub/FISHtrees
GeneCards database	http://www.genecards.org
Gene Cluster, Version 3.0	Laboratory of DNA Information Analysis, University of Tokyo, Japan
Gene Expression Omnibus (GEO) database	https://www.ncbi.nlm.nih.gov/geo/
Gene Set Enrichment Analysis (GSEA)/MSigDB v6.1	http://software.broadinstitute.org/gsea/msigdb/annotate.jsp
GenomeStudio® Methylation Module, Version 1.8	Illumina, San Diego, CA, USA
GraphPad Prism 6.0 Statistics	GraphPad Software, La Jolla, CA, USA
Inkscape, Version 0.92.1	http://www.inkscape.org/en/
Java Tree View, Version 1.1.6r4	http://www.jtreeview.sourceforge.net
NEXUS Copy Number, Version 8.0	BioDiscovery, El Segundo, CA, USA
Photoshop, CS2	Adobe Systems, San Jose, CA, USA
PowerPoint 2016	Microsoft, Redmond, WA, USA
Pyro Q-CpG Software	Qiagen, Hilden, Germany
PyroMark Assay Design Software	Qiagen, Hilden, Germany
REST, Relative Expression Software Tool 2009	Qiagen, Hilden, Germany
Venn Diagram Plotter, Version 1.5.5	https://omics.pnl.gov/software/
Word 2016	Microsoft, Redmond, WA, USA

2.2 Patient samples

Formalin-fixed paraffin embedded (FFPE) tissue samples were collected from the tissue archive of the Institute of Pathology of the University Medical Centre Mannheim between 2002 and 2014. In total, 87 (68 biological individuals) FFPE colorectal adenoma specimens (and three additional normal mucosae samples) were obtained. The selection was performed by reviewing the electronic clinical database of the Central Interdisciplinary Endoscopy Unit of the University Medical Centre Mannheim of patients who underwent endoscopic resection of adenomas during sigmoidoscopies or colonoscopies (Table 7). Polyp size was determined by the endoscopist or by pathological measurement. Due to the fragmentation of nearly all samples, pathological evaluation of *in toto* polypectomy was not feasible. However, endoscopic removal of polyps was indicated as complete for all lesions by experienced endoscopists. H&E-stained tissue sections were utilized for histopathological diagnosis by experienced board-certified pathologists blinded for the clinical data and based on the current WHO classification (Bosman et al., 2010). Thus, conventional colorectal adenomas were classified by histological subtypes into tubular, tubulo-villous or villous with low-grade dysplasia (LGD) (Figure 14A) or high-grade dysplasia (HGD) (Figure 14B). The endoscopists defined and categorized the localizations of polyps by segments: the right hemicolon (cecum, ascending colon, hepatic flexure, transverse colon), the left hemicolon (splenic flexure, descending colon and sigmoid colon) and the rectum, respectively.

Colorectal adenoma samples were categorized by the status of recurrence into three groups (Table 8): (*i*) primary adenomas without documented recurrence in the follow-up period (*n*=30), (*ii*) primary adenomas presenting adenoma recurrence at the same location within the follow-up period (*n*=19) and (*iii*) matched pairs (*n*=19) consisting of the primary adenoma and the corresponding recurrent adenoma, respectively.

The study was approved by the local board of ethics (Medizinische Ethikkommission II, University of Heidelberg, ethics approval identifier: *2012-608R-MA*) and was consistent with the Declaration of Helsinki (World Medical Association, 2013). Furthermore, the use of the subsequently coded and de-identified data was granted by the National Institutes of Health's "Office of Human Subjects Research" (OHSR *#13220*).

With regards to limited tissue material and increased labor-intensity of the methods, not all assays utilized in this study could be performed on each specimen (Figure 15).

Table 7. Clinicopathological features of specimens of the adenoma cohort. Clinical parameters of primary adenomas without recurrence, primary adenomas with recurrence, recurrent adenomas (matched pairs) and normal mucosae are listed below. F, female; M, male; ND, not determined.

|	Sample ID	Gender	Age at di-agnosis (y)	Grade of dysplasia	Recurrence	Observation time / recur-rence (m)	Histology	Localization	Size (mm)
Primary adenoma *w/o recurrence*	A 1	M	66.7	Low-grade	no	27.1	Tubulo-villous	Ascending	20
	A 2	M	66.0	Low-grade	no	12.1	Tubular	Cecum	18
	A 3	M	50.3	Low-grade	no	14.1	Tubulo-villous	Rectum	35
	A 4	M	67.2	Low-grade	no	38.0	Tubular	Descending	20
	A 5	M	67.3	Low-grade	no	47.2	Tubulo-villous	Hepatic flexure	20
	A 6	M	68.6	Low-grade	no	19.6	Villous	Rectum	40
	A 7	F	63.9	Low-grade	no	23.9	Tubulo-villous	Cecum	25
	A 8	M	70.3	Low-grade	no	45.4	Tubulo-villous	Sigmoid	12
	A 9	F	68.2	Low-grade	no	21.9	Tubulo-villous	Ascending	20
	A 10	M	62.8	Low-grade	no	20.4	Villous	Rectum	10
	A 11	F	68.5	Low-grade	no	47.5	Tubulo-villous	Sigmoid	30
	A 12	M	66.3	Low-grade	no	3.4	Tubular	Cecum	40
	A 13	F	71.8	Low-grade	no	12.5	Tubulo-villous	Ascending	12
	A 14	M	50.9	Low-grade	no	10.9	Tubular	Transverse	15
	A 15	M	66.6	Low-grade	no	36.5	Tubular	Ascending	25
	A 31	F	66.3	Low-grade	no	ND	Tubulo-villous	Rectum	20
	A 32	F	83.6	Low-grade	no	50.6	Tubulo-villous	Transverse	20
	A 33	F	69.7	Low-grade	no	ND	Tubulo-villous	Cecum	30
	A 34	M	67.3	Low-grade	no	ND	Tubular	Rectum	10
	A 35	F	75.4	Low-grade	no	ND	Tubular	Sigmoid	8
	A 36	M	76.0	Low-grade	no	ND	Tubulo-villous	Sigmoid	5
	A 37	M	67.9	Low-grade	no	ND	Tubular	Rectum	12
	A 38	M	74.1	Low-grade	no	ND	Tubulo-villous	Ascending	20
	A 39	F	84.4	Low-grade	no	ND	Tubulo-villous	Caecum	20
	A 40	F	62.9	Low-grade	no	ND	Tubulo-villous	Ascending	15
	A 41	F	63.4	Low-grade	no	ND	Tubulo-villous	Hepatic flexure	20
	A 42	F	58.2	Low-grade	no	ND	Tubular	Cecum	12
	A 43	F	55.7	Low-grade	no	ND	Tubulo-villous	Rectum	23
	A 44	F	71.0	Low-grade	no	ND	Tubular	Sigmoid	12
	A 45	M	69.0	Low-grade	no	ND	Tubular	Descending	20
Primary adenoma *w/ recurrence*	A 16	M	69.2	Low-grade	Primary	24.2	Tubulo-villous	Ascending	15
	A 17	M	68.5	Low-grade	Primary	7.7	Tubulo-villous	Rectum	20
	A 18	M	55.1	Low-grade	Primary	30.4	Tubulo-villous	Sigmoid	42
	A 19	M	64.9	Low-grade	Primary	6.6	Tubulo-villous	Cecum	20
	A 20	F	70.2	Low-grade	Primary	41.8	Tubulo-villous	Hepatic flexure	40
	A 21	M	62.7	Low-grade	Primary	21.0	Tubular	Ascending	20
	A 22	M	73.0	Low-grade	Primary	8.0	Tubulo-villous	Rectum	25
	A 23	F	64.9	Low-grade	Primary	17.2	Tubular	Transverse	6
	A 25	M	61.9	Low-grade	Primary	5.3	Tubulo-villous	Transverse	ND
	A 26	F	79.0	Low-grade	Primary	60.1	Tubulo-villous	Transverse	20
	A 28	F	55.9	Low-grade	Primary	8.4	Tubulo-villous	Rectum	35
	A 30	M	78.0	Low-grade	Primary	14.5	Tubulo-villous	Rectum	30
	A 46	F	52.3	Low-grade	Primary	3.1	Tubulo-villous	Rectum	63
	A 47	F	58.8	Low-grade	Primary	54.2	Tubulo-villous	Cecum	40
	A 48	M	52.0	Low-grade	Primary	7.0	Tubulo-villous	Ascending	ND
	A 49	M	63.6	Low-grade	Primary	52.7	Tubular	Ascending	7
	A 50	F	72.9	Low-grade	Primary	26.6	Tubulo-villous	Cecum	12
	A 51	F	70.7	Low-grade	Primary	36.3	Tubular	Cecum	18
	A 52	M	63.5	Low-grade	Primary	3.2	Tubulo-villous	Rectum	40
Normal	N 1	M	51.7	Low-grade	Normal	ND	ND	Descending	ND
	N 2	M	52.6	Low-grade	Normal	ND	ND	Transverse	ND
	N 3	M	69.1	Low-grade	Normal	ND	ND	Transverse	ND

Continued next page.

Table 7. (continued).

	Sample ID	Gender	Age at di-agnosis (y)	Grade of dysplasia	Recurrence	Observation time / recurrence (m)	Histology	Localiza-tion	Size (mm)
	A 24	F	45.3	Low-grade	Primary	19.9	Tubulo-villous	Ascending	25
	A 24b	F	46.8	Low-grade	Recurrent		Tubular	Ascending	10
	A 27	M	87.9	Low-grade	Primary	19.1	Tubulo-villous	Rectum	94
	A 27b	M	89.3	Low-grade	Recurrent		Tubulo-villous	Rectum	ND
	A 29	M	37.9	Low-grade	Primary	24.9	Tubulo-villous	Cecum	40
	A 29b	M	39.8	Low-grade	Recurrent		Tubular	Cecum	ND
	A 53	F	69.8	Low-grade	Primary	32.6	Tubulo-villous	Cecum	35
	A 53b	F	72.5	Low-grade	Recurrent		Tubulo-villous	Cecum	15
	A 54	F	64.3	Low-grade	Primary	1.9	Tubulo-villous	Rectum	20
	A 54b	F	64.4	Low-grade	Recurrent		Tubulo-villous	Rectum	ND
	P 1a	F	81.9	Low-grade	Primary	26.4	Tubulo-villous	Rectum	84
	P 1b	F	84.1	Low-grade	Recurrent		Tubulo-villous	Rectum	20
	P 2a	F	53.6	Low-grade	Primary	29.1	Tubulo-villous	Sigmoid	45
	P 2b	F	56.0	Low-grade	Recurrent		Tubulo-villous	Sigmoid	ND
	P 3a	F	54.4	Low-grade	Primary	12.8	Tubulo-villous	Sigmoid	20
	P 3b	F	55.4	Low-grade	Recurrent		Tubulo-villous	Sigmoid	ND
	P 4a	M	63.8	Low-grade	Primary	47.4	Tubulo-villous	Cecum	30
	P 4b	M	67.6	Low-grade	Recurrent		Tubulo-villous	Cecum	30
	P 5a	F	73.8	Low-grade	Primary	11.8	Tubulo-villous	Sigmoid	30
	P 5b	F	74.7	Low-grade	Recurrent		Tubulo-villous	Sigmoid	20
	P 6a	M	87.1	Low-grade	Primary	9.5	Tubulo-villous	Cecum	40
	P 6b	M	87.9	Low-grade	Recurrent		Tubulo-villous	Cecum	19
	P 7a	F	72.1	Low-grade	Primary	6.0	Tubulo-villous	Ascending	20
	P 7b	F	72.6	Low-grade	Recurrent		Tubulo-villous	Ascending	20
	P 8a	F	64.2	High-grade	Primary	7.6	Tubulo-villous	Rectum	50
	P 8b	F	64.9	High-grade	Recurrent		Tubulo-villous	Rectum	15
	P 9a	F	78.7	High-grade	Primary	20.8	Tubulo-villous	Rectum	15
	P 9b	F	80.4	High-grade	Recurrent		Villous	Rectum	20
	P 10a	M	69.9	High-grade	Primary	9.8	Tubulo-villous	Sigmoid	15
	P 10b	M	70.8	High-grade	Recurrent		Tubulo-villous	Sigmoid	15
	P 11a	F	72.7	High-grade	Primary	5.0	Tubular	Ascending	30
	P 11b	F	73.0	High-grade	Recurrent		Tubular	Ascending	12
	P 12a	M	64.1	High-grade	Primary	15.1	Tubulo-villous	Cecum	15
	P 12b	M	65.3	High-grade	Recurrent		Tubulo-villous	Cecum	10
	P 13a	F	68.2	High-grade	Primary	24.1	Tubulo-villous	Rectum	25
	P 13b	F	70.2	High-grade	Recurrent		Tubular	Rectum	8
	P 14a	M	52.3	High-grade	Primary	95.9	Tubulo-villous	Sigmoid	60
	P 14b	M	60.2	High-grade	Recurrent		Tubular	Sigmoid	10

Matched pairs: primary adenoma and corresponding recurrent adenoma (left side label)

Table 8. Patient groups. ND, not determined.

Group	Definition	Gender		Age at diagnosis (y) mean ± SD median (IQR)	Size (mm) mean ± SD median (IQR)	Observation time (m) mean ± SD median (IQR)
		Male	Female			
I (*n*=30)	Primary adenomas *without recurrence*	16	14	67.3 ± 7.5 / 67.3 (63.8-70.5)	19.6 ± 8.8 / 20 (12-23.5)	26.9 ± 15.2 / 22.9 (12.9-43.6)
II (*n*=38)	Primary adenomas *with recurrence*	18	20	65.8 ± 10.9 / 64.9 (58.1-72.8)	31.8 ± 19.6 / 27.5 (20-40)	22.3 ± 19.6 / 18.2 (7.7-29.4)
III (*n*=19)	Recurrent adenomas (matched pairs)	7	12	68.2 ± 13.0 / 70.2 (60.2-74.7)	16 ± 5.9 / 15 (10-20)	22.1 ± 21.1 / 19.1 (9.5-26.4)
IV (*n*=3)	Normal mucosa	3	0	57.8 ± 9.8 / 52.6 (51.7-69.1)	ND / ND	ND / ND
Σ (*n*=90)	Cohort	44	46	66.5 ± 10.4 / 67.3 (62.5-72.5)	24.5 ± 15.8 / 20 (15-30)	23.3 ± 19 / 19.9 (9.5-29.8)
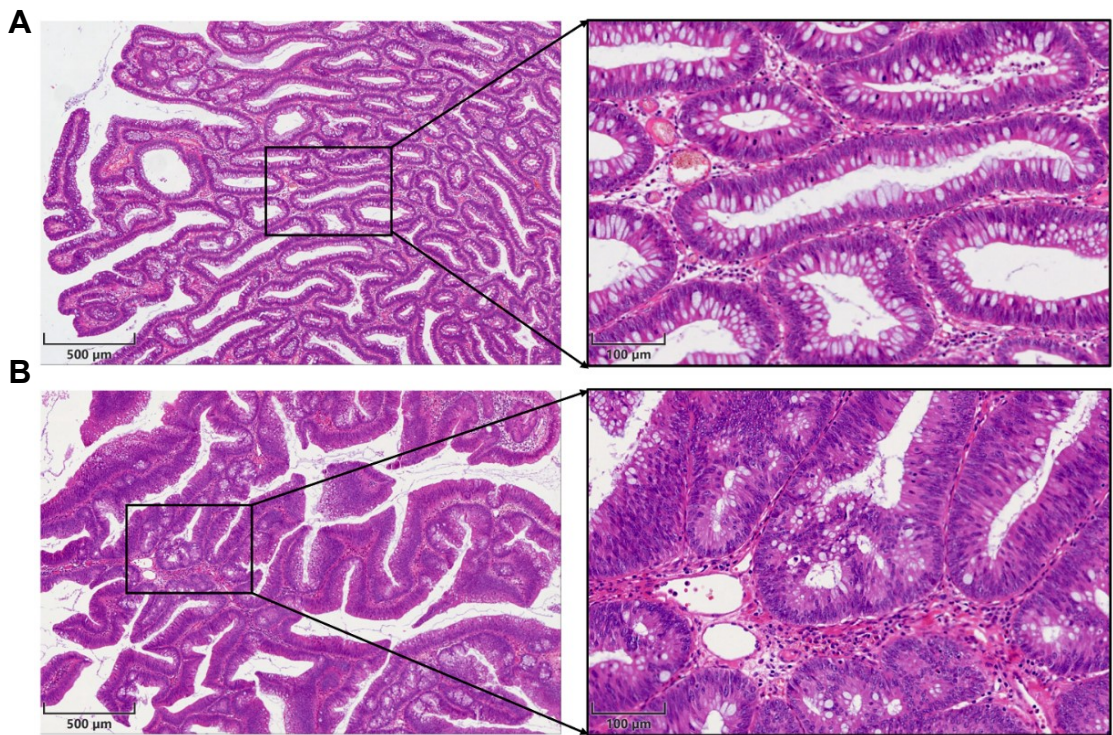
Figure 14. Histological images of colorectal adenomas. (**A**) Representative overview (*left*) of a conventional colonic adenoma (A24) with tubulo-villous histologic architecture and low-grade dysplasia. Zoom-in (*right*) highlights the hyperchromatic cells with a markedly shifted nuclei-to-cytoplasm ratio. (**B**) Representative overview (*left*) of a conventional colonic adenoma (P14a) of tubulo-villous histology and high-grade dysplasia. Zoom-in (*right*) magnifies the multilayered irregular nuclei in atypical crypts with a visibly distorted cytology. Images were captured with *PreciPoint Microscope & Scanner*. Scale bars overview, 500 µm; Scale bars zoom, 100 µm.
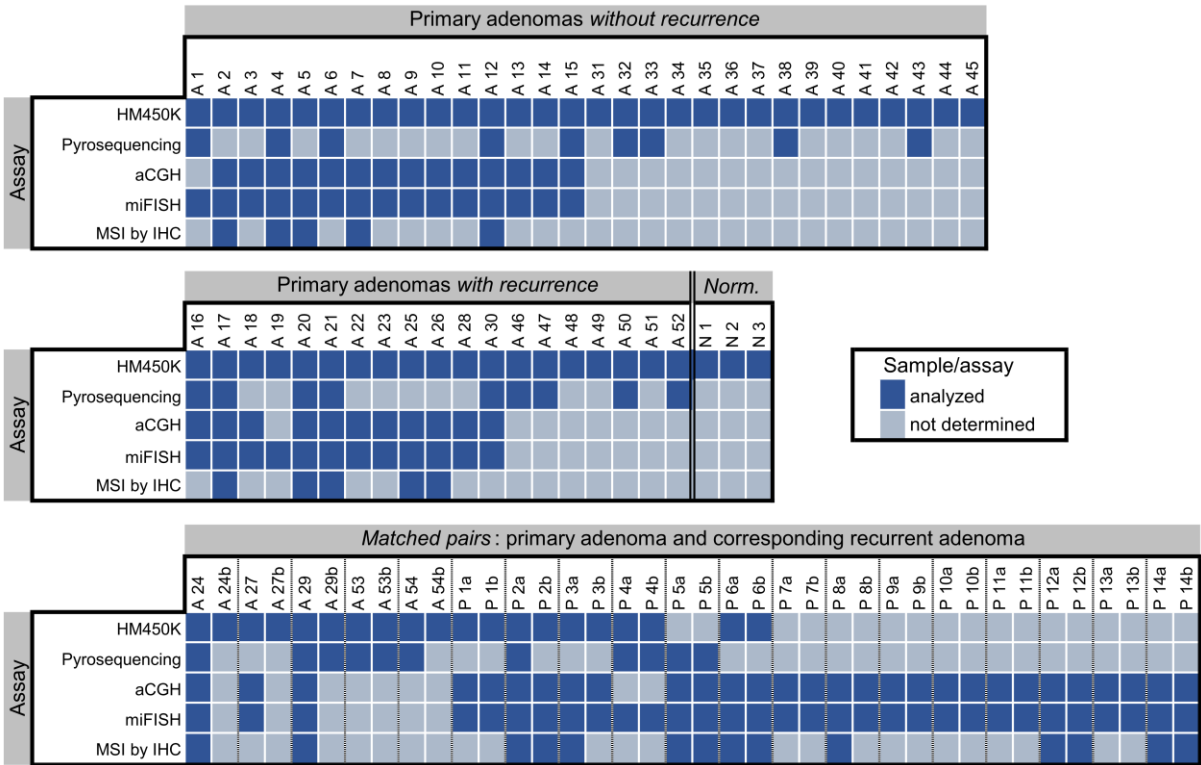
Figure 15. Patient samples analyzed by individual genetic and epigenetic assays. Adenoma (*n*=87) and normal (*n*=3) specimens were targets of analysis by two epigenetic approaches, *i.e.*, HM450K BeadChip array and pyrosequencing of *GREM2*, and additionally by two (cyto-)genetic assays, *i.e.*, aCGH and miFISH, and IHC to determine microsatellite status. Norm., normal mucosa.

2.3   DNA extraction from FFPE tissue

Archived FFPE tissue blocks were cut into four consecutive sections by a microtome. Tissue sections were carefully applied onto Duran® microscopy slides with the aid of a paintbrush prior to incubation at 62°C for 30 min for drying. The first section (4 µm-thick) was stained by H&E to mark the region of interest, comprising at least 70% of adenoma cells, by an experienced board-certified pathologist. The next three un-stained sections (each 10 µm-thick) were deparaffinized by repeated incubation in xylene for 10 min. Dehydration was continued by exposing the tissue twice to 100% EtOH for 10 min. Tissues slides were then taken for conscientious macro-dissection with a scalpel guided by the markings of the formerly H&E-stained slide. Scraped tissues were collected in 1.5 ml tubes and air-dried overnight at RT. DNA was extracted utilizing the Gentra Puregene Tissue Kit (Qiagen) following the manufacturer's protocol. The procedure started by adding 300 µl cell lysis solution into the tube and followed by incubation (65°C for 60 min at 1,100 rpm). Subsequently, 1.5 µl Puregene proteinase K was added to the tube which was then 25x inverted until mixed thoroughly. The vial was incubated in a Thermomixer (55°C overnight at 450 rpm). The next day 1.5 µl RNase A was added into the solution which was 25 times inverted prior to incubation (37°C for 60 min). Samples were chilled on ice for 1 min. 100 µl protein precipitation solution were added and the tube was mixed on a "Vortex" for 2 min. Proteins were pelleted by spinning the tube (13,000 g for 3 min at RT). If the formed pellet was not tight, the sample was incubated on ice for 5 min and centrifugation was subsequently repeated. The supernatant was subsequently poured into another vial which contained 300 µl isopropanol. The tube was inverted up to 50 times and spun (13,000 g for 3 min at RT). The supernatant was discarded while the remaining pellet was washed by adding 500 µl of 70% EtOH and thorough mixing. The tube was centrifuged (13,000 g for 1 min at RT), the supernatant was pipetted off and the tube was inversely drained on absorbent cleaning tissue for 15 min. Residual liquid was taken off by carefully wiping an absorbent lens paper around the vial's walls. DNA was eluted by repeatedly pipetting up and down 22 µl of ddH$_2$O. The vial was mixed ("Vortex" for 0.5 min) and incubated on a Thermomixer (65°C for 60 min at 450 rpm) to completely dissolve the DNA. Extracted DNAs were gently shaken at RT overnight. The following day, extraction-yield and 260/230-ratio were spectrophotometrically quantified by NanoDrop ND-1000. The DNAs were stored at -20°C.

2.4 CpG dinucleotides methylation analysis

2.4.1 Real-time-PCR quality check

The quality of extracted DNAs (chapter 2.3) was assessed by performing the Infinium HD FFPE QC Real-time PCR assay (Illumina). Briefly, DNAs were diluted to 1 ng/µl in ddH$_2$O, and the 100-fold QC-template reagent was aliquoted (10 µl QC-template and 990 µl ddH$_2$O). The assay was performed in triplicates (3 wells/sample). Wells were either pipetted with 2 µl 100-fold QC-template (standardized control template), 2 µl genomic DNA (sample DNA) or 2 µl ddH$_2$O (NTC), respectively. To each well, 8 µl of the qPCR premix (5 µl SYBR Green MM, 1 µl QC-primer, 2 µl ddH$_2$O) were added. The plate was sealed and spun at 280 g prior to placing it in the qPCR thermal cycler (Table 9). The quality of the amplification was evaluated by the melting curve. If more than one specific PCR product existed, the different GC content would generate two independent melting curves. Obtained data on threshold cycles (C$_t$) was analyzed following the MIQE guidelines (Bustin et al., 2009; Livak and Schmittgen, 2001). Sample DNA would pass the QC test to be selected for further experiments if the ΔC$_t$ value ranged below 5, which was recommended by the manufacturer's protocol (Illumina).

Table 9. Thermal profile real-time-PCR.

Cycles	Temperature	Time
1 x	50°C	2 min
1 x	95°C	10 min
40 x	95°C	30 sec
	57°C	30 sec
	72°C	30 sec
1 x	4°C	∞

$$Average\ C_t = \frac{C_t1 + C_t2 + C_t3}{3} \tag{1}$$

$$\Delta C_t = average\ C_t\text{Sample} - average\ C_t\text{QC template} \tag{2}$$

2.4.2 Bisulfite conversion

Extracted FFPE DNAs that passed the quality check were treated by sodium bisulfite conversion using the EpiTect Plus DNA Bisulfite Kit to convert unmethylated cytosines. The bisulfite mix was dissolved in 800 µl ddH$_2$O and heated to 60°C. The bisulfite reactions were set up in PCR tubes in a distinct order (Table 10). The vials were mixed thoroughly and incubated in a thermal cycler (Table 11). Subsequently, bisDNAs were

transferred into 1.5 ml reaction tubes and cleaned-up by adding 310 µl buffer BL (including 10 µg/ml carrier RNA) and 250 µl 100% EtOH. Upon mixing on a "Vortex" and spinning of the tubes, the content was transferred onto the MiniElute DNA spin columns prior to centrifugation (13,000 g for 1 min). Flow through was discarded, the columns were washed with 500 µl buffer BW and spun (13,000 g for 1 min). Again, the flow through was discarded and 500 µl buffer BD incubated on the column for 15 min at RT. The columns were spun (13,000 g for 1 min) and a repetitive wash step by adding 500 µl buffer BW and subsequent centrifugation (13,000 g for 1 min) were performed. 250 µl 100% EtOH were added and the tubes centrifuged (13,000 g for 1 min). Spin columns were incubated at 60°C for 5 min to evaporate residual liquids. DNAs were eluted with 15 µl buffer EB for 1 min at RT and centrifugation (15,000 g for 1 min). Purified bisDNA was stored at -20°C.

Table 10. Bisulfite reaction setup.

Reagent	Per reaction
DNA solution	X µl (maximum 20 µl)
ddH$_2$O	DNA solution ad 20 µl
Bisulfite mix	85 µl
DNA protect buffer	35 µl
Final volume	140 µl

Table 11. Thermal profile bisulfite conversion.

Cycles	Temperature	Time
1 x	95°C	5 min
1 x	60°C	25 min
1 x	95°C	5 min
1 x	60°C	85 min
1 x	95°C	5 min
1 x	60°C	175 min
1 x	20°C	∞

2.4.3 DNA methylation microarray: HumanMethylation450K BeadChip

DNA methylation array

Sixty-nine colorectal adenoma DNAs and three normal tissue DNAs were selected for genome-wide methylation analysis by HM450K array (Figure 15). The assay comprises 485,577 probes of CpG sites which allows the quantification of DNA methylation in >99% of RefSeq genes with an average correlation of technical replicates of R^2=0.992 (Bibikova et al., 2011). Performed steps were conducted in strict accordance

41

with the detailed manufacturer's standard protocol for FFPE samples (Illumina user guide, 2012). Very briefly, DNA samples were quantified by Qubit fluorometer, bisulfite-converted (approx. 250 ng), whole-genome amplified (WGA), enzymatically fragmented and then precipitated prior to re-suspension in the hybridization buffer (Bibikova et al., 2011). Subsequently, DNAs were hybridized onto six Infinium HumanMethylation450K BeadChip arrays (Illumina) in a hybridization oven for 20h. Aiming to minimize the technical bias of the array (Dedeurwaerder et al., 2011), sample DNAs were hybridized, subsequently after bisulfite conversion took place, in a random distribution to avoid batch effects. After the hybridization procedure, the array was treated with a primer extension and immunohistochemically detected before a coating step was performed. The arrays' probe signals were detected by an iScan array scanner (Illumina), and the raw data *idat*-files were exported with GenomeStudio software prior to data processing.

(Bioinformatical) data processing and analysis

Idat-files were loaded into GenomeStudio software to assess quality control metrics for both specimens and probes. Normalization of signal intensities, background level correction, and color adjustment were performed via Subset-quantile within array normalization (SWAN) method (Maksimovic et al., 2012), which is incorporated in the Bioconductor *R*-package *minfi* (Aryee et al., 2014). Methylation of CpG positions was measured on the array by fluorescent signals from the methylated (M) and unmethylated (U) alleles. Levels of methylation were then assessed per sample and CpG site by calculation of the ratio of signal intensities (beta) as described by Bibikova et al. (2011):

$$Signal\ intensity\ \beta = \frac{intensity\ (methylated)}{intensity\ (unmethylated + intensity\ (methylated) + 100} \qquad (3)$$

The β-value reflects a quantitative measure ranging from zero to one, *i.e.*, fully unmethylated to fully methylated (Martin-Subero et al., 2009).

Prior to DNA methylation analysis, probes which harbor potential biological and/or technical bias had to be bioinformatically excluded (Dedeurwaerder et al., 2014). A total of 156,004 probes were removed as the following criteria applied (multiple probes overlap across lists): (*i*) cross-reactive polymorphic CpG positions (Chen et al., 2013), (*ii*) non-specific probes (Chen et al., 2013), (*iii*) age-dependent CpG sites (Rakyan et al., 2010) and (*iv*) probes containing SNPs at/near the targeted CpG position (Wang et al., 2012). Thus, 329,573 probes remained for further analysis.

Statistical analysis

Methylation levels were evaluated to identify significantly differentially methylated CpG positions (DMPs) when comparing primary adenomas without recurrence, primary adenomas with recurrence, recurrent adenomas, and normal mucosa. Significant differentially hyper- or hypomethylated sites were defined by applying Benjamini Hochberg-corrected (*i.e.*, multiple testing corrected) false discovery rates (FDR q-value) below 0.05 and 0.01 (Mann-Whitney *U* test), respectively, and at the same time a fold change (Δ β-value) showing a difference of >10% (see Galamb et al. 2016). Although this threshold seems comparably low (Dedeurwaerder et al., 2011; Sharma et al., 2016), the sample set comprised an adequate number (comparisons included tumor vs tumor) to reduce biased (*i.e.*, false positive) discoveries. Additionally, pre-selected CpG sites were further categorized into "*top genes*," *i.e.*, most differentially hyper/-hypomethylated sites, which not only applied to criteria mentioned above but also showed absolute changes in Δ β exceeding ±0.1. Among the *top genes* probes targeting CpGs on sex-chromosomes were excluded to avoid gender-specific bias. Calculations were performed utilizing the bioinformatics software *R*.

DMPs were compared for overlaps across defined sample groups using Venn Diagram Plotter software. The distribution of DMPs was studied by plotting the CpG positions according to their linear location within CpG islands (CGI; stretch of 0.5-2kb), CpG shores (<2kb from CGI), CpG shelves (2-4kb from CGI) or open sea (>4kb from CGI), respectively (Bibikova et al., 2011; Schneider et al., 2016). The distribution of DMPs was also analyzed by association within gene regions (proximal promoter regions, gene body, and 3'UTR) and by chromosomal location (Schneider et al., 2016, Sharma et al., 2016). Identified *top gene*-DMPs were further used to analyze whether site-specific DNA methylation correlates with distinct biological pathways. A gene set enrichment analysis (GSEA) using the molecular signatures database (MSigDB of the Broad Institute) was carried out to compute the overlaps of gene-associated DMPs with curated gene sets involving pathways listed in BioCarta, KEGG, NABA, PID, REACTOME, and SIG, respectively. Evaluated were pathways only which scored FDR q<0.001 to guarantee highly significant discoveries.

Unsupervised hierarchical clustering of samples was performed by probes reaching *P*<0.05 using two-way cluster analysis (Euclidean distance and Ward's group linkage method). Cluster associations were correlated with clinicopathological data applying Spearman's correlation coefficient. Principal component analysis (PCA) was carried out to investigate the distance of samples via average DNA methylation.

2.4.4 Pyrosequencing

Assay-design and pre-PCR

Results obtained by methylation analysis of the HM450K (2.4.3) were orthogonally validated using pyrosequencing (Tost et al., 2003) by quantifying the methylation levels of selected CpG sites within the gene body of *GREM2* (Figure 16A). The gene was chosen due to the pivotal role in the BMP pathway whose alteration influences tumorigenesis (Kosinski et al., 2007; Loh et al., 2008). The *PyroMark Assay Design Software* was used to design assays by incorporating the selected differential CpG dinucleotides in the sequence to analyze (Table 12). Pre-PCR for initial fragment amplification was performed by adding the reagents into a PCR tube (Table 13). Samples were amplified in a PCR thermal cycler (Table 14). PCR products were verified by running 5 µl per sample mixed with 1 µl loading dye on an agarose-gel electrophoresis (1% agarose in 50 ml TAE buffer with 3 µl GelRed) for 30 min with 120 V. GeneRuler 100 bps DNA ladder (1 µl) was used as reference and the gel was analyzed using an UV-transilluminator. If a PCR product were visible, the sample would qualify for the sequencing reaction.

Table 12. Pyrosequencing primers for pre-PCR and sequencing.

Assay	Primer	DNA sequence 5'→ 3' (°biotinylated)	Length
GREM2 (FOR-Assay)	*Forward*	AGTATTTTAAGAGTGATTGGTGTAA	158 bp
	Reverse	°CTAAAAAAACTCCTCCTCCTTCTTCA	
	Sequencing	TGAGYGAGGAGGGTTG	
GREM2 (REV-Assay)	*Forward*	°AAGGAGGAGGAGTTTTTTTAGT	135 bp
	Reverse	ACCRACACTACTTCACCTTCTAAATTTTCT	
	Sequencing	CTTCACCTTCTAAATTTTCTTA	

Table 13. Pyrosequencing pre-PCR setup.

Reagent	Per reaction
REDTaq ready mix	12.5 µl
Primer mix	1 µl (10 µM For/10 µM Rev)
MgCl$_2$	1 µl (25 µM)
bisDNA	2 µl (~100 ng)
HPLC-grade ddH$_2$O	8.5 µl
Final volume	25 µl

Table 14. Thermal profile pyrosequencing pre-PCR.

Cycles	Temperature	Time
1 x	95°C	15 min
45 x	94°C	30 sec
	56°C	30 sec
	72°C	30 sec
1 x	72°C	10 min
1 x	4°C	∞

Sequencing

PCR product was purified utilizing the biotinylated strand which was then analyzed in a sequencing reaction. Sequencing of *GREM2* covered by the FOR-assay a 37-bp sequence of interest (TYGGAGTYGTATTATTTTTAATYGTTTTTGTTAYGGT) analyzing four sequential CpG sites, while the REV-assay covered a 47-bp sequence of interest (AATCRAAAAAATAAATCCAAACCRAAACACTCRAACTCCACRAAAAC) incorporating four sequential CpG sites (Figure 16A). Bisulfite-converted DNA of glioblastoma cell lines (LN229; T98G) served as technical replicates for assay validation. Each pyrosequencing run was additionally performed with these controls, *e.g.*, hypermethylated CpGs, to confirm the reliability and robustness of the obtained pyrograms.

Performing the sequencing reactions started by setting up (multiplied by *n* samples) the bead-solution (5 µl beads with 35 µl binding buffer) and the primer-solution (0.8 µl primer [20 pM] with 39.2 µl annealing buffer), respectively. The latter one was applied (40 µl) into the wells of the sequencing plate. Instead, bead-solution (40 µl) and pre-PCR product (40 µl) were mixed in the wells of a cell culture plate and shaken for 5 min. The mix was soaked up by the vacuum pump of the PyroMark vacuum prep station, subsequently and sequentially immersed in EtOH (70%), NaOH and washing buffer prior to setting off the vacuum pump to release the bead-bound product onto the sequencing plate, which was heated at 80°C for 2 min and cooled to RT. Meanwhile, suitable volumes (determined by software Pyro-Q-CpG) of enzyme, substrate, and dNTPs were added into the cartridge which was then placed in the PyroMark Q24 device. The sequencing plate was set into the PyroMark device, and the reaction was started by programming the assay using Pyro Q CpG software. Once the sequencing reaction was completed, "pyrograms" were evaluated for mean CpG methylation of the sequenced CpG positions (Figure 16B).

A

B
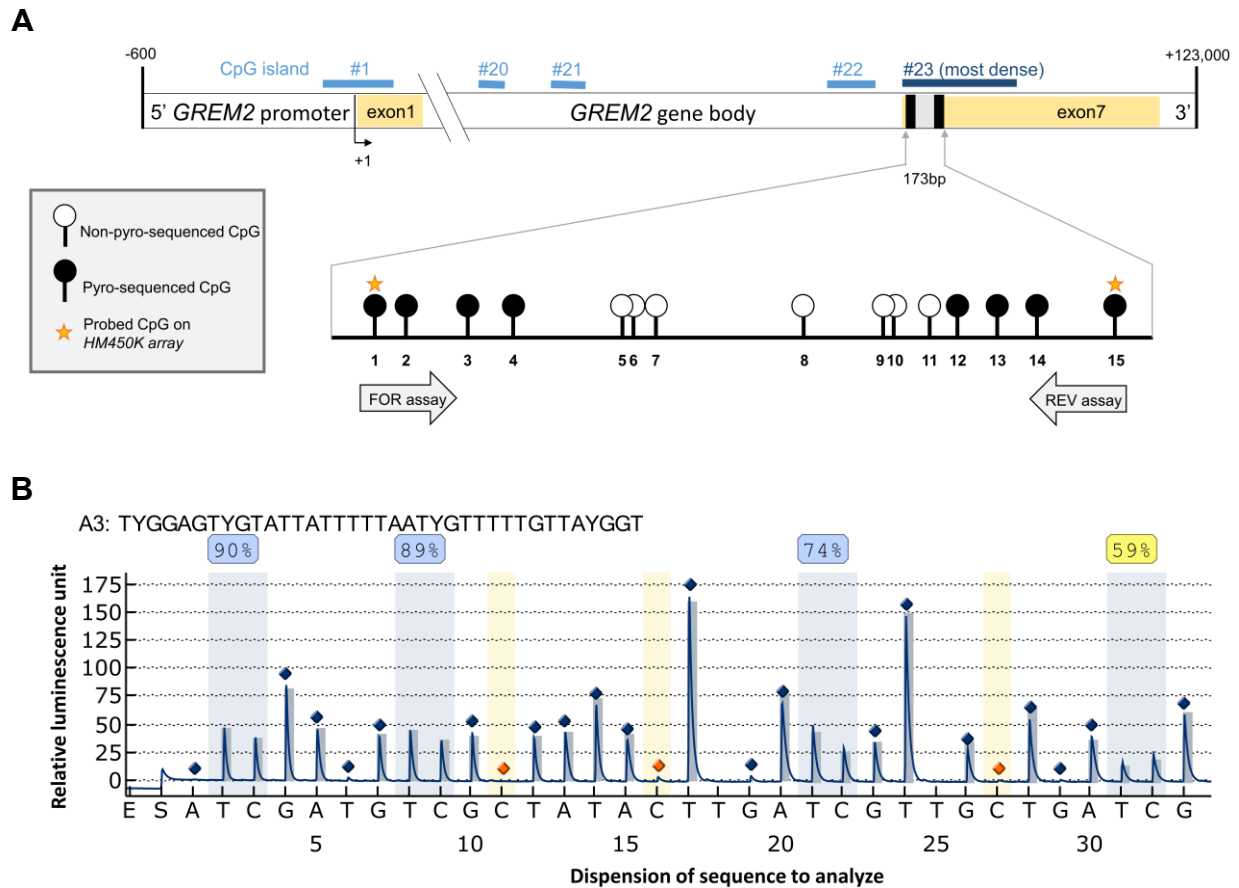
Figure 16. *GREM2* **gene card and representative "pyrogram".** (**A**) Schematic map of the human *GREM2* gene displaying excerpts of both gene body and promoter including the relative positions of pyro-sequenced CpG sites (*black boxes* and *circles*). *Grey boxes* and *white circles* show non-pyro-sequenced CpGs. *GREM2* consists of seven exons and 23 predicted CpG islands within 122,577 bases from transcription start site (*+1*). CpG1, cg.id=cg01809217; CpG15, cg.id=cg02577267. (**B**) Representative "pyrogram" of the *GREM2* Forward-assay displaying the relative luminescence in relation to the dispersed sequence to analyze. *Blue* highlighted peaks show the expected ratio. CpG sites are highlighted in *blue* while highlighted bases in *orange* reflect bisulfite controls. Percentage values atop CpG sites display the average degree of methylation detected by sequencing (*blue*, very good confidence; *yellow*, medium confidence; *red*, poor confidence/needs to be evaluated). CpGs within "pyrograms" of not evaluable dispension peaks were excluded from analysis.

2.5 Array-based comparative genomic hybridization (aCGH)

2.5.1 DNA purification

Extracted gDNAs were purified from salt/wash buffers using the Genomic DNA Clean & Concentrator Kit to obtain >500 ng of DNAs (DNA-recovery rate 70-95%) eligible for downstream analysis. Consequently, 2 µg of extracted DNA (1x volume) were mixed briefly in a 1.5 ml tube with DNA binding buffer (7x volume). The mixture was applied to the spin column placed in collection tubes. Samples were spun (13,000 g for 0.5 min at RT), the flow-through was discarded, and the column was repeatedly washed with 200 µl of DNA wash buffer prior to centrifugation (13,000 g for 0.5 min at RT). The column was transferred to a new collection tube. Afterward, 21 µl of ddH$_2$O was applied to the column matrix to eluate the bound DNA. The column tube was incubated (3 min at RT) prior to centrifugation of the column (13,000 g for 1 min). Ultra-pure DNA was spectrophotometrically quantified by NanoDrop ND-1000.

2.5.2 DNA labeling

Adenoma samples were prepared for hybridization onto the SurePrint G3 Human Genome CGH microarray 8x60K (comprising probes covering 55,077 distinct biological features, 1,000 replicated biological features and 3,886 internal QC) with a median overall probe spacing of 41 kb. Following the manufacturer's protocol (Agilent user guide, 2015), previously purified gDNAs were prepared for hybridization onto the microarray by non-enzymatically labeling both the reference DNA and the sample gDNA with the fluorescent dyes Cy3 (excitation peak=550 nm) and Cy5 (excitation peak=650 nm), respectively. 500 ng of purified sample gDNA and corresponding human reference DNA (female/male) were added to sterile PCR tubes, respectively. Volumes were adjusted to 8 µl per sample by either adding ddH$_2$O or vacuum concentration of the DNA. Reference DNAs only were then incubated at 95°C in a thermal cycler with heated lids for 10 min to heat fragment the DNAs. Subsequently, samples were chilled on ice for 3 min. Tubes were then spun in a microfuge (6,000 g for 0.5 min). Labeling reactions were started by adding 2 µl of the respective labeling master mix (Cy3-dUTPs for reference DNA, Cy5-dUTPs for gDNA) to each tube (Table 15). A thermal cycler incubated the samples at 85°C for 30 min prior to cool down on the ice (3 min). Non-reacted ULS-cyanines were removed with KREApure purification columns to avoid interference (*e.g.*, background noise) when hybridizing the DNAs onto the microarray. KREApure columns were mixed, bottom closures snapped off, and

columns were placed in collection tubes before centrifugation was performed (16,000 g for 1 min). While the flow-through was discarded, 300 µl of ddH$_2$O was applied to the column matrix. Again, the vial was spun (16,000 g for 1 min) prior to the discarding of the collection tube. Columns were placed into sterile collection tubes, and 10 µl of the ULS-labeled DNAs were added onto the matrix. A centrifugation step followed (16,000 g for 1 min). Here, the flow-through contained the purified ULS-DNA. By spectrophotometrically quantifying 1.5 µl of purified ULS-DNA per sample using NanoDrop ND-1000 ("MicroArray Measurement"/sample type: DNA-50) *degree of labeling*, *specific activity* and *yield* were determined. Matching ULS-Cy3- and ULS-Cy5-labeled DNAs were combined equaling 17 µl per pair of gDNA and reference DNA. Volumes were vacuum concentrated to 9 µl per couple utilizing a SpeedVac centrifuge.

Table 15. aCGH labeling master mix.

Reagent	Per reaction	Per 9x array
ddH$_2$O	0.5 µl	4.5 µl
ULS Cy3 or ULS Cy5	0.5 µl	4.5 µl
10x Labeling solution	1 µl	9 µl
Final volume	2 µl	18 µl

$$Degree\ of\ labeling = \frac{340 \text{ x pmol per µl dye}}{\text{ng per µl gDNA x 1000}} \text{ x } 100\% \qquad (4)$$

$$Specific\ activity = \frac{\text{pmol per µl dye}}{\text{ng per µl gDNA}} \qquad (5)$$

$$Yield\ (\text{µg}) = \frac{\text{DNA concentration (ng/µl) x sample volume (µl)}}{1000 \text{ ng/µl}} \qquad (6)$$

2.5.3 Microarray processing

Hybridization

As described in the manufacturer's protocol (Agilent user guide, 2015), ULS-labeled genomic and reference DNAs (9 µl) were thoroughly mixed with 25 µl hybridization buffer (Table 16). Samples were incubated in a thermal cycler (95°C for 3 min, then 37°C for 30 min). Tubes were spun in a microfuge (6,000 g for 1 min) prior to addition of 11 µl of Agilent-CGHblock (equilibrated to RT). The sample-hybridization mixture (40 µl) was dispensed onto the respective well on the gasket slide. The probe-labeled surface of the microarray slide was aligned onto the gasket slide without causing air bubbles. The "slide-sandwich" was sealed with the cover of the "SureHyb chamber" and set in the rotisserie of the hybridization oven (65°C for 40 hours at 20 rpm).

Table 16. aCGH hybridization master mix.

Reagent	Per hybridization	Per 9x array
Cot-1 DNA	2 µl	18 µl
100x aCGH blocking agent	0.5 µl	4.5 µl
2x Hybridization buffer	22.5 µl	202.5 µl
Final volume	25 µl	225 µl

Washing

When the hybridization was completed, the SureHyb chamber was disassembled in an ozone-free environment performing the following steps under an ozone-controlled enclosure due to cyanines being susceptible to ozone degradation. The "microarray-gasket sandwich" was removed from the SureHyb chamber and quickly submerged into the slide-staining dish #1 containing oligo aCGH/ChIP-on-chip wash buffer 1 (RT). After disassembly, the microarray-slide was washed in dish #2 containing oligo aCGH/ChIP-on-chip wash buffer 1 (RT) for 5 min with a magnetic stir. The second wash was performed by transferring the microarray into dish #3 containing oligo aCGH/ChIP-on-chip wash buffer 2 (warmed to 37°C) for 1 min. The slide was slowly removed from the dish to minimize droplets and loaded in the slide holder.

Scanning

The slide holder loaded with the microarray was placed in the SureScan microarray scanner. Scanning of the slide was performed using protocol AgilentG3_CGH with the following parameters: dye channel R+G, scan region 61x21.6 mm, scan resolution 3 µm and Tiff file dynamic range 16 bit. Agilent Feature Extraction software was applied for data extraction from the scanned microarray image. Agilent's Microarray QC Metrics (*e.g.,* background noise, signal intensity, and derivative log ratio spread) were assessed per sample. However, the quality of starting gDNA was drastically reduced for several samples due to FFPE tissue storage. Thus, eligibility for analysis was further evaluated by loading the processed sample data into Nexus Copy Number. The software was used for visualization by applying the rank segmentation algorithm on the CN profiles, which were manually reviewed according to the baseline noise. Overall, a rather conservative CNA threshold was applied >0.25. Data were filtered by excluding (*i*) CNAs comprising ≤1.5 Mbp, (*ii*) sex-determining chromosomes and (*iii*) aberrations in copy number variant (CNV)-overlapping regions. Microarray data was stored online at NCBI's Gene Expression Omnibus database (data accession no. GSE110221).

2.6 Multiplex-interphase fluorescence *in situ* hybridization (miFISH)

2.6.1 DNA probe production

The single-cell analysis of adenoma cells by miFISH was performed with self-made, non-commercially available fluorophore probes. The selection of locus-specific identifier probes was based on (*i*) published data on chromosomal aberrations in CRC (Bomme et al., 1994; Hermsen et al., 2002; Meijer et al., 1998; Ried et al., 1996) and on the (*ii*) aCGH landscape of copy number alterations identified in this study. Consequently, data assessment obtained the following genes: *COX2* (1q31.1), *PIK3CA* (3q26.32), *APC* (5q22.2), *CLIC1* (6p21.33), *EGFR* (7p11.2*), MYC* (8q24.21), *CCND1* (11q13.3), *CDX2* (13q12.2), *CDH1* (16q22.1), *TP53* (17p13.1), *HER2/ERBB2* (17q12), *SMAD7* (18q21.1), *SMAD4* (18q21.2), and *ZNF217* (20q13.2). Additionally, a commercially available probe (CytoTest) targeting the centromere of chromosome 10 (CEP10) was included for ploidy estimation since this chromosome is rarely altered in the colorectal tumorigenesis according to aCGH analyses (see above for references). The procedures of design and production of the probe sets are described below.

Bacterial artificial chromosome (BAC) clone contigs

BAC clones were picked from the University of California, Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu/) querying the genome assemblies dated to Feb.09/hg19 and Dec.13/hg39. Each probe contig comprised approximately 300-700 kb and was assembled by three to four overlapping BAC clones which overlapped and centered around the genes of interest (Supplemental Table 1). BAC clones were purchased from BACPAC Resources Center (http://bacpacresources.org). Probe contigs were assorted into three panels consisting of five genes each. The clone contigs were directly labeled with fluorophores (Dyomics) with an excitation peak of 415 nm (blue; DY-415-dUTP), 505 nm (green; DY-505-dUTP), 547 nm (gold; DY-547P1-dUTP), 590 nm (red; DY-590-dUTP), and 651 nm (far red; DY-651-dUTP), respectively. Panel 1 (Figure 17) consisted of *MYC* (blue), *CCND1* (green), *SMAD4* (orange), *CDX2* (red), and *PIK3CA* (far red), whereas panel 2 (Figure 18) was composed of *EGFR* (blue), *COX2* (green), *APC* (gold), *CLIC1* (red), and *SMAD7* (far red). Finally, panel 3 (Figure 19) combined the locus-specific probes for *ZNF217* (blue), *CDH1* (green), *TP53* (gold), *HER2* (red), and centromere probe CEP10 (far red).
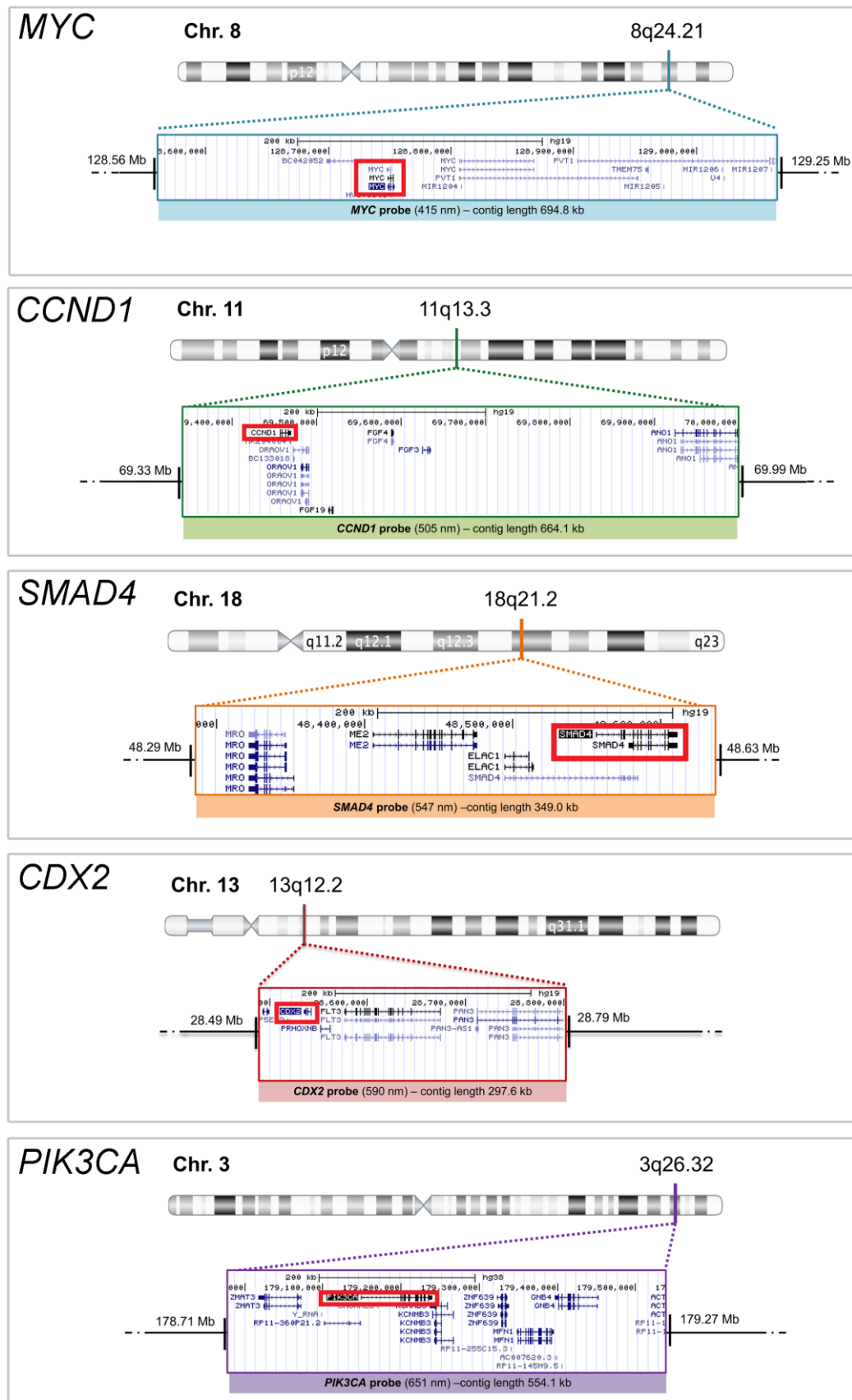
Figure 17. miFISH probe panel 1. Chromosomal location of each probe (*MYC*, *CCND1*, *SMAD4*, *CDX2*, *PIK3CA*) within panel 1 sorted by fluorophore wavelengths ranging from blue (415 nm) to far red (651 nm). A detailed map for the respective probe is illustrated in the zoomed-in box (proximal to distal flanking BAC clone sequences; 12.5 kb per column). The target gene is highlighted by a bold red frame. The box underneath gives an approximation of the final FISH probe size. Gene annotation/prediction is shown by UCSC Genes. Chromosome sizes are schematic and not to scale.

51
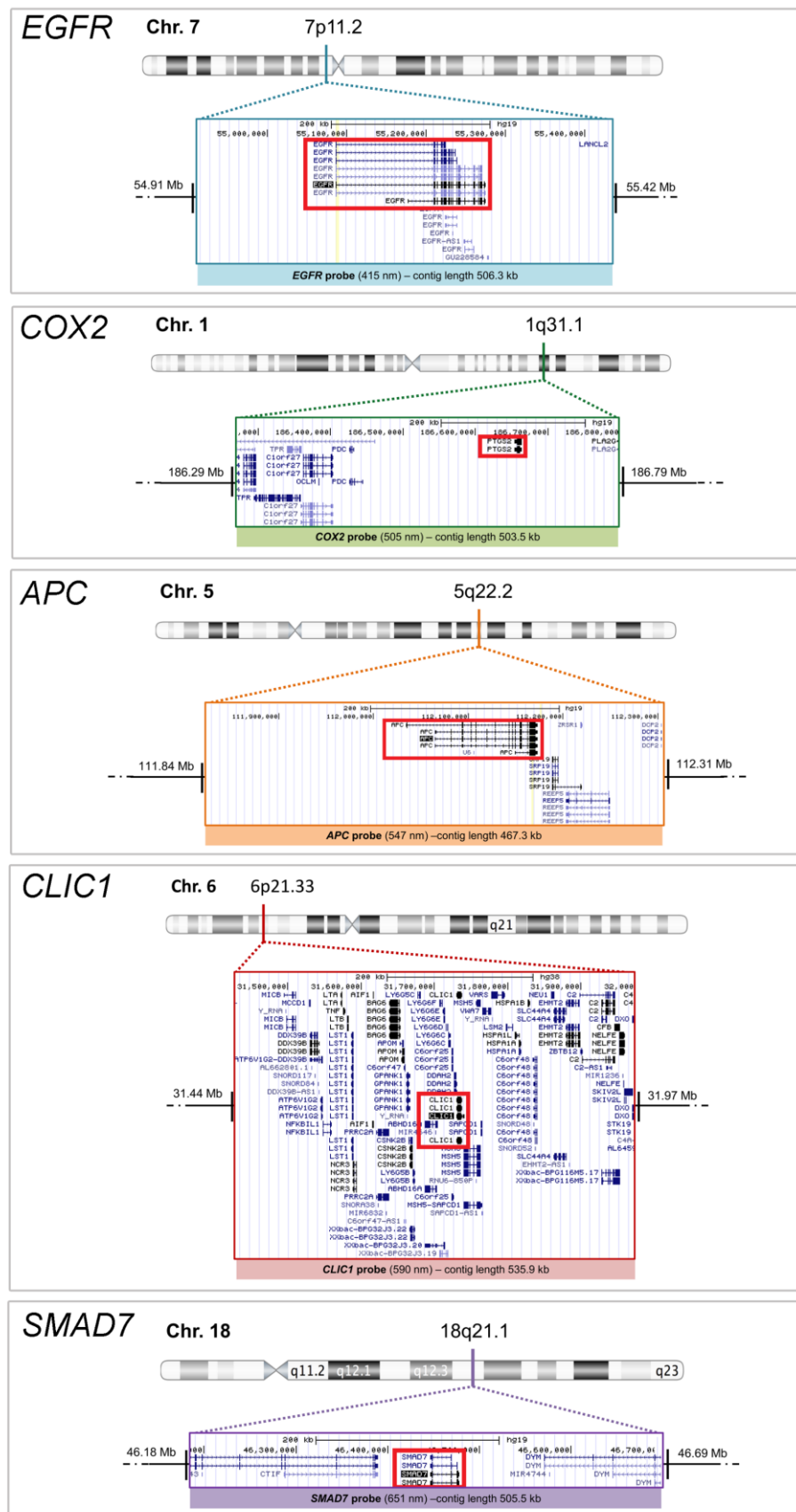
Figure 18. miFISH probe panel 2. Chromosomal location of each probe (*EGFR*, *COX2*, *APC*, *CLIC1*, *SMAD7*) within panel 2 sorted by fluorophore wavelengths ranging from blue (415 nm) to far red (651 nm). A detailed map for the respective probe is illustrated in the zoomed-in box (proximal to distal flanking BAC clone sequences; 12.5 kb per column). The target gene is highlighted by a bold red frame. The box underneath gives an approximation of the final FISH probe size. Gene annotation/prediction is shown by UCSC Genes. Chromosome sizes are schematic and not to scale.

52
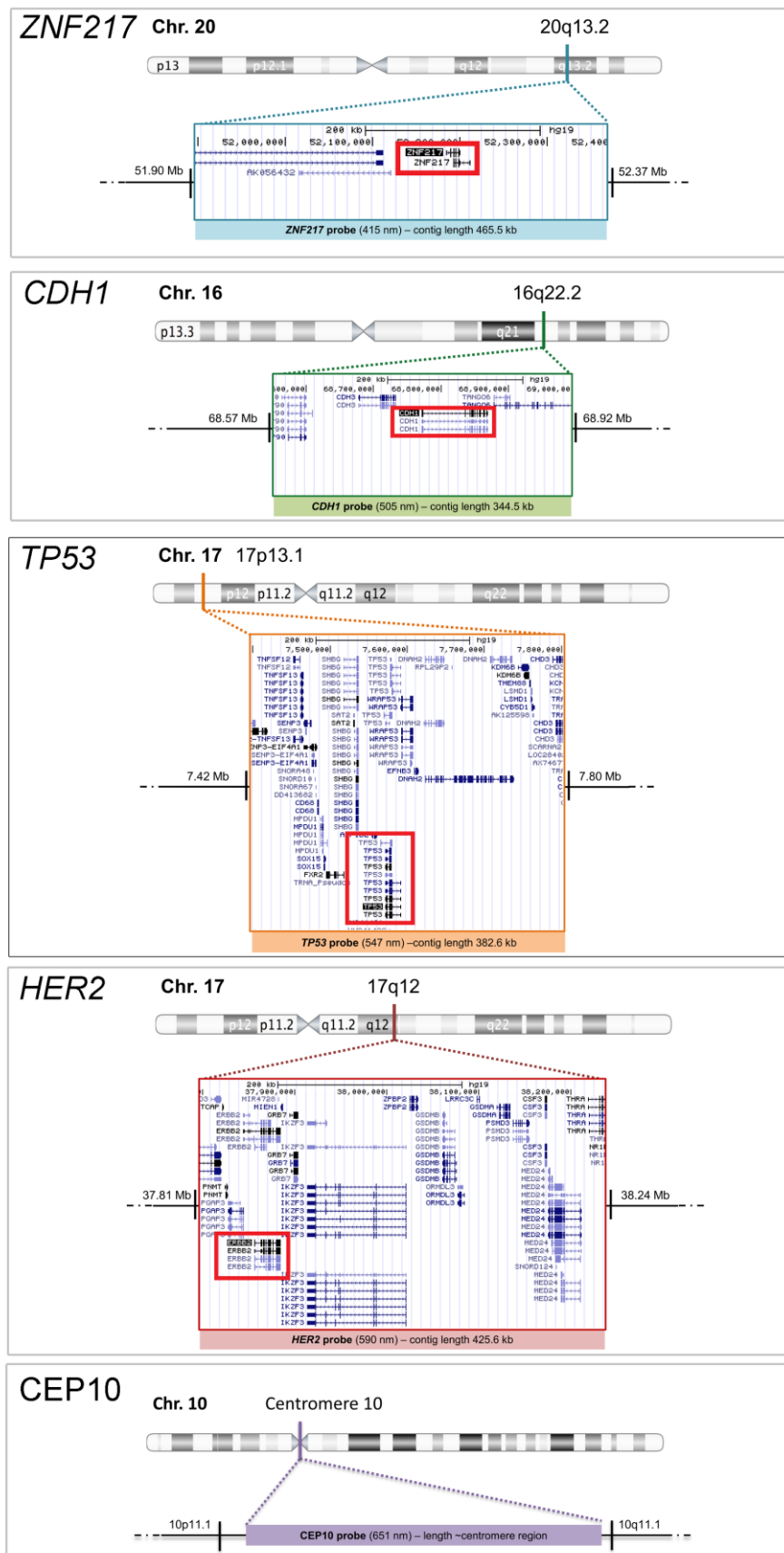
Figure 19. miFISH probe panel 3. Chromosomal location of each probe (*ZNF217*, *CDH1*, *TP53*, *HER2*, CEP10) within panel 3 sorted by fluorophore wavelengths ranging from blue (415 nm) to far red (651 nm). A detailed map for the respective probe is illustrated in the zoomed-in box (proximal to distal flanking BAC clone sequences; 12.5 kb per column). The target gene is highlighted by a bold red frame. The box underneath gives an approximation of the final FISH probe size. Gene annotation/prediction is shown by UCSC Genes. Chromosome sizes are schematic and not to scale.

53

BAC clone cultivation and DNA extraction

BAC clones were propagated in DH10B *E.coli* and shipped in LB agar (containing 12.5 µg/ml chloramphenicol) stab cultures at 4°C distributed by BACPAC Resources Center (Children's Hospital Oakland Research Institute, Oakland, CA, USA). Bacterial cultures were streaked to single colonies with a sterile loop on LB agar plates (containing 12.5 µg/ml chloramphenicol). Plates were incubated agar-side up at 37°C overnight. The day after a single colony was picked with a sterile loop to inoculate a starter culture in a 250-ml-flask containing 25 ml of LB medium and 25 µl chloramphenicol. The flask was incubated (37°C for 5 hours shaking at 300 rpm). Subsequently, 1 ml of the starter culture was taken off to inoculate an overnight culture in a 1 l flask with 250 ml LB medium and 250 µl chloramphenicol incubated overnight (37°C for 16-18 hours at 350 rpm). The overnight culture was pelleted (6,000 g for 15 min at 4°C) prior to pouring off the supernatant. The pellet was frozen at -80°C.

DNA extraction was performed using the Plasmid Maxi Kit starting by resuspending the bacterial pellet with 10 ml of chilled buffer P1. Subsequently, 10 ml buffer P2 was added, the tube was inverted 4-6 times and incubated at RT for 5 min. The mixture turned blue as an indicator of proper homogeneity. Buffer P3 was added (10 ml), the tube was repeatedly inverted and left on ice for 30 min when the mixture's color changed to white. The tube was centrifuged (20,000 g for 30 min at 4°C).

Meanwhile, the genomic-tip was equilibrated with 10 ml buffer QBT by gravity flow. The flow-through was discarded, and 30 ml of the supernatant were applied to the column. When the supernatant had entered the column matrix by gravity flow, the tip was washed twice with 30 ml of buffer QC. The flow-through was discarded, and the plastic tube was exchanged for a glass collection tube. DNA was eluted with 15 ml of the pre-warmed (65°C) buffer QF. DNA was precipitated by adding 10.5 ml isopropanol to the eluate. The tube was sealed with Parafilm M, inverted until entirely mixed and spun (15,000 g for 30 min at 4°C). The supernatant was aspirated, the residual pellet was washed with 5 ml cold 70% EtOH and subsequently centrifuged (15,000 g for 10 min at 4°C). When the supernatant was aspirated, the pellet was air-dried on a napkin for 5 min upside-down and right-side-up, respectively. The DNA pellet was re-dissolved in 200 µl dH_2O by pipetting up and down for 10 min. Dissolved DNA was transferred into a 1.5 ml microcentrifuge tube, whereas the glass tube was centrifuged (5,000 g for 3 min) before the repetition of the previous step. DNA yield was measured on a spectrophotometer and stored at -20°C prior to further use.

Nick translation

Probe DNA was simultaneously fragmented and labeled with fluorophores by nick translation. In this procedure, exonuclease activity of the added DNA polymerase induces single-strand breaks in the probe-DNA. Consequently, the DNA was nicked and a fluorophore-labeled dUTP was incorporated using the intact DNA strand as a template. Additionally, human Cot-1 DNA (containing repetitive sequences) was added to the reaction to reduce the hybridization of FISH probes onto repetitive sequences within the genome and therefore reducing background signals.

The nick translation reaction was performed in small-scale if the clone-DNA had to be tested, or in large-scale, if the clone-DNA was previously tested before. Consecutively, probe clone-DNA, NT master mix, and dUTP mixture were added into a tube (Table 17). The volume was adjusted by ddH$_2$O prior to addition of polymerase and 1:1,000 DNase I (1 mg/ml). The solution was mixed on a "Vortex", spun down and incubated in a thermocycler (2 hours at 15°C). DNA fragments were verified by running 5 µl per sample mixed with 1 µl loading dye on an agarose-gel electrophoresis (1% agarose in 50 ml TAE buffer with 3 µl GelRed) for 30-40 min with 100 V. DNA molecular weight marker II (1 µl mixed with 1 µl loading dye) and GeneRuler 100 bps DNA ladder (1 µl) were used as references. If the product displayed appropriate length of approximately 300-600 bps using an UV-transilluminator, nick translation would be terminated by adding 1-2 µl EDTA (0.5 M, pH=5.2). The tube was incubated (65°C for 10 minutes), and translated DNAs were frozen at -20°C.

Table 17. Nick translation input volumes per probe clone.

Reagent	Small scale	Large scale
DNA	2 µg	10 µg
NT master mix	30 µl	150 µl
dUTP	10 µl	50 µl
ddH$_2$O	X µl	X µl
Polymerase	2 µl	10 µl
DNase (1:1,000)	4 µl	14 µl
Final volume	100 µl	500 µl

10x NT buffer: 500 µl Tris-HCl (1 M), pH = 8.0

100 µl MgCl$_2$ (0.5 M)

350 µl ddH$_2$O

55

DNase I stock solution:	10 mg DNase
	1.5 ml NaCl (1 M)
	5 ml Glycerol
	ad 10 ml ddH$_2$O
dUTP mixture (0.5 mM):	1.25 µl dUTP (1 nmol/µl)
	3.75 µl dTTP (1 mM)
	5 µl ddH$_2$O
Master mix (single reaction):	10 µl NT buffer (10x)
	10 µl dATP-dCTP-dGTP mix (0.5 mM)
	10 µl β-Mercaptoethanol (0.1 M)

Precipitation

Fluorophore-labeled clone DNAs were pooled per gene probe and further high salt concentrated by DNA precipitation. Clone DNAs labeled by nick translation were inserted to the reaction in equal volumes per clones and probe. However, due to the varying intensity and detectability of probe fluorophores, the volume of probe clone DNA diverged for probes with an excitation peak at 415 nm and 651 nm, respectively (Table 18). In general, human Cot-1 DNA was added comprising 1/3 of the labeled-probe DNA and 3 µl salmon testes DNA. Moreover, sodium acetate (1/10 of the DNA mixture's volume) and 100% EtOH (threefold of the DNA mixture's volume) were added to the reaction tube which was subsequently incubated at -20°C overnight. The vial was centrifuged (14,000 rpm for 30 min at 4°C) prior to removal of the supernatant. The pellet was washed twice (2 ml cold 70% EtOH). After thorough mixing, the tube was spun (14,000 rpm for 15 min at 4°C). The supernatant was taken off carefully, and the tube was dried upside-down at 15°C. If residual liquid remained, a clean lens cleaning tissue was used to wipe the tube dry. Furthermore, the tube was incubated on a Thermomixer at 37°C for 1 min with an open lid. Probes labeled with green- or gold-fluorophores were then resuspended with equal volumes of formamide and master mix to obtain a final contig volume of 30 µl. Other probes were concentrated more rigorously with 6 µl and 20 µl, respectively (Table 18). An appropriate volume of pre-warmed deionized formamide (3-15 µl) was added for resuspension. The tube was incubated on a Thermomixer (700 rpm for 30 min at 37°C). Resuspension was completed by adding the appropriate volume of precipitation master mix (3-15 µl). The tube was shaken on a Thermomixer (700 rpm for 30 min at 37°C). Hybridizing the probe on a lymphocyte test slide evaluated the probe quality.

Table 18. Large-scale DNA precipitation scheme. Note, if the number of BAC clones diverged from three the input volumes for the additional reagents would be adjusted.

Fluorophore excitation	415 nm	505 nm	547 nm	590 nm	651 nm
Number of clones	3	3	3	3	3
Labeled-probe DNA per clone	400 µl	300 µl	300 µl	300 µl	420 µl
Sum	1,200 µl	900 µl	900 µl	900 µl	1,260 µl
Human Cot-1 DNA	400 µl	300 µl	300 µl	300 µl	420 µl
Salmon testes DNA	3 µl	3 µl	3 µl	3 µl	3 µl
Na-Acetate	160.3 µl	120.3 µl	120.3 µl	120.3 µl	168.3 µl
100% EtOH	4,408.3 µl	3,308.3 µl	3,308.3 µl	3,308.3 µl	4,628.3 µl
Sum volume	6,171.6 µl	4,631.6 µl	4,631.6 µl	4,631.6 µl	6,479.6 µl
Final contig volume	20 µl	30 µl	20 µl	30 µl	6 µl

Master mix:

40 ml Dextran sulfate (50%)

20 ml SSC (20x)

ad 100 ml ddH$_2$O

pH adjusted to 7.0 by HCl (1 M)

2.6.2 Cytospin preparation

The procedure was a modification of the previously published *Hedley* technique (Heselmeyer-Haddad et al., 2012; Heselmeyer-Haddad et al., 2014): FFPE tissue blocks were cut into three consecutive sections. The first section (4 µm-thick) was stained by H&E to mark the region of interest by an experienced board-certified pathologist. The next two sections (each 50 µm-thick) were taken for conscientious macro-dissection with a scalpel guided by the formerly H&E-stained slide. Scraped tissue was collected in a 1.5 ml tube. Deparaffinization began by adding 1 ml xylene (20 min at RT). Xylene was removed by pipetting off the supernatant after spinning the tube (3,000 rpm for 3 min). These steps were repeated three times. The tissue was dehydrated by subsequently adding 1 ml of 100% (twice), 90% and 70% EtOH, respectively. The tube was respectively incubated for 5 min at RT prior to centrifugation (3,000 rpm for 3 min). The supernatant was removed. Finally, 1 ml 50% EtOH was added and the tube incubated at RT for 5 min. The tissue was spun (14,000 rpm at 4°C for 15 min) before EtOH was taken off and 1 ml dH$_2$O was added (20 min at RT). The tube was spun (4,000 rpm at 4°C for 15 min) prior to removal of the dH$_2$O. Macro-dissected tissue was then digested with 500 µl 0.1% protease in 1x PBS (800 rpm at

45°C for 60-120 min; case-specific). The degree of disintegration was periodically (every 15 min) monitored by taking off 5 µl of the tissue solution and mixing it with 5 µl of staining solution on a microscope slide. The tissue sample was evaluated under a fluorescence microscope using the 16x objective field view. Optimal criteria for disintegration included approximately 30 nuclei per field and intact nuclei exhibiting little to no cytoplasm while the DAPI stain was intense. In case of incomplete disintegration, the incubation of the tissue in the tube was continued. When the nuclei were of digested as desired, the reaction was stopped by adding 500 µl of 1x PBS before the vial was spun (2,500 rpm for 5min at RT). The supernatant was removed leaving 100 µl of residual volume. Disintegrated single cells were resuspended in a case-specific volume of sterile 1x PBS to obtain a concentration which allows the production of ten or more cytospin-slides. Meanwhile, cytology funnel filters were filled with 80 µl 1x PBS and loaded onto the cytospin centrifuge (1,400 rpm for 5 min) accompanied by a microscopy adhesion slide (SuperFrost™). Filters were then filled with 80-160 µl cell suspension and spun in the cytospin centrifuge (1,400 rpm for 5 min). The density of the cytospin-slides was evaluated with a phase contrast microscope, and the cell suspension was further diluted or concentrated as needed, respectively. Obtained single-layer cytospin-slides were dehydrated by subsequent exposure to 100 ml Coplin jars containing 70% EtOH (5 min), 90% EtOH (5 min) and 100% EtOH (10 min), respectively. Slides were air-dried and stored (4°C) until used further.

DAPI (1 mM):		1 mg DAPI in 27 ml dH$_2$O
Sulforhodamine (5 mM):		50 mg Sulforhodamine in 16.5 ml dH$_2$O
Staining solution:		2 ml DAPI solution
		2 ml Sulforhodamine solution
	Ad	98 ml PBS (1x), pH = 7.4

2.6.3 MiFISH

The following chapter will depict the steps of the procedure of multiplex-interphase fluorescence *in situ* hybridization (miFISH) including the slide's pretreatment, hybridization, detection, scanning, and re-hybridization, respectively.

Pretreatment

Prior to hybridize the locus-identifier probes onto the previously prepared FFPE adenoma cell single-layer cytospin slides the nuclei had to be pretreated to obtain analyzable results of the hybridization procedure. Hereinafter, the pretreatment procedure is described for cytospin case slides first, while the modified pretreatment of lymphocyte slides (used to test probe quality) are described in the second part of this chapter.

Desired slides were examined under a phase contrast microscope to identify the appearance and shape of the nuclei. Additionally, the area of interest comprising the nuclei was marked by a glass pen. Slides were equilibrated in a Coplin jar with 2x SSC for 10 min at RT and transferred into 1x PBS for 2 min. Subsequently, slides were placed in a hybridization chamber on a plastic grid covering paper tissues soaked in dH$_2$O. The proteolysis of peptide bonds in membrane proteins of the adenoma cells was performed by applying 120 µl proteinase solution (0.1%) on top of the cytospin slide. The hybridization chamber was incubated (37°C for 15 min). Slides were transferred into a Coplin jar containing 1x PBS to stop the digest temporarily and to check the nuclei under the phase contrast microscope. If the cells exhibited an excess of cytoplasm which was determined by a grey haze surrounding the nuclei, the slide was pre-incubated with 120 µl proteinase solution (0.1%; 100 mg/ml dissolved in ddH$_2$O) for a more extended period. The specific pretreatment intervals ranged from 15-120 min in regard of the case and slide age.

In contrast, if the cells showed a significant fading of the grey haze by reduction of cytoplasm, the reaction would be stopped. Slides were washed three times in Coplin jars containing 1x PBS. Finally, the slides were dehydrated in an increasing EtOH series comprising of 70%, 90% and 100% for 5 min, respectively. Slides were air-dried prior to performing the hybridization procedure.

Nevertheless, the pretreatment protocol for lymphocyte slides was slightly modified to ensure somewhat moderate proteolysis to avoid over-digest. Lymphocyte slides were equilibrated in 2x SSC for 20 min while the proteolysis was performed by adding 5 µl pepsin solution (0.1%; 100 mg/ml dissolved in ddH$_2$O) in a pre-warmed (37°C) Coplin jar which was filled up with 100 ml HCl (0.01 M; pH=2.0). The slide was incubated for 30-120 sec (variance per batch). The additional steps were performed as described beforehand.

Hybridization

Previously prepared locus-identifier probes were assorted into three panels of five probes (see 2.6.1 DNA probe production). Panels were set up in 1.5-ml-tubes in 10 µl batch volume comprising a 415 nm-probe (2 µl), a 505 nm-probe (2 µl), a 547 nm-probe (2 µl), a 590 nm-probe (2 µl), a 651 nm-probe (1 µl) and, additionally, forma-mide/precipitation master mix solution (1 µl). The desired probe was prepared for hy-bridization by denaturing the probe DNA (73°C for 5 min). Subsequently, the probe was pre-annealed by shaking on a Thermomixer (37°C for 1 hour at 350 rpm) covered with aluminum foil for light-protection. Noteworthy, if probe panel 3 was prepared for hybridization, centromere probe CEP10 was solely denatured and added to the probe mix when pre-annealed.

Meanwhile, pretreated and air-dried single-cell adenoma suspension cytospin-slides were denatured. Therefore, 120 µl of 70% formamide/2x SSC were applied on a 24x60 mm^2 coverslip. The cytospin-slide was touched to the coverslip strictly avoid-ing the formation of air bubbles. The slide was incubated at 72°C on a ThermoBrite StatSpin hot plate for 3 min. As soon as this procedure was performed, the coverslip was slipped off gently. The slide was quickly transferred into an ice-cold increasing ethanol series comprising of 70%, 90% and 100% for 3 min, respectively. Slides were air-dried under the laboratory fume hood. The pre-annealed probe (2 µl) was applied to the slide's nuclei and covered by a round coverslip strictly avoiding the formation of bubbles. Drops of rubber cement sealed the edges after drying for approximately 3 min. Cytospin-slides were placed in the humid, light-impenetrable hybridization chamber and incubated at 37°C for 16 hours.

Contrary to the adenoma FFPE cytospin slides, lymphocyte slides were denatured at 65°C for 30-60 sec, which was batch-dependent.

70% formamide/2x SSC:	10 ml SSC (20x)
	20 ml ddH$_2$O
	ad 100 ml Formamide, deionized; pH = 7.0

Detection

Overnight-hybridized cytospin-slides were taken off the hybridization chamber to per-form a stringent wash. Rubber cement was removed gently utilizing a tweezer. The slides were immersed into a Coplin jar of pre-warmed (48°C) 2x SSC/0.3% NP-40 buffer. With the aid of a tweezer, the coverslips were slid carefully into the solution.

The slides were washed by gentle and periodical shakes while incubated in another Coplin jar with pre-warmed (48°C) 2x SSC/0.3% NP-40 buffer for 2 min. Subsequently, slides were transferred into 2x SSC/0.1% NP-40 buffer for 1 min at RT prior to a wash in 2x SSC for 0.5 min and 1x PBS for 2 min, respectively. The Coplin jars were kept away from light. Slides were then exposed to the DAPI solution for 1 min at RT. Washing the slides three times with 2x SSC for 2 min removed excessive residues with additional agitation. An increasing EtOH series comprising of 70%, 90% and 100% for 2 min, respectively, dehydrated the slides at RT. Stored in a drawer slides were dried while hidden from light-exposure. Finally, 1-2 drops Vectashield fluorescence mounting medium without DAPI were applied on a coverslip which was subsequently touched onto the cytospin-slide under strict avoidance of bubble formation. The slides were ready for imaging and scanning.

Nevertheless, lymphocyte slides were washed less stringently. The utilization of the detergent NP-40 was eluded. Instead, the slides were washed three times in 2x SSC.

DAPI solution:	50 µl of DAPI stock (5 mg/ml)
	50 ml PBS (1x)
	mix and store light-protected at 4°C

Scanning

The detected cytospin-slides were subsequently placed onto the DUET scanning imaging workstation (BioView Ltd.). This workstation comprised of a fluorescence microscope accompanied by an automated stage with a loading capacity of up to 8 slides (Olympus). Ready-to-image slides were prepared for scanning by adjusting the scanning parameters using an exemplary field of view with several nuclei of the respective slide. Therefore, the plane of focus was adjusted by exciting the slide with the excitation peak of DAPI (λ=358 nm; ultra-violet) by the mercury fluorescence lamp integrated into the microscope. Consecutively, all probes were tested by reviewing the chosen field of view. Probe excitation and image acquisition were adjusted by brightness, exposure time, intensity and plane of focus to obtain precise and intense probe signals with a reduced background, respectively. The modified parameters were adapted for the respective channels and applied for the entire scan. A 2448x2048 monochrome digital camera took high-quality images. Recorded probe signals were automatically re-colored by the software according to the probe's respective excitation peak (*i.e.*, blue=415 nm; green=505 nm; gold=547 nm; red=590 nm; far red=651 nm). The

software located an upper limit of 12,000 targets per cytospin-slide. If the scanned slide were too sparse, the whole slide would be scanned.

If a slide were re-scanned after re-hybridization with panel 2 or 3, respectively, the additional scan would start via re-localizing the previously scanned targets by the DUET software. Nevertheless, the process had to be monitored and reviewed manually to assure exact overlap. If the target nuclei were successfully retrieved, scanning of the probe panel would be conducted by revisiting all targets formerly scanned by the fluorescence microscope. Prior to the scan parameters were adjusted as described before. Images of the current signals were taken before sorting these images to the corresponding image of each target nucleus. Thus, signal counts for all 15 miFISH probes could be collected for each targeted nucleus assorted into three panels.

Re-Hybridization

After scanning and imaging a previously hybridized slide, the probe could be washed off to perform a novel miFISH on the same slide (up to four times). The desired probe panel was set up and prepared for hybridization by denaturing the probe DNA at 73°C for 5 min. Subsequently, the probe was pre-annealed by shaking in a Thermomixer (37°C for 1 hour at 350 rpm) covered with aluminum foil.

Meanwhile, the slide was immersed into 2x SSC, and the coverslip was slid off gently (with a tweezer). Slides were washed three times (2x SSC for 2 min). By combining washing and denaturation, the slide was exposed to pre-warmed (80°C) 50% formamide/2xSSC for 1 min. Slides were then dehydrated in an ice-cold increasing EtOH series comprising of 70%, 90% and 100% for 2 min, respectively. Slides were air-dried in the laboratory fume hood prior to application of 2 µl of the pre-annealed probe panel. As described before (see 2.6.3 miFISH: Hybridization), slides were covered by a round coverslip strictly avoiding the formation of bubbles. Drops of rubber cement sealed the edges after drying for approximately 3 min. Cytospin-slides were placed in the humid, light-impenetrable hybridization chamber and incubated (37°C for 16 hours) until slide detection was performed.

50% formamide/2x SSC:	10 ml SSC (20x)
	40 ml ddH$_2$O
	ad 100 ml Formamide, deionized; pH = 7.0

2.6.4   Data assessment

Signal enumeration

The recorded images of the probe signals within the three probe panels were analyzed when the scanning of all probes was successfully performed. The nuclei were sorted by target number with each target comprising three panels. Each panel showed a merged image of the nucleus which combined all five probe channels within a scanned panel. Moreover, the panel presented a DAPI image and a single image for each of the five probes. The signal counts were manually counted for accuracy using the SOLO satellite workstation (BioView) with the customized software. Counting was performed consecutively reviewing the scanned targets starting with target number one. When reviewing the targets, a nucleus was excluded from analysis if a criterion for exclusion applied (Table 19). Split probe signals were cautiously enumerated concerning the underlying ploidy. Targets were assorted into the following categories:

(i)   nuclei with an unchanged signal count for each probe (*i.e.*, non-aneuploid)

(ii)   nuclei with any changed signal count (*i.e.*, aneuploidy)

(iii)   nuclei with signal counts suggesting a triploid/tetraploid set of chromosomes

(iv)   nuclei excluded from analysis.

Counting proceeded until 350 altered nuclei were identified (cut-off). If this were not applicable, the whole slide would be enumerated. Consequently, counting was stopped when the last target was reviewed no matter how many altered cells were identified. On average, 10,398 targets were scanned per adenoma cytospin-slide (limit of 12,000 targets). Signal evaluation within all specimens revealed a mean final count of 180 aberrant (range, 5-350) and 342 non-aberrant interphase nuclei (range, 18-867), respectively. Enumerated signal counts were recorded in an excel-spreadsheet.

Table 19. Criteria of an exclusion of nuclei from the miFISH analysis.

criteria	description
i	the nucleus overlapped with other nuclei
ii	the nucleus was of non-epithelial morphology (*i.e.*, non-ellipsoid nucleus) suggesting an immune cell
iii	the nucleus was visibly damaged indicating a truncated or cut nucleus
iv	the nucleus exhibited indistinguishable probe signals indicating an artifact or failed panel hybridization
v	the nucleus lacked a centromere signal which was likely due to a truncated artifact

Processing of data files

Signal counts were recorded in raw data tab-delimited text files saved separately per case. The raw data file contained each scanned target which was further assorted by the assigned categories of nuclei while counting. Analyzed nuclei possessed a signal count denoted separately for each probe. Based on these raw data files, analyzed cells were re-sorted by an excel macro application. Rows were lexicographically ordered ranging from low signal numbers towards increased signal numbers. Columns were sorted by the probe's chromosome location in ascending order. The combination of probe signal counts is called the signal pattern. The macro application grouped multiple nuclei which exposed the same signal pattern into one row and annotated the count of nuclei in the respective column. Nuclei with a diploid signal count within each analyzed probe (non-altered) were also listed in the table. Data files were used for the further analysis described at a later point.

Ploidy annotation and determination of gain and loss

An overall ploidy was assigned to each analyzed nucleus to determine copy number gains and losses further. Cells were classified as either diploid (2N), triploid (3N) or tetraploid (4N). Higher polyploidies (>4N) were not observed in our cohort. Centromere 10 (CEP10) was chosen as an indicator of DNA ploidy since aCGH analyses have revealed that copy number alterations exceptionally rarely targeted this chromosome in the colorectal tumorigenesis. The annotation of ploidy per signal pattern was inferred by assessing the signal count of CEP10. However, the landscape of additional gene probe counts was also considered for ploidy estimation using the empirical knowledge of which probes represented oncogenes and tumor suppressors, respectively.

Moreover, an average ploidy was calculated for each adenoma case by dividing the sum of ploidy values per signal pattern from the sum of cells analyzed. Among adenomas analyzed, 94.8% (55/58) were inferred to possess a diploid (2N) average ploidy whereas 5.1% (3/58) of cases showed conclusive evidence for the presence of polyploid sets (4N) of chromosomes (case 13, 30 and P7b). Albeit, polyploid samples were excluded from further statistical analysis due to their severity of genomic instability compared to diploid lesions (2N).

CN gain and loss, respectively, were annotated for each gene probe regarding the underlying ploidy of the respective signal pattern (Figure 20). Consistently, copy numbers exceeding two signals would not, *per se*, indicate a gain. Neither would the observation of two signals inevitably imply an unchanged signal number if the ploidy of the signal pattern was not considered. CN gains and losses were not only separately evaluated per nucleus but also in total per adenomatous lesion. As such, copy number alterations were considered if the respective alterations were present in at least 10% (threshold) of the counted cell population per adenoma lesion.

Figure 20. Exemplarily translation of probe signal counts into gain and loss signal pattern displays. The flow chart exemplarily depicts the first 15 analyzed nuclei of adenoma case A1. First, signal probe counts were reviewed and stored in an excel spreadsheet for each nucleus. Second, probe counts were translated into copy number status neutral (N), gain (G) or loss (L) in accordance to the annotated and inferred ploidy, respectively. Finally, copy number information were color-coded to display hundreds of single cells within a graph quickly assessing clonal populations. Columns represent all probe signal counts per individual nucleus while rows display the counts for a certain probe within the nuclei fraction. Copy number color-code: neutral (*blue*), gain (*green*), loss (*red*).

Tumor heterogeneity by (ecological) measures of diversity

With the scope of a quantitative assessment of inter- and intratumor heterogeneity within the adenomas the following four measures of diversity were calculated:

- *(i)* Instability index (Heselmeyer-Haddad et al., 2012 and 2014)
- *(ii)* Shannon entropy (Martinez et al., 2016; Wangsa et al., 2016),
- *(iii)* Simpson diversity index (Martinez et al., 2016; Wangsa et al., 2016)
- *(iv)* Accumulated pairwise genetic diversity (Rogers, 2015).

First, the frequencies of signal patterns (*i.e.*, "genetically divergent clones") were determined. To calculate the diversity measures, let p_i (*i.e.*, allele frequency) be the frequency of the ith count pattern of kth (*i.e.*, allele number) different loci expressed by probe-specific copy numbers. The total pattern count was defined as N of n nuclei. Additionally, the number of altered count patterns was expressed as X_K while the number of diploid (non-altered) count patterns was given as Y_K.

- *(i)* The ecological diversity measure of species richness was transferred into clonal richness, here defined as "instability index" *I*. The index comprises the number of observed patterns per sample divided by the total number of enumerated clones within a sample. The value is multiplied by 100 to obtain a range of 0-100 patterns.

$$genomic\ instability\ index\ I = \frac{(N*100)}{n} \qquad (7)$$

- *(ii)* Shannon entropy *H'* is a measure of richness and evenness in information theory which puts more weight on the sheer number of genetically different clones. The dimensionless index usually gives values in the range of 1-3 (low to high heterogeneity), although the scale is unlimited.

$$Shannon\ entropy\ H' = -\sum p_i log_2(p_i) \qquad (8)$$

- *(iii)* Simpson diversity index *D'* is a measure of dominance in population genetics which puts more weight on more frequently observed clones. The dimensionless index is in the range of 0-1 (low to high heterogeneity).

$$Simpson\ diversity\ index\ D' = 1 - \sum p_i^2 \qquad (9)$$

(iv)   Genetic diversity metrics can assess genetic distance, here defined as accumulated pairwise genetic diversity *D*. The dimensionless index estimates the genetic distance among clones within a tumor sample typically in a range of 1-6 (low to high clonal distance). Each copy number alteration increases the genetic distance away from the diploid signal pattern - no matter whether caused by copy number gain or loss, respectively.

$$Accumulated\ pairwise\ genetic\ diversity\ D = \frac{2}{N(N-1)} \sum_{k=1}^{15}(x_K * y_K) \quad (10)$$

Phylogenetic miFISH trees

The clonal evolution of subclones identified by single-cell miFISH can be modeled by tumor phylogenies using the algorithm of the publicly available FISHtree software (Chowdhury et al., 2013; Chowdhury et al., 2014). The updated version of FISHtree 3.0 inferred a model with the possible clonal evolution not only within a single tumor but also from the primary to the secondary tumor by building a consensus tree (Gertz et al., 2016). In the study presented herein, the clonal evolution of subclones within paired primary adenoma and recurrent adenoma was modeled to obtain a consensus graph. However, each paired adenoma lesion (*i.e.*, primary and recurrent adenoma) was also modeled separately to obtain the probabilities of copy number changes for the respective edge. The heuristic algorithm considers the change of probe counts which represent different copy number alterations within the clonal population of the adenoma lesions. Thus, the algorithm was designed to infer the most likely relation and distance of copy number changes across the entire lesion based on the total number of copy number alterations observed. The approach used in this study modeled the distance of copy number alterations irrespective of the ploidy. Excel data files separated per lesion contained the probe signal counts for each analyzed nucleus. These nuclei were incorporated by the FISHtree algorithm which obtained a *dot*-file per adenoma lesion. The respective graph description file was exported to a vector graph file by *DotEditor* and visualized by *Inkscape*. Tree models were then manually translated into *PowerPoint* graphs precisely adopting the clone distances and tree levels.

Created tree models display the diploid (all probe counts equal 2) copy number-neutral signal pattern ("root") at the top of the graph (Figure 21). Edges leading away from the root signal pattern represent the minimum of copy number alterations required to form the copy number-wise distantly closest clone populations, whose tree-nodes are defined as children of the root. Edges within the tree correspond to the

accumulation of an additional copy number alteration connecting the nodes of the progenitor and the descendant, respectively. Nodes within the tree represent a particular signal pattern based on the observed CNA. While trees derived from a single lesion also displayed the probability value for the connection between two copy number changes annotated at the respective edge, consensus trees did not provide this information as a matter of lucidity. With the aim to simplify the consensus tree, clone fractions exclusive to the primary adenoma or the recurrent adenoma were marked by a circularly or rectangularly shaped node, respectively. If both lesions shared a population, the node would be shown as a diamond. Additionally, major clone populations of both the primary and the recurrent adenoma were color-coded. Moreover, if the most significant clone population within the primary tumor corroborated by the most substantial fraction of cells within the recurrent tumor, the node would be highlighted. A dotted node marked signal patterns which were observed once within an individual adenoma. A bold node showed signal patterns counted twice (or more).

Statistical analysis of the tree-based model provided the opportunity to assess the degree of tumor heterogeneity and the amassment (*i.e.*, complexity) of copy number changes within the clones. Consequently, the number of tree levels away from the neutral root node defined the "node depth" which was previously identified as a potential measure for the prediction of tumor progression (Heselmeyer-Haddad et al., 2014).
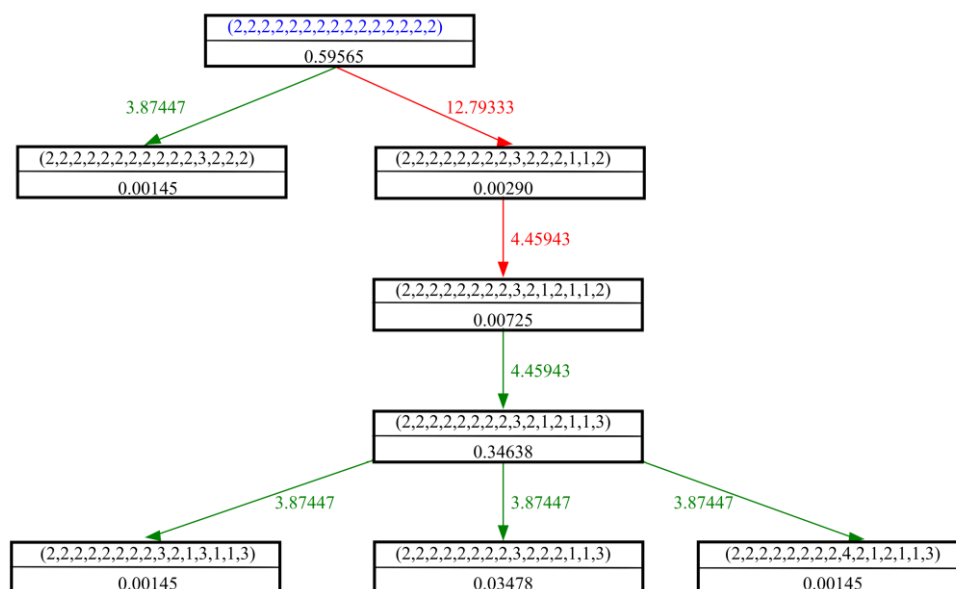
Figure 21. A representative example of a signal pattern miFISH tree. The phylogenetic tree shows the inferred clonal evolution of an adenoma. Top node ("root of the tree") represents the diploid, unchanged CN pattern (*blue*). The two children of the root are nodes with the distantly closest alteration pattern (*quasi,* "trunk of the tree") Every node shows a signal pattern present in the adenomatous lesion and each edge stands for a change in CNs away from the root cell. The probability value for a certain alteration is annotated to the edge. CN gains are shown in *green* while copy number losses are shown in *red*. Signal pattern frequencies are shown below the line separating each node-box. Probe order within the signal pattern: *COX2, PIK3CA, APC, CLIC1, EGFR, MYC,* CEP10*, CCND1, CDX2, CDH1, TP53, HER2, SMAD7, SMAD4, ZNF217.*

2.7 Immunohistochemistry

2.7.1 Microsatellite instability detection

Although adenomatous polyps are reported to be rarely affected by losses of MMR proteins (Gaiser et al., 2013; Petko et al., 2005), the microsatellite phenotypes of colorectal adenoma samples not displaying any CNAs in the previous analyses were assessed indirectly by the loss of MMR protein expression (Gaiser et al., 2013).

Therefore, four FFPE tissue sections (3 µm-thick) per case were mounted on SuperFrost microscope slides followed by deparaffinization in xylene (3x 5 min). Slides were rehydrated in a decreasing ethanol series comprising 100% (2x), 96% and 80% for 2 min, respectively, before rinsing in dH$_2$O (2x). Antigen retrieval was performed via heat-induction by immersing the slide in a Coplin jar of Target Retrieval Solution (pH 9.0; Tris/EDTA; 1:10) in a steamer (95°C for 20 min). Subsequently, sections were washed with TBS buffer. Peroxidase Blocking Solution was added for 7 min which was then removed by washing with TBS buffer. Primary antibodies MLH1 (1:25) and PMS2 (1:50) were diluted with Antibody Diluent, while primary antibodies MSH2 and MSH6 were ready-to-use and applied to the tissue slides for 30 min. Sections were washed with TBS buffer and detection was performed with the EnVision Detection System, Peroxidase/DAB, rabbit/mouse. The peroxidase-labeled secondary antibody was incubated for 30 min. Sections were rinsed, and chromogen-DAB was applied for 6 min prior to rinsing the slides. Slides were counterstained by immersion into hematoxylin solution for 2 min followed by rinsing in tap-water for 2 min. Dehydration of the sections was carried out in an increasing EtOH series comprising of 80% (2x), 96% and 100% for 2 min, respectively, and subsequent incubation in xylene for 2 min. A coverslip and Pertex mounting medium covered the sections.

Experienced pathologists evaluated the immunohistochemical staining. Evaluation criteria followed Gaiser et al. (2013): positive nuclear staining in basal crypts of normal epithelium and inflammatory cells was considered for expression evaluation of the mismatch repair genes MLH1, MSH2, MSH6, and PMS2. Distinct expression of each marker in tumor cells was classified as microsatellite stability, whereas a loss of MLH1, MSH2, MSH6 or PMS2 indicated microsatellite instability for the respective adenoma. Staining was validated in comparison to positive and negative controls, respectively. Slides were imaged by a digital microscope and an accompanied scanner.

2.7.2 CDX2 expression

Immunohistochemistry on tumor samples was performed with a primary antibody (mouse) targeting CDX2. In general, the procedure was carried out as described before (see 2.7.1 Microsatellite instability detection) with slight modifications using the CDX2 primary antibody (dilution 1:50) and performing an antigen retrieval with EDTA (pH=9). An experienced pathologist evaluated immunohistochemical staining. Positive nuclear staining in normal epithelium and nuclear/cytoplasmic staining in tumor cells were evaluated as controls. Intensities of staining were evaluated by the pathologist as an indication of putative relatively high or low expression of *CDX2* per adapted immunoreactive score (Remmele and Stegner, 1987).

2.8 Statistical analysis

Statistical identification and interpretation of differentially methylated CpG positions (DMPs) in adenoma specimens analyzed by HM450K were described in detail before (see 2.4.3 DNA methylation microarray: Human Methylation 450K Array).

In general, statistical analyses were performed using the statistical software *GraphPad Prism*, and *R*. Continuous data were tested by Shapiro-Wilk-normality test to assess whether values were Gaussian distributed. Unpaired, parametric variables were then analyzed by Student's *t* test comparing the columns by the mean and the standard deviation (SD), while non-parametric data were analyzed by Mann-Whitney *U* test comparing the columns by medians with interquartile ranges (IQR). Comparisons of paired variables were performed by Wilcoxon matched-pairs signed rank test. If three or more columns were compared, significance was tested by either one-way ANOVA (parametric) or Kruskal-Wallis (non-parametric) for non-grouped data, whereas two-way ANOVA compared grouped variables. Categorical variables were tested in 2x2-contingency tables by Fisher's exact test while more variables were tested with the Freeman-Halton test. Recurrence-free time was statistically assessed using Kaplan-Meier analysis calculating the significance via log-rank (Mantel-Cox) test and hazard ratio (HR) with 95% confidence intervals (CI). Cohen's kappa (with 95% CI) quantitatively determined inter-observer agreement of CN status detected by aCGH and miFISH. Clinicopathological features were correlated by Spearman's non-parametric correlation coefficient. Fractions of a whole were displayed by frequencies.

Statistical significance was defined as $P<0.05$ and displayed throughout the thesis with three significant digits. Performed tests were carried out exploratory and two-sided without multiple test correction (if not mentioned otherwise: *e.g.*, HM450K data) demanding careful interpretation of the data.

The profile of copy number alterations (CNAs) detected by aCGH was tested for fitting Gaussian distribution or rather Poisson distribution when $P>0.10$. To compare CNAs among different sample groups the average number of copy alterations (ANCA) was calculated (Ried et al., 1999).

Chromosomal instability and tumor heterogeneity were quantitatively assessed (see 2.6.4 Data Assessment: Tumor heterogeneity by (ecological) measures of diversity). Tumor heterogeneity and clonal evolution in recurring adenomas were further investigated by modeling phylogenetic trees based on single-cell miFISH utilizing FISHtree software (Chowdhury et al., 2013). The node depth of the respective trees was measured as a parameter to compare the degree of copy number alterations within the clonal population of the samples. Based on observed CNAs by single-cell miFISH, the average signal number (ASN) per gene identifier probe and per sample were calculated. Adenoma samples were unsupervised clustered by *Gene Cluster* software comparing the ASNs of probe markers to generate a heat map with the following criteria (Wangsa et al., 2016):

(i)   genes were normalized
(ii)   genes were centered by the mean
(iii)   genes and arrays were non-centered correlated
(iv)   genes and arrays were clustered via complete linkage

The obtained tree was visualized by *Java Tree View* which color-coded copy number gain, loss, and neutrality, respectively. The cluster tree was divided into distinct cluster groups.

Moreover, the bp-length of CNAs in relation to affected chromosome arms of colorectal adenomas detected by aCGH were clustered as heat map (applying the above-mentioned criteria). Neutral copy numbers were annotated as 0%, while focal CNAs were expressed as ≤10%. More extensive alterations were annotated in accordance to the respective bp-length in steps of 20%, *i.e.*, whole-chromosome aberrations equaled 100%.

3 RESULTS

3.1 CpG methylation in colorectal (recurring) adenomas

3.1.1 Methylation array analysis

Sixty-nine colorectal adenoma samples and three additional normal mucosa specimens of 62 biological individuals were evaluated for DNA methylation utilizing the HM450K. Prior to analysis, the data was normalized and checked for technical artifacts (2.4.3). The array's 485,577 probes were filtered, leaving 329,573 CpGs for analysis. Excluding putative batch effects, samples were correlated with one of the six HM450K assays. Indeed, a batch-effect was ruled out as confirmed by random distribution.

First, a principal component analysis (PCA) of the mean β-value per sample showed a wide-spread distribution (*i.e.*, the distance of component scores) (Figure 22A). Based on PCA, it was not possible to differentiate between primary adenoma without recurrence, primary adenoma with recurrence, recurrent adenoma, and normal mucosa.

Second, an unsupervised hierarchical cluster analysis based on the mean β-value per sample was performed seeking to distinguish the adenoma subgroups (Figure 22B). Cluster analysis revealed three cluster groups; however, cluster 3 was sharply different from the remaining clusters 1 and 2. The sample annotation was tested for recurrence association, *i.e.*, primary adenomas without recurrence, primary adenomas with recurrence and recurrent adenomas, and showed only a tendency (P=0.171; Figure 22C). The strong divergence of cluster 3 was, however, not associated with adenoma recurrence. Aiming to reveal an explanation for this divergence cluster associations based on mean β-values were dichotomized (*i.e.*, cluster 1 and 2 versus 3) and correlated with clinical parameters (Spearman's correlation coefficient). Neither patient age (P=0.496), nor adenoma localizations (P=0.837), nor right- and left-sided location (P=0.869), nor adenoma size (P=0.671), nor gender (P=0.511) showed an association with the separation between clusters 1 and 2 versus 3. Nevertheless, dichotomized clustering was significantly associated with the histological subtype (P=0.008; Figure 22D). This finding implies that the histological phenotype is also associated with a specific epigenetic DNA methylation pattern. As a side note, trichotomized cluster assignment showed a trend of divergent distribution of females and males (P=0.052; Figure 22E); although patient genders were balanced throughout the cohort (n=34 vs n=38). Altogether, the mean β-value per sample was not capable of discriminating the adenoma subgroups. This finding demanded further analysis of the DNA methylation data.
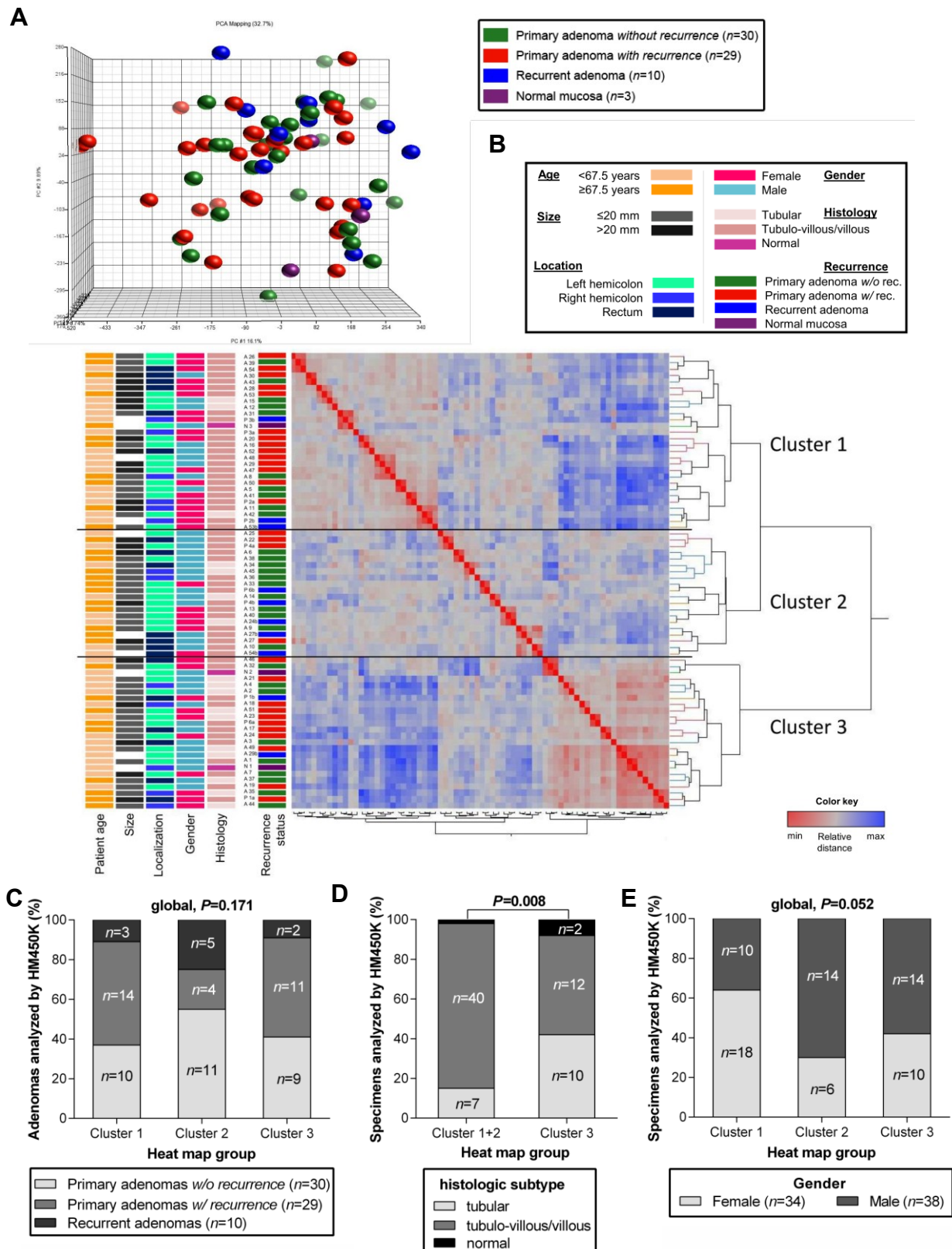
Figure 22. DNA methylation in colorectal adenomas: PCA, cluster analysis, and correlations. (**A**) PCA projection of the entire sample set analysed by HM450K (mean β-value). Defined is an orthogonal coordinate system in which the variance of the samples is plotted and given by the component score. The analysis did not reveal personal associations with the adenoma subgroups. (**B**) Unsupervised hierarchical clustering (Euclidean distance and Ward's group linkage) of the sample set was performed by mean β-values per sample. Cluster analysis exhibited three cluster groups. Cluster 3 diverged stronger from the other clusters. (**C**) Separation of clusters failed short of an association with the adenoma subgroups (Freeman-Halton test). (**D**) Dichotomized clustering was significantly associated with the histologic subtype (Freeman-Halton test). Tubular adenomas were enriched in cluster 3. (**E**) Cluster groups displayed a trend of a non-equal distribution of males and females (Freeman-Halton test).

3.1.2 Identification of differentially methylated CpG positions

The filtered 329,573 probes were used for the detection of differentially methylated CpG positions (DMPs) when comparing the adenoma groups (Table 20). For exploratory analysis, false discovery rates (FDR; multiple-testing adjusted *P*-values) were detected using stringent criteria of *q*≤0.01, *q*≤0.05, and *q*≤0.1, respectively. Neither did computing the differences of primary adenomas without recurrence compared to primary adenomas with recurrence unveil any significant DMP, nor did the comparison of matched pair adenomas, *i.e.*, primary versus corresponding recurrent adenomas (Table 20). Comparing primary adenomas without recurrence versus adenomas associated with recurrence (primary and recurrent) failed to reveal any DMP (Table 20).

Nevertheless, the comparison of recurrent adenomas versus primary adenomas with recurrence (matched and non-matched) exhibited 2,415 (0.7%) DMPs of 329,573 assessed sites (*q*≤0.05; Table 20). Twenty-two DMPs were very significantly different (*q*≤0.01) in this comparison. 99% of DMPs were hypermethylated in recurrent adenomas compared to the primary adenomas with recurrence (Figure 23A). Next, recurrent adenomas were compared versus the complete set of primary adenomas, *i.e.*, adenomas without and with recurrence, which resulted in 575 (0.2%) DMPs (*q*≤0.05; Table 20). Fifty-four DMPs differed very significantly (*q*≤0.01). Hypermethylation of recurrent adenomas versus all primary adenomas was observed in 99% of these 575 DMPs (Figure 23B). Finally, colorectal adenomas were compared to normal mucosa which exhibited 9,266 (2.8%) DMPs (*q*≤0.05; Table 20), of which 5,457 (1.6%) were very significant differential (*q*≤0.01). Notably, 73% of DMPs were hypomethylated in adenomas compared to normal mucosa (Figure 23C).

Table 20. Significant differentially methylated CpG positions (DMPs) in adenoma groups.

| #	Specimen groups	Sample number	Description	Significant DMPs (*n*)		
				FDR ≤0.01	FDR ≤0.05	FDR ≤0.1
1	Primary adenomas *without recurrence* vs Primary adenomas *with recurrence*	30 vs 29		0	0	0
2	Matched pairs: Primary adenomas *with recurrence* vs Recurrent adenomas	10 vs 10		0	0	0
3	Primary adenomas *without recurrence* vs Primary adenomas *with recurrence* and recurrent adenomas	30 vs 39		0	0	0
4	Recurrent adenomas vs Primary adenomas *with recurrence* (matched and non-matched)	10 vs 29	Comparison A	22	2,415	14,855
5	Recurrent adenomas vs Primary adenomas *with and without recurrence*	10 vs 59	Comparison B	54	575	2,667
6	Colorectal adenomas vs Normal mucosae	69 vs 3	Comparison C	5,457	9,266	19,065
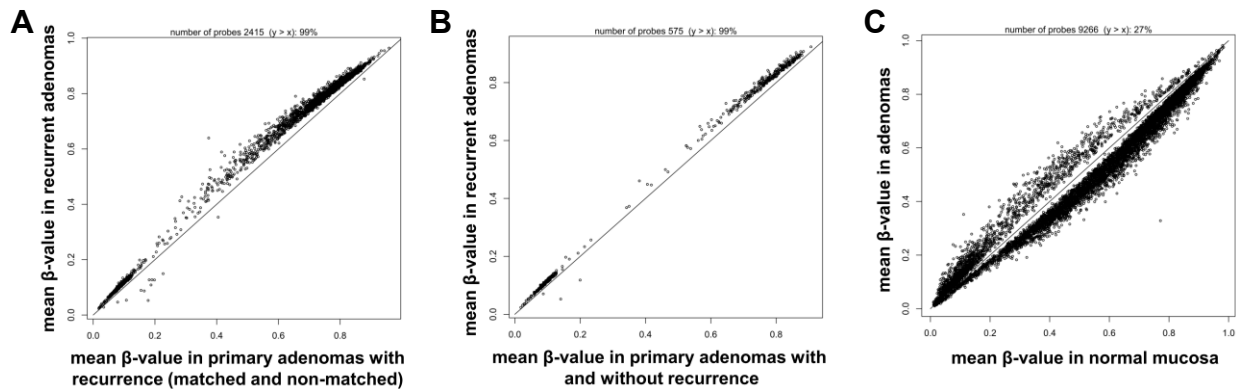
Figure 23. Scatter plot distribution of DMPs identified by HM450K. (**A**) Scatter plot of DMPs in recurrent adenomas versus primary adenomas with recurrence (matched and non-matched). Each DMPs is plotted by the ratio of the mean β-value per group. The apparent shift of probes towards the y-axis is the result of hypermethylation in recurrent adenomas compared to the primary adenomas. (**B**) Scatter plot of DMPs in recurrent adenomas versus the complete set of primary adenomas. Each DMPs is plotted by the ratio of the mean β-value per group. Most probes were stronger methylated (*i.e.*, hypermethylated) in recurrent adenomas. (**C**) Scatter plot of DMPs in adenomas versus normal mucosa. Each DMPs is plotted by the ratio of the mean β-value per group. Here, 73% of DMPs are hypomethylated (dots below the diagonal line) and 27% are hypermethylated in adenomas compared to normal mucosa.

The provided number of 12,256 DMPs (total of all comparisons; $q \leq 0.05$) was tested for CpG sites which were shared or non-shared across the comparisons (Figure 24A). Seventeen probes were common across the three comparisons with the following ten most significant ($q < 0.04$) genes: *MPPED1*, *AACS*, *FARP1*, *SMARCA4*, *TTC25*, *SIDT2*, *NUMBL*, *ABCC12*, *TBC1D12*, and *ZNF655*. However, these common CpG dinucleotides did not necessarily overlap regarding hyper-/hypomethylation as this was defined depending on the comparisons. For instance, gene *ZNF655* was hypermethylated in colorectal adenomas versus normal mucosa (20.8% vs 13.5%), but hypomethylated in recurrent adenomas compared to primary adenomas (5.3% vs 17.8%).

With the objective to identify the most differential and biologically truly relevant methylated CpG positions, obtained DMPs were filtered once again excluding probes with fold changes <10%. Reassuringly, 12,256 DMPs were reduced to 5,094 DMPs of which 2,824 (55.4%) probes were hypermethylated, and 2,270 (44.6%) were hypomethylated. DMPs composed of *comparison A* (674; 98% hypermethylated), *comparison B* (241; 98% hypermethylated) and *comparison C* (4,179; 54% hypomethylated), respectively (Figure 24B). These DMPs were finally stringently filtered for an absolute change (*i.e.*, Δβ) ≥|0.1| providing 35 *top gene*-DMPs (97% hypermethylated) for *comparison A* (Supplemental Table 2), 7 *top gene*-DMPs (71% hypermethylated) for *comparison B* (Supplemental Table 3) and 347 *top gene*-DMPs (68% hypomethylated) for *comparison C* (top50: Supplemental Table 4). Candidates were subsequently selected from the list of *top gene*-DMPs for technical validation via pyrosequencing (see 2.4.4).

Among the *top genes*-list of the comparison of recurrent adenomas versus primary adenomas with recurrence (matched and non-matched; comparison A), 30 of 35 (85.7%) DMP-associated genes were listed only once. Two DMPs in an open sea were associated with *TMEM85* (transmembrane protein 85). Interestingly, three DMPs were in a CpG island which is located in the gene body of the *GREM2* (gremlin 2) gene, a member of the differential screening-selected gene in neuroblastoma (DAN) family of bone morphogenetic protein (BMP) antagonists and TGF-β modulators. Additionally, another member of the TGF-β/BMP-axis was among the 35 identified *top gene*-DMPs, namely *BMP3* (bone morphogenetic protein 3). This supported the ultimate selection of the *GREM2*-associated CpGs cg01809217 (Δβ=0.103; *q*=0.026) and cg02577267 (Δβ=0.101; *q*=0.048) as validation targets. The biological function, the relevance, and the statistical significance of three DMPs (*i.e.*, hypermethylated in recurrent adenomas) in this locus seemed striking (see 3.1.5 Validation of *GREM2* by pyrosequencing).
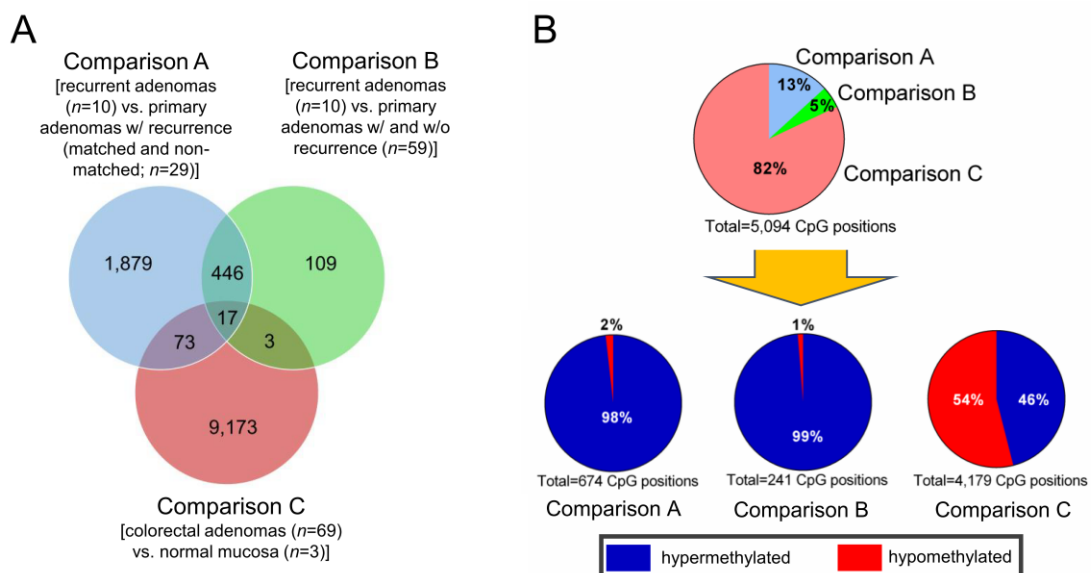
Figure 24. Venn-diagram and distribution of DMPs by multiple comparisons. (**A**) Venn-diagram depicts the overlaps of DMPs across the three comparisons (A, B and C) of sample groups. (**B**) The upper pie-chart displays the distribution of DMPs within the three comparisons. Included are DMPs only which scored FDR≤0.05 with fold change >10%. These DMPs are further distributed on the three comparisons as shown in the lower panel of pie-charts. *Comparisons A* and *B* exhibited an evident hypermethylation in recurrent adenomas. *Comparison C* was more balanced.

3.1.3 Genome-wide distribution of DMPs

Genomic localizations of the DMPs were classified in regions defined by the closest CpG island (see 2.4.3), including information about the genomic direction as upstream (*i.e.*, north) or downstream (*i.e.*, south), to get an impression of the distribution of DNA methylation. The total count of 5,094 DMPs (4,917 when overlaps between comparisons were filtered out), identified by FDR≤0.05 and fold change >10% (Figure 24B),

demonstrated an unequal distribution across the genome with an enrichment of hyper-methylated DMPs in CpG islands (33.7%; 1,655/4,917). Contrary, DMPs in open sea regions were hypomethylated (24.1%; 1,186/4,917; Figure 25A). Shore and shelf regions exhibited a low number of DMPs independent whether hyper- or hypomethylated. Note, the average number of probes located in shelf regions was lower than CpG islands (4.1 vs 5.6). Comparing the DMP-localizations by ratios of hypermethylation versus hypomethylation revealed that more than 75% of DMPs were hypomethylated in open sea and shelf regions, respectively (Figure 25B). However, the most frequent abundance (86%) was documented for hypermethylated CpG island DMPs, while shore regions seemed balanced with a slight excess of hypermethylation (Figure 25B). The distribution of DMPs was analyzed per comparison which exhibited an excess of DMPs located in CpG islands of *comparison A* (*i.e.*, recurrent adenomas versus primary adenomas with recurrence; 47%) and *comparison B* (*i.e.*, recurrent adenomas vs the complete set of primary adenomas; 66%; Figure 25C). Interestingly, *comparison C* (*i.e.*, adenomas vs normal mucosa) showed a more homogeneous distribution of DMPs across islands (37%), shores (27%) and open sea (31%; Figure 25C).

The distribution of the 4,917 DMPs was further analyzed regarding functional genomic regions: proximal promoter (*i.e.*, upstream of the transcription start site, 5'UTR and 1st exon), gene body and 3'UTR (Figure 25D). Proximal promoter DMPs were most frequently hypermethylated, whereas DMPs in the gene body were in total numbers most abundant and more frequently hypomethylated (Figure 25D). However, the mean number of probes per functional region on the HM450K needs to be considered: proximal promoters 15.2 probes, gene bodies 9.9 probes and 3'UTR 1.5 probes. The non-random distribution of hyper-and hypomethylation was further confirmed by analyzing the ratio of hypermethylated versus hypomethylated DMPs per functional region (Figure 25E). DMPs of the proximal promoter were more frequently hypermethylated (exceeding 50% in each subregion), and DMPs in the gene body and 3'UTR were more frequently hypomethylated. The DMP-distribution of gene regions per individual comparison confirmed a large excess of DMPs located in promoter regions with the highest percentage (74%) in *comparison B* (*i.e.*, recurrent adenomas vs the complete set of primary adenomas; Figure 25F).

Finally, the distribution of DMPs was plotted by chromosomal location (sex-chromosomes excluded) which exhibited the highest frequencies of DMPs located on chr1 (10.8% hypermethylated, 11.3% hypomethylated; Figure 25G). However, this is also the largest chromosome and has the highest gene-density.
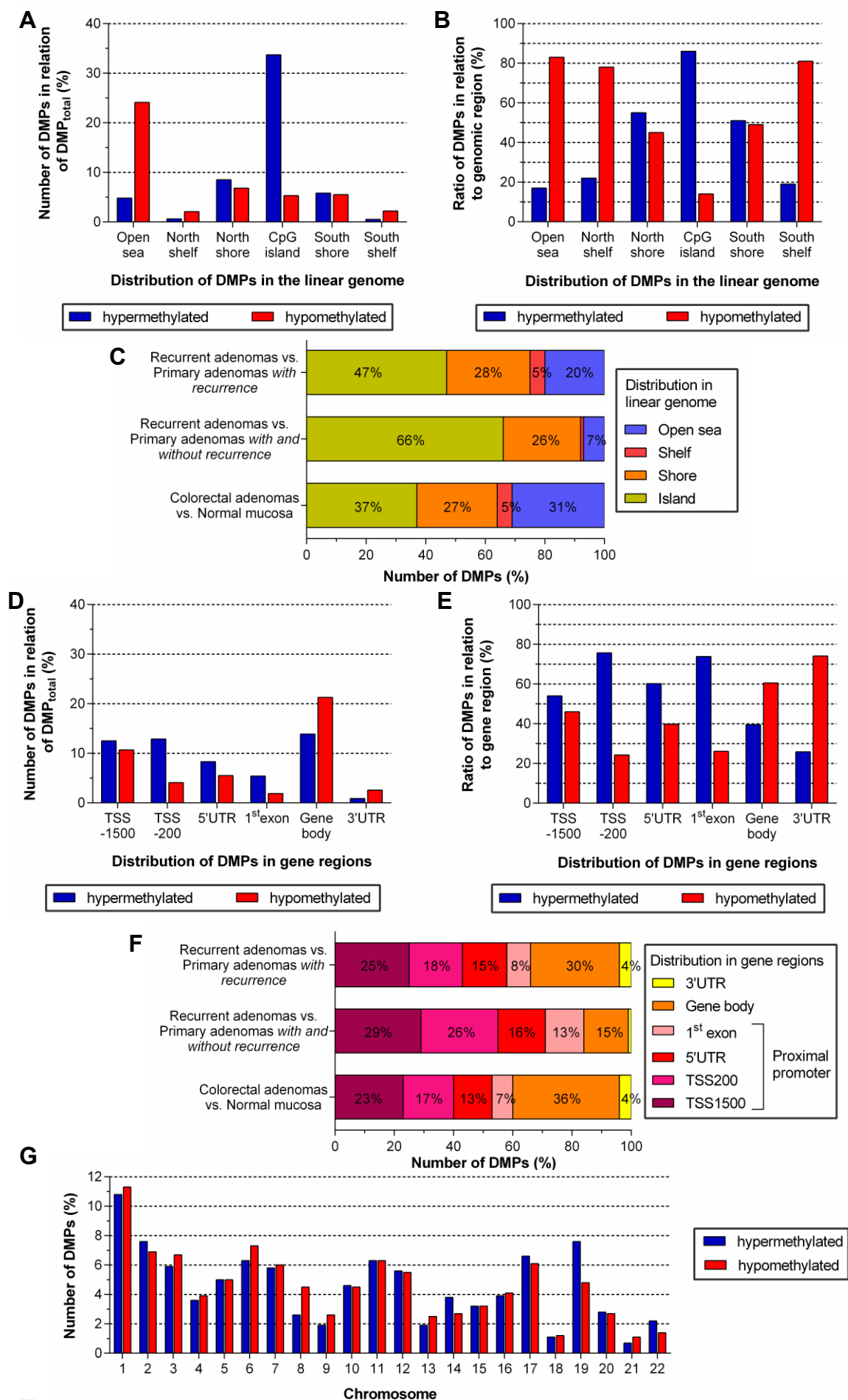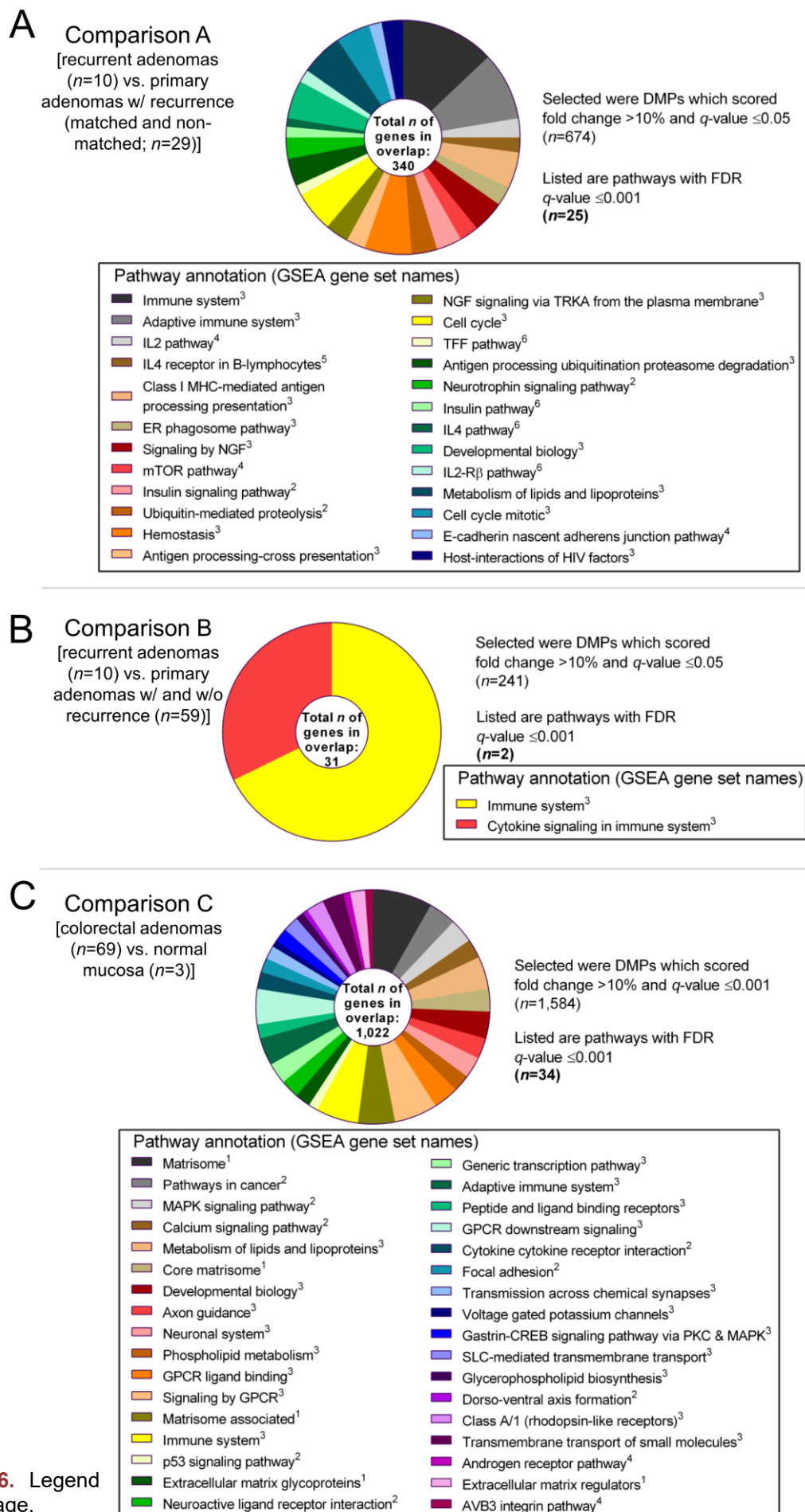
Figure 25. Legend on next page.

Figure 25. Distribution of DMPs by genome localizations and chromosomes. (**A**) Distribution of DMPs in regards of CpG content in the linear genome. Shown percentages are computed with the total number of DMPs (n_{total}=4,917). (**B**) Depicted is the ratio of hypermethylated vs hypomethylated DMPs in multiple genomic regions defined by CpG content. (**C**) Distribution of DMPs in regard to genomic regions defined by CpG content and sorted by individual *comparisons*. (**D**) Distribution of DMPs in gene-associated regions defined as 200-1,500 bp upstream from transcription start site (TSS1500), up to 200 bp upstream from transcription start site (TSS200), 5'untranslated region (5'UTR), 1st exon, gene body and 3'untranslated region (3'UTR), respectively. Shown percentages are computed in relation to the total number of DMPs (n_{total}=4,917). (**E**) Ratio of hypermethylated versus hypomethylated DMPs in different gene-associated regions. (**F**) Distribution of DMPs in regards of gene-associated regions and sorted by individual *comparisons*. (**G**) Chromosomal distribution (sex chromosomes excluded) of DMPs categorized as hyper- or hypomethylated, respectively. Note that chr8 showed the lowest ratio of hypermethylation/hypomethylation (*i.e.*, hypomethylation enriched), while chr19 showed the highest ratio.

3.1.4 Functional pathway enrichment of DMPs

The DMPs identified per individual comparison (Figure 24B) were annotated by pathways and biological functions via gene set enrichment analysis (GSEA; Figure 26). The input of 674 DMPs of comparison A resulted in 632 individual genes found in the database which provided 25 different pathways and gene ontologies highly significantly (q≤0.001) associated with 340 genes (Figure 26A). The most significant were genes involved in the immune system (n=44; q=2.9x10^{-9}) and the adaptive immune system (n=32; q=4.8x10^{-9}) as well as in inflammation-mediated chemokine and cytokine pathways, *e.g.*, IL2 pathway (n=9; q=2.1x10^{-5}; Figure 26A). Interesting targets in adenoma recurrence might also be genes associated with cellular survival, growth and differentiation via signaling by neural growth factor (NGF; n=15; q=6.9x10^{-5}), mTOR pathway (n=9; q=6.9x10^{-5}) and cell cycle pathways (n=19; q=7.3x10^{-4}; Figure 26A).

Comparison B provided of 241 DMPs of 233 individual genes which were found in the database, however only two pathways/ontologies were highly significantly associated with 31 genes (Figure 26B): again, genes involved in the immune system (n=21; q=2.1x10^{-5}) and cytokine signaling in the immune system (n=10; q=9.3x10^{-4}; Figure 26B). Note that eighteen other pathways, *e.g.*, cell cycle checkpoints, reached q≤0.01.

Finally, 4,179 DMPs of comparison C had to be filtered by q≤0.001 (resulting in n=1584) to reduce the data input. These DMPs provided 1,396 individual genes of which 34 pathways were highly significantly associated with 1,022 genes (Figure 26C). Genes encoding not only extracellular matrix (ECM) and ECM-associated proteins (n=84; q=5.8x10^{-13}) and core ECM-proteins (n=32; q=2.4x10^{-8}) but also "classical targets of tumorigenesis" such as pathways in cancer (n=39; q=3.4x10^{-10}), MAPK signaling (n=34; q=1.0x10^{-9}) or p53 signaling (n=13; q=1.3x10^{-5}; Figure 26C) appeared to be the most striking. Genes involved in the immune system (n=60; q=5.7x10^{-6}) and adaptive immune system (n=39; q=5.1x10^{-5}) were also enriched as described before.

A Comparison A

[recurrent adenomas (*n*=10) vs. primary adenomas w/ recurrence (matched and non-matched; *n*=29)]

Total *n* of genes in overlap: 340

Selected were DMPs which scored fold change >10% and *q*-value ≤0.05 (*n*=674)

Listed are pathways with FDR *q*-value ≤0.001 (**n=25**)

Pathway annotation (GSEA gene set names)

- Immune system[3]
- Adaptive immune system[3]
- IL2 pathway[4]
- IL4 receptor in B-lymphocytes[5]
- Class I MHC-mediated antigen processing presentation[3]
- ER phagosome pathway[3]
- Signaling by NGF[3]
- mTOR pathway[4]
- Insulin signaling pathway[2]
- Ubiquitin-mediated proteolysis[2]
- Hemostasis[3]
- Antigen processing-cross presentation[3]
- NGF signaling via TRKA from the plasma membrane[3]
- Cell cycle[3]
- TFF pathway[6]
- Antigen processing ubiquitination proteasome degradation[3]
- Neurotrophin signaling pathway[2]
- Insulin pathway[6]
- IL4 pathway[6]
- Developmental biology[3]
- IL2-Rβ pathway[6]
- Metabolism of lipids and lipoproteins[3]
- Cell cycle mitotic[3]
- E-cadherin nascent adherens junction pathway[4]
- Host-interactions of HIV factors[3]

B Comparison B

[recurrent adenomas (*n*=10) vs. primary adenomas w/ and w/o recurrence (*n*=59)]

Total *n* of genes in overlap: 31

Selected were DMPs which scored fold change >10% and *q*-value ≤0.05 (*n*=241)

Listed are pathways with FDR *q*-value ≤0.001 (**n=2**)

Pathway annotation (GSEA gene set names)

- Immune system[3]
- Cytokine signaling in immune system[3]

C Comparison C

[colorectal adenomas (*n*=69) vs. normal mucosa (*n*=3)]

Total *n* of genes in overlap: 1,022

Selected were DMPs which scored fold change >10% and *q*-value ≤0.001 (*n*=1,584)

Listed are pathways with FDR *q*-value ≤0.001 (**n=34**)

Pathway annotation (GSEA gene set names)

- Matrisome[1]
- Pathways in cancer[2]
- MAPK signaling pathway[2]
- Calcium signaling pathway[2]
- Metabolism of lipids and lipoproteins[3]
- Core matrisome[1]
- Developmental biology[3]
- Axon guidance[3]
- Neuronal system[3]
- Phospholipid metabolism[3]
- GPCR ligand binding[3]
- Signaling by GPCR[3]
- Matrisome associated[1]
- Immune system[3]
- p53 signaling pathway[2]
- Extracellular matrix glycoproteins[1]
- Neuroactive ligand receptor interaction[2]
- Generic transcription pathway[3]
- Adaptive immune system[3]
- Peptide and ligand binding receptors[3]
- GPCR downstream signaling[3]
- Cytokine cytokine receptor interaction[2]
- Focal adhesion[2]
- Transmission across chemical synapses[3]
- Voltage gated potassium channels[3]
- Gastrin-CREB signaling pathway via PKC & MAPK[3]
- SLC-mediated transmembrane transport[3]
- Glycerophospholipid biosynthesis[3]
- Dorso-ventral axis formation[2]
- Class A/1 (rhodopsin-like receptors)[3]
- Transmembrane transport of small molecules[3]
- Androgen receptor pathway[4]
- Extracellular matrix regulators[1]
- AVB3 integrin pathway[4]

Figure 26. Legend on next page.

Figure 26. Distribution of DMPs according to pathway analysis/gene ontology. (**A**) Circle plot shows the frequency of genes (picked from DMPs of *comparison A*) associated with the respective pathways/gene ontologies, which are listed in the legend. Results show that genes involved in inflammatory-pathways were differentially methylated in this comparison. (**B**) Circle plot shows the frequency of genes (picked from DMPs of *comparison B*) associated with the respective pathways/gene ontologies, which are listed in the legend. Here, immunologic pathways only were identified to be highly significantly associated with the DMPs. (**C**) Circle plot shows the frequency of genes (picked from DMPs of *comparison C*) associated with the respective pathways/gene ontologies, which are listed in the legend. Results document the tumor-related differential methylation of multiple cellular signaling pathways and extracellular matrix proteins. Not that genes may be assorted to several pathways.

All pathways were ordered by significance and not by the frequency of genes involved. Pathway annotations relate to database entry: [1]=NABA; [2]=KEGG; [3]=REACTOME; [4]=PID; [5]=SIG; [6]=BioCarta.

3.1.5 Validation of CpG methylation by pyrosequencing

Methylation frequencies detected by HM450K were sought to be validated by selection (*i.e.*, highly significant *p*-values across group comparisons, differential methylation fold change exceeding 10%, located in CpG island) of a candidate gene. Pyrosequencing validated the locus-specific methylation alterations of CpG sites in CpG islands of the gene. Three CpG sites within *GREM2* (Gremlin 2) diverged when comparing primary adenomas with/without recurrence to recurrent adenomas (see 3.1.2) suggesting this antagonist of the BMP (bone morphogenetic protein) pathway as a potential target.

Highly specific probe design incorporated two CpG sites (*i.e.*, CpG position 1=cg01809217 and CpG position 15=cg02577267) covered by HM450K (Figure 16A). Assays were firstly validated by technical replicates to test the array reproducibility (Figure 27A; see 2.4.4 Pyrosequencing): methylation status of *GREM2* forward- and reverse-assay were 8-times and 4-times, respectively, measured in glioblastoma cell lines, which confirmed a high validity (Figure 27A). Bisulfite-converted DNAs of primary adenomas with (*n*=15) and without (*n*=9) recurrence and recurrent adenomas (*n*=4) were analyzed for *GREM2* gene body methylation frequencies (Figure 15; Figure 27B). While *GREM2* methylation frequencies were highly variable across samples, methylation ratios of analyzed CpG sites were detected to be low (*i.e.*, hypomethylated; ≤33%) in 14.6% (28/192), medium (>33-66%) in 49.5% (95/192) and high (*i.e.*, hypermethylated; >66%) in 35.9% (69/192) of specimens, respectively. Addressing the comparability of *GREM2* CpG dinucleotide methylation patterns across adenomas with and without recurrence the mean methylation frequencies per CpG site was calculated (Figure 27C, D). Interestingly, primary adenomas showed very similar ratios across CpG sites 1-4 whereas recurrent adenomas tended to exhibit slightly higher mean methylation frequencies (Figure 27C). On the contrary methylation ratios of CpG sites 12-15 were found to be more heterogeneous across adenoma groups: recurrent

adenomas unveiled the lowest frequencies while primary adenomas showed, on average, the higher methylation frequencies (Figure 27D). As a side note, methylation ratios of CpG 2 and CpG 13, respectively, had the highest (mean, 70% ± 18%) and the lowest (mean, 38% ± 15%) ratios across specimens.

In-depth analysis conducted by calculating the average methylation frequency per sample based on cumulated *GREM2* CpG dinucleotides 1-4 showed no difference comparing the adenoma groups (P=0.814; Figure 28A). Despite that, mean methylation ratios of cumulated *GREM2* CpG dinucleotides 12-15 seemed to tend weakly towards lower average methylation frequencies in primary adenomas with recurrence compared to primary adenomas without recurrence (P=0.172; Figure 28B). Comparing the median methylation frequencies of individual CpG dinucleotides 1-4 sorted by adenoma groups failed to reveal a difference across primary adenomas with and without recurrence and recurrent adenomas (P=0.590; Figure 28C), although recurrent adenomas seemed to be hypermethylated compared to primary adenomas without (P=0.406) and with (P=0.312) recurrence. However, the comparison of median methylation frequencies of individual CpG dinucleotides 12-15 unveiled a significant difference comparing primary adenomas with recurrence to adenomas without recurrence (P=0.042; Figure 28D). Moreover, methylation ratios of CpG sites 12-15 were highly significantly different across primary adenomas with recurrence, primary adenomas without recurrence and recurrent adenomas (P=0.002; Figure 28D). Here, recurrent adenomas were hypomethylated compared to primary adenomas without (P=0.004) and with (P=0.003) recurrence. According to our data, the combination of *GREM2* CpG dinucleotides 12-15 provides the most reliable statistical power to discriminate primary adenomas without recurrence, primary adenomas with recurrence and recurrent adenomas via differential methylation.

Average percentages of *GREM2* DNA methylation of CpG position 1 (cg01809217) were measured on the HM450K as β=0.716 for recurrent adenomas (n=10) and β=0.613 for primary adenomas with recurrence (matched and non-matched; n=29). Pyrosequencing showed values of β=0.753 for recurrent adenomas (n=4) and β=0.687 for primary adenomas with recurrence (matched and non-matched; n=14) (Figure 27C). These values are in line with the array results, although a smaller number of samples was investigated by pyrosequencing ($\Delta\beta_{HM450K}$=0.103 versus $\Delta\beta_{Pyroseq}$=0.066) (Figure 27C, D).
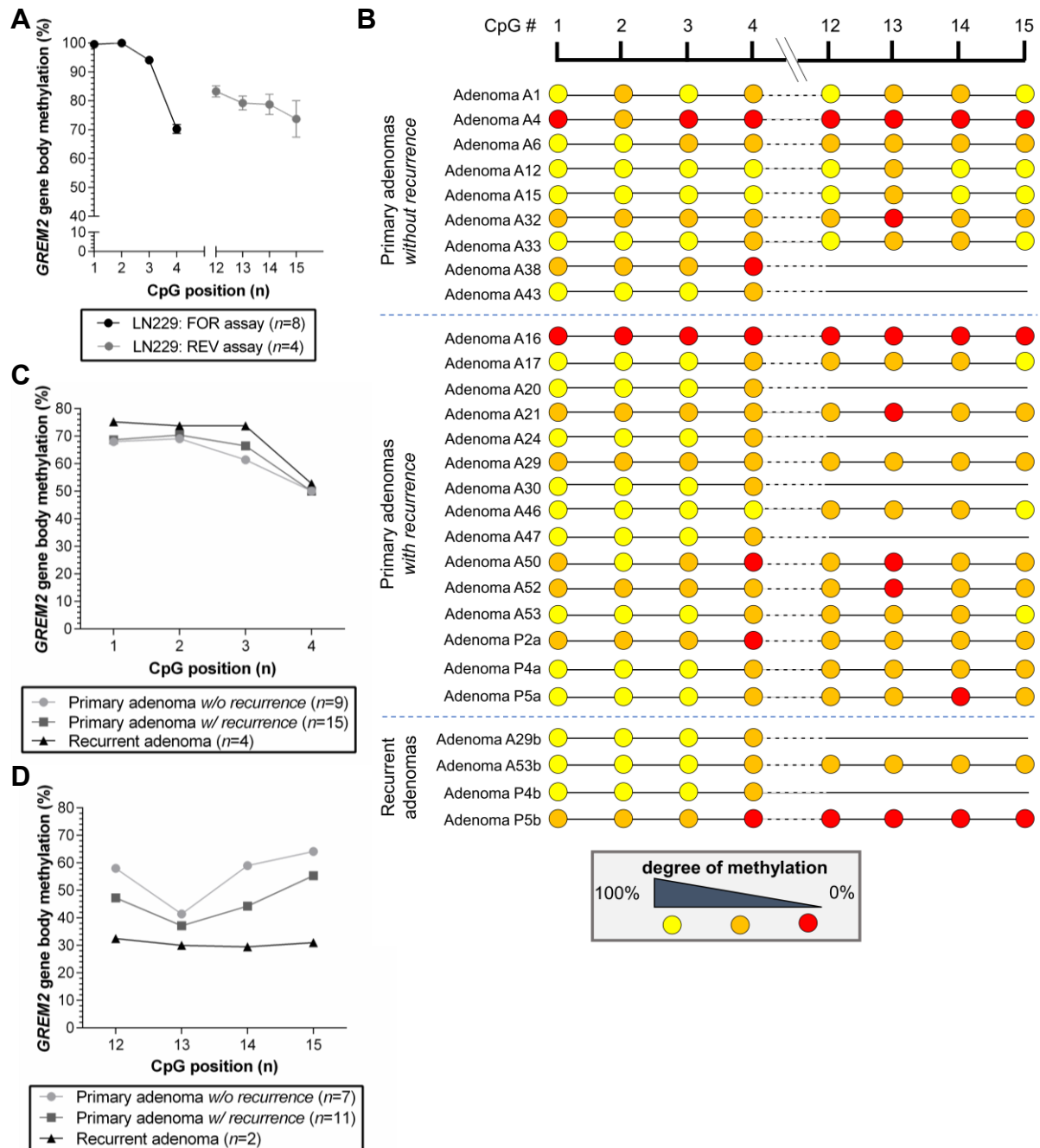
Figure 27. CpG methylation validation by pyrosequencing. (**A**) Quantitative *GREM2* gene body methylation analysis (mean ± SEM) of glioblastoma cell line LN229 (served as the technical control) confirms the reliability and reproducibility of results obtained by pyrosequencing. (**B**) Detected methylation values are depicted as lollipops resembling the color-coded degree of methylation from high to low (*i.e.*, hypermethylation to hypomethylation) in trichotomized order (*yellow > orange > red*). CpG sites of the reverse-assay whose methylation levels were not determined (N.D.) are shown as straight lines. CpG number 1-4 were determined by the forward-assay whereas CpG number 12-15 were analyzed by the reverse-assay. (**C**) Quantitative *GREM2* gene body methylation analysis of four CpG dinucleotides (for-assay) in primary adenomas with and without recurrence and recurrent adenomas, respectively. Data points in the pyrogram represent the mean methylation frequency per CpG and subgroup. (**D**) Quantitative *GREM2* gene body methylation analysis of four CpG dinucleotides (rev-assay) in primary adenomas with and without recurrence and recurrent adenomas, respectively. Data points in the pyrogram represent the mean methylation frequency per CpG and subgroup.
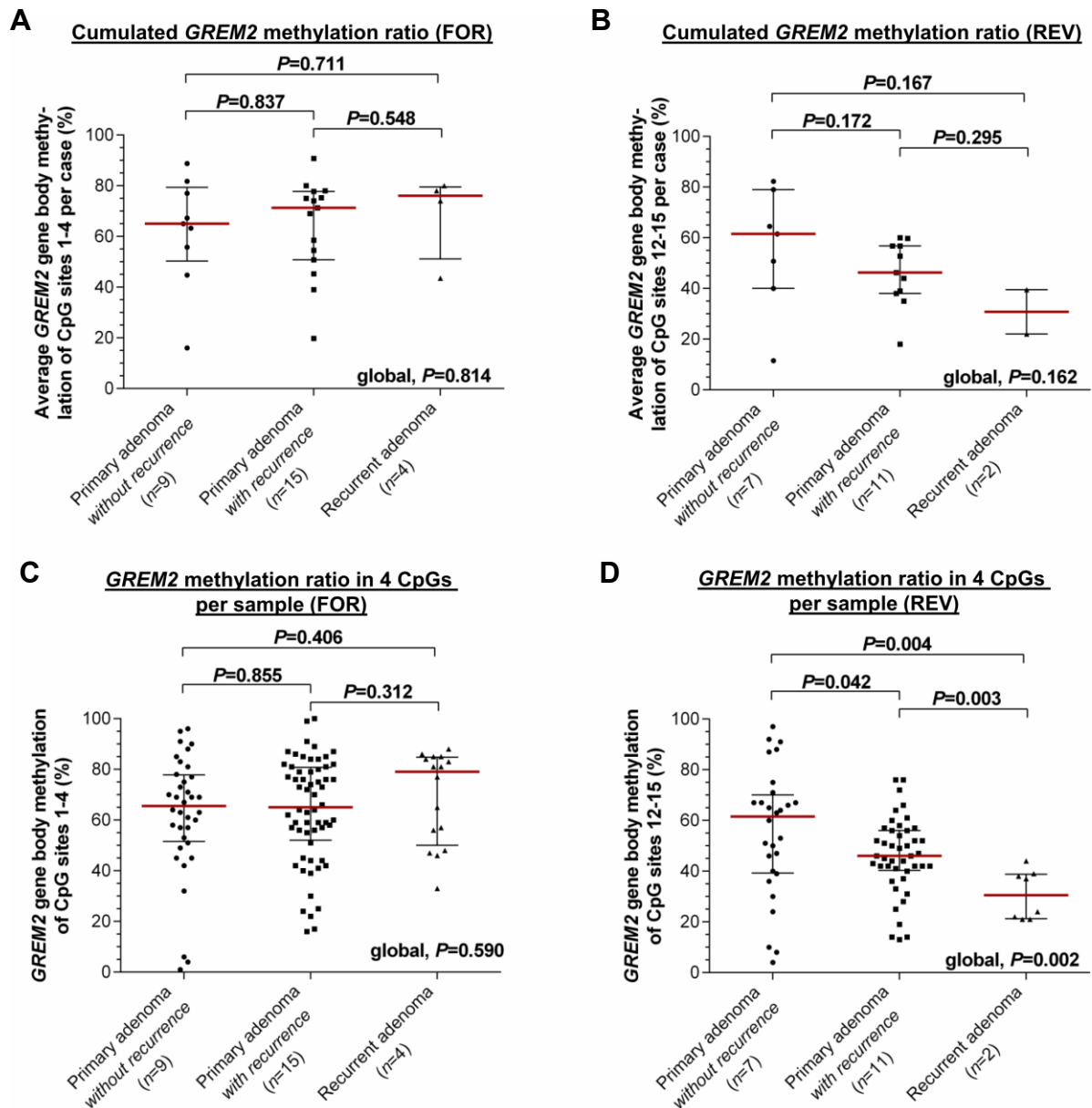
Figure 28. *GREM2* gene body methylation in colorectal adenoma subgroups. (**A**) Computing the mean methylation frequency of cumulated CpG sites 1-4 of the *GREM2* pyrosequencing assay per adenoma did not reveal differences (median with IQR) across primary adenomas with and without recurrence (Mann-Whitney *U* test) or recurrent adenomas (global, Kruskal-Wallis test), respectively. (**B**) Mean methylation frequencies of cumulated CpG sites 12-15 of the *GREM2* pyrosequencing assay showed a trend to differ (median with IQR) not only when comparing primary adenomas with and without recurrence (Mann-Whitney *U* test), but also when including recurrent adenomas in the comparison (Kruskal-Wallis test). (**C**) Plotting the methylation frequencies of CpG sites 1-4 individually per primary adenoma with and without recurrence and recurrent adenoma did not show differences (median with IQR; Mann-Whitney *U* test; global, Kruskal-Wallis test). (**D**) Median (with IQR) methylation frequencies of GpG sites 12-15 plotted individually per subgroup showed a significant difference across primary adenomas with and without recurrence (Mann-Whitney *U* test). Differences were strongly divergent when comparing primary adenomas with and without recurrence and recurrent adenomas (Kruskal-Wallis test).

3.2 Landscape of CNAs in colorectal (recurring) adenomas

3.2.1 Overall aCGH analysis

Analysis of aCGH results detected a total of 178 copy number changes among the non-tetraploid colorectal adenomas (*n*=51) (Figure 15; Table 21). Overall, gains were more frequently observed than losses (means 1.9 ± 2.3 vs 1.6 ± 2.4; *P*=0.113, Mann-Whitney *U* test). The proportion of whole-chromosome changes (defined as ≥90%; 22.5%; 40/178) per adenoma was significantly lower than part-chromosome changes (defined as ≤50%; 77.5%; 138/178) (means 0.78 ± 0.20 vs 2.71 ± 0.50; *P*=0.004; Figure 29A). Frequencies of both whole-chromosome CNAs (*z*=0.88, *P*=0.379) and part-chromosome CNAs (*z*=1.20, *P*=0.230) suggested a distribution by Poisson (*P*>0.100) indicating that CNAs are non-randomly distributed throughout the genome. However, there was no tendency visible that adenomas with larger numbers of whole-chromosome alterations also showed larger numbers of part-chromosome alterations (*r*=0.112; *P*=0.433; Spearman correlation). ANCA (average number of copy alterations) values ranged from 0-16 with a mean of 3.5 ± 4.1 (Table 21). Evaluation of the clinicopathological parameters revealed a positive correlation between the presence of CNAs and the adenoma location (*P*=0.037; Figure 29B). In other words, rectal adenomas showed more CNAs than colonic adenomas. However, none of the other clinicopathological features (adenoma size, histology, gender, grading, time of recurrence, patient age) reached statistical significance when correlated with the number of CNAs.
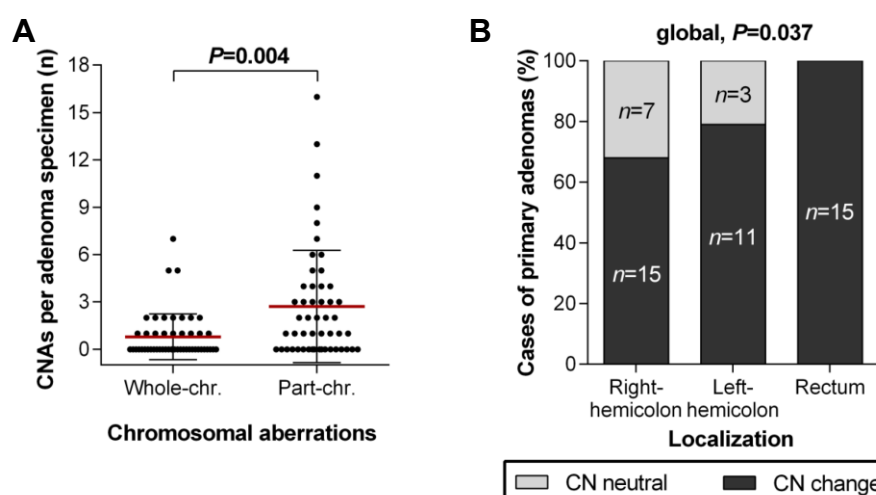
Figure 29. CNAs detected by aCGH. (**A**) Comparison (mean ± SD) of whole-chromosome aberrations and part-chromosome aberrations within all samples analysed (*n*=51) revealed that whole-chromosome changes occurred less frequently (Mann-Whitney *U* test). (**B**) CNAs by aCGH were not randomly distributed among colorectal adenomas (*n*=51) when separated by trichotomized localization (Freeman-Halton test). All analyzed rectal adenomas displayed CNAs.

Table 21. Summary of copy number alterations detectable by aCGH. The table depicts the copy number changes in colorectal adenomas expressed as gain or loss of bps. Changes were of small magnitude, *i.e.*, no amplifications or deep-deletions were detected. Aberrations <1,500kb, copy number variant (CNV)-overlapping regions and sex chromosomes were excluded. F, female; M, male; N.D., not determined; †, tetraploid adenoma; % genome changed, extracted from *Nexus Copy Number*.

Sample ID	Sex	Gains	Losses	No. of CNAs	Whole-chr. aberr.[a]	Part-chr. aberr.[b]	% genome changed
A 1	M	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
A 2	M	-	4q28.3; 4q32.1-q32.3	2	0	2	0.66
A 3	M	3p21.31-p21.1; 7q22.1; 11q12.2-q13.4; 12q13.11-q13.2	1pter-p35.2; 1p33-p32.2; 5q15-q21.1; 16q21	8	0	8	6.25
A 4	M	-	-	0	0	0	0.69
A 5	M	-	-	0	0	0	0.00
A 6	M	6p22.1; 17q12-q21.33; 19pter-p13.11	-	3	0	3	4.01
A 7	F	-	-	0	0	0	0.00
A 8	M	6p22.2-22.1; **8**; **19**	-	3	2	1	6.47
A 9	F	**8**	-	1	1	0	4.80
A 10	M	**5**; **6**; **7**; **14**; **20**	2p12-p11.2; 3q26.1-q26.2; 4q28.3; 10q25.1; 16pter-13.2; 16q21	11	5	6	24.92
A 11	F	6p22.2-22.1; **7**; 20p13-p12.3	-	3	1	2	5.58
A 12	M	6p22.1-p21.32; 7q22.1	14q21.1-q21.3	3	0	3	0.60
A 13†	F	20q	**1**; **4**; **6**; **10**; **16**; **17**; **18**	16	15	1	37.95
A 14	M	**7**	3q26.1-q26.2; 16q21; 18p11.31-p11.23; 18q11.2-q12.1	5	1	4	8.37
A 15	M	7q11.23; 7q22.1; 19q	5q22.1-q23.2	4	0	4	0.66
A 16	M	1q; 6p22.1; 6p21.33-p21.32; 7q22.1; **8**; 19q; **20**	13q21.1-q21.33; 16q21	9	2	7	10.93
A 17	M	3p21.31-p21.1; 6p22.2; 17p13.1	-	3	0	3	2.18
A 18	M	1pter-p33; 6p22.1; 6p21.33-6p21.2; **7**; 12q13.11-q14.1; **13**; 17pter-17q21.33; 19q	2p12-p11.2; 4q28.3; 5p14.3-p13.3; 8p12; 11p12; 14q24.3-q31.3; 16q21	15	2	13	13.54
A 19	M	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
A 20	F	-	-	0	0	0	1.07
A 21	M	6p22.1	-	1	0	1	0.06
A 22	M	6p22.1-p21.32; 7q22.1; 20q	1pter-p35.2; 5q21.1- q22.3; 11q23.1-q24.2; 16p13.3-p13.2; 16q21; **18**; 20p	10	1	9	6.69
A 23	F	**3**; 6p22.1; 7p15.2-p15.1; 11q12.2-q13.3; 17q	5pter-p15.31; 5p14.3-p14.1; 6p12.3-p12.2; 13q21.1-q21.33; 16q21; 17p13.1-p12; 18q12.2-q12.3	12	1	11	10.60
A 24	F	-	-	0	0	0	0.00
A 25	M	-	-	0	0	0	0.00
A 26	F	6p22.1; 12p13.2-p12.3	16q21	3	0	3	2.73
A 27	M	6p21.33; 17q12-q21.32; 19p; 20q	16q21	5	0	5	5.48
A 28	F	6p22.1	16q21	2	0	2	1.21
A 29	M	-	16q21	1	0	1	0.51
A 30†	M	6p22.2-p21.2	**1**; **2**; **5**; **8**; **11**; **15**; **18**	8	7	1	30.98

Case ID	Sex	Gains	Losses	No. of CNAs	Whole-chr. aberr.[a]	Part-chr. aberr.[b]	% genome changed
P 1a	F	6p21.33; 7q22.1; 11q12.2-q13.2	16q21	4	0	4	2.52
P 1b	F	6p22.1-p21.33; 7q22.1; 11q12.2-q13.2; 17q21.2-q21.31	16q21	5	0	5	1.62
P 2a	F	-	-	0	0	0	0.00
P 2b	F	13q21.33-q22.2	4p15.33-p15.1; 16pter-13.2; **19**	4	1	3	4.01
P 3a	F	12q12-q13.2	-	1	0	1	0.50
P 3b	F	3q26.2-q26.31; 4q21.1-q21.21; 5q13.2-q13.3; 5q15-q21.1; 6p21.33; 6q23.3; 8p22-p21.3; 8p21.1-p12; 8q24.21; 12p13.2-p13.1; 13q21.33-q22.1	5pter-p15.31; 15q11.2-q13.1; 17q11.2-q12; 19q13.33-qter; 22q13-qter	16	0	16	5.61
P 4a	M	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
P 4b	M	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
P 5a	F	-	-	0	0	0	0.12
P 5b	F	-	11q22.1-q22.3	1	0	1	1.04
P 6a	M	6p22.1-6p21.33; 7q22.1	18q22.3	3	0	3	1.03
P 6b	M	6p22.1	5p14.1-p13.3	2	0	2	0.71
P 7a	F	**7**	-	1	1	0	5.05
P 7b†	F	6p22.2-p21.1; 7q22.1; **17**; **19**	1p33-p13.3; 1q; **2**; **4**; 11q; **15**; 16q; **18**	12	6	6	25.46
P 8a	F	-	1pter-p35.2	1	0	1	0.81
P 8b	F	-	1pter-p35.2; 16q13-q24.3	2	0	2	1.96
P 9a	F	**13**	-	1	1	0	3.18
P 9b	F	**13**	-	1	1	0	3.11
P 10a	M	**13**; **20**	1pter-p35.1; 1p34.1-p21.2; **4**; 6q16.1; **12**; **14**; 17p; **18**; **22**	11	7	4	25.62
P 10b	M	5q; **13**; **20**	1pter-p35.1; 1p34.1-p21.2; **4**; 9q32-q33.2; 12q; **15**; 17p; **18**	11	5	6	27.29
P 11a	F	**7**; **9**	-	2	2	0	8.55
P 11b	F	**7**; **9**	-	2	2	0	9.55
P 12a	M	-	-	0	0	0	0.19
P 12b	M	-	-	0	0	0	0.39
P 13a	F	**13**; **20**	-	2	2	0	5.18
P 13b	F	**13**; **20**	-	2	2	0	5.23
P 14a	M	-	5q14.3-q22.1	1	0	1	0.77
P 14b	M	-	5q14.3-q22.1	1	0	1	0.73

[a]Whole-chromosome aberrations defined as alteration comprising ≥90 of the chromosome.
[b]Part-chromosome aberrations defined as alteration comprising ≤50% of the chromosome.

3.2.2 Genomic locations of CNAs

CNAs were detected in 41/51 (80.4%) of the colorectal adenomas, *i.e.*, 10/51 (19.6%) of the colorectal adenomas did not reveal a detectable copy number change. The most frequent changes observed applied to focal (≤10 Mb) copy number gains of 6p22.1-p21.33 (17/51; 33.3%) and 7q22.1 (16/51; 31.4%; Table 21) followed by the focal loss of 16q21 (15/51; 29.4%). In contrast, chromosomal aberrations affecting entire

chromosomes and chromosome-arms were less frequent with chromosomes 7 (7/51; 13.7%), 13q (7/51; 13.7%) and 20q (7/51; 13.7%) being most commonly changed in copy numbers (Figure 30A-D). The frequency plot revealed, in general, a low incidence (not exceeding 50%) of genomic imbalances in adenomas without and with recurrence.
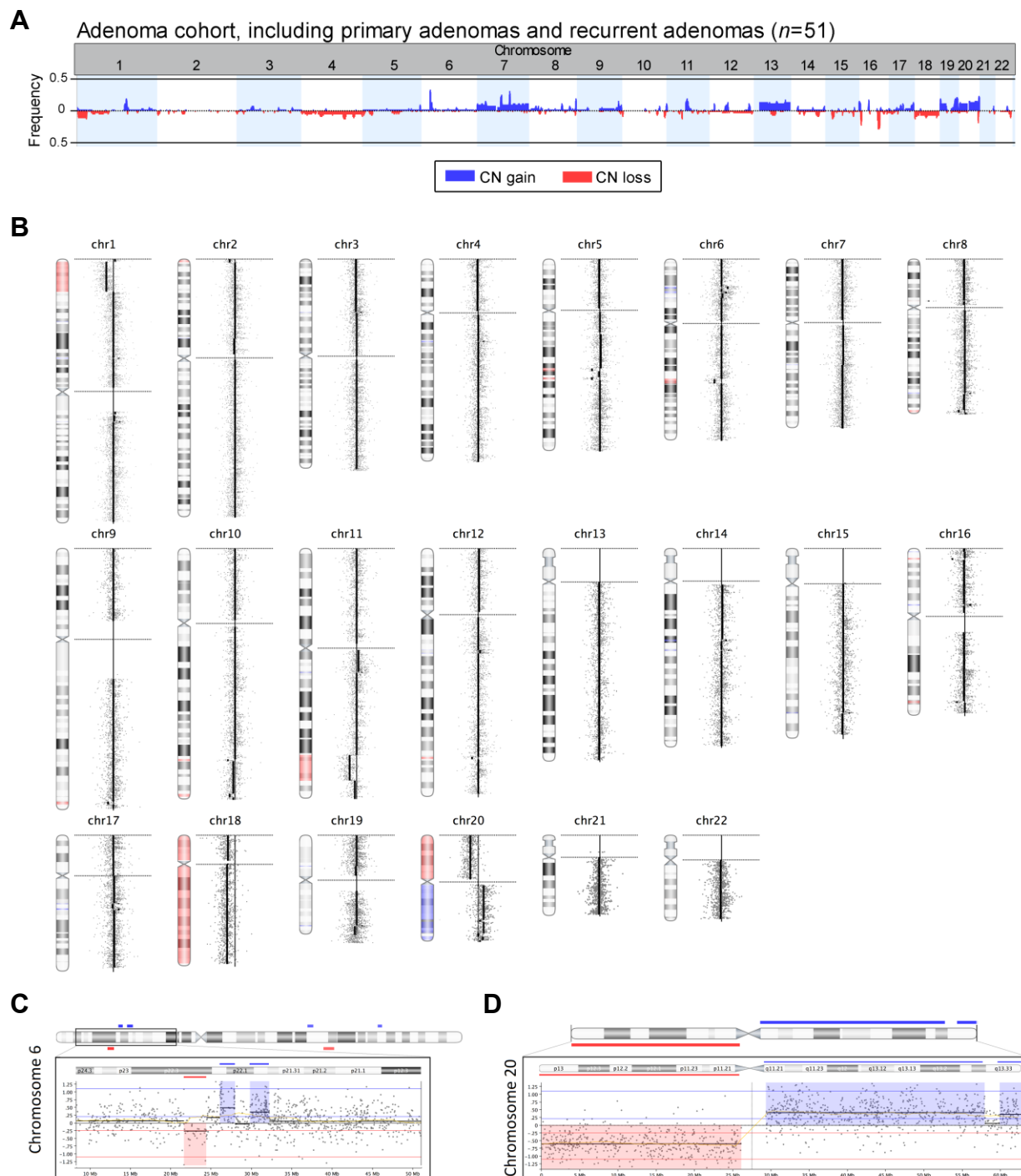
Figure 30. Chromosomal aberrations in colorectal adenomas identified by aCGH. (**A**) Frequency plot of CN gains (*blue*) and losses (*red*) distributed in the adenoma cohort (*n*=51). The most frequent focal alterations were observed on chr6, chr7 and chr16, respectively, while whole-chromosome CNAs were detected on chr7, chr13, chr18 and chr20, respectively. (**B**) Representative aCGH ideogram of genome-wide distributed chromosomal gains and losses (case A22). (**C**) Chromosome view of chr6 which was affected by focal aberrations (zoomed-in: p-arm). Graph shows the distribution of probes via log2-ratio. (**D**) Chromosome view of chr20 which displayed arm-level aberrations (-20p, +20q).

3.2.3 Differential CNAs in primary (recurring) adenomas

The investigated set of primary colorectal adenomas (n=39) exhibited at least one copy number alteration in 76.9% (10/13) of samples without recurrence whereas 80.7% (21/26) of adenomas with recurrence were affected by CNAs (Figure 31A; Table 21). Consequently, neither the average number of copy alterations (ANCA) of primary adenomas without recurrence compared to primary adenomas with recurrence did show any difference (means 3.3 ± 3.2 SD vs 3.4 ± 4.3 SD; P=0.955, Figure 31B), nor did the percentage of changed genome per sample, which reflects the length of respective chromosomal aberrations (means 4.8 ± 6.7 SD vs 4.2 ± 5.8 SD; P=0.748, Figure 31C). Although the alteration patterns shared an immense proportion of chromosomal aberrations across colorectal adenomas without and with recurrence (Figure 31A),
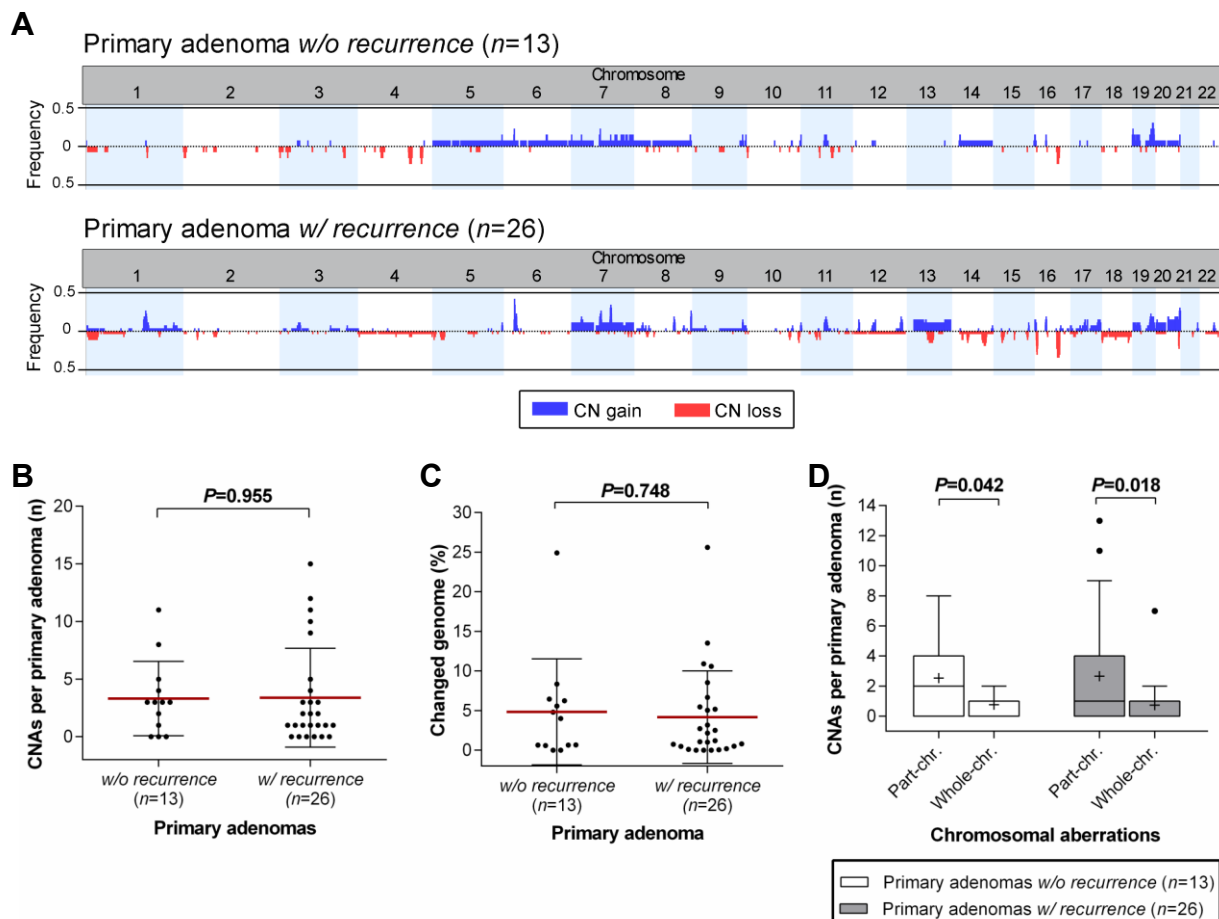
Figure 31. Chromosomal aberrations in primary colorectal adenomas without and with recurrence. (**A**) Frequency plot of CN gains (*blue*) and losses (*red*) of autosomal chromosomes distributed among primary adenomas without recurrence (upper panel) and with recurrence (lower panel). Despite visible differences on chr6, chr13, chr18 and chr20, none of the alteration patterns reached statistical significance (P>0.05). (**B**) Primary adenomas without recurrence and primary adenomas with recurrence did not show a difference in the number of copy alterations expressed by the ANCA index (mean ± SD; Student's *t* test). (**C**) Percentage of changed genome (mean ± SD) failed to distinguish primary adenomas without recurrence from the ones with recurrence (Student's *t* test). (**D**) Part-chromosome aberrations were more frequently observed than whole-chromosome aberrations (Mann-Whitney *U* test). +, mean; whiskers, Tukey.

characteristic imbalances of the colorectal carcinogenesis were identified in both groups, *e.g.*, copy number gains of chromosome 7 and chromosome 20. Of note, none of the chromosomal aberration loci reached statistical significance to distinguish primary adenomas with recurrence from adenomas without recurrence (Table 22). However, copy number events affecting chromosome 13 showed a trend to distinguish primary adenomas with recurrence with 23.1% (6/26) from non-recurring primary adenomas with 0% (0/13; *P*=0.081, Fisher's exact test).

Interestingly, the number of whole-chromosome CNAs was lower than the number of observed part-chromosome aberrations per sample in both groups: primary adenomas without (means 0.8 ± 1.4 vs 2.5 ± 2.5) and with recurrence (means 0.7 ± 1.5 vs 2.7 ± 3.6), respectively (Figure 31D). However, the comparison was more significant in primary adenomas with recurrence (*P*=0.018, Student's paired *t* test).

Table 22. Differential CNAs in primary adenomas without and with recurrence. Genome wide differential copy number alterations occurring in 2 or more cases in any of the groups.

Cytoband	Event	Primary adenoma *w/o recurrence*	Trend	Primary adenoma *w/ recurrence*	*P* value°
1pter-p35.2	CN loss	7.7% (1/13)	<	11.5% (3/26)	1.000
3p21.31-p21.1	CN gain	15.4% (2/13)	>	3.8% (1/26)	0.253
3q26.1-q26.2	CN loss	15.4% (2/13)	>	0% (0/26)	0.105
4q28.3	CN loss	15.4% (2/13)	>	7.7% (2/26)	0.589
5p14.3-p14.1	CN loss	0% (0/13)	<	7.7% (2/26)	0.544
6p22.1-p21.33	CN gain	30.8% (4/13)	<	42.3% (11/26)	0.728
chr7	CN gain	23.1% (3/13)	>	11.5% (3/26)	0.381
7q22.1	CN gain	38.5% (5/13)	>	30.8% (8/26)	0.725
chr8	CN gain	15.4% (2/13)	>	3.8% (1/26)	0.253
chr13	CN gain	0% (0/13)	<	15.4% (4/26)	0.281
13q21.1-q21.33	CN loss	0% (0/13)	<	7.7% (2/26)	0.544
14q21.1-q21.3	CN loss	0% (0/13)	<	7.7% (2/26)	0.544
14q24.3-q31.3	CN loss	0% (0/13)	<	7.7% (2/26)	0.544
16q21	CN loss	23.1% (3/13)	<	30.8% (8/26)	0.719
17p	CN gain	0% (0/13)	<	7.7% (2/26)	0.544
17p13.1-p12	CN loss	0% (0/13)	<	7.7% (2/26)	0.544
17q12-q21.32	CN gain	7.7% (1/13)	<	15.4% (4/26)	0.648
chr18	CN loss	0% (0/13)	<	7.7% (2/26)	0.544
18q12.2-q12.3	CN loss	0% (0/13)	<	11.5% (3/26)	0.538
18q22.3	CN loss	0% (0/13)	<	11.5% (3/26)	0.538
19pter-p13.11	CN gain	15.4% (2/13)	>	3.8% (1/26)	0.253
19q	CN gain	15.4% (2/13)	>	7.7% (2/26)	0.589
20p	CN gain	7.7% (1/13)	<	11.5% (3/26)	1.000
20q	CN gain	7.7% (1/13)	<	19.2% (5/26)	0.643

°, Fisher's exact test.

Furthermore, genome-wide aCGH profiles of primary adenomas were correlated with the clinicopathological features of these samples. Copy number changes correlated with the dichotomized localization, *i.e.*, colon or rectum, of the primary adenomas (*P*=0.040; Figure 32A), as seen in the complete sample-set (Figure 29B). Consequently, at least one CNA was detected in each rectal adenoma analyzed by aCGH. Subsequently, numbers of CNAs per sample were tested for association with the dichotomized localization which revealed a strong trend (*P*=0.052; Figure 32B). Surprisingly, primary adenomas of male patients harbored a more significant load of CNAs than their female counterpart (*P*=0.029; Figure 32C). None of the other clinicopathological parameters (adenoma size, histology, grading, time of recurrence, patient age) correlated significantly.
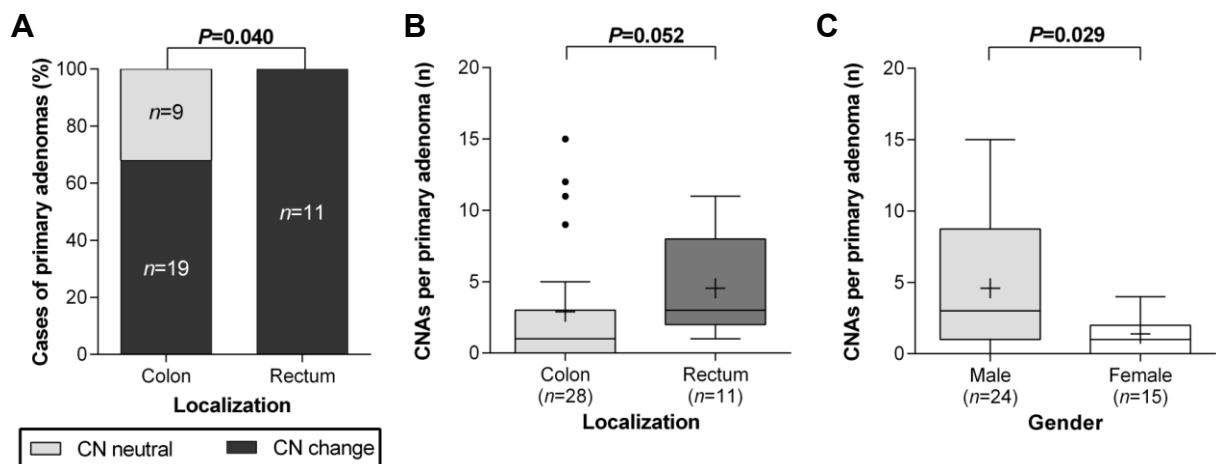
Figure 32. CNAs detected by aCGH correlate with the localization of the primary adenomas. (**A**) CN changes/neutrality in primary adenomas correlated with the dichotomized localization of either the colon or the rectum (Fisher's exact test). (**B**) Numbers of CNAs per sample in primary adenomas were almost significantly different when separated by dichotomized localization (Mann Whitney *U* test). (**C**) Primary adenomas of male patients exhibited a larger number of CNAs (Mann Whitney *U* test). +, mean; whiskers, Tukey.

3.2.4 Hierarchical clustering of colorectal adenoma CNA-profiles

Unsupervised hierarchical clustering of aCGH-detected CNA-profiles expressed as chromosome arm-lengths separated the 51 samples into three distinct groups (Figure 33). One of the groups (Cluster 1) contained samples with either several chromosome-arm losses, *e.g.*, chromosome 18 or whole-chromosome gains of chromosome 7 and 8, respectively. The second group (Cluster 2) was defined by other copy number alterations, *i.a.*, whole-chromosome gains of 13 or part-chromosome aberrations on chromosome arms 11q or 19p (Figure 33). The third group (Cluster 3) was composed of adenoma samples which either showed focal aberrations only or samples without any detectable CNA at all. Thus, the number of copy alterations per sample was strongly

associated with the unsupervised clustering of the adenoma samples into Cluster 1 (mean 6.5 ± 4.8), Cluster 2 (mean 3.7 ± 3.0) and Cluster 3 (mean 1.7 ± 3.3) ($P<0.001$; Figure 34A). Cluster-based groups were further evaluated by consideration of the status of recurrence. However, cluster assignment did not discern primary adenomas without recurrence from primary adenomas with recurrence ($P=0.449$; Freeman-Halton test). Additionally, recurrent adenomas were not associated with a distinct cluster separating these samples from primary adenomas ($P=0.653$, Freeman-Halton test; Figure *33* 33).
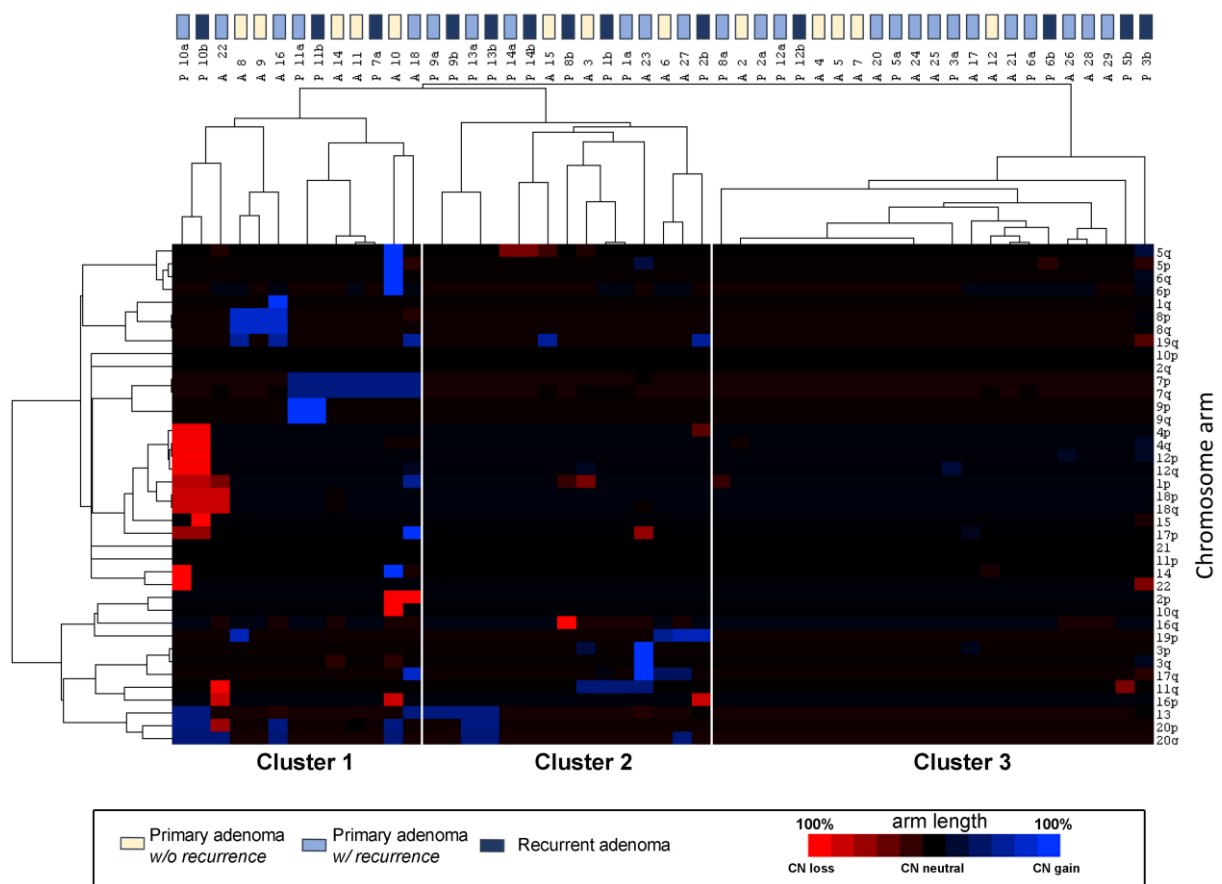
Figure 33. Heat map of CNAs in colorectal adenomas detected by aCGH. Heat map cluster analysis of 51 colorectal adenoma specimens including primary adenomas without recurrence ($n=13$), primary adenomas with recurrence ($n=26$) and recurrent adenomas ($n=13$). Unsupervised hierarchical clustering was performed by calculating the bp-length of CNAs detected by aCGH expressed as a percentage of the affected chromosome arm p and q, respectively. *Blue*, *red* and *black* show CN gain, loss, and neutrality. Adenoma cases are plotted along the *x-axis* with chromosome arms plotted along the y-axis. However, acrocentric chromosomes (*i.e.*, chr13, chr14, chr15, chr21 and chr22) are represented by the q-arm only. The heat map is divided into three well-defined clustering groups. However, these groups were not associated with the status of recurrence but with the number of CNAs per sample.

Aiming to identify putative additional correlations of the cluster assignment, besides the number of CNAs per sample (Figure 34A), other clinicopathologic features of the cluster groups were evaluated. The sample distribution suggested by the heat map clustering failed short of a trend to be correlated with dichotomized lesion size separating the tumors at the cohort's median of 20 mm (IQR, 15-35 mm; *P*=0.108; Figure 34B). Cluster group 3 consisted of more than 80% of adenomas which exceeded the median adenoma size of 20 mm (Figure 34B). More striking, cluster groups correlated negatively with dichotomized patient age considering the cohort's median of 67.3 years (IQR, 61.9-72.7 years; *P*=0.027; Figure 34C). Smaller lesions tended to have fewer CNAs than larger lesions. While Cluster 1 was dominated by samples above the median of 67.3 years, Cluster 3 was dominated by samples from younger patients, *i.e.*, below the median. Apart from that none of the other tested clinicopathological features (adenoma histology, gender, grading, time of recurrence) demonstrated a correlation with the heat map cluster assignment.
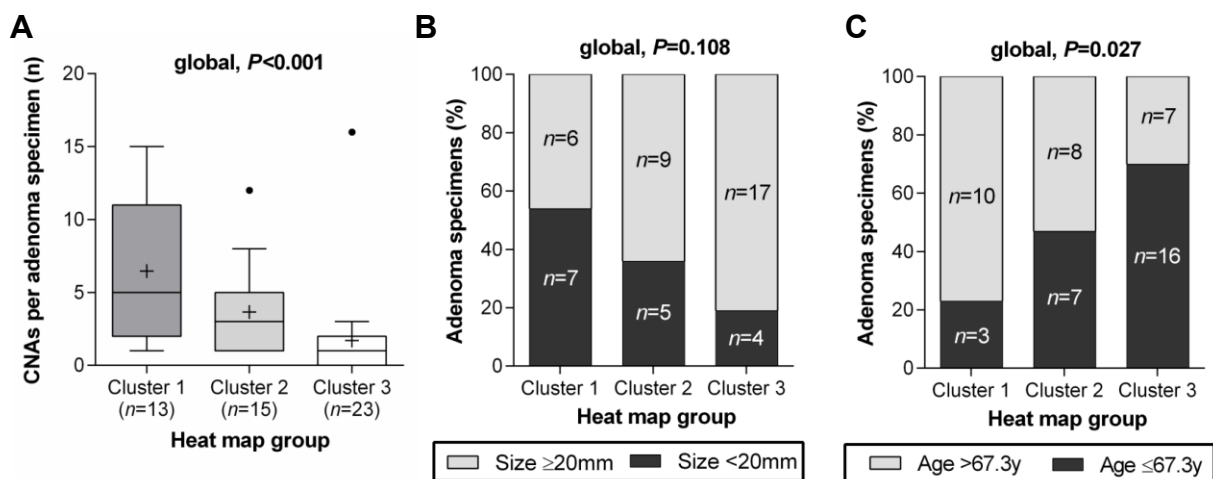
Figure 34. Heat map cluster group correlations. (**A**) Heat map cluster adenoma groups were most significantly associated with the number of CNAs per sample, with Cluster 1 exhibiting the largest numbers of CNAs (Kruskal-Wallis test). +, mean; whiskers, Tukey. (**B**) Heat map cluster groups failed short of a trend being correlated with the dichotomized size of the adenomatous lesions (Freeman-Halton test). (**C**) Distribution of samples across the cluster groups was associated with dichotomized patient age (Freeman-Halton test). While cluster 1 was dominated by samples equaling or exceeding 67.3 years, *vice versa*, cluster group 3 was dominated by samples of patients below 67.3 years (y).

3.2.5 Comparison of CN profiles of match-paired colorectal adenomas

Matched adenoma pairs, consisting of the primary tumor and the corresponding recurrent tumor from the same patient, were compared between both samples by arm-level copy number aberration profiles identified via aCGH (Figure 35A). Among analyzed adenoma pairs, 38.5% (5/13) showed an identical copy number profile in the primary tumor and the corresponding recurrent tumor (*P*=1.000; Figure 35A). 23.1% (3/13) of

adenoma pairs differed, yet not statistically significant, genome-wide in a single CNA in the primary and the recurrent tumor (Figure 35A). Interestingly, the different aberration patterns comprised an additional focal CNA in the recurrent adenoma of pair P1 and P5, respectively. Pair P8 showed an arm-level CN loss of 80% compared to the corresponding lesion, whereas P10 exhibited more changes (Figure 35A). However, the number of arm-level CNAs per patient varied drastically across these paired samples (range, 0-14) implicating the presence of intertumor heterogeneity. Remarkably, 46.2% (P1, P5, P6, P8, P12, P14) of the pairs harbored none or only a few focal CNAs, while 30.8% (P9, P10, P11, P13) of the pairs were affected by various focal and arm-level CNAs. Based on the detected CN profiles, ten adenoma pairs suggested none to a shallow level of variance between the primary tumor and the corresponding recurrent tumor. CN profiles were classified as either "stable" CN patterns in 53.8% (7/13) of cases, or as "simplification/complexity" patterns (if the patterns were more divergent) (Figure 35B). The most recurrent alterations (n>2) were the CN gains of 7q22 (30%) and 13q (30%), respectively.

Although most match-paired cases exhibited a substantial overlap of the aCGH CN profiles across the primary and the recurrent tumor, 46.2% (6/13) of adenoma pairs revealed an acquisition of CNAs in the recurrent lesion (apart from P6 and P10): the simplification/complexity pattern (Figure 35B). Even though no CNA was detected in the primary adenoma of pair P2, the recurrent lesion harbored three focal aberrations and a gain of chromosome 19. Statistically, the comparison of CNAs of both lesions showed a trend for an accumulation of CNAs in the recurrent tumor (P=0.078; Figure 35A). An even more significant increase in the number of CNAs was observed in the recurrent adenomas of pair P3 (P=0.006) and P7 (P<0.001), respectively (Figure 35A). Herein, pair P3 exhibited a significant acquisition of additional focal CNAs which indicate an increase of the genomic instability from the primary tumor to the recurrent adenoma. More striking, adenoma pair P7 exposed a primary lesion with a single gain of chr7 while the recurrent tumor was dominated by ten chromosome arm-level losses and four arm-level gains (Figure 35A).

An association of any pattern with a longer time interval of adenoma recurrence was not observed (P=0.231; Figure 35C).

A
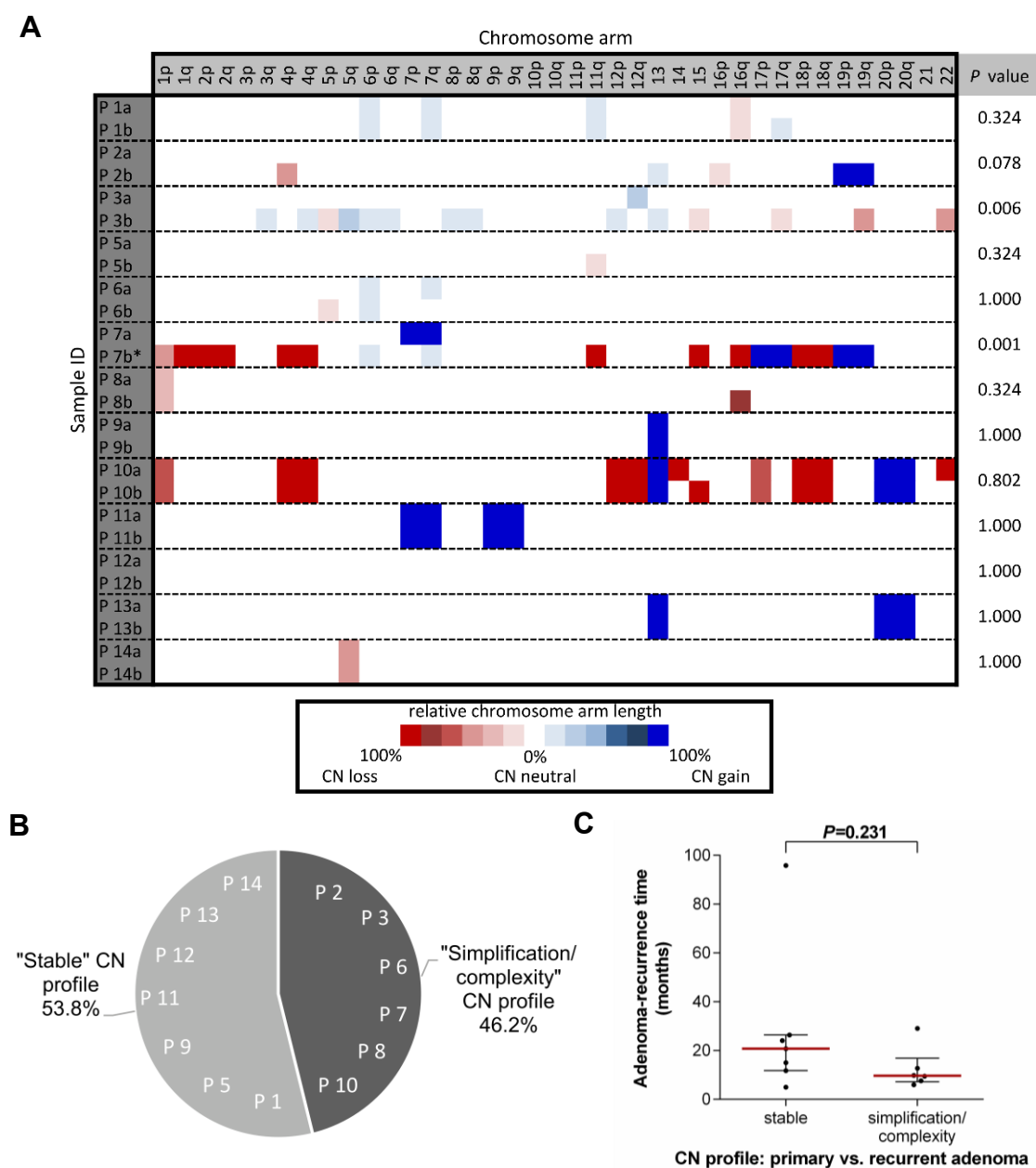
Figure 35. Comparison of chromosomal aberrations in matched adenoma pairs. (**A**) Matched adenoma pairs (consisting of the primary adenoma and the corresponding recurrent adenoma) are plotted by the bp-length of CNAs detected by aCGH. CNAs are expressed as a percentage of the affected chromosome arm p and q, respectively. Neutral copy numbers were annotated as 0%. Focal alterations were expressed as ≤10%. More extensive alterations are plotted in accordance to the respective bp-length in steps of 20%, *i.e.*, a whole-chromosome arm aberration equaled 100%. Acrocentric chromosomes are represented by the q-arm only. *P* value, student's paired *t* test**. (B**) Distribution of matched pair adenomas by CN profiles comparing primary adenomas to the corresponding recurrent adenoma. (**C**) There was no correlation observed between the time to recur and the CNA profiles detected via aCGH in the matched pair adenoma samples (median with IQR; Mann-Whitney *U* test).

3.3 Single-cell analyses of CNAs in colorectal (recurring) adenomas

3.3.1 Distribution of clinicopathological features

Clinicopathological parameters were collected for all colorectal adenoma samples (Table 7). The two-different recurrence groups of primary adenomas were matched in terms of the histopathological and clinical features (Table 23).

Table 23. Clinicopathological features of primary adenomas without and with recurrence.

Variable	Primary adenomas w/o recurrence (n=15)	Primary adenomas w/ recurrence (n=29)	P value
Age at diagnosis (y)			
Mean ± SD	65.0 ± 6.3	66.6 ± 11.7	0.633[a]
Median (IQR)	66.7 (63.9-68.5)	68.2 (57.4-72.9)	
Sex			
Female	4	14	0.208[b]
Male	11	15	
Localization			
Right hemicolon	9	14	0.774[c]
Left hemicolon	3	6	
Rectum	3	9	
Histology			
Tubular	5	3	0.099[b]
Tubulo-villous / Villous	10	26	
Size (mm)			
Mean ± SD	22.8 ± 9.7	32.5 ± 20.2	0.086[a]
Median (IQR)	20.0 (15.0-30.0)	27.5 (20.0-40.0)	
Observation time/ recurrence-free time (m)			
Mean ± SD	25.4 ± 14.4	21.7 ± 19.5	0.528[a]
Median (IQR)	21.9 (12.5-38.0)	17.2 (8.2-25.7)	

[a] Student's t test; [b] Fisher's exact test; [c] Freeman-Halton test

3.3.2 Chromosomal instability in colorectal adenomas

Fifty-eight colorectal adenoma samples were sequentially hybridized with three miFISH probe panels which allowed the simultaneous enumeration of fifteen probes per individual nucleus of the same specimen (Figure 15; Figure 36). Signals were manually counted/enumerated for on average 523 nuclei per case (range, 326-1164 nuclei) including nuclei with two signals for all probes. Ploidy estimation was inferred based on CEP10 and on the average signal number of all probes, which were balanced regarding tumor suppressor genes and oncogenes.

The enumeration of clonal aberration patterns was performed on a cell-to-cell basis. Each cell fitting a respective pattern had the identical signal count for all fifteen probe markers. These signal pattern clones were visualized by depicting each cell of the adenoma according to its copy number status of gain, loss or neutrality, respectively.

The clonal fractions are compared within each adenoma and between the individual specimens. Calculating the frequency of gains and losses for each probe marker in percentages of the total cell population aided to investigate recurrent patterns of CNAs (Figure 37; Figure 38-44).
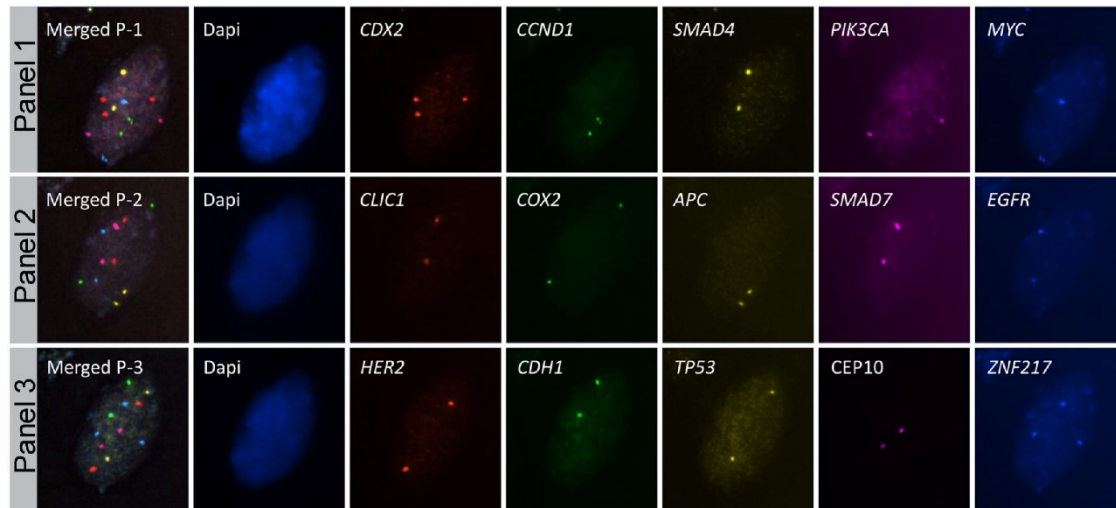
Figure 36. Representative image of a colorectal adenoma nucleus hybridized by miFISH. The image shows a nucleus which was consecutively hybridized with panel 1 (*top row*), panel 2 (*middle row*) and panel 3 (*bottom row*). The merged image in each row displays the overlay of all respective channels. Gene probe names are annotated to each box. The nucleus presented here reveals a copy number gain for *CDX2* and *ZNF217* (signal count: *three*) while the other markers remained neutral (signal count: *two*).
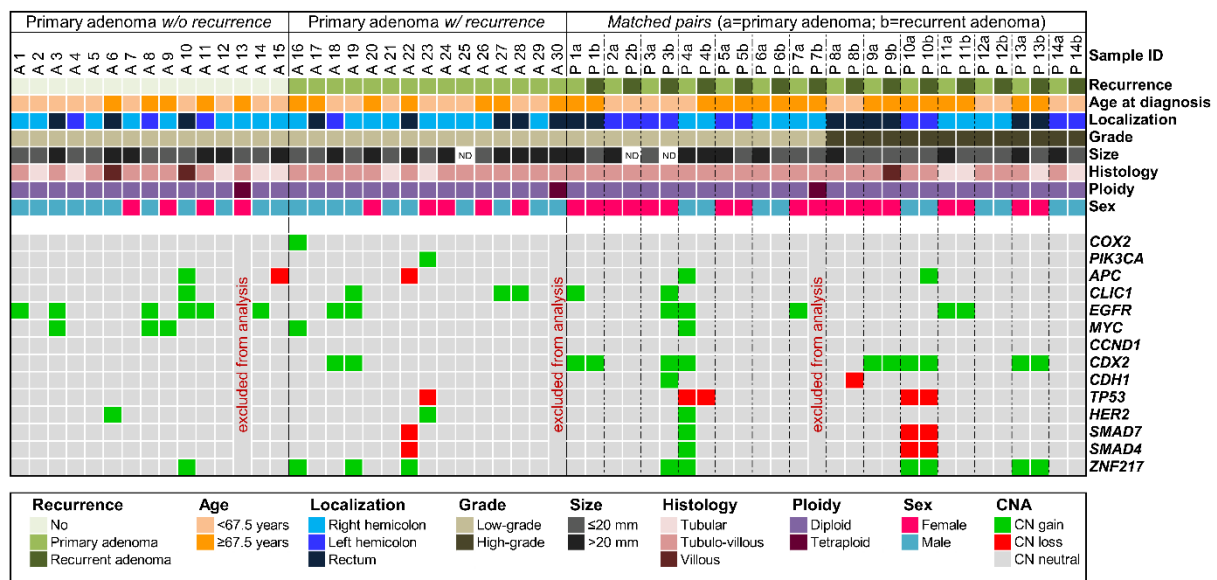
Figure 37. CNAs in colorectal adenomas identified by miFISH. Clinicopathological features (*upper columns*) and CNA status (*lower columns*) are plotted per individual adenoma. Polyploid adenomas were excluded from analysis due to the increased degree of genomic instability. Age at diagnosis and lesion size were dichotomized by the median considering the sample cohort, respectively. ND, not determined.
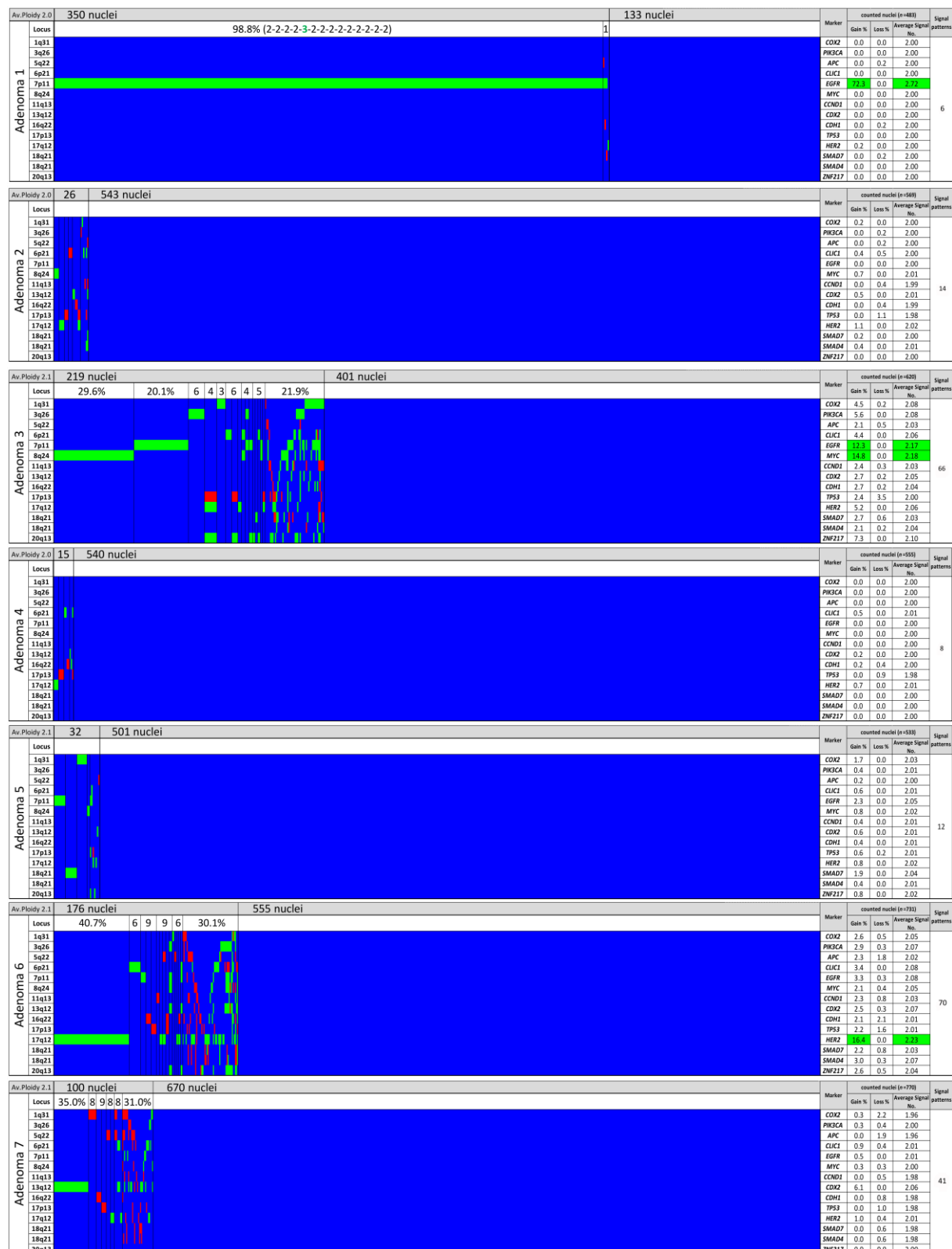
Figure 38. Summary of color displays of signal pattern clones in primary colorectal adenomas without recurrence (*n*=15). Depicted are the adenoma cases without recurrence (A1-A15). *Green*, CN gain; *red*, CN loss; *blue*, CN neutral. Nuclei are sorted horizontally according to their signal pattern from left to right by frequency and with the gene probes assorted vertically by chromosomal location from top to bottom. Vertical lines separate the clonal fractions whose prevalence of the aberrant clone population are noted in percentages. CN gains and losses are calculated as percentages of the total population. When applicable the pattern of the most abundant clone is annotated in brackets. Aberrant nuclei and nuclei without detected CNAs are displayed as separated clonal fractions. Average signal numbers, the number of observed signal patterns and the average ploidy are shown in the respective columns. Noteworthy, case A13 displayed near-tetraploid (4N) signal counts compared to the other cases (2N).
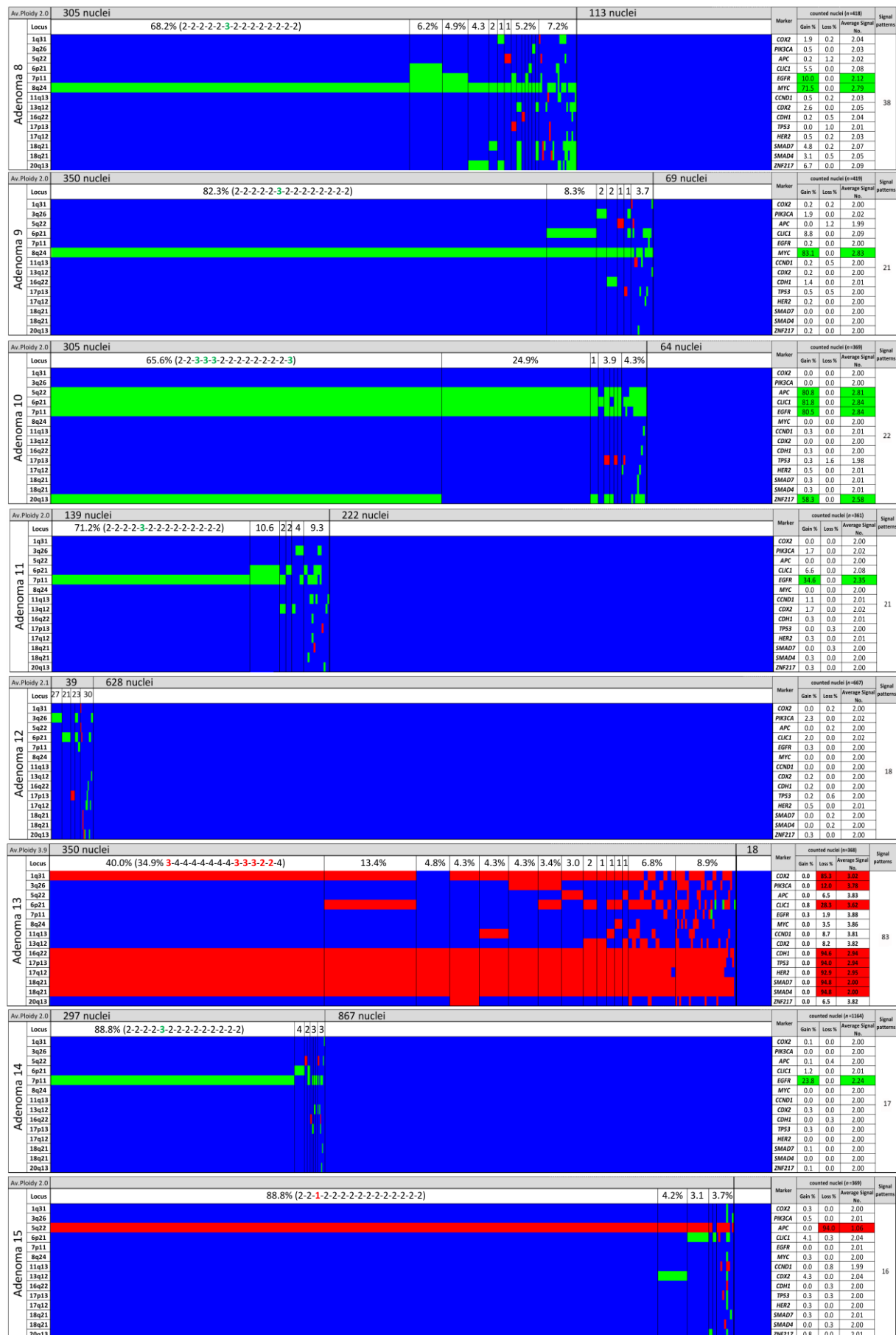
Figure 38. (continued).
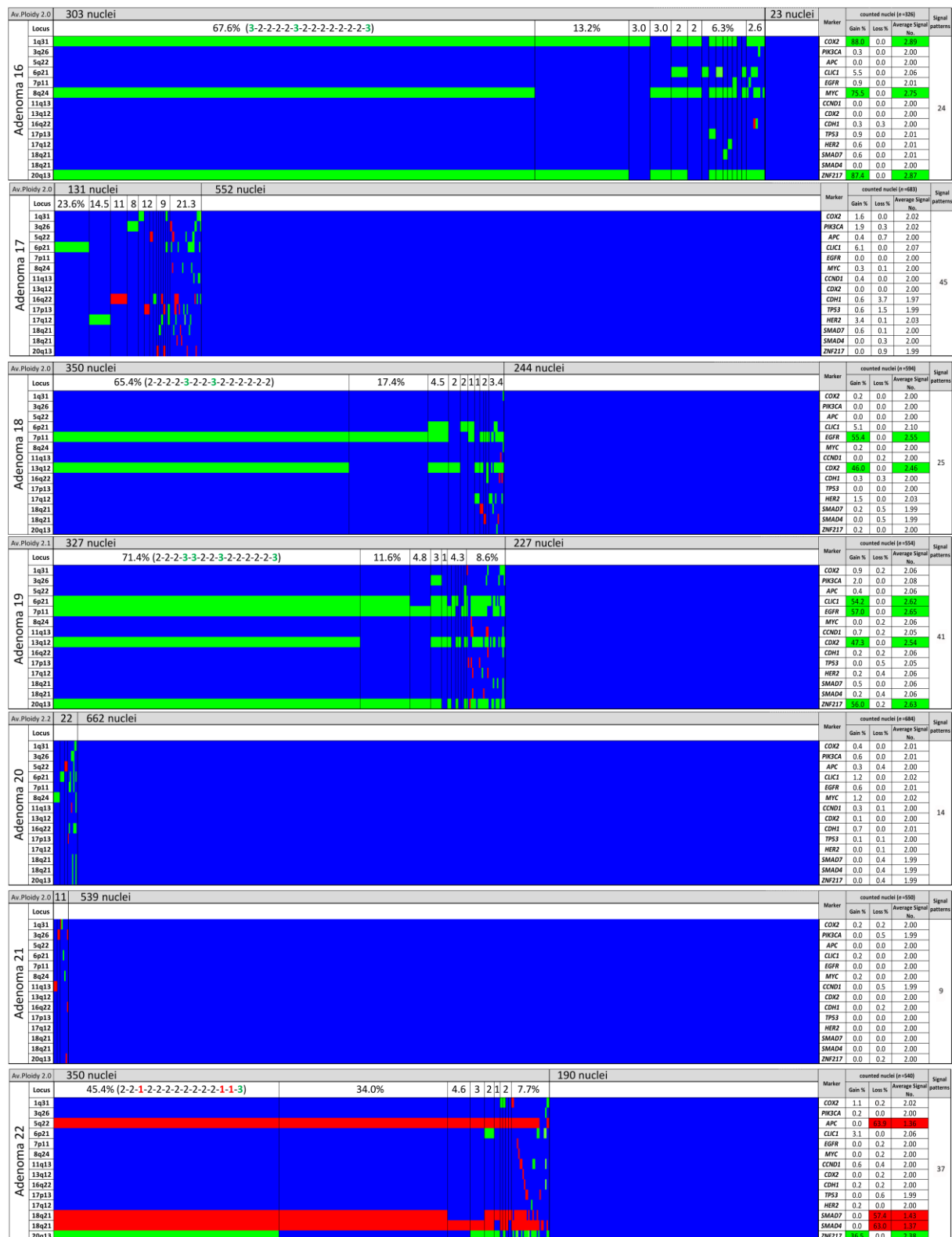
Figure 39. Summary of color displays of signal pattern clones in primary colorectal adenomas with recurrence (*n*=15). Depicted are the adenoma cases without recurrence (A16-A30). *Green*, CN gain; *red*, CN loss; *blue*, CN neutral. Nuclei are sorted horizontally according to their signal pattern from left to right by frequency and with the gene probes assorted vertically by chromosomal location from top to bottom. Vertical lines separate the clonal fractions whose prevalence of the aberrant clone population are noted in percentages. CN gains and losses are calculated as percentages of the total population. When applicable the pattern of the most abundant clone is annotated in brackets. Aberrant nuclei and nuclei without detected CNAs are displayed as separated clonal fractions. Average signal numbers, the number of observed signal patterns and the average ploidy are shown in the respective columns. Noteworthy, case A30 displayed near-tetraploid (4N) signal counts compared to the other cases (2N).

100
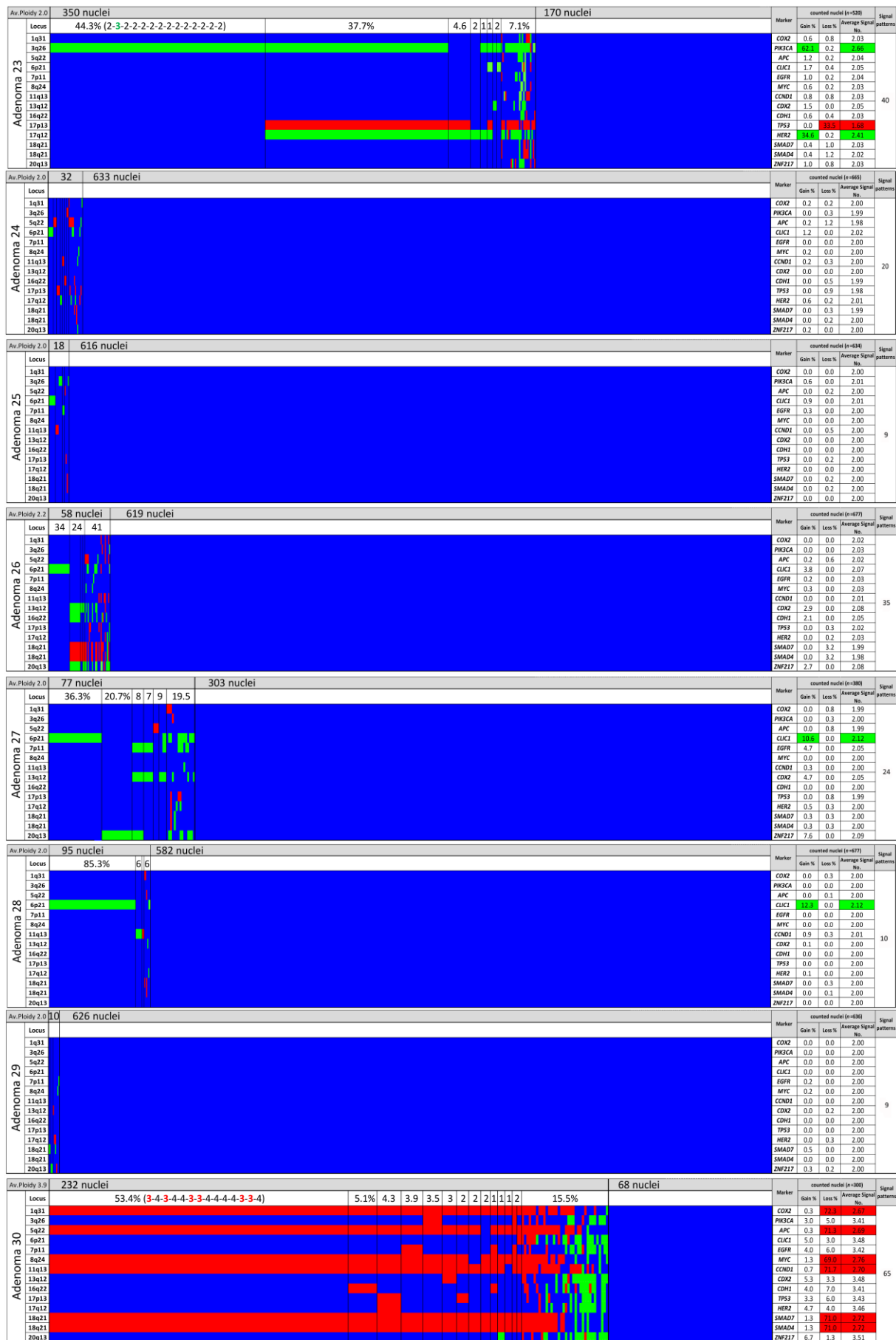
Figure 39. (continued).

A
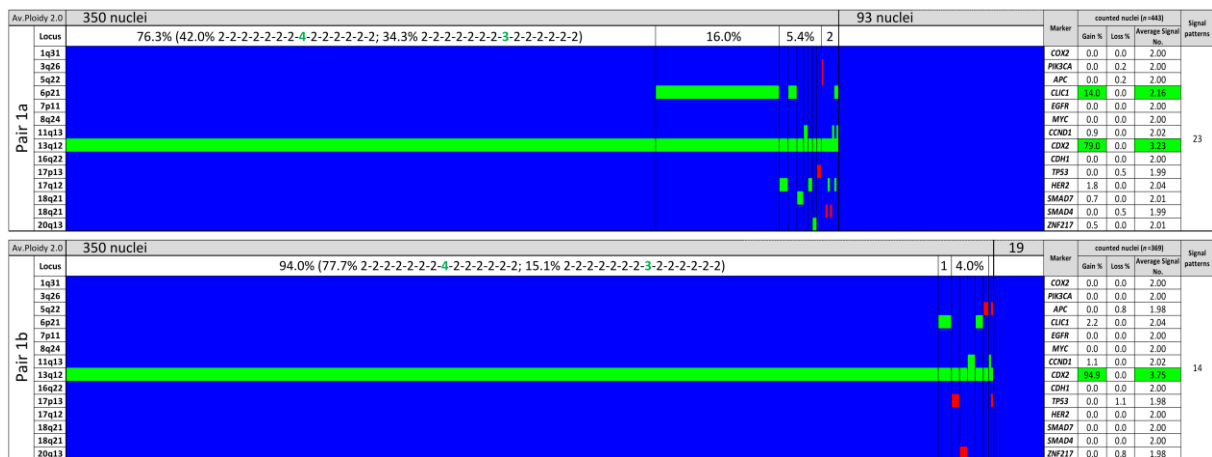
B

C
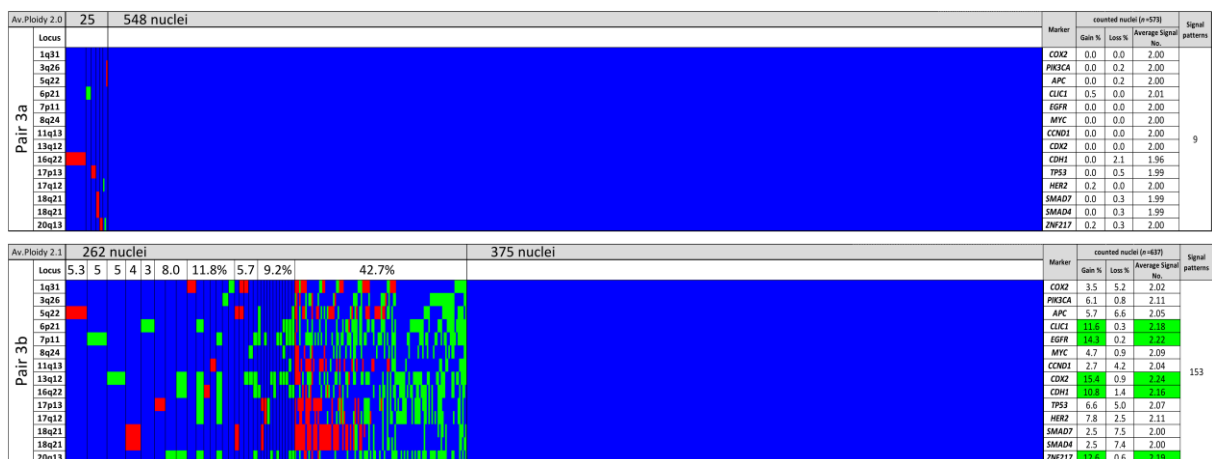
Figure 40. Summary of color displays of signal pattern clones in match-paired colorectal adenomas (Part I). Depicted are the adenoma pairs P1 (**A**), P2 (**B**), and P3 (**C**) consisting of the primary tumor (a) and the corresponding recurrent (b) lesion. *Green*, CN gain; *red*, CN loss; *blue*, CN neutral. Nuclei are sorted horizontally according to their signal pattern from left to right by frequency and with the gene probes assorted vertically by chromosomal location from top to bottom. Vertical lines separate the clonal fractions whose prevalence of the aberrant clone population are noted in percentages. CN gains and losses are calculated as percentages of the total population. When applicable the pattern of the largest clone is annotated in brackets. Aberrant nuclei and nuclei without detected CNAs are displayed as separated clonal fractions. Average signal numbers, the number of observed signal patterns and the average ploidy are shown in the respective columns. The clonal distribution of the primary and the recurrent adenoma of pair 1 and pair 2, respectively, were similar. Pair 3 showed an accumulation of clones in the recurrent lesion.
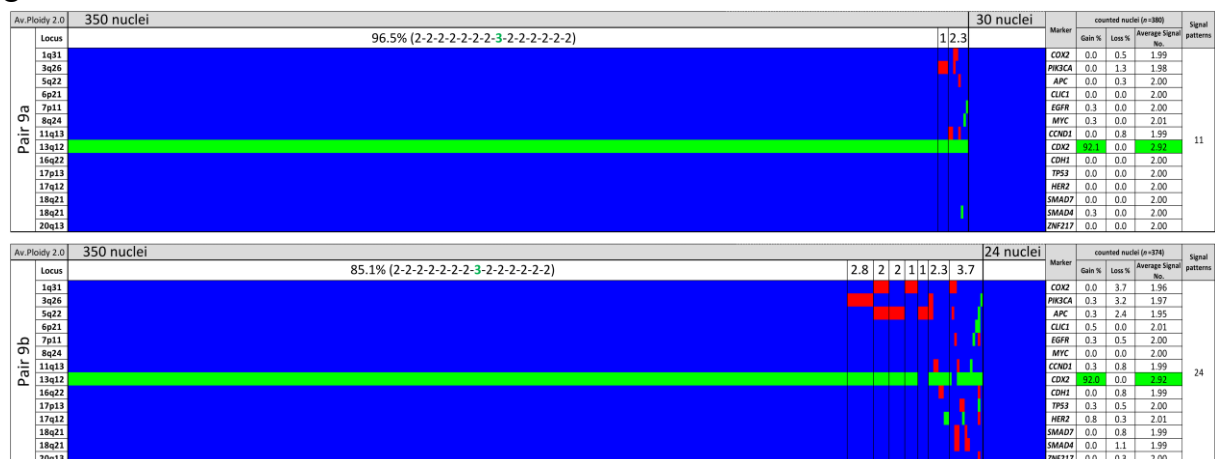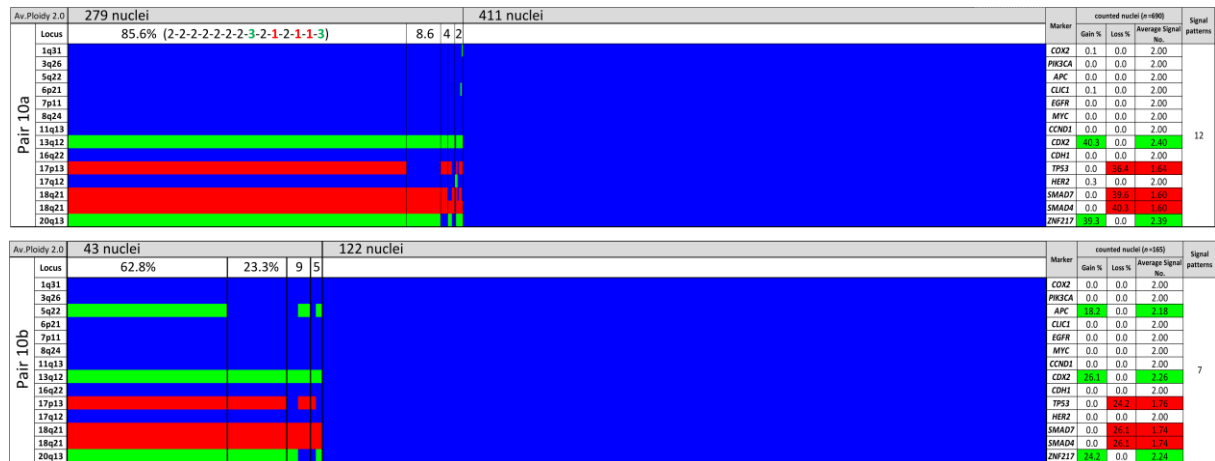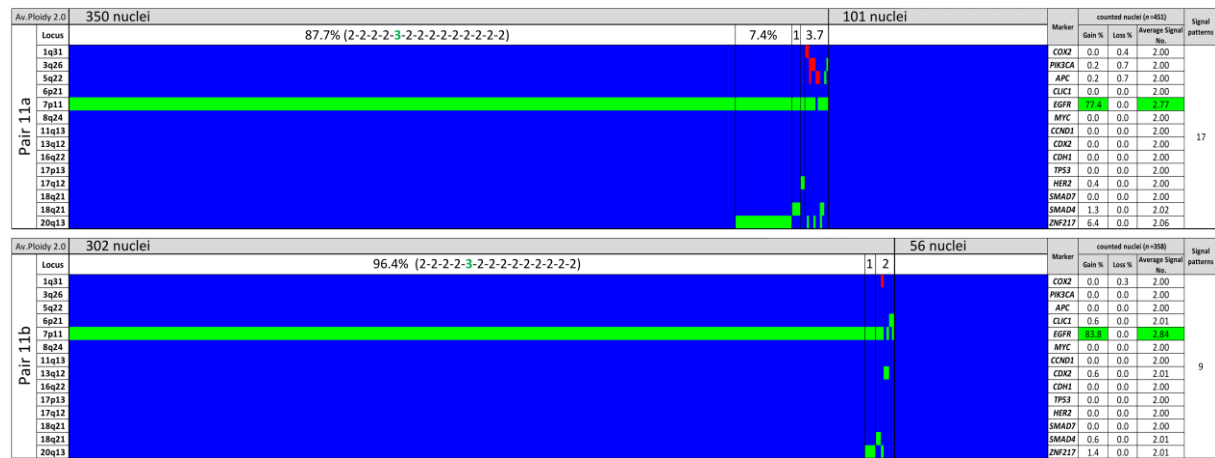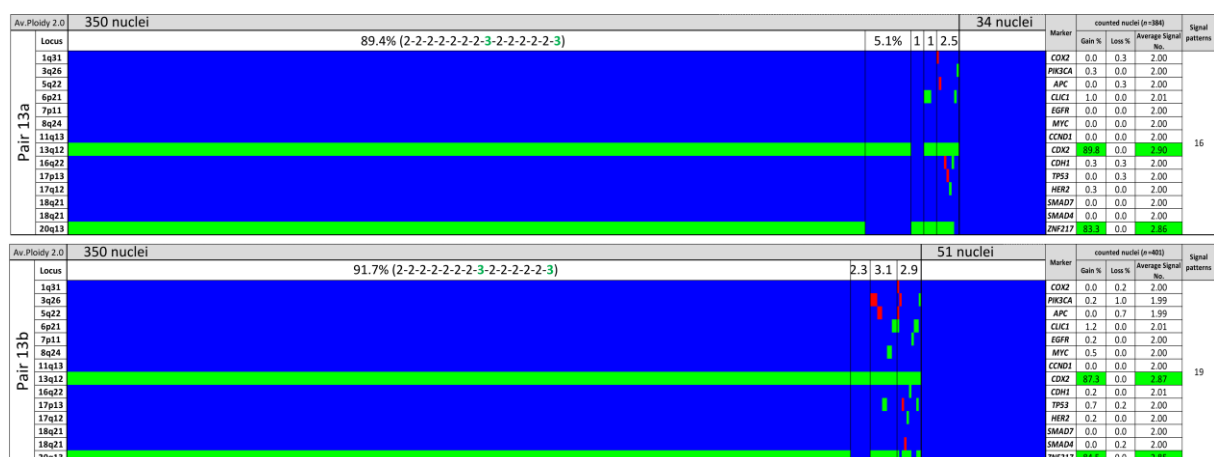
A
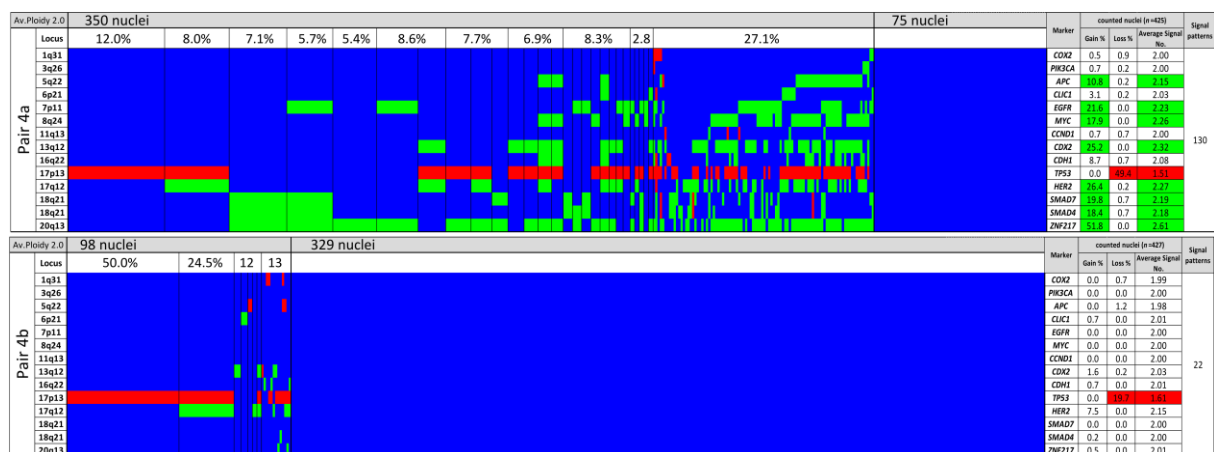
B

C

Figure 41. Summary of color displays of signal pattern clones in match-paired colorectal adenomas (Part II). Depicted are the adenoma pairs P4 (**A**), P5 (**B**), and P6 (**C**) consisting of the primary tumor (a) and the corresponding recurrent (b) lesion. *Green*, CN gain; *red*, CN loss; *blue*, CN neutral. Nuclei are sorted horizontally according to their signal pattern from left to right by frequency and with the gene probes assorted vertically by chromosomal location from top to bottom. Vertical lines separate the clonal fractions whose prevalence of the aberrant clone population are noted in percentages. CN gains and losses are calculated as percentages of the total population. If applicable the pattern of the most abundant clone is annotated in brackets. Aberrant nuclei and nuclei without detected CNAs are displayed as separated clonal fractions. Average signal numbers, the number of observed signal patterns and the average ploidy are shown in the respective columns. Pair 4 exhibited a decrease of clone fractions from the primary towards the recurrent adenoma. The clonal distribution of the primary and the recurrent adenoma of pair 5 and pair 6, respectively, were similar.

A
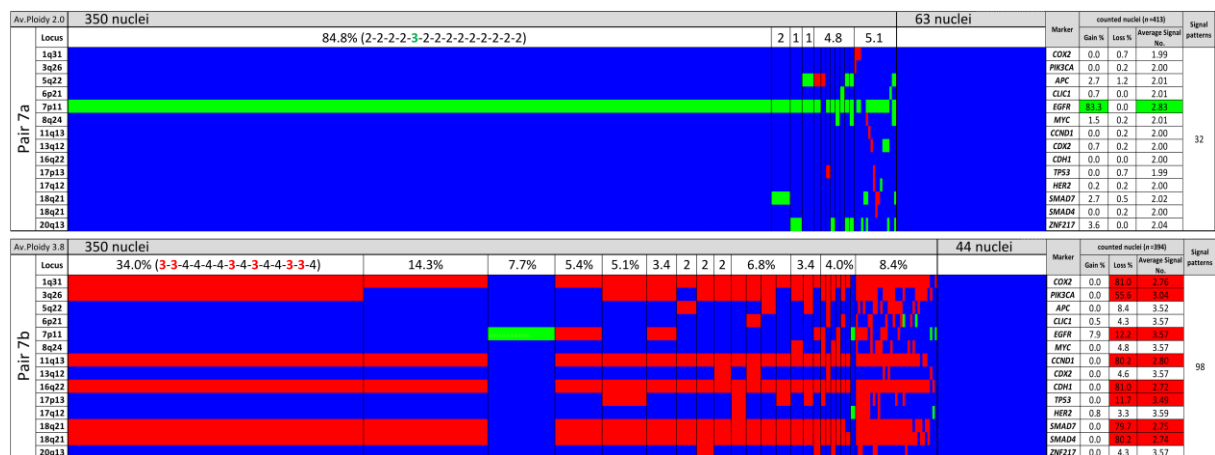
B

C

Figure 42. Summary of color displays of signal pattern clones in match-paired colorectal adenomas (Part III). Depicted are the adenoma pairs P7 (**A**), P8 (**B**), and P9 (**C**) consisting of the primary tumor (a) and the corresponding recurrent (b) lesion. *Green*, CN gain; *red*, CN loss; *blue*, CN neutral. Nuclei are sorted horizontally according to their signal pattern from left to right by frequency and with the gene probes assorted vertically by chromosomal location from top to bottom. Vertical lines separate the clonal fractions whose prevalence of the aberrant clone population are noted in percentages. CN gains and losses are calculated as percentages of the total population. When applicable the pattern of the most abundant clone is annotated in brackets. Aberrant nuclei and nuclei without detected CNAs are displayed as separated clonal fractions. Average signal numbers, the number of observed signal patterns and the average ploidy are shown in the respective columns. Pair 7 showed a dramatic increase in CNAs from the primary towards the recurrent adenoma. Pair 8 exhibited an additional clonal loss of *CDH1* in the recurrent adenoma. Pair 9 demonstrated a very similar pattern across primary and recurrent adenoma.

A

B

C

Figure 43. Summary of color displays of signal pattern clones in match-paired colorectal adenomas (Part IV). Depicted are the adenoma pairs P10 (**A**), P11 (**B**), and P12 (**C**) consisting of the primary tumor (a) and the corresponding recurrent (b) lesion. *Green*, CN gain; *red*, CN loss; *blue*, CN neutral. Nuclei are sorted horizontally according to their signal pattern from left to right by frequency and with the gene probes assorted vertically by chromosomal location from top to bottom. Vertical lines separate the clonal fractions whose prevalence of the aberrant clone population are noted in percentages. CN gains and losses are calculated as percentages of the total population. When applicable the pattern of the most abundant clone is annotated in brackets. Aberrant nuclei and nuclei without detected CNAs are displayed as separated clonal fractions. Average signal numbers, the number of observed signal patterns and the average ploidy are shown in the respective columns. Pair 10 showed stable clonal fractions across the primary and the recurrent adenoma, which slightly progressed with an additional CNA of *APC.* Pair 11 and pair 12 displayed similar patterns of clonal compositions across the primary towards the recurrent adenoma, respectively.

A

B

Figure 44. Summary of color displays of signal pattern clones in match-paired colorectal adenomas (Part V). Depicted are the adenoma pairs P13 (**A**) and P14 (**B**) consisting of the primary tumor (a) and the corresponding recurrent (b) lesion. *Green*, CN gain; *red*, CN loss; *blue*, CN neutral. Nuclei are sorted horizontally according to their signal pattern from left to right by frequency and with the gene probes assorted vertically by chromosomal location from top to bottom. Vertical lines separate the clonal fractions whose prevalence of the aberrant clone population are noted in percentages. CN gains and losses are calculated as percentages of the total population. When applicable the pattern of the most abundant clone is annotated in brackets. Aberrant nuclei and nuclei without detected CNAs are displayed as separated clonal fractions. Average signal numbers, the number of observed signal patterns and the average ploidy are shown in the respective columns. Pair 13 and pair 14 displayed similar patterns of clonal compositions across the primary towards the recurrent adenoma, respectively.

3.3.3 Patterns of genomic imbalances in colorectal adenomas

Non-diploid (here: tetraploid) adenomas were excluded from overall analysis (samples A13, A30, P7b). Fifty-five adenoma samples were analyzed to reveal patterns of chromosomal gains and losses. In general, probes targeting the oncogenes *COX2* (1q), *PIK3CA* (3q), *CLIC1* (6p), *EGFR* (7p), *MYC* (8q), *CDX2* (13q), *HER2* (17q), and *ZNF217* (20q) were found to be affected by copy number gains (Figure 45A; Figure 46A). *Vice versa*, probes which targeted the tumor suppressor genes *APC* (5q), *CDH1* (16q), *TP53* (17p), *SMAD7* (18q), and *SMAD4* (18q) were rather lost than gained

(Figure 45A). The frequencies of aberrant (*i.e.,* abnormal signal count which combines both gain and loss) cells separated per miFISH markers displayed an unequal distribution in primary colorectal adenomas (Figure 45A). While CNAs most often affected the probes targeting *EGFR* (13.0%), *CDX2* (10.9%), and *ZNF217* (10.9%), followed by *APC* (6.6%), *CLIC1* (6.1%), and *MYC* (6.5%). The remaining probes were affected by CNA by less than 5% with *CCND1* (0.5%) being the probe most rarely altered in primary colorectal adenomas (Figure 45A). The comparison of abnormal signal counts in primary adenomas without recurrence to primary adenomas with recurrence revealed no significant difference (*P*=0.735, Wilcoxon signed rank test; Figure 45B).
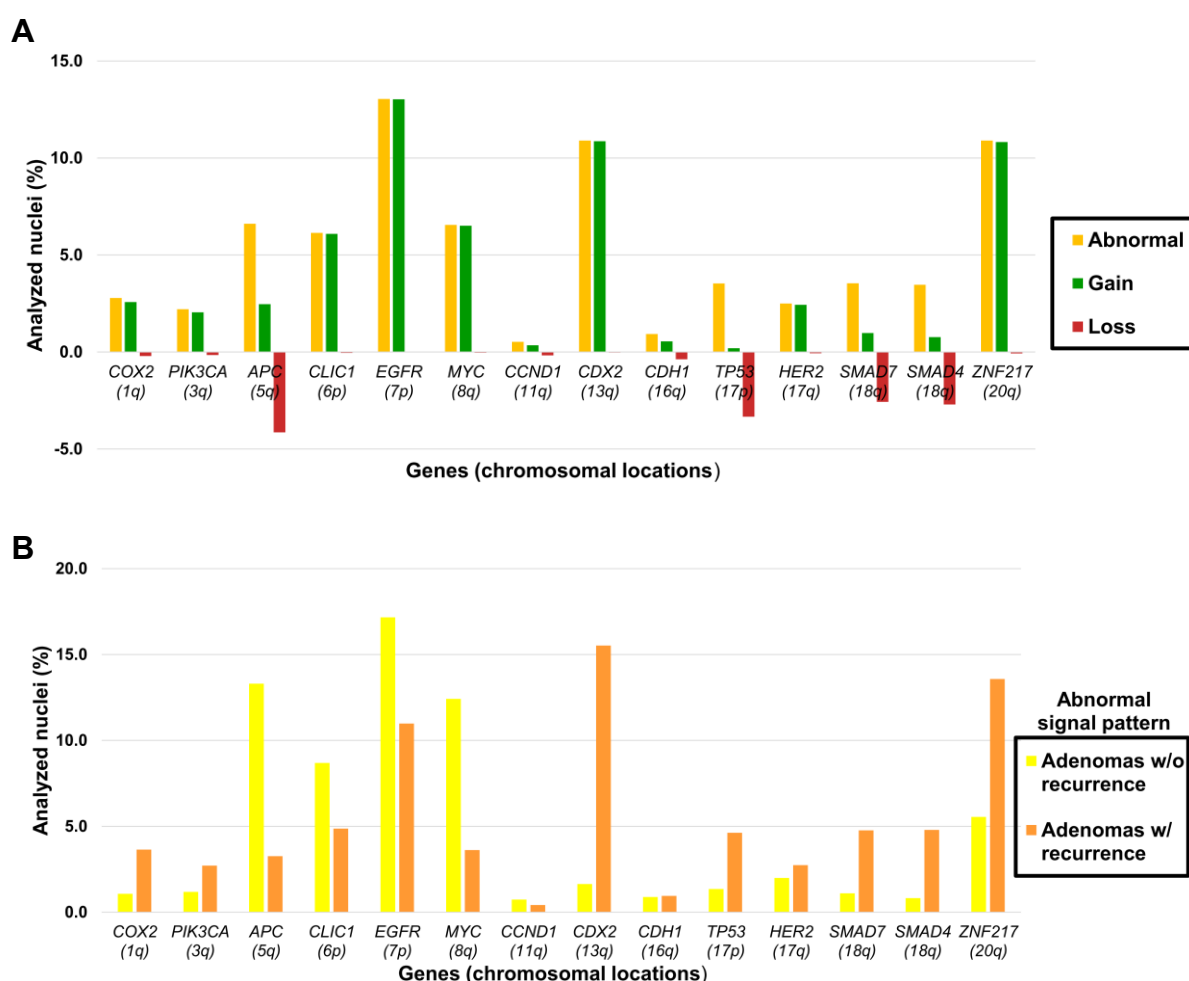
Figure 45. Average gain and loss frequencies of all miFISH gene markers. (**A**) The graph depicts the mean percentage values of aberrant cells in primary adenomas (*n*=42). Aberrant counts are expressed as gain (*green*), loss (*red*) and the combined abnormal counts (*blue*) sorted by gene markers and chromosomal location. (**B**) Abnormal signal counts (combining gain and loss) in primary adenomas without recurrence (*n*=14) compared to primary adenomas with recurrence (*n*=28). None of the analyzed probes reached statistical significance. However, the difference of the mean in abnormal nuclei counts for *CDX2* in adenomas without recurrence compared to adenomas with recurrence showed a trend (*P*=0.082, Student's *t* test).

Despite visible differences in the frequencies of abnormal miFISH markers when comparing primary adenomas without recurrence to primary adenomas with recurrence, none of the probes reached statistical significance ($P>0.05$, multiple t tests; Figure 45B). However, the frequencies of cells displaying an abnormal signal count of *CDX2* showed a trend to discriminate primary adenomas with recurrence from primary adenomas without recurrence (means 15.5 ± 21.7 vs 1.6 ± 1.5, $P=0.082$; Figure 45B).

Nevertheless, the percentage values of aberrant cells are an insufficient parameter biased by cases of adenomas formed of large dominant clone fractions. Thus, analyzing the colorectal adenoma cohort ($n=55$) regarding the average signal number (ASN) per miFISH gene marker provides a more robust, yet arguably biased by dominant clones, overview of the frequency of CNAs (Figure 46A). Here, *CDX2* revealed the highest ASN (mean 2.17 ± 0.36), followed by *EGFR* (mean 2.12 ± 0.25) and *ZNF217* (mean 2.11 ± 0.24). These ASNs demonstrate that these loci were prone to be gained (diverging from 2.00, *i.e.*, 2N). While the ASNs of *CLIC1* (mean 2.06 ± 0.13) and *MYC* (mean 2.06 ± 0.18) ranged as fourth and fifth highest, the remaining probes displayed ASNs which hardly deviated from 2.00 (Figure 46A). In general, ASNs below 2.00 were distributed in a narrow range not exceeding 1.96. The distribution barely changed when analyzing the ASNs of primary colorectal adenomas ($n=42$) only (Figure 46B). Interestingly, *EGFR* (mean 2.13 ± 0.25) and *CDX2* (mean 2.12 ± 0.28) had the highest ASNs. More important, frequencies of ASNs were plotted to separate the colorectal adenoma samples by primary adenomas without recurrence, primary adenomas with recurrence and recurrent adenomas, respectively, and statistical differences were evaluated (Figure 46C; Table 24). Obtained ASNs of *CDX2* tended to be capable of discriminating the groups from each other ($P=0.102$, one-way ANOVA; Figure 46C): primary adenomas without recurrence exposed a lower ASN (mean 2.02 ± 0.03) compared to primary adenomas with recurrence (mean 2.18 ± 0.33) and recurrent adenomas (mean 2.32 ± 0.54). On the contrary, the ASN of *CDX2* was slightly higher in recurrent adenomas compared to primary adenomas with recurrence, though not reaching statistical significance ($P=0.382$, one-way ANOVA with multiple comparisons; Figure 46C). Based on ASN it was not possible to discern the adenoma groups with statistical significance although ASNs of *CLIC1*, *EGFR*, *MYC*, and *ZNF217* slightly differed between the adenoma subgroups (Figure 46C; Table 24).

Table 24. Statistical evaluation of ASNs of miFISH analysis on adenomas. Comparison of primary adenomas without recurrence (*n*=14), with recurrence (*n*=28) and recurrent adenomas (*n*=13). Adjustment for multiple testing was performed by Tukey's multiple comparison test.

Marker	One-way ANOVA *p*-value		Tukey **Primary adenoma *w/o recurrence* vs Primary adenoma *w/ recurrence***		Tukey **Primary adenoma *w/ recurrence* vs Recurrent adenoma**		Tukey **Primary adenoma *w/o recurrence* vs Recurrent adenoma**	
	not adjusted for multiple testing	adjusted for multiple testing	not adjusted for multiple testing	adjusted for multiple testing	not adjusted for multiple testing	adjusted for multiple testing	not adjusted for multiple testing	adjusted for multiple testing
COX2	0.622	0.798	0.808	0.900	0.615	0.900	0.900	0.900
PIK3CA	0.780	0.910	0.900	0.900	0.743	0.900	0.900	0.900
APC	0.941	0.941	0.900	0.900	0.900	0.900	0.900	0.900
CLIC1	0.365	0.798	0.628	0.900	0.714	0.900	0.334	0.900
EGFR	0.534	0.798	0.638	0.900	0.900	0.900	0.533	0.900
MYC	0.158	0.798	0.256	0.900	0.833	0.900	0.169	0.900
CCND1	0.627	0.798	0.598	0.900	0.890	0.900	0.900	0.900
CDX2	0.102	0.798	0.382	0.900	0.463	0.900	0.084	0.900
CDH1	0.321	0.798	0.900	0.900	0.350	0.900	0.392	0.900
TP53	0.454	0.798	0.471	0.900	0.900	0.900	0.551	0.900
HER2	0.928	0.941	0.900	0.900	0.900	0.900	0.900	0.900
SMAD7	0.508	0.798	0.493	0.900	0.900	0.900	0.659	0.900
SMAD4	0.485	0.798	0.457	0.900	0.900	0.900	0.704	0.900
ZNF217	0.569	0.798	0.547	0.900	0.842	0.900	0.899	0.900
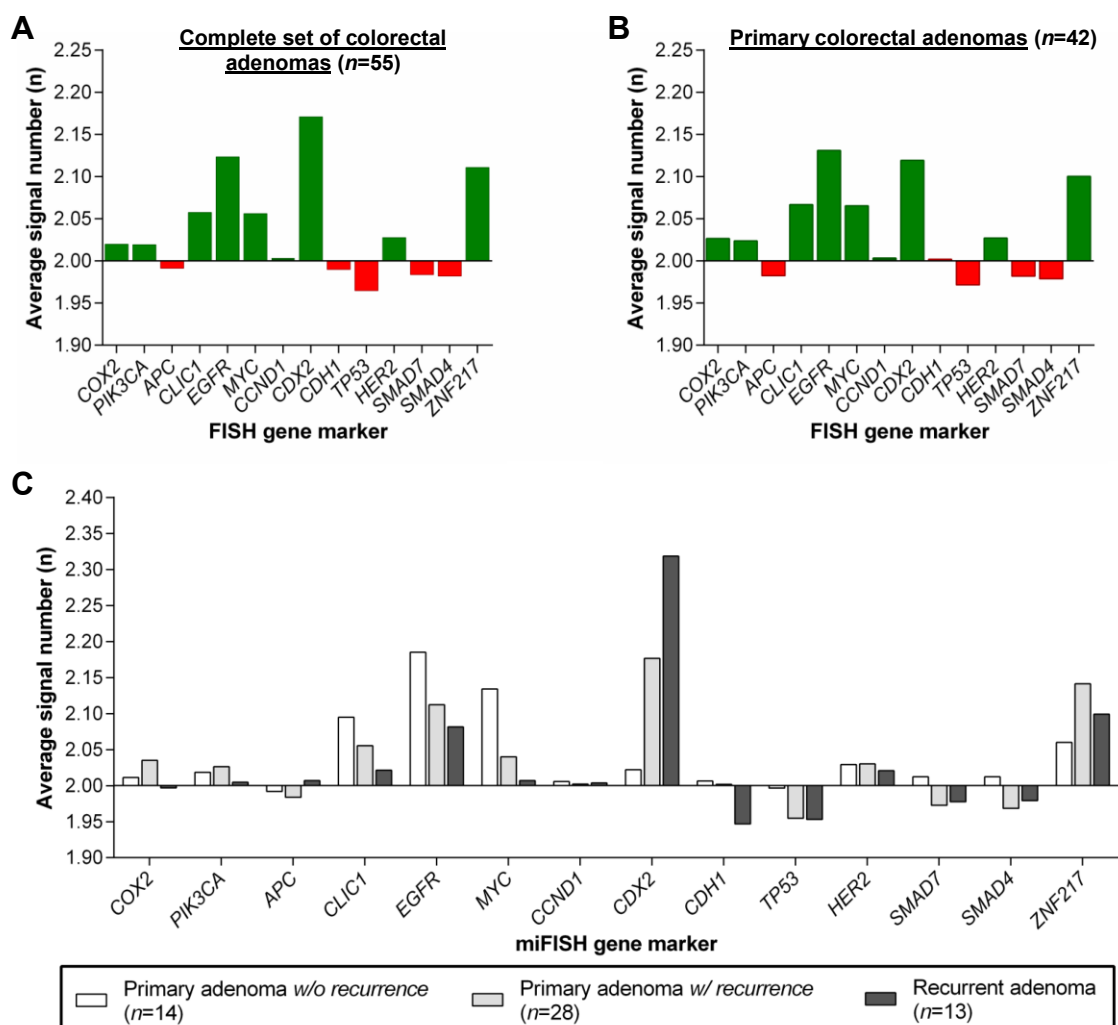
Figure 46. Legend on next page.

Figure 46. Average signal numbers of miFISH markers in colorectal adenomas. (**A**) ASNs (mean) per miFISH marker (sorted by chromosome location) in the cohort of colorectal adenomas (*n*=55). Signal numbers above the baseline of 2.00 represent an average copy number gain (*green*) while signal numbers below resemble an average copy number loss (*red*). As expected, probes targeting oncogenes were, on average, gained whereas TSG probes were lost. (**B**) ASNs of primary colorectal adenomas (*n*=42) revealed a similar distribution. However, observed signal numbers of *EGFR* were, on average, higher than for *CDX2*. (**C**) Distribution of ASNs (mean) per miFISH marker in colorectal adenomas without recurrence, primary adenomas with recurrence and recurrent adenomas, respectively.

To assess the number of CNAs per case, we applied a 10% threshold, *i.e.*, signal patterns present in less than 10% of the counted nuclei were not considered (Table 25; Figure 47A). In the entire sample set the amplitude displayed that *EGFR* CN gain (*n*=13) was most abundant followed by gains of copy numbers of *CDX2* (*n*=12) and *ZNF217* (*n*=10), respectively (Figure 47A). While alterations of *CLIC1* (*n*=6), *MYC* (*n*=5), and *TP53* (*n*=5) were less common in the adenoma cohort, other probes displayed an even rarer abundance of CNAs (Figure 47A). The performance of gene markers among colorectal adenomas grouped by the status of recurrence was not linked with differential patterns of CNAs (*P*=0.372; Figure 47B). Differences in the percentages of CNAs in primary adenomas without and with recurrence were detected for *CLIC1* (7.1% vs 14.3%), *EGFR* (42.9% vs 17.9%), *MYC* (21.4% vs 7.1%), and *ZNF217* (7.1% vs 21.4%). Remarkably, CNAs of six of the gene markers (*COX2*, *PIK3CA*, *CDX2*, *TP53*, *SMAD7*, *SMAD4*) were exclusively reported in primary adenomas with recurrence (Figure 47B). Thus, primary adenomas without recurrence (*n*=14) reported a less broad distribution of different alterations across the miFISH markers than primary adenomas with recurrence (*n*=28). 6 of 14 miFISH-markers were affected in primary adenomas without recurrence versus 12 of 14 in primary adenomas with recurrence (*P*=0.046, Fisher's exact test). Despite that, comparing the average number of CNAs per sample failed to discern primary adenomas without recurrence from primary adenomas with recurrence (means 1.0 ± 1.1 vs 1.39 ± 2.1, *P*=0.917; Figure 47C). CN status of *CDX2* on 13q12.2, in contrast, evolved potential to discriminate 25% (sensitivity) of the primary adenomas with recurrence from the non-recurring adenomas (100% specificity) (*P*=0.040; Figure 47D). Although the sensitivity seemed arguably low, the gain of this marker concurrently provided a specificity of 100% due to the absence of this CNA in primary adenomas without recurrence (*n*=14). Additionally, a probe set comprising *CLIC1*, *CDX2*, and *ZNF217* would have detected 39% (sensitivity) of the primary adenomas with recurrence via CN gain of at least one of the markers (*P*=0.036; Figure 47E). A decrease of test specificity to 93%, due to an adenoma without recurrence showing positivity for both *CLIC1* and *ZNF217* (case A10; Figure 38), would have admittedly altered the clinical value of this panel.
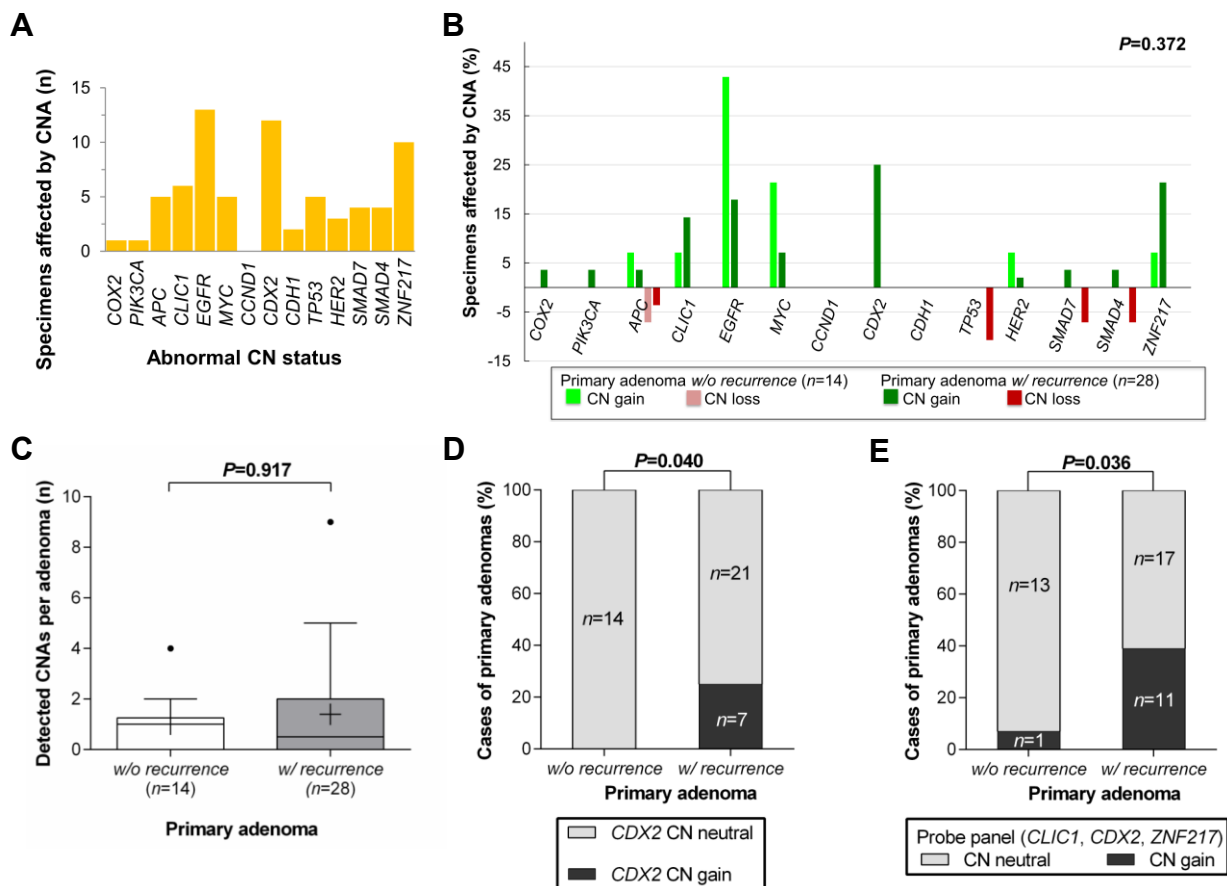
Figure 47. CNAs in primary adenomas without and with recurrence. (**A**) Distribution of CNAs across the miFISH gene markers in colorectal primary and recurrent adenomas (*n*=55). (**B**) Percentage of gains and losses for all miFISH gene markers observed in primary adenomas without recurrence and primary adenomas with recurrence (Wilcoxon signed-rank test). Percentages of gains are shown above the 0% line, and losses are shown below the 0% line. (**C**) Comparing primary adenomas without recurrence to primary adenomas with recurrence did not show a difference in the number of detected CNAs (Mann-Whitney *U* test). +, mean; whiskers, Tukey. (**D**) Copy numbers of *CDX2* were exclusively gained in primary adenomas with recurrence compared to primary adenomas without recurrence (Chi-square test), providing a 25% sensitivity and 100% specificity. (**E**) A probe panel consisting of *CLIC1*, *CDX2*, and *ZNF217* detected primary adenomas with recurrence from primary adenomas without recurrence with 39% sensitivity while providing 93% specificity (Fisher's exact test).

Table 25. Aberration frequencies detected by miFISH.

Marker	Gene location	Primary adenomas and recurrent lesions	Primary adenomas	Primary adenomas *w/o recurrence*	Primary adenomas *w/ recurrence*	Recurrent adenomas
COX2	1q31	1/55 (1.8%)	1/42 (2.4%)	0/14 (0%)	1/28 (3.6%)	0/13 (0%)
PIK3CA	3q26	1/55 (1.8%)	1/42 (2.4%)	0/14 (0%)	1/28 (3.6%)	0/13 (0%)
APC	5q22	5/55 (9.1%)	4/42 (9.5%)	2/14 (14.3%)	2/28 (7.1%)	1/13 (7.7%)
CLIC1	6p21	6/55 (10.9%)	5/42 (11.9%)	1/14 (7.1%)	4/28 (14.3%)	1/13 (7.7%)
EGFR	7p11	13/55 (23.6%)	11/42 (26.2%)	6/14 (42.8%)	5/28 (17.8%)	2/13 (15.4%)
MYC	8q24	5/55 (9.1%)	5/42 (11.9%)	3/14 (21.4%)	2/28 (7.1%)	0/13 (0%)
CCND1	11q13	0/55 (0%)	0/42 (0%)	0/14 (0%)	0/28 (0%)	0/13 (0%)
CDX2	13q12	12/55 (21.8%)	7/42 (16.6%)	0/14 (0%)	7/28 (25.0%)	5/13 (38.5%)
CDH1	16q22	2/55 (3.6%)	0/42 (0%)	0/14 (0%)	0/28 (0%)	2/13 (15.4%)
TP53	17p13	5/55 (9.1%)	3/42 (7.1%)	0/14 (0%)	3/28 (10.7%)	2/13 (15.4%)
HER2	17q12	3/55 (5.4%)	3/42 (7.1%)	1/14 (7.1%)	2/28 (7.1%)	0/13 (0%)
SMAD7	18q21	4/55 (7.3%)	3/42 (7.1%)	0/14 (0%)	3/28 (10.7%)	1/13 (7.7%)
SMAD4	18q21	4/55 (7.3%)	3/42 (7.1%)	0/14 (0%)	3/28 (10.7%)	1/13 (7.7%)
ZNF217	20q13	10/55 (18.2%)	7/42 (16.6%)	1/14 (7.1%)	6/28 (21.4%)	3/13 (23.1%)
Ratio of CNAs per lesion		1.29	1.21	1.00	1.39	1.57

3.3.4 Hierarchical clustering of CNAs in colorectal adenomas

Performing unsupervised clustering based on the fourteen miFISH gene identifier probes generated a heatmap of 55 colorectal adenoma samples which separated into four clustering groups (Figure 48A). Further examination of the assignment of clusters exhibited a strong relationship with the occurrence of CNAs (P<0.001; Figure 48B). Cluster 1 predominately combined adenomas with a gain of *EGFR* (r=-0.328, P=0.015, Spearman correlation) whereas Cluster 2 was mainly composed of adenomas without any detected CNAs (Figure 48A). Cluster 3 included adenomas with multiple different CNAs and Cluster 4 was strongly associated with the CN gain of *CDX2* (r=0.497, P<0.001, Spearman correlation). Although the losses of *SMAD7*/*SMAD4* were present only in a subset of samples of Cluster 4, there was no significant correlation (r=-0.101, P=0.463, Spearman correlation). Noteworthy, the CNA of *ZNF217* correlated positively with the cluster assignment (r=-0.377, P=0.005, Spearman correlation), although both Cluster 3 and 4 contained samples affected by this alteration. Moreover, heatmap cluster groups correlated positively with genetic heterogeneity expressed by the following measures of diversity (see 2.6.4 Data Assessment): instability index (r=0.360, P=0.021, Spearman correlation), Shannon entropy (r=0.390, P=0.011, Spearman correlation) and Simpson diversity index (r=0.359, P=0.019, Spearman correlation). Cluster 3 exhibited the highest diversity indices due to the increased number of CNAs (see 3.3.5 Clonal composition).

The distribution of primary adenomas without recurrence and with recurrence across the heatmap exhibited a very weak trend to separate both groups (P=0.193; Figure 48C). Remarkably, cluster 4 was exclusively composed of primary adenomas with recurrence. Further evaluation of the histopathological features and the cluster assignment unveiled a relationship between the dichotomized localization, *i.e.*, either colon or rectum, of the primary adenoma specimens (P=0.021; Figure 48D). Rectal samples were stronger associated with Cluster 3 and 4 compared to colonic adenomas dominating Cluster 1 and 2. This observation is in line with the higher average number of CNAs in rectal compared to colonic adenomas detected by aCGH (see 3.2.1 Overall aCGH analysis). Nevertheless, consideration of additional clinicopathologic parameters such as patient age at diagnosis, adenoma size, histology or grading did not uncover a statistically significant correlation with the heatmap cluster assignment based on the ASN of miFISH probe markers.
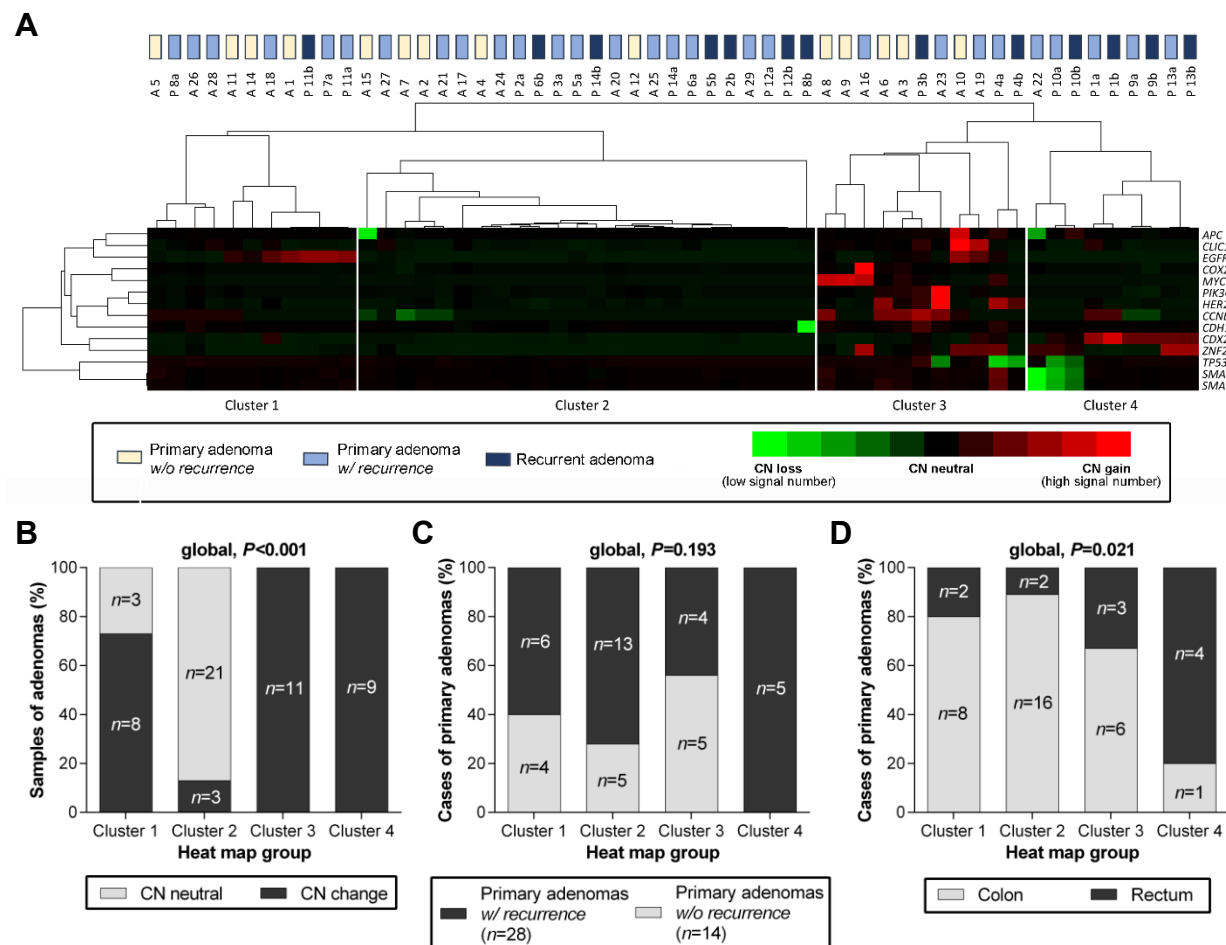
Figure 48. Heat map clustering of 55 colorectal adenomas by CNAs of miFISH. (**A**) The heat map was generated by calculating the average signal number (ASN) per marker in each sample. Dendrogram correlation sorted genes and specimens. ASNs were assigned a *green* color representing CN losses, while *black* indicated CN neutrality and a *red* color depicted CN gains, respectively. The heat map separated the adenomas in four clustering groups (Cluster 1-4). (**B**) The heat map cluster assignment was apparently based on the occurrence of copy number changes within the colorectal adenoma specimens (Freeman-Halton test). (**C**) Cluster assignment tended to separate primary adenomas without recurrence from primary adenomas with recurrence (Freeman-Halton test). (**D**) Primary colorectal adenomas localized in the rectum or colon revealed a relationship with the cluster assignment (Freeman-Halton test).

3.3.5 Clonal composition of primary (recurring) adenomas

Tumor heterogeneity assessed by signal patterns and instability index

Comparing the clonal composition within primary colorectal adenomas without recurrence and with recurrence, color displays of signal pattern clones were assessed. These visualized the clonal frequencies and appearances of CNAs (see Figure 38-44 of chapter 3.3.2). The population size of the major clone fraction comprised on average 62.8% (IQR, 39-88%) of the aberrant cell population within all analyzed adenoma (*n*=55) samples (Table 26). The median size of the major clone population did not differ when comparing primary adenomas without recurrence to primary adenomas with recurrence (median 69.0% vs 68.0%; *P*=0.795; Figure 49A). Interestingly, recurrent

adenomas seemed to exhibit a slightly larger size of the major clone population (median 85.0%) than the primary adenomas. It might indicate a higher level of clonal selection: however, the difference was statistically not significant (P=0.696; Figure 49A).

The analysis of CNAs (only considering clones >10%) within the group of primary adenomas without recurrence unveiled that 35.7% (5/14) of cases were not affected by a CNA of any of the gene markers. 42.8% (6/14) of cases showed a single alteration and 21.4% (3/14) of adenomas were detected with multiple CNAs (Figure 49B). Nevertheless, within the primary adenomas without recurrence a major clone with four CNAs, *i.e.*, *APC*, *CLIC1*, *EGFR*, and *ZNF217*) was present in adenoma case A10 only (Figure 38). In contrast, primary adenomas with recurrence were not affected by any CNA in 50.0% (14/28) of cases. However, 17.9% (5/28) of adenomas revealed a single CNA and 32.1% (9/28) of cases demonstrated multiple alterations (Figure 49B). The comparison of CNA frequencies across primary adenomas with and without recurrence showed merely a weak trend to separate the groups (P=0.263; Figure 49B). Yet, primary adenomas with recurrence harbored a substantially higher average number of CNAs per clone than primary adenomas without recurrence when only cases with clonal aberrations were considered (Figure 49C). Three and four CNAs were accumulated in clones of primary adenomas with recurrence. Clone populations displaying more than four CNAs were exclusively observed in primary adenomas with recurrence.

With the objective to quantitate the degree of genomic instability, an instability index was calculated by the number of aberrant FISH signals per 100 cells divided by the total number of nuclei (Table 26). Within the analyzed primary and recurrent adenomas (n=55) the average chromosomal instability index equaled 4.9 ± 5.1 patterns, suggesting a low-level intratumor heterogeneity (ITH) across the adenoma samples. The instability index as a measure of ITH could not discern primary adenomas without recurrence from primary adenomas with recurrence (means 4.8 ± 5.6 vs 4.8 ± 3.2; P=0.603; Figure 49D). Among the cohort of primary and recurrent adenomas, a fraction of 29.1% (16/55) samples did not harbor any aberrant clonal population exceeding 5% of the aberrant cells. However, in 23.6% (13/55) of samples, only a single clonal population was present while 47.3% (26/55) of adenomas were composed of multiple clonal populations (Figure 49E). The number of clone populations was not a measure to discern primary adenomas without and with recurrence (P=0.675; Figure 49F). Noteworthy, more than half of the adenomas harbored multiple clonal populations indicating the presence of intratumor heterogeneity.

Table 26. Signal patterns of the largest aberrant clone populations and instability indices.

	Sample ID	Recurrence	Largest clone populations with signal pattern°	Instability index§
Primary adenoma *w/o recurrence*	A 1	no	99% 22223222222222	1.2
	A 2	no	No	2.5
	A 3	no	30% 22222322222222; 20% 22223222222222; 6% 23222222222222	10.7
	A 4	no	No	1.4
	A 5	no	No	2.3
	A 6	no	41% 22222222223222; 6% 22232222222222	9.6
	A 7	no	35% 22222223222222; 8% 12222222222222	5.3
	A 8	no	69% 22222322222222; 6% 22233322222222	9.1
	A 9	no	82% 22222322222222; 8% 22232322222222	5.0
	A 10	no	66% 22333222222223; 25% 22333222222222	6.0
	A 11	no	71% 22223222222222; 11% 22233322222222	5.8
	A 12	no	27% 23222222222222; 22% 22232222222222	2.7
	A 13 †	no	40% 34444444333224; 13% 34434444333224	22.6
	A 14	no	89% 22223222222222	1.5
	A 15	no	89% 21222222222222	4.3
Primary adenoma *w/ recurrence*	A 16	Primary	68% 32222322222223; 13% 32222222222223	7.4
	A 17	Primary	24% 22232222222222; 15% 22222222223222; 11% 22222222122222	6.6
	A 18	Primary	65% 22223223222222; 17% 22223222222222	4.2
	A 19	Primary	71% 22233223222223; 11% 22233222222223	7.4
	A 20	Primary	No	2.1
	A 21	Primary	No	1.6
	A 22	Primary	45% 22122222222113; 34% 22122222222112	6.9
	A 23	Primary	44% 23222222222222; 38% 23222222213222	7.7
	A 24	Primary	No	3.0
	A 25	Primary	No	1.4
	A 26	Primary	34% 22232222222222; 17% 22222223322113	7.3
	A 27	Primary	31% 22232222222222; 21% 22222222222223; 8% 22223223222223	6.3
	A 28	Primary	85% 22232222222222; 6% 22222233222222	1.5
	A 29	Primary	No	1.4
	A 30 †	Primary	53% 34344334444334; 5% 34344334344334	21.7
***Matched pairs:* primary adenoma and corresponding recurrent adenoma**	P 1a	Primary	**76% 22222224222222**; 16% 22232224222222	5.2
	P 1b	Recurrent	**94% 22222224222222**	3.8
	P 2a	Primary	No	2.7
	P 2b	Recurrent	No	2.2
	P 3a	Primary	48% 22222222122222	1.6
	P 3b	Recurrent	5% 22122222222222	24.0
	P 4a	Primary	**12% 22222222212222; 8% 22222222213222**; 7% 22222222222333; 6% 22223222222333	30.6
	P 4b	Recurrent	**50% 22222222212222; 25% 22222222213222**	5.2
	P 5a	Primary	No	2.4
	P 5b	Recurrent	No	2.1
	P 6a	Primary	**69% 22232222222222**	1.4
	P 6b	Recurrent	15% 22222223222222; **12% 22232222222222**; 10% 22222222212222; 9% 22122222222222	3.9
	P 7a	Primary	**85% 22223222222222**	7.8
	P 7b †	Recurrent	34% 33444434344334; 14% 34444434344334; **8% 22223222222222**; 5% 33443434344334	24.9
	P 8a	Primary	39% 22232222222222; **21% 22222222122222;** 18% 22222223222222	2.3
	P 8b	Recurrent	**96% 22222222122222**	2.7
	P 9a	Primary	**97% 22222223222222**	2.9
	P 9b	Recurrent	**85% 22222223222222**	6.4
	P 10a	Primary	**86% 22222223212113**; 9% 22222223222113	1.7
	P 10b	Recurrent	63% 22322223212113; **23% 22222223212113**	4.2
	P 11a	Primary	**88% 22223222222222**; 7% 22223222222223	3.8
	P 11b	Recurrent	**96% 22223222222222**	2.5
	P 12a	Primary	No	0.8
	P 12b	Recurrent	No	1.8
	P 13a	Primary	**89% 22222223222223**; 5% 22222223222222	4.2
	P 13b	Recurrent	**92% 22222223222223**	4.7
	P 14a	Primary	No	2.6
	P 14b	Recurrent	No	2.0

° Signal patterns are displayed in the following gene order: *COX2, PIK3CA, APC, CLIC1, EGFR, MYC, CCND1, CDX2, CDH1, TP53, HER2, SMAD7, SMAD4, ZNF217*. Clone populations (shown are clones exceeding 5%) are depicted as percentage of aberrant cells. Identical clone patterns in primary and recurrent adenomas are shown in bold.

§ The instability index is calculated as described in Materials and Methods. † excluded from analysis due to DNA ploidy status
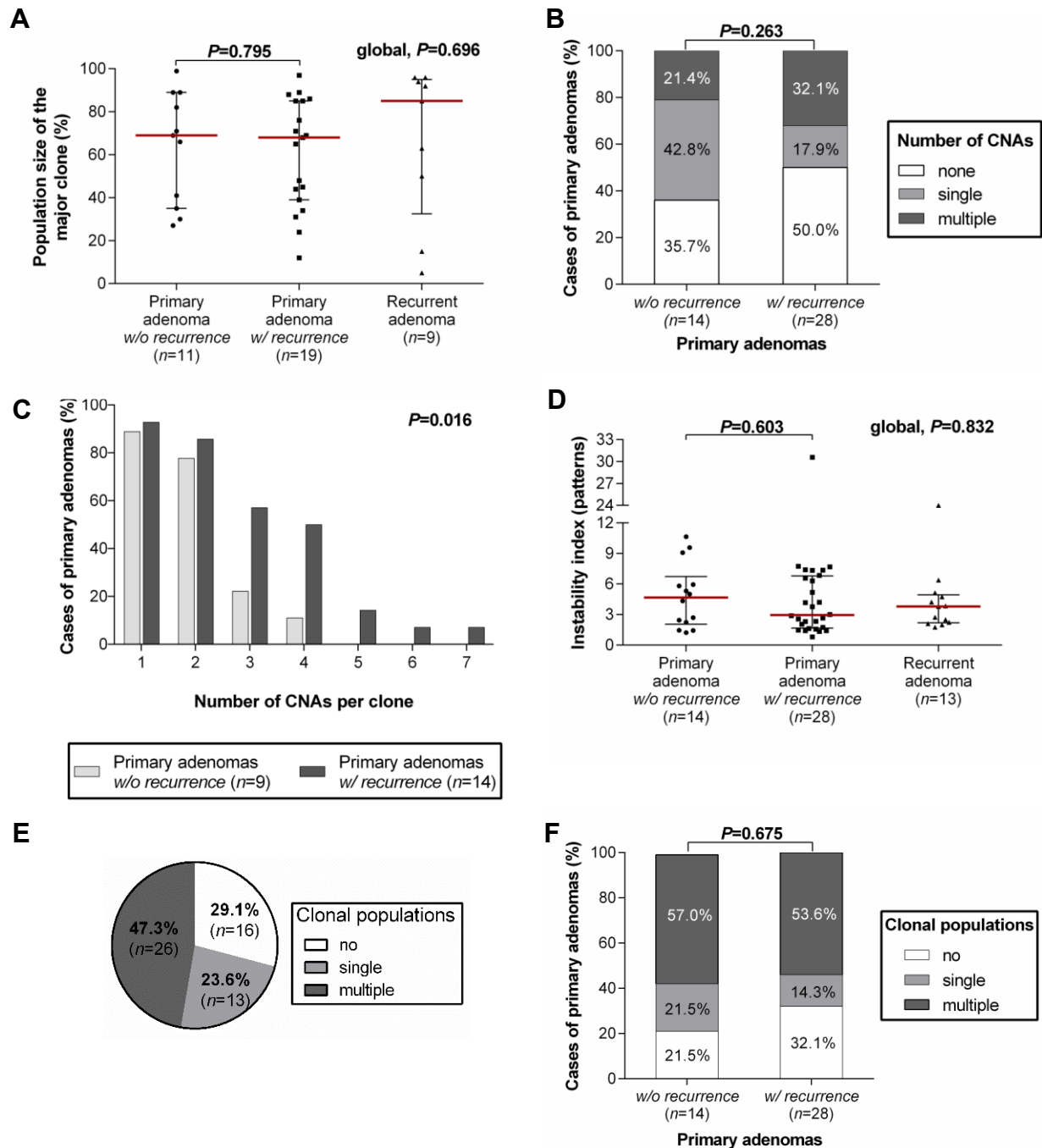
Figure 49. Clonal populations and instability indices of colorectal adenomas. (**A**) Major clone populations are on average (median with IQR) of comparable size in primary adenomas without and with recurrence (Mann-Whitney *U* test). Recurrent adenomas displayed a larger major clone size, however statistically not relevant (Kruskal-Wallis test). Adenoma samples without identifiable clone population (≤35 nuclei) were not considered. (**B**) Distribution of the number of CNAs across primary adenomas without recurrence and with recurrence (Freeman-Halton test). (**C**) Shown is the distribution of CNAs per clone across primary colorectal adenomas without and with recurrence, respectively (Wilcoxon signed-rank test). Considered were cases with at least a single CNA. (**D**) Instability indices (median with IQR) of primary adenomas with and without recurrence were similar (Mann-Whitney *U* test). Recurrent adenomas did not show a different average instability index (Kruskal-Wallis test). (**E**) Fractions of clonal populations within primary and recurrent adenomas (*n*=55) uncovered that approximately half of the samples comprised of two or more clone populations. (**F**) Clonal populations in primary adenomas without and with recurrence exposed similar fractions (Freeman-Halton test).

Tumor heterogeneity assessed by measures of diversity

Seeking to quantitatively characterize and compare the colorectal adenomas regarding genetic intra- and intertumor heterogeneity, ecological diversity measures were applied (see 2.6.4 Data Assessment: Tumor heterogeneity). All analyzed adenomas (*n*=55) revealed genetically different clones, *i.e.*, signal patterns (Figure 38-44) whose median number equaled 17 patterns (IQR, 10-25 patterns) in the cohort (*n*=55). A median of 17 patterns (IQR, 11-33 patterns) was noted in primary adenomas (*n*=42), while recurrent adenomas (*n*=13) showed a median of 12 patterns (IQR, 9-23 patterns).

Genetic diversity expressed by Shannon entropy (SE) considers the signal pattern quantity and their abundance. The values exhibited a heterogeneous distribution among primary adenomas with and without recurrence, respectively, and recurrent adenomas (*P*=0.648; Figure 50A). Average SE of primary adenomas equaled a median of 1.1 (IQR, 0.4-1.9). SE was not capable of distinguishing primary adenomas without recurrence from primary adenomas with recurrence (*P*=0.516, Figure 50A; *P*=0.219, Figure 50B). Primary adenomas with recurrence tended to be associated with SE exceeding 2.0 (above 75% quartile) compared to primary adenomas without recurrence when only cases were considered which revealed CNAs (*P*=0.086; Figure 50C).

Next, the Simpson diversity index (SI; range 0-1) was calculated which benefits abundant clones by more weight within the formula. Comparing the sample subgroups did not expose differential median SI (*P*=0.766; Figure 50D). The median SI of primary adenomas equaled 0.31 (IQR, 0.09-0.55). Median indices of primary adenomas with and without recurrence failed to distinguish the subgroups (*P*=0.531; Figure 50D). Despite that, the sample distribution range varied stronger in primary adenomas with recurrence compared to primary adenomas without recurrence (grouped SI in intervals of 0.1, *P*=0.086; Figure 50E). Adenoma cases ranging in the top third range (0.66-1.00) of the index were exclusively associated with recurrence (*P*=0.052; Figure 50F).

Additional to the diversity measures developed by Shannon and Simpson, the genetic distance was assessed by calculating the accumulated pairwise genetic diversity for the adenoma subsets (*P*=0.648; Figure 51A). The utilized distance metric equaled a median of 2.65 (IQR, 2.17-4.36) in primary adenomas. However, comparing primary adenomas by recurrence status did not reveal statistical differences (*P*=0.516; Figure 51A). Grouping primary adenomas with and without recurrence by the accumulated pairwise genetic diversity in intervals of 1.0 did not show a relevant difference (*P*=0.156; Figure 51B). Of note, the occurrence of tetraploid cells within diploid (2N) adenomas was identified to be very strongly correlated with the accumulated pairwise

genetic diversity, *i.e.*, genetic distance (*P*<0.001; Figure 51C). Noteworthy, average numbers of signal patterns differed across primary adenomas without tetraploid cells (*n*=31; median 16, IQR 9-23) compared to primary adenomas with one or more tetraploid cells (*n*=9; median 37, IQR 14-47) (*P*=0.016, Mann-Whitney *U* test).

Ploidy was identified to be a fundamental cause for differential genetic diversity measures: adenomas with a tetraploid (4N) average ploidy showed higher values of instability index, SI and SE, respectively (*P*<0.001; Figure 51D-F), compared to adenomas of diploid character (2N). This justified the exclusion of the polyploid adenoma samples (*i.e.*, A13, A30 and P7b) regarding the advanced degree of genomic changes.
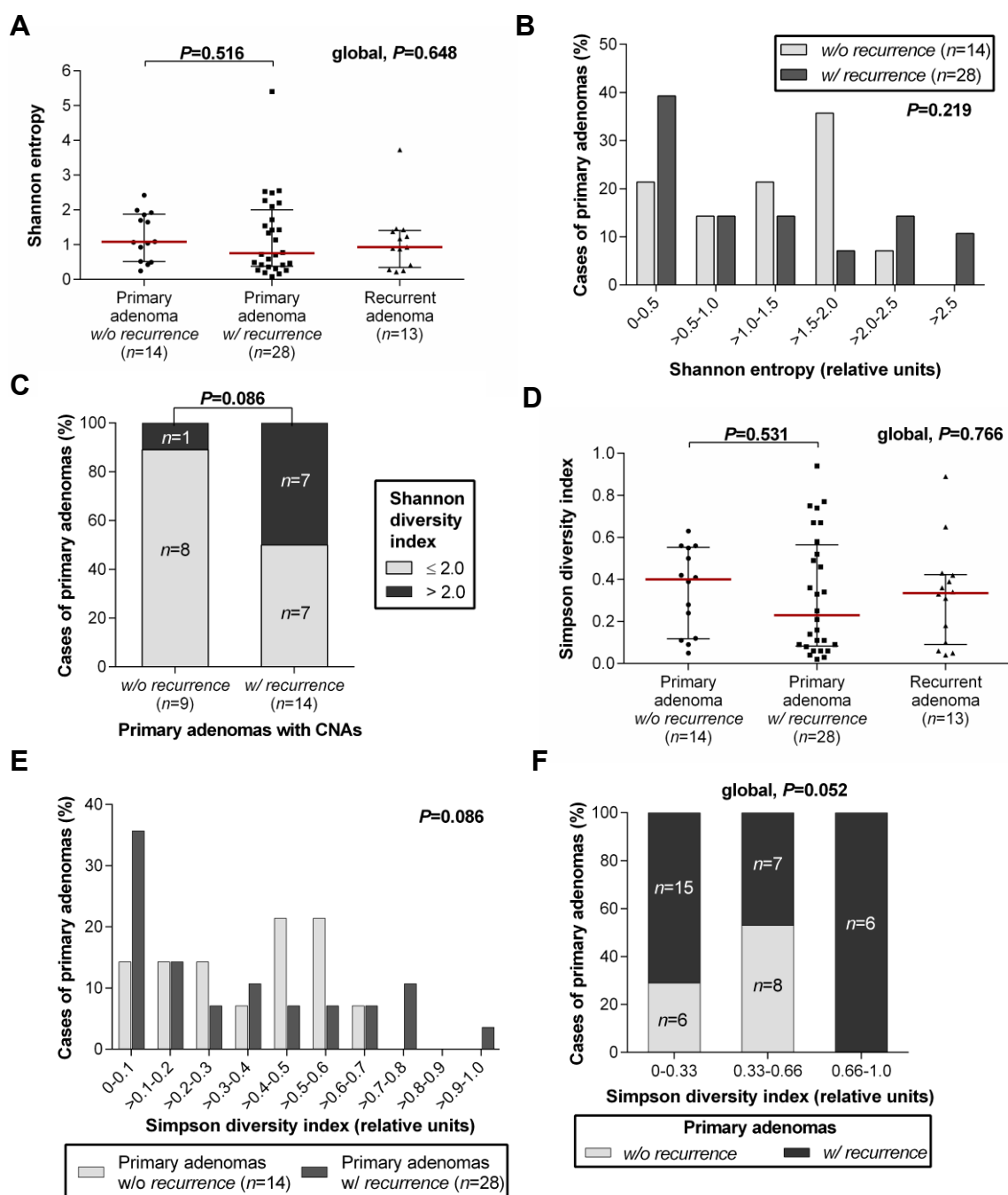
Figure 50. Legend on next page.

Figure 50. Tumor heterogeneity assessed by Shannon and Simpson diversity indices. (**A**) Shannon entropies (median with IQR) of primary adenomas with and without recurrence were similar (Mann-Whitney *U* test). Recurrent adenomas did not show a different average Shannon index (Kruskal-Wallis test). (**B**) Distribution of Shannon indices (grouped by intervals of 0.5) across primary adenomas without and with recurrence, respectively (Wilcoxon signed-rank test). (**C**) Primary adenomas with recurrence tended to be associated with higher Shannon indices (dichotomized) compared to primary adenomas without recurrence (Freeman-Halton test). Considered were cases with at least a single CNA. (**D**) Simpson diversity indices (median with IQR) of primary adenomas with and without recurrence were similar (Mann-Whitney *U* test). Recurrent adenomas did not show a different average Simpson index (Kruskal-Wallis test). (**E**) Distribution of Simpson indices (grouped by intervals of 0.1) across primary adenomas without and with recurrence, respectively (Wilcoxon signed-rank test). (**F**) Simpson indices in the upper third (trichotomized) of the range were exclusively associated with primary adenomas with recurrence (Freeman-Halton test).
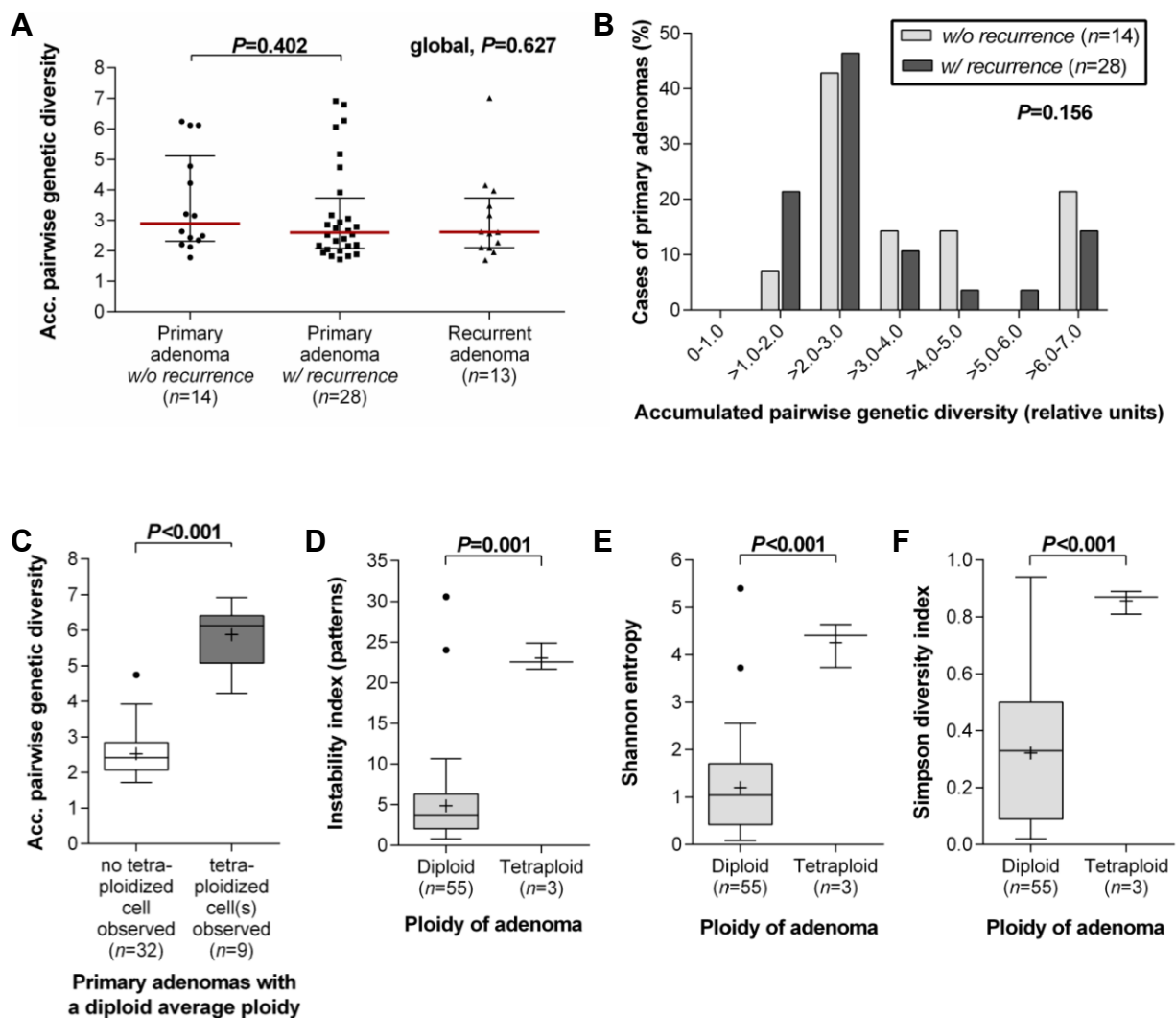
Figure 51. Tumor heterogeneity assessed by genetic distance. (**A**) Accumulated pairwise genetic diversity values (median with IQR) of primary adenomas with and without recurrence were similar (Mann-Whitney *U* test). Recurrent adenomas did not show a different average genetic distance index (Kruskal-Wallis test). (**B**) Distribution of accumulated pairwise genetic diversity (grouped by intervals of 1.0, range 0-7.0) across primary adenomas without and with recurrence, respectively (Wilcoxon signed-rank test). (**C**) Accumulated pairwise genetic diversity correlated strongly with the abundance of tetraploidized cells within the amassment of nuclei across primary adenomas with an average ploidy of diploid (2N) character (Mann-Whitney *U* test). (**D**) Adenomas with a near-tetraploid average ploidy (4N) exposed higher instability indices compared to diploid adenomas (2N; Mann-Whitney *U* test). (**E**) Near-tetraploid adenomas (4N) showed higher Shannon entropies compared to diploid adenomas (2N; Mann-Whitney *U* test). (**F**) Near-tetraploid adenomas (4N) exhibited higher Simpson diversity indices compared to diploid adenomas (2N; Mann-Whitney *U* test). Box plots in (**C**-**F**): +, mean; whiskers, Tukey.

119

### 3.3.6 Clonal evolution of paired primary and recurrent adenomas

Subanalysis of miFISH ASNs of matched pair adenomas did not show a significant difference comparing primary to recurrent lesions (P=0.383; Figure 52). Nevertheless, 43% (6/14) of pairs differed in CNAs across both lesions (Supplemental Table 5). Shedding light on the clonal evolution within the primary colorectal adenoma and the corresponding recurrent adenoma was a key feature of the presented study: CNAs observed in matched adenoma pairs were visualized by constructing consensus phylogenetic trees per matched pair utilizing the FISHtree software (Figure 53-56). The algorithm created FISH tress by applying a heuristic approach to minimize the number of CNAs across the tree. Emanating from the diploid root cell observed CNA-signal patterns were incorporated to form the tree nodes which are connected by edges resembling the underlying genomic events. The primary adenoma-recurrence consensus trees suggested four distinct clonal evolution patterns (Figure 53-56): first, a simplification pattern was observed which is characterized by a primary tumor with a broad level of ITH expressed by the presence of multiple different signal pattern clone populations (Figure 53A). On the contrary, the recurrent tumor depicted a clonal composition which is dominated by a previously rare minor clone of the primary adenoma while other clones vanished and became extinct. Thus, there was a shift in ITH (here represented by different clone populations) decreasing from the primary adenoma towards the recurrent adenoma (Figure 53B). For instance, the primary adenoma of pair P4 showed multiple clones, whereas only two fractions larger 5% dominated the recurrent lesion. Interestingly, a complexity pattern was observed which, *vice versa* to the simplification pattern, was distinguished by a major clone in the primary tumor (Figure 54A, 55A). However, the clone shrank to a minor clone in the recurrent adenoma which was instead dominated by several newly emerged clone populations (Figure 54B, 55B). Accordingly, the intratumor heterogeneity within the recurrent lesions increased drastically. While the primary adenoma of pair P7 showed a very clonal pattern with a major clone affected by *EGFR* CN gain, the recurrent tumor was dominated by vast numbers of clones being targets of a whole genome duplication (WGD) event (Figure 54A, B). Clonal aberrations exhibited complex patterns of CN counts, *i.e.*, trees developed multiple branches which increase the node depth.

In opposition to these two clonal evolution patterns which show a shift in clone numbers and intratumor heterogeneity, adenoma pairs were observed which were extraordinarily stable for measures mentioned above. Consequently, adenomas of the

stabilization pattern showed (relatively well) superimposing aberrant clone populations across the primary and the recurrent tumor considering both quantities of differential clones and population frequencies (Figure 56, 57A). Stabilization patterns were characterized by trees with low node depths indicating less complex patterns of CNAs with the identical major clone present in the primary and recurrent adenoma, respectively (Figure 57A). Despite that, pair P10 resembled a modification of this pattern type: the major clone of the primary lesion is recognizable in the recurrent tumor (stabilization); however, a clone evolved with an additional CNA (progression) (Figure 56, 57A).

Lastly, adenoma pairs which failed to expose copy number changes in more than 10% of the counted total cell population were categorized into "zero" pattern (Figure 57B). The lack of clonality and presence of few aberrant cells are emphasized by the failure to model phylogenetic trees for these cases. Since CNAs do not play a pivotal role in adenoma pairs of the "zero" clonal evolution pattern putatively other (epi)genetic mechanisms are drivers in these lesions (see 3.5.1 Immunohistochemical detection).

Addressing the question whether primary adenomas are, in general, differential compared to the recurrent counterpart, non-consensus phylogenetic trees were analyzed considering the number of tree levels, *i.e.*, the node depth (Figure 58A). As expected, a general trend of differential numbers of tree levels in primary lesions compared to recurrent lesions was not confirmed ($P=0.336$). This fact was underpinned by the distribution of clonal evolution patterns across the fourteen adenoma matched pairs (Figure 58B), which showed an imbalance towards adenoma cases of "stability" (64%; stabilization and "zero" patterns) versus "shift" (36%; simplification and complexity patterns). Interestingly, shifting patterns seemed to be associated with shorter adenoma-recurrence free time than stable patterns ($P=0.166$; Figure 58C). In-depth analysis of CNAs in match-paired adenomas unveiled that copy number gains of *CDX2* were exclusively associated with the stabilization pattern ($P=0.005$; Figure 58D).
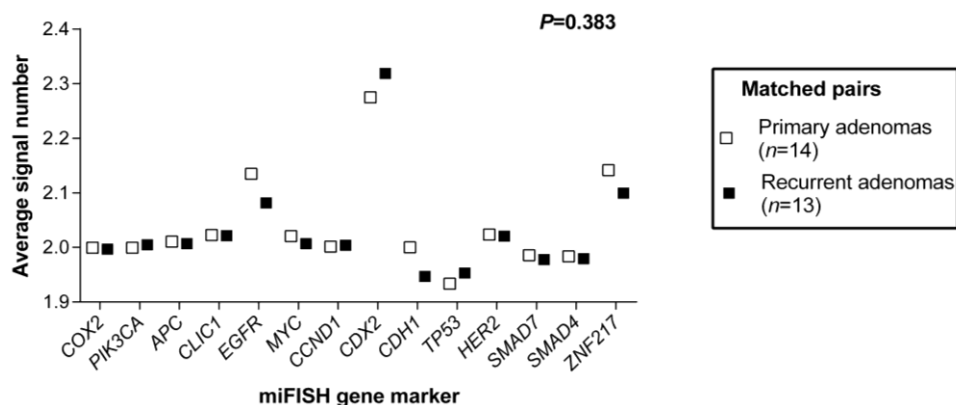
Figure 52. ASNs in match-paired adenomas. Although ASNs varied across the probes, primary adenomas did not show alternating ASNs compared to recurrent lesions (Wilcoxon signed-rank test).
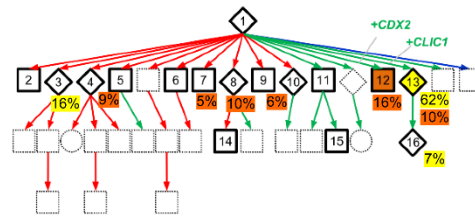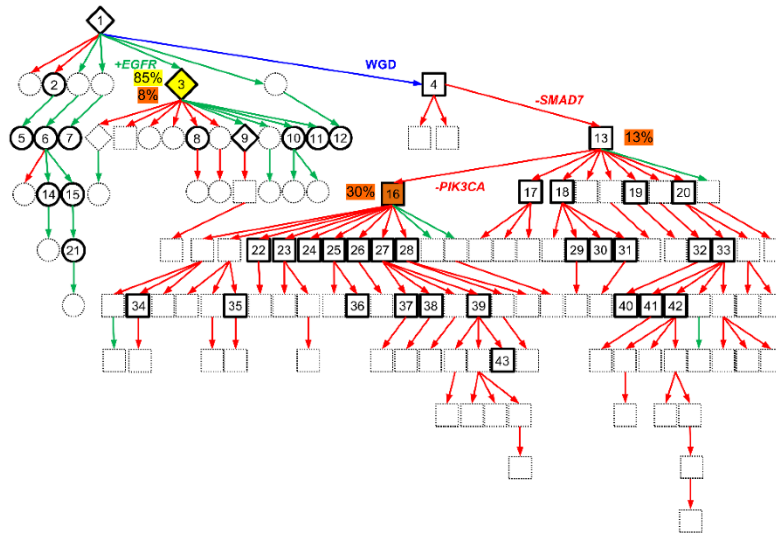
A

B
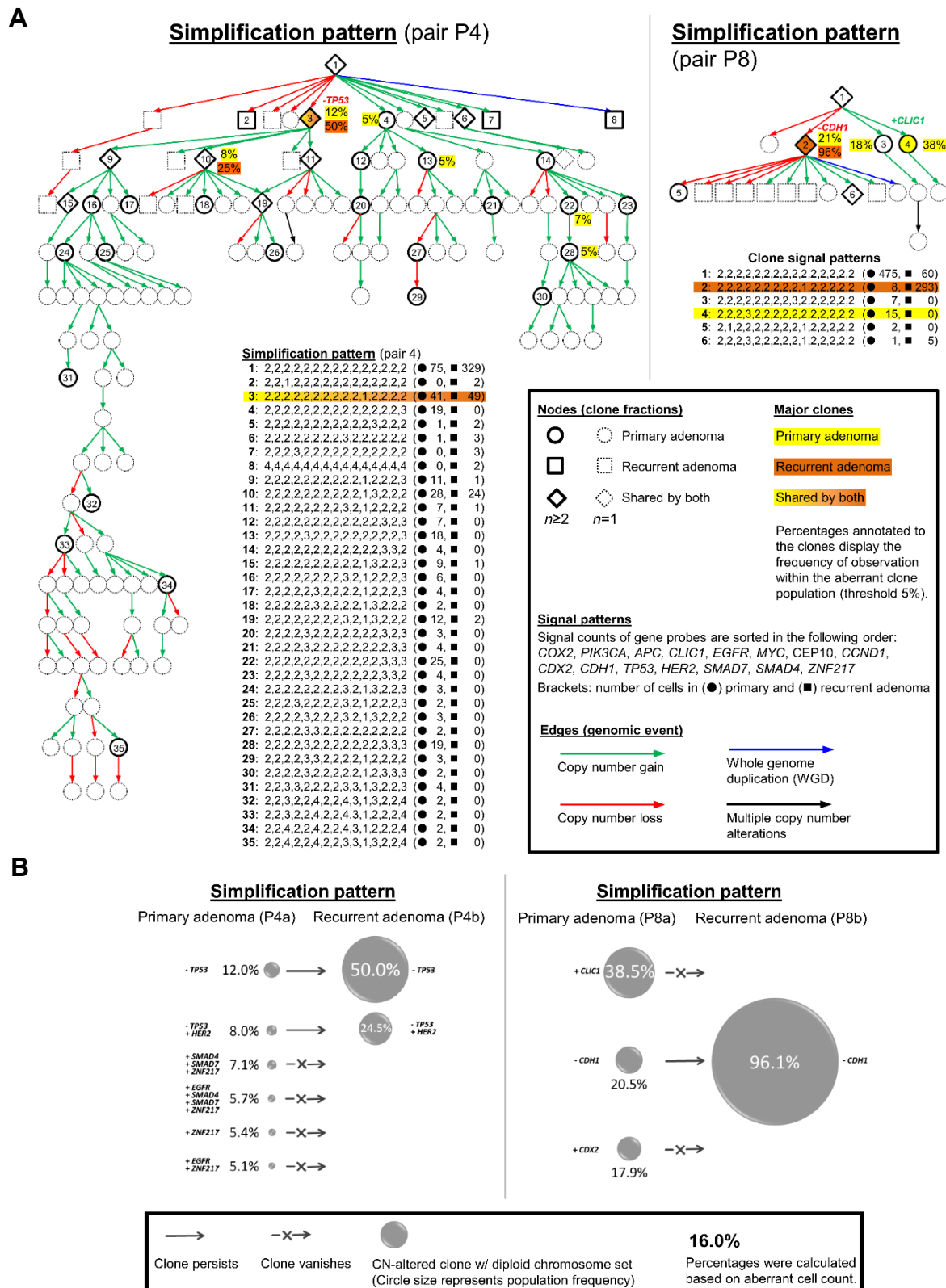
Figure 53. Phylogenetic trees of simplification pattern-adenomas. (**A**) Consensus phylogenetic trees depict the clonal distribution across the primary adenoma and the corresponding recurrent adenoma. Shown are pairs P4 and P8, respectively. Nodes (*circle*, *square*, *rhomb*) reflect the copy number profile of the fourteen gene identifier probes along CEP10. Genomic events which led to the formation of the major clone population (*highlighted*) are remarked to the edges. Adenoma pairs exhibiting the simplification pattern were characterized by vast numbers of diverging clone populations in the primary tumor compared to the recurrent adenoma (**B**) Schematic presentation of the simplification evolution pattern. The decline from multiple populations in the primary adenoma towards fewer numbers in the recurrent adenoma are visible in both cases (pairs P4 and P8, respectively). Graphs show populations only which exceeded 5% of the aberrant cells in the primary adenoma.
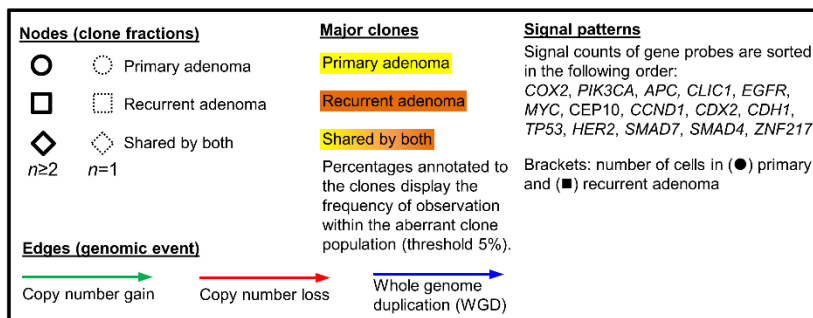
Complexity pattern (pair P3)

A

Clone signal patterns

1: 2,2,2,2,2,2,2,2,2,2,2,2,2,2,2 (● 548, ■ 375)
2: 1,2,2,2,2,2,2,2,2,2,2,2,2,2,2 (●   0, ■   4)
3: **2,2,1,2,2,2,2,2,2,2,2,2,2,2,2** (●   0, ■ 14)
4: 2,2,2,2,2,2,2,1,2,2,2,2,2,2,2 (●   0, ■   4)
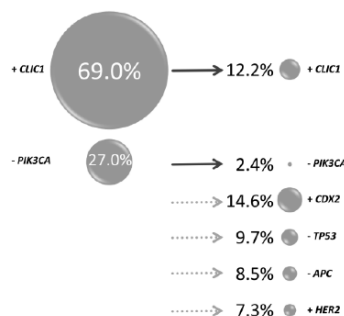5: **2,2,2,2,2,2,2,2,1,2,2,2,2,2** (● 12, ■   4)
6: 2,2,2,2,2,2,2,2,2,2,1,2,2,2,2 (●   3, ■   7)
7: 2,2,2,2,2,2,2,2,2,2,2,2,2,2,1 (●   2, ■   0)
8: 2,2,2,2,2,2,2,2,2,2,2,2,2,2,3 (●   1, ■   7)
9: 2,2,2,2,2,2,2,2,3,2,2,2,2,2,2 (●   0, ■   9)
10: 2,2,2,2,3,2,2,2,2,2,2,2,2,2,2 (●   0, ■ 13)
11: 2,2,2,3,2,2,2,2,2,2,2,2,2,2,2 (●   3, ■   9)
12: 2,3,2,2,2,2,2,2,2,2,2,2,2,2,2 (●   0, ■   4)
13: 3,2,2,2,2,2,2,2,2,2,2,2,2,2,2 (●   0, ■   4)
14: 0,2,2,2,2,2,2,2,2,2,2,2,2,2,2 (●   0, ■   2)
15: 2,2,1,2,2,2,2,2,2,2,2,2,1,1,2 (●   0, ■   3)
16: 1,2,1,2,2,2,2,2,2,2,2,2,2,2,2 (●   0, ■   3)
17: 2,2,2,2,2,2,2,4,2,2,2,2,2,2,2 (●   0, ■   3)
18: 2,2,2,2,3,2,2,3,2,2,2,2,2,2,2 (●   0, ■   2)
19: 2,2,2,3,2,2,2,3,2,2,2,2,2,2,2 (●   0, ■   2)
20: 2,2,2,2,2,2,2,2,3,3,2,2,2,2,2 (●   0, ■   2)
21: 2,2,2,3,2,2,2,2,2,2,2,2,2,2,3 (●   0, ■   2)
22: 2,2,2,2,2,2,2,2,2,2,2,2,1,1,2 (●   2, ■ 10)
23: 2,2,2,3,2,2,2,2,2,1,1,2,2,2 (●   0, ■   2)
24: 2,2,2,2,2,2,2,4,2,2,2,2,2,2,3 (●   0, ■   2)
25: 2,2,2,3,3,2,2,2,3,2,2,2,2,2,2 (●   0, ■   2)
26: 2,2,2,2,2,2,2,3,3,2,2,2,2,2,2 (●   0, ■   7)
27: 2,2,2,2,2,2,2,2,2,2,1,2,1,1,2 (●   0, ■   2)
28: 0,2,2,2,2,2,2,2,4,2,2,2,2,2,2 (●   0, ■   4)
29: 2,2,2,2,3,2,2,2,3,3,2,2,2,2,3 (●   0, ■   3)
30: 2,2,2,2,3,2,2,2,3,3,3,3,2,2,3 (●   0, ■   3)
31: 2,2,2,3,3,2,2,3,2,3,2,2,2,2,3 (●   0, ■   2)
32: 2,2,2,3,2,2,2,2,2,3,3,3,3,2,2,3 (●   0, ■   5)
33: 2,2,2,2,2,2,2,2,2,2,2,3,3,2,2 (●   0, ■   2)
34: 3,3,4,4,4,3,2,3,3,4,3,3,3,3,3 (●   0, ■   2)

Nodes (clone fractions)

○ ⬭ Primary adenoma

□ ⬚ Recurrent adenoma

◇ ◈ Shared by both

n≥2   n=1

Signal patterns

Signal counts of gene probes are sorted in the following order:
COX2, PIK3CA, APC, CLIC1, EGFR, MYC, CEP10, *CCND1, CDX2, CDH1, TP53, HER2, SMAD7, SMAD4, ZNF217*
Brackets: number of cells in (●) primary and (■) recurrent adenoma

Major clones

Primary adenoma

Recurrent adenoma

Percentages annotated to the clones display the frequency of observation within the aberrant clone population (threshold 5%).

Edges (genomic event)

→ Copy number gain

→ Copy number loss

→ Whole genome duplication (WGD)

→ Multiple copy number alterations

B

Complexity pattern

Primary adenoma (P3a)   Recurrent adenoma (P3b)

- CDH1 **48.0%** → 2.0% · - CDH1

+ CLIC1   8.0% → 3.4% + CLIC1

⇢ 5.3% - APC

⇢ 4.9% + EGFR

⇢ 4.6% + CDX2

→ Clone persists   ⇢ Clone emerges   ● CN-altered clone w/ diploid chromosome set (Circle size represents the population frequency.)

16.0% Percentages were calculated based on aberrant cell count.

Figure 54. Phylogenetic tree of complexity pattern-adenoma P3. (**A**) Consensus phylogenetic tree depicts the clonal distribution across the primary adenoma and the corresponding recurrent adenoma of pair P3. Nodes (*circle, square, rhomb*) reflect the copy number profile of the fourteen gene identifier probes and CEP10. Genomic events which led to the formation of the major clone population (*highlighted*) are remarked to the edges. Although the primary adenoma was composed of a small number of clone populations the recurrent adenoma revealed an evident increase of clonal populations. (**B**) Schematic presentation of the complexity evolution pattern. Several newly emerged populations were present in the recurrent adenoma.
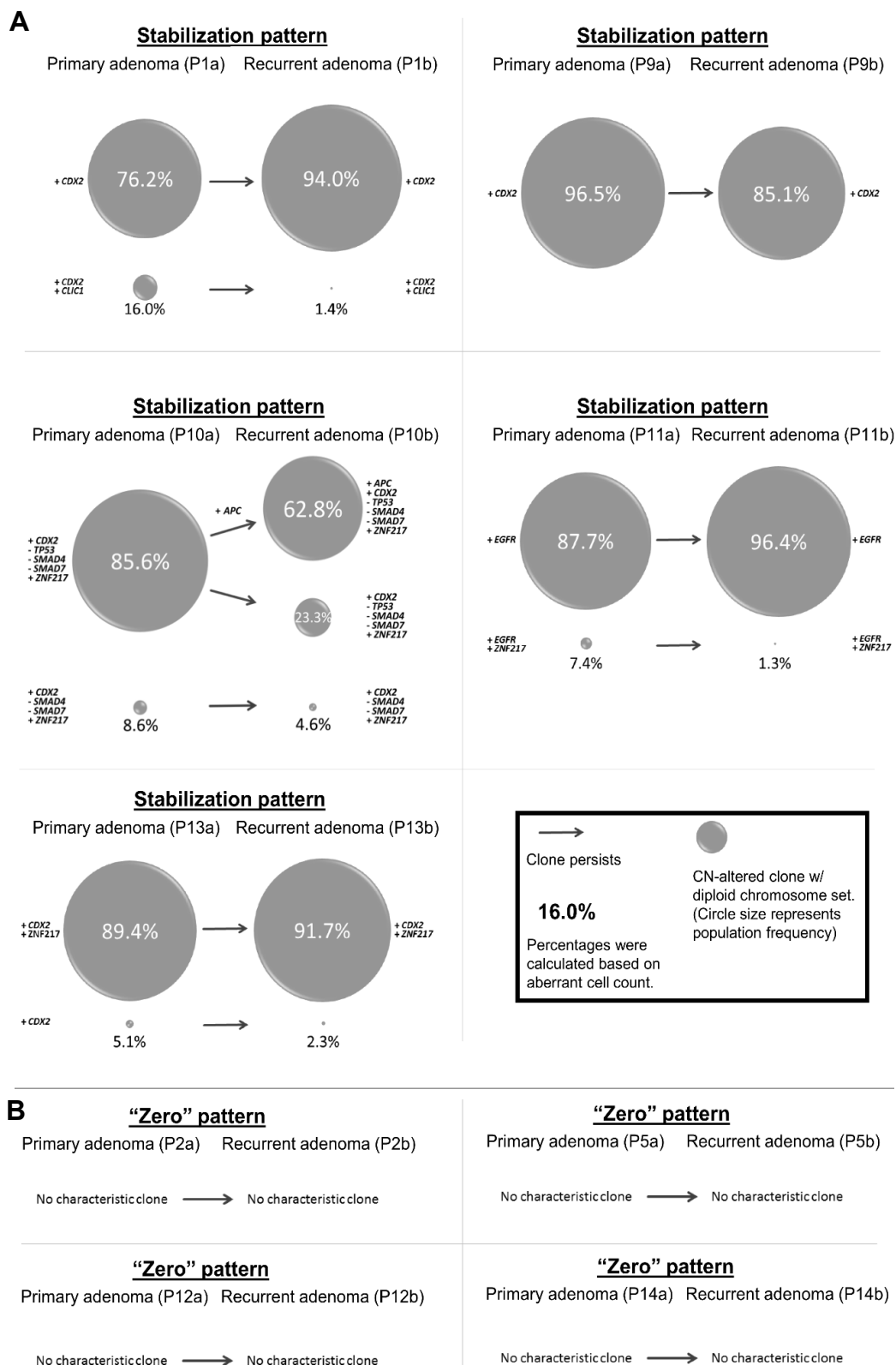
Figure 55. Phylogenetic trees of complexity pattern-adenomas. (**A**) Consensus phylogenetic trees display the clonal distribution across the primary adenoma and the corresponding recurrent adenomas of the pairs P6 and P7, respectively. Nodes (*circle*, *square*, *rhomb*) reflect the copy number profile of the fourteen gene identifier probes and CEP10. Genomic events which led to the formation of the major clone population (*highlighted*) are remarked to the edges. Pairs of the complexity pattern show an (drastically) increasing number of clone populations from the primary to the recurrent adenoma. Notably, pair P7 was affected by a whole genome duplication in the recurrent adenoma. (**B**) Schematic presentation of the complexity clonal evolution pattern. The shift from large major clone populations in the primary tumor towards multiple emerging clone populations in the recurrent adenoma were visible in both cases.
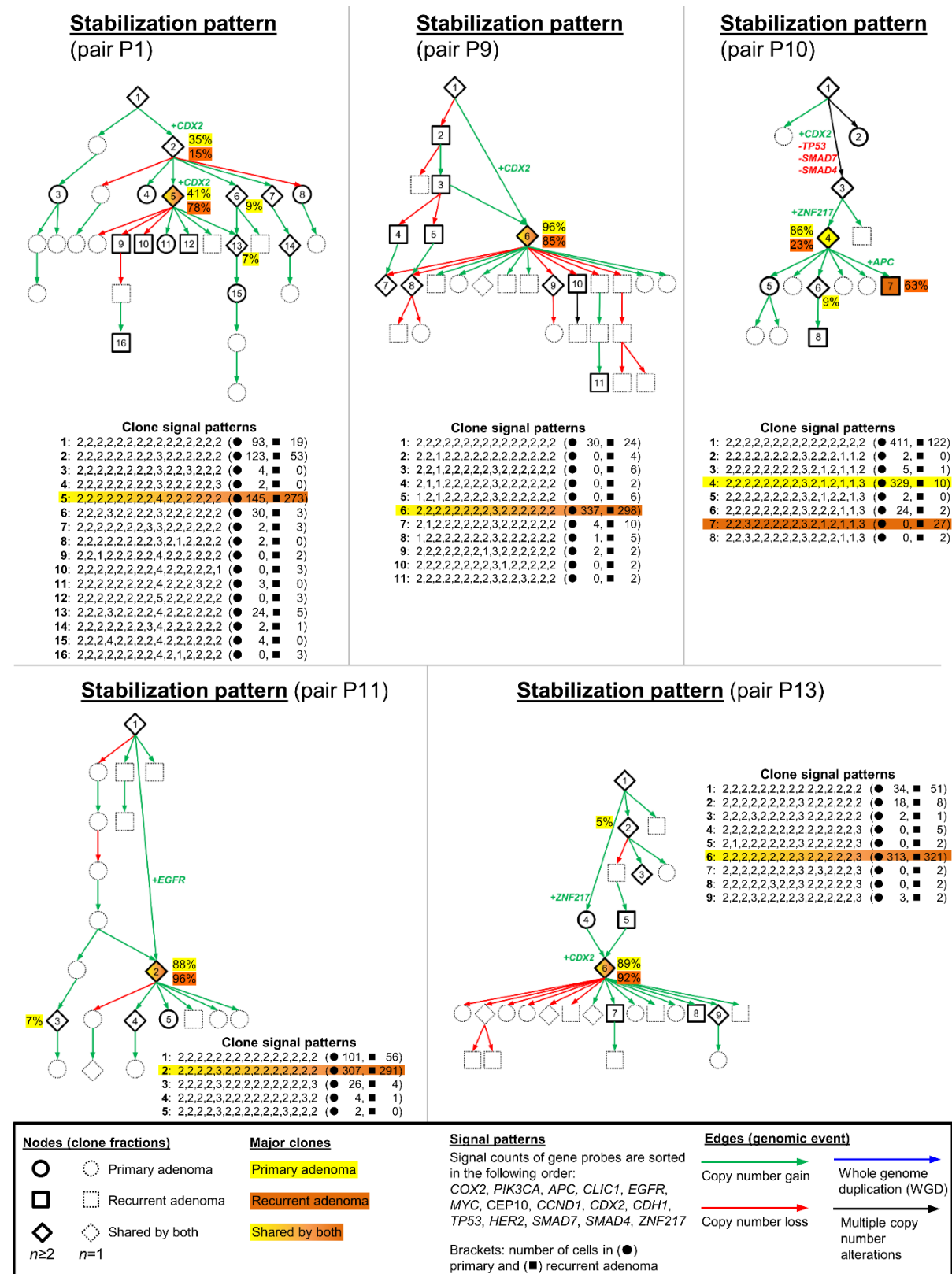
Figure 56. Phylogenetic trees of stabilization pattern-adenomas. Consensus phylogenetic trees depict the clonal distribution along the primary adenoma and the corresponding recurrent adenoma of the pairs P1, P9, P10, P11 and P13, respectively. Nodes (*circle*, *square*, *rhomb*) reflect the copy number profile of the fourteen gene identifier probes and CEP10. Genomic events which led to the formation of the major clone population (*highlighted*) are remarked to the edges. Stabilization pattern-adenomas revealed similar clonal populations within the primary and the recurrent tumor underpinned by shared major clones in both lesions. However, pair P10 was a slight exception due to a clone population which accumulated an additional CNA. Therefore, this matched pair could be more precisely termed stabilization(-progression) pattern.

A

Stabilization pattern

Primary adenoma (P1a)   Recurrent adenoma (P1b)

76.2% → 94.0%
+ CDX2   + CDX2
+ CDX2   + CDX2
+ CLIC1   + CLIC1
16.0% → 1.4%

Stabilization pattern

Primary adenoma (P9a)   Recurrent adenoma (P9b)

+ CDX2 96.5% → 85.1% + CDX2

Stabilization pattern

Primary adenoma (P10a)   Recurrent adenoma (P10b)

+ APC
+ CDX2
+ CDX2   - TP53
- TP53   + APC - SMAD4
- SMAD4   62.8% - SMAD7
- SMAD7   + ZNF217
+ ZNF217
85.6%
                 + CDX2
                 - TP53
            23.3% - SMAD4
                 - SMAD7
                 + ZNF217
+ CDX2   + CDX2
- SMAD4   - SMAD4
- SMAD7   - SMAD7
+ ZNF217 8.6% → 4.6% + ZNF217

Stabilization pattern

Primary adenoma (P11a)   Recurrent adenoma (P11b)

+ EGFR 87.7% → 96.4% + EGFR

+ EGFR   + EGFR
+ ZNF217 7.4% → 1.3% + ZNF217

Stabilization pattern

Primary adenoma (P13a)   Recurrent adenoma (P13b)

+ CDX2 89.4% → 91.7% + CDX2
+ ZNF217   + ZNF217

+ CDX2 5.1% → 2.3%

Clone persists

16.0%

Percentages were calculated based on aberrant cell count.

CN-altered clone w/ diploid chromosome set. (Circle size represents population frequency)

B

"Zero" pattern

Primary adenoma (P2a)   Recurrent adenoma (P2b)

No characteristic clone → No characteristic clone

"Zero" pattern

Primary adenoma (P5a)   Recurrent adenoma (P5b)

No characteristic clone → No characteristic clone

"Zero" pattern

Primary adenoma (P12a)   Recurrent adenoma (P12b)

No characteristic clone → No characteristic clone

"Zero" pattern

Primary adenoma (P14a)   Recurrent adenoma (P14b)

No characteristic clone → No characteristic clone

Figure 57. Schematic presentations of the stabilization and "zero" clonal evolution patterns.
(**A**) Adenoma pairs whose primary tumor clone populations persist in the recurrent tumor were classified as stabilization patterns. Graphs depict populations only which exceeded 5% of the aberrant cells in the primary adenoma. Although pair P10 showed characteristics of the stabilization evolution-type, the recurrent adenoma accumulated an additional CNA of the previous major clone population in the primary tumor suggesting a stabilization(-progression) pattern. (**B**) Matched pairs of colorectal adenomas which did not reveal any CNA across the miFISH panels were, therefore, allocated into the "zero" pattern group.

126
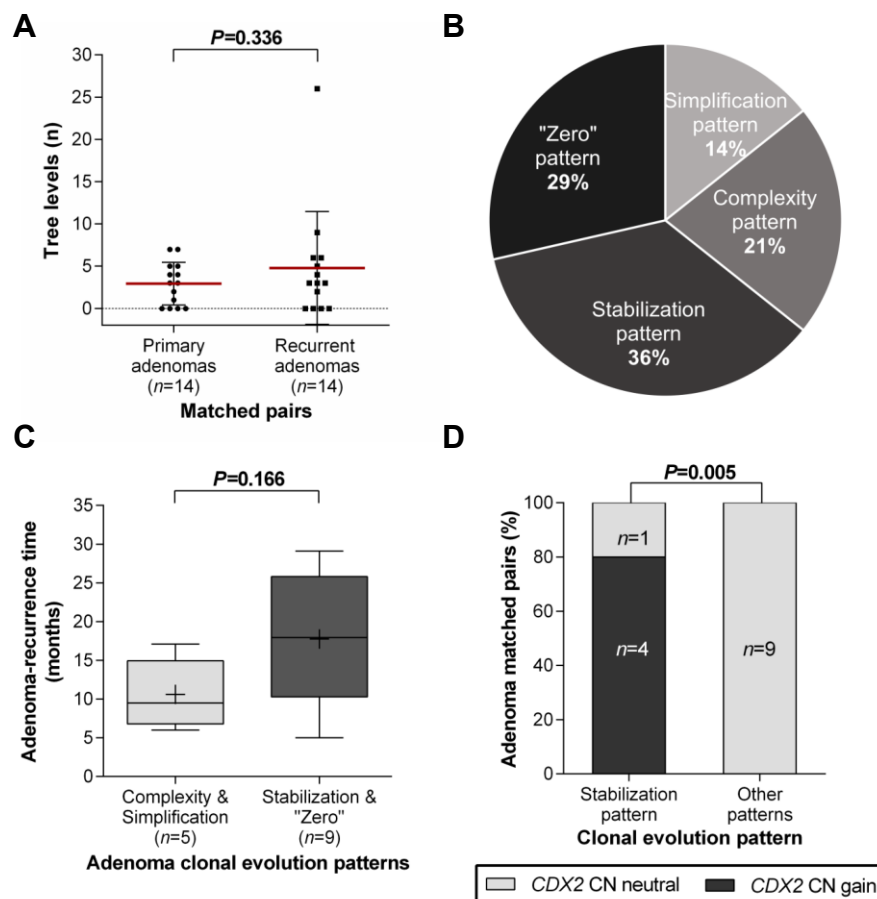
Figure 58. Tree-levels and clonal evolution patterns in matched pairs. (**A**) Phylogenetic non-consensus trees of primary adenomas and recurrent adenomas were compared by the depth of the trees (*i.e.*, tree levels). Recurrent adenomas showed on average (mean ± SD) a statistically not significant higher number of tree levels (Student's *t* test). (**B**) Distribution of adenoma clonal evolution patterns in matched pair adenomas (*n*=14) (**C**) Comparison of adenoma-recurrence time in adenoma cases of clonal evolution patterns which are differential across primary and recurrent adenoma (complexity and simplification patterns) and patterns of "stable" adenoma cases (stabilization and "zero" pattern). Remarkably, non-shifted patterns, *i.e.*, stabilization and zero patterns, weakly tended to be associated with a longer adenoma-recurrence time (Mann-Whitney *U* test). +, mean; whiskers, Tukey. (**D**) Copy number gains of *CDX2* were exclusively observed in adenomas presenting the stabilization pattern (Fisher's exact test).

127

3.3.7 Correlation of signal enumeration and clinicopathological data

The clinical data annotated to the adenoma samples which were investigated by miFISH were analyzed (see Table 23 in chapter 3.3.1 and Figure 37 in chapter 3.3.2). The following passages present the highlights of these correlations per clinical parameter. However, correlations of the clinicopathological features and the signal enumeration expressed by the calculated ASNs were not discussed in detail due to arguably low informative value (Supplemental Table 6).

Size

Adenoma size correlated with the histological subtypes (villous histologic architecture was not considered): smaller adenomas were more frequently observed with tubular histopathology while lesions of larger diameters were associated with tubulo-villous histology (medians 18 mm vs 20 mm, P=0.034; Figure 59A). Furthermore, lesion sizes correlated markedly with the assignment as primary or recurrent adenoma, respectively, in matched pair samples (medians 30 mm vs 15 mm, P=0.003; Figure 59B). Larger diameters were frequently seen in primary adenomas compared to the recurrent adenomas being suggestive of tight surveillance intervals and immediate polypectomy.

Histology

The cohort of primary adenomas investigated in this study comprised three adenomas of villous histology indicating an underrepresentation compared to the other histological subtypes. Accordingly, villous adenomas were combined into one group with adenomas of tubulo-villous histology. The comparison of both histological subtype groups (*i.e.*, tubular and tubulo-villous/villous) revealed that tubular adenomas were once observed in left hemicolonic locations, and not present in the rectum (P=0.059; Figure 59C).

Grading

Separating the primary colorectal adenomas by the grade of dysplasia exhibited a difference in the average number of clone populations within the tumors. Surprisingly, low-grade dysplastic lesions (LGD) frequently showed a higher number of clones compared to adenomas with high-grade dysplasia (HGD; medians 20 clones vs 10 clones, P=0.016; Figure 59D).
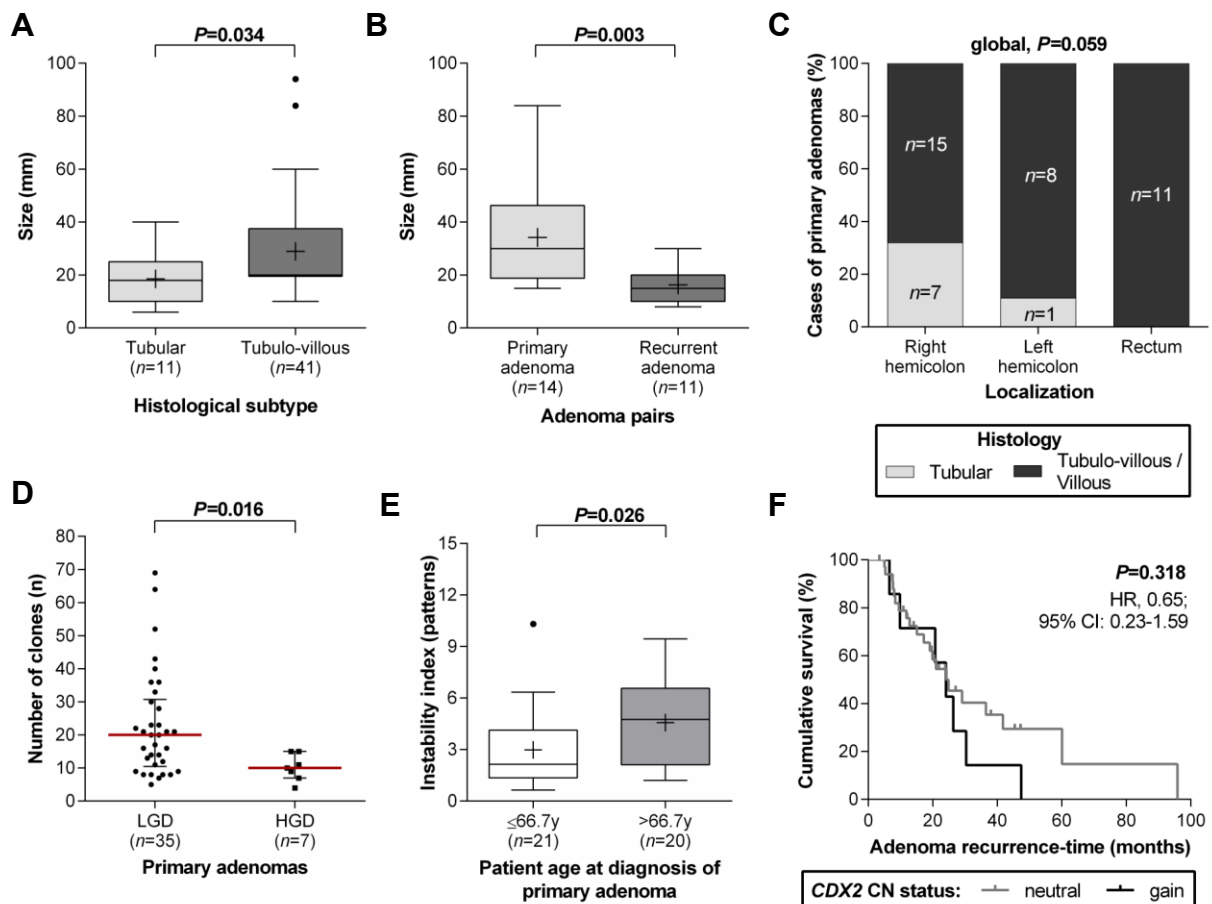
Figure 59. Correlations of clinical features in colorectal adenomas. (**A**) Larger adenoma samples were associated with tubulo-villous histology, whereas tubular adenomas were smaller (Mann-Whitney *U* test). (**B**) The average size of primary adenomas was larger than the corresponding recurrent adenoma size (Mann-Whitney *U* test). (**C**) Histological subtypes were associated with the trichotomized localization of the adenomas (Freeman-Halton test). (**D**) The number of clones (median with IQR) differed in adenomas with LGD compared to adenomas with HGD (Mann-Whitney *U* test). (**E**) Patients of advanced age displayed increased instability indices in the adenomatous lesions than younger patients (Mann-Whitney *U* test). Adenoma P4A was excluded as an outlier. (**F**) Kaplan-Meier analysis of adenoma recurrence-free survival time depending on the copy number status of *CDX2* (χ=0.998, Log-rank test). CI, confidence interval; HR, hazard ratio. Box plots in (**A**), (**B**) and (**E**): +, mean; whiskers, Tukey.

Age at diagnosis

Patient age at the time of primary adenoma diagnosis correlated with the genomic instability index (P=0.026; Figure 59E). An advanced age (above the median of 66.7 years, IQR 62.5-70.2 years) was associated with an increasing instability in the lesions compared to younger patients (medians 4.8 patterns vs 2.1 patterns).

Time to recur

Addressing the question whether the CNA of *CDX2* was associated with an earlier recurrence of adenomas, adenoma samples were analyzed by Kaplan-Meier (Figure 59F). An increase of *CDX2* copy numbers seemed, arguably, not associated with adenoma-recurrence free survival (P=0.318).

Localization

A potential unequal distribution of CNAs in adenomas across various colonic segments was assessed. Separating the large intestine into the colonic and the rectal fraction, respectively, exhibited that adenomas resected from the rectum were frequently affected by CNAs (81.8%) while copy numbers in colonic adenomas remained predominantly neutral (54.8%; P=0.036; Figure 60A). Moreover, copy number gain of *CLIC1* was associated with the dichotomized location: 80% (4/5) of adenomas affected by *CLIC1* gains were found in the rectum (P=0.013; Figure 60B).

Intratumor heterogeneity

Diversities in the composition of intratumor populations were measured by the observed number of divergent signal pattern clone populations. Considering a CNA to be present when more than 10% of the counted cells harbored an alteration, primary adenomas were categorized into groups of copy numbers neutral and altered, respectively. As expected, adenomas with distinct CNAs revealed a higher number of different clonal populations compared to adenomas without CNAs suggesting an increase of intratumor heterogeneity (medians 21 clones vs 11 clones, P=0.006; Figure 60C). Computing an instability index as a measure of diversity helped to assess intratumor heterogeneity while providing intertumoral comparability (see chapter 2.6.4 Data assessment: Tumor heterogeneity). Primary adenomas which harbored a CNA of *MYC* were clearly associated with an increase of the intratumor heterogeneity indicating that this CNA accelerates chromosomal instability (medians 7.9 patterns vs 2.4 patterns, P=0.016; Figure 60D). Moreover, dichotomizing primary adenomas via low or high instability index (median 2.92 patterns, IQR 1.69-6.46) revealed that CN gains of *EGFR* were predominately (80%) in line with a higher intratumor heterogeneity in the cases (P=0.036; Figure 60E). Interestingly, an increased average instability index was also observed in primary adenomas resected from the rectum in comparison to colonic locations (medians 6.0 patterns vs 2.6 patterns, P=0.045; Figure 60F). A similar observation was made when Shannon entropy was compared to the adenoma location confirming the increased intratumor heterogeneity across rectal cases (P=0.029; Figure 60G). The larger the population size of the major clone within an adenoma, the lower was the value of the Simpson diversity index (P<0.001; Figure 60H). Dichotomizing primary adenomas by the median major clone population size (69%, IQR 38-85%) revealed a negative linkage with accumulated pairwise genetic diversity (P=0.014; Figure 60I).

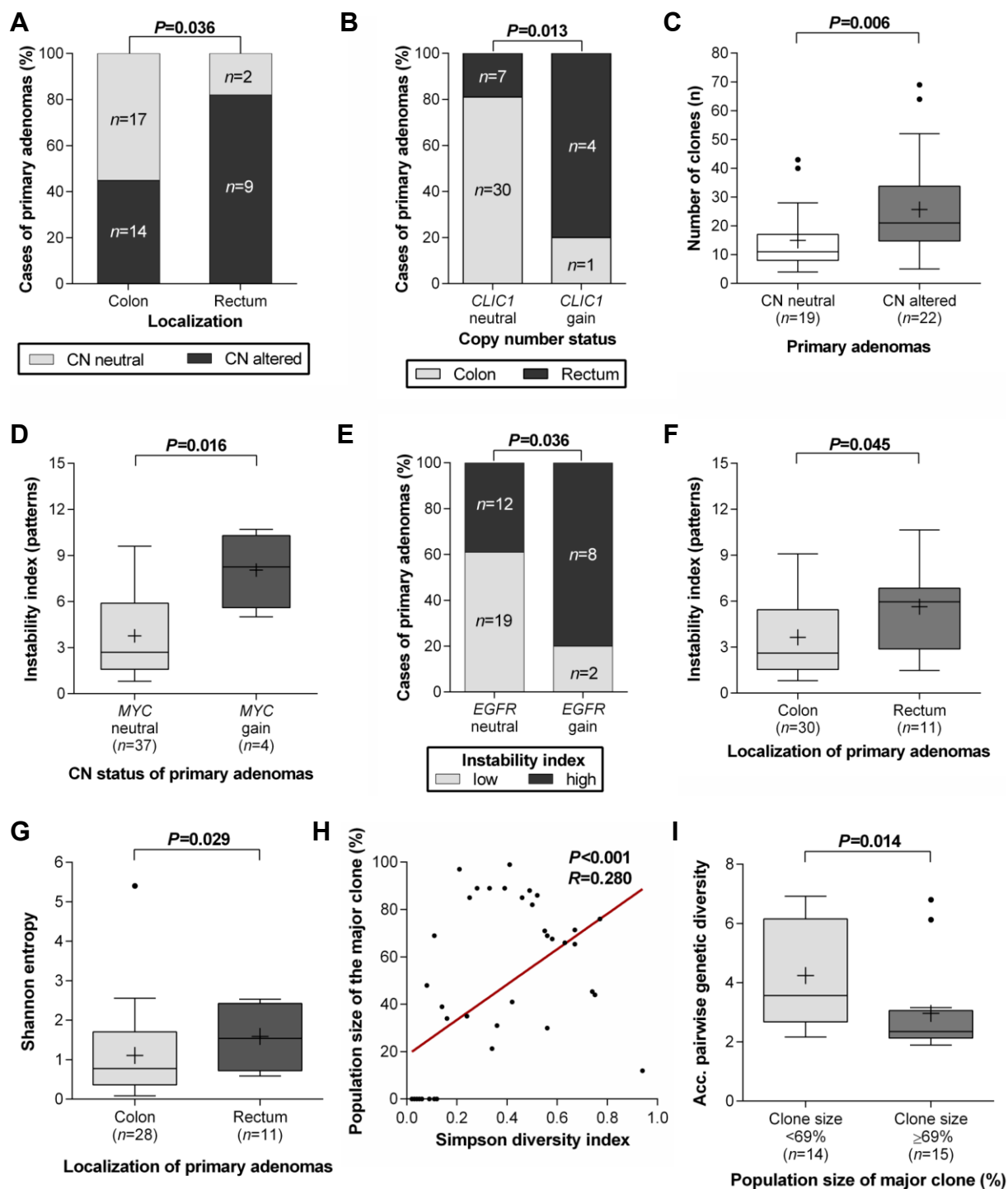**Figure 60.** Legend on next page.

3.4 Comparison of CNAs detected by aCGH and miFISH

The results of copy number counts called by aCGH and miFISH in adenomas were correlated with an applied detection threshold of >10% of the cells for miFISH (Figure 61A). Copy number status of aCGH findings and miFISH counts exhibited a "very good" strength of agreement of 97.3% (753/774 observations) which equaled κ=0.86 (CI, 0.80-0.92; Cohen's kappa coefficient). The few discrepancies in CNA-calls were predominantly observed for gains of *CLIC1* called by aCGH but not seen (in equal numbers) by miFISH. In contrast, CN gains of *CDX2*, *EGFR*, *MYC*, and *ZNF217* were less abundantly observed by aCGH as by miFISH analysis (Figure 61A).

Next, CN profiles and signal patterns in match paired adenomas (*n*=13) detected by aCGH and miFISH were correlated (Figure 61B). Note that aCGH profiles were classified as stable or simplification/complexity due to the lack of clonal information. 84.6% (11/13) of pairs showed superimposable trends when comparing primary and recurrent adenoma by aCGH and miFISH, respectively. However, pair P2 showed a shift from primary adenoma towards an increase of CNAs in the recurrent adenoma by aCGH while miFISH detected no CNAs. However, called CNAs by aCGH were not covered, *i.e.*, represented, by miFISH probe markers. Pair P10 resembled to be a complicated case as the aCGH profile showed differential CN losses of chr14 and chr15, respectively. Any other alteration was shared across both lesions. However, miFISH targeted none of these chromosomes and clonal populations within this adenoma case seemed similar across the primary and the recurrent adenoma.

Both methods call copy numbers of probe markers. However, aCGH data merely provides an average copy number profile (in comparison to a baseline) of the tumor bulk. In contrast, miFISH represents a direct single-cell approach whose quantification is not (ultimately) superimposable to aCGH (Lassmann et al., 2007).
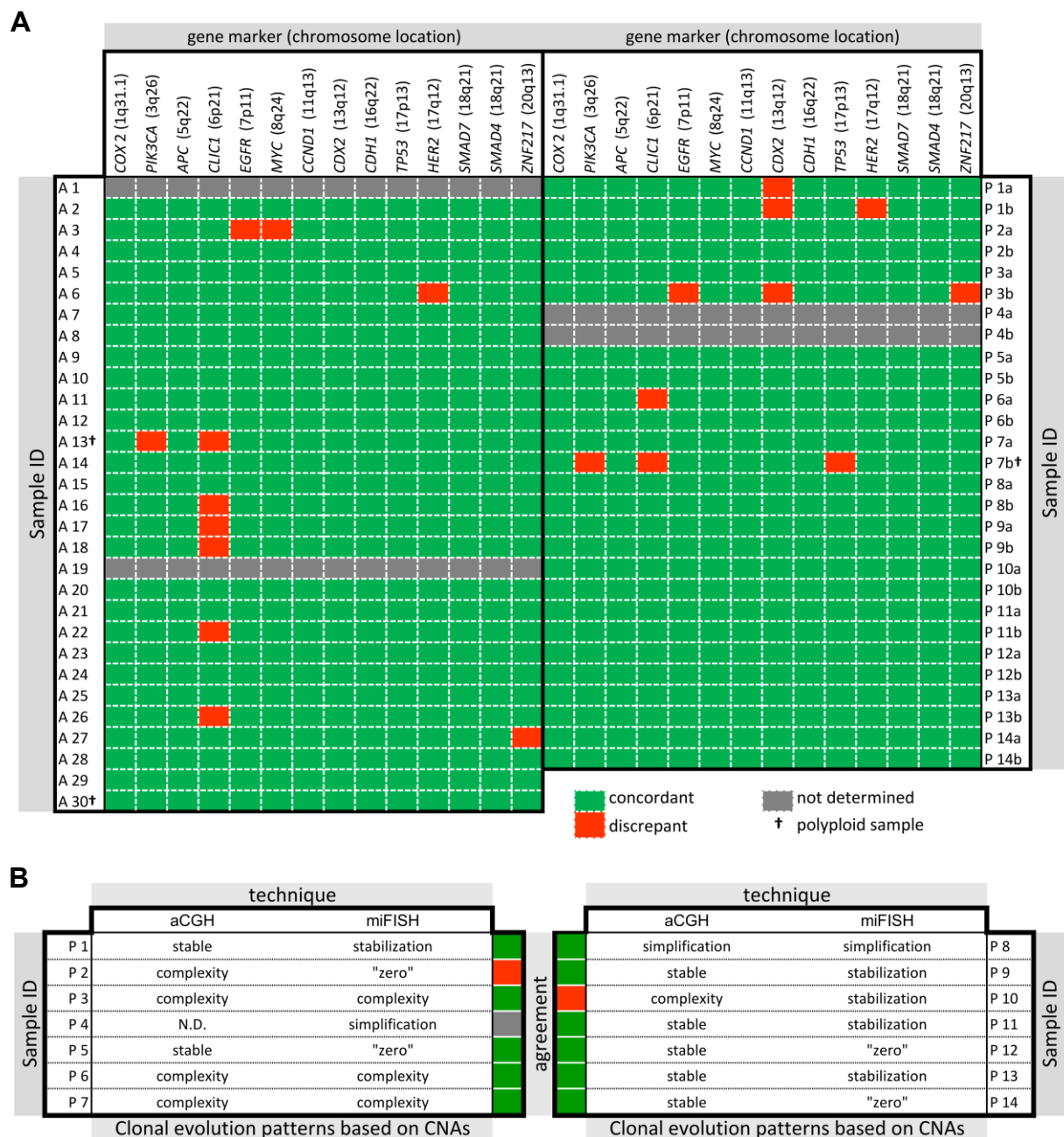
Figure 61. Correlation matrix of copy number results by aCGH and miFISH. (**A**) Obtained results of CNAs by aCGH and miFISH, respectively, were correlated with each other. A match would be concordant (*green*) if both techniques showed the same results concerning the copy number status (gain, loss, neutral). Accordingly, an overlap would be discrepant (*red*) if copy number counts were unequal. CN status was statistically evaluated confirming a very good agreement of 97.3% (κ=0.86, Cohen's kappa coefficient). (**B**) Comparison of clonal evolution patterns determined by aCGH CN profiles and miFISH probe counts, respectively. CN tendencies between the primary and recurrent adenoma were classified as either stable or simplification/complexity via aCGH. Patterns were compared to the miFISH recurrence patterns of stabilization, "zero," simplification and complexity. "Zero" patterns were introduced as no CNA was detected by miFISH and, accordingly, corresponding aCGH profiles are stable. An agreement was marked as either concordant (*green*), discrepant (*red*), or not determined (*dark grey*). N.D.; not determined.

3.5 Protein expression of MMR proteins and CDX2 in (recurring) adenomas

3.5.1 Immunohistochemical detection of microsatellite instability

The cohort of colorectal adenomas analyzed by miFISH comprised a subgroup of 24 samples (19 biological individuals) with no detectable CNAs in any of the fourteen gene identifier probes and centromere 10 probe, respectively (see chap. 3.3). Addressing the question whether these cases were instead affected by microsatellite instability instead of copy number alterations and/or chromosomal instability, expression status of mismatch repair proteins (namely MLH1, MSH2, MSH6, and PMS2) were immuno-histochemically evaluated as indirect evidence of putative microsatellite instability (Figure 62). None of the tested DNA MMR proteins was lost in any of the specimens (n=24; Figure 15): IHC revealed distinct and clear nuclear expression of the proteins targeted by the antibodies (Table 27). Thus, immunohistochemical detection confirmed a microsatellite stable (MSS) phenotype in 100% (24/24) of the tested adenomas.
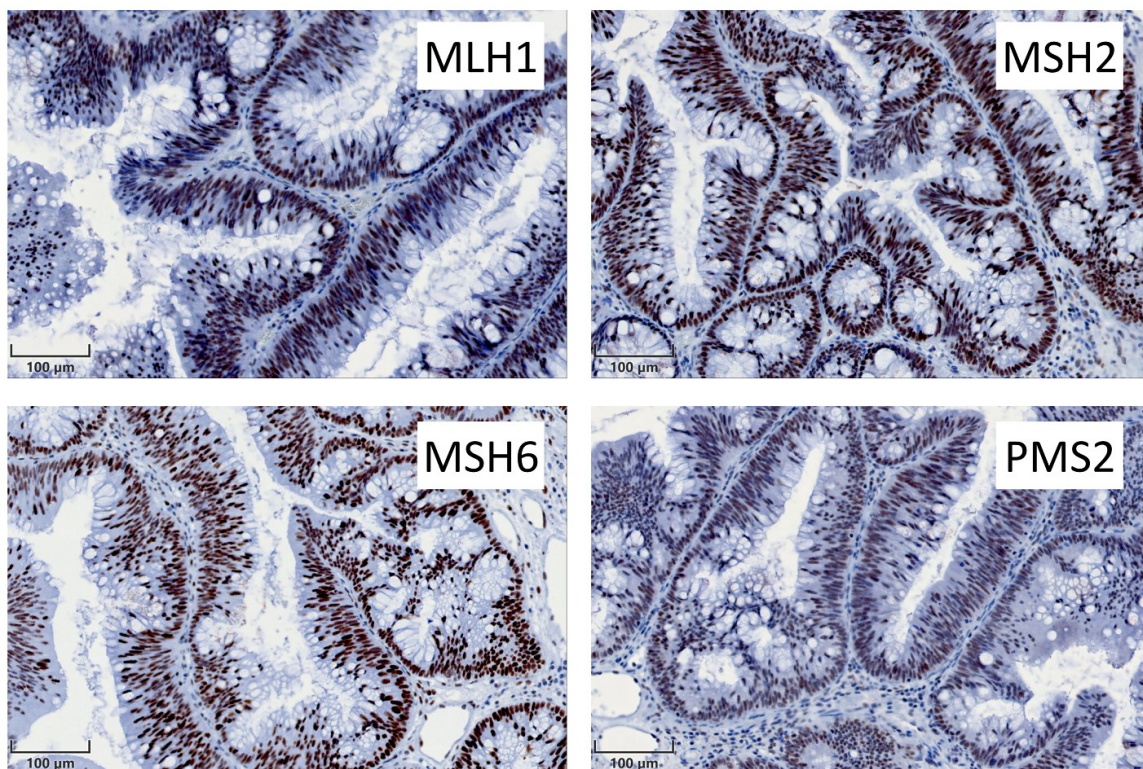
Figure 62. Immunohistochemical detection of MMR proteins in colorectal adenomas. Representative images of immunohistochemistry (IHC) detecting DNA mismatch repair proteins, *i.e.*, MLH1, MSH2, MSH6, PMS2, in adenoma sample P14a. Positive immunoreaction was observed for all four protein markers. Scale bar, 100 µm.

Table 27. IHC of DNA MMR proteins in colorectal adenomas. Adenomas not displaying any CNA by miFISH were tested for DNA MMR protein expression. +, stable nuclear protein expression; MSS, microsatellite stable.

		A 2	A 4	A 5	A 7	A 12	A 17	A 20	A 21	A 24	A 25	A 26	A 29	P 2a	P 2b	P 3a	P 5a	P 5b	P 6a	P 6b	P 8a	P 12a	P 12b	P 14a	P 14b
	Sample ID																								
MMR protein expression	MLH1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	MSH2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	MSH6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	PMS2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
		MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
	DNA microsatellite status																								

3.5.2 CDX2 protein expression

Detected CN gains of *CDX2* in colorectal adenomas raised the question whether these genetic changes have a direct (measurable) impact on CDX2 expression. Therefore, IHC was applied to representative adenomas (*n*=8) comprising neutral and gained *CDX2* CNs, respectively (Figure 63). IHC performance presented abundant nuclear CDX2 expression in 100% (8/8) of samples: not only in the adenomas with CN gain (*i.e.*, A19, P9a, P10a, P13a) but also in the adenomas with neutral CN count for *CDX2* (*i.e.*, A3, A9, A21, P5a). Statistical quantification of expression levels by immunoreactive score was not performed due to the similarly strong intensities of IHC-staining and homogenous staining of tumor cells.
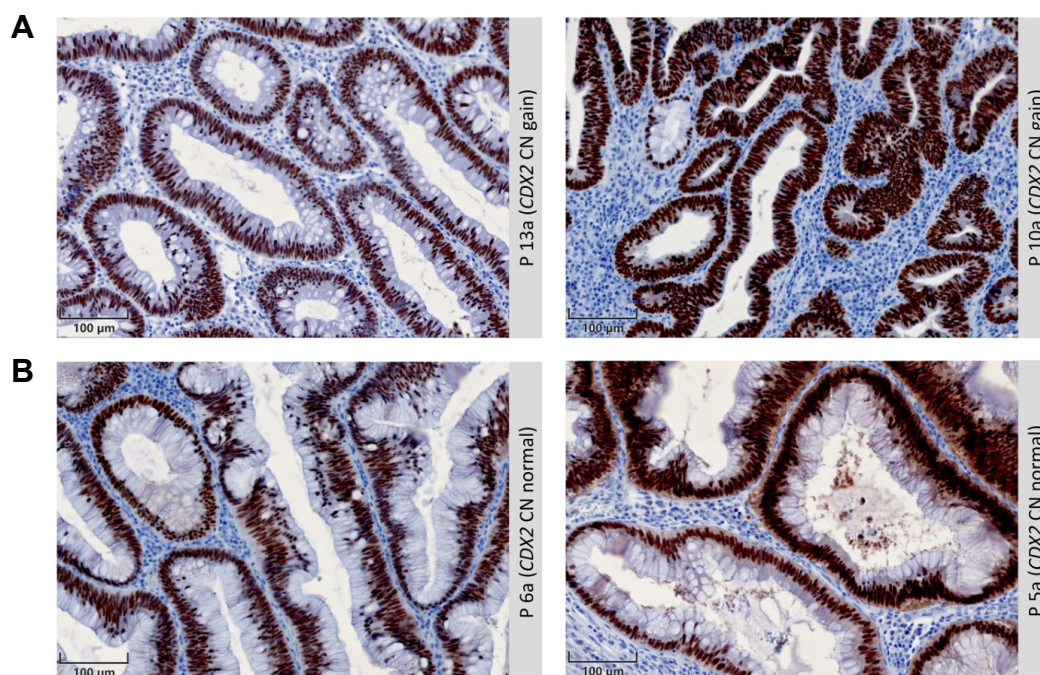
Figure 63. CDX2 immunostaining of primary colorectal adenomas. (**A**) Strongly positive nuclear staining of CDX2 in epithelial cells of two representative adenoma cases with detected copy number gain of *CDX2*. (**B**) Similarly, a distinct expression of CDX2 in epithelial cells of two representative adenoma cases which exhibited neutral copy numbers concerning CDX2. A differential staining level across adenomas with and without CDX2 gain was not detectable. Scale bar, 100 µm.

4 DISCUSSION

4.1 The lack of colorectal adenoma recurrence biomarkers

Colorectal cancer (CRC) is the third most common cancer in men and the second most common cancer in females in Germany in 2013 (Zentrum für Krebsregisterdaten, 2016). Thus, it denotes to be a dramatic health burden for patients and a tremendous cost factor for the health care system. As CRC is believed to develop in an interval of years to decades, a characteristic stepwise transition (from normal mucosa to adenoma and CRC) was proposed: the adenoma-carcinoma sequence (Fearon and Vogelstein, 1990). Representing the most common neoplastic lesion in the colorectum (Eide, 1991; Levine and Ahnen, 2006), adenomatous polyps bear the opportunity to reduce CRC risk to up to 90% by resection, *i.e.*, polypectomy, due to their intermediate role as precursor lesions (Winawer et al., 1993; Zauber et al., 2012). Although less than 5% of adenomas progress into CRCs (Eide, 1991; Shinya and Wolff, 1979), most CRCs indeed arise from adenomas. This circumstance confirms the rationale, and currently performed routine (Leitlinienprogramm Onkologie, 2014), to resect all adenomas after their detection during colonoscopy. However, recurrence of adenomatous polyps affects more than every third patient (Avidan et al., 2002; Bonithon-Kopp et al., 2004; Neugut et al., 1995; Winawer et al., 1993). This elevates both CRC risk and the risk of complications associated with the invasive colonoscopy (Corley et al., 2014). This emphasizes an urgent need for the identification of epigenetic and genetic biomarkers to infer and predict the individual risk for adenoma recurrences ("personalized medicine").

4.2 The role of CpG methylation in adenoma recurrence

4.2.1 DMPs in colorectal adenomas

The genetic landscape of cancers such as CRC is an extensively studied field ever since alterations were associated with cancer. It is well accepted that cancer is not caused by a "single hit" but rather by multiple hits, *i.e.*, alterations (Knudson, 1971; Vogelstein et al., 2013). Since epigenetic phenomena were recognized to alter gene product expression without changing the DNA sequence, more and more studies shed light on epigenetic landscapes. Epigenetic variation in cancers is much higher than genetic variation (Bennett-Baker et al., 2003). The DNA methylation on CpG dinucleotides, an epigenetic modification, is a frequent target of alterations in colorectal neoplastic lesions (Esteller, 2008; Fadda et al., 2018; Galamb et al., 2016). DNA

methylation patterns were shown to be highly dynamic with specific variance across tissues, cells and diseases, respectively (Sharma et al., 2016). In the advent of studies to identify and validate specific biomarkers, differential DNA methylation consequently became the focus of attention due to their pivotal role in gene expression. Advances in array technologies such as HM450K then allowed the simultaneous analysis not only of multiple CpG loci but also of numbers of CpGs on a genome-wide scale, *i.e.*, the global methylome (Irizarry et al., 2009; Bibikova et al., 2011).

Dozens of studies focused on unveiling epigenetic biomarkers and differentially methylated CpG positions (DMPs) across the methylome of CRC, adenomas and normal epithelium (Lam et al., 2016). To this day there is still an imperative for informative epigenetic biomarkers. This study sought to discover aberrantly methylated CpG dinucleotides in the methylome of primary colorectal adenomas without recurrence (*n*=30) compared to primary colorectal adenomas with recurrence (*n*=29), recurrent adenomas (*n*=10) and normal epithelium (*n*=3). By utilizing the HM450K array, 329,573 probes were available after filtering to seek differences across the sample groups. However, principal component analysis and unsupervised hierarchical clustering of mean β-values were not capable of separating the specimens by recurrence groups. Cluster assignment by mean β-values was, instead, associated with adenoma histology (*P*=0.008). This finding leads to the assumption that the observed epigenetic alterations also manifest themselves in phenotypical variation. A previous study linked adenoma location, *i.e.*, right- or left-sided, to distinct differential methylation patterns (Koestler et al., 2014). This observation did not account for the adenoma samples analyzed in our study. Remarkably, Luo and colleagues (2014) identified two epigenotypes of adenomas defined by low or high methylation frequency. The authors additionally linked the high methylation phenotype significantly to adenomas progressing to CRC. Such a separation was not found in our study.

Nevertheless, the different mean β-value methylation patterns for the different histologies (tubular or tubule-villous/villous) suggest a crucial role in pathogenesis. This epigenetic evidence substantiates previous linkages of recurrence with histologic features to assess colonoscopy surveillance intervals (Levine and Ahnen, 2006; Noshirwani et al., 2000). In contrast, other studies did not find an association of histology with recurrence and, instead, reported adenoma number and size as risk factors (Avidan et al., 2002; Bonithon-Kopp et al., 2004; Martinez et al., 2001).

Besides average β-values, we sought to unveil DMPs by comparing the grouped adenomas. Significant DMPs were not identified when comparing primary adenomas without recurrence to primary adenomas with recurrence. This initially suggests that the methylation of CpGs is not informative in predicting adenoma recurrence. One explanation for the observation could be the broad level of heterogeneity in epigenetic patterns across the primary lesions irrespective of recurrence status. Arguably, while recurrence might be triggered by alterations in DNA methylation patterns in several patients, the individual impact of DNA methylation across primary adenomas and patients could (*i.e.*, ITH) prevent the discovery of DMPs. Nevertheless, a previous study found the promoters of *MLH1*, *ATM*, and *FHIT* to be hypermethylated in adenomas with recurrence (Rengucci et al., 2014; De Maio et al., 2017). While this seems contradictory to our findings, their results must be interpreted with caution as hypermethylation was considered when methylation exceeded 20% (evaluated by methylation-specific ligation probe assay). Additionally, FDR was not corrected for multiple testing. The failed appearance of DMPs in our dataset of primary adenomas with recurrence versus adenomas without recurrence could, theoretically, be due to incomplete resection and not caused genetically. In this scenario, adenomatous polyps recur out of a remaining residue, and the epigenetic pattern will not reveal substantial differences compared to non-recurring adenomas. Corley et al. (2014) named incomplete removal, in fact, a biasing factor. Absent identification of DMPs also draws the conclusion that differential DNA methylation might be exclusively associated with serrated and CIMP[+] lesions (serrated pathway of carcinogenesis) (Andrew et al., 2017; Conesa-Zamora et al., 2015), and not with sporadic adenomatous polyps.

Similarly, the comparison of matched pairs did not reveal significant DMPs. This suggests that ITH combined with small sample counts (*n*=10 pairs) were, arguably, responsible in this matter. However, recurrent adenomas were aberrantly methylated (FDR<0.05 and FC>10%) in 674 and 221 CpG sites compared to the set of primary adenomas with recurrence (*comparison A*) and the complete set of primary adenomas (*comparison B*), respectively. Results demonstrated that a significant portion of recurrent adenomas underwent an epigenetic shift towards genomic hypermethylation. According to the scientific consensus, genome-wide methylation decreases while CGI methylation increases as tumors progress (Esteller et al., 2008). As most of the assessed hypermethylated DMPs are located within CGIs or nearby, *i.e.*, shores and shelves, the results suggest adhering to this biological rationale. This is supported by the fact that recurrent adenomas were resected at later time points. This argues in

favor of a dynamic progression of the landscape of methylation alterations which is in concordance to age-related changes (Ahuja et al., 1998). Speculatively, the procedure of endoscopic removal might pave the ground for hypermethylated adenomas.

The list of *top genes* across recurrent adenomas versus primary adenomas with recurrence provided 34 DMPs associated with 31 individual genes. These comprised, *e.g.*, *STK32B* (serine/threonine kinase 32B), *LCMT1* (leucine carboxyl methyltransferase 1), *MCC* (mutated in colorectal cancers), *JAK3* (Janus kinase 3), *BMP3* (bone morphogenetic protein 3), and *GREM2* (gremlin 2). Although methylation of *BMP3* was previously associated with neoplastic lesions of the colon via stool detection (Ashktorab et al., 2014; Kisiel et al., 2013), solely *GREM2* harbored three DMPs within a CGI of the gene body. Thus, the gene resembles a preferential validation candidate. *GREM2*, expressed in basal epithelial stem cells of the crypt (Roy and Majumdar, 2012), is a member of the DAN gene family of bone morphogenetic protein (BMP) antagonists which belong to the TGF-β superfamily. Due to BMPs' involvement in development, morphogenesis, cell proliferation and apoptosis (Bragdon et al., 2011; Pelli et al., 2016), a dysregulation of these proteins by antagonists can promote both colorectal polyp and cancer formation (Hardwick et al., 2008; Kodach et al., 2008). Pyrosequencing of two of these three DMPs validated the differential methylation in primary adenomas and recurrent adenomas. Additionally, more informative CpG sites located within the DMPs were unveiled. Functionally, hypermethylation of CpGs in the gene body of *GREM2* in recurrent adenomas may supposedly deactivate the gene transcription resulting in a decreased protein expression. Research on stored datasets in the Oncomine database (2014) documented a highly significant downregulation of *GREM2* mRNA in colorectal adenomas compared to normal mucosa (Sabates-Bellver et al., 2007). While *GREM2* mRNA was strongly expressed in myofibroblasts, a complete knock-down was observed in CRC cell lines (Kosinski et al., 2007). As previously hypothesized, this could result in a deregulation of BMP signaling. In turn, WNT signaling is disrupted and promotes colorectal carcinogenesis (Kosinski et al., 2007).

Comparing colorectal adenomas (*n*=69) to the normal colorectal epithelium (*n*=3) unveiled 9,266 significant DMPs (FDR<0.05) of which 73% were hypomethylated. Filtering led to 4,179 DMPs (FDR<0.05 and FC>10%) in this setting. Another study identified 43,999 DMPs when comparing adenomas to normal mucosa (Fadda et al., 2018). However, the count of tumor samples was less than a third (*n*=21) compared to the here presented study. Our observation of genome-wide hypomethylation of adenomas is in line with the findings by Fadda and coworkers (2018). Additionally, their suggested

panel of methylation markers, comprising *GRIA4*, *SLC8A1*, and *SYN3*, for detection of CRC and adenomas (in stool) was confirmed in our approach. We discovered significant DMPs associated with the respective set of genes. Nevertheless, solely *GRIA4* ranged in the list of 347 *top genes* identified in our work while the other markers did not show Δβ>0.1. Furthermore, there are currently two methylation markers commercially available for noninvasive detection of colorectal adenomas and cancer (Gyparaki et al., 2013; Sameer and Nissar, 2016), namely *SEPT9* (ColoVantage®) and *VIM* (ColoSure™). Only the latter was among the list of 4,179 significant DMPs, although Δβ was lower than any of the 347 *top genes.* Moreover, from a set of ten cancer-related genes frequently aberrantly methylated in CRC (Sugai et al., 2017), probes located in *SFRP1*, *DKK3*, *IGFBP7*, and *MLH1* were found among the significant DMPs in our study. This indicates an early alteration of methylation markers in the carcinogenesis. Interestingly, Galamb and colleagues (2016) identified differential methylation of WNT pathway genes in adenomas versus normal mucosa. Of these genes, our study confirmed *AXIN2* (two probes among *top genes*) and *SFRP1*. These results underline that the list of *top genes* harbors potential for informative CpG dinucleotide methylation biomarkers for the detection of adenomas.

4.2.2 Distribution of DMPs in the genome

The literature first mentioned the methylation of cytosines, *i.e.*, resulting in 5-methylcytosines. This epigenetic modification was discovered to function as a dynamic regulatory element involved in gene repression. Then, studies primarily reported on CpG dense regions (CpG islands) in gene promoters being increasingly methylated (Saxonov et al., 2006). In recent years, attention has shifted towards epigenetic alterations in non-promoter regions with an even sparser density of CpG dinucleotides in neighboring regions of CGIs. Indeed, these regions were positively linked to gene transcription (Esteller, 2008; Jones, 2012; Price et al., 2013). The high-throughput DNA methylation array HM450K assessed CpGs not only located in proximal promoter regions such as upstream of the transcription start sites, 5'UTR, and 1st exon but also in the gene body and 3'UTR (Bibikova et al., 2011). Besides gene regions, CpG dinucleotides distributed in neighboring regions up- and downstream of CpG islands, *i.e.*, the shore and shelf regions and the open sea, can be targeted (Bibikova et al., 2011).

In the present study, a non-random distribution of hyper- and hypomethylated CpGs was uncovered when investigating the set of 4,917 DMPs. In the comparison of

recurrent adenomas versus primary adenoma or tumor versus normal, one-third of the DMPs were located within a CpG island and hypermethylated, while 24% of DMPs were localized in the open sea and hypomethylated. This observation is in line with the current understanding of alterations in tumors (and their progression). According to Esteller (2008), the genome-wide level of 5-methylcytosines decreases whereas the level of methylation in CGIs, conversely, increases. Comparing recurrent adenomas to primary adenomas with recurrence (*comparison A*) showed a fraction of 47% of DMPs being in CGIs, whereas recurrent adenomas compared to all primary adenomas (*comparison B*) showed 66% of DMPs being in CGIs. In contrast, *comparison C* (adenomas vs normal mucosa) just revealed 37% of DMPs being in CGIs. This could indicate that aberrant methylation of CGIs inclines from normal epithelium to adenomatous lesions with a linkage to recurrence. Our study unveiled broad hypomethylation of DMPs in open sea and shelf regions, whereas CpG island-DMPs were mostly hypermethylated and DMPs in shores were found to be balanced. Since recent analysis has shown that aberrant methylation of shore regions occurs in colorectal lesions (Irizarry et al., 2009), a vast spectrum of CpG positions may account for imbalances in gene regulation.

Analysis revealed that DMPs located within the proximal promoter were excessively hypermethylated. Conversely, the majority of DMPs found in the gene body and 3'UTR were hypomethylated. This observation indicates a crucial role of methylation in the regulation of transcription. Previous studies have shown that transcription start sites, 5'UTRs and 1st exons are more prone to be hypermethylated than other regions (Hinoue et al., 2012; Naumov et al., 2013; Sharma et al., 2016). Hypermethylation of CpGs in the 1st exon is strongly correlated with gene silencing (Brenet et al., 2011). In combination with our findings, it supports the assumption that recurrent adenomas were affected by more substantial degrees of gene silencing than primary adenomas with and without recurrence. However, DMPs in gene body regions have a tremendous relevance in colorectal lesions as more than 90% of WNT pathway-associated DMPs were not located within the proximal promoter (Galamb et al., 2016).

The distribution of DMPs with regards to their chromosomal location unveiled that more than 20% of DMPS were located on chromosome 1. Although this observation was similarly made in CRC cell lines (Sandoval et al., 2011), chromosome 1 is also the largest and gene-densest chromosome which arguably increases the odds of a biased DMP-identification. However, the results also suggest that DMPs were non-randomly distributed throughout the genome with chromosomes 2, 6, 11, 12, 17 and 19 harboring more differential CpGs than others such as chromosomes 9, 13, 18, 21 and 22.

While the low gene content is a likely explanation for the few DMP counts on chromosomes 13, 18 and 21 (Schneider et al., 2016), other frequencies did not necessarily correlate with chromosome size.

4.2.3 Association of DMPs with distinct pathways

Genes associated with DMPs and identified in *comparisons A-C* were subject to a pathway ontology analysis with stringent criteria to unravel the most significant association between molecular and cellular pathways. Identified genes of *comparison A* were involved in processes of the (adaptive) immune system, cytokine signaling in the immune system and inflammation-mediated cytokines such as IL2 and IL4. Further, identified genes were also related to antigen processing, *i.e.*, class I MHC-mediated antigen processing presentation and antigen processing-cross presentation. This highly significant accumulation of genes related to immunologic processes leads to the assumption that the microenvironment of the recurrent adenomas was dominated by inflammatory conditions in comparison to the primary lesions. Evidence confirms the promotion of tumor development by cancer-related inflammation aiding in the proliferation and survival of malignant cells (Mantovani et al., 2008). Porta and colleagues (2009) termed these conditions "smoldering inflammation" which, to their knowledge, is supported by the presence of myelomonocytic cells of the immune system secreting pro-inflammatory cytokines. The differential methylation of immunologically-related genes could both repress or activate gene transcription with respect to the genomic location of the CpG site. However, the findings are in line with observations of increased tumor rates in patients suffering from chronic inflammations of the bowel, *i.e.*, ulcerative colitis and Crohn's disease (Sartor, 2006). Hanahan and Weinberg (2011) count chronic inflammatory conditions as one of the hallmarks of cancer.

The second class of pathways was associated with (*i*) the proteasomal degradation of proteins, namely the ER phagosome pathway, (*ii*) ubiquitin-mediated proteolysis and (*iii*) antigen processing ubiquitination proteasome degradation. Aberrant methylation in these genes might be an indicator of a disruption in the proteolytic system of recurrent adenoma cells. Deregulation of the proteasome and ubiquitin ligases results in uncontrolled proliferation and inclined genomic instability in cancers (Nakayama and Nakayama, 2006). These are hallmarks which likely promote the recurrence of an adenomatous polyp. Interestingly, the cell cycle is activated by cyclin-dependent kinases (CDKs) which are strictly regulated by ubiquitin-mediated proteolysis (Nakayama and

Nakayama, 2006). By considering the aberrant methylation of genes of the cell cycle and mTOR pathway, the analysis suggests a tumor micro-environment beneficial to proliferation by cell-cycle disruption, growth promotion, and modulation of inflammatory responses (Ciechanover et al., 2000). Third, DMPs associated with genes of the insulin pathway were identified suggesting a correlation between colon neoplasia and diabetes, which was previously named as a risk factor for CRC (Tsilidis et al., 2015). Insulin production is common in neoplastic transformation (Ashktorab et al., 2014), which hints towards a more advanced methylation pattern of recurrent adenomas. Interestingly, insulin signaling was found to influence the oncogene *Ras* (Giovannucci et al., 2010).

Finally, pathways were unveiled which are associated with aberrantly methylated genes across adenomas versus normal mucosa. Multiple genes involved in the formation of the matrisome, the core matrisome, extracellular matrix glycol-proteins, and matrisome associated pathway were linked to the DMPs. The extracellular matrix (ECM) is responsible for the development and homeostasis of tissues by influencing cellular growth, survival, migration, and differentiation (Pickup et al., 2014). Consequently, dysregulation of the matrisome (sum of ECM constituents) was shown to be a critical initiation factor for neoplastic lesions as these ECM proteins are empowered to modulate and accelerate the multifaceted hallmarks of cancer (Hanahan and Weinberg, 2011; Pickup et al., 2014). Besides ECM pathways, genes involved in heterotrimeric G-protein coupled receptor (GPCR)-ligand binding, GPCR signaling and GPCR downstream signaling were strongly positively correlated with the DMPs. GPCRs are linked to various intracellular processes, *e.g.*, EGFR-mediated growth signals, and heterogeneous expression of these receptors was demonstrated in multiple cancer entities (Bhola and Grandis, 2008). Remarkably, multiple genes were listed in our comparison which are members of pathways in cancer, MAPK signaling, Ca^{2+} signaling, p53 signaling and CREB-signaling via PKC and MAPK. These pathways are classical molecular signaling targets in the transition of normal tissue towards the formation of a neoplastic tumor. This highlights their importance as intracellular receptors are linked to growth signaling, cell survival, tumor-repression and cellular homeostasis (Bonni et al., 1999; Hanahan and Weinberg, 2011; Vogelstein et al., 2000). Comparing DNA methylation of adenomas with normal mucosa revealed genes which were associated with the immune system. This fact suggests that an inflammatory microenvironment is beneficial for the transition towards a neoplastic lesion (Mantovani et al., 2008).

4.3 The role of CNAs in adenoma recurrence

Unraveling the molecular dynamics leading to recurrence of colorectal adenomas is a prerequisite for defining valuable biomarkers for prediction. While the traditional perspective of clonal evolution in cancer has been reflected by gradualism (Nowell, 1976), evidence suggests that subclonal populations preferably caused ITH. These emerge in branched patterns which occur with highly "variable evolutionary tempos" (Burrell and Swanton, 2014; Burrell and Swanton, 2016; Cross et al., 2016). The advent of single-cell approaches has unveiled the broad existence of ITH which is crucial for tumorigenesis and carcinogenesis (Heselmeyer-Haddad et al., 2012; Heselmeyer-Haddad et al., 2014; Wangsa et al., 2016). A dominant role in promoting ITH of many solid tumors attributes to CNAs, *i.e.*, gain or loss of whole chromosomes or fractions, leading to chromosome instability (CIN) (Cross et al., 2016). As karyotypical aberrations are frequently identified in colorectal lesions (Bomme et al., 1994; Muleris et al., 1994; Ried et al., 1996), this study has been conceived to analyze the potential of CNAs as a significant genetic signature which distinguishes primary adenomas by their status of recurrence. Moreover, the aspect of a putative dynamic landscape of CNAs in the clonal evolution of match-paired primary adenomas compared to recurrent adenomas has been a matter of analysis with an emphasis on ITH.

4.3.1 Characteristic patterns of CNAs in colorectal adenomas

Conducting aCGH on the set of colorectal adenomas unveiled the global landscape of CNAs present in the tumor bulk per patient. The analysis exhibited copy number gains most frequently affecting whole-chromosomes and chromosome arms, namely 7, 13q and 20q. Additionally, the most frequently reported copy number loss was documented for chr18. These aberrations are generally in line with the characteristic patterns of genomic copy number landscapes in colorectal neoplastic lesions and the carcinogenesis (Bomme et al., 1994; Borras et al., 2016; Meijer et al., 1998; Quintanilla, 2017; Ried et al., 1996). The results underpin the pivotal role of these aberrations as early events, *i.e.*, frequently observed in low-grade adenomatous polyps, in the adenoma-carcinoma sequence (Fearon and Vogelstein, 1990). Although extensive chromosomal alterations do not alter single genes only, chromosomal gain tremendously impacts the transcriptional dosage of oncogenes (Cross et al., 2016; Murugaesu et al., 2015). Defined as "evolutionary karyotypic theory" (Stepanenko and Kavsan, 2012), a dysregulation of copy numbers of whole-chromosomes and chromosome-arms affects

hundreds of genes and induces changes of both the transcriptome and proteome (Cross et al., 2016). CN gain of chromosome 7 comprises well-known proto-oncogenes which modulate Ras- and WNT signaling, *e.g.*, *EGFR* and *MET* (Flora et al., 2012; Roberts et al., 2002; Zeng et al., 2008). Chromosome 13 CNAs are, similarly, mapped to colorectal lesions by alteration of proto-oncogenes regulating β-catenin activity, *i.e.*, *CDK8* (Firestein et al., 2008), and epithelial growth, *i.e.*, *CDX2* (Bai et al., 2003; Salari et al., 2012). Analogical to the aforementioned CNAs, chromosome-arm 20q harbors various oncogenes, *e.g.*, *AURKA*, *ADRM1*, *NCOA3*, and *ZNF217*, which are characteristic for the colorectal carcinogenesis due to altering effects on cellular growth and migration (Carvalho et al., 2008; Habermann et al., 2013; Hirsch et al., 2012; Quinlan et al., 2007). Conversely, copy number loss of chromosome 18 has a crucial impact as the chromosome accommodates TSGs such as *SMAD2* and *SMAD4* which are regulators of the TGF-β pathway and, therefore, repress tumorous growth (Fearon and Vogelstein, 1990; Heldin et al., 1997). Whereas previous studies on colorectal adenomas and carcinomas had also defined chromosome 8 as a recurring CNA in colorectal carcinogenesis (Grade et al., 2006, Ried et al., 2012), adenomas investigated in this study were not common targets of chr8-associated aberrations. This cicumstance supports the perception of the gain of chr8 as a late event in the colorectal tumorigenesis which might be associated with CRC and increased malignancy. As a side note, chromosomal aberrations affecting chr19 were recurrently observed in the present and previous studies on colorectal adenomas. However, Jones et al. (2007) have reported that probes targeting this chromosome might be error-prone and would have introduced artifactual bias, justifying their non-consideration.

Besides alterations of whole-chromosomes and chromosome-arms, the most frequent CNAs in colorectal adenomas were identified to be three focal CNAs:

(*i*) the gain of 6p22.1-21.33 harboring a cluster of histone genes, namely *HIST1H*, whose alteration was recently associated with CRC (Brosens et al., 2010; Camps et al., 2008). Additionally, the focal gain also affects *CLIC1* (chloride intracellular channel protein 1) which was identified as a putative biomarker in CRC (Petrova et al., 2008). Santos and collaborators (2007) linked aberrations within the entire region of 6p21-p23 with growth advantages and tumor progression in several cancers;

(*ii*) the gain of 7q22 harboring a cluster of membrane-bound mucin genes, *i.e.*, *MUC3, MUC11, MUC12,* and *MUC17*. Their proteins contain an EGF-like domain, suggesting a role in epithelial growth modulation in colorectal pathogenesis (Byrd and Bresalier, 2004; Walsh et al., 2007);

(*iii*) the loss of 16q21 comprising *CDH11* (cadherin 11) which was previously reported as a silenced pro-apoptotic tumor suppressor modulating WNT signaling in colorectal lesions (Li et al., 2012; van Roy, 2014).

The distribution of CNAs across the colorectal adenomas was investigated. The results demonstrated a Poisson distribution of CNAs frequencies with a few adenomas exhibiting higher numbers of CNAs. This observation alludes to the existence of CIN in only a small fraction of adenomas, which presented an elevated frequency of CNAs. This pairs well to recent findings by Jones and colleagues (2007) who concluded that CIN occurs in only a few cases of adenomas. It is tempting to speculate whether the tetraploid adenomas were first affected by CIN leading to aneuploidy (Sheltzer et al., 2011), or *vice versa* (Yuen, 2010). Notably, the unequal distribution of whole- and part-chromosome alterations implicates the presence of CIN (McGranahan et al., 2012).

The average number of copy alterations (ANCA) equaled 3.5 CNAs in the adenoma cohort, which is comparatively low but in good agreement with previous studies on adenomatous polyps (Hirsch et al., 2012; Ried et al., 1999). Furthermore, rectal adenomas frequently displayed CNAs, correlated with tubulo-villous histology and were increasingly genomically instable expressed by the genomic instability index and Shannon entropy. This evidence implies that CNAs are a key driver of pathogenesis in rectal adenomas. They seem genetically more advanced than colonic adenomas which might hint at the slightly different anatomy compared to the colon (Ponz de Leon and Di Gregorio, 2001). Genomic and epigenomic differences in left- or right-sided adenomas were recently documented (Koestler et al., 2014).

While aCGH provides an overview of the landscape of chromosomal aberrations of the tumor bulk by comparative analysis with non-tumor DNA (Lassmann et al., 2007), the present study was conceived to quantify CNAs additionally with a single-cell-specific approach. We analyzed the copy number status of fourteen cancer-associated genes (*i.e.*, nine oncogenes and five tumor suppressor genes) along centromere 10 in the set of colorectal adenoma samples. Except for *CCND1*, all targeted loci showed aberrant copy numbers across the sample set. Specifically, nine of the gene markers, *i.e.*, *COX2*, *PIK3CA*, *CLIC1*, *EGFR*, *MYC*, *CDX2*, *TP53*, *HER2*, and *ZNF217*, demonstrated unidirectional CNAs across the cohort. This suggests that these alterations did not occur randomly. Instead, they seem to play a causative role as being targeted by providing a beneficial genomic state to the colorectal tumor (Carvalho et al., 2009; Habermann et al., 2013; Jehan et al., 2009; Petrova et al., 2008; Wang et al., 2016). Conversely, the observation that four markers, *i.e.*, *APC*, *CDH1*, *SMAD7*, and *SMAD4*,

showed bidirectional CNAs could be attributed to genomic instability being randomly triggered in these regions. This assumption is substantiated by the fact that neither of the markers was homozygously lost nor amplified (Poorman et al., 2015). However, the tumor suppressive character was dominant in the complete sample set.

Abnormal signal counts across the set of adenomas were most frequently found to affect the markers *EGFR* (13%), *CDX2* (10.9%), and *ZNF217* (10.9%). As these markers reflect chromosomal aberrations of chromosome-arms 7p, 13q and 20q (Ried et al., 1996), the landscape of CNAs detected in single cells of the cohort overlapped with the findings by aCGH. Since the total number of altered copy number counts and ASNs harbored potential for sampling bias, a 10% threshold was used to call a specific alteration across the sum of single cells. Comparing CNAs called by miFISH and aCGH confirmed the valuable threshold since it provided a very good concordance, *i.e.*, 97.3% (κ=0.861). Herein, the present study achieved high overlaps compared to Heselmeyer-Haddad and coworkers (2014) using a FISH detection-threshold of 30% and found agreement in 83.3% of aCGH-miFISH comparisons.

Unsupervised hierarchical clustering based on CNAs separated the adenomas into four cluster groups. Cluster distribution 1-4 correlated with aberrations of *EGFR*, no detectable CNA, various CNAs, and alteration of *CDX2*. The results indicate that *EGFR* and *CDX2* gains occurred mutually exclusive. This underpins the relevance of both markers for colorectal tumorigenesis (Habermann et al., 2011; Roberts et al., 2002; Salari et al., 2012). It leads to the assumption that a gain of any of these markers could initiate and promote the formation of adenomatous polyps, irrespective of accumulation of additional alterations.

4.3.2 CNAs for prediction of adenoma recurrence

Utilizing the information of copy number landscapes obtained by aCGH, primary adenomas without recurrence were compared to primary adenomas with recurrence. Herein, neither the sheer presence of CNAs nor the ANCA index was capable of distinguishing both groups from each other. This notion indicates that the quantity of CNAs in adenomas is not exclusively linked to adenoma recurrence. It further implies that copy number alterations play a dominant role in most adenomas, though irrespective of their status of recurrence, in accordance with the adenoma-carcinoma sequence (Fearon and Vogelstein, 1990). Interestingly, part-chromosome alterations were significantly more often observed than whole-chromosome alterations in both groups. This

leads to the conclusion that neither the quantity of CNAs nor the base-pair length is capable to predict recurrence reliably. When dissecting the frequency plot of primary adenomas without and with recurrence, in total 24 differential regions of CNAs became evident. Sixteen CNAs were enriched in primary adenomas with recurrence. CNAs affecting chromosome 13q were exclusively associated with primary adenomas with recurrence. The proto-oncogenes, *e.g.*, *CDK8* and *CDX2* (Firestein et al., 2008; Salari et al., 2012), located on this chromosome might provide growth advantages and promote cellular survival, which facilitates a recurrence of the adenoma after polypectomy.

The aCGH data was unsupervised hierarchically clustered and showed no association with the adenoma groups. This observation confirms the initial results that the number of CNAs fails as a measure to discern primary adenomas with recurrence from primary adenomas without recurrence. It underpins that adenomas, irrespective of the status of recurrence, harbor various loads of chromosomal alterations. These separate individual fractions of specimens ranging from no-detectable CNA to multiple CNAs. The observation corroborates previous studies (Hermsen et al., 2002). Adenomas of *cluster 1*, however, showed complex chromosomal aberration patterns with alterations in chromosome number and focal CNAs. This complex pattern suggests a reduced genomic stability putatively indicative of an underlying CIN (Grady and Carethers, 2008). Cluster grouping was also associated with the median patient age. Adenomas of patients of advanced age also carried a higher CNA burden. These findings support the relevance of the adenoma-carcinoma sequence in which tumors accumulate mutations and chromosomal alterations over time (Fearon and Vogelstein, 1990).

Single-cell miFISH-analysis of CNAs in the set of adenomas revealed that *EGFR* was most-frequently affected by alterations (13/55). However, the marker was not statistically differentially distributed when comparing primary adenomas without recurrence (6/14) to primary adenomas with recurrence (7/28). Although *EGFR* gain was not related to recurrence, the results support the finding of previous studies that *EGFR* CNA is the most common alteration in adenomas and an early, causative factor in tumor development (Habermann et al., 2013; Ried et al., 1996). This view is documented by the correlation of *EGFR* gains with high instability indices, which implies to lead to an accelerated accumulation of instable clones. Functionally, the copy number gain gives rise to an overexpression of the transmembrane receptor which comprises an intrinsic tyrosine kinase activity. This overexpression confers cellular resistance to apoptosis and results in proliferation and expansion of tumor cells (Grünwald and Hidalgo, 2003; Roberts et al., 2002).

The second most common alteration observed by miFISH in colorectal adenomas was the copy number gain of *CDX2* (12/55). Considering that *CDX2* is located on chromosome 13q, this finding substantiates previous works revealing that this is a frequent target in the colorectal carcinogenesis (Bomme et al., 1994; Borras et al., 2012; Hirsch et al., 2012; Meijer et al., 1998; Ried et al., 2012; Xie et al., 2012). Unlike *EGFR,* the gain of *CDX2* was non-equally distributed across primary adenomas without and with recurrence. Primary adenomas without recurrence failed to present a CNA of *CDX2* (0/14), whereas 25% (7/28) of the primary adenomas with recurrence exhibited an extra copy of this gene. This result is equivalent to a sensitivity of 25% and a test specificity of 100% (*P*=0.040). Mechanistically, a *CDX2* gain is attributed with accelerating the proliferation capability and with promoting survival of CRC cells (Dang et al., 2006; Salari et al., 2012). It seems therefore reasonable to assume that this CN gain provides a selective evolutionary advantage to the intestinal cells. It further allows an independent proliferation which then empowers the lesion to recur after polypectomy. However, controversial findings also showed that *CDX2* expression was lost in a small subgroup of CRCs, suggesting *CDX2* as a prognostic biomarker (Choi et al., 2006; Dalerba et al., 2016). *CDX2* encodes a protein which is expressed in intestinal epithelial cells and is involved in intestinal development and differentiation (German et al., 1994; Liu et al., 2007). The analysis of the expression pattern in a subset of samples of the adenoma cohort did not unveil significant differences in expression levels across samples with and without *CDX2* gain. Nevertheless, this observation could be the correlate of low-level gains, *i.e.*, no amplifications were detected. Consequently, these gains solely induce alterations in the protein expression pattern which are not distinguishable by immunohistochemistry. Previous studies reported that both allelic imbalances and DNA methylation of *CDX2* were not directly correlated with the gene's protein expression (Koestler et al., 2014; Subtil et al., 2007).

The predictive value of the *CDX2* CNA could be consolidated by an "adenoma recurrence-risk" panel combining the marker with *CLIC1* and *ZNF217*. Copy number gain of any of these markers identified primary adenomas with recurrence with 39% sensitivity (*P*=0.036). However, specificity was decreased to 93% which is a sign of lesser clinical value. This combined leads to the conclusion that adenoma recurrence is a complex process orchestrated by various CNAs and other genetic processes. With the advent of "personalized medicine," analyzing the copy number status of *CDX2* might aid to improve the assessment of an individual risk for adenoma recurrence.

4.3.3 CNAs unveil intratumor heterogeneity

The present study utilized miFISH to assess the copy number status of well-accepted oncogenes and tumor suppressor genes in single cells of primary colorectal adenomas without and with recurrence, and in recurrent adenomas. Single-cell analysis of genetic subclones complemented the analysis of the genomic landscape of the tumor bulk by aCGH. The relevance of single cells in "intratumor diversification" of colorectal lesions was very recently highlighted (Roerink et al., 2018).

The set of colorectal adenoma specimens exhibited a major clone population size of on average 62.8%. This fraction size implies a stable clonal development of the colorectal adenomas with a low-level of clone diversity. This perspective, in turn, might reflect the clinical observation of a slow progression of adenomas towards CRC (Sillars-Hardebol et al., 2012). While major clone fractions of primary adenomas without and with recurrence did not differ, recurrent adenomas showed a mean major clone population size of 85%. This fact indicates that the dominant clone fractions of recurrent adenomas had a growth and selection advantage. Interestingly, colorectal adenomas comprised sizeable major clone fractions in comparison to prostate cancers (average size 20.8%; Heselmeyer-Haddad et al., 2014) and breast lesions (average size 17.5%; Heselmeyer-Haddad et al., 2012).

Addressing the average number of CNAs per case, primary adenomas without and with recurrence did not exhibit a difference across the subgroups. This finding underpins previous studies which report that chromosomal aberrations in colorectal lesions are rather heterogenous (The Cancer Genome Atlas, 2012). Nevertheless, dissection of primary adenomas with CNAs uncovered that adenomas with recurrence harbored clones with more complex signal patterns compared to adenomas without recurrence (P=0.016). Hence, it seems reasonable to assume that primary adenomas with complex CNA-pattern clones are more prone to recur. Various selection advantages required for recurrence might have equipped these adenomas. The number of CNAs per clone could arguably be used as a predictor for adenoma recurrence.

Besides the dominant clone populations, 47.3% of adenoma specimens possessed multiple minor clone populations. Although no significant differences were discovered between primary adenomas without and with recurrence, more than two-thirds of the specimens comprised of one or more clones. These observations corroborate recent works which claim the existence of polyclonality in colorectal adenomas (Novelli et al., 1996; Thirwell et al., 2010) and the presence of ITH in primary CRC (Kim et al., 2015).

Multiple CNAs contribute to an ITH. Thus, this study aimed to elucidate the relevance of ITH for adenoma recurrence by an assessment of diversity measures.

First, a chromosomal instability index was calculated which considers the number of signal patterns per 100 cells. This measure ranked on average 4.9 patterns in the adenoma cohort, which suggests a low degree of ITH compared to what was demonstrated in DCIS (62.3 patterns) and IDC (70.6 patterns; Heselmeyer-Haddad et al., 2012). Although the index failed to discern primary adenomas with recurrence from adenomas without recurrence, this accords with the stable clonality of the adenomas reflected by dominant and minor clones. Additionally, an association of older patients with an increased genomic instability index was documented - highlighting the impact of aging on the cellular and genetic homeostasis (Carvalho et al., 2012).

Second, the diversity measures of Shannon entropy (cohort mean 1.1) and Simpson index (cohort mean 0.31) were computed for the adenomatous specimens. No significant differences were documented across primary adenomas with and without recurrence. Despite that, trends were unveiled linking the most heterogeneous samples, *i.e.*, a large number of divergent clones, to primary adenomas with recurrence. This association points towards a similar direction, *i.e.*, a correlation with primary adenomas with recurrence, as the number of CNAs per clone does. It could be interpreted in the way that recurrence becomes more likely as the clonal composition of the adenoma becomes more diverse. This perspective complements previous findings which showed that a high ITH goes in line with poorer prognosis of cancers (Burrel and Swanton, 2016; Wangsa et al., 2016). Accordingly, tumor cells may have become resistant to immune surveillance and therapeutic intervention as the clone count elevated (Schwartz and Schäffer, 2017; Sottoriva et al., 2013).

Third, the accumulated pairwise genetic diversity was computed as a distance metric which is a simplified adaption of the average pairwise genetic divergence (Martinez et al., 2016; Rogers, 2015). The formula estimates the genetic distance of a set of given clone populations to the normal (*i.e.*, diploid) signal pattern. It showed a wide range across the adenomatous polyps. This result supports the impression of a heterogeneous distribution of CNAs beneficial to inter- and intratumor heterogeneity. The metric appeared not capable of distinguishing primary adenomas with recurrence from adenomas without recurrence. Instead, it was associated with tetraploid clone fractions. This association suggests that computing the genetic distance aids to identify whole-genome duplication-affected clone populations, which might eventually gain selective advantages due to tolerance to stress and promotion of transformation (Gordon

et al., 2012; Quintanilla, 2017; Storchová, 2012). The considerably higher scores in the applied diversity measures further emphasized the extraordinary role of tetraploidization. Consequently, it justified the exclusion of adenomas with clearly tetraploid clone fractions. Rutledge and Cimini (2016) postulated that aneuploidy is beneficial to tumor cells. This notion permits the conjecture that the risk of adenoma recurrence might be increased. However, the small number of tetraploid adenomas in our cohort could not confirm this assumption.

The analysis of diversity metrics in colorectal adenomas with and without recurrence leads to the conclusion that these measures can identify ITH with good precision in a wide range of samples. Strongly heterogeneous adenomas regarding clone fractions seemed more prone to recur than low-level ITH counterparts.

4.3.4 Clonal evolution via phylogenetic analysis of CNAs

As the understanding of the relevance of ITH in solid cancer populations advanced, more and more studies were conducted targeting the clonal evolution of neoplastic lesions by single-cell approaches (Heselmeyer-Haddad et al., 2012; Heselmeyer-Haddad et al., 2014; Russnes et al., 2011). The advanced miFISH method provided the possibility to infer the phylogenetic evolution from primary adenoma towards the recurrent adenoma by utilizing the FISHtree algorithm (Chowdhury et al., 2013; Schwartz and Schäffer, 2017). The present study unveiled four distinct clonal evolution patterns of adenoma recurrence: (*i*) simplification pattern, (*ii*) complexity pattern, (*iii*) stabilization pattern and (*iv*) "zero" pattern.

First, matched pairs demonstrating a simplification pattern (14%) suggest the following biological background. While subclonal populations of the primary adenoma were competing, the polypectomy of the bulk of the polyp supposedly removed or damaged clonal fractions of the tumor. That, conversely, gave rise to a new dominant clone. Additionally, it seems reasonable to assume that a dominant clone has overgrown less competitive clones because of positive selection. This linear evolutionary event, termed "clonal sweep," is an essential part of the "selective sweep model" which is characterized by a step-wise selection of fitter clones (Burrell and Swanton, 2014; Cross et al., 2016). Burrell and Swanton (2016) proposed that early divergence of the tumor cells might have caused recurrences of tumors. The fact that the dominant clone of the recurrent adenoma was already present in the primary adenoma supports this perspective.

Second, the complexity pattern, accounting for 21% of the matched pairs, showed a shift of the major clone population. This shift implies that the previous dominant clone was outcompeted as the selection pressure did not favor an expansion of the given clone population. It further suggests that both the tumor microenvironment and the growth conditions shifted which facilitated an increased ITH. These findings are in line with the scenario of a branched evolution (Burrell et al., 2013). Moreover, the appearance of multiple clone fractions balanced in population sizes underpins the existence of a neutral growth as proposed by Cross et al. (2016). Taken together, evidence suggests that the complexity pattern shares characteristics of the "Big Bang growth model" defined for CRC (Sottoriva et al., 2015). Recently, a study also demonstrated a branched evolution in CRC accompanied by excessive ITH via genomic and epigenomic assessment of spatially distant tumor regions (Uchi et al., 2016). In our study, pair P7 showed a distinctive feature as aneuploid clones with a 4N karyotype were observed in the recurrent adenoma. It implies a whole-genome duplication (WGD) in the process of recurrence. This WGD granted multiple advantageous growth features to these clones empowering them to outcompete clones with 2N karyotypes (Kuznetsova et al., 2015; Rutledge and Cimini, 2016).

Third, the stabilization pattern, accounting for 36% of the matched pairs, describes adenoma pairs in which CNAs and clonal populations of the primary lesion have not changed compared to the recurrent lesion. This stability buttresses the presence of an evolutionary tempo termed "stasis," according to which the tumor has adapted to the environment and developed a "fitness peak" (Cross et al., 2016). It was reported that adenomas might remain stable for years (Hofstad et al., 1996). An interesting aspect is added by the clonally dominant gains of *EGFR* and *CDX2*, respectively, being exclusively associated with the stabilization pattern. This exclusivity might indicate that chromosomal aberrations comprising any of these genes provide a growth advantage to the clones (Lockhart et al., 2005; Salari et al., 2012). Presumably, a negative selection then penalizes clones with additional or diverging CNAs.

Fourth, the "zero" pattern, seen in 29% of matched pairs, combined matched adenoma pairs which did not display clonal CNA present in more than 10% of the cells. The lack of CNAs further indicates that other genetic drivers might have played a role in the tumorigenesis of these adenomas (Snover, 2011). Nevertheless, expression of DNA MMR proteins was found to be retained in these samples, suggesting that the formation of adenomas did not occur via the MSI pathway. This microsatellite stability leads to the assumption that these adenomas might be affected by other driver-gene

mutations, *e.g.,* affecting *APC* or *KRAS* (Vogelstein et al., 2015). However, somatic mutational screening was not part of this study.

The presented identification of the clonal evolution patterns of adenoma recurrence seems to support and complement the emerging paradigm of an evolutionary "punctuated equilibrium" in CRCs, *i.e.*, variable tempos of stasis, gradualism, and punctuation (Cross et al., 2016). This notion would explain why adenomas following any of the shifting patterns, *i.e.*, simplification and complexity pattern, tended to recur earlier than adenomas of rather stable types, *i.e.*, stabilization and "zero" pattern. A punctuated evolution will show an accelerated tempo compared to a stasis (Cross et al., 2016).

Comparing the adenoma recurrence patterns with the progression from ductal carcinoma *in situ* (DCIS) to invasive ductal carcinoma (IDC) unveils an overlap in clonal evolution patterns resembling the stabilization and the complexity type (Heselmeyer-Haddad et al., 2012). Notwithstanding, the presence of neither the simplification nor the "zero" pattern was demonstrated in these breast lesions. An explanation is an observation that these tumors are far more progressed. This view is underpinned by the manifestation of utterly higher ITH measured via genomic instability index.

The role of incomplete resection of polyps seems a critical point of this study. On the one hand, the identification of adenoma pairs sharing concordant clonal populations across the primary and the recurrent adenoma encourages suspicion about an incomplete resection of the polyps. Indeed, the study was only comprised of adenomas which were endoscopically completely resected. Histopathologic complete resection is arguably tricky to diagnose when tissue has been fragmented. This complex of problems supports a perspective in which the adenomas might have been incompletely resected as this is a likely assumption of recurrent adenomas (Corley et al., 2014). On the other hand, a body of evidence suggests that incomplete resection is not irrevocably evident in our cohort. A longer timespan of recurrence in stable versus shifting patterns and the exclusive association of the stabilization pattern with *EGFR* or *CDX2* gains argue in favor of a non-relatedness to incomplete resection. The effect of field cancerization affecting normal histopathological epithelium is another possible contributor to recurrence (De Maio et al., 2017). Altogether, we cannot rule out that incomplete resection gave rise to recurrence but do see much more molecular evidence that four characteristic clonal evolution patterns dominate adenoma recurrence.

5 SUMMARY

Colorectal adenomas are precursors to colorectal cancer, one of the most common types of cancer diagnosed in both men and women in the Western world. Therefore, endoscopic resection of adenomatous polyps via polypectomy performed during colonoscopy provides the opportunity to reduce cancer risk. However, recurrence of adenomas after polypectomy is a common finding and affects one-third of the patients. This fact manifests both a cost and a health burden for the affected patients. For cost reduction and to individualize endoscopic surveillance strategies, identifying epigenetic and genetic alterations which predict adenoma recurrence would be of importance. Additionally, our knowledge of understanding the evolutionary processes of recurrence needs improvement.

Here, formalin-fixed paraffin-embedded tissues of primary adenomas without recurrence (n=30), primary adenomas with recurrence (n=19), matched-pair samples (n=19; primary adenoma and corresponding recurrent adenoma) and normal epithelium (n=3) were collected. Global epigenetic analysis of the methylome of the colorectal adenomas was conducted by Illumina's HumanMethylation450K BeadChip array (HM450K) to identify differentially methylated positions (DMPs). Array validation was performed via pyrosequencing. The genetic landscape of copy number alterations (CNAs) was investigated by array-comparative genomic hybridization (aCGH) and single-cell multiplex-interphase fluorescence *in situ* hybridization (miFISH). This assay comprised fifteen probes, 14 gene probes targeting colorectal cancer-related oncogenes and tumor suppressor genes along with a centromere 10 probe.

Filtering of methylation data provided 329,573 probes. Unsupervised clustering demonstrated an association of methylation patterns with the histologic subtype (P=0.008). 5,094 DMPs could be revealed across the comparisons of subgroups: 2,824 (55.4%) DMPs were hypermethylated and 2,270 (44.6%) DMPs were hypomethylated. DMPs located in CpG islands were strongly hypermethylated (86%), while DMPs in open sea regions and shelves were hypomethylated (83% and 79%, respectively). DMP-associated genes were enriched in inflammatory- and cancer-related pathways such as the MAPK signaling pathway, for instance. DMP discovery unveiled panels of 35 and 347 *top gene* DMPs across recurrent adenomas versus primary adenomas with recurrence and adenomas versus normal epithelium, respectively. Despite that, no DMPs could be identified across primary adenomas with and without recurrence. Array validation via pyrosequencing used the methylation of the *GREM2*

gene. The genetic landscape of adenomas unveiled by aCGH exhibited typical CNAs in colorectal tumorigenesis: chromosomes 7, 13q, 18 and 20. Most frequent focal aberrations (≤10 Mb) were located on 6p22.1-p21.33 (33.3%), 7q22.1 (31.4%) and 16q21 (29.4%). Single-cell miFISH showed frequent copy number gains within the colorectal adenomas affecting the markers *EGFR* (23.6%), *CDX2* (21.8%) and *ZNF217* (18.2%). Remarkably, copy number gains of *CDX2* were exclusively observed in primary adenomas with recurrence (25%) and recurrent adenomas (38.5%), while this CNA was absent in primary adenomas without recurrence (0%). The presence of major clone (defined as >40%) populations (average size 62.8%) accompanied by multiple minor clones was present in most adenomas (52.7%). While few adenomas were mainly composed of diploid cells (29.1%), a few adenomas were highly heterogeneous without a major clone population (16.4%) as confirmed by four quantitative diversity measures. Inferring of phylogenetic trees unraveled four distinct patterns of clonal evolution from primary adenomas towards recurrent adenomas: simplification pattern, complexity pattern, stability pattern and zero pattern. Noteworthy, copy numbers detected by aCGH and miFISH were concordant in 97.3% of observations ($\kappa=0.861$).

Collectively, the findings underpin that adenoma development and recurrence are complex processes orchestrated by genetic and epigenetic alterations. Whereas CpG methylation patterns appeared to be inconclusive for adenoma recurrence, the evaluation of CNAs affecting *CDX2* in single cells via miFISH might bear potential as a predictor of recurrence.

6 BIBLIOGRAPHY

Agilent (2015). Oligonucleotide Array-Based CGH for Genomic DNA Analysis. In Manual Protocol, (Santa Clara, CA, USA: Agilent Technologies), pp. 1–94.

Ahuja, N., Li, Q., Mohan, A.L., Baylin, S.B., and Issa, J.P. (1998). Aging and DNA methylation in colorectal mucosa and cancer. Cancer Res. *58*, 5489–5494.

American Cancer Society (2017). Colorectal Cancer - Facts and Figures 2017-2019 (Atlanta, Georgia, USA: American Cancer Society). Online: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/colorectal-cancer-facts-and-figures/colorectal-cancer-facts-and-figures-2017-2019.pdf, accessed: 08.06.17.

Andrew, A.S., Baron, J.A., Butterly, L.F., Suriawinata, A.A., Tsongali, G.J., Robinson, C.M., and Amos, C.I. (2017). Hyper-methylated loci persisting from sessile serrated polyps to serrated cancers. Int. J. Mol. Sci. *18*, 1–12.

Andreyev, H.J., Norman, A.R., Cunningham, D., Oates, J.R., and Clarke, P.A. (1998). Kirsten ras mutations in patients with colorectal cancer: the multicenter "RASCAL" study. J. Natl. Cancer Inst. *90*, 675–684.

Armaghany, T., Wilson, J.D., Chu, Q., and Mills, G. (2012). Genetic alterations in colorectal cancer. Gastrointest. Cancer Res. *5*, 19–27.

Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369.

Ashktorab, H., Daremipouran, M., Goel, A., Varma, S., Leavitt, R., Sun, X., and Brim, H. (2014). DNA methylome profiling identifies novel methylated genes in African American patients with colorectal neoplasia. Epigenetics *9:4*, 503–512.

Aust, D.E., and Baretton, G.B. (2010). Serrated polyps of the colon and rectum (hyperplastic polyps, sessile serrated adenomas, traditional serrated adenomas, and mixed polyps)-proposal for diagnostic criteria. Virchows Arch. *457*, 291–297.

Avidan, B., Sonnenberg, A., Schnell, T.G., Leya, J., Metz, A., and Sontag, S.J. (2002). New occurrence and recurrence of neoplasms within 5 years of a screening colonoscopy. Am. J. Gastroenterol. *97*, 1524–1529.

Bai, Y.-Q., Miyake, S., Iwai, T., and Yuasa, Y. (2003). CDX2, a homeobox transcription factor, upregulates transcription of the p21/WAF1/CIP1 gene. Oncogene *22*, 7942–7949.

Bakhoum, S.F., and Compton, D.A. (2012). Chromosomal instability and cancer: a complex relationship with therapeutic potential. J. Clin. Invest. *122*, 1138–1143.

Bakhoum, S.F., Thompson, S.L., Manning, A.L., and Compton, D.A. (2009). Genome stability is ensured by temporal control of kinetochore-microtubule dynamics. Nat. Cell Biol. *11*, 27–35.

Barrasa, J.I., Olmo, N., Lizarbe, M.A., and Turnay, J. (2013). Bile acids in the colon, from healthy to cytotoxic molecules. Toxicol. Vitr. *27*, 964–977.

Bennett-Baker, P.E., Wilkowski, J., and Burke, D.T. (2003). Age-associated activation of epigenetically repressed genes in the mouse. Genetics *165*, 2055–2062.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899–905.

Bestor, T.H. (2005). Transposons reanimated in mice. Cell *122*, 322–325.Bettington, M., Walker, N., Clouston, A., Brown, I., Leggett, B., and Whitehall, V. (2013). The serrated pathway to colorectal carcinoma: current concepts and challenges. Histopathology *62*, 367–386.

Bhola, N.E., and Grandis, J.R. (2008). Crosstalk between G-protein-coupled receptors and epidermal growth factor receptor in cancer. Front. Biosci. *13*, 1857–1865.

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., et al. (2011). High density DNA methylation array with single CpG site resolution. Genomics *98*, 288–295.

Bomme, L., Bardi, G., Pandis, N., Fenger, C., Kronborg, O., and Heim, S. (1994). Clonal karyotypic abnormalities in colorectal adenomas: clues to the early genetic events in the adenoma-carcinoma sequence. Genes. Chromosomes Cancer *10*, 190–196.

Bond, J.H. (2000). Polyp guideline: diagnosis, treatment, and surveillance for patients with colorectal polyps. Practice Parameters Committee of the American College of Gastroenterology. Am. J. Gastroenterol. *95*, 3053–3063.

Bonithon-Kopp, C., Piard, F., Fenger, C., Cabeza, E., O'Morain, C., Kronborg, O., Faivre, J., and European Cancer Prevention Organisation Study Group (2004). Colorectal adenoma characteristics as predictors of recurrence. Dis. Colon Rectum *47*, 323–333.

Bonni, A., Brunet, A., West, A.E., Datta, S.R., Takasu, M.A., and Greenberg, M.E. (1999). Cell survival promoted by the Ras-MAPK signaling pathway by transcription-dependent and -independent mechanisms. Science *286*, 1358–1362.

Borras, E., San Lucas, F.A., Chang, K., Zhou, R., Masand, G., Fowler, J., Mork, M.E., You, Y.N., Taggart, M.W., McAllister, F., et al. (2016). Genomic Landscape of Colorectal Mucosa and Adenomas. Cancer Prev. Res. *9*, 417–427.

Bosman, F., Carneiro, F., Hruban, R., and Theise, N. (2010). WHO Classification of Tumors of the Digestive System, 4th edition (Lyon: IARC Press).

Boveri, T. (1914). Zur Frage der Entstehung maligner Tumoren (Jena: Gustav Fischer).

Boyle, T., Keegel, T., Bull, F., Heyworth, J., and Fritschi, L. (2012). Physical activity and risks of proximal and distal colon cancers: a systematic review and meta-analysis. J. Natl. Cancer Inst. *104*, 1548–1561.

Bragdon, B., Moseychuk, O., Saldanha, S., King, D., Julian, J., and Nohe, A. (2011). Bone morphogenetic proteins: a critical review. Cell. Signal. *23*, 609–620.

Brenet, F., Moh, M., Funk, P., Feierstein, E., Viale, A.J., Socci, N.D., and Scandura, J.M. (2011). DNA methylation of the first exon is tightly linked to transcriptional silencing. PLoS One *6*, e14524.

Brosens, R.P.M., Haan, J.C., Carvalho, B., Rustenburg, F., Grabsch, H., Quirke, P., Engel, A.F., Cuesta, M.A., Maughan, N., Flens, M., et al. (2010). Candidate driver genes in focal chromosomal aberrations of stage II colon cancer. J. Pathol. *221*, 411–424.

Buecher, B., Cacheux, W., Rouleau, E., Dieumegard, B., Mitry, E., and Lièvre, A. (2013). Role of microsatellite instability in the management of colorectal cancers. Dig. Liver Dis. *45*, 441–449.

Burrell, R.A., McGranahan, N., Bartek, J., and Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. Nature *501*, 338–345.

Burrell, R.A., and Swanton, C. (2016). Re-Evaluating Clonal Dominance in Cancer Evolution. Trends in Cancer *2*, 263–276.

Burrell, R.A., and Swanton, C. (2014). The evolution of the unstable cancer genome. Curr. Opin. Genet. Dev. *24*, 61–67.

Burt, R.W. (2000). Colon cancer screening. Gastroenterology *119*, 837–853.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. *55*, 611–622.

Butterworth, A.S., Higgins, J.P.T., and Pharoah, P. (2006). Relative and absolute risk of colorectal cancer for individuals with a family history: a meta-analysis. Eur. J. Cancer *42*, 216–227.

Byrd, J.C., and Bresalier, R.S. (2004). Mucins and mucin binding proteins in colorectal cancer. Cancer Metastasis Rev. *23*, 77–99.

Cahill, D.P., Lengauer, C., Yu, J., Riggins, G.J., Willson, J.K., Markowitz, S.D., Kinzler, K.W., and Vogelstein, B. (1998). Mutations of mitotic checkpoint genes in human cancers. Nature *392*, 300–303.

Camps, J., Armengol, G., del Rey, J., Lozano, J.J., Vauhkonen, H., Prat, E., Egozcue, J., Sumoy, L., Knuutila, S., and Miró, R. (2006). Genome-wide differences between microsatellite stable and unstable colorectal tumors. Carcinogenesis *27*, 419–428.

Camps, J., Grade, M., Nguyen, Q.T., Hörmann, P., Becker, S., Hummon, A.B., Rodriguez, V., Chandrasekharappa, S., Chen, Y., Difilippantonio, M.J., et al. (2008). Chromosomal breakpoints in primary colon cancer cluster at sites of structural variants in the genome. Cancer Res. *68*, 1284–1295.

Camps, J., Ried, T., and Castells, A. (2013). Intratumor heterogeneity: finding the needle in a haystack for cancer treatment. Gastroenterology *145*, 242–244.

Cardoso, J., Boer, J., Morreau, H., and Fodde, R. (2007). Expression and genomic profiling of colorectal cancer. Biochim. Biophys. Acta *1775*, 103–137.

Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. *30*, 413–421.

Carvalho, B., Postma, C., Mongera, S., Hopmans, E., Diskin, S., van de Wiel, M.A., van Criekinge, W., Thas, O., Matthäi, A., Cuesta, M.A., et al. (2009). Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. Gut *58*, 79–89.

Carvalho, B., Sillars-Hardebol, A.H., Postma, C., Mongera, S., Terhaar Sive Droste, J., Obulkasim, A., van de Wiel, M., van Criekinge, W., Ylstra, B., Fijneman, R.J.A., et al. (2012). Colorectal adenoma to carcinoma progression is accompanied by changes in gene expression

associated with ageing, chromosomal instability, and fatty acid metabolism. Cell. Oncol. *35*, 53–63.

Center, M.M., Jemal, A., and Ward, E. (2009). International trends in colorectal cancer incidence rates. Cancer Epidemiol. Biomarkers Prev. *18*, 1688–1694.

Chan, D.S.M., Lau, R., Aune, D., Vieira, R., Greenwood, D.C., Kampman, E., and Norat, T. (2011). Red and processed meat and colorectal cancer incidence: meta-analysis of prospective studies. PLoS One *6*, e20456.

Chang, H.-F., Wu, C.-C., Sun, C.-A., Chu, C.-M., Lin, F.-G., Hsieh, J.-F., Hsu, C.-H., Huang, C.-H., Yang, T., Tsai, Y.-M., et al. (2016). Clinical stage and risk of recurrence and mortality: interaction of DNA methylation factors in patients with colorectal cancer. J. Investig. Med. *64*, 1200–1207.

Chen, Y., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J., and Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics *8*, 203–209.

Cho, E., Smith-Warner, S.A., Ritz, J., van den Brandt, P.A., Colditz, G.A., Folsom, A.R., Freudenheim, J.L., Giovannucci, E., Goldbohm, R.A., Graham, S., et al. (2004). Alcohol intake and colorectal cancer: a pooled analysis of 8 cohort studies. Ann. Intern. Med. *140*, 603–613.

Choi, B.J., Kim, C.J., Cho, Y.G., Song, J.H., Kim, S.Y., Nam, S.W., Lee, S.H., Yoo, N.J., Lee, J.Y., and Park, W.S. (2006). Altered expression of CDX2 in colorectal cancers. Acta Pathol. Microbiol. Immunol. Scand. *114*, 50–54.

Chowdhury, S.A., Shackney, S.E., Heselmeyer-Haddad, K., Ried, T., Schäffer, A.A., and Schwartz, R. (2013). Phylogenetic analysis of multiprobe fluorescence in situ hybridization data from tumor cell populations. Bioinformatics *29*, 189–198.

Chowdhury, S.A., Shackney, S.E., Heselmeyer-Haddad, K., Ried, T., Schäffer, A.A., and Schwartz, R. (2014). Algorithms to model single gene, single chromosome, and whole genome copy number changes jointly in tumor phylogenetics. PLoS Comput. Biol. *10*, e1003740.

Ciechanover, A., Orian, A., and Schwartz, A.L. (2000). Ubiquitin-mediated proteolysis: biological regulation via destruction. BioEssays *22*, 442–451.

Compton, D.A. (2000). Spindle assembly in animal cells. Annu. Rev. Biochem. *69*, 95–114.

Conesa-Zamora, P., García-Solano, J., Turpin, M.D.C., Sebastián-León, P., Torres-Moreno, D., Estrada, E., Tuomisto, A., Wilce, J., Mäkinen, M.J., Pérez-Guillermo, M., et al. (2015). Methylome profiling reveals functions and genes which are differentially methylated in serrated compared to conventional colorectal carcinoma. Clin. Epigenetics *7*, 101.

Cooper, H.S. (1983). Surgical pathology of endoscopically removed malignant polyps of the colon and rectum. Am. J. Surg. Pathol. *7*, 613–623.

Corley, D.A., Jensen, C.D., Marks, A.R., Zhao, W.K., Lee, J.K., Doubeni, C.A., Zauber, A.G., de Boer, J., Fireman, B.H., Schottinger, J.E., et al. (2014). Adenoma detection rate and risk of colorectal cancer and death. N. Engl. J. Med. *370*, 1298–1306.

Costello, J.F., Frühwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomäki, P., Lang, J.C., et al. (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nat. Genet. *24*, 132–138.

Cotran, R., Collins, K., and Robbins, T. (1999). Pathologic Basis of Disease (Philadelphia, Pennsylvania, USA: W.B. Saunders Company).

Cremer, T., Landegent, J., Brückner, A., Scholl, H.P., Schardin, M., Hager, H.D., Devilee, P., Pearson, P., and van der Ploeg, M. (1986). Detection of chromosome aberrations in the human interphase nucleus by visualization of specific target DNAs with radioactive and non-radioactive in situ hybridization techniques: diagnosis of trisomy 18 with probe L1.84. Hum. Genet. *74*, 346–352.

Cross, W.C., Graham, T.A., and Wright, N.A. (2016). New paradigms in clonal evolution: punctuated equilibrium in cancer. J. Pathol. *240*, 126–136.

Dalerba, P., Sahoo, D., Paik, S., Guo, X., Yothers, G., Song, N., Wilcox-Fogel, N., Forgó, E., Rajendran, P.S., Miranda, S.P., et al. (2016). CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer. N. Engl. J. Med. *374*, 211–222.

Dang, L.H., Chen, F., Ying, C., Chun, S.Y., Knock, S.A., Appelman, H.D., and Dang, D.T. (2006). CDX2 has tumorigenic potential in the human colon cancer cell lines LOVO and SW48. Oncogene *25*, 2264–2272.

De Angelis, P.M., Clausen, O.P., Schjølberg, A., and Stokke, T. (1999). Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their associations with tumour DNA ploidy, genotypes and phenotypes. Br. J. Cancer *80*, 526–535.

De Filippo, C., Luceri, C., Caderni, G., Pacini, M., Messerini, L., Biggeri, A., Mini, E., Tonelli, F., Cianchi, F., and Dolara, P. (2002). Mutations of the APC gene in human sporadic colorectal cancers. Scand. J. Gastroenterol. *37*, 1048–1053.

De Maio, G., Zama, E., Rengucci, C., and Calistri, D. (2017). What influences preneoplastic colorectal lesion recurrence? Oncotarget *8*, 12406–12416.

De Meyer, T., Bady, P., Trooskens, G., Kurscheid, S., Bloch, J., Kros, J.M., Hainfellner, J.A., Stupp, R., Delorenzi, M., Hegi, M.E., et al. (2015). Genome-wide DNA methylation detection by MethylCap-seq and Infinium HumanMethylation450 BeadChips: an independent large-scale comparison. Sci. Rep. *5*, 15375.

Dedeurwaerder, S., Defrance, M., Bizet, M., Calonne, E., Bontempi, G., and Fuks, F. (2014). A comprehensive overview of Infinium HumanMethylation450 data processing. Brief. Bioinform. *15*, 929–941.

Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C., and Fuks, F. (2011). Evaluation of the Infinium Methylation 450K technology. Epigenomics *3*, 771–784.

Dephoure, N., Hwang, S., O'Sullivan, C., Dodgson, S.E., Gygi, S.P., Amon, A., and Torres, E.M. (2014). Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. Elife *3*, e03023.

DiSario, J.A., Foutch, P.G., Mai, H.D., Pardy, K., and Manne, R.K. (1991). Prevalence and malignant potential of colorectal polyps in asymptomatic, average-risk men. Am. J. Gastroenterol. *86*, 941–945.

Douglas, E.J., Fiegler, H., Rowan, A., Halford, S., Bicknell, D.C., Bodmer, W., Tomlinson, I.P.M., and Carter, N.P. (2004). Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. Cancer Res. *64*, 4817–4825.

Duesberg, P., Rausch, C., Rasnick, D., and Hehlmann, R. (1998). Genetic instability of cancer cells is proportional to their degree of aneuploidy. Proc. Natl. Acad. Sci. U. S. A. *95*, 13692–13697.

Duval, A., and Hamelin, R. (2002). Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. Cancer Res. *62*, 2447–2454.

Eads, C.A., Danenberg, K.D., Kawakami, K., Saltz, L.B., Blake, C., Shibata, D., Danenberg, P. V, and Laird, P.W. (2000). MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res. *28*, E32.

East, J.E., Saunders, B.P., and Jass, J.R. (2008). Sporadic and syndromic hyperplastic polyps and serrated adenomas of the colon: classification, molecular genetics, natural history, and clinical management. Gastroenterol. Clin. North Am. *37*, 25–46.

Eden, A., Gaudet, F., Waghmare, A., and Jaenisch, R. (2003). Chromosomal instability and tumors promoted by DNA hypomethylation. Science *300*, 455.

Eide, T.J. (1991). Natural history of adenomas. World J. Surg. *15*, 3–6.

Elmunzer, B.J., Hayward, R.A., Schoenfeld, P.S., Saini, S.D., Deshpande, A., and Waljee, A.K. (2012). Effect of flexible sigmoidoscopy-based screening on incidence and mortality of colorectal cancer: a systematic review and meta-analysis of randomized controlled trials. PLoS Med. *9*, e1001352.

Esteller, M., Fraga, M.F., Guo, M., Garcia-Foncillas, J., Hedenfalk, I., Godwin, A.K., Trojan, J., Vaurs-Barrière, C., Bignon, Y.J., Ramus, S., et al. (2001). DNA methylation patterns in hereditary human cancers mimic sporadic tumorigenesis. Hum. Mol. Genet. *10*, 3001–3007.

Esteller, M., Sparks, A., Toyota, M., Sanchez-Cespedes, M., Capella, G., Peinado, M.A., Gonzalez, S., Tarafa, G., Sidransky, D., Meltzer, S.J., et al. (2000). Analysis of adenomatous polyposis coli promoter hypermethylation in human cancer. Cancer Res. *60*, 4366–4371.

Esteller, M. (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. Nat. Rev. Genet. *8*, 286–298.Esteller, M. (2011). Cancer Epigenetics for the 21st Century: What's Next? Genes Cancer *2*, 604–606.

Esteller, M. (2008). Epigenetics in cancer. N. Engl. J. Med. *358*, 1148–1159.Ethier, S.P. (2002). Signal transduction pathways: the molecular basis for targeted therapies. Semin. Radiat. Oncol. *12*, 3–10.

Fadda, A., Gentilini, D., Moi, L., Barault, L., Leoni, V.P., Sulas, P., Zorcolo, L., Restivo, A., Cabras, F., Fortunato, F., et al. (2018). Colorectal cancer early methylation alterations affect the crosstalk between cell and surrounding environment, tracing a biomarker signature specific for this tumor. Int. J. Cancer *143*, 907–920.

Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell *61*, 759–767.Fearon, E.R. (2011). Molecular genetics of colorectal cancer. Annu. Rev. Pathol. *6*, 479–507.

Feinberg, A.P., and Vogelstein, B. (1983). Hypomethylation of ras oncogenes in primary human cancers. Biochem. Biophys. Res. Commun. *111*, 47–54.

Feinberg, A.P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature *301*, 89–92.

Firestein, R., Bass, A.J., Kim, S.Y., Dunn, I.F., Silver, S.J., Guney, I., Freed, E., Ligon, A.H., Vena, N., Ogino, S., et al. (2008). CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. Nature *455*, 547–551.

Flora, M., Piana, S., Bassano, C., Bisagni, A., De Marco, L., Ciarrocchi, A., Tagliavini, E., Gardini, G., Tamagnini, I., Banzi, C., et al. (2012). Epidermal growth factor receptor (EGFR) gene copy number in colorectal adenoma-carcinoma progression. Cancer Genet. *205*, 630–635.

Fodde, R., Smits, R., and Clevers, H. (2001). APC, signal transduction and genetic instability in colorectal cancer. Nat. Rev. Cancer *1*, 55–67.

Fraga, M.F., and Esteller, M. (2002). DNA methylation: a profile of methods and applications. Biotechniques *33*, 632, 634, 636–649.

Fraga, M.F., Herranz, M., Espada, J., Ballestar, E., Paz, M.F., Ropero, S., Erkek, E., Bozdogan, O., Peinado, H., Niveleau, A., et al. (2004). A mouse skin multistage carcinogenesis model reflects the aberrant DNA methylation patterns of human tumors. Cancer Res. *64*, 5527–5534.

Frattini, M., Saletti, P., Romagnani, E., Martin, V., Molinari, F., Ghisletta, M., Camponovo, A., Etienne, L.L., Cavalli, F., and Mazzucchelli, L. (2007). PTEN loss of expression predicts cetuximab efficacy in metastatic colorectal cancer patients. Br. J. Cancer *97*, 1139–1145.

Gaiser, T., Meinhardt, S., Hirsch, D., Killian, J.K., Gaedcke, J., Jo, P., Ponsa, I., Mirõ, R., Rüschoff, J., Seitz, G., et al. (2013). Molecular patterns in the evolution of serrated lesion of the colorectum. Int. J. Cancer *132*, 1800–1810.

Galamb, O., Kalmár, A., Péterfia, B., Csabai, I., Bodor, A., Ribli, D., Krenács, T., Patai, Á. V, Wichmann, B., Barták, B.K., et al. (2016). Aberrant DNA methylation of WNT pathway genes in the development and progression of CIMP-negative colorectal cancer. Epigenetics *11*, 588–602.

Ganem, N.J., Godinho, S.A., and Pellman, D. (2009). A mechanism linking extra centrosomes to chromosomal instability. Nature *460*, 278–282.

German, M.S., Wang, J., Fernald, A.A., Espinosa, R., Le Beau, M.M., and Bell, G.I. (1994). Localization of the genes encoding two transcription factors, LMX1 and CDX3, regulating insulin gene expression to human chromosomes 1 and 13. Genomics *24*, 403–404.

Gertz, E.M., Chowdhury, S.A., Lee, W.-J., Wangsa, D., Heselmeyer-Haddad, K., Ried, T., Schwartz, R., and Schäffer, A.A. (2016). FISHtrees 3.0: Tumor Phylogenetics Using a Ploidy Probe. PLoS One *11*, e0158569.

Giovannucci, E. (2002). Modifiable risk factors for colon cancer. Gastroenterol. Clin. North Am. *31*, 925–943.

Giovannucci, E., Harlan, D.M., Archer, M.C., Bergenstal, R.M., Gapstur, S.M., Habel, L.A., Pollak, M., Regensteiner, J.G., and Yee, D. (2010). Diabetes and cancer: a consensus report. Diabetes Care *33*, 1674–1685.

Gordon, D.J., Resio, B., and Pellman, D. (2012). Causes and consequences of aneuploidy in cancer. Nat. Rev. Genet. *13*, 189–203.

Gorgoulis, V.G., Vassiliou, L.-V.F., Karakaidos, P., Zacharatos, P., Kotsinas, A., Liloglou, T., Venere, M., DiTullio, R.A., Kastrinakis, N.G., Levy, B., et al. (2005). Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. Nature *434*, 907–913.

Grade, M., Becker, H., Liersch, T., Ried, T., and Ghadimi, B.M. (2006). Molecular cytogenetics: Genomic imbalances in colorectal cancer and their clinical impact. Cell. Oncol. *28*, 71–84.

Grady, W.M., and Carethers, J.M. (2008). Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology *135*, 1079–1099.

Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L., and Robertson, M. (1991). Identification and characterization of the familial adenomatous polyposis coli gene. Cell *66*, 589–600.

Grünwald, V., and Hidalgo, M. (2003). Development of the epidermal growth factor receptor inhibitor OSI-774. Semin. Oncol. *30*, 23–31.

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., et al. (2015). The consensus molecular subtypes of colorectal cancer. Nat. Med. *21*, 1350–1356.

Gulwani, H. (2016). Colon nontumor normal histology. PathologyOutlines.com website, Online: http://www.pathologyoutlines.com/topic/colonhistology.html, accessed: 08.25.17.

Gupta, A.K., Pinsky, P., Rall, C., Mutch, M., Dry, S., Seligson, D., and Schoen, R.E. (2009). Reliability and accuracy of the endoscopic appearance in the identification of aberrant crypt foci. Gastrointest. Endosc. *70*, 322–330.

Gupta, A.K., Pretlow, T.P., and Schoen, R.E. (2007). Aberrant crypt foci: what we know and what we need to know. Clin. Gastroenterol. Hepatol. *5*, 526–533.

Gyparaki, M.-T., Basdra, E.K., and Papavassiliou, A.G. (2013). DNA methylation biomarkers as diagnostic and prognostic tools in colorectal cancer. J. Mol. Med. *91*, 1249–1256.

Haan, J.C., Labots, M., Rausch, C., Koopman, M., Tol, J., Mekenkamp, L.J.M., van de Wiel, M.A., Israeli, D., van Essen, H.F., van Grieken, N.C.T., et al. (2014). Genomic landscape of metastatic colorectal cancer. Nat. Commun. *5*, 5457.

Habermann, J.K., Brucker, C.A., Freitag-Wolf, S., Heselmeyer-Haddad, K., Krüger, S., Barenboim, L., Downing, T., Bruch, H.-P., Auer, G., Roblick, U.J., et al. (2011). Genomic instability and oncogene amplifications in colorectal adenomas predict recurrence and synchronous carcinoma. Mod. Pathol. *24*, 542–555.

Hamilton, S., Vogelstein, B., Kudo, S., Riboli, E., Nakamura, S., and Hainout, P. (2000). Tumours of the Colon and Rectum. In Pathology and Genetics of Tumours of the Digestive System, (Lyon, France: IARC Press), p. 103–129.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Hardwick, J.C., Kodach, L.L., Offerhaus, G.J., and van den Brink, G.R. (2008). Bone morphogenetic protein signalling in colorectal cancer. Nat. Rev. Cancer *8*, 806–812.

Hassan, C., Quintero, E., Dumonceau, J.-M., Regula, J., Brandão, C., Chaussade, S., Dekker, E., Dinis-Ribeiro, M., Ferlitsch, M., Gimeno-García, A., et al. (2013). Post-polypectomy colonoscopy surveillance: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. Endoscopy *45*, 842–864.

He, Q.-J., Zeng, W.-F., Sham, J.S.T., Xie, D., Yang, X.-W., Lin, H.-L., Zhan, W.-H., Lin, F., Zeng, S.-D., Nie, D., et al. (2003). Recurrent genetic alterations in 26 colorectal carcinomas and 21 adenomas from Chinese patients. Cancer Genet. Cytogenet. *144*, 112–118.

He, T.C., Sparks, A.B., Rago, C., Hermeking, H., Zawel, L., da Costa, L.T., Morin, P.J., Vogel-stein, B., and Kinzler, K.W. (1998). Identification of c-MYC as a target of the APC pathway. Science *281*, 1509–1512.

Heldin, C.H., Miyazono, K., and ten Dijke, P. (1997). TGF-beta signalling from cell membrane to nucleus through SMAD proteins. Nature *390*, 465–471.

Herbst, R.S. (2004). Review of epidermal growth factor receptor biology. Int. J. Radiat. Oncol. Biol. Phys. *59*, 21–26.

Herman, J.G., Graff, J.R., Myöhänen, S., Nelkin, B.D., and Baylin, S.B. (1996). Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc. Natl. Acad. Sci. U. S. A. *93*, 9821–9826.

Hermsen, M., Postma, C., Baak, J., Weiss, M., Rapallo, A., Sciutto, A., Roemen, G., Arends, J., Williams, R., Giaretti, W., et al. (2002). Colorectal adenoma to carcinoma progression fol-lows multiple pathways of chromosomal instability. Gastroenterology *123*, 1109–1119.

Herzog, F., Primorac, I., Dube, P., Lenart, P., Sander, B., Mechtler, K., Stark, H., and Peters, J.-M. (2009). Structure of the anaphase-promoting complex/cyclosome interacting with a mi-totic checkpoint complex. Science *323*, 1477–1481.

Heselmeyer-Haddad, K.M., Garcia, L.Y.B., Bradley, A., Hernandez, L., Hu, Y., Habermann, J.K., Dumke, C., Thorns, C., Perner, S., Pestova, E., et al. (2014). Single-cell genetic analysis reveals insights into clonal development of prostate cancers and indicates loss of PTEN as a marker of poor prognosis. Am. J. Pathol. *184*, 2671–2686.

Heselmeyer-Haddad, K., Berroa Garcia, L.Y., Bradley, A., Ortiz-Melendez, C., Lee, W.J., Christensen, R., Prindiville, S.A., Calzone, K.A., Soballe, P.W., Hu, Y., et al. (2012). Single-cell genetic analysis of ductal carcinoma in situ and invasive breast cancer reveals enormous tumor heterogeneity yet conserved genomic imbalances and gain of MYC during progression. Am. J. Pathol. *181*, 1807–1822.

Hinoi, T., Tani, M., Lucas, P.C., Caca, K., Dunn, R.L., Macri, E., Loda, M., Appelman, H.D., Cho, K.R., and Fearon, E.R. (2001). Loss of CDX2 expression and microsatellite instability are prominent features of large cell minimally differentiated carcinomas of the colon. Am. J. Pathol. *159*, 2239–2248.

Hinoue, T., Weisenberger, D.J., Lange, C.P.E., Shen, H., Byun, H.-M., Van Den Berg, D., Ma-lik, S., Pan, F., Noushmehr, H., van Dijk, C.M., et al. (2012). Genome-scale analysis of aberrant DNA methylation in colorectal cancer. Genome Res. *22*, 271–282.

Hirsch, D., Camps, J., Varma, S., Kemmerling, R., Stapleton, M., Ried, T., and Gaiser, T. (2012). A new whole genome amplification method for studying clonal evolution patterns in malignant colorectal polyps. Genes. Chromosomes Cancer *51*, 490–500.

Hofstad, B., Vatn, M.H., Andersen, S.N., Huitfeldt, H.S., Rognum, T., Larsen, S., and Osnes, M. (1996). Growth of colorectal polyps: redetection and evaluation of unresected polyps for a period of three years. Gut *39*, 449–456.

Holland, A.J., and Cleveland, D.W. (2009). Boveri revisited: chromosomal instability, aneu-ploidy and tumorigenesis. Nat. Rev. Mol. Cell Biol. *10*, 478–487.

Holliday, R. (1987). The inheritance of epigenetic defects. Science *238*, 163–170.

Hornick, J., and Odze, R. (2015). Polyps of the large intestine. In Surgical Pathology of the GI Tract, Liver, Billary Tract, and Pancreas., (Philadelphia, Pennsylvania, USA: Elsevier/Saunders), pp. 621–642.

Huret, J.-L., Ahmad, M., Arsaban, M., Bernheim, A., Cigna, J., Desangles, F., Guignard, J.-C., Jacquemot-Perbal, M.-C., Labarussias, M., Leberre, V., et al. (2013). Atlas of genetics and cytogenetics in oncology and haematology in 2013. Nucleic Acids Res. *41*, D920-4.

Iacopetta, B., Li, W.Q., Grieu, F., Ruszkiewicz, A., and Kawakami, K. (2006). BRAF mutation and gene methylation frequencies of colorectal tumours with microsatellite instability increase markedly with patient age. Gut *55*, 1213–1214.

Ibrahim, A.E.K., Arends, M.J., Silva, A.-L., Wyllie, A.H., Greger, L., Ito, Y., Vowler, S.L., Huang, T.H.-M., Tavaré, S., Murrell, A., et al. (2011). Sequential DNA methylation changes are associated with DNMT3B overexpression in colorectal neoplastic progression. Gut *60*, 499–508.

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

Illumina (2012). Infinium HD FFPE Methylation Assay, #15027309-Rev-B. In Manual Protocol, (San Diego, CA, USA), pp. 1–28. Online: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/infinium_assays/infinium-hd-ffpe-methylation/infinium-hd-methylation-ffpe-auto-euc-15027309-c.pdf, accessed 07.15.18.

Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D., and Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature *363*, 558–561.

Irby, K., Anderson, W.F., Henson, D.E., and Devesa, S.S. (2006). Emerging and widening colorectal carcinoma disparities between Blacks and Whites in the United States (1975-2002). Cancer Epidemiol. Biomarkers Prev. *15*, 792–797.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. *41*, 178–186.

Issa, J.-P. (2008). Colon cancer: it's CIN or CIMP. Clin. Cancer Res. *14*, 5939–5940.Itoh, F., Asao, H., Sugamura, K., Heldin, C.H., ten Dijke, P., and Itoh, S. (2001). Promoting bone morphogenetic protein signaling through negative regulation of inhibitory Smads. EMBO J. *20*, 4132–4142.

Jabbari, K., and Bernardi, G. (2004). Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. Gene *333*, 143–149.

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. *33 Suppl*, 245–254.

Jasperson, K.W., Tuohy, T.M., Neklason, D.W., and Burt, R.W. (2010). Hereditary and familial colon cancer. Gastroenterology *138*, 2044–2058.

Jass, J.R. (2007). Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology *50*, 113–130.

Jass, J.R., Barker, M., Fraser, L., Walsh, M.D., Whitehall, V.L.J., Gabrielli, B., Young, J., and Leggett, B.A. (2003). APC mutation and tumour budding in colorectal cancer. J. Clin. Pathol. *56*, 69–73.

Jass, J.R., Biden, K.G., Cummings, M.C., Simms, L.A., Walsh, M., Schoch, E., Meltzer, S.J., Wright, C., Searle, J., Young, J., et al. (1999). Characterisation of a subtype of colorectal cancer combining features of the suppressor and mild mutator pathways. J. Clin. Pathol. *52*, 455–460.

Jehan, Z., Bavi, P., Sultana, M., Abubaker, J., Bu, R., Hussain, A., Alsbeih, G., Al-Sanea, N., Abduljabbar, A., Ashari, L.H., et al. (2009). Frequent PIK3CA gene amplification and its clinical significance in colorectal cancer. J. Pathol. *219*, 337–346.

Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Murray, T., and Thun, M.J. (2008). Cancer Statistics, 2008. CA Cancer J Clin. *58*, 71–96.

Jen, J., Powell, S.M., Papadopoulos, N., Smith, K.J., Hamilton, S.R., Vogelstein, B., and Kinzler, K.W. (1994). Molecular determinants of dysplasia in colorectal lesions. Cancer Res. *54*, 5523–5526.

Jhawer, M., Goel, S., Wilson, A.J., Montagna, C., Ling, Y.-H., Byun, D.-S., Nasser, S., Arango, D., Shin, J., Klampfer, L., et al. (2008). PIK3CA mutation/PTEN expression status predicts response of colon cancer cells to the epidermal growth factor receptor inhibitor cetuximab. Cancer Res. *68*, 1953–1961.

Johnson, V., Volikos, E., Halford, S.E., Eftekhar Sadat, E.T., Popat, S., Talbot, I., Truninger, K., Martin, J., Jass, J., Houlston, R., et al. (2005). Exon 3 beta-catenin mutations are specifically associated with colorectal carcinomas in hereditary non-polyposis colorectal cancer syndrome. Gut *54*, 264–267.

Jones, A.M., Thirlwell, C., Howarth, K.M., Graham, T., Chambers, W., Segditsas, S., Page, K.M., Phillips, R.K.S., Thomas, H.J.W., Sieber, O.M., et al. (2007). Analysis of copy number changes suggests chromosomal instability in a minority of large colorectal adenomas. J. Pathol. *213*, 249–256.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Jones, P.A., and Baylin, S.B. (2007). The epigenomics of cancer. Cell *128*, 683–692.

Khanna, K.K., and Jackson, S.P. (2001). DNA double-strand breaks: signaling, repair and the cancer connection. Nat. Genet. *27*, 247–254.

Kim, T.-M., Jung, S.-H., An, C.H., Lee, S.H., Baek, I.-P., Kim, M.S., Park, S.-W., Rhee, J.-K., Lee, S.-H., and Chung, Y.-J. (2015). Subclonal Genomic Architectures of Primary and Metastatic Colorectal Cancer Based on Intratumoral Genetic Heterogeneity. Clin. Cancer Res. *21*, 4461–4472.

Kinzler, K.W., and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. Cell *87*, 159–170.Kinzler, K.W., and Vogelstein, B. (1997). Cancer-susceptibility genes. Gatekeepers and caretakers. Nature *386*, 761, 763.

Kisiel, J.B., Yab, T.C., Nazer Hussain, F.T., Taylor, W.R., Garrity-Park, M.M., Sandborn, W.J., Loftus, E. V, Wolff, B.G., Smyrk, T.C., Itzkowitz, S.H., et al. (2013). Stool DNA testing for the detection of colorectal neoplasia in patients with inflammatory bowel disease. Aliment. Pharmacol. Ther. *37*, 546–554.

Knudson, A.G. (1971). Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. U. S. A. *68*, 820–823.

Kodach, L.L., Wiercinska, E., de Miranda, N.F.C.C., Bleuming, S.A., Musler, A.R., Peppelenbosch, M.P., Dekker, E., van den Brink, G.R., van Noesel, C.J.M., Morreau, H., et al. (2008). The bone morphogenetic protein pathway is inactivated in the majority of sporadic colorectal cancers. Gastroenterology *134*, 1332–1341.

Koestler, D.C., Li, J., Baron, J.A., Tsongalis, G.J., Butterly, L.F., Goodrich, M., Lesseur, C., Karagas, M.R., Marsit, C.J., Moore, J.H., et al. (2014). Distinct patterns of DNA methylation in conventional adenomas involving the right and left colon. Mod. Pathol. *27*, 145–155.

Kontogeorgos, G. (2004). Molecular cytogenetics of pituitary adenomas, assessed by FISH technique. Front. Horm. Res. *32*, 205–216.

Kosinski, C., Li, V.S.W., Chan, A.S.Y., Zhang, J., Ho, C., Tsui, W.Y., Chan, T.L., Mifflin, R.C., Powell, D.W., Yuen, S.T., et al. (2007). Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. Proc. Natl. Acad. Sci. U. S. A. *104*, 15418–15423.

Krimpenfort, P., Song, J.-Y., Proost, N., Zevenhoven, J., Jonkers, J., and Berns, A. (2012). Deleted in colorectal carcinoma suppresses metastasis in p53-deficient mammary tumours. Nature *482*, 538–541.

Kushi, L.H., Doyle, C., McCullough, M., Rock, C.L., Demark-Wahnefried, W., Bandera, E. V, Gapstur, S., Patel, A. V, Andrews, K., Gansler, T., et al. (2012). American Cancer Society Guidelines on nutrition and physical activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. CA Cancer J Clin. *62*, 30–67.

Kuznetsova, A.Y., Seget, K., Moeller, G.K., de Pagter, M.S., de Roos, J.A.D.M., Dürrbaum, M., Kuffer, C., Müller, S., Zaman, G.J.R., Kloosterman, W.P., et al. (2015). Chromosomal instability, tolerance of mitotic errors and multidrug resistance are promoted by tetraploidization in human cells. Cell Cycle *14*, 2810–2820.

Lam, K., Pan, K., Linnekamp, J.F., Medema, J.P., and Kandimalla, R. (2016). DNA methylation based biomarkers in colorectal cancer: A systematic review. Biochim. Biophys. Acta *1866*, 106–120.

Lane, D.P. (1992). Cancer. p53, guardian of the genome. Nature *358*, 15–16.

Lassmann, S., Weis, R., Makowiec, F., Roth, J., Danciu, M., Hopt, U., and Werner, M. (2007). Array CGH identifies distinct DNA copy number profiles of oncogenes and tumor suppressor genes in chromosomal- and microsatellite-unstable sporadic colorectal carcinomas. J. Mol. Med. *85*, 293–304.

Leitlinienprogramm Onkologie (AWMF, Deutsche Krebsgesellschaft und Deutsche Krebshilfe): Pox, C., Aretz, S., Bischoff, S.C., Graeven, U., Hass, M., Heußner, P., Hohenberger, W., Holstege, A., Hübner, J., Kolligs, F., et al. (2014). S3-Leitlinie Kolorektales Karzinom Langversion 1.1 – August 2014. AWMF-Registernummer: 021/007OL. Online: https://leitlinienprogramm-onkologie.de/uploads/tx_sbdownloader/LL_KRK_Langfassung_1.1.pdf, accessed 07.16.18.

Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1997). Genetic instability in colorectal cancers. Nature *386*, 623–627.

Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. Nature *396*, 643–649.

Leslie, A., Carey, F.A., Pratt, N.R., and Steele, R.J.C. (2002). The colorectal adenoma-carcinoma sequence. Br. J. Surg. *89*, 845–860.

Levine, J.S., and Ahnen, D.J. (2006). Clinical practice. Adenomatous polyps of the colon. N. Engl. J. Med. *355*, 2551–2557.

Li, L., Ying, J., Li, H., Zhang, Y., Shu, X., Fan, Y., Tan, J., Cao, Y., Tsao, S.W., Srivastava, G., et al. (2012). The human cadherin 11 is a pro-apoptotic tumor suppressor modulating cell stemness through Wnt/β-catenin signaling and silenced in common carcinomas. Oncogene *31*, 3901–3912.

Liu, Q., Teh, M., Ito, K., Shah, N., Ito, Y., and Yeoh, K.G. (2007). CDX2 expression is progressively decreased in human gastric intestinal metaplasia, dysplasia and cancer. Mod. Pathol. *20*, 1286–1297.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408.

Lockhart, A.C., Lockhart, C., and Berlin, J.D. (2005). The epidermal growth factor receptor as a target for colorectal cancer therapy. Semin. Oncol. *32*, 52–60.

Loh, K., Chia, J.A., Greco, S., Cozzi, S.-J., Buttenshaw, R.L., Bond, C.E., Simms, L.A., Pike, T., Young, J.P., Jass, J.R., et al. (2008). Bone morphogenic protein 3 inactivation is an early and frequent event in colorectal cancer development. Genes. Chromosomes Cancer *47*, 449–460.

Longacre, T.A., and Fenoglio-Preiser, C.M. (1990). Mixed hyperplastic adenomatous polyps/serrated adenomas. A distinct form of colorectal neoplasia. Am. J. Surg. Pathol. *14*, 524–537.

Luo, Y., Wong, C., Kaz, A.M., Dzieciatkowski, S., Kelly, T., Lutterbaugh, J.D., Shrubsole, M.J., Zheng, W., and Sanford, D. (2014). Differences in DNA methylation signatures reveal multiple pathways of progression from adenoma to colorectal cancer. Gastroenterology *147*, 418–429.

Ma, Y., Yang, Y., Wang, F., Zhang, P., Shi, C., Zou, Y., and Qin, H. (2013). Obesity and risk of colorectal cancer: a systematic review of prospective studies. PLoS One *8*, e53916.

Maiato, H., DeLuca, J., Salmon, E.D., and Earnshaw, W.C. (2004). The dynamic kinetochore-microtubule interface. J. Cell Sci. *117*, 5461–5477.

Maksimovic, J., Gordon, L., and Oshlack, A. (2012). SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips. Genome Biol. *13*, R44.

Mann, B., Gelos, M., Siedow, A., Hanski, M.L., Gratchev, A., Ilyas, M., Bodmer, W.F., Moyer, M.P., Riecken, E.O., Buhr, H.J., et al. (1999). Target genes of beta-catenin-T cell-factor/lymphoid-enhancer-factor signaling in human colorectal carcinomas. Proc. Natl. Acad. Sci. U. S. A. *96*, 1603–1608.

Mantovani, A., Allavena, P., Sica, A., and Balkwill, F. (2008). Cancer-related inflammation. Nature *454*, 436–444.

Markowitz, S.D., and Bertagnolli, M.M. (2009). Molecular origins of cancer: Molecular basis of colorectal cancer. N. Engl. J. Med. *361*, 2449–2460.

Martinez, P., Timmer, M.R., Lau, C.T., Calpe, S., Sancho-Serra, M.D.C., Straub, D., Baker, A.-M., Meijer, S.L., Kate, F.J.W. Ten, Mallant-Hent, R.C., et al. (2016). Dynamic clonal equilibrium and predetermined cancer risk in Barrett's oesophagus. Nat. Commun. *7*, 12158.

Martin-Subero, J.I., Ammerpohl, O., Bibikova, M., Wickham-Garcia, E., Agirre, X., Alvarez, S., Brüggemann, M., Bug, S., Calasanz, M.J., Deckert, M., et al. (2009). A comprehensive microarray-based DNA methylation study of 367 hematological neoplasms. PLoS One *4*, e6986.

McGranahan, N., Burrell, R.A., Endesfelder, D., Novelli, M.R., and Swanton, C. (2012). Cancer chromosomal instability: therapeutic and diagnostic challenges. EMBO Rep. *13*, 528–538.

Meijer, G.A., Hermsen, M.A.J.A., Baak, J.P.A., Van Diest, P.J., Meuwissen, S.G.M., Beliën, J.A.M., Hoovers, J.M.N., Joenje, H., Snijders, P.J.F., and Walboomers, J.M.M. (1998). Progression from colorectal adenoma to carcinoma is associated with non- random chromosomal gains as detected by comparative genomic hybridisation. J. Clin. Pathol. *51*, 901–909.

Meling, G.I., Lothe, R.A., Børresen, A.L., Graue, C., Hauge, S., Clausen, O.P., and Rognum, T.O. (1993). The TP53 tumour suppressor gene in colorectal carcinomas. I. Genetic alterations on chromosome 17. Br. J. Cancer *67*, 88–92.

Miyaki, M., Iijima, T., Kimura, J., Yasuno, M., Mori, T., Hayashi, Y., Koike, M., Shitara, N., Iwama, T., and Kuroki, T. (1999). Frequent mutation of beta-catenin and APC genes in primary colorectal tumors from patients with hereditary nonpolyposis colorectal cancer. Cancer Res. *59*, 4506–4509.

Miyaki, M., Konishi, M., Kikuchi-Yanoshita, R., Enomoto, M., Igari, T., Tanaka, K., Muraoka, M., Takahashi, H., Amada, Y., and Fukayama, M. (1994). Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors. Cancer Res. *54*, 3011–3020.

Morson, B.C. (1962). Precancerous lesions of the colon and rectum. Classification and controversial issues. JAMA *179*, 316–321.

Morson, B.C., Whiteway, J.E., Jones, E.A., Macrae, F.A., and Williams, C.B. (1984). Histopathology and prognosis of malignant colorectal polyps treated by endoscopic polypectomy. Gut *25*, 437–444.

Muto, T., Bussey, H.J., and Morson, B.C. (1975). The evolution of cancer of the colon and rectum. Cancer *36*, 2251–2270.

Murugaesu, N., Wilson, G.A., Birkbak, N.J., Watkins, T., McGranahan, N., Kumar, S., Abbassi-Ghadi, N., Salm, M., Mitter, R., Horswell, S., et al. (2015). Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. Cancer Discov. *5*, 821–831.

Muleris, M., Zafrani, B., Validire, P., Girodet, J., Salmon, R.J., and Dutrillaux, B. (1994). Cytogenetic study of 30 colorectal adenomas. Cancer Genet. Cytogenet. *74*, 104–108.

Mysliwiec, P., Cronin, K., and Schatzkin, A. (2006). Chapter 5: New Malignancies Following Cancer of the Colon, Rectum, and Anus. In New Malignancies among Cancer Survivors: SEER Cancer Registries, 1973-2000, (Bethesda (Maryland): National Cancer Institute).

Nagaoka, S.I., Hassold, T.J., and Hunt, P.A. (2012). Human aneuploidy: mechanisms and new insights into an age-old problem. Nat. Rev. Genet. *13*, 493–504.

Nakayama, K.I., and Nakayama, K. (2006). Ubiquitin ligases: cell-cycle control and cancer. Nat. Rev. Cancer *6*, 369–381.

Naumov, V.A., Generozov, E. V, Zaharjevskaya, N.B., Matushkina, D.S., Larin, A.K., Chernyshov, S. V, Alekseev, M. V, Shelygin, Y.A., and Govorun, V.M. (2013). Genome-scale analysis of DNA methylation in colorectal cancer using Infinium HumanMethylation450 BeadChips. Epigenetics *8*, 921–934.

Neugut, A.I., Jacobson, J.S., Ahsan, H., Santos, J., Garbowski, G.C., Forde, K.A., Treat, M.R., and Waye, J. (1995). Incidence and recurrence rates of colorectal adenomas: A prospective study. Gastroenterology *108*, 402–408.

Nicholson, J.M., Macedo, J.C., Mattingly, A.J., Wangsa, D., Camps, J., Lima, V., Gomes, A.M., Dória, S., Ried, T., Logarinho, E., et al. (2015). Chromosome mis-segregation and cytokinesis failure in trisomic human cells. Elife *4*.

Noshirwani, K.C., van Stolk, R.U., Rybicki, L.A., and Beck, G.J. (2000). Adenoma size and number are predictive of adenoma recurrence: implications for surveillance colonoscopy. Gastrointest. Endosc. *51*, 433–437.

Novelli, M.R., Williamson, J.A., Tomlinson, I.P., Elia, G., Hodgson, S. V, Talbot, I.C., Bodmer, W.F., and Wright, N.A. (1996). Polyclonal origin of colonic adenomas in an XO/XY patient with FAP. Science *272*, 1187–1190.

Nowell, P.C. (1976). The clonal evolution of tumor cell populations. Science *194*, 23–28.

O'Connell, J.B., Maggard, M.A., and Ko, C.Y. (2004). Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. J. Natl. Cancer Inst. *96*, 1420–1425.

O'Hagan, R.C., Chang, S., Maser, R.S., Mohan, R., Artandi, S.E., Chin, L., and DePinho, R.A. (2002). Telomere dysfunction provokes regional amplification and deletion in cancer genomes. Cancer Cell *2*, 149–155.

Oltmann, J., Heselmeyer-Haddad, K., Hernandez, L.S., Meyer, R., Torres, I., Hu, Y., Doberstein, N., Killian, J.K., Petersen, D., Zhu, Y.J., et al. (2018). Aneuploidy, TP53 mutation, and amplification of MYC correlate with increased intratumor heterogeneity and poor prognosis of breast cancer patients. Genes. Chromosomes Cancer *57*, 165–175.

Oncomine database. Thermo-Fisher Scientific. Online: https://www.oncomine.org/resource/login.html, accessed 11.18.14.

Otori, K., Konishi, M., Sugiyama, K., Hasebe, T., Shimoda, T., Kikuchi-Yanoshita, R., Mukai, K., Fukushima, S., Miyaki, M., and Esumi, H. (1998). Infrequent somatic mutation of the adenomatous polyposis coli gene in aberrant crypt foci of human colon tissue. Cancer *83*, 896–900.

Pardue, M.L., and Gall, J.G. (1969). Molecular hybridization of radioactive DNA to the DNA of cytological preparations. Proc. Natl. Acad. Sci. U. S. A. *64*, 600–604.

Paspatis, G.A., Papanikolaou, N., Zois, E., and Michalodimitrakis, E. (2001). Prevalence of polyps and diverticulosis of the large bowel in the Cretan population. An autopsy study. Int. J. Colorectal Dis. *16*, 257–261.

Passerini, V., Ozeri-Galai, E., de Pagter, M.S., Donnelly, N., Schmalbrock, S., Kloosterman, W.P., Kerem, B., and Storchová, Z. (2016). The presence of extra chromosomes leads to genomic instability. Nat. Commun. *7*, 10754.

Patai, Á. V, Valcz, G., Hollósi, P., Kalmár, A., Péterfia, B., Patai, Á., Wichmann, B., Spisák, S., Barták, B.K., Leiszter, K., et al. (2015). Comprehensive DNA Methylation Analysis Reveals a Common Ten-Gene Methylation Signature in Colorectal Adenomas and Carcinomas. PLoS One *10*, e0133836.

Patel, S.G., and Ahnen, D.J. (2012). Familial colon cancer syndromes: an update of a rapidly evolving field. Curr. Gastroenterol. Rep. *14*, 428–438.

Pelli, A., Väyrynen, J.P., Klintrup, K., Mäkelä, J., Mäkinen, M.J., Tuomisto, A., and Karttunen, T.J. (2016). Gremlin1 expression associates with serrated pathway and favourable prognosis in colorectal cancer. Histopathology *69*, 831–838.

Petko, Z., Ghiassi, M., Shuber, A., Gorham, J., Smalley, W., Washington, M.K., Schultenover, S., Gautam, S., Markowitz, S.D., and Grady, W.M. (2005). Aberrantly methylated CDKN2A, MGMT, and MLH1 in colon polyps and in fecal DNA from patients with colorectal polyps. Clin. Cancer Res. *11*, 1203–1209.

Petrova, D.T., Asif, A.R., Armstrong, V.W., Dimova, I., Toshev, S., Yaramov, N., Oellerich, M., and Toncheva, D. (2008). Expression of chloride intracellular channel protein 1 (CLIC1) and tumor protein D52 (TPD52) as potential biomarkers for colorectal cancer. Clin. Biochem. *41*, 1224–1236.

Pickup, M.W., Mouw, J.K., and Weaver, V.M. (2014). The extracellular matrix modulates the hallmarks of cancer. EMBO Rep. *15*, 1243–1253.

Pino, M.S., and Chung, D.C. (2010). The chromosomal instability pathway in colon cancer. Gastroenterology *138*, 2059–2072.

Platzer, P., Upender, M.B., Wilson, K., Willis, J., Lutterbaugh, J., Nosrati, A., Willson, J.K. V, Mack, D., Ried, T., and Markowitz, S. (2002). Silence of chromosomal amplifications in colon cancer. Cancer Res. *62*, 1134–1138.

Poeaim, S., Rerkamnuaychoke, B., Jesdapatarakul, S., and Campiranon, A. (2005). Chromosome alterations in colorectal cancer in Thai patients. Cancer Genet. Cytogenet. *160*, 152–159.

Ponz de Leon, M., and Di Gregorio, C. (2001). Pathology of colorectal cancer. Dig. Liver Dis. *33*, 372–388.

Poorman, K., Borst, L., Moroff, S., Roy, S., Labelle, P., Motsinger-Reif, A., and Breen, M. (2015). Comparative cytogenetic characterization of primary canine melanocytic lesions using array CGH and fluorescence in situ hybridization. Chromosom. Res. *23*, 171–186.

Popat, S., Hubner, R., and Houlston, R.S. (2005). Systematic review of microsatellite instability and colorectal cancer prognosis. J. Clin. Oncol. *23*, 609–618.

Popkin, B.M. (2015). Nutrition Transition and the Global Diabetes Epidemic. Curr. Diab. Rep. *15*, 64.

Porta, C., Larghi, P., Rimoldi, M., Totaro, M.G., Allavena, P., Mantovani, A., and Sica, A. (2009). Cellular and molecular pathways linking inflammation and cancer. Immunobiology *214*, 761–777.

Postma, C., Hermsen, M.A.J.A., Coffa, J., Baak, J.P.A., Mueller, J.D., Mueller, E., Bethke, B., Schouten, J.P., Stolte, M., and Meijer, G.A. (2005). Chromosomal instability in flat adenomas and carcinomas of the colon. J. Pathol. *205*, 514–521.

Potten, C.S., Wilson, J.W., and Booth, C. (1997). Regulation and significance of apoptosis in the stem cells of the gastrointestinal epithelium. Stem Cells *15*, 82–93.

Powell, S.M., Zilz, N., Beazer-Barclay, Y., Bryan, T.M., Hamilton, S.R., Thibodeau, S.N., Vogelstein, B., and Kinzler, K.W. (1992). APC mutations occur early during colorectal tumorigenesis. Nature *359*, 235–237.

Pretlow, T.P., Barrow, B.J., Ashton, W.S., O'Riordan, M.A., Pretlow, T.G., Jurcisek, J.A., and Stellato, T.A. (1991). Aberrant crypts: putative preneoplastic foci in human colonic mucosa. Cancer Res. *51*, 1564–1567.

Pretlow, T.P., Brasitus, T.A., Fulton, N.C., Cheyer, C., and Kaplan, E.L. (1993). K-ras mutations in putative preneoplastic lesions in human colon. J. Natl. Cancer Inst. *85*, 2004–2007.

Price, M.E., Cotton, A.M., Lam, L.L., Farré, P., Emberly, E., Brown, C.J., Robinson, W.P., and Kobor, M.S. (2013). Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics Chromatin *6*, 4.

Pruitt, K., and Der, C.J. (2001). Ras and Rho regulation of the cell cycle and oncogenesis. Cancer Lett. *171*, 1–10.

Quinlan, K.G.R., Verger, A., Yaswen, P., and Crossley, M. (2007). Amplification of zinc finger gene 217 (ZNF217) and cancer: when good fingers go bad. Biochim. Biophys. Acta *1775*, 333–340.

Quintanilla, I. (2017). New insights into the colorectal carcinogenesis: from early precursor lesions to the role of aneuploidy. Medical dissertation. University of Barcelona, Spain.

Radtke, F., and Clevers, H. (2005). Self-renewal and cancer of the gut: two sides of a coin. Science *307*, 1904–1909.

Rakyan, V.K., Down, T.A., Maslau, S., Andrew, T., Yang, T.-P., Beyan, H., Whittaker, P., McCann, O.T., Finer, S., Valdes, A.M., et al. (2010). Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res. *20*, 434–439.

Reish, O., Brosh, N., Gobazov, R., Rosenblat, M., Libman, V., and Mashevich, M. (2006). Sporadic aneuploidy in PHA-stimulated lymphocytes of Turner's syndrome patients. Chromosom. Res. *14*, 527–534.

Remmele, W., and Stegner, H.E. (1987). Recommendation for uniform definition of an immunoreactive score (IRS) for immunohistochemical estrogen receptor detection (ER-ICA) in breast cancer tissue. Pathologe *8*, 138–140.

Ren, J., Kirkness, C.S., Kim, M., Asche, C. V, and Puli, S. (2016). Long-term risk of colorectal cancer by gender after positive colonoscopy: population-based cohort study. Curr. Med. Res. Opin. *32*, 1367–1374.

Rengucci, C., De Maio, G., Casadei Gardini, A., Zucca, M., Scarpi, E., Zingaretti, C., Foschi, G., Tumedei, M.M., Molinari, C., Saragoni, L., et al. (2014). Promoter methylation of tumor suppressor genes in pre-neoplastic lesions; potential marker of disease recurrence. J. Exp. Clin. Cancer Res. *33*, 65.

Rex, D.K., Ahnen, D.J., Baron, J.A., Batts, K.P., Burke, C.A., Burt, R.W., Goldblum, J.R., Guillem, J.G., Kahi, C.J., Kalady, M.F., et al. (2012). Serrated lesions of the colorectum: review and recommendations from an expert panel. Am. J. Gastroenterol. *107*, 1315-29.

Richter, H., Slezak, P., Walch, A., Werner, M., Braselmann, H., Jaramillo, E., Ost, A., Hirata, I., Takahama, K., and Zitzelsberger, H. (2003). Distinct chromosomal imbalances in nonpolypoid and polypoid colorectal adenomas indicate different genetic pathways in the development of colorectal neoplasms. Am. J. Pathol. *163*, 287–294.

Ried, T., Heselmeyer-Haddad, K., Blegen, H., Schröck, E., and Auer, G. (1999). Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: a phenotype/genotype correlation. Genes. Chromosomes Cancer *25*, 195–204.

Ried, T., Knutzen, R., Steinbeck, R., Blegen, H., Schröck, E., Heselmeyer, K., du Manoir, S., and Auer, G. (1996). Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. Genes. Chromosomes Cancer *15*, 234–245.

Ried, T. (2009). Homage to Theodor Boveri (1862-1915): Boveri's theory of cancer as a disease of the chromosomes, and the landscape of genomic imbalances in human carcinomas. Environ. Mol. Mutagen. *50*, 593–601.

Ried, T., Hu, Y., Difilippantonio, M.J., Ghadimi, B.M., Grade, M., and Camps, J. (2012). The consequences of chromosomal aneuploidy on the transcriptome of cancer cells. Biochim. Biophys. Acta *1819*, 784–793.

Rieder, C.L., Cole, R.W., Khodjakov, A., and Sluder, G. (1995). The checkpoint delaying anaphase in response to chromosome monoorientation is mediated by an inhibitory signal produced by unattached kinetochores. J. Cell Bol. *130*, 941–948.

Rieder, C.L., Schultz, A., Cole, R., and Sluder, G. (1994). Anaphase onset in vertebrate somatic cells is controlled by a checkpoint that monitors sister kinetochore attachment to the spindle. J. Cell Biol. *127*, 1301–1310.

Roberts, R.B., Min, L., Washington, M.K., Olsen, S.J., Settle, S.H., Coffey, R.J., and Threadgill, D.W. (2002). Importance of epidermal growth factor receptor signaling in establishment of adenomas and maintenance of carcinomas during intestinal tumorigenesis. Proc. Natl. Acad. Sci. U. S. A. *99*, 1521–1526.

Roelen, B.A.J., Cohen, O.S., Raychowdhury, M.K., Chadee, D.N., Zhang, Y., Kyriakis, J.M., Alessandrini, A.A., and Lin, H.Y. (2003). Phosphorylation of threonine 276 in Smad4 is involved in transforming growth factor-beta-induced nuclear accumulation. Am. J. Physiol. Cell Physiol. *285*, C823-30.

Roerink, S.F., Sasaki, N., Lee-Six, H., Young, M.D., Alexandrov, L.B., Behjati, S., Mitchell, T.J., Grossmann, S., Lightfoot, H., Egan, D.A., et al. (2018). Intra-tumour diversification in colorectal cancer at the single-cell level. Nature *556*, 457–462.

Rogers, A.R. (2015). Descriptive Statistics for DNA Sequences. In Lecture Notes on Gene Genealogies, (University of Salt Lake City, Utah, USA), pp. 5–7.

Rothwell, P.M., Wilson, M., Elwin, C.-E., Norrving, B., Algra, A., Warlow, C.P., and Meade, T.W. (2010). Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. Lancet *376*, 1741–1750.

Roy, S., and Majumdar, A.P. (2012). Signaling in colon cancer stem cells. J. Mol. Signal. *7*, 11.

Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S., and Polakis, P. (1996). Binding of GSK3beta to the APC-beta-catenin complex and regulation of complex assembly. Science *272*, 1023–1026.

Rudkin, G.T., and Stollar, B.D. (1977). High resolution detection of DNA-RNA hybrids in situ by indirect immunofluorescence. Nature *265*, 472–473.

Rüschoff, J., Aust, D., and Hartmann, A. (2007). [Colorectal serrated adenoma: diagnostic criteria and clinical implications]. Verh Dtsch Ges Pathol. *91*, 119–125.

Russnes, H.G., Navin, N., Hicks, J., and Borresen-Dale, A.L. (2011). Insight into the heterogeneity of breast cancer through next-generation sequencing. J. Clin. Invest. *121*, 3810–3818.

Rutledge, S.D., and Cimini, D. (2016). Consequences of aneuploidy in sickness and in health. Curr. Opin. Cell Biol. *40*, 41–46.

Sabates-Bellver, J., Van der Flier, L.G., de Palo, M., Cattaneo, E., Maake, C., Rehrauer, H., Laczko, E., Kurowski, M.A., Bujnicki, J.M., Menigatti, M., et al. (2007). Transcriptome profile of human colorectal adenomas. Mol. Cancer Res. *5*, 1263–1275.

Salari, K., Spulak, M.E., Cuff, J., Forster, A.D., Giacomini, C.P., Huang, S., Ko, M.E., Lin, A.Y., van de Rijn, M., and Pollack, J.R. (2012). CDX2 is an amplified lineage-survival oncogene in colorectal cancer. Proc. Natl. Acad. Sci. *109*, E3196–E3205.

Sameer, A.S., and Nissar, S. (2016). Epigenetics in diagnosis of colorectal cancer. Mol. Biol. Res. Commun. *5*, 49–57.

Samuels, Y., and Velculescu, V.E. (2004). Oncogenic mutations of PIK3CA in human cancers. Cell Cycle *3*, 1221–1224.

Samuels, Y., Wang, Z., Bardelli, A., Silliman, N., Ptak, J., Szabo, S., Yan, H., Gazdar, A., Powell, S.M., Riggins, G.J., et al. (2004). High frequency of mutations of the PIK3CA gene in human cancers. Science *304*, 554.

Sandoval, J., Heyn, H. a., Moran, S., Serra-Musach, J., Pujana, M.A., Bibikova, M., and Esteller, M. (2011). Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics *6*, 692–702.

Santini, D., Loupakis, F., Vincenzi, B., Floriani, I., Stasi, I., Canestrari, E., Rulli, E., Maltese, P.E., Andreoni, F., Masi, G., et al. (2008). High concordance of KRAS status between primary colorectal tumors and related metastatic sites: implications for clinical practice. Oncologist *13*, 1270–1275.

Santos, G.C., Zielenska, M., Prasad, M., and Squire, J.A. (2007). Chromosome 6p amplification and cancer progression. J. Clin. Pathol. *60*, 1–7.

Sartor, R.B. (2006). Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. Nat. Clin. Pract. Gastroenterol. Hepatol. *3*, 390–407.Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc. Natl. Acad. Sci. U. S. A. *103*, 1412–1417.

Schlegel, J., Stumm, G., Scherthan, H., Bocker, T., Zirngibl, H., Rüschoff, J., and Hofstädter, F. (1995). Comparative genomic in situ hybridization of colon carcinomas with replication error. Cancer Res. *55*, 6002–6005.

Schmid, D., and Leitzmann, M.F. (2014). Television viewing and time spent sedentary in relation to cancer risk: a meta-analysis. J. Natl. Cancer Inst. *106*.

Schneider, E., Dittrich, M., Böck, J., Nanda, I., Müller, T., Seidmann, L., Tralau, T., Galetzka, D., El Hajj, N., and Haaf, T. (2016). CpG sites with continuously increasing or decreasing methylation from early to late human fetal brain development. Gene *592*, 110–118.

Schwartz, R., and Schäffer, A.A. (2017). The evolution of tumour phylogenetics: principles and practice. Nat. Rev. Genet. *18*, 213–229.

Secretan, B., Straif, K., Baan, R., Grosse, Y., El Ghissassi, F., Bouvard, V., Benbrahim-Tallaa, L., Guha, N., Freeman, C., Galichet, L., et al. (2009). A review of human carcinogens--Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. Lancet Oncol. *10*, 1033–1034.

Sharma, P., Bhunia, S., Poojary, S.S., Tekcham, D.S., Barbhuiya, M.A., Gupta, S., Shrivastav, B.R., and Tiwari, P.K. (2016). Global methylation profiling to identify epigenetic signature of gallbladder cancer and gallstone disease. Tumour Biol. *37*, 14687–14699.

Sheltzer, J.M., Blank, H.M., Pfau, S.J., Tange, Y., George, B.M., Humpton, T.J., Brito, I.L., Hiraoka, Y., Niwa, O., and Amon, A. (2011). Aneuploidy drives genomic instability in yeast. Science *333*, 1026–1030.

Sheltzer, J.M., Ko, J.H., Replogle, J.M., Habibe Burgos, N.C., Chung, E.S., Meehl, C.M., Sayles, N.M., Passerini, V., Storchova, Z., and Amon, A. (2017). Single-chromosome Gains Commonly Function as Tumor Suppressors. Cancer Cell *31*, 240–255.

Sheltzer, J.M., Torres, E.M., Dunham, M.J., and Amon, A. (2012). Transcriptional consequences of aneuploidy. Proc. Natl. Acad. Sci. U. S. A. *109*, 12644–12649.

Shinya, H., and Wolff, W.I. (1979). Morphology, anatomic distribution and cancer potential of colonic polyps. Ann. Surg. *190*, 679–683.

Sillars-Hardebol, A.H., Carvalho, B., Van Engeland, M., Fijneman, R.J., and Meijer, G.A. (2012). The adenoma hunt in colorectal cancer screening: Defining the target. J. Pathol. *226*, 1–6.

Silva, A.-L., Dawson, S.N., Arends, M.J., Guttula, K., Hall, N., Cameron, E.A., Huang, T.H.-M., Brenton, J.D., Tavaré, S., Bienz, M., et al. (2014). Boosting Wnt activity during colorectal cancer progression through selective hypermethylation of Wnt signaling antagonists. BMC Cancer *14*, 891.

Snover, D.C. (2011). Update on the serrated pathway to colorectal carcinoma. Hum. Pathol. *42*, 1–10.

Snover, D.C., Jass, J.R., Fenoglio-Preiser, C., and Batts, K.P. (2005). Serrated polyps of the large intestine: a morphologic and molecular review of an evolving concept. Am. J. Clin. Pathol. *124*, 380–391.

Soetikno, R.M., Kaltenbach, T., Rouse, R. V, Park, W., Maheshwari, A., Sato, T., Matsui, S., and Friedland, S. (2008). Prevalence of nonpolypoid (flat and depressed) colorectal neoplasms in asymptomatic and symptomatic adults. JAMA *299*, 1027–1035.

Song, M., Garrett, W.S., and Chan, A.T. (2015). Nutrients, foods, and colorectal cancer prevention. Gastroenterology *148*, 1244–60.e16.

Sottoriva, A., Kang, H., Ma, Z., Graham, T.A., Salomon, M.P., Zhao, J., Marjoram, P., Siegmund, K., Press, M.F., Shibata, D., et al. (2015). A Big Bang model of human colorectal tumor growth. Nat. Genet. *47*, 209–216.

Sottoriva, A., Spiteri, I., Piccirillo, S.G.M., Touloumis, A., Collins, V.P., Marioni, J.C., Curtis, C., Watts, C., and Tavaré, S. (2013). Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc. Natl. Acad. Sci. U. S. A. *110*, 4009–4014.

Sparks, A.B., Morin, P.J., Vogelstein, B., and Kinzler, K.W. (1998). Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. Cancer Res. *58*, 1130–1134.

Stepanenko, A.A., and Kavsan, V.M. (2012). Evolutionary karyotypic theory of cancer versus conventional cancer gene mutation theory. Biopolym. Cell *28*, 267–280.

Stingele, S., Stoehr, G., Peplowska, K., Cox, J., Mann, M., and Storchova, Z. (2012). Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. Mol. Syst. Biol. *8*, 608.

Stock, C., Ihle, P., Schubert, I., and Brenner, H. (2011). Colonoscopy and fecal occult blood test use in Germany: results from a large insurance-based cohort. Endoscopy *43*, 771–781.

Storchová, Z. (2012). The causes and consequences of aneuploidy in eukaryotic cells. In Aneuploidy in Health and Disease, Z. Storchová (InTech Open), pp. 1–22. Online: https://www.intechopen.com/books/aneuploidy-in-health-and-disease/the-effect-of-aneuploidy-on-physiology-of-eukaryotic-cell, accessed 07.16.18.

Strum, W.B. (2016). Colorectal Adenomas. N. Engl. J. Med. *374*, 1065–1075.

Subtil, C., Guérin, E., Schneider, A., Chenard, M.-P., Martin, E., Domon-Dell, C., Duluc, I., Brabletz, T., Kedinger, M., Duclos, B., et al. (2007). Frequent rearrangements and amplification of the CDX2 homeobox gene in human sporadic colorectal cancers with chromosomal instability. Cancer Lett. *247*, 197–203.

Sugai, T., Yoshida, M., Eizuka, M., Uesugii, N., Habano, W., Otsuka, K., Sasaki, A., Yamamoto, E., Matsumoto, T., and Suzuki, H. (2017). Analysis of the DNA methylation level of cancer-related genes in colorectal cancer and the surrounding normal mucosa. Clin. Epigenetics *9*, 55.

Takagi, Y., Kohmura, H., Futamura, M., Kida, H., Tanemura, H., Shimokawa, K., and Saji, S. (1996). Somatic alterations of the DPC4 gene in human colorectal cancers in vivo. Gastroenterology *111*, 1369–1372.

Takayama, T., Katsuki, S., Takahashi, Y., Ohi, M., Nojiri, S., Sakamaki, S., Kato, J., Kogawa, K., Miyake, H., and Niitsu, Y. (1998). Aberrant crypt foci of the colon as precursors of adenoma and cancer. N. Engl. J. Med. *339*, 1277–1284.

The Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330–337.

Thirlwell, C., Will, O.C.C., Domingo, E., Graham, T.A., McDonald, S.A.C., Oukrif, D., Jeffrey, R., Gorman, M., Rodriguez-Justo, M., Chin-Aleong, J., et al. (2010). Clonality assessment and clonal ordering of individual neoplastic crypts shows polyclonality of colorectal adenomas. Gastroenterology *138*, 1441–1454, 1454.e1-7.

Thompson, S.L., Bakhoum, S.F., and Compton, D.A. (2010). Mechanisms of chromosomal instability. Curr. Biol. *20*, R285-95.

Thompson, S.L., and Compton, D.A. (2008). Examining the link between chromosomal instability and aneuploidy in human cells. J. Cell Biol. *180*, 665–672.

Torlakovic, E., Skovlund, E., Snover, D.C., Torlakovic, G., and Nesland, J.M. (2003). Morphologic reappraisal of serrated colorectal polyps. Am. J. Surg. Pathol. *27*, 65–81.

Torres, E.M., Sokolsky, T., Tucker, C.M., Chan, L.Y., Boselli, M., Dunham, M.J., and Amon, A. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science *317*, 916–924.

Tost, J., Dunker, J., and Gut, I.G. (2003). Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. Biotechniques *35*, 152–156.

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., and Issa, J.P. (1999). CpG island methylator phenotype in colorectal cancer. Proc. Natl. Acad. Sci. U. S. A. *96*, 8681–8686.

Tsilidis, K.K., Kasimis, J.C., Lopez, D.S., Ntzani, E.E., and Ioannidis, J.P. (2015). Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. BMJ *350*, g7607.

Tycko, B. (2000). Epigenetic gene silencing in cancer. J. Clin. Invest. *105*, 401–407.

Uchi, R., Takahashi, Y., Niida, A., Shimamura, T., Hirata, H., Sugimachi, K., Sawada, G., Iwaya, T., Kurashige, J., Shinden, Y., et al. (2016). Integrated Multiregional Analysis Proposing a New Model of Colorectal Cancer Evolution. PLoS Genet. *12*, e1005778.

van Heijningen, E.-M.B., Lansdorp-Vogelaar, I., Kuipers, E.J., Dekker, E., Lesterhuis, W., Ter Borg, F., Vecht, J., De Jonge, V., Spoelstra, P., Engels, L., et al. (2013). Features of adenoma and colonoscopy associated with recurrent colorectal neoplasia based on a large community-based study. Gastroenterology *144*, 1410–1418.

van Rijnsoever, M., Grieu, F., Elsaleh, H., Joseph, D., and Iacopetta, B. (2002). Characterisation of colorectal cancers showing hypermethylation at multiple CpG islands. Gut *51*, 797–802.

van Roy, F. (2014). Beyond E-cadherin: roles of other cadherin superfamily members in cancer. Nat. Rev. Cancer *14*, 121–134.Vatn, M.H., and Stalsberg, H. (1982). The prevalence of polyps of the large intestine in Oslo: an autopsy study. Cancer *49*, 819–825.

Väyrynen, S. (2016). Histological and molecular features of serrated colorectal adenocarcinoma and its precursor lesions. Academic dissertation in health and bioscience. University of Oulu, Finland.

Vilar, E., and Gruber, S.B. (2010). Microsatellite instability in colorectal cancer-the stable evidence. Nat. Rev. Clin. Oncol. *7*, 153–162.

Visvader, J.E. (2011). Cells of origin in cancer. Nature *469*, 314–322.

Vogelstein, B., and Kinzler, K.W. (2015). The Path to Cancer --Three Strikes and You're Out. N. Engl. J. Med. *373*, 1895–1898.

Vogelstein, B., Lane, D., and Levine, A.J. (2000). Surfing the p53 network. Nature *408*, 307–310.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558.

von Hansemann, D. (1890). Ueber asymmetrische Zelltheilung in Epithelkrebsen und deren biologische Bedeutung. Virchows Arch. Pathol. Anat. Physiol. Klin. Med. *191*, 299–326.

Voorham, Q.J.M., Janssen, J., Tijssen, M., Snellenberg, S., Mongera, S., van Grieken, N.C.T., Grabsch, H., Kliment, M., Rembacken, B.J., Mulder, C.J.J., et al. (2013). Promoter methylation of Wnt-antagonists in polypoid and nonpolypoid colorectal adenomas. BMC Cancer *13*, 603.

Vousden, K.H., and Lane, D.P. (2007). p53 in health and disease. Nat. Rev. Mol. Cell Biol. *8*, 275–283.

Walsh, M.D., Young, J.P., Leggett, B.A., Williams, S.H., Jass, J.R., and McGuckin, M.A. (2007). The MUC13 cell surface mucin is highly expressed by human colorectal carcinomas. Hum. Pathol. *38*, 883–892.

Walsh, R.M., Ackroyd, F.W., and Shellito, P.C. (1992). Endoscopic resection of large sessile colorectal polyps. Gastrointest. Endosc. *38*, 303–309.

Walther, A., Johnstone, E., Swanton, C., Midgley, R., Tomlinson, I., and Kerr, D. (2009). Genetic prognostic and predictive markers in colorectal cancer. Nat. Rev. Cancer *9*, 489–499.

Wang, D., Yan, L., Hu, Q., Sucheston, L.E., Higgins, M.J., Ambrosone, C.B., Johnson, C.S., Smiraglia, D.J., and Liu, S. (2012). IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. Bioinformatics *28*, 729–730.

Wang, H., Liang, L., Fang, J.-Y., and Xu, J. (2016). Somatic gene copy number alterations in colorectal cancer: new quest for cancer drivers and biomarkers. Oncogene *35*, 2011–2019.

Wangsa, D., Chowdhury, S.A., Ryott, M., Gertz, E.M., Elmberger, G., Auer, G., Lundqvist, E.Å., Küffer, S., Ströbel, P., Schäffer, A.A., et al. (2016). Phylogenetic analysis of multiple FISH markers in oral tongue squamous cell carcinoma suggests that a diverse distribution of copy number changes is associated with poor prognosis. Int. J. Cancer *138*, 98–109.

Wangsa, D., Heselmeyer-Haddad, K., Ried, P., Eriksson, E., Schäffer, A.A., Morrison, L.E., Luo, J., Auer, G., Munck-Wikland, E., Ried, T., et al. (2009). Fluorescence in situ hybridization markers for prediction of cervical lymph node metastases. Am. J. Pathol. *175*, 2637–2645.

Weaver, B.A.A., and Cleveland, D.W. (2006). Does aneuploidy cause cancer? Curr. Opin. Cell Biol. *18*, 658–667.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat. Genet. *39*, 457–466.

Weisburger, J.H., and Wynder, E.L. (1987). Etiology of colorectal cancer with emphasis on mechanism of action and prevention. Important Adv. Oncol. 197–220.

Weisenberger, D.J., Siegmund, K.D., Campan, M., Young, J., Long, T.I., Faasse, M.A., Kang, G.H., Widschwendter, M., Weener, D., Buchanan, D., et al. (2006). CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat. Genet. *38*, 787–793.

Welch, H.G., and Robertson, D.J. (2016). Colorectal Cancer on the Decline - Why Screening Can't Explain it All. N. Engl. J. Med. *374*, 1605–1607.

Winawer, S.J., Zauber, A.G., Ho, M.N., O'Brien, M.J., Gottlieb, L.S., Sternberg, S.S., Waye, J.D., Schapiro, M., Bond, J.H., and Panish, J.F. (1993). Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N. Engl. J. Med. *329*, 1977–1981.

Wippold, F.J., and Perry, A. (2007). Neuropathology for the neuroradiologist: fluorescence in situ hybridization. Am. J. Neuroradiol. *28*, 406–410.

World Medical Association (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA *310*, 2191–2194.

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., and Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer Cell *26*, 577–590.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. *13*, 335–340.

Yuen, K.W.Y. (2010). Chromosome Instability (CIN), Aneuploidy and Cancer. In Encyclopedia of Life Sciences, (Chichester, UK: John Wiley and Sons), pp. 1–10.

Zauber, A.G., Winawer, S.J., O'Brien, M.J., Lansdorp-Vogelaar, I., van Ballegooijen, M., Hankey, B.F., Shi, W., Bond, J.H., Schapiro, M., Panish, J.F., et al. (2012). Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N. Engl. J. Med. *366*, 687–696.

Zeng, Z.-S., Weiser, M.R., Kuntz, E., Chen, C.-T., Khan, S.A., Forslund, A., Nash, G.M., Gimbel, M., Yamaguchi, Y., Culliford, A.T., et al. (2008). c-Met gene amplification is associated with advanced stage colorectal cancer and liver metastases. Cancer Lett. *265*, 258–269.

Zentrum für Krebsregisterdaten im Robert Koch Institut (2016). Bericht zum Krebsgeschehen in Deutschland 2016 (Berlin, Germany: Robert Koch Institut). Online: https://www.krebsdaten.de/Krebs/DE/Content/Publikationen/Krebsgeschehen/Krebsgeschehen_download.pdf?__blob=publicationFile, accessed 08.08.17.

Zhang, N., Ge, G., Meyer, R., Sethi, S., Basu, D., Pradhan, S., Zhao, Y.-J., Li, X.-N., Cai, W.-W., El-Naggar, A.K., et al. (2008). Overexpression of Separase induces aneuploidy and mammary tumorigenesis. Proc. Natl. Acad. Sci. U. S. A. *105*, 13033–13038.

7   SUPPLEMENTARY TABLES

Supplemental Table 1. BAC clone contig probe panels for miFISH.

Panel	Fluorophore	Gene	Locus	BAC clone	Build (UCSC)	Gene location	Start	Stop	Total bp	Contig size°
I	DY-415-dUTP (Dyomics)	MYC	8q24.21	RP11-1136L8	GRCh37/hg19	128,748,315 - 128,753,680	128,555,811	128,714,938	159,128	694,889
				CTD-3056O22	GRCh37/hg19		128,660,772	128,808,586	147,815	
				RP11-55J15	GRCh37/hg19		128,870,291	129,064,392	194,102	
				RP11-709E21	NCBI 37/hg3		129,154,373	129,250,700	96,328	
	DY-505-dUTP (Dyomics)	CCND1	11q13.3	CTD-2507F7	GRCh37/hg19	69,228,803 -69,242,171	69,332,248	69,523,715	191,468	664,147
				RP11-300I6	GRCh37/hg19		69,453,281	69,614,785	161,505	
				RP11-186D19	NCBI 37/hg3		69,846,911	69,996,395	149,485	
	DY-547-dUTP (Dyomics)	SMAD4	18q21.2	CTD-3236H18	GRCh37/hg19	48,494,387 -48,584,614	48,285,441	48,479,444	194,004	349,039
				RP11-729G3	GRCh37/hg19		48,478,296	48,634,480	156,185	
	DY-590-dUTP (Dyomics)	CDX2	13q12.2	RP11-328P22	GRCh37/hg19	28,536,274 -28,543,505	28,488,636	28,517,161	28,526	297,570
				RP11-136G6	GRCh37/hg19		28,502,226	28,683,446	181,221	
				RP11-486I19	GRCh37/hg19		28,629,619	28,786,206	156,588	
	DY-651-dUTP (Dyomics)	PIK3CA	3q26.32	RP11-355N16	GRCh37/hg19	178,866,311 - 178,952,497	178,711,249	178,905,842	194,594	554,087
				RP11-245C23	GRCh37/hg19		178,830,412	179,014,819	184,408	
				RP11-145M9	GRCh37/hg19		179,091,279	179,265,336	174,058	
II	DY-415-dUTP (Dyomics)	EGFR	7p11.2	RP11-23F4	GRCh37/hg19	55,086,725 - 55,275,031	54,912,535	55,082,244	169,710	506,308
				RP11-815K24	GRCh37/hg19		55,049,736	55,246,161	196,426	
				RP11-97C21	GRCh37/hg19		55,239,327	55,418,843	179,517	
	DY-505-dUTP (Dyomics)	COX2 (PTGS2)	1q31.1	RP11-1149C23	GRCh37/hg19	186,640,944 - 186,649,559	186,290,778	186,428,689	137,912	503,451
				CTD-2509N15	GRCh37/hg19		186,448,188	186,651,391	203,204	
				RP11-809N5	GRCh37/hg19		186,606,706	186,794,229	187,524	
	DY-547-dUTP (Dyomics)	APC	5q22.2	RP11-159K7	GRCh37/hg19	112,043,202 - 112,181,936	111,842,788	112,021,220	178,433	467,340
				RP11-124K18	GRCh37/hg19		112,014,487	112,198,755	184,269	
				RP11-1104F23	GRCh37/hg19		112,134,536	112,310,128	175,593	
	DY-590-dUTP (Dyomics)	CLIC1	6p21.33	RP11-184F16	GRCh38/hg38	31,698,358 - 31,704,341	31,437,976	31,628,870	190,895	535,964
				RP11-1104F14	GRCh38/hg38		31,627,621	31,797,832	170,212	
				CTD-2652P14	GRCh38/hg38		31,757,045	31,973,940	216,896	
	DY-651-dUTP (Dyomics)	SMAD7	18q21.1	RP11-164C15	GRCh37/hg19	46,446,223 - 46,469,177	46,182,137	46,332,593	150,457	505,472
				RP11-956P18	GRCh37/hg19		46,357,409	46,552,744	195,336	
				RP11-15F12	GRCh37/hg19		46,516,511	46,687,609	171,099	
III	DY-415-dUTP (Dyomics)	ZNF217	20q13.2	RP4-669H2	GRCh37/hg19	51,617,017 - 51,633,043	51,901,748	52,021,567	119,820	465,525
				RP11-299B10	GRCh37/hg19		52,026,197	52,239,348	213,152	
				CTD-2573N1	GRCh37/hg19		52,228,296	52,367,273	138,978	
	DY-505-dUTP (Dyomics)	CDH1	16q22.2	RP11-615I2	NCBI 37/hg3	68,771,195 - 68,869,444	68,576,846	68,693,621	116,776	344,455
				RP11-354M1	NCBI 37/hg3		68,693,622	68,875,324	181,703	
				RP11-354N7	GRCh37/hg19		68,761,021	68,921,301	160,281	
	DY-547-dUTP (Dyomics)	TP53	17p13.1	CTD-3028L6	GRCh37/hg19	7,512,469 - 7,531,642	7,420,243	7,596,553	176,311	382,551
				RP11-89D11	GRCh37/hg19		7,495,729	7,663,042	167,314	
				RP11-709J3	GRCh37/hg19		7,631,225	7,802,794	171,570	
	DY-590-dUTP (Dyomics)	HER2 (ERBB2)	17q12	RP11-94L15	GRCh37/hg19	37,856,254 - 37,884,915	37,811,809	37,973,609	161,801	425,589
				CTD-3211L18	GRCh37/hg19		37,953,501	38,194,528	241,028	
				CTD-2248E4	GRCh37/hg19		38,115,365	38,237,398	122,034	
	CCP10 (CytoTest)	CEP10	10							

Clones covering the gene are marked in **bold**.
°Length in base pairs (bp). Contig size is defined by stop-base of the distal clone minus start-base of the proximal clone.
Locations were depicted from Genome Browser (https://ucsc.genome.edu) and BAC clones were provided by BACPAC Resources Center (http://bacpacresources.org).

Supplemental Table 2. List of 35 *top genes*-DMPs of *comparison A*. Listed are genes with q≤0.05, fold change >10% and absolute change ≥|0.1|. Sex-chromosome probes were excluded. ●, more than one probe was significant within this gene.

CpG ID	β-value Recurrent adenomas	β-value Primary adenomas	Fold change	Absolute change	P-value	Q-value (adjusted P-value)	Chr.	Coordinate	Gene symbol (UCSC)	Methylation status	Gene region (UCSC)	Relation to CGI (UCSC)
cg07362278	0.728	0.623	1.168	0.105	1.61E-08	0.002	19	45737011	EXOC3L2	hypermeth.	5'UTR	N shore
cg17263323	0.649	0.535	1.213	0.114	4.42E-07	0.007	4	5238459	STK32B	hypermeth.	Body	Open sea
cg14796563	0.791	0.655	1.206	0.135	5.89E-06	0.017	19	1287576	EFNA2	hypermeth.	Body	Island
cg17001652	0.615	0.513	1.198	0.101	1.36E-05	0.022	4	78432403	CXCL13	hypermeth.	TSS1500	Open sea
cg01452189	0.603	0.492	1.227	0.112	1.60E-05	0.023	16	25189426	LCMT1	hypermeth.	3'UTR	Open sea
cg19880608	0.714	0.605	1.181	0.110	2.97E-05	0.025	5	112568208	MCC	hypermeth.	Body	Open sea
cg24671202	0.589	0.416	1.416	0.173	3.45E-05	0.026	6	116938100	RSPH4A	hypermeth.	1stExon	Open sea
cg01809217	0.716	0.613	1.169	0.103	3.93E-05	0.026	1	240656342	GREM2●	hypermeth.	Body	Island
cg14847483	0.589	0.456	1.291	0.133	4.93E-05	0.027	15	34516640	TMEM85●	hypermeth.	TSS1500	Open sea
cg07573078	0.667	0.560	1.191	0.107	5.34E-05	0.028	1	207626214	CR2	hypermeth.	TSS1500	N shore
cg03206537	0.455	0.350	1.300	0.105	6.24E-05	0.028	20	44521739	CTSA	hypermeth.	Body	S shore
cg26699569	0.497	0.377	1.319	0.120	6.96E-05	0.030	4	81952653	BMP3	hypermeth.	1stExon	Island
cg06402590	0.671	0.568	1.182	0.103	7.18E-05	0.030	16	2820155	SRRM2	hypermeth.	Body	S shelf
cg12301695	0.569	0.454	1.255	0.116	8.27E-05	0.030	2	165812159	SLC38A11	hypermeth.	TSS200	Open sea
cg16895026	0.693	0.569	1.218	0.124	7.90E-05	0.030	4	30745743	PCDH7	hypermeth.	Body	Open sea
cg25198275	0.645	0.524	1.232	0.122	8.09E-05	0.030	12	29650692	OVCH1	hypermeth.	TSS200	Open sea
cg08496964	0.595	0.495	1.203	0.100	1.14E-04	0.032	16	8729419	C16orf68	hypermeth.	Body	Open sea
cg24553547	0.755	0.654	1.154	0.101	1.22E-04	0.032	17	19247920	MIR1180	hypermeth.	TSS200	Island
cg00608965	0.629	0.506	1.243	0.123	1.33E-04	0.033	12	91572191	DCN	hypermeth.	1stExon	Open sea
cg01946191	0.473	0.373	1.270	0.101	1.52E-04	0.035	12	54444463	HOXC4	hypermeth.	5'UTR	N shore
cg00606880	0.607	0.505	1.203	0.102	1.91E-04	0.037	10	63213784	TMEM26	hypermeth.	TSS1500	S shore
cg12949483	0.589	0.477	1.233	0.111	2.28E-04	0.039	15	34516777	TMEM85●	hypermeth.	TSS1500	Open sea
cg22758700	0.053	0.178	0.295	-0.125	2.26E-04	0.039	7	99156126	ZNF655	hypometh.	TSS1500	Island
cg08325898	0.737	0.623	1.184	0.114	2.50E-04	0.040	8	71096899	NCOA2	hypermeth.	Body	Open sea
cg23629643	0.680	0.579	1.174	0.101	2.47E-04	0.040	4	77547320	SHROOM3	hypermeth.	Body	Open sea
cg11612852	0.707	0.597	1.185	0.110	3.09E-04	0.042	4	21699745	KCNIP4	hypermeth.	5'UTR	Open sea
cg17078427	0.473	0.354	1.336	0.119	3.57E-04	0.044	3	170137552	CLDN11	hypermeth.	Body	Island
cg24811290	0.579	0.477	1.213	0.102	3.58E-04	0.044	4	159092553	FAM198B	hypermeth.	5'UTR	Open sea
cg17423207	0.411	0.309	1.331	0.102	3.79E-04	0.044	19	17958892	JAK3	hypermeth.	TSS200	S shore
cg18411108	0.494	0.382	1.295	0.112	3.91E-04	0.045	3	63955727	ATXN7	hypermeth.	Body	Open sea
cg04391463	0.642	0.533	1.205	0.109	3.97E-04	0.045	1	240656667	GREM2●	hypermeth.	Body	Island
cg09826050	0.691	0.557	1.240	0.134	4.19E-04	0.046	1	217310174	ESRRG	hypermeth.	5'UTR	N shore
cg04845649	0.647	0.544	1.190	0.103	4.37E-04	0.046	16	67199463	HSF4	hypermeth.	Body	Island
cg02577267	0.781	0.680	1.148	0.101	5.30E-04	0.048	1	240656513	GREM2●	hypermeth.	Body	Island
cg19063972	0.576	0.426	1.351	0.150	5.59E-04	0.049	13	95364908	SOX21	hypermeth.	TSS1500	Island

Supplemental Table 3. List of 7 *top genes*-DMPs of *comparison B*. Listed are genes with q≤0.05, fold change >10% and absolute change ≥|0.05|. Sex-chromosome probes were excluded.

CpG ID	β-value Recurrent adenomas	β-value Primary adenomas	Fold change	Absolute change	P-value	Q-value (adjusted P-value)	Chr.	Coordinate	Gene symbol (UCSC)	Methylation status	Gene region (UCSC)	Relation to CGI (UCSC)
cg07362278	0.728	0.646	1.127	0.082	9.06E-09	0.001	19	45737011	EXOC3L2	hypermeth.	5'UTR	N shore
cg17263323	0.649	0.566	1.146	0.082	1.22E-06	0.008	4	5238459	STK32B	hypermeth.	Body	Open sea
cg09635768	0.581	0.525	1.107	0.056	7.18E-06	0.018	1	8601318	RERE	hypermeth.	Body	Open sea
cg13771471	0.582	0.525	1.110	0.057	1.23E-05	0.022	10	1505595	ADARB2	hypermeth.	Body	N shore
cg22177309	0.659	0.591	1.115	0.068	3.00E-05	0.033	19	48260285	GLTSCR2	hypermeth.	3'UTR	Open sea
cg22758700	0.053	0.141	0.374	-0.088	4.44E-05	0.036	7	99156126	ZNF655●	hypometh.	TSS1500	Island
cg21950525	0.118	0.200	0.591	-0.082	5.79E-05	0.040	7	99155767	ZNF655●	hypometh.	TSS1500	N shore

Supplemental Table 4. List of best-50 _top genes_-DMPs of _comparison C_. Listed are genes with q≤0.001, fold change >10% and absolute change ≥|0.1|. Sex-chromosome probes were excluded. Total number of DMPs in this comparison were 347. ●, more than one probe was significant within this gene.

CpG ID	β-value Recurrent adenomas	β-value Primary a-denomas	Fold change	Abso-lute change	_P_-value	_Q_-value (adjusted _P_-value)	Chr	Coordinate	Gene symbol (UCSC)	Methyla-tion status	Gene re-gion (UCSC)	Relation to CGI (UCSC)
cg16879115	0.275	0.421	0.654	-0.146	3.18E-27	8.64E-22	12	7819180	_APOBEC1_	hypometh.	TSS1500	Open sea
cg19378216	0.369	0.524	0.704	-0.155	1.77E-25	1.61E-20	20	31870775	_C20orf114_	hypometh.	TSS200	Open sea
cg07074042	0.366	0.504	0.727	-0.138	3.87E-23	2.63E-18	22	30783944	_RNF215_	hypometh.	TSS1500	S shore
cg23875663	0.329	0.447	0.735	-0.118	4.93E-23	2.68E-18	20	17299350	_PCSK2_	hypometh.	Body	S shelf
cg02265239	0.336	0.467	0.721	-0.130	9.73E-23	4.41E-18	14	59748448	_DAAM1_	hypometh.	Body	Open sea
cg25730020	0.458	0.586	0.781	-0.128	1.13E-21	3.40E-17	18	56701518	_LOC390858_	hypometh.	TSS1500	Open sea
cg15676837	0.646	0.771	0.838	-0.125	7.45E-20	1.28E-15	1	202137199	_PTPRV_	hypometh.	Body	Open sea
cg08590987	0.484	0.586	0.825	-0.102	1.02E-18	1.27E-14	8	110702751	_GOLSYN_	hypometh.	5'UTR	N shore
cg01156373	0.435	0.545	0.798	-0.110	1.87E-17	1.54E-13	2	213248884	_ERBB4_	hypometh.	Body	Open sea
cg23696618	0.426	0.537	0.793	-0.111	2.37E-17	1.74E-13	18	61583699	_SERPINB10_	hypometh.	Body	Open sea
cg14479014	0.495	0.621	0.798	-0.125	3.20E-17	2.23E-13	6	49460031	_CENPQ_	hypometh.	3'UTR	Open sea
cg02320784	0.619	0.752	0.822	-0.134	4.54E-17	2.94E-13	12	26195988	_RASSF8_	hypometh.	5'UTR	Open sea
cg13302785	0.191	0.323	0.592	-0.132	1.19E-16	6.90E-13	3	51104051	_DOCK3_	hypometh.	Body	Open sea
cg12768447	0.465	0.616	0.755	-0.151	2.24E-16	1.22E-12	10	69425930	_CTNNA3_	hypometh.	TSS1500	Open sea
cg15651267	0.470	0.590	0.797	-0.120	3.92E-16	2.01E-12	5	167028535	_ODZ2_	hypometh.	Body	Open sea
cg13455326	0.283	0.394	0.718	-0.111	5.68E-16	2.71E-12	19	10043341	_OLFM2_	hypometh.	Body	N shelf
cg15133719	0.624	0.507	1.231	0.117	7.64E-16	3.53E-12	2	223163804	_PAX3•_	hypermeth	TSS200	Island
cg23530543	0.512	0.621	0.824	-0.110	9.70E-16	4.19E-12	11	124788414	_HEPN1_	hypometh.	TSS1500	N shelf
cg01687040	0.283	0.391	0.724	-0.108	3.46E-15	1.21E-11	12	55043559	_DCD_	hypometh.	TSS1500	Open sea
cg07000713	0.328	0.439	0.747	-0.111	7.19E-15	2.22E-11	12	62463441	_FAM19A2_	hypometh.	5'UTR	Open sea
cg06119894	0.408	0.282	1.448	0.126	1.00E-14	2.75E-11	11	105480347	_GRIA4_	hypermeth	TSS1500	N shore
cg24242823	0.574	0.441	1.299	0.132	1.01E-14	2.75E-11	7	24323675	_NPY_	hypermeth	TSS200	Island
cg10075287	0.513	0.619	0.829	-0.106	2.07E-14	4.94E-11	2	180727677	_ZNF385B_	hypometh.	TSS1500	S shore
cg11019211	0.526	0.366	1.436	0.160	2.47E-14	5.42E-11	17	56833043	_PPM1E_	hypermeth	TSS200	Island
cg07999090	0.456	0.562	0.810	-0.107	2.77E-14	6.02E-11	13	28882299	_FLT1_	hypometh.	Body	Open sea
cg06186245	0.429	0.570	0.752	-0.142	5.57E-14	1.09E-10	3	100245062	_TMEM45A_	hypometh.	5'UTR	Open sea
cg21049958	0.338	0.442	0.764	-0.104	6.00E-14	1.16E-10	19	42134128	_CEACAM4_	hypometh.	TSS1500	Open sea
cg26929012	0.409	0.264	1.548	0.145	1.09E-13	1.89E-10	19	57988860	_ZNF772_	hypermeth	1stExon	Open sea
cg26822330	0.607	0.710	0.855	-0.103	1.22E-13	2.10E-10	13	36867876	_C13orf38•_	hypometh.	Body	N shelf
cg07784084	0.417	0.518	0.804	-0.101	1.48E-13	2.43E-10	3	26751141	_LRRC3B_	hypometh.	5'UTR	Open sea
cg03646889	0.516	0.402	1.284	0.114	2.30E-13	3.46E-10	1	99729807	_LPPR4_	hypermeth	TSS200	N shore
cg01866434	0.612	0.718	0.852	-0.106	2.89E-13	4.23E-10	5	55479377	_ANKRD55_	hypometh.	Body	Open sea
cg24389611	0.624	0.761	0.821	-0.136	7.56E-13	9.61E-10	4	39056150	_KLHL5_	hypometh.	5'UTR	Open sea
cg13464924	0.350	0.234	1.500	0.117	1.38E-12	1.61E-09	1	190447603	_FAM5C_	hypermeth	TSS1500	Open sea
cg17735563	0.557	0.701	0.794	-0.145	1.70E-12	1.96E-09	2	33738079	_RASGRP3_	hypometh.	5'UTR	Open sea
cg22545168	0.314	0.417	0.753	-0.103	1.73E-12	1.98E-09	19	54866514	_LAIR1_	hypometh.	3'UTR	Open sea
cg21983531	0.491	0.616	0.797	-0.125	3.63E-12	3.70E-09	7	74075317	_GTF2I_	hypometh.	5'UTR	S shelf
cg16545821	0.470	0.585	0.802	-0.116	6.16E-12	5.74E-09	1	204118840	_ETNK2_	hypometh.	Body	N shore
cg06577005	0.484	0.375	1.290	0.109	6.28E-12	5.81E-09	1	26758228	_DHDDS_	hypermeth	TSS1500	N shore
cg13888600	0.485	0.381	1.272	0.104	1.26E-11	1.04E-08	15	89911317	_MIR9-3_	hypermeth	Body	Island
cg10208609	0.595	0.696	0.856	-0.100	1.62E-11	1.27E-08	1	90331229	_LRRC8D_	hypometh.	5'UTR	Open sea
cg08567913	0.590	0.732	0.807	-0.141	1.94E-11	1.49E-08	11	114082594	_ZBTB16_	hypometh.	Body	Open sea
cg08638917	0.520	0.641	0.811	-0.121	2.12E-11	1.59E-08	16	58299153	_CCDC113_	hypometh.	Body	Open sea
cg14544666	0.526	0.414	1.273	0.113	2.59E-11	1.89E-08	5	179634789	_RASGEF1C_	hypermeth	5'UTR	Island
cg10201297	0.466	0.279	1.667	0.186	2.69E-11	1.94E-08	1	32930832	_ZBTB8B_	hypermeth	5'UTR	Island
cg08321129	0.587	0.451	1.300	0.136	2.79E-11	1.99E-08	13	78493878	_EDNRB•_	hypermeth	TSS1500	S shore
cg00341204	0.676	0.776	0.871	-0.100	3.11E-11	2.17E-08	19	11520483	_RGL3_	hypometh.	Body	S shelf
cg14154381	0.453	0.560	0.808	-0.108	3.41E-11	2.33E-08	15	25476386	_SNORD115-15_	hypometh.	Body	Open sea
cg04658772	0.466	0.317	1.469	0.149	3.81E-11	2.56E-08	16	77822433	_VAT1L_	hypermeth	TSS200	Island
cg13780718	0.316	0.467	0.676	-0.152	4.69E-11	3.00E-08	20	44513768	_ZSWIM1_	hypometh.	3'UTR	Open sea

Supplemental Table 5. ASNs per miFISH gene marker in matched pair-samples. ASNs represent the mean signal count per probe of the total number of analyzed nuclei within the respective sample.

Sample ID	COX2	PIK3CA	APC	CLIC1	EGFR	MYC	CCND1	CDX2	CDH1	TP53	HER2	SMAD7	SMAD4	ZNF217
P 1a	2.00	2.00	2.00	2.16	2.00	2.00	2.02	3.23	2.00	1.99	2.04	2.01	1.99	2.01
P 1b	2.00	2.00	1.98	2.04	2.00	2.00	2.02	3.75	2.00	1.98	2.00	2.00	2.00	1.98
P 2a	2.00	2.01	2.00	2.02	2.00	2.00	2.00	2.01	1.99	1.97	2.00	1.99	1.99	2.00
P 2b	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.04	2.00	2.00	2.00	2.00	1.99	2.00
P 3a	2.00	2.00	2.00	2.01	2.00	2.00	2.00	2.00	1.96	1.99	2.00	1.99	1.99	2.00
P 3b	2.02	2.11	2.05	2.18	2.22	2.09	2.04	2.24	2.16	2.07	2.11	2.00	2.00	2.19
P 4a	2.00	2.00	2.15	2.03	2.23	2.26	2.00	2.32	2.08	1.51	2.27	2.19	2.18	2.61
P 4b	1.99	2.00	1.98	2.01	2.00	2.00	2.00	2.03	2.01	1.61	2.15	2.00	2.00	2.01
P 5a	2.00	2.00	1.99	1.99	2.05	2.00	2.00	2.05	1.99	1.98	2.00	1.99	1.99	2.00
P 5b	1.99	2.00	1.99	2.00	2.00	2.00	2.00	2.01	2.00	2.00	2.00	2.01	2.00	2.01
P 6a	2.00	1.99	2.00	2.04	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
P 6b	2.00	2.00	1.98	2.01	2.00	2.00	2.00	2.01	1.99	1.98	1.99	2.00	2.00	2.00
P 7a	1.99	2.00	2.01	2.01	2.83	2.01	2.00	2.00	2.00	1.99	2.00	2.02	2.00	2.04
P 7b†	2.76	3.04	3.52	3.57	3.57	3.57	2.80	3.57	2.72	3.49	3.59	2.75	2.74	3.57
P 8a	2.01	2.00	2.00	2.04	2.01	2.01	2.01	2.02	1.98	2.00	2.01	2.01	2.01	2.01
P 8b	2.00	2.00	1.99	2.01	2.00	2.00	2.00	2.00	1.16	2.00	2.00	2.00	2.00	2.00
P 9a	1.99	1.98	2.00	2.00	2.00	2.01	1.99	2.92	2.00	2.00	2.00	2.00	2.00	2.00
P 9b	1.96	1.97	1.95	2.01	2.00	2.00	1.99	2.92	1.99	2.00	2.01	1.99	1.99	2.00
P 10a	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.40	2.00	1.64	2.00	1.60	1.60	2.39
P 10b	2.00	2.00	2.18	2.00	2.00	2.00	2.00	2.26	2.00	1.76	2.00	1.74	1.74	2.24
P 11a	2.00	2.00	2.00	2.00	2.77	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.02	2.06
P 11b	2.00	2.00	2.00	2.01	2.84	2.00	2.00	2.01	2.00	2.00	2.00	2.00	2.01	2.01
P 12a	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
P 12b	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.99	2.00	2.00	2.00	2.00	2.00
P 13a	2.00	2.00	2.00	2.01	2.00	2.00	2.00	2.90	2.00	2.00	2.00	2.00	2.00	2.86
P 13b	2.00	1.99	1.99	2.01	2.00	2.00	2.00	2.87	2.01	2.00	2.00	2.00	2.00	2.85
P 14a	2.00	2.01	2.00	2.01	2.00	2.00	2.00	2.00	2.00	2.00	2.01	2.00	2.00	2.00
P 14b	2.00	1.99	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.99	2.01	1.97	2.00	2.00

Matched pairs: primary adenoma (a) and corresponding recurrent adenoma (b)

†, excluded from analysis; *green*, CN gain detected in more than 10% of the total number of nuclei; *red*, CN loss detected in more than 10% of the total number of nuclei.

Supplemental Table 6. Correlation matrix of ASNs and clinicopathological features. Observed average signal numbers of the gene markers were correlated with the clinicopathological data using Spearman rank correlation coefficient.

	COX2	PIK3CA	APC	CLIC1	EGFR	MYC	CCND1	CDX2	CDH1	TP53	HER2	SMAD7	SMAD4	ZNF217
COX2		0.000	0.010	0.022	0.367	0.001	0.000	0.818	0.009	0.007	0.002	0.026	0.064	0.046
PIK3CA	0.000		0.015	0.005	0.124	0.013	0.001	0.285	0.001	0.070	0.008	0.040	0.165	0.433
APC	0.010	0.015		0.137	0.004	0.003	0.000	0.198	0.000	0.397	0.001	0.000	0.000	0.009
CLIC1	0.022	0.005	0.137		0.117	0.214	0.000	0.169	0.043	0.360	0.009	0.053	0.552	0.043
EGFR	0.367	0.124	0.004	0.117		0.127	0.036	0.173	0.017	0.352	0.169	0.007	0.021	0.007
MYC	0.001	0.013	0.003	0.214	0.127		0.022	0.176	0.000	0.056	0.002	0.001	0.002	0.038
CCND1	0.000	0.001	0.000	0.000	0.036	0.022		0.298	0.001	0.122	0.000	0.002	0.022	0.024
CDX2	0.818	0.285	0.198	0.169	0.173	0.176	0.298		0.163	0.743	0.020	0.341	0.870	0.005
CDH1	0.009	0.001	0.000	0.043	0.017	0.000	0.001	0.163		0.006	0.034	0.002	0.018	0.013
TP53	0.007	0.070	0.397	0.360	0.352	0.056	0.122	0.743	0.006		0.984	0.093	0.218	0.759
HER2	0.002	0.008	0.001	0.009	0.169	0.002	0.000	0.020	0.034	0.984		0.004	0.008	0.195
SMAD7	0.026	0.040	0.000	0.053	0.007	0.001	0.002	0.341	0.002	0.093	0.004		0.000	0.003
SMAD4	0.064	0.165	0.000	0.552	0.021	0.002	0.022	0.870	0.018	0.218	0.008	0.000		0.061
ZNF217	0.046	0.433	0.009	0.043	0.007	0.038	0.024	0.005	0.013	0.759	0.195	0.003	0.061	
Age at diagnosis	0.775	0.898	0.588	0.465	0.572	0.209	0.685	0.107	0.202	0.425	0.703	0.946	0.341	0.113
Age (dichot. median=67y)	0.426	0.989	0.581	0.793	0.582	0.170	0.339	0.491	0.103	0.189	0.677	0.845	0.375	0.294
CN change (yes vs no)	0.809	0.563	0.142	0.005	0.001	0.209	0.102	0.021	0.005	0.680	0.850	0.013	0.149	0.000
Gender (F vs M)	0.171	0.693	0.385	0.919	0.719	0.599	0.524	0.323	0.384	0.308	0.459	0.097	0.091	0.337
Grading (LGD vs HGD)	0.284	0.122	0.790	0.008	0.215	0.501	0.445	0.354	0.375	0.631	0.155	0.576	0.733	0.424
Heatmap Cluster group (1-4)	0.259	0.698	0.437	0.256	0.079	0.177	0.878	0.018	0.283	0.150	0.473	0.319	0.843	0.004
Histology (tubular vs tub-vil)	0.665	0.973	0.773	0.075	0.868	0.475	0.084	0.209	0.875	0.520	0.932	0.539	0.120	0.438
Instability Index	0.036	0.005	0.041	0.000	0.003	0.001	0.038	0.004	0.018	0.765	0.000	0.008	0.102	0.000
Instability Index (dichot. median=2.92 pattern)	0.156	0.013	0.136	0.000	0.001	0.067	0.083	0.009	0.016	0.760	0.011	0.028	0.326	0.000
Localization (Cecum, Ascendens, Transverse, Descendens, Sigmoid, Rectum)	0.249	0.576	0.682	0.019	0.922	0.900	0.002	0.079	0.915	0.908	0.355	0.980	0.734	0.296
Localization (Right vs Left)	0.821	0.925	0.876	0.071	0.555	0.292	0.029	0.046	0.320	0.298	0.387	0.522	0.444	0.542
Localization (Right vs Left vs Rectum)	0.654	0.941	0.788	0.019	0.549	0.470	0.017	0.044	0.391	0.463	0.346	1.000	0.849	0.275
Localization (Colon vs Rectum)	0.468	0.726	0.687	0.007	0.644	0.973	0.035	0.127	0.685	0.988	0.399	0.257	0.384	0.093
Number of clone patterns	0.028	0.004	0.050	0.000	0.002	0.011	0.041	0.004	0.058	0.571	0.000	0.069	0.303	0.001
Recurrence (time in months)	0.656	0.408	0.708	0.549	0.600	0.126	0.563	0.116	0.006	0.277	0.038	0.696	0.743	0.956
Recurrence (yes vs no)	0.652	0.033	0.822	0.587	0.070	0.244	0.460	0.790	0.132	0.302	0.084	0.071	0.025	0.823

7.1 Lists of figures, tables and formulas

List of figures

List of tables

List of formulas

8   CURRICULUM VITAE

Personal details

Name and surname:	David Fiedler
Date of birth:	August 9th, 1985
Place of birth:	Berlin-Friedrichshain (Germany)
Marital status:	Single
Father:	Thorsten Andreas Fiedler
Mother:	Susanne Moesner (née Hertel)

Education

1992 – 1995	Elementary school: Grundschule am Brandenburger Tor, Berlin-Mitte (Germany)
1995 – 1998	Elementary school: Otfried-Preußler-Grundschule, Großbeeren (BRB, Germany)
1998 – 2005	Comprehensive secondary school: Peter-Joerres-Gymnasium, Ahrweiler (RLP, Germany)
Mar. 12th, 2005	Abitur (ø 1.6)

Academic education

Fall semester 2005/2006	Start of studies in *Biology* at the Rheinisch-Westfälisch Technische Hochschule (RWTH) Aachen (NRW, Germany)
Sep. 24th, 2007	Graduation of the basic study period: intermediate diploma (mark: *good*)
2007 – 2011	Advanced study period
Apr. – Oct. 2010	Research internship at the *Cell and Molecular Biology Group*, NTNU Trondheim (Norway) ("Allyl iso-thiocyanate affects beta-tubulin in *A.thaliana*")
2011	Diploma thesis at the *Institute of Pathology*, University Hospital Aachen ("Characterization of the DNA methylation of *ECRG4* and *EDN3* as biomarkers in human bladder cancer")
Sep. 19th, 2011	Diploma (mark: *excellent*)
2012 -	Start of doctorate studies at the *Institute of Pathology*, University Hospital Mannheim, Heidelberg University (Germany)

9   PROOF OF PUBLICATION

D. Fiedler, K. Heselmeyer-Haddad, D. Hirsch, L. Hernandez, I. Torres, D. Wangsa, Y. Hu, L. Zapata, J. Rüschoff, S. Belle, T. Ried, and T. Gaiser
"Single-cell genetic analysis of clonal dynamics in colorectal adenomas indicates CDX2 gain as a predictor of recurrence"
Publication; *Epub:* 19.09.2018;
Volume 144, Issue 7, pp.1561-1573, 01.04.2019; DOI: 10.1002/ijc.31869
International Journal of Cancer, Germany (IF-2017: 7.36)

D. Fiedler, K. Heselmeyer-Haddad, D. Hirsch, L. Hernandez, I. Torres, S. Belle, T. Ried, and T. Gaiser
"Genomische Instabilität in rezidivierenden und nicht-rezidivierenden Kolonadenomen"
101. Annual Meeting of the German Society of Pathology (Deutsche Gesellschaft für Pathologie, DGP)
Oral Presentation; June 22nd – 24th, 2017
Heinrich Lades Halle, Erlangen (BY), Germany

D. Fiedler, K. Heselmeyer-Haddad, D. Hirsch, L. Hernandez, I. Torres, S. Belle, T. Ried, and T. Gaiser
"Genomic Instability in Colorectal Adenomas with and without Recurrence"
Symposium on Chromosomal Aberrations in Cancer
Oral Presentation; February 13th, 2017
Institute of Pathology, University Hospital Mannheim of Heidelberg University, Mannheim (BW), Germany

D. Fiedler, K. Heselmeyer-Haddad, D. Hirsch, L. Hernandez, I. Torres, S. Belle, T. Ried, and T. Gaiser
"Genomic Instability in Colorectal Adenomas with and without Recurrence"
International Workshop on Translating Colorectal Cancer Research
Oral Presentation; February 9th-10th, 2017
Instituto Português de Oncologia, Porto, Portugal

D. Fiedler, K. Heselmeyer-Haddad, N. El Hajj, H. Yang, S. Belle, T. Haaf, A. Marx, T. Ried, T. Gaiser
"Genomic Instability in Recurring and Nonrecurring Adenomas of the Colon"
16th Annual NCI CCR-Fellows and Young Investigators Colloquium
Oral Presentation; March 31st – April 1st, 2016
National Cancer Institute, Shady Grove Campus, Bethesda (MD), USA

D. Fiedler, K. Heselmeyer-Haddad, N. El Hajj, H. Yang, S. Belle, T. Haaf, A. Marx, T. Ried, T. Gaiser
"Genomic Instability in Recurring and Nonrecurring Adenomas of the Colon"
Fourth Symposium on Translational Genomics with a Special Focus on Liver Cancer
Poster Presentation; March 17th -18th, 2016
National Institutes of Health / National Cancer Institute, Bethesda (MD), USA

10 ACKNOWLEDGEMENTS