ANALYSES OF REGULATORY MECHANISMS AND IDENTIFICATION OF MICRORNAS AS BIOMARKERS IN COLORECTAL CANCER

INAUGURAL-DISSERTATION

TO OBTAIN THE ACADEMIC DEGREE

'*DOCTOR RERUM NATURALIUM*'

(DR. RER. NAT.)

SUBMITTED TO THE DEPARTMENT OF BIOLOGY, CHEMISTRY AND PHARMACY

OF FREIE UNIVERSITÄT BERLIN

BY

CHRISTINA RÖHR

FROM BERLIN

2012

Diese Arbeit wurde unter der Leitung von Frau Dr. Dr. Michal-Ruth Schweiger in der Zeit von 2008 bis 2012 am Max-Planck-Institut für Molekulare Genetik in Berlin erstellt.

1. Gutachter: Prof. Dr. Hans Lehrach, Max-Planck-Institut für Molekulare Genetik

2. Gutachter: Prof. Dr. Rupert Mutzel, Freie Universität Berlin

Disputation am: 12.10.2013

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Prof. Dr. Hans Lehrach for giving me the opportunity to accomplish this work at the Max Planck Institute for Molecular Genetics. In addition, I would especially like to thank Dr. Dr. Michal-Ruth Schweiger for giving me the opportunity to work on this interesting PhD project and for her guidance and continuous support throughout the last years. Her expertise added considerably to my graduate experience.

I am also thankful to the thesis committee of the Freie Universität Berlin, in particular Prof. Dr. Rupert Mutzel for taking the time to be my second supervisor.

During this work I collaborated with many external and internal colleagues and I want to thank all of them for their support, especially the Institute of Pathology of the Medical University of Graz, the Next Generation Sequencing Group and the Otto Warburg Laboratories at the Max Planck Institute for Molecular Genetics, the Department of Gastroenterology at the Charité, the German Center for Neurodegenerative Diseases and the Institute for Medical Biometry, Informatics and Epidemiology at the University of Bonn.

I would like to thank the current and former members of the 'Cancer Genomics Group', Stefan Börno, Andrea Wunderlich, Melanie Isau, Martin Kerick, Axel Fischer, Michelle Hussong, Nada Kumer and Anna Kosiura, for their scientific assistance and for creating such an enjoyable atmosphere in the lab.

Moreover, I would particularly like to thank Christoph Wierling, Martin Kerick, Axel Fischer and Thomas Meinel for their bioinformatical work and support.

I am thankful to all the other colleagues from the MPI who helped me with my work in different ways and who provided me with technical assistance, knowledge and reagents from time to time.

I would like to thank all my fantastic friends for encouraging and supporting me along this journey. Thank you for being my friends beyond thesis matters.

Last but not least, I wish to thank my family, especially my parents for the unconditional support they provided me through my entire life and their faith in me. To them I dedicate this thesis.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A	Adenine
ATP	adenosine triphosphate
BSA	Bovine serum albumin
C	Cytosine
cDNA	Complementary deoxyribonucleic acid
C. elegans	Caenorhabditis elegans
dATP	Deoxyadenosine triphosphate
ddH_2O	Double distilled water
DIG	Digoxigenin
DMEM	Dulbeco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
E	Efficiency
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced green fluorescent protein
EMT	Epithelial-to-mesenchymal transition
FACS	Fluorescence-activated cell sorting
FCS	Fetal calf serum
FFPE	Formaldehyde fixed-paraffin embedded
G	Guanine
gDNA	Genomic deoxyribonucleic acid
GDP	Guanosine diphosphate
GFP	Green fluorescent protein
GTP	Guanosine triphosphate
IHC	Immunohistochemistry
ISH	*In-situ* hybridization
LNA	Locked nucleic acid
MeDIP	Methylated DNA immunoprecipitation
MET	Mesenchymal-to-epithelial transition
miR	MicroRNA
mRNA	Messenger ribonucleic acid
MSI	Microsatellite instability
MSS	Microsatellite stability
NGS	Next generation sequencing
PAGE	Polyacrylamide gel electrophoresis
PCA	Principal component analyses

PBS	Phosphate buffered saline
PBST	0.1% Tween 20 in phosphate buffered saline
PCR	Polymerase chain reaction
Pen/Strep	Penicillin/streptomycin
PFA	Paraformaldehyde
Pol II/ Pol III	Polymerase II or III
qRT-PCR	Quantitative real-time polymerase chain reaction
RISC	RNA induced silencing complex
RNA	Ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
SDS	Sodium dodecyl sulfate
snRNA	Small nuclear RNA
SNP	Single-nucleotide polymorphisms
SNV	Single nucleotide variations
SOM	Self-Organizing Map
SSC	Solution of sodium chloride and sodium citrate
T	Thymine
TAE	Tris-acetate-ethylenediaminetetraacetic acid
TBE	Buffer Tris-borate-EDTA buffer
TNM	Tumor-node-metastasis
Tris	Tris(hydroxymethyl)aminomethane
Tween20	Polyoxyethylensorbitanmonolaurat
U	Uracil
UTR	Untranslated region
UV	Ultraviolet

bp	Base pairs
kb	Kilo bases
nt	Nucleotides
°C	Degrees Celsius
Ct	Threshold cycle
g	Gram
mg	Milligram
µg	Microgram
ng	Nanogram
l	Liter
ml	Milliliter
µl	Microliter
mol	Mole

pmol	Picomole
M	Molar concentration, 1 mol per liter
mM	Millimolar
µM	Micromolar
nM	Nanomolar
s	Second
min	Minute
h	Hour
µm	Micrometer
mm	Millimeter
U	Enzyme unit
xg	Times gravity

ABSTRACT

The investigation of regulatory mechanisms that form the basis of colorectal cancer and the identification of molecular biomarkers were the main goals of this work. Therefore, we performed a genome-scale analysis of the microRNA expression patterns of eight colon cancer patients using Illumina next generation sequencing technologies each with matching normal, tumor and metastasis tissues. We validated different expression patterns using the TaqMan assay technology. In addition, we analyzed the microRNA expression in colon cancer cell lines SW480 and their corresponding metastasis cell line SW620. On the basis of this genome-scale microRNA profiling we selected and validated different microRNA candidates as potential molecular biomarkers. These candidates were analyzed in a large screen containing 330 samples from 16 different tumor entities. We identified *miR-1* as a top candidate differentially expressed in tumor and metastasis of colorectal cancers, but also in 16 additional tumor entities. For this microRNA we performed functional analyses, including cell viability assays and wound healing experiments and we were able to show a tumor suppressive activity of *miR-1*. These results provided the basis for further investigations where we identified interesting target genes of *miR-1* and analyzed them by GFP reporter assay. We also addressed the question of camptothecin treatment in the presence of a deregulated *miR-1* and observed a synergistic effect with camptothecin in colorectal cancer cells. Since microRNAs are discussed as diagnostic and therapeutic molecules, we tried to identify potential therapeutic effects of microRNA drugs in colorectal cancer and used a systems-biology simulation software (PyBioS) of cellular cancer models. We modeled target regulations based on the results of transcriptome sequencings and simulated the impact of a *miR-1* deregulation. In particular we evaluated the consequences of a *miR-1* reduction and overexpression as a potential treatment option. Within further studies we identified not yet annotated smallRNAs and investigated the mutational landscape of colorectal cancers using whole exome re-sequencing approaches. Here we uncovered significant differences between the mutation patterns of microsatellite stable (MSS) and instable (MSI) tumors. This might have an additional effect, besides the gene expression patterns, on the therapy response of these tumors.

1 INTRODUCTION

1.1 Colon CANCER

Cancer is a leading cause of death worldwide, accounting for 7.6 million deaths (around 13% of all deaths) in 2008. According to the World Health Organization (WHO) in 2008 colorectal cancer (CRC) is the third most prevalent cancer and a major cause of cancer mortality worldwide with an incidence of approximately 1.2 million cases out of 12.6 Mio general cancer cases (Table **1.1**, Figure **1.1**) (www.who.int/en).

Table 1.1: List of most frequent cancers in both sexes according to the International Agency for Research on Cancer (IARC), part of the World Health Organization (WHO) in 2008.

Cancer	***Incidence***	***Mortality***
Lung	1,608,055	1,376,579
Breast	1,384,155	458,503
Colorectum	1,235,108	609,051
Stomach	988,602	737,419
Prostate	899,102	258,133
Liver	749,744	695,726
Cervix uteri	530,232	275,008
Oesophagus	481,645	406,533
Bladder	382,660	150,282
Non-Hodgkin lymphoma	356,431	191,599
Leukaemia	350,434	257,161
Corpus uteri	288,387	73,854
Pancreas	278,684	266,669
Kidney	273,518	116,368
Lip, oral cavity	263,020	127,654
Brain, nervous system	237,913	174,880
Ovary	224,747	140,163
Thyroid	213,179	35,383
Melanoma of skin	199,627	46,372
Larynx	150,677	81,892
Gallbladder	145,203	109,587
Other pharynx	136,622	95,550
Multiple myeloma	102,826	72,453
Nasopharynx	84,441	51,609
Hodgkin lymphoma	67,919	29,902
Testis	52,322	9,874
All cancers excl. non-melanoma skin cancer	**12,662,554**	**7,564,802**

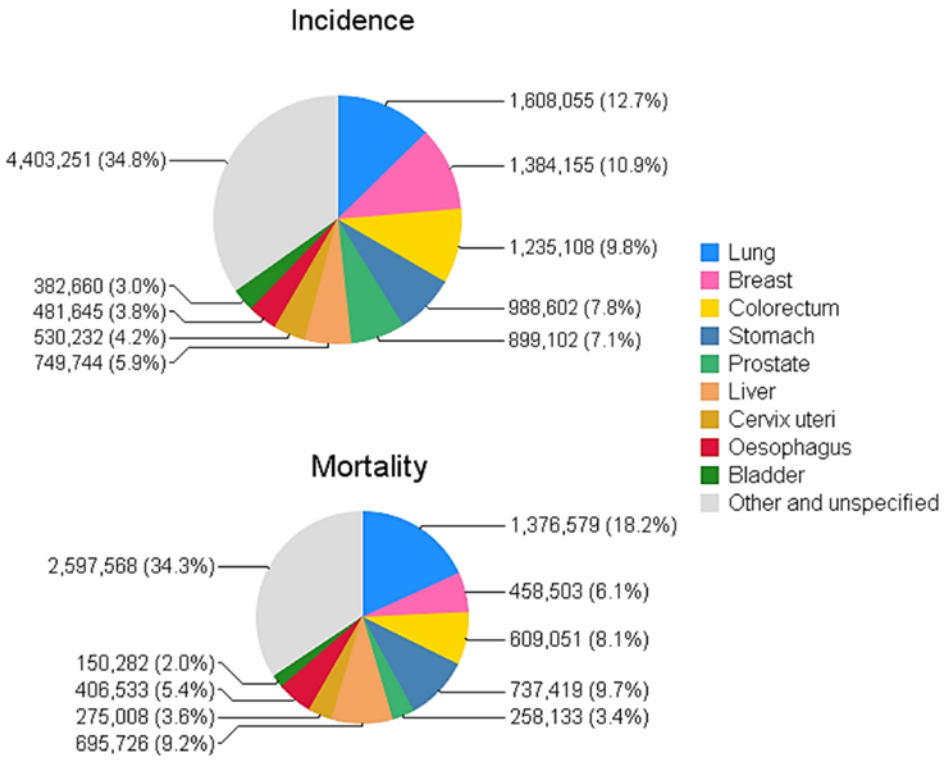

Figure 1.1: Graphics for cancer incidence and mortality in most frequent cancers in both sexes according to the International Agency for Research on Cancer (IARC), part of the World Health Organization (WHO) in 2008.

Cancer is a broad term for a large group of diseases. A central feature of all cancers is the rapid formation of abnormal cells that grow beyond their usual boundaries, and to spread and invade into distant parts of the body for forming metastases in other organs (Figure **1.2**). Metastasis is the major cause of death from cancer (www.who.int/en).

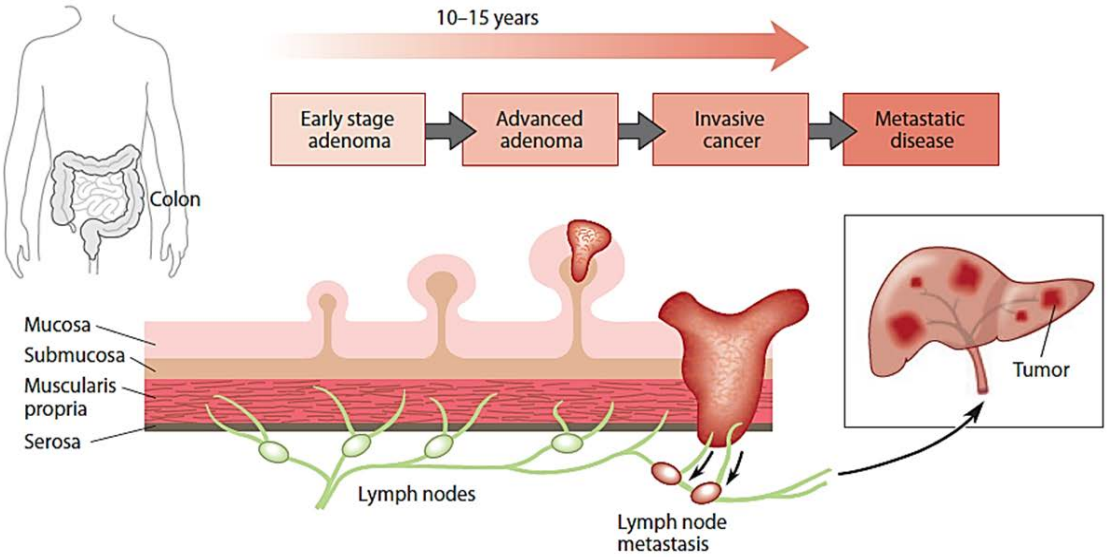

Figure 1.2: Cascade for colorectal cancer progression and formation of distal metastases (Bretthauer et al. 2011 [1]).

An early detection and treatment of cancer can reduce the mortality. At early stages of CRC a curative treatment is achieved by surgical resection or by chemotherapy, which is often applied in an adjuvant or neo adjuvant manner [2] (American Cancer Society, 2010). If the tumor exhibits features of a more advanced cancer, characterized by metastasizes that spread into distant sites it is usually incurable [3]. Therefore the detection of early symptoms, an early diagnosis and immediate treatment are relevant to lower the mortality of cancer.

There are some lifestyle risk factors which appear to promote the progression of CRC, such as a diet rich in unsaturated fatty acids and red meat, total energy consumption, excessive alcohol abuse and reduced physical activity [4,5]. In contrast, non-steroidal anti-inflammatory drugs, estrogen and calcium show some protecting effects against CRC [6,7]. Beside lifestyle and environmental risk factors, genetic aberrations can predispose to CRC. In the present paradigm inherited defects as well as somatic alterations promote the pathogenesis of colorectal cancer.

On a molecular level colorectal cancer is a heterogeneous disease that is described by changes in different pathogenic pathways, such as chromosomal instability (CIN), CpG island methylation phenotype (CIMP) and microsatellite instability (MSI) [8] (Figure **1.3**). One early theory by Vogelstein and Fearon which is still valid today states that it is the accumulation of genetic changes which contributes to the progression of CRC [9].

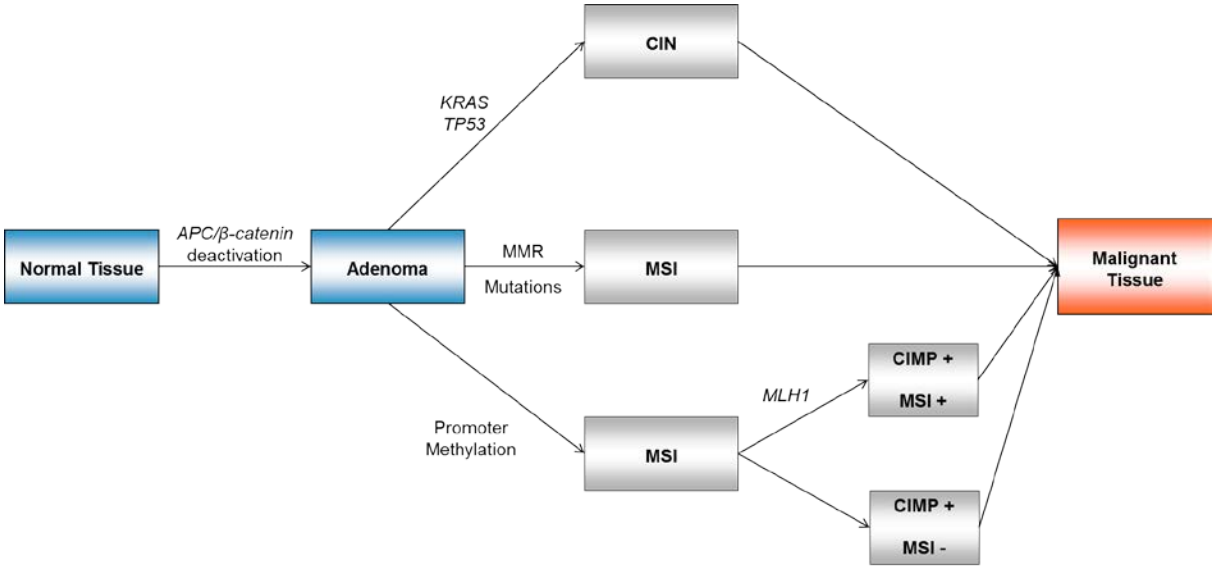

Figure 1.3: Schematic for main molecular pathways identified for malignant transformation of colon tissue. The pathways are described as the Chromosomal Instability Pathway (CIN), CpG Island Methylator Phenotype (CIMP) pathway and Microsatellite Instability pathway (MSI), (MMR = mismatch repair).

Chromosome instability (CIN) is the most common type of genomic imbalance to occur in colorectal cancers. It is present in almost 85% of all cases and could lead to the physical loss of a functional copy of one or many tumor-suppressor genes [9,10] (Figure **1.3**). A relatively limited number of

oncogenes and tumor-suppressor genes, most prominently the *APC* (adenomatous polyposis coli), *KRAS* (v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog) and *TP53* (tumor protein p53) gene are affected by critical alterations that could contribute to CRC progression. A deletion on chromosome 5, 18q or 17p is associated to a mutation of the *APC* gene. Especially the long arm of chromosome 18 contains genes that are often found deleted in CRC, for example *SMAD2* (SMAD family member 2) and *SMAD4* (SMAD family member 4) which play an important role in TGF-β signaling pathway [11,12]. Germ-line mutations in the tumor-suppressor gene *APC* are known to promote an autosomal dominant disorder known as familial adenomatous polyposis coli (FAP) which accounts for ~0,5% of all CRC cases [13,14]. The incidence of untreated FAP patients to develop CRC during lifetime is 100% [15]. In addition to the important role of *APC* mutations in the heredity of CRC it is even more prominent for the progression of sporadic colorectal adenomas and carcinomas, which account for approximately 70% of all clinical cases. The pathogenesis is associated to somatic mutations in the *APC* gene, which is characterized by a variety of functions[16-19]. The normal APC protein regulates cell-cell adhesion, cell migration, chromosomal segregation and apoptosis in the colonic crypt [18,20,21]. It interacts with different proteins, such as β-catenin, glycogen synthase kinase (GSK) 3β and Bub kinases. When APC binds to β-catenin it promotes the cytosolic degradation of β-catenin in a tandem with AXIN (Axis inhibition protein), casein kinase I (CKI) and glycogen synthase kinase [22]. A mutational inactivation of *APC* increases the level of β-catenin and leads to the activation of the Wnt signaling pathway, where the oncoprotein β-catenin binds to nuclear partners to act as a transcription factor for genes that are involved in activating cellular processes. Wnt signaling is an initiating event for CRC progression [9,23] (Figure **1.4**).

Another key feature for the development of colorectal cancer is the inactivation of the TP53 pathway by somatic mutations. In CRC, this inactivation contributes to continued growth and the acquisition of invasive properties under certain cellular stress conditions and is a key event for the transformation of large adenomas into invasive carcinomas [9,24,25]. The Ras family members, such as *KRAS*, *HRAS* (v-Ha-ras Harvey rat sarcoma viral oncogene homolog) and *NRAS* (neuroblastoma RAS viral (v-ras) oncogene homolog) are common targets for somatic mutations in many human cancers. Mutations in the proto-oncogene *KRAS* are found in almost 50% of all colorectal cancer cases [26]. KRAS is a cellular membrane protein of the inner surface, which plays a major role in the transduction of intercellular signals by acting as a molecular switch. The GTP bound form of KRAS activates downstream signaling targets, while the GDP bound form does not. Mutations in the GTPase domain of KRAS lead to a decreased GTPase activity and a constitutive activation of KRAS which produces uncontrolled proliferation and the progression of polyps [27].

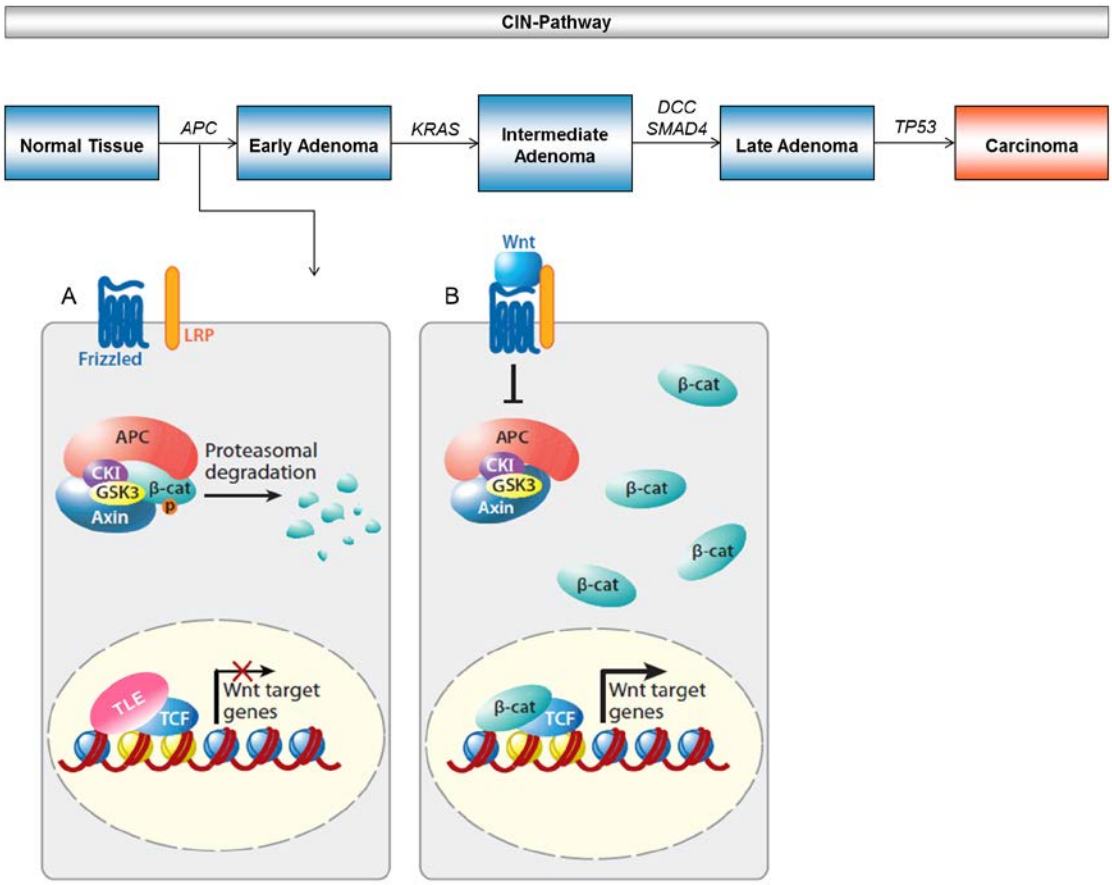

Figure 1.4: Schematic representation of CIN pathway in the progression of colon carcinoma and model of adenomatous polyposis coli (APC), β-catenin function and Wnt signaling pathway. (**A**) The APC protein forms along with other proteins, such as glycogen synthase kinase 3β (GSK3β), AXIN and casein kinase I (CKI) a destruction complex that induces the proteasomal degradation of β-catenin. The transcription of T cell factor (TCF)-regulated target genes is then repressed by an interaction of TCFs with repressor proteins of the transducin-like enhancer of split (TLE) family. (**B**) Inactivating mutations of the APC protein are similar to the action of Wnt ligands which bind to Frizzled receptors and low-density lipoprotein receptor–related protein (LRP). This can lead to the accumulation of free β-catenin and an increased free β-catenin level in the cell. The β-catenin can enter the nucleus where it activates transcription of TCF-regulated target genes (modified from van der Flier et al., 2009 and Fearon et al., 2011).

Epigenetic instability of the genome is another molecular subtype of CRC. Here an epigenetic silencing of genes is frequently mediated by an altered DNA methylation. The mechanism most often occurs at CpG site in the genome by converting a cytosine to 5-methylcytosine. The CpG island methylator phenotype (CIMP) is present in almost 15% of all sporadic colorectal cancers [28,29] (Figure **1.3**). In a large subset of CRCs the tumor suppressor gene *HIC1* (hypermethylated in cancer 1) is transcriptionally inactivated by hypermethylation. This is also known for the *SFRPs* (secreted Frizzled-related proteins), which function as Wnt signaling antagonists [30]. In some CRCs hypermethylation is observed in various genes, which leads to the assumption, that an aberrant DNA methylation is globally present in this subset of CRCs. The CIMP pathway can also affect promoter methylation and subsequent expression of the *MLH1* (mutL homolog 1, colon cancer, nonpolyposis type 2) gene, which

results in sporadic CRCs with microsatellite instability (MSI) [31]. MSI is a genomic instability where aberrations occur in short repeated nucleotide sequences (microsatellites) throughout the genome. The discrepancy of local repeat counts between tumor and normal DNA are known as microsatellite instability [32]. The MSI pathway is strongly associated to the mismatch repair (MMR) system, which recognizes errors made by the DNA polymerase during replication processes and eliminates them [33]. Mutations that occur in the MMR system can contribute to the progression of MSI in cancers [34] (Figure **1.3**). These mutations can impair one or more of the seven genes associated to the MMR system, such as *MLH1*, *MLH3* (mutL homolog 3), *MSH2* (mutS homolog 2, colon cancer, nonpolyposis type 1), *MSH3* (mutS homolog 3), *MSH6* (mutS homolog 6), *PMS1* (postmeiotic segregation increased 1) and *PMS2* (postmeiotic segregation increased 2) [35]. Germline mutations in the MMR system are highly frequent in Lynch Syndrome or Hereditary Non Polyposis Colorectal Cancer (HNPCC) and often lead to the development of MSI in those subgroups of CRC. Alternatively, the MMR system can be affected by promoter hypermethylation, which results in the progression of sporadic MSI tumors [36,37]. MSI positive CRC can be subdivided into MSI-high and MSI-low tumors, whereas the degree of instability is the decisive factor for the separation [38,39]. Using whole exome next generation sequencing technology, Timmermann et al. [40] profiled MSI and MSS colorectal cancer patients and identified eight times more somatic variations in MSI than in MSS tumors, which underlines the potential of this approach to classify solid tumors and to predict clinically relevant mutations.

1.2 NEXT GENERATION SEQUENCING TECHNOLOGY AND TARGET ENRICHMENT

The ‘first generation’ sequencing technology known as Sanger sequencing was published by Maxam and Sanger in 1977 [41,42]. From 1990 on the automated sanger sequencing system based on capillary electrophoresis combined with DNA fluorescent detection was used and led to the completion of the human genome project (HGP) in 2001 [43,44,45]. Sanger sequencing attained a capacity of 1-1.2kb length per read with a high specificity but very low throughput. With the aim of high-throughput sequencing of larger DNA fragments a new approach was developed during the HGP. Shotgun sequencing technologies were established and paved the way for ‘next generation sequencing’ (NGS) [46-48]. This second generation of sequencing technology is able to produce a massively parallel throughput. The sequencing output is generated from pools of fragmented DNA called ‘libraries’ which provide reads of 50-500 continuous basepairs. This short read length is a limiting factor for the NGS technology and will be overcome soon by the third generation of sequencing technologies like nanopore sequencing (Oxford nanopore technologies).

Currently five platforms of NGS technology are commercially available: the genome sequencer GS-FLX Titanium (Roche), the Illumina HiSeq 2000, the Applied Biosystems SOLID 5500xl, the Polonator G.007 and Helicos HeliScope, with Roche, Illumina and ABI dominating the market. The methodology used by the available NGS systems share common features such as library preparation, the sequencing and imaging process and bioinformatic data analysis, where digital information about DNA sequences are assembled and compared to reference genomes. For the Illumina platform the

sequencing-by-synthesis approach is used to produce sequencing reads (Figure **1.5**). DNA library fragments are immobilized in clusters on the surface of a glass flow cell for simultaneous bridge amplification. During this process polymerase and fluorescent labeled nucleotides that have their 3'-OH chemically modified by a reversible terminator are supplied to each fragment of a cluster. This ensures that only a single base is incorporated per cycle during the sequencing procedure [49] (www.illumina.com).

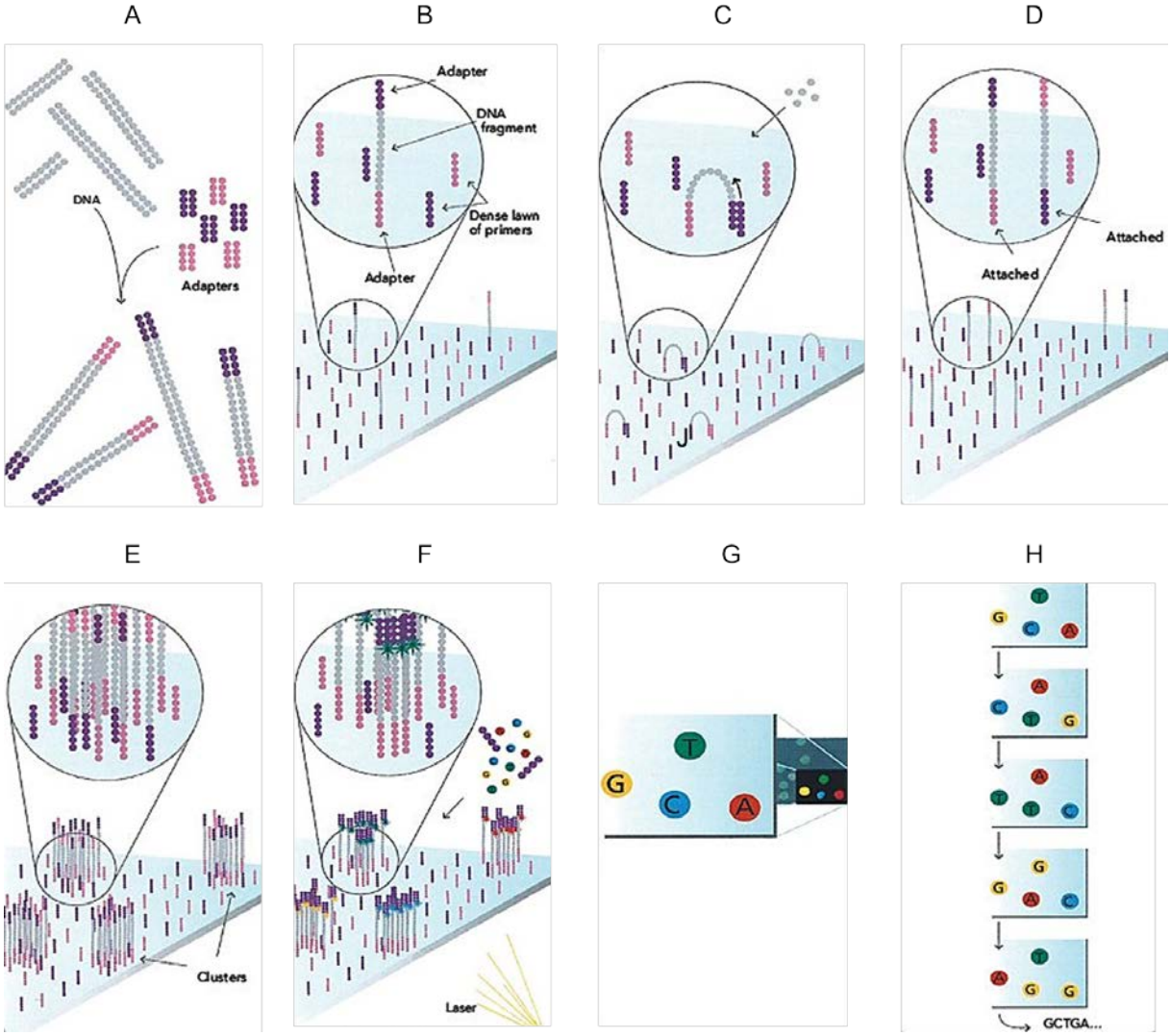

Figure 1.5: Illumina sequencing-by synthesis workflow. (**A**) Preparation of genomic DNA sample beginning from DNA fragmentation and adapter ligation. (**B**) Binding of the single-stranded DNA library to the surface of the flow cell. (**C**) Bridge amplification to generate double-stranded DNA-fragments. (**D**) Denaturation of double-stranded fragments to leave single-stranded fragments on the flow cell surface. (**E**) Completion of amplification. Each fragment is amplified into clusters. (**F**) Sequencing steps: starting with the first chemistry cycle to determine the first base. All four labeled reversible terminators, primers and DNA polymerase enzyme are added to the flow cell. (**G**) Imaging. After laser excitation the image of emitted fluorescence is captured from each cluster. (**H**) Determination of sequencing reads over multiple chemistry cycles. Cycles of sequencing are repeated to determine the sequence of a DNA fragment a single base at a time (modified from Illumina).

The technology of next generation sequencing is applied to a variety of genetic questions. Its main uses are whole genome sequencing and target re-sequencing, for large scale analysis of DNA methylation, for mapping of genetic rearrangements, for chromatin immunoprecipitation-sequencing and for transcriptome and small RNA analysis. One promising application of NGS may be the target re-sequencing of human genomes for personalized medicine.

Next generation sequencing platforms are constantly evolving. The 'third-generation-sequencing' technologies are based on ultra-fast DNA sequencing in real time without previous amplification steps, which provides more accurate sequencing data with potentially longer reads. Diverse approaches of those most recent DNA sequencing technologies have been build up, such as fluorescence-based single-molecule sequencing, nano-technologies for single-molecule sequencing and electronic detection for single-molecule sequencing. Several companies are working on those technologies as for example Pacific BioSciences, which build a system of single-molecule real time (SMRT) DNA sequencing technology and Visigen Biotechnology, which developed a similar approach using internal reflectance fluorescence (TIRF) technology to measure time-dependent fluorescence signals of single DNA fragments in parallel. Both systems are based on optical detection. Nanopore sequencing is based on monitoring the movement of a single DNA strand through a nano-tunnel and different concepts of data readout are under development like electrical current fingerprints which differ for each nucleotide base. The company Oxford Nanopore Technologies developed an exonuclease-coupled nanopore sequencing approach for the sequential identification of DNA bases as the processing enzyme passes through the nanopore. The concept of Pyro-sequencing was advanced and transformed by Ion Torrent Technology. They used an ion sensor to detect hydrogen ions that are released when a nucleotide is incorporated into a DNA strand during synthesis. The chemical information is then transcribed into digital sequence information [50-53]. The development of less expensive and time saving concepts and technologies with a higher accuracy of the sequencing results is the main concern of every player in the sequencing market.

1.2.1 TARGET ENRICHMENT TECHNOLOGIES

Different target enrichment approaches were developed during the past years, each with unique parameters that affect sensitivity, specificity, reproducibility, costs and the amount of DNA required for an experiment. The key feature of each strategy is to lower the amount of sequencing needed. Genomic subsets that are highly relevant for downstream analysis can be analyzed using NGS platforms. This strategy is used to examine specific genomic regions like exons of specific candidate genes that are associated to disease–related pathways or phenotypes or the complete protein-coding DNA sequence. The concept of target enrichment is well established. Based on PCR technology Frazer et al. and RainDance Technologies developed a microdroplet PCR enrichment strategy in 2009. Droplets containing forward-and reverse specific targeting primer and droplets with genomic DNA and the associated PCR reagent are dispensed into a channel of a microfluidic chip and paired together in a 1 to 1 ratio to perform separate PCR reactions in microdroplets in a single reaction tube.

This method is able to achieve 84% capture efficiency with 90% of the target bases showing a uniform coverage [54].

There are additional widely used strategies for target enrichment based on hybridization either to custom-designed oligonucleotide microarrays or to strepatavidin-coated magnetic beads in solution. In 2007 the microarray-based-genomic selection was developed by Roche NimbleGen and collaborators. They developed an oligonucleotide microarray which provides complementary sequences for solid phase hybridization to enrich for exons or specific genomic regions [55-57]. The basis for this hybrid capture method is an adapter modified genomic DNA library which is able to hybridize to the target specific probes. Specific DNA fragments are immobilized and the non-specific hybrids are removed by washing steps before the enriched DNA can be eluted. The capture efficiency varies for the Roche 454 platform and the Illumina/Solexa platform. The main disadvantage of an array-based enrichment is the large amount of starting material required for the performance of one capture experiment. Therefore an alternative method was established by Nussbaum et al in 2009 [58] where they generate biotinylated RNA sequences complementary to the genomic sequences of interest and captured those using strepatavidin-coated magnetic beads. Nussbaum et al achieved a capture efficiency of 60% for exons and 80% for selected genomic regions. Another multiplex exon capture method, established by Shendure and Church et al in 2007, is the multiplex exon capture approach based on the application of molecular inversion probes [55,56,58-60]. For the enrichment a library of 100mer targeting oligo precursors are custom-designed on a microarray. These probes consist of a universal spacer region flanked by the specific target sequence, designed for each amplicon which hybridize upstream and downstream of each specific sequence. The target region is enzymatically gap-filled and ligated, so that the sequence of interest is incorporated into a circle that can be linearly amplified by rolling circle amplification. Improvements of the protocol by Davis et al lead to a capture efficiency of 90% [61]. Target enrichment technologies are an effective way to reduce not only the complexity of the genome but also to reduce the sequencing time and costs and thereby enable experiments on an unprecedented scale like the detection of disease causing variants for a multitude of complex diseases.

1.3 MICRORNAS

MicroRNAs are evolutionarily conserved, small (20-25 nucleotides) non-protein-coding molecules that regulate gene expression at the post-transcriptional level. These single strand RNAs participate in the regulation of various crucial cellular processes such as cell differentiation, cell cycle progression, metabolism and apoptosis. Therefore, they influence major biological systems such as immunity and cancer [62].

1.3.1 THE DISCOVERY OF MICRORNAS

With the initial discovery of microRNAs in 1993, a new class of gene regulation was found. The group led by Ambros and Lee revealed that the gene *lin-4* is transcribed into a 22 nucleotide RNA molecule,

which inhibits the protein synthesis of *lin-14* by nearly perfect complementary binding to its 3' untranslated region (3'UTR). This process affects temporal progression of cell differentiation in the nematode *C. elegans* [63]. At present time (2012), more than 1600 microRNA sequences have been identified in the human genome (miRBase v19, 2012). It is predicted that approximately 3% of the human genome encodes for microRNAs and that ~30% of the protein-encoding genes are regulated by microRNAs [64,65].

1.3.2 MICRORNA BIOGENESIS AND MICRORNA-MEDIATED TRANSLATION CONTROL

MicroRNA maturation is a mechanism which can be divided into the nuclear processing by DROSHA (drosha, ribonuclease type III), and the cytoplasmatic processing by DICER (dicer 1, ribonuclease type III). The nuclear export is enabled by exportin-5 in a RAS-related, nuclear protein–guanosine triphosphate (GTP)-dependent manner. The transcription of the primary microRNA (pri-microRNA), which contains a local fold back structure, is mediated predominantly by RNA Polymerase II in the nucleus. Polymerase III was first believed to transcribe microRNA genes but pri-microRNAs comprise several kilobases and contain poly-uracil stretches that would terminate the activity of Pol III. The stem-loop structure of the primary transcript is recognized by the nuclear RNase III enzyme DROSHA and its partner DGCR8 (DiGeorge syndrome critical region gene 8; known as Pasha in *D. melanogaster*) [66]. The resulting hairpin structure, called pre-microRNA, is transported to the cytoplasm, where it undergoes further processing by *Dicer* enzyme and the Dicer cofactor HIV-1 TAR RNA-binding protein (TRBP) which have various roles in microRNA stability and effector complex formation, liberating a ~19 - 24 bp double stranded microRNA duplex structure that consists of the mature and the complementary mature star sequence. One of these sequences, generally the mature microRNA, incorporates into the miRISC complex (microRNA-containing RNA-induced silencing complex) which contains Argonaute (AGO) proteins, and binds preferentially to the 3'UTR of mRNA target molecules [67-69]. The RISC-bound microRNA binds with complementary or near-complementary sequence to their mRNA targets and induces cleavage or inhibits the translation of the target mRNA. The RISC-bound mRNAs can be localized to sub-cytoplasmatic compartments, known as P-bodies, where they are reversibly stored or degraded. MicroRNAs that bind to their mRNA targets with perfect complementarity induce mRNA target destabilization, but microRNAs that bind imperfect modulate target gene expression by inhibiting protein translation. A number of mechanisms are discussed by which microRNAs can inhibit translation, either by stopping the initiation or the elongation process of protein translation [70] (Figure **1.6**).

The regulatory potential of microRNAs is due to the fact that each microRNA targets multiple mRNAs and one mRNA can be regulated by different microRNAs. Through base pairing with microRNA-binding sites in the 3' or 5' UTR of mRNA targets they negatively regulate gene expression at the level of translation to control processes as diverse as cellular growth rates, proliferation, differentiation and cell death [71].

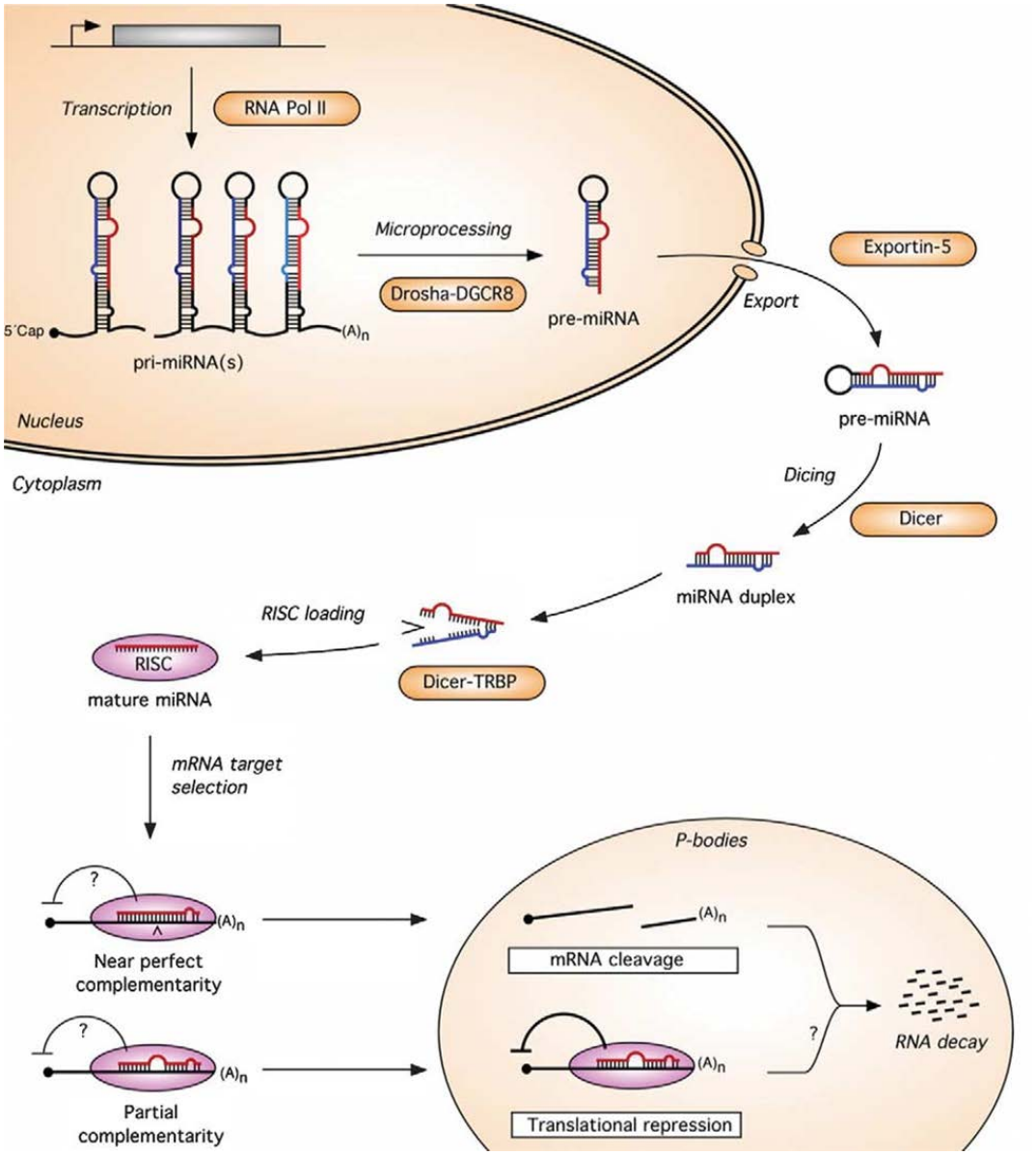

Figure 1.6: Model of microRNA biogenesis and function in animal cells. MicroRNA genes are transcribed by RNA polymerase II (Pol II) to generate a long primary microRNA (pri-microRNA). In the nucleus the pri-microRNA recognized and cleaved by the RNase III enzyme DROSHA in complex with DGCR8, resulting in an hairpin precursor form called pre-microRNA of ~70 nucleotides in length. The pre-microRNA is exported into the cytoplasm by exportin-5 and further processed by the RNase enzyme DICER in the cytoplasm where it incorporates as a ~21 nucleotide long mature microRNA into the RNA-induced silencing complex (RISC) and binds to specific mRNA targets (picture taken from www.yale.edu/giraldezlab).

1.3.3 MICRORNAS IN CANCER

Based on current research microRNAs are involved in a variety of disease processes, such as cancer. They play a critical role in human oncogenic signaling pathways including oncogenesis, progression, invasion, metastasis and angiogenesis [72]. For various cancers microRNA profiling data are available and the expression pattern of specific microRNAs can be correlated with cancer type, stage and other

clinical variables. Therefore, microRNAs provide the possibility for application in cancer diagnosis and prognosis, and they are discussed as a tool for cancer therapy. In 2002 Calin et al. [73] were the first who described the link between microRNAs and cancer. They demonstrated that in over 65% of B cell chronic lymphocytic leukemia (B-CLL) patients, the microRNA genes *miR-15* and *miR-16* are located within a deletion region on chromosome 13, which is the most frequent genetic abnormality in B-CLL. Both microRNAs were subsequently not expressed or down-regulated in the majority of B-CLL cases. These findings led to further investigations of microRNAs in genomic regions associated with cancer. It is now known that approximately 50% of the annotated microRNAs are located inside or close to fragile sites, in regions of loss of heterozygosity in amplified regions and common breakpoints associated with cancer [74]. For example, *miR-145* and *miR-143* are located in a region on chromosome 5q33 which is known to be frequently deleted in myelodisplastic syndromes, or the *miR-17-92* cluster is located at 13q31, a region commonly amplified in lymphomas [75].

Based on the deregulation of microRNAs they can either act as oncogenes or tumor suppressors. MicroRNAs are negative regulators of gene expression. Consequently, if a specific microRNA is down-regulated in cancer and targets an oncogene it might act as tumor suppressor, whereas an up-regulation of a microRNA that targets a tumor suppressor gene will depict this microRNA as an oncogene [72]. The initial evidence for the existence of tumor suppressor microRNA gene was identified by the observed down-regulation of *miR-15* and *miR-16* in B-CLL [73]. Cimmino et al showed that both microRNAs act as tumor suppressors by targeting B-cell CLL/lymphoma 2 (*BCL-2*), which is frequently up-regulated in CLL [76]. Additional microRNAs that act as tumor suppressors are *miR-145* and *miR-143* which are located in fragile sites and known as down-regulated in colon- and breast-cancer patients [77-79] and *let-7*, which was one microRNA first to be discovered. There are 12 human paralogs of *let-7* which are all observed to be down-regulated in lung carcinomas compared to normal lung tissues. *Let-7* is experimentally validated to be a negative regulator of the *RAS* (p21 protein activator (GTPase activating protein) 1) proto-oncogene expression [80].

In contrast, the cluster *miR-17-92* acts as an oncogene. This cluster consists of six microRNAs (*miR-17, -18a, -19a, -20a, -19b-1* and *92a-1*). Because of its location in a commonly amplified region in B-cell lymphoma the expression level of the microRNA genes within this cluster is increased in 65% of B-cell lymphoma patients. MicroRNAs not only regulate the expression of protein-coding genes, they can also be transcriptional targets of oncogenes or tumor suppressor genes. A regulatory feedback mechanism is reported for the *miR-17-92* cluster, which is transcriptionally activated by the proto-oncogene *MYC* (v-myc myelocytomatosis viral oncogene homolog). Two microRNAs *miR-17* and *miR-20a* target the transcription factor *E2F1 (*E2F transcription factor 1), which is involved in cell cycle progression and apoptosis. *E2F1* is a direct target of *MYC*. Accordingly, *MYC* simultaneously activates *E2F1* and inhibits its translation through a microRNA expression control mechanism [81].

Another example for the oncogenic function of microRNAs is *miR-21*, which is overexpressed in different cancer types including colorectal cancer [82-88] and promotes cell growth by inhibiting the tumor suppressor gene tropomyosin 1 (Table **1.2**).

As with *miR-21* and the *miR-17-92* cluster, *miR-155* is an oncogenic microRNA. It is located in the non-coding *BIC* (*miR-155* host gene) gene region and is known to be consistently up-regulated in pediatric Burkitt lymphoma [89], classical Hodgkin disease [90] B-cell CLL [91] as well as in lung and breast cancer [78,92] (Table **1.2**). The *BIC* gene and *miR-155* seem to be regulated by nuclear factor-kappa B (*NF-kB*) through the B-cell receptor-mediated signaling pathway [90] or through the Toll-like receptor-activated signaling pathway [93].

Table 1.2: MicroRNA expression in different cancer types.

microRNA	*Cancer type*	*Regulation*	*Reference*
miR-15	Leukemia	down-regulated	73
miR-16	Leukemia	down-regulated	73
miR-21	Breast cancer	up-regulated	68
	Colon cancer	up-regulated	82-88
	Pancreatic cancer	up-regulated	82-88
miR-145	Breast cancer	down-regulated	67
	Colon cancer	down-regulated	67
miR-143	Breast cancer	down-regulated	67
	Colon cancer	down-regulated	67
miR-155	Breast cancer	up-regulated	68
	Lung cancer	up-regulated	82
miR-17-92 cluster	Breast cancer	up-regulated	68
	Lymphoma	up-regulated	75
miR-1	Lung cancer	down-regulated	94

Using microRNAs as a diagnostic and prognostic tool offers the advantage that microRNAs are relatively stable and resistant towards RNase degradation and therefore they could be more accurate biomarkers than mRNA counterparts [95,96]. Additionally, microRNAs are discussed for therapeutic applications in cancer therapy. To manipulate cancer progression in cells using microRNAs it is necessary to overexpress or inhibit microRNAs. The silencing of microRNA expression can be achieved using various modified antisense oligonucleotides, such as locked nucleic acid (LNA)-modified oligonucleotides. These novel class of chemically engineered oligonucleotides are named antagomirs [97]. To overexpress specific microRNAs, synthetic microRNA mimics including siRNA-like oligonucleotide duplex can be applied [98]. Several studies showed a successful knockdown of microRNAs using antagomirs and indicated for a long lasting silencing effect accompanied with a relatively high bioavailability. It was possible to silence a specific microRNA in different tissues except the brain because of the blood-brain barrier [99]. Improving the selective tumor-specific and tissue-specific delivery, antagomirs could function as therapeutic tools for cancer treatment. In the context of cancer treatment, microRNAs are also known to influence the drug sensitivity of specific cancer cells, described by Nasser et al. in 2008 [94]. It is also known that some chemical compounds modify the expression of multiple microRNAs simultaneously, which makes it reasonable to screen for drugs that

alter the expression of a group of microRNAs and do not only modulate the expression of a single candidate [100].

1.3.4 MICRORNAS FUNCTION IN METASTASIS

Metastasis of primary cancers is a multi-step process preceded by progressive growth and vascularization of the primary tumor followed by invasion of adjacent tissues and detachment to enter the systemic circulation. After surviving the circulation the tumor cells arrest in distant organs and extravasate into the surrounding tissues where they proliferate [101]. MicroRNAs are likely to be involved transiently or permanently in these processes. For example, *miR-125* is known to have an influence on cell motility and invasion of breast cancer cell lines [102]. *MiR-21*, which is well known to be up-regulated in a variety of tumors regulates motility, invasion, intravasation processes by down-regulating *PTEN* (phosphatase and tensin homolog) and *PDCD4* (programmed cell death 4 (neoplastic transformation inhibitor)) [85,103]. Another important microRNA family which is involved in metastasis progression is the *miR-200* family (*miR-200a, -200b, -200c, -141, -429*), in particular *miR-200c* and *miR-200b* which target *ZEB1* (zinc finger E-box biding homeobox 1) and *ZEB2* (zinc finger E-box biding homeobox 2). Both target genes promote the epithelial-mesenchymal transition (EMT) by repressing E-cadherin and vimentin transcription [104,105]. During EMT, a molecular reprogramming of epithelial cells to mesenchymal cells occurs and this process enables not only the remodeling of tissues during embryogenesis it also aids tumor invasion and metastasis. The posttranscriptional down-regulation of *ZEB1* and *ZEB2* by *miR-200* family members leads to an up-regulation of *E-cadherin* and inhibits the transformation from epithelial cells to mesenchymal cells and therefore it represses tumor metastasis [106,107]. On the other hand, *ZEB1* and *ZEB2* are able to inhibit the transcription of the *miR-200* family by binding to their putative promoter region. Interestingly, transforming growth factor-β (*TGF-β*) activates the expression of *ZEB1* and *ZEB2* and is also translational inhibited by *miR-141*, a member of the *miR-200* family. Hence, there is a reciprocal negatively feedback loop pathway regulating the expression of *miR-200* family, *ZEB1* and *ZEB2* which contributes to the EMT process (Figure **1.7**) [108,109].

1.3.5 MICRORNA FUNCTION IN CRC

MicroRNAs are known as important factors for tumorigenesis in many types of tumors including colon cancer. Functional studies of microRNAs in CRC indicated that microRNAs may act as tumor suppressors or oncogenes. Currently two approaches are of main interest in the investigation of microRNAs in CRC, the analysis of microRNAs that are involved in important pathogenic pathways of CRC and the large scale screening of microRNA expression profiles in CRC to identify novel potential biomarkers.

Deregulation of the Wnt/β-catenin pathway is one of the earliest events during CRC development and the pathway is activated by an inactivation of the adenomatous polyposis coli gene (*APC*). This occurs

in approximately 60% of all colorectal adenoma and carcinoma [9]. A range of microRNAs are known to regulate this pathogenic pathway in CRC. For example, *miR-135a* and *miR-135b* were identified by Nagel et al in 2008 to target *APC* with the ability to inhibit the translation of *APC in vitro* (Figure **1.7**). They also found an increased expression level of both microRNAs in colorectal cancer patients that showed a decreased level of *APC* [110]. In this pathogenic pathway β-catenin acts as a transcriptional activator and enhances the expression of *Wnt* target genes, such as *MYC*. Interestingly, Dews et al. showed that an activation of *MYC* contributes to a downstream activation of the *miR-17-92* cluster in mouse models for colon cancer, which regulates the expression of anti-angiogenic factors thrombospondin 1 (*TSP1*) and connective tissue growth factor (*CTGF*) (Figure **1.7**). An induced expression of *MYC* or *miR-17-92* cluster can lead to a vascularization of tumors [111]. *MiR-145* is supposed to regulate the expression of *MYC* indirectly by targeting the IGF-signaling pathway, where it inhibits the translation of *IRS-1* (insulin receptor substrate 1) and *IGF-1R* (insulin-like growth factor-1 receptor) in colon cancer cells [112,113]. Both genes play a major role in cell growth and cell proliferation processes [114].

Another pathogenic pathway which is influenced by the expression of microRNAs in CRC is the epidermal growth factor receptor (EGFR) pathway, which mediates tumor growth, survival, angiogenesis and metastasis by activating *KRAS* and *PI-3K* signaling. The epidermal growth factor receptor (*EGFR*) also serves as an important anti-cancer drug target [115]. The *KRAS* oncogene is known to be a direct target of the *let-7* microRNA family, of *miR-143* and *miR-18a* (Figure **1.7**). A low expression of *let-7* and *miR-18* was observed in colon cancer cells [116,117] and for *miR-143* an inverse correlation between *KRAS* protein levels and microRNA expression level was detected in clinical CRC samples [118]. The activation of the phosphatidylinositol-3-kinase (*PI-3-K*) pathway induces a cascade of anti-apoptotic and pro-survival signals. The *p85β* (phosphoinositide-3-kinase, regulatory subunit 2 (beta)) regulatory subunit of the *PI-3-K* signaling network is proven to be a direct target of *miR-126*, which is often found to be down-regulated in CRC cell lines and in primary colon tumors accompanied with an increased level of *p85β* [119]. The *miR-21*, which is frequently up-regulated in CRC, represses the tumor suppressor gene *PTEN* (phosphatase and tensin homolog), which is another important member of the *PI-3-K* pathway [120-122].

Recently, a strong connection between the TP53 pathway and the regulatory microRNA network was discovered [123]. The tumor suppressor gene *TP53* was found to mediate the expression of the highly conserved *miR-34* family genes. Target genes of *miR-34a* are involved in cell cycle progression, cellular proliferation, apoptosis, DNA repair and angiogenesis [124]. An experimental overexpression of *miR-34* family genes caused a down-regulation in well-known *TP53* targets, such as *CDK4* (cyclin-dependent kinase 4), *CDK6* (cyclin-dependent kinase 6), *CCNE2* (cyclin E2), *E2F5* (E2F transcription factor 5), *BIRC3* (baculoviral IAP repeat containing 3) and *BCL2* (B-cell CLL/lymphoma 2) [124,125] (Figure **1.7**). The overexpression of *miR-34a* induced also a down-regulation of *SIRT1* (sirtuin 1), which is a negative regulator of apoptosis in response to cellular stress [126]. When *SIRT1* was translationally inhibited by *miR-34a* in colon cancer cells, an apoptotic response was observed. In contrast, *TP53* mutant cells showed no response. Therefore, it is assumed that a positive regulatory

feedback loop between *miR-34* and the tumor suppressor gene *TP53* exists [126]. The observed down-regulation of *miR-34* in many tumors, including CRC is probably caused by a deletion of the genomic region on chromosome 1p36, where *miR-34a* is located and which is common in a variety of human tumors. Another potential explanation for the down-regulation of *miR-34a* is that it is brought about by a hypermethylation of the *miR-34a* promoter region. For the genomic regions of *miR-34b* and *miR-34c* a hypermethylation was also found with high frequency in CRC [127,128].

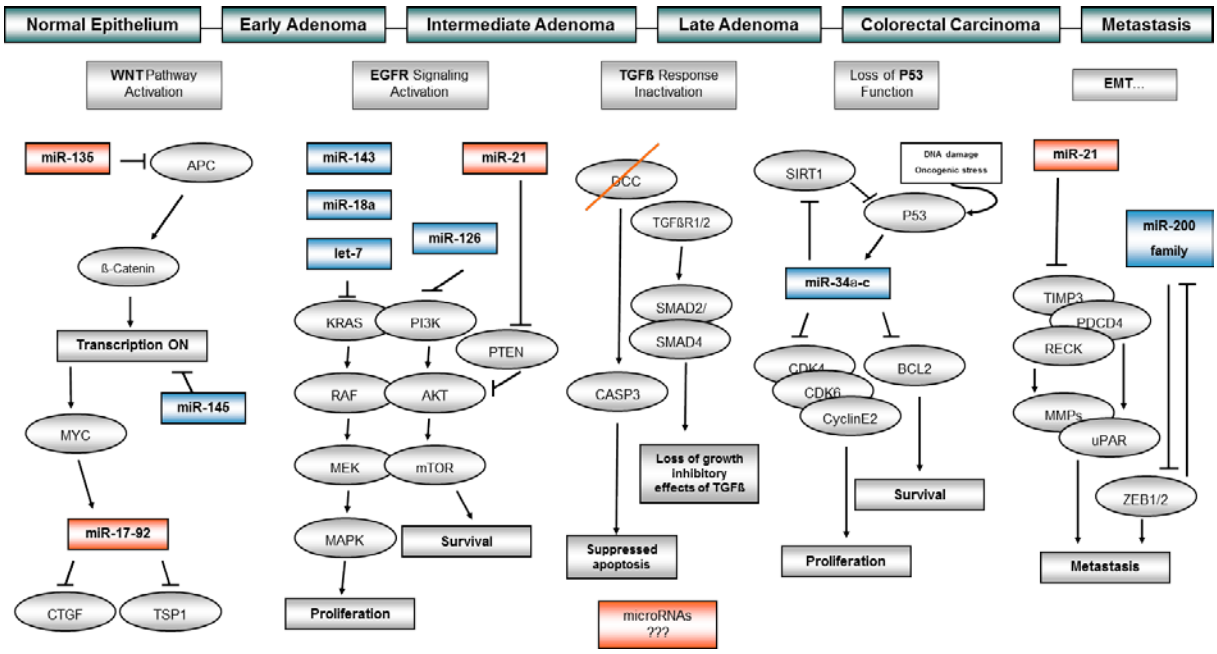

Figure 1.7: MicroRNAs in colorectal cancer pathogenesis. The influence of microRNA dysregulation on target genes described in the carcinogenesis of CRC is represented in a model (modification of Vogelstein's model of colorectal cancer pathogenesis). The development from normal epithelium to early adenoma, intermediate adenoma, late adenoma, colorectal carcinoma and metastasis is shown considering different pathways that are involved in the progression of CRC, such as Wnt signaling pathway, EGFR signaling pathway, TGFβ response, the loss of *TP53* and EMT - epithelial-mesenchymal transition. The deregulation of *miR-135, miR-145, miR-17-92 cluster, miR-143, miR-18a, let-7, miR126, miR-21, miR-34a-c* and *miR-200* family genes are displayed in different colors, blue corresponds to a down-regulation in CRC and red represents an up-regulation in CRC.

The association between the deregulation of microRNAs and metastasis progression in CRC was mainly analyzed in vitro and in mouse models. The remodeling of the extracellular matrix (ECM) is a key feature of tumor growth, survival, invasiveness and metastasis progression, where proteinases, such as urokinase plasminogen activator (uPA) and matrix metalloproteinases (MMPs) play a crucial role. Based on a study in brain tumors it was found, that *miR-21* is associated to cellular motility and ECM remodeling, while targeting multiple genes that are involved in these processes, such as *RECK* (reversion-inducing-cysteine-rich protein with kazal motifs) and *TIMP3* (TIMP metallopeptidase inhibitor 3). These target genes are known to inhibit MMPs and suppress malignancy. An up-regulation of *miR-21* was also determined in CRC cell lines with an accompanied increased migration and

invasion rate of the cells [129]. *MiR-21* is also predicted as a negative regulator of intravasation through inhibiting *uPAR*, which is crucial for invasion events during metastasis. *PDCD4* is another direct target of *miR-21*, and was strongly associated to invasive processes in CRC cell lines [103] (Figure **1.7**).

The epithelial-mesenchymal transition (EMT) is a crucial event for tumor invasion and metastasis in CRC, where epithelial cells convert into mesenchymal cells. This process triggers the expression of *ZEB1* and *ZEB2*, whereas *ZEB1* represses the expression of *E-cadherin* and promotes *vimentin* transcription and is able to suppress the transcription of the highly conserved *miR-200* family. The down-regulation of *miR-200* family genes is predicted to be a crucial event for metastasis not only in CRC [109 130] (Section **1.3.4**, Figure **1.7**).

1.4 BIOMARKERS IN COLORECTAL CANCER

For the identification of microRNAs as potential biomarkers which could enable prediction and prognosis of colorectal cancers it is necessary to investigate the dysregulation of microRNAs in malignant colon tissue. Changes in microRNA expression levels can be caused by epigenetic mechanisms, alterations in the microRNA processing machinery and cancer associated copy number aberrations of microRNA loci [131].

Cancer biomarkers enable the measurement of specific molecular alterations in order to evaluate pathogenic processes and determine predispositions and the response to therapeutic interventions for individual patients. There are various types of genetic biomarkers which can be applied in cancer, such as SNVs, CNVs, microsatellite instability, mitochondrial DNA or epigenetic status. Additional groups of biomarkers are proteins and RNA molecules including microRNAs. A range of potential non-invasive biomarkers have been discovered recently.

Among the available screening approaches colonoscopy and fecal occult blood test (FOBT) are most commonly used and recommended for individuals at high-risk. For the FOBT the hemoglobin of the colorectal and upper gastrointestinal tract can be detected either enzymatically by measuring the peroxidase-like activity of hemoglobin (Table **1.3**). This test only detects the heme group of hemoglobin which is also present in hemoglobin and myoglobin from meat which can lead to a high number of false-positive results. The FOBT can also be performed immunologically, using antibodies against human hemoglobin. FOBT shows a relatively poor sensitivity for the detection of adenomas (~10%) and CRC (40-85%). Clinical studies showed a weak reliability for the FOBT. Nevertheless, it reduces cancer mortality by about 30% [132,133,134]. In contrast, genes and epigenetic markers, such as *KRAS*, *APC*, *TP53*, long-form DNA (L-DNA) or microsatellite instability (MSI) are promising but not widely used for the routine detection of CRC (Table **1.3**). Alterations in the DNA markers can be determined non-invasively, by analyzing cell material from cryptal cells, which are shed continuously from the colonic mucosa or invasively from a tissue sample [135].

As an early-genetic event in the progression of colorectal adenomas and carcinomas, approximately 70-80% of all CRC cases show somatic mutations in the tumor suppressor gene *APC* with an accompanying inactivation of the gene [19,17]. *KRAS*, a member of the Ras family is a common target for somatic mutations in approximately 40% of CRCs [9,136]. Mutations in the *TP53* tumor suppressor gene facilitate continuous growth, invasive properties of the cells and tumor cell survival under varied stress conditions [137]. Half of all human cancers and approximately 30-60% of CRCs contain a *TP53* mutation which appears to be a relatively late genetic event in colorectal carcinogenesis and therefore it is limited in its use as a biomarker [138]. Microsatellite instability (MSI) is a condition in which microsatellite tandem repeats of 1-5 nucleotides, located throughout the entire genome, differ in local repeat counts between normal and malignant cells of an individual [139,8,140]. In 15% of all CRC cases microsatellite instability is observed, which can be due to epigenetic silencing or mutations of DNA mismatch repair genes, such as *MLH1* [141,142]. The occurrence of long DNA (L-DNA) in feces of CRC patients can be used as an additional DNA marker. Colonocytes usually undergo apoptosis which is accompanied by DNA fragmentation. This process is reduced in tumor cells which enables the detection of L-DNA for different genetic loci, such as *APC*, *TP53*, *BRCA1* (breast cancer 1, early onset) and *BRCA2* (breast cancer 2, early onset) in feces of CRC patients with a sensitivity of 57% and specificity of 94% [143]. During a large population study in 2004 a panel of 21 mutations in the *KRAS* gene (3), in the *APC* gene (10) and in the *TP53* gene (8) in combination with the common MSI marker *BAT-26* and L-DNA was used to detect CRC. Compared to the FOBT the usage of this fecal DNA panel showed a better sensitivity for invasive CRC without affecting specificity[144]. However, these fecal DNA markers are not yet in clinical use but at least in the clinical validation process.

Beside fecal biomarkers there are a range of potential serum and blood markers for CRC, such as Carcinoembryonic antigen (*CEA*) and Carbohydrate antigen 19-9 (*CA 19-9*) (Table **1.3**). Both of them are in clinical use, where *CEA* is employed for the prognosis of CRC patients, in that higher levels of *CEA* are associated with a worse outcome [145]. Another serum marker, the tissue inhibitor of metalloproteinase type 1 (*TIMP-1*) is being validated clinically (Table **1.3**). A number of serum markers are in the clinical development, for example a five-serum-marker panel for Spondin-2, tumor necrosis factor receptor superfamily member 6B (DcR3), TRAIL receptor 2 (TRAIL-R2), Reg IV and macrophage inhibitory cytokine 1 (MIC1) all of which are increased in CRC patients. Beside this panel there are additional promising protein candidates with the possibility to function as gastrointestinal markers, such as protesome activator complex subunit 3 (PSME3), nicotinamide N-methyltrasferase (NNMT) and collapsin response mediator protein-2 (CRMP-2). Using the surface-enhanced laser desorption/ionization (SELDI) mass spectrometry for apolipoprotein C1, complement C3a-desArg, α1-antitrypsin and transferrin helped to identify their diagnostic potential with a sensitivity of 95% and specificity of 91% in a clinical study performed by Martin et al. in 2006 [146] (Table **1.3**). Other serum proteins in clinical development are α-Defensin, macrophage migration inhibitory factor (MIF), macrophage-colony stimulating factor (M-CSF), prolactin, M2-pyruvate kinase (M2-PK), matrix metallopeptidase-9,-7 and laminin. The gene expression profiles of cytidinedeaminase (*CDA*), B-cell scaffold protein with anakyrin repeats 1 (*BANK1*), B-cell novel protein 1 (*BCNP1*) and membrane-spanning 4-domains, subfamily A, member 1 (*MS4A1*) in white blood cells are interesting candidate

genes for the detection of colorectal carcinomas. Beside genomic markers epigenomic markers are under development. Of these, a promising DNA blood marker for early epigenetic events in CRC is Septin-9 (*SEPT9*) which is frequently hypermethylated in CRC patients.

Table 1.3: Examples of molecular biomarkers in use or suggested for the detection of colorectal cancer.

Molecular biomarker	***Type***	***Subject***	***Status***
Fecal hemoglobin	Protein	Stool	as FOBT in use
CEA	Protein	Serum	In use
CA 19-9	Carbohydrate	Serum	In use
APC	DNA	Stool	Clinical validation
KRAS	DNA	Stool	Clinical validation
TP53	DNA	Stool	Clinical validation
L-DNA	DNA	Stool	Clinical validation
TIMP-1	Protein	Serum	Clinical validation
SELDI	Protein	Serum	Preclinical development
SEPT9	DNA	Plasma	Preclinical development

1.4.1 MICRORNAS AS BIOMARKERS IN COLORECTAL CANCER

The development of genomic technologies enables global scale analysis of genetic events and led to the discovery of microRNAs as new candidates for biomarkers. Altered microRNA expressions were identified to be potentially useful for prognosis and diagnosis of colorectal cancer [147,148,149,150]. In 2008, Mitchell et al. [151] discovered a high concentration of microRNAs in the peripheral blood of cancer patients where they were found highly stable in a cell free form. In contrast to mRNAs and proteins, circulating microRNAs are packed in complexes, such as exomes or microvesicles. They are released by normal and tumor cells, where they take place in different cell-to-cell signal transduction pathways and they are also responsible for the exchange of different genetic information [152]. This led to the assumption that circulating microRNAs could serve as potential biomarkers. For colorectal cancer microRNAs are available from various resources, such as blood plasma and feces. MicroRNAs belonging to the *miR-17-92* cluster showed an increased expression in the plasma of CRC patients whereas *miR-92a* was able to distinguish gastrointestinal cancer from inflammatory bowel disease [153] and advanced adenoma from normal control tissue [154]. The elevated expression level of *miR-29a* in tumor-derived plasma was associated with an advanced TNM (tumor-node-metastases) stage [154] (Table **1.4**). Several fecal microRNAs were identified as potential diagnostic markers, such as *miR-21, mir-106a, miR-96, miR-203, miR-20a, miR-326, miR-92a* and *miR-135* which were all found up-regulated in feces of CRC patients. Down-regulated microRNAs, for example, *miR-320, miR-126, miR-484-5p, miR-125b, miR-16, miR-143* and *miR-145* showed a strong correlation to an advanced colorectal cancer stage. The decreased expression of *miR-143* and *miR-145* are also correlated to a larger tumor size of CRC patients [122,155-161] (Table **1.4**).

Table 1.4: Biomarker candidates of microRNAs in colorectal cancer.

microRNA	*Regulation in CRC*	*Function of biomarker*	*Reference*
miR-92	up-regulated	diagnostic biomarker	153,154
miR-29a	up-regulated	diagnostic biomarker	154
miR-21	up-regulated	diagnostic and prognostic biomarker	122,158,159
mir-106a	up-regulated	diagnostic and prognostic biomarker	121
miR-96	up-regulated	diagnostic biomarker	121
miR-203	up-regulated	diagnostic biomarker	121
miR-20a	up-regulated	diagnostic biomarker	121
miR-326	up-regulated	diagnostic biomarker	121
miR-135b	up-regulated	diagnostic biomarker	157
miR-320	down-regulated	prognostic biomarker	159
miR-126	down-regulated	prognostic biomarker	160
miR- 484-5p	down-regulated	prognostic biomarker	159
miR-125b	down-regulated	prognostic biomarker	159
miR-16	down-regulated	prognostic biomarker	159
miR-143	down-regulated	prognostic biomarker	158,148 122
miR-145	down-regulated	prognostic biomarker	148 122 162
miR-31	up-regulated	diagnostic biomarker	161,122,162
miR-215	down-regulated	prognostic biomarker	163,164
let-7g	up-regulated	diagnostic and prognostic biomarker	165
miR-181b	up-regulated	diagnostic and prognostic biomarker	165
miR-18a	up-regulated	diagnostic and prognostic biomarker	166
miR-320	down-regulated	prognostic biomarker	159
miR-498	down-regulated	prognostic biomarker	159

The advantage of stool based microRNAs as diagnostic markers compared to plasma derived microRNAs is that they can be readily detected at earlier stages of CRC [167]. In several clinical studies *miR-21* was postulated as a potential prognostic marker. The observed up-regulation of *miR-21* in CRC was strongly correlated to the development of distant metastases [122,159] and a shorter disease-free interval, but not with overall survival [158] (Table **1.4**). A significantly increased expression level of *miR-31* in CRC primary tumor tissues and cell lines was detected by Bandres et al. [162] and Wang et al. [161] and associated to an advanced pathological TNM stage (Table **1.4**). As a predictive marker for treatment outcome in CRC *miR-215* could be an interesting candidate, since it is down-regulated in clinical colon cancer specimen and up-regulated in colon cancer stem cells. Interestingly, this microRNA affects an increased chemoresistance of the colon cancer cell line HCT116 to methotrexate and tomudex but not to cisplatin and doxorubicin [163] (Table **1.4**). Additional microRNAs that could serve as prognostic and predictive markers for CRC are, for example, *let-7g, miR-181b* [165] and *miR-18a* [166], which are all expressed at higher levels in CRC and associated with a poor prognosis. The aberrant expression of the two microRNAs *miR-320* and *miR-498* in CRC was shown to be correlated with the microsatellite status of CRC patients. Both of them are down-regulated in

microsatellite stable CRC patients, a finding which is associated with a shorter progression-free survival [159] (Table **1.4**). Taken together with recent findings, they confirm the potential of microRNAs as biomarkers for diagnosis, prognosis and susceptibility in CRC.

1.5 AIM OF THESIS

Driven by the motivation to provide a more detailed understanding of the molecular-biological mechanisms of colorectal cancer (CRC), we wanted to exhibit the putative relevance of microRNAs in the progression of CRC, since those molecules are discussed as novel diagnostic and therapeutic tools. Previous studies of microRNA expression patterns in CRC elucidated a strong association between expression levels of microRNAs and the tumor stage as well as the survival and prognosis for cancer patients [87,168,169]. These findings led us to explore alterations in the microRNA expression of colon cancer patients and cell lines to determine which individual microRNAs appear to have diagnostic, prognostic or therapeutic significance in colorectal cancer. Furthermore we wanted performed mutation analyses to gain further insight into the pathomechanisms of CRC. We applied next generation sequencing technologies in combination with standard molecular biological and cell biological methods to evaluate microRNAs as potential biomarkers and in order to find new and encouraging tools for an early detection of colorectal cancer, for diagnostic or therapy.

2 MATERIAL AND METHODS

2.1 EQUIPMENT

- BD FACSAria II flow cytometry cell sorter (*BD Biosciences*)
- BioAnalyzer 2100 (*Agilent*)
- Benchtop microcentrifuge (*Eppendorf*)
- Centrifuge (Beckman J-6B) (*Beckman*)
- Clean scalpels (*B. Braun*)
- Coplin jar (*IHC World*)
- CP100 Pipette Tips (*Gilson*)
- Cryovials (*VWR*)
- DAKO-Pen (*Dako*)
- Dark reader transilluminator or UV transilluminator (*Clare Chemical Research*)
- Development folder
- Electrophoresis power supply (*BioRad*)
- Elution chamber EL1 (*NimbleGen*)
- Elution System Retaining Ring (*NimbleGen*)
- Eppendorf Research pipettes (*Eppendorf*)
- Fluorescence spectrophotometer (*Agilent*)
- Humid chamber (*IHC World*)
- Magnetic rack *(Dynal MPC-S)*
- Microtome (*Olympus*)
- Mixer port seals (*NimbleGen*)
- NanoDrop (*Thermofisher Scientific*)
- Nebulizer Apparatus (*Roche*)
- NimbleGen 2.1M arrays (*NimbleGen*)
- NimbleGen HX1 Mixer for 2.1M arrays (*NimbleGen*)
- Nylon membrane plus-positively charged (*Invitrogen*)
- Paint brush (fine)
- Plastic coverslips 24mm x 60mm (*IHC World*)
- Precision Mixer Alignment Tool (PMAT) (*NimbleGen*)
- Sonifier cell disruptor B-30 (*G. Heinemann*)
- SpeedVac (*Thermofisher Scientific*)
- Stainless steel or tungsten carbide beads (3-7mm) (*Qiagen*)
- SuperFrost Plus slides (*Thermofisher Scientific*)
- Thermal cycler (*Harlow Scientific*)
- Thermal shaker (*Eppendorf*)
- TipOne RPT Filter Tips (*STARLAB*)
- TissueLyser II (*Qiagen*)

- TissueLyser Adapter Set 2 x 24 (*Qiagen*)
- Tube rotator (*Stuart*)
- T-25 cell culture flask (*TPP*)
- T-75 cell culture flask (*TPP*)
- UV cross linker (*Lehmann scientific*)
- Water bath
- XCell Sure Lock Mini-Cell electrophoresis unit (*Invitrogen*)
- X-Ray film (*Fuji*)
- 0.2µm syringe filter (*Sigma Aldrich*)
- 12-well plate for cell culture (*TPP*)
- 15ml falcon tubes (*BD Biosciences*)
- 21-gauge needles (*BD Microlance*)
- 50ml falcon tubes (*BD Biosciences*)
- 96-well plate for cell culture (*TPP*)
- 96-well plate for qPCR (*Applied Biosystems*)
- 384-well plate for qPCR (*Applied Biosystems*)
- 7900HT Fast Real-Time PCR system (*Applied Biosystems*)

2.2 DNA AND RNA OLIGONUCLEOTIDES

Table 2.1: Sequences of oligonucleotides used for different methods.

Oligonucleotide	***Sequence***	***Method***
Adapters	5'- P-GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG -3' 5' - ACACTCTTTCCCTACACGACGCTCTTCCGATCT - 3'	Primer for gDNA Library preparation
PCR primer 1.1	5' - AATGATACGGCGACCACCGACACTCTTTCCCTACACGACGCTCTTC CGATCT - 3'	
PCR primer 2.1	5' - CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT - 3'	
RT Primer	5' - CAAGCAGAAGACGGCATACGA - 3'	Primer for smallRNA Library preparation
5' RNA Adapter	5' - GUUCAGAGUUCUACAGUCCGACGAUC - 3'	
3' RNA Adapter	5' - P-UCGUAUGCCGUCUUCUGCUUGUidT - 3'	
Small RNA PCR Primer 1	5' - CAAGCAGAAGACGGCATACGA - 3'	
Small RNA PCR Primer 2	5' - AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA - 3'	
qPCR VEGFA fw	5' - AGGGCAGAATCATCACGAAG - 3'	Specific gene primer for qPCR
qPCR VEGFA rev	5' - GGTCTCGATTGGATGGCAG - 3'	
qPCR PDGFA fw	5' - GGGTCCATGCCACTAAGC - 3'	
qPCR PDGFA rev	5' - AATCTCGTAAATGACCGTCCTG - 3'	
qPCR PTPLAD1 fw	5' - GACGTGCATTGACATGGATTG - 3'	
qPCR PTPLAD1 rev	5' - TGGAATGGACTGAATCACTGAG - 3'	
qPCR HIF1A fw	5' - CCGCTGGAGACACAATCATATC - 3'	
qPCR HIF1A rev	5' - ACTTCCTCAAGTTGCTGGTC - 3'	
qPCR HPRT-F	5' - AGGAAAGCAAAGTCTGCATTG - 3'	

qPCR HPRT-R	5' - GGTGGAGATGATCTCTCAACT - 3'	
AP1	5' - A*T*C* AAGCCGAAGACAGTGT/3deoxyU (*: phosphorothioate) - 3'	Oligos for multiplex exon amplification
AP2	5' - 5Phos/TAGCGCGAATTTCCTGGATC - 3'	
Guide-DpnII	5' - GCGCGAATTTTCCTGGATCNNN - 3'	
CP-2-FA	5' - GCACGATCCGACGGTAGTGT - 3'	
CP-2-RA	5' - CCGTAATCGGGAAGCTGAAG - 3'	
CP-2-CIRC-A	5' - ACACTACCGTCGGATCGTGCACCGTAATCGGGAAGCTGAAG - 3'	
CP-2CIRC-NO-A	5' - ACACTACCGTCGGATCGTGCCCGTAATCGGGAAGCTGAA G – 3'	
gSel3 DNA Oligo	5' - CTC GAG AATTCTGGATCCTC - 3'	Oligos for microarray-based enrichment
gSel4-Pi DNA Oligo	5' - Phos/GAGGATCCAGAATTCTCGAGTT - 3'	
qPCR Oligo NSC-0237 fw	5' - CGCATTCCTCATCCCAGTATG - 3'	
qPCR Oligo NSC-0237 rv	5' - AAAGGACTTGGTGCAGAGTTCAG - 3'	
qPCR Oligo NSC-0247 fw	5' - CCCACCGCCTTCGACAT- 3'	
qPCR Oligo NSC-0247 rv	5' - CCTGCTTACTGTGGGCTCTTG - 3'	
qPCR Oligo NSC-0268 fw	5' - CTCGCTTAACCAGACTCATCTACTGT - 3'	
qPCR Oligo NSC-0268 rv	5' - ACTTGGCTCAGCTGTATGAAGGT - 3'	
qPCR Oligo NSC-0272 fw	5' - CAGCCCCAGCTCAGGTACAG - 3'	
qPCR Oligo NSC-0272 rv	5' – ATGATGCGAGTGCTGATGATG - 3'	
novel-miR-Chr6	5' – TGTGACAGGCTCCACTTTCAGTTCTCTTG - 3'	Probes for northern blot
novel-miR-Chr13	5' – ATCCAAACCCAACCATCTTAA - 3'	
novel-miR-Chr19	5' – GGAAATGATGAGCAGTTACTGATGCCA - 3'	
novel-miR-Chr17	5' – CTCTCGGCTCCTCGCGGCTCGCGGCGG - 3'	
novel-miR-Chr4	5' – GCTATTAAAGGTTCGTTTGTTCGACAAT - 3'	
U60-control	5' – ACCCTCAAGCCTCAGTCTTGCTAAATAATCAGACTG - 3'	
PTPLAD1-3UTR-XhoI-Fw	5'- TTTCTCGAGAGATGGCTTCTTGCCAGTTTGAGC -3'	Primer for to generate GFP-reporter constructs
PTPLAD1-3UTR-HindIII-Rev	5'- TTTAAGCTTTCTACTGCAAAGCTGATCTGCCCA -3'	
VEGFA-3UTR-XhoI-Fw	5'- TTTCTCGAGAAGGAGCCTCCCTCAGGGTTT -3'	
VEGFA-3UTR-HindIII-Rev	5'- TTTAAGCTTAACTGCACTAGAGACAAAGACGTG -3'	
PDGFA-3UTR-XhoI-Fw	5'- TTTCTCGAGCAACCAGATGTGAGGTGAGGATGA -3'	
PDGFA-3UTR-EcoRI-Rev	5'- TTTGAATTCAGAACATAGCAGGCTCTCAGGT -3'	
HIF1A-3UTR-XhoI-Fw	5'- TTTCTCGAGGAAAGTGGATTACCACAGCTGACC -3'	
HIF1A-3UTR-HindIII-Rev	5'- TTTAAGCTTGCCTGGTCCACAGAAGATGTTT -3'	

2.3 COLORECTAL CANCER TISSUE SAMPLES AND CELL LINES FOR NGS EXPERIMENTS

The primary colon carcinoma tissue and matched normal colonic epithelium as well as liver metastases tissue applied for the next generation sequencing (NGS) experiments were obtained from patients diagnosed with colorectal cancer and undergoing surgical resection (Table **A.11**). The samples were snap frozen in liquid nitrogen immediately after surgery and stored in liquid nitrogen. This was performed by the Medical University of Graz. We analyzed eight human colorectal cancer patients with four male and four female patients between the age of 44 and 80 years and with tumor-node-metastasis (TNM) state TNM IV. The set encompasses microsatellite instable (MSI) and stable (MSS) patients.

Colon cancer cell lines SW480 and SW620 were cultured in a humified atmosphere of 90% air, 5% CO_2 using Dulbecco's Modified Eagle's Medium (DMEM) medium and 10% fetal calf serum (FCS) and 1% penicillin/streptomycin (Section **2.5**).

2.4 TISSUE SAMPLES FOR CANCER SCREEN

For the tumor tissue screen 243 tumor samples and 87 normal tissue samples were selected from a tissue bank of the Institute of Pathology from the Medical University of Graz, Austria, to estimate the microRNA expression in a variety of human cancer types. This archive consists of tumor and normal tissue samples obtained from different organs, such as adipose tissue, brain, breast, colon, endometrium, kidney, liver, lung, lymphatic system, muscle, ovary, pancreas, prostate, stomach, testis and thyroid gland (Table **3.4**). All tissue samples were reviewed and diagnosed histopathologically and RNA quality was assessed by automated electrophoresis.

2.5 CELL LINES

SW480 CELLS:

The cell line was derived from a primary adenocarcinoma of a 50 year old male colon cancer patient (Tumor Stage: Dukes' type B). In this cell line a G to A mutation in codon 273 of the *TP53* gene (tumor protein p53) is present and results in an arginine to histidine substitution. A mutation from C to T in codon 309 generates a proline to serine substitution. The cells express an elevated level of *TP53* and they are positive for expression of *MYC* (v-myc myelocytomatosis viral oncogene homolog), *KRAS* (v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog), *HRAS* (v-Ha-ras Harvey rat sarcoma viral oncogene homolog), *NRAS* (neuroblastoma RAS viral (v-ras) oncogene homolog), *MYB* (v-myb myeloblastosis viral oncogene homolog), *PDGFB* (platelet-derived growth factor beta polypeptide) and *FOS* (FBJ murine osteosarcoma viral oncogene homolog) oncogenes (ATCC number CCL-228).

SW620 CELLS:

This cell line was established from a lymph node metastasis taken from the same patient as for the SW480 cell line one year later (51-year-old Caucasian male). The same mutations in in codon 273 of the *TP53* gene is present and it is also positive for the expression of *MYC*, *KRAS*, *HRAS*, *NRAS*, *MYB*, *PDGFB* and *FOS* oncogenes (ATCC number CCL-227).

HEK-293 CELLS:

The cells were derived from human embryonic kidney cells and generated by transformation with sheared adenovirus 5 DNA. The cells were originally obtained from a single apparently healthy fetus (ATCC number CRL-1573).

2.6 MOLECULAR BIOLOGICAL METHODS

2.6.1 BASIC MOLECULAR BIOLOGICAL METHODS

POLYMERASE CHAIN REACTION

MATERIAL:

10X PCR reaction buffer	500mM Tris-HCl, pH 8.8	(*Merck*)
	200mM $(NH4)_2SO_4$	(*Merck*)
	15mM $MgCl_2$	(*Merck*)
	0.1% Tween 20	(*Applichem*)
	in ddH_2O	

Reaction master mix, 25 µl reaction:	ddH_2O 15.75 µl
	10X PCR reaction buffer 2.5 µl
	25mM dNTP mix 0.25 µl
	1 µM forward primer 0.5 µl
	1 µM reverse primer 0.5 µl
	Pfu-Polymerase (10U/µl) 0.1 µl
	Taq-Polymerase (10U/µl) 0.4 µl
	template DNA (15 ng/µl) 5.0 µl

PROTOCOL

The polymerase chain reaction is common molecular biology technique, which enables the exponential amplification of a DNA fragment. This scientific method was first described by Saiki and Mullis et al. in 1985 [170]. A standard PCR reaction needs a thermo-stable DNA polymerase, a DNA template, forward and reverse primer, nucleotides and reaction buffer. The typical PCR program consists of a denature step at 95°C, annealing step at approximately 55°C and an elongation step at 72°C. Those steps will be repeated in 25 to 30 cycles. The standard program is as follows:

Initial denature step	95°C for 5 minutes
First step	95°C for 1 minute
Second step	primer Tm (melting temperature) – 5°C for 1 minute
Third step	72°C for 1-2 minutes per kb of the PCR target
Final elongation step	72°C for 5 minutes
Storage	4°C

To amplify the DNA template the first to third step are repeated 25 to 30 times. The amplified DNA product can be visualized on by agarose gel electrophoresis.

AGAROSE GEL ELECTROPHORESIS

MATERIAL:

TAE (50X)	2mol/Tris Base	(*Merck*)
	57ml/l Acetic acid	(*Merck*)
	50mmol/l EDTA (0.5M) pH 8.0	(*Roth*)
	in ddH_2O	
TAE running buffer	1X TAE	
Agarose		(*Invitrogen*)
Ethidium bromide-stock solution		(*Roth*)
Marker	1kb DNA ladder	(*Fermentas*)
10X DNA loading dye	400ml/l Glycerol	(*Thermofisher Scientific*)
	50mmol/l Tris	(*Merck*)
	Tip-off Bromphenol blue	(*Sigma Aldrich*)
	in ddH_2O	

PROTOCOL:

Using this scientific technology it is possible to separate DNA fragments between 0.5 and 25 kb in length. Therefore the appropriate concentration of the gel was determined according to the expected length of the DNA fragments and prepared in the corresponding amount of 1X TAE running buffer and heated in the microwave. The solution was then cooled down to approximately 60°C before Ethidium Bromide (EtBr) was added and the gel was transferred to an agarose gel chamber filled with 1X TAE running buffer. After complete polymerization of the gel, it was loaded with the marker DNA and DNA sample all combined with 1X DNA loading dye. Then, the gel was attached to 100V for approximately 1h or until the bromphenol blue reached the lower part of the gel.

PURIFICATION OF PCR PRODUCTS USING QIAQUICK PCR PURIFICATION KIT (QIAGEN)

MATERIAL:

QIAquick PCR Purification Kit (# 28106)	(*Qiagen*)
Binding buffer PB	(*Qiagen*)
Wash buffer PE	(*Qiagen*)
QIAquick Spin Column	(*Qiagen*)
Elution buffer EB	(*Qiagen*)

PROTOCOL

For purification of PCR products the Qiagen QIAquick PCR Purification Kit was applied. The method was established to cleanup single- or double-stranded DNA fragments of 100bp to 10kb length from PCR and other enzymatic reactions. Therefore 5 volumes of Buffer PB were added to the sample and mixed well before it was applied to the QIAquick column placed in 2ml collection tube and centrifuged for 1 minute at 10000xg. The flow-through was discarded and to wash the DNA 750µl of Buffer PE were added and centrifuged through the membrane of the QIAquick column for 1 minute at 10000xg. An additional centrifugation step for 1 minute at 10000xg removed remaining ethanol from the washing buffer and afterwards the DNA was eluted with 30-50µl Elution Buffer (EB) in a new 1.5ml collection tube and the DNA concentration was determined using NanoDrop technology.

ISOLATION OF DNA FRAGMENTS FROM AGAROSE GEL USING QIAGEN PROTOCOL

MATERIAL:

QIAquick Gel Extraction Kit (# 28706)	(*Qiagen*)
Binding buffer QG	(*Qiagen*)
Wash buffer PE	(*Qiagen*)
QIAquick Spin Column	(*Qiagen*)
Elution buffer EB	(*Qiagen*)

PROTOCOL:

To extract DNA from a low-melt TAE or TBE agarose gel, generally the Qiagen QIAquick Gel Extraction Kit was applied according to the manufacture's protocol. The protocol was optimized for the extraction of 70bp to 10 kb DNA. Therefore a gel slice containing the corresponding DNA fragment was excised from the gel using a UV transilluminator and a clean scalpel. The weight of the gel slice was measured and 3 volumes of Buffer QG were added to 1 volume of the gel. To dissolve the gel it was incubated by vigorous shaking at 50°C for 10 minutes until the gel slice was completely melted. For purification of DNA fragments that were below 500 bp in length and larger than 4 kb 1 gel volume of isopropanol was added to increase the yield of DNA. For all the remaining DNA fragments this step was optional and had no effect on the yield. The sample was applied to QIAquick spin column, which was placed in a 2ml collection tube and centrifuged for 1 minute at 10000xg. The flow-through was discarded and 500µl of QG Buffer was added to the QIAquick spin column and centrifuged for 1 minute at 10000xg. To wash the DNA, 750µl of PE Buffer was added to the membrane of the column and centrifuged for 1 minute at 10000xg. The flow-through was discarded and the column was centrifuged for an additional minute at 10000xg to remove remaining wash buffer. To elute the DNA 30-50µl of EB Buffer were added directly to the QIAquick membrane and centrifuged in a new 1.5ml collection tube at 10000xg for 1 minute. The concentration of the DNA was measured using NanoDrop technology.

2.6.2 PREPARATION OF TOTAL RNA FROM HUMAN CELL LINES

MATERIAL:

Trizol reagent	(*Invitrogen*)
Chloroform	(*Applichem*)
RNase-free DNase Digestion Kit (# 79254)	(*Qiagen*)
Phenol/Chloroform/Isopropyl alcohol 25/24/1	(*Applichem*)
100% ethanol	(*VWR*)
75% ethanol	(*VWR*)
10M Ammonium acetate	(*Merck*)
GenElute-LPA (linear polyacrylamide)	(*Sigma Aldrich*)
Phase lock tubes (heavy)	(*5 Prime*)
BioAnalyzer Agilent DNA 7500 Kit (# 5067-1506)	(*Agilent*)

PROTOCOL:

RNA was extracted applying Trizol reagent. For RNA extraction from cells harvested on a 12-well plate 150µl were added to the cells, which were previously washed with PBS. Using a 1 ml pipette tip the cells were scraped off and collected in a 1.5 ml tube and immediately placed on dry ice and stored at -80°C. After thawing the cells on ice 30µl chloroform was added. The samples were mixed vigorously and incubated for 2-3 minutes on ice before the nucleic acids were separated from insoluble material by centrifugation for 15 minutes at 12000xg at 4°C. The upper phase was transferred to a new 1.5 ml collection tube and 1µl GenElute-LPA and 150µl of isopropyl alcohol were added and incubated for 30 minutes at -80°C to precipitate the RNA and then incubated for 10 minutes at room temperature before the RNA was pelleted by centrifugation for approximately 10 minutes at 12000xg and the supernatant was discarded. The pellet was washed in in 500µl 75% ethanol air dried and diluted in 87.5µl RNase-free water. To eliminate the remaining DNA from the RNA solution a DNase digestion step was performed using the RNase-free DNase set from Qiagen. Therefore 10µl of the RDD buffer and 2.5µl DNase I were added to a final volume of 100µl of the solution. The sample was incubated for 10 minutes at room temperature and afterwards cleaned. For cleaning procedure the 100µl solution was transferred to a Phase lock tube (heavy) and phenol/chloroform/isopropyl alcohol 25/24/1 was added to the sample mixed vigorously and incubated for 5 minutes at room temperature. To separate the organic phase from the aqueous RNA containing phase the samples were centrifuged for 5 minutes at 8400xg. The upper phase was transferred to a new 1.5 ml collection tube. For ethanol precipitation 1µl GenElute-LPA 0.2 volume 10M ammonium acetate and 2 volumes 100% ethanol were added to the sample and mixed vigorously before they were incubated at -80°C for approximately 30 minutes to precipitate the RNA. After incubation the samples were centrifuged for 15 minutes at full speed at 4°C to pellet the RNA. The supernatant was discarded and the pellet was washed in 500µl 75% ethanol and air dried. The RNA pellet was finally resuspended in 30µl RNase-free water. The integrity was assessed using the Agilent BioAnalyzer 2100 technology.

2.6.3 PREPARATION OF GENOMIC DNA, TOTAL RNA AND TOTAL PROTEINS FROM HUMAN TISSUE SAMPLES

MATERIAL:

Allprotect Tissue Reagent	(*Qiagen*)
TissueLyser II	(*Qiagen*)
TissueLyser Adapter Set 2 x 24	(*Qiagen*)
Stainless steel or tungsten carbide beads (3-7mm)	(*Qiagen*)
AllPrep DNA/RNA/Protein Mini Kit (# 80204)	(*Qiagen*)
Lysis buffer RLT	(*Qiagen*)
AllPrep DNA spin column	(*Qiagen*)
RNeasy spin column	(*Qiagen*)
Wash buffer RW1	(*Qiagen*)
Protein precipitation buffer APP	(*Qiagen*)
70% ethanol	(*VWR*)
Elution buffer ALO	(*Qiagen*)
Wash buffer AW2	(*Qiagen*)
Elution buffer EB	(*Qiagen*)

PROTOCOL:

Disruption and homogenization of human tissue samples is fundamental for RNA extraction. Therefore the TissueLyser system is an effective instrument. For the procedure 5-10 mg of Allprotect stabilized tissue were placed in an RNase-free 2 ml microcentrifuge tube that contained 1 stainless steel bead (3–7 mm mean diameter) and an appropriate volume (350µl) of lysis buffer (Buffer RLT) was added. The tubes were placed to the TissueLyser Adapter Set 2 x 24 and disruption was carried out for 2 minutes at 20–30 Hz. After rearranging the tubes in the rack of the TissueLyser it was activated for additional 2 minutes at 20-30 Hz until the tissue was completely homogenized. Subsequently genomic DNA, total RNA and total proteins were purified using AllPrep® DNA/RNA/Protein Mini Kit. In general, preparation was performed according to the manufacturer's protocol, where lysate of homogenized tissue and RLT buffer were centrifuged for 3 minutes at full speed. The supernatant was transferred to an AllPrep DNA spin column, which was placed to a 2 ml collection tube and centrifuged for 30 seconds at 7500xg. The AllPrep spin column was stored in a new 2 ml collection tube at 4°C later DNA purification steps. For total RNA purification the flow through was mixed with 250µl 96-100% ethanol and 700µl of the sample was transferred to an RNeasy spin column placed in a 2 ml collection tube. Afterwards it was centrifuged for 15 seconds at 7500xg and the flow through was stored for later protein isolation. To wash the spin column membrane 700µl RW1 buffer was added to the RNeasy spin column and centrifuged for 15 seconds at 7500xg. For an additional washing step 500µl RPE buffer was added to the RNeasy spin column and centrifuged for 15 seconds at 7500xg. This step was repeated. To dry the spin column completely it was placed in a new 2 ml collection tube and centrifuged at full speed for 1 minute. The RNeasy spin column was placed in a new 1.5 ml collection tube and 50µl RNase-free water was pipetted directly on the membrane and incubated for 1 minute. To elute the RNA from the membrane the spin column was centrifuged for 1 minute at 7500xg. The

quality of the eluted RNA was checked using BioAnalyzer technology and stored at -80°C. The protein containing flow-through from previous cleaning steps was used for a total protein extraction. Therefore 600µl of APP buffer was added to the sample and incubated for 10 minutes at room temperature in order to precipitate the protein, before it was centrifuged at full speed for 10 minutes. To the protein pellet 500µl of 70% ethanol was added and centrifuged at full speed for 1 minute and the supernatant was removed. Afterwards the protein pellet was completely dried for approximately 10 minutes at room temperature. To dissolve the pellet up to 100µl ALO buffer was added and mixed vigorously. In order to denature the protein sample, it was incubated for 5 minutes at 95°C and then cooled to room temperature before it was centrifuged for 1 minute at full speed to pellet any remaining material. The protein containing supernatant was stored at -20°C. The AllPrep DNA spin column stored at 4°C was used to proceed with the genomic DNA isolation. So 500µl AW1 buffer were added to the spin column and centrifuged for 15 seconds at 7500xg. Then, 500µl of buffer AW2 were added and centrifuged for 2 minutes at full speed to wash the column membrane. To elute the DNA from the AllPrep spin column membrane it was placed in a new 1.5 ml collection tube and 100µl EB buffer were added directly to the center of the membrane and incubated for 2 minutes at room temperature before it was centrifuged for 1 minute at 7500xg. The eluted DNA was measured using NanoDrop technology.

2.6.4 PREPARATION OF GENOMIC DNA FROM HUMAN BLOOD SAMPLES

MATERIAL:

Lysis buffer	10mM Tris, ph 8.0	(*Merck*)
	0.1M EDTA, ph 8.0	(*Roth*)
	0.5% SDS	(*Roth*)
	DNase-free pancreatic RNase (20µg/ml) in ddH_2O	
Proteinase K (20mg/ml)		(*Roche*)
10M ammonium acetate		(*Merck*)
GenElute-LPA (linear polyacrylamide)		(*Sigma Aldrich*)
Phenol equilibrated with 0.1M Tris		(*Applichem*)
100% ethanol		(*VWR*)
70% ethanol		(*VWR*)
TE buffer (pH 8.0)		(*Illumina*)

PROTOCOL:

To isolate high-molecular-weight DNA from mammalian cells, a method originally described by Stafford et al. [171] was applied in a modified manner. Approximately 250µg mammalian genomic DNA, 100-150 kb in length could be yielded from 20 ml of human blood sample. Therefore the freshly drawn blood sample was transferred to a falcon collection tube and centrifuged at 2000xg for 15 minutes at

4°C to separate the different layer of cells. The plasma containing supernatant fluid was removed by aspiration and discarded. Then, the buffy coat, which is characterized as a broad band of white blood cells of heterogeneous density, was carefully removed using a Pasteur pipette and transferred to a fresh collection tube. The remaining layer and the pellet of red blood cells were discarded. Compared to the plasma and the red blood cells, the white blood cells contain more DNA. To acquire a maximum amount of those cells from the buffy coat the centrifugation step could be repeated. The buffy coat was then resuspended in 15 ml of lysis buffer and incubated for 1 h at 37°C. After incubation the lysate was transferred to a 50 ml centrifuge tube and proteinase K was added to a final concentration of 100µg/ml. The viscous solution was incubated overnight at 55°C and swirled continuously. The following day the solution was cooled down to room temperature and an equal volume of phenol equilibrated with 0.1M Tris (pH 8.0) was added. The solution was mixed slowly by turning the tube end-over-end until the two phases formed an emulsion. This could last between 10 to 60 minutes. To separate the phenolic and the DNA containing aqueous phase the mixture was centrifuged at 5000xg for 15 minutes at room temperature. The upper aqueous phase was transferred to a new collection tube. Repeating the phenol extraction could yield a higher amount of genomic DNA. Then, the obtained aqueous phases were pooled before the DNA was isolated by ethanol precipitation. For the ethanol precipitation of DNA with an average size of 100-150kb, 1µl LPA and 0.2volume of 10M ammonium acetate were added to the pooled aqueous phases, as well as 2 volumes of 100% ethanol at room temperature. The solution was mixed well and incubated at -80°C for approximately 20 minutes. Centrifugation at 5000xg for 30 minutes collected the DNA as a pellet on the bottom of tube. The pellet was washed twice in 70% ethanol and air dried before the DNA was resuspended in 100-200µl TE buffer and stored at -20°C. The concentration of the DNA was measured using NanoDrop technology.

2.6.5 RIBOMINUS TRANSCRIPTOME ISOLATION

MATERIAL:

RiboMinus Human/Mouse Transcriptome Isolation Kit (#K1550-02)	(*Invitrogen*)
Magnetic rack	(*Dynal*)
RiboMinus streptavidin coated magnetic beads	(*Invitrogen*)
RNase-free water	(*Invitrogen*)
RiboMinus probe	(*Invitrogen*)
Hybridization Buffer B5	(*Invitrogen*)
GenElute-LPA (linear polyacrylamide)	(*Sigma Aldrich*)
10M ammonium acetate	(*Merck*)
Isopropanol	(*Sigma Aldrich*)
70% ethanol	(*VWR*)

PROTOCOL:

For isolation of human transcriptome the RiboMinus Human/Mouse Transcriptome Isolation Kit was used. It enabled the elimination of abundant large ribosomal RNA molecules from total RNA which is used for further analyses. Therefore human rRNA sequence-specific 5'-biotinylated oligonucleotide probes and streptavidin coated magnetic beads were used. As input amount of total RNA 2µg was diluted in 20µl RNase-free water and 8µl of the RiboMinus probe were added to the RNA sample 300µl Hybridization Buffer B5 and incubated for 5 minutes at 75°C. Then, the sample was cooled down slowly to 37°C over the time of 30 minutes to hybridize the rRNA to the 5'-biotin labeled probe. While the sample was cooling down the streptavidin coated RiboMinus Magnetic Beads were prepared. Therefore 500µl of the beads per sample were pipetted into a 1.5 ml microcentrifuge tube and placed for 1 minute on a magnetic stand. During this procedure the beads settled to the bottom of the tube. The supernatant was aspirated and discarded. To wash the beads, 500µl of RNase-free water was pipetted to the beads and was resuspended by inverting and gently tapping the tube. The tube with the bead suspension was placed on a magnetic rack for 1 minute and the supernatant was removed. The water cleaning step was repeated and followed by re-suspension step in 500µl Hybridization Buffer B5. The tube was placed on a magnetic rack for 1 minute and the supernatant removed. Until hybridization the bead pellet was diluted in 200µl Hybridization Buffer B5 and incubated at 37°C. The hybridization sample was collected by briefly centrifuging, and the 328µl hybridization sample was added to the prepared beads and mixed well. The solution was incubated for 15 minutes at 37°C. During incubation the rRNA/5'-biotin labeled probe and bead complex was gently mixed occasionally. The tube was placed on a magnetic rack for 1 minute and the supernatant (~528µl) was transferred to a new 2 ml collection tube. To precipitate the RNA from this solution, 1µl GenElute LPA 50µl of 10M ammonium acetate and 600µl Isopropanol were added and incubated for 30 minutes at -80°C. Afterwards it was centrifuged for 15 minutes at 4°C at full speed. The pellet was washed in 70% ethanol and air dried. To resolve the pellet 10µl RNase-free water was added. The RNA was stored at -80°C.

2.6.6 REVERSE TRANSCRIPTION FOR CDNA SYNTHESIS

MATERIAL:

SuperScript II – Reverse Transcription Kit (# 18064-022)	(*Invitrogen*)
Random hexamer 50ng/µl	(*Invitrogen*)
10mM dNTP mix (each)	(*Invitrogen*)
0.1M DDT	(*Invitrogen*)
5X first strand buffer	(*Invitrogen*)
RNaseOut (40U/µl)	(*Invitrogen*)
SuperScript reverse transcriptase	(*Invitrogen*)
5X Second strand buffer	(*Invitrogen*)
E. coli DNA Ligase (10U/µl)	(*Invitrogen*)

E. coli DNA Polymerase I (10U/µl)	(*Invitrogen*)
E. coli RNaseH (2U/µl)	(*Invitrogen*)
Phase Lock tube (light)	(*5 Prime*)
Phenol/Chloroform/Isopropyl alcohol 25/24/1	(*Applichem*)
GenElute-LPA (linear polyacrylamide)	(*Sigma Aldrich*)
10M ammonium acetate	(*Merck*)
100% ethanol	(*VWR*)
70% ethanol	(*VWR*)

PROTOCOL:

For cDNA synthesis two differing protocols were applied. One, which is based on the reverse transcription of RiboMinus transcriptome isolated RNA, and the other one is based on the use of Trizol isolated total RNA normally extracted from cell line material.

For double strand cDNA synthesis of RNA which was freed from ribosomal RNA molecules, we applied the 10µl RNA solution isolated by RiboMinus Human/Mouse Transcriptome Isolation Kit, which was thawed on ice before adding 4.7µl random hexamer primer of a 50ng/µl stock solution and 1.25µl dNTPS (10mM each). The solution was heated to 65°C for 5 minutes and quick chilled on ice. After cooling down the sample 2.5µl DDT (0.1M) 5µl of 5X first strand buffer and 0.5µl RNaseOut were added and incubated for 2 minutes at 25°C before 1µl SuperScript was added. The sample was incubated for additional 10 minutes at 25°C and then heated up to 42°C for 50 minutes to initiate the activity of the reverse transcriptase before the sample was heated up to 70°C for 15 minutes and afterwards quick chilled on ice. To the 25µl of the first strand reaction 85µl water 30µl 5X second strand buffer 3µl dNTPs (10mM each) 2µl E. coli DNA Ligase (10U/µl) 4µl E. coli DNA Polymerase I (10U/µl) and 1µl E. coli RNaseH (2U/µl) were added. The solution was incubated at 16°C for 2h and cooled on ice. For final sample purification the 150µl of the synthesized dsDNA were transferred to a Phase Lock tube (light) and 150µl Phenol/Chloroform/Isopropyl alcohol 25/24/1 was added and mixed vigorously and incubated for 5 minutes at room temperature before being centrifuged for 5 minutes at 8500xg. The upper dsDNA containing phase was transferred to a new 1.5 ml collection tube and 1µl GenElute LPA 0.2 volume 10M ammonium acetate and 100% ethanol were added. The solution was incubated at -80°C for 30 minutes and afterwards centrifuged. The pellet was washed with 500µl 70% ethanol and air dried. The pellet was finally resuspended in 10µl water and the concentration was measured using NanoDrop technology. The dsDNA generated by this protocol was mainly used for further NGS investigations.

To generate single strand cDNA that is mainly used for quantitative real-time applications we applied another reverse transcription protocol. Therefore 500ng total RNA were used as input amount and 250ng random hexamer primer, 1µl 10mM dNTPs and RNase-free water were added to a final volume of 13µl. The reaction was heated up to 65°C for 5 minutes and quick chilled on ice. Then, 4µl of 5X first strand buffer, 1µl 0.1M DTT, 1µl of a 40U/µl concentrated RNaseOut solution and 1µl of were

added. The reaction was incubated at 25°C for 5 and then heated up at 50°C for 60 minutes and to inactivate the enzyme it was incubated at 70°C for 15 minutes. To eliminate the remaining RNA molecules 2U of RNaseH were added to the reaction and incubated for 20 minutes at 37°C. With an additional incubation step at 65°C for 20 minutes the RNaseH was incubated. The final sample purification of the generated single stranded cDNA corresponds to the already described final cleaning procedure for the synthesized double stranded cDNA. The pellet was resolved in 30µl water and the concentration determined by NanoDrop technology.

2.6.7 PREPARATION OF SEQUENCING LIBRARY FOR DNA SOLEXA SEQUENCING

MATERIAL:

Genomic Sample Prep Kit (# FC-102-1001)	(*Illumina*)
TE buffer	(*Illumina*)
Bioanalyzer Agilent DNA 1000 Kit (# 5067-1504)	(*Agilent*)
T4 DNA ligase buffer	(*Illumina*)
10mM ATP	(*Illumina*)
dNTP mix	(*Illumina*)
T4 DNA polymerase	(*Illumina*)
Klenow DNA polymerase	(*Illumina*)
T4 Polynucleotide Kinase	(*Illumina*)
Klenow buffer	(*Illumina*)
dATP	(*Illumina*)
Klenow exo with polymerase activity (3' to 5' minus)	(*Illumina*)
DNA ligase buffer	(*Illumina*)
Adapter oligo mix	(*Illumina*)
DNA ligase	(*Illumina*)
Phusion DNA polymerase	(*Illumina*)
PCR primer 1.1	(*Illumina*)
PCR primer 1.2	(*Illumina*)
QIAquick PCR Purification Kit (# 28106)	(*Qiagen*)
MinElute PCR Purification Kit (# 28004)	(*Qiagen*)
Binding buffer PB	(*Qiagen*)
Wash buffer PE	(*Qiagen*)
QIAquick Spin Column	(*Qiagen*)
Elution buffer EB	(*Qiagen*)
Certified low-range Ultra Agarose	(*BIO-RAD*)

TAE (50X)	2mol/Tris Base	(*Merck*)
	57ml/l Acetic acid	(*Merck*)
	50mmol/l EDTA (0.5M) pH 8.0	(*Roth*)
	in ddH_2O	
TAE running buffer	1X TAE	
Ethidium bromide-stock solution		(*Roth*)
DNA Loading buffer		(*Illumina*)
Low molecular weight (LMW) DNA ladder		(*New England Biolabs*)
QIAquick Gel Extraction Kit (# 28706)		(*Qiagen*)

PROTOCOL:

To prepare a sequencing DNA library it is essential to fragment the genomic DNA or generated cDNA to a general size of less than 800bp. Therefore an ultrasonic bath (Sonifier cell disruptor B-30) was used. Up to 5 µg of the DNA was made in a final volume of 30µl using TE buffer and sonicated for 1.5 minutes. The achieved fragment size was verified on a DNA Chip using BioAnalyzer technology. To perform the following end repair reaction, which converts overhangs resulting from fragmentation procedure into blunt ends 45µl water was added to the sample as well as 10µl T4 DNA ligase buffer with 10mM ATP (Illumina), 4µl dNTP mix (Ilumina), 5µl T4 DNA polymerase (Illumina), 1µl Klenow DNA polymerase and T4 Polynucleotide Kinase (Illumina) to a total volume of 100µl. The mixture was incubated in a thermal cycler for 30 minutes at 20°C and purified using QIAquick PCR Purification Kit (Qiagen). To elute the bounded DNA from the QIAquick column 32µl EB was used. In the following protocol step 'A' bases were added to the 3'-end of the blunt phosphorylated DNA fragments. Therefore 5µl Klenow buffer (Illumina), 10µl dATP (Illumina) and 3µl Klenow exo with polymerase activity (3' to 5' minus) were added. The reaction was incubated at 37°C for 30 minutes and purified using MinElute PCR Purification Kit (Qiagen). For Elution 10µl EB was applied to the QIAquick MinElute column. This procedure prepared the DNA fragments for the adapter ligation that had a single 'T' base overhang at their 3'-end. For adapter ligation 25µl of DNA ligase buffer (Illumina), 10µl adapter oligo mix (Illumina) and 5µl DNA ligase (Illumina) were added to the 10µl purified DNA sample and incubated for 15 minutes at room temperature. The DNA was purified using QIAquick PCR Purification Kit and eluted in 30µl EB. For size selection of the ligation products, gel purification was applied. Therefore a 2% agarose gel in 1X TAE was prepared. The gel was loaded with 8µl low molecular weight DNA ladder in combination with 3µl loading buffer. To load the samples 10µl loading buffer was added to the 30µl purified ligation reaction and the gel was run at 120V for 60 minutes. Depending on the NGS application the size was selected and the corresponding band was removed from the gel using a sterile scalpel and a UV transilluminator table. For transcriptome sequencing the fragments between 150 and 200bp were extracted. To purify the DNA from the gel the MinElute Gel Extraction Kit (Qiagen) was used and eluted in 30µl elution buffer. To enrich the adapter-modified DNA fragments a PCR step was performed. For PCR amplification 1µl of the purified DNA was added to 25µl Phusion DNA polymerase and 1µl of PCR primer 1.1 and 1.2 respectively. In 18 cycles the

adapter modified DNA of PCR mix was amplified. An initiating denaturing step at 98°C for 30 seconds was followed by 18 cycles of 10 seconds at 98°C, 30 seconds at 65°C and 30 seconds at 72°C. A terminal elongation step at 72°C for 5 minutes finished the amplification, after which, the PCR product was held at 4°C. The product was cleaned up using QIAquick PCR Purification Kit (Qiagen) and eluted in 50µl elution buffer. The purified DNA was quantified by NanoDrop and diluted to 10nM for cluster generation and sequencing on Illumina Genome Analyzer. For those experiments, single-end reads of 36bp in length were produced.

2.6.8 PREPARATION OF SEQUENCING LIBRARY FOR SMALLRNA SOLEXA SEQUENCING

MATERIAL:

DGE Small RNA Sample Prep Kit (# FC-102-1009)	(*Illumina*)
SRA 0.3M NaCl	(*Illumina*)
Glycogen	(*Illumina*)
RNase-free water	(*Illumina*)
SRA 5' RNA adapter	(*Illumina*)
SRA 3' RNA adapter	(*Illumina*)
10X T4 RNA ligase buffer	(*Illumina*)
RNase OUT	(*Illumina*)
T4 RNA ligase	(*Illumina*)
SRA gel loading dye	(*Illumina*)
SRA RT primer	(*Illumina*)
12.5mM dNTP mix	(*Illumina*)
25mM dNTP mix	(*Illumina*)
Primer GX1	(*Illumina*)
Primer GX2	(*Illumina*)
Phusion DNA Polymerase	(*Illumina*)
Ultra-pure ethidium bromide	(*Roth*)
100% ethanol	(*VWR*)
100% ethanol (-20°C)	(*VWR*)
75% ethanol	(*VWR*)
70% ethanol	(*VWR*)
SuperScript II reverse transcriptase	(*Invitrogen*)
100nM DTT	(*Illumina*)
5X first strand buffer	(*Invitrogen*)
3M NaOAc, pH 5.2	(*Merck*)
6X DNA loading dye	(*Promega*)
15% TBE-urea PAGE gel, 1mm, 10 well	(*SequaGel*)
10% TBE-urea PAGE gel, 1mm, 10 well	(*SequaGel*)

6% TBE PAGE gel, 1mm, 10 well		
5X TBE buffer	1.1M Tris	(*Merck*)
	900mM Borate	(*Merck*)
	25mM EDTA (pH 8.3)	(*Roth*)
	in ddH_2O	

PROTOCOL:

For smallRNA sequencing from human colon tissue and cell lines, total RNA was isolated according to the protocols described in section **2.6.2** and **2.6.3.** The smallRNA preparation was performed by Illumina's DGE SmallRNA Sample Prep Kit following the manufacture's instructions. In general this protocol encompasses the following steps: smallRNA isolation, cDNA library preparation and sequencing. To isolate the smallRNA 10µg total RNA were made up to a final volume of 10µl and combined with the same amount of SRA gel loading dye. The mixture was heated up to 65°C for 5 minutes before it was loaded on a denaturing 15% TBE-urea polyacrylamide gel. Before running the gel at 200V for 1h it was pre-run for 15 minutes. The loaded RNA was size-fractionated by extracting a band of smallRNA fragments of 18-30 bases in length using a clean scalpel. Using a punctured 0.5 ml collection which is placed in another 2 ml collection tube the RNA was purified from the gel. Then, the gel was centrifuged at 10700xg for 2 minutes through the punctured bottom of the 0.5 ml tube into the 2 ml tube to crush the piece of gel. 300µl of SRA 0.3M NaCl was added to the gel debris in the 2ml tube, and the RNA was eluted by rotating the tube gently at room temperature for 4 h. The eluate and the gel were transferred to the top of a Spin X cellulose acetate filter and centrifuged for 2 minutes at 10700xg to separate the gel from the eluate. To precipitate the RNA 1µl glycogen and 750µl of room temperature 100% ethanol were added and incubated at -80°C for 30 minutes. Then, the sample was centrifuged immediately at 10700xg for 25 minutes at 4°C. The supernatant was removed and discarded before the pellet was washed in 750µl of 75% room temperature ethanol. After letting the pellet dry, it was resuspended in 5.7µl RNase-free water. To ligate the 5' RNA adapter to the size selected RNA, the sample was transferred to a 200µl PCR tube and 1.3µl of the SRA 5' adapter was added as well as 1µl of 10X T4 RNA ligase buffer, 1µl RNase OUT and 1µl T4 RNA ligase up to a total volume of 10µl. The ligation mix was incubated for at least 6 h or overnight at 20°C. To stop the reaction 10µl of SRA gel loading dye was added and to select the corresponding size of the generated ligation product the sample was loaded on 15%-TBE polyacrylamide urea gel and run for 1h at 200V. Before this, the gel was pre-running for 15-30 minutes at 200V and the sample was incubated at 65°C for 5 minutes. Using a clean scalpel the corresponding nucleotide band of 40-60 bases was cut out and cleaned according to the protocol already described. After elution and precipitation the RNA pellet was resuspended in 6.4µl RNase-free water. To ligate the 3' RNA adapter to the size selected purified RNA 0.6µl SRA 3' adapter, 1µl of 10X T4 RNA ligase buffer, 1µl of RNase OUT and 1µl of T4 RNA ligase were added up to final volume of 10µl. The ligation mix was incubated at 20°C for at least 6 h or overnight. The next day, the reaction was stopped by adding 10µl of SRA gel loading dye. To remove unligated adapters, the ligation products of the desired size range between 70-90 bases in length were purified from a 10% TBE-urea polyacrylamide gel as described previously. The purified RNA was resuspended in 4.5µl RNase-free water. The ligation to both adapters enabled the reverse

transcription and amplification of the smallRNA. Therefore the 5' and 3' adapter-ligated RNA was combined with 0.5µl SRA RT primer and heated up to 65°C in a thermal cycler for 10 minutes and immediately cooled down on ice. For reverse transcription reaction 2µl of 5X first strand buffer, 0.5µl of 12.5mM dNTP mix, 1µl DTT and 0.5µl RNase OUT were added to the 5µl RNA and primer mix. The sample was heated up to 48°C for 3 minutes in a thermal cycler. Afterwards 1µl of SuperScript II reverse transcriptase was added and the mixture was incubated for 1 h at 44°C. To amplify the generated single stranded cDNA 28µl ultra-pure water was added to the 10µl of cDNA as well as 10µl of 5X Phusion HF buffer, 0.5µl of Primer GX1, 0.5µl of Primer GX2, 0.5µl of 25mM dNTP mix and 0.5µl of Phusion DNA Polymerase up to a final volume of 40µl. The PCR product was amplified in a thermal cycler according to the following PCR protocol: 30 seconds at 98°C, 15 cycles of 10 seconds at 98°C, 30 seconds at 60°C and 15 seconds at 72°C followed by a single step of 10 minutes at 72°C. Then, the PCR was held at 4°C. The amplification product was separated on a 6% TBE polyacrylamide gel and a corresponding gel fragment of approximately 92 bases in length was excised. The gel slice was centrifuged through a punctured 0.5 ml tube in a new 2 ml collection tube and 100µl of prepared 1X gel elution buffer was added to the gel debris. To elute the DNA from the gel the sample was rotated for 2 h at room temperature and afterwards centrifuged through a Spin-X-filter for 2 minutes at 10700xg. Then, the DNA was precipitated by adding 1µl glycogen, 10µl of 3M NaOAc and 325µl of -20°C 100% ethanol. Immediately after adding the ice-cooled ethanol, the mixture was centrifuged at 10700xg for 20 minutes. The resulting pellet was washed in 500µl 70% ethanol and air dried. For the final resolution of the DNA pellet 10µl ultra-pure water was added. As a quality control the DNA could be checked on an Agilent Technology 2100 Bioanalyzer. Then, the purified DNA was quantified and diluted to 10nM for cluster generation and sequencing on Illumina Genome Analyzer.

2.6.9 ANALYSIS OF DEEP SEQUENCING DATA

Different analytical steps were required to analyze deep sequencing data. After sequencing, the obtained raw reads were filtered based on their attained quality. Regarding the circumstance that the average read length of Ilumina sequencer read (~36nt) is longer than the average size of a microRNA (~25nt), all reads were mapped against Illumina adapter sequences using blat. If the reads passed quality filtering and adapter filtering the adapter sequences were clipped from the reads subsequently. The clipping process started from one of the ends of a read and reads with a length less than 16 bases after clipping were omitted from analysis. All remaining clipped sequence reads were mapped to the golden path version GRCh37 (hg19) using the bwa 0.5.8 alignment tool with default parameters. Subsequently, sequencing reads were aligned towards the precursor microRNA sequences in the reference list of microRNA precursors provided by the latest version of miRBase (http://www.mirbase.org/). The output file was used to perform expression analysis of annotated microRNAs. Therefore reads on target were calculated and a read had to have at least one base within the target region to be evaluated 'on target'. Log2 ratios were normalized for each comparison within each patient by centering the median of log2 ratios to zero. The differential expression was evaluated using student's t-test on the log2 ratios (samples versus controls). To compare the obtained

NGS data for microRNA expression to the expression values determined by TaqMan technology normalized log2ratios from sequencing and TaqMan data were plotted against each other and a linear model (y = mx + b) was fitted for both comparisons: tumor versus benign and metastasis versus benign. The correlation between both experiments was calculated as the Pearson product-moment correlation for both comparisons.

2.6.10 REAL-TIME QUANTIFICATION OF MICRORNAS USING STEM-LOOP REAL-TIME PCR

MATERIAL:

TaqMan MicroRNA Reverse Transcription Kit (# 4366596)	(*Applied Biosystems*)
5X reverse transcription primer	(*Applied Biosystems*)
10X reverse transcription buffer	(*Applied Biosystems*)
dNTPs mix (25mM each)	(*Applied Biosystems*)
MultiScribe reverse transcriptase (50U/µl)	(*Applied Biosystems*)
RNase inhibitor (20U/µl)	(*Applied Biosystems*)
RNase-free water	(*Applied Biosystems*)
TaqMan Universal PCR Master Mix (# 4324018)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-1 (# 4427975, ID 000385)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-129 (# 4427975, ID 001184)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-133a (# 4427975, ID 000458)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-133b (# 4427975, ID 000591)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-135b (# 4427975, ID 002261)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-145 (# 4427975, ID 002278)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-150 (# 4427975, ID 000473)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-183 (# 4427975, ID 000484)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-215 (# 4427975, ID 000518)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-31 (# 4427975, ID 001100)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-375 (# 4427975, ID 000564)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-497 (# 4427975, ID 001043)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-493 (# 4427975, ID 001040)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-9 (# 4427975, ID 000583)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, hsa-miR-96 (# 4427975, ID 000434)	(*Applied Biosystems*)
TaqMan MicroRNA Assay, RNU44 (# 4427975, ID 001094)	(*Applied Biosystems*)

PROTOCOL:

For real-time quantification of mature microRNAs, we used TaqMan MicroRNA Assays (Applied Biosystems) and performed a two-step RT-PCR according to the manufacturer's protocol. In the reverse transcription reaction, single-stranded cDNA was synthesized from 25ng total RNA using

50nM looped RT primer, 1X Reverse Transcription buffer, 0,25mM each of dNTPs, 0,25U/µl RNase Inhibitor and 3,33U/µl MultiScribe reverse transcriptase per 15µl reaction. The reaction was incubated in a thermal cycler for 30 min at 16°C, 30 min at 42°C, followed by 5 min incubation at 85°C to inactivate the reverse transcriptase and then held at 4°C. The cDNA was diluted 10-fold before preparing for PCR reaction.

To quantify the RT-products, TaqMan minor groove binding (MGB) probes were used in a 10µl PCR reaction, containing 1µl of the 10-fold diluted cDNA, 1X TaqMan MicroRNA Assay (PCR Primer/Probe/dNTP mix) and 1X TaqMan Universal PCR master mix. The real-time PCR was performed in a 384-well plate using standard protocols on an Applied Biosystems 7900HT Fast Real-Time PCR System, including 10 min incubation at 95°C to activate the AmpliTaq Gold enzyme, followed by 40 amplification cycles of 95°C for 15 s and 60°C for 1 min. All reactions were typically run in triplicates.

2.6.11 REAL-TIME QUANTIFICATION FOR MICRORNA TARGETS

MATERIAL:

Specific gene primer – forward (20µM)	(*Metabion*)
Specific gene primer - reverse (20µM)	(*Metabion*)
SYBRGreen PCR Master Mix (# 4309155)	(*Applied Biosystems*)

PROTOCOL:

For messengerRNA quantification total RNA from previous preparations steps was applied (Section **2.6.2, 2.6.3**) and cDNA was synthesized (Section **2.6.6**) using gene specific primers. The concentration of the cDNA was quantified using NanoDrop Spectophotometer. A 10µl real-time PCR master mix was prepared with 5µl of SYBR-Green Mastermix, 1µl of a 5ng/µl cDNA template stock solution and 1.5µl of a primer mix solution consisting 20µM of specific gene primer forward and reverse. The real-time PCR was performed in a 96-well plate using standard protocols on an Applied Biosystems 7900HT Fast Real-Time PCR System, including 5 min incubation at 95°C, followed by 40 amplification cycles of 95°C for 15 s and 60°C for 1 min. All reactions were typically run in triplicates.

2.6.12 DATA ANALYSIS FOR REAL-TIME QUANTIFICATION

The relative quantification of microRNA and mRNA expression was calculated using the comparative Ct method or ΔΔCt method. Therefore, default threshold settings were applied to determinate the Ct value, which is defined as the fractional cycle number at which the fluorescence passes the fixed threshold [172]. For normalization of the Ct value the difference to an endogen reference will be used which results in ΔCt value. As an internal reference we applied *HPRT* (hypoxanthin-phosphoribosyl-transferase 1) for the mRNA real-time PCRs and *RNU6B* for the microRNA real-time PCRs.

$$\Delta Ct = Ct\ (target) - Ct\ (endogenous\ reference)$$

The normalization towards an internal control enables the minimization of mistakes caused by volume differences in the pipetting processes. The ΔCT value expresses how weak or strong the expression of a target is compared to the internal reference. To assess whether the expression of a target differs throughout an experiment (treated and non-treated samples) or in different tissues, for example, normal and tumor tissue, it is necessary to calculate the difference of the determined ΔCT values of a test sample towards a calibrator sample to receive the ΔΔCT value.

$$\Delta\Delta Ct = \Delta Ct\ (test\ sample) - \Delta Ct\ (calibrator\ sample)$$

The ΔΔCt expresses the relative difference of fractional cycle numbers for both samples. The ΔΔCt indicates the differential of expression in a test sample compared to a calibrator sample as an n-time expression. To convert this relative value in a more descriptive fold change value we used the following formula:

$$fold\ change = 2^{(-\Delta\Delta Ct)}$$

A fold change value < 1 expresses a decreased expression of the test sample and fold change value > 1 an increased expression.

2.6.13 MULTIPLEX EXON CAPTURE

MATERIAL:

10X PCR Buffer	(*Invitrogen*)
AP1 primer (100µM)	(*Metabion*)
AP2 primer (100µM)	(*Metabion*)
dNTP mix (25mM each)	(*Bioline*)
25mM $MgCl_2$	(*New England Biolabs*)
Platinum Taq (5U/µl)	(*Invitrogen*)
100nM template oligos	
100X SybrGreen	(*Applied Biosystems*)
Ultra-pure ethidium bromide	(*Roth*)
Phase lock tubes (light)	(*5 Prime*)
Phenol/chloroform	(*Applichem*)
10M ammonium acetate	(*Merck*)
GenElute-LPA (linear polyacrylamide)	(*Sigma Aldrich*)
100% ethanol	(*VWR*)
75% ethanol	(*VWR*)

Lambda DNA 10X reaction buffer	(*New England Biolabs*)
Lambda Exonuclease (5U/µl)	(*New England Biolabs*)
USER (1U/µl)	(*New England Biolabs*)
10X DpnII buffer	(*New England Biolabs*)
DpnII (5U/µl)	(*New England Biolabs*)
500pmol DpnII-guide	(*Metabion*)
1U/µl USER enzyme	(*New England Biolabs*)
Filter Spin-X column (0.2µm)	(*Costar*)
TE buffer (pH 8.0)	(*Roche*)
10X AmpLigase buffer	(*Epicentre*)
AmpLigase (5U/µl)	(*Epicentre*)
Taq Stoffel Fragment (10U/µl)	(*Applied Biosystems*)
10X ExoI Buffer	(*New England Biolabs*)
Exonuclease I (20U/µl)	(*New England Biolabs*)
Exonuclease III (100U/µl)	(*New England Biolabs*)
10X Accuprime Buffer	(*Invitrogen*)
Accuprime Pfx (2.5U/µl)	(*Invitrogen*)
CP-2-FA primer (100µM)	(*Metabion*)
CP-2-RA primer (100µM)	(*Metabion*)
CP-2-CIRC-A (10µM)	(*Metabion*)
CP-2-CIRC-NO-A (10µM)	(*Metabion*)
10mM ATP	(*New England Biolabs*)
20mM DTT	(*Invitrogen*)
T4 Polynucleotide Kinase (10U/µl)	(*New England Biolabs*)
10X Phi29 buffer	(*Epicentre*)
Phi29 polymerase (100U/µl)	(*Epicentre*)
1mM random hexamer primer (phosphorothioate protected)	(*Invitrogen*)
10% TBE-urea PAGE gel	(*SequaGel*)
Low molecular weight (LMW) DNA ladder	(*New England Biolabs*)
6% TBE PAGE gel	

5X TBE buffer	1.1M Tris	(*Merck*)
	900mM Borate	(*Merck*)
	25mM EDTA (pH 8.3)	(*Roth*)
	in ddH_2O	

PROTOCOL:

The multiplex exon capture method is a target capture and amplification method using molecular inversion probes (MIPs). As a first step, a library of 55000 100mer oligos were synthesized and released in parallel from a custom designed microarray. This led to a pool of oligos, which consisted of a common 15 nucleotide flanking sequence, and a unique 20 nucleotide segment also described as

targeting arms. They hybridize upstream and downstream of a genomic target region that wants to be captured. A universal linker sequence of 30 nucleotides is localized between the targeting arms. Using specific primers (AP1, AP2) towards the 15 nucleotide flanking sequence, the oligos could be amplified. For PCR amplification a master mix which consisted of 10µl of 10X PCR Buffer, 0.5µl of 100µM AP1 primer, 0.5µl of 100µM AP2 primer, 0.8µl of 25mM dNTP mix, 6µl of 25mM $MgCl_2$, 0.5µl of 5U/µl Platinum Taq, 1µl 100nM template oligos (MIPs) and 80.2µl water in a total volume of 100µl for one reaction was prepared. To amplify those MIPs 8-12 cycles were run on a qPCR machine under the following conditions, 5 minutes at 95°C followed by cycles of 30 seconds at 95°C, 1 minute at 58°C and 1 minute at 72°C. A qPCR machine allows one to observe the amplification and to stop the reaction before the primers are used up, which normally occurred after approximately 9 cycles. A longer amplification could lead to the formation of chimeras of the longer MIPs. After the amplification step the MIPs were purified using phenol/chloroform extraction and ethanol precipitation as described in section **2.6.6**. The purified DNA was checked on a 6% TBE polyacrylamide gel. A clean band of DNA should run at 110bp. After quantifying the DNA, it was diluted to a 100nM solution. Amplification was repeated until an amount of 100pmol was achieved. For one capture experiment an input amount of MIPs of 2.5pmol was used. The samples of several amplification steps were pooled, phenol/chloroform extracted, ethanol precipitated and resuspended in a volume of 85µl of water. In following restriction digest steps a single-stranded 70-mer oligo was generated. Therefore the 85µl DNA sample from previous step was combined with 10µl of Lambda DNA 10X Reaction Buffer and 5µl of 5U/µl Lambda Exonuclease. The lambda exonuclease digest was incubated at 37°C for 60 minutes and for heat-inactivation the reaction was held at 75°C for 10 minutes. An USER digest followed and to the 100µl DNA from the previous step 290µl of TE buffer (pH 8.0) and 10µl of 1U/µl USER enzyme were added and incubated for 120 minutes at 37°C. Afterwards the DNA was purified by phenol/chloroform extraction and ethanol precipitation, as previously described, and the DNA was resuspended in 38µl water. To the purified DNA 5µl of 10X DpnII Buffer and 5µl of 500pmol DpnII-guide were added. To hybridize the primer, the sample was heated up to 95°C for minutes and the thermal cycler was then turned off, leaving the sample in the block so that the sample could cool down slowly to 25°C. Then, 2µl of 5U/µl DpnII enzyme were added and incubated at 37°C for 120 minutes. At this point 5µl of 1U/µl USER enzyme were added and incubated at 37°C overnight for an additional restriction digestion. The sample was gel purified over a 6% TBE polyacrylamide gel. The correct band was located at 70bp. The corresponding gel slice was excised from the gel using a clean scalpel and a dark reader. To obtain the DNA, the gel was transferred to a punctured 0.5 ml collection tube which was placed in another 2 ml collection tube and centrifuged at maximum speed for 1 minute. This procedure crushed the polyacrylamide slice completely in the 2 ml tube and an approximately equal amount of TE buffer was added and incubated at 65°C for more than 60 minutes. Eluate and gel debris were transferred to 0.2µm filter Spin-X column and centrifuged at maximum speed for 1 minute. The DNA was then purified by ethanol precipitation and resuspended in 20µl water. For a final quantification of the generated 70-mer single-stranded oligos a 10% denatured TBE-Urea polyacrylamide gel was run, loaded with 1% of DNA. As a ladder, a low molecular weight marker (LMW) should be used. For the capture protocol 3µg of genomic DNA and 150X Molecular Inversion Probes was used. The hybridization of the MIPs to specific regions of the genomic DNA was

performed with 1µl of 2.5pmol MIPs combined with 2µl of 10X AmpLigase Buffer and the corresponding volume for 3µg gDNA. The reaction was brought up to final volume of 20µl by adding water. Before the sample was hybridized for more than 40 h at 60°C it was heated up to 95°C for 5 minutes to denature the gDNA. For probe capture and enzymatically gap filling a master mix was generated with 0.25µl of 2mM dNTPs, 2.5µl of 10X AmpLigase Buffer, 5µl of 10U/µl Taq Stoffel Fragment, 12.5µl of 5U/µl AmpLigase and 4.75µl water for one capture reaction. From this mixture 1µl was added to the capture sample and incubated for more than 10 h at 60°C. The probe capture reaction was followed by an exonuclease digestion. Therefore, 2µl of 10X ExoI Buffer, 2µl of 20U/µl Exonuclease I and 2µl of 100U/µl Exonuclease III were added to the 21µl sample from the previous step. The reaction was incubated at 37°C for 60 minutes and heat-inactivated at 80°C for 15 minutes. After exonuclease digestion of linear material the desired products were PCR amplified. Therefore, 5µl of 10X Accuprime Buffer, 0.15µl of 100µM CP-2-FA primer, 0.15µl of 100µM CP-2-RA primer, 0.4µl of 25mM dNTPs, 0.15µl of 100X SybrGreen, 0.4µl of 2.5U/µl Accuprime Pfx and 41.75µl of water were added to 2µl of template from the MIP capture reaction. The PCR was run on a real-time PCR machine under the following conditions, 5 minutes at 95°C, 8-12 cycles of 95°C for 30 seconds, 1 minute at 58°C, 1 minute at 72°C and a single step at 72°C for 5 minutes. The most important thing to consider was not to stop amplification reaction before primers run out. To estimate the success of capture experiment it was essential to run a no-template control, no-probe control and a water control beside the sample. The amplified products were purified by phenol/chloroform extraction and ethanol precipitation before they were resuspended in an appropriate volume to load the whole sample on a 6% TBE polyacrylamide gel. The corresponding DNA band from 150bp to 250bp was cut out using a clean scalpel and DNA was extracted as described before using 0.2µm filter Spin-X column and resuspended in 20µl of TE. In the following step the linear DNA products were prepared for rolling circle amplification. For circularization reaction, 9µl of the linear DNA from previous step were combined with 0.5µl of 10µM splint CP-2-CIRC-A oligo, 0.5µl of 10µM splint CP-2-CIRC-NO-A oligo, 2µl of 10X AmpLigase Buffer, 2µl of 10mM ATP, 1µl of 20mM DTT, 1µl of 5U/µl AmpLigase, 0.5µl of 10U/µl T4 Polynucleotide Kinase and 3.5µl water and incubated for 10 cycles of 30 minutes at 37°C, 30 seconds at 95°C and 10 minutes at 60°C in a thermal cycler. To digest linear material, 2µl of 20U/µl Exonuclease I and 2µl of 100U/µl Exonuclease III were added to the mixture and incubated for 60 minutes at 37°C and heat-inactivated at 80°C for 15 minutes. In the following step, the DNA was amplified by hyperbranched rolling circle amplification (hRCA). For this reaction 5µl of 10X Phi29 Buffer, 2µl of 25mM each dNTP mix, 0.2µl of 100X SybrGreen, 2.5µl of 1mM random hexamer primer (phosphorothioate protected – NNNN*N*N), 0.25µl of 100U/µl Phi29 Polymerase and 30µl water were added to 10µl of the DNA from the previous step and amplified overnight using a real-time machine at 30°C. For heat-inactivation the reaction was incubated at 65°C for 10 minutes. When the parallel negative control started to register after approximately 5 h of amplification the real-time was stopped.

2.6.14 MICROARRAY-BASED ENRICHMENT

MATERIAL:

NimbleGen Sequence Capture Hybridization Kit (# 05853257001)		(*NimbleGen*)
2X SC Hybridization Buffer		(*NimbleGen*)
SC Hybridization Component A		(*NimbleGen*)
NimbleGen Sequence Wash and Elution Kit (# 05853257001)		(*NimbleGen*)
SC wash buffer I		(*NimbleGen*)
SC wash buffer II		(*NimbleGen*)
SC wash buffer III		(*NimbleGen*)
Stringent wash buffer		(*NimbleGen*)
Elution reagent		(*NimbleGen*)
Nebulization buffer		(*Roche*)
TE buffer		(*Roche*)
Oligo Annealing Buffer (OAB)	100µl of 1M-HCl (pH 7.8)	(*Sigma Aldrich*)
	20µl of 0.5M EDTA (pH 8.0)	(*Roth*)
	100µl of 5M NaCl	(*Merck*)
	9.78 ml ddH_2O	
gSel3 oligo		(*Metabion*)
gSel4-Pi oligo		(*Metabion*)
4% 1X TAE agarose gel		
10X NEB buffer 2		(*New England Biolabs*)
100X BSA		(*New England Biolabs*)
100mM ATP (ribonucleotide)		(*Roche*)
0.5M EDTA, (pH 8.0)		(*Roth*)
1M MgCl2		(*Sigma Aldrich*)
1M Tris HCl (pH 7.8)		(*Sigma Aldrich*)
Glycerol		(*Thermofisher Scientific*)
5M NaCl		(*Merck*)
70% ethanol		(*VWR*)
Tween 20		(*Applichem*)
T4 DNA Polymerase		(*New England Biolabs*)
DNA Clean and Concentrator-25 Kit (# D4005)		(*Zymo Research*)
Zymo-Spin Column		(*Zymo Research*)
Binding buffer		(*Zymo Research*)
Wash buffer		(*Zymo Research*)
T4 Polynucleotide Kinase		(*New England Biolabs*)
T4 DNA Ligase		(*New England Biolabs*)
AMPure DNA Purification Beads		(*Beckman Coulter*)

BioAnalyzer Agilent DNA 7500 Kit (# 5067-1506)	(*Agilent*)
Bioanalyzer Agilent DNA 1000 Kit (# 5067-1504)	(*Agilent*)
10X ThermoPol reaction buffer	(*New England Biolabs*)
dNTP mix (25mM each)	(*Bioline*)
Taq DNA Polymerase	(*Qiagen*)
PfuTurbo DNA Polymerase	(*Stratagene*)
SybrGreen	(*Applied Biosystems*)
Human Cot-1 DNA	(*Invitrogen*)
10bp DNA step ladder	(*Roche*)
10M Sodium Hydroxide	(*Promega*)
Acetic Acid, Glacial	(*Sigma Aldrich*)
QIAquick PCR Purification Kit (# 28106)	(*Qiagen*)
MinElute PCR Purification Kit (# 28004)	(*Qiagen*)
NSC-0237 primer	(*Metabion*)
NSC-0247 primer	(*Metabion*)
NSC-0268 primer	(*Metabion*)
NSC-0272 primer	(*Metabion*)

PROTOCOL:

The microarray-based-enrichment or microarray-based genomic selection is a method to capture specific regions of the human genome using NimbleGen Sequence Capture arrays and amplification by ligation-mediated PCR (LM-PCR). Applying the protocol described by NimbleGen, 20µg of high-quality, unamplified gDNA were required to perform a single Sequence Capture experiment. The optimal concentration of the DNA should lie between 250ng/µl and 500ng/µl. To proceed with the protocol supplied by NimbleGen, fragmentation of gDNA was an essential step. Therefore nebulization of gDNA sample was recommended. This procedure is highly reproducible method to get a uniform population of DNA fragments but it requires a great amount of input material. For nebulization 20µg of gDNA were diluted in TE buffer up to a final volume of 200µl and combined with 400µl Nebulization buffer. The mixture was transferred to the nebulizer cup and the nebulizer apparatus was put together and 3 bar of gas were directed through the nebulizer apparatus for 1 minute. The nebulizer apparatus was taped on the bench to collect the dispersed liquid. Approximately 300-400µl of the input amount could be recovered for the following cleaning up step. To purify the fragmented gDNA, the DNA Clean & Concentrator-25 Kit was applied. Therefore 2 volumes of DNA Binding Buffer were added to the recovered sample and mixed well before being transferred to a DNA Clean & Concentrator-25 column. The sample was centrifuged through the Zymo-Spin Column at ≥ 10000xg for 1 minute. The flow-through was discarded and the column washed with 200µl Wash Buffer. After centrifugation at ≥ 10000xg for 1 minute, the washing step was repeated and the fragmented DNA was eluted in 80µl TE buffer and centrifuged at full speed for 1 minute to a new 1.5 ml collection tube. To analyze the results of the fragmentation an Agilent Bioanalyzer 2100 DNA Chip was applied. The nebulized gDNA should be normally distributed between 300bp and 500bp to continue with the linker annealing process for library preparation. Therefore an Oligo Annealing Buffer (OAB) was prepared where 100µl

of 1M-HCl (pH 7.8), 20µl of 0.5M EDTA (pH 8.0), 100µl of 5M NaCl and 9.78 ml water were combined to a total volume of 10 ml OAB. The buffer was sterilized by filtration with a 0.2µ syringe filter. The amount needed for a final concentration of 4000µM for gSel3 oligo and gSel4-Pi oligo was calculated using the IDT Scitools resuspension calculator (www.idtdna.com/analyzer/Applications/resuspensioncalc/). The oligos were resolved in the appropriate amount of OAB and 5µl of the 4000µM gSel3 and the 4000µM gSel4-Pi oligo were mixed together 0.2 ml stripes. The primers had to be annealed to generate linkers according to the following thermo cycler protocol, 5 minutes at 95°C followed by a ramp cool with 0.1°C per second down to 12°C, where it was held. To create a working solution of linkers, 30µl OAB was added to the annealed primer and stored at -20°C. For quality control, the linkers were checked on a 4% 1X TAE agarose gel. Two different ligation mixes were prepared, one which contains 1µl of NEB Buffer 2 (10X), 1µl of 100mM ATP (ribonucleotide), 5µl water and 1µl ligase for one reaction and one ligation mix where the ligase was replaced with 1µl water as a negative control. To every 8µl ligation mix, 2µl of the prepared linker set were added and incubated at 25°C for 45 minutes in a thermal cycler. Every mix was transferred to the 4% 1X TAE agarose gel and evaluated how much of the linker complex will be found at 40bp and 20bp. Ideally most of the linker complex should be present at 40pb. If more than 50% of the linker was un-ligated, the linker should be re-annealed. The negative control should only show a band at 20bp. For the following library preparation, a polishing master mix was prepared which contains 12µl of 10X T4 DNA Polymerase Buffer, 9µl of water, 1µl of 100X BSA, 5µl of 25mM dNTP, 1µl of 100mM ATP (ribonucleotide), 6µl of 3U/µl T4 DNA Polymerase and 6µl of 10U/µl T4 Polynucleotide Kinase for one reaction. To the 40µl master mix, the 80µl of purified and fragmented gDNA were transferred up to total volume of 120µl. The polishing reaction was incubated in a thermal cycler according to the following protocol, 20 minutes at 12°C, 20 minutes at 25°C and 20 minutes at 75°C. At this point, the samples could be stored at 4°C overnight. For later quality control 1µl of the solution was saved to be checked on an Agilent Bioanalyzer 2100 DNA Chip. For ligation the remaining 119µl of polished DNA were combined with 8µl of the working annealed linker solution and 51µl of water, 8µl of 10X NEB Buffer 2, 4µl of 100mM ATP (ribonucleotide) and 10µl of T4 DNA Ligase were added up to a total volume of 200µl. The ligation mix was incubated for 90 minutes at 25°C. Here, too, 1 µl was saved for later quality check on an Agilent Bioanalyzer 2100 DNA Chip. To remove small non-ligated fragments from the library, AMPure Beads were applied. Therefore 2µl of 10% Tween 20 were added to the sample and mixed by pipetting. The exact volume was measured and brought up to a final volume of 300µl by adding approximately 115µl of TE buffer (pH 8.0). Then, 0.7x volume of AMPure Beads were added to the sample, vortexed and incubated at room temperature for 10 minutes. The solution was spun down for 1-2 seconds to collect beads at the bottom of the collection tube. The tube was placed in a magnetic rack (Dynal MPC-S). When the solution clarified the supernatant was transferred in a new collection tube and stored to use it as a non-captured sample control for later LM-PCR. The beads were washed twice with 500µl of 70% ethanol and dried with open lid at 37°C for 3-5 minutes. When the beads appeared as ‘cracked,’ 50µl of water were added to the tube and vortexed, before the tube was placed at the magnetic rack for approximately 30-60 seconds. The library containing supernatant was transferred to a new collection tube. To evaluate the size and quality of the generated DNA library 1µl was used to check it on an Agilent

Bioanalyzer 2100 DNA Chip. To verify the efficiency of the linker ligation reaction, a linker mediated PCR was performed to amplify the DNA sample. The amplification products were used later on to estimate the amount of enrichment attained from the capture experiment. For LM-PCR a LM-PCR master mix was prepared with 5µl of 10X ThermoPol Reaction Buffer, 0.5µl of 25mM dNTP, 1.25µl of 40µM LMPCR3 Oligo, 16.25µl water, 1µl of 5U/µl Taq DNA Polymerase and 1µl of 0.05U/µl Pfu Turbo DNA Polymerase for one reaction. To 1µl of the linker-ligated DNA 103µl of water were added. The 105µl of diluted DNA were divided into 4 samples for 4 PCR reactions, each with a 25µl. To the each sample 25µl of LM-PCR master mix were added. The samples were amplified using the following thermal cycler protocol, 2 minutes at 95°C followed by 30 cycles of 1 minute at 95°C, 1 minute at 60°C and 2 minutes at 72°C. After 30 cycles the samples were incubated for additional 5 minutes at 72°C and held on 4°C. The amplified samples were cleaned up using DNA Clean & Concentrator-25 Kit and according to the protocol already described in this section. Therefore 2 reactions for each sample were combined. Every sample was eluted in 25µl TE buffer and the eluate of 2 samples was combined to a 50µl sample. Afterwards the quantity of the samples was determined using NanoDrop technology. A concentration of more than 125ng/µl was required to continue with microarray hybridization. For the hybridization procedure, the NimbleGen Hybridization System should set to 42°C for at least 3 h before starting the hybridization. For a 2.1M array 300µl of 1mg/ml Cot-1 DNA was added to prepared sample and completely dried in a SpeedVac on high heat at 60°C. To rehydrate the sample 11.2µl water were added, vortexed for 30 seconds and heated to 70°C for 10 minutes to fully solubilize the DNA. Then, 18.5µl of 2X SC Hybridization Buffer and 7.3µl of SC Hybridization Component A were added. The sample was vortexed for 30 seconds, denatured at 95°C for 10 minutes, centrifuged down at full speed for 30 second and immediately stored at 42°C until hybridization. Before hybridization, the NimbleGen Mixer had to be prepared while inserting it to the Precision Mixer Alignment Tool (PMAT) following the manufacturer instructions. The 2.1M array was also placed in the according position of the PMAT and the PMAT was closed to combine mixer and slide in the right position with each other. Afterwards the PMAT was opened carefully and the combined mixer-slide assembly was removed. It was placed in a 42°C heat block chamber of the hybridization System and if possible the complete sample was pipetted into the fill port of the mixer with due regard that no bubbles appeared between mixer and slide. The fill port was closed by a port seal and the Hybridization System was closed to initiate the mixing procedure by turning on the mixing panel and setting the mixing mode to B. Hybridization was performed for at least 64 h (not more than 72 h) at a constant temperature of 42°C. After hybridization, the slide-mixer complex was taken from the hybridization chamber and washed and eluted according to the following protocol. Therefore, the slide which is still combined with the mixer was placed in disassembly basin containing 100ml of SC wash buffer II heated to 42°C and equilibrated for 10 seconds before the mixer was peeled off from the slide. Then, the slide was washed in a tube containing fresh 42°C SC wash buffer II. The tube was inverted 10 times and afterwards the slide was placed in a tube containing 32 ml of 47.5°C Stringent wash buffer. The tube was inverted 10 times and the slide was incubated for 5 minutes at 47.5°C. The stringent wash step was repeated once and the slide was placed in tube containing 32ml of SC wash buffer I. The tube was slowly inverted for 2 minutes before the slide was washed in 32 ml of SC wash buffer II for 1 minute by inverting the washing tube for 1 minute. Then, slide was placed in tube with 32 ml SC wash

buffer III and inverted 10 times. After this washing step, the slide was prepared to elute the hybridized DNA using the NimbleGen Elution System which was heated up previously to a constant temperature of 95°C. The slide was placed within the 95°C Elution System a disposable and previously assembled Elution Chamber EL1 was placed on the top of the slide and locked by an Elution System Retaining Ring. Then, 2 times 900µl of 95°C elution reagent was pipetted into the bottom filling port of the elution chamber. The slide was incubated with the elution reagent for approximately 5 minutes at 95°C. After incubation, the eluted DNA was transferred by pipetting into two 1.5 ml collection tubes each containing 900µl eluted DNA. The eluted DNA samples were completely dried in a SpeedVac set to 60°C and rehydrated with water to a combined volume of 320µl for the LM-PCR on captured samples. The sample was vortexed, incubated for 10-12 minutes at 70°C, again vortexed and centrifuged at full speed for 3 seconds. For LM-PCR, the master mix already described in this section was prepared and an 25µl aliquot of the captured DNA was combined with 25µl of the master mix before the reactions were amplified according to the LM-PCR thermal cycler protocol already mentioned. During the following step the samples were cleaned up by DNA Clean & Concentrator-25 Kit with slightly modifications to the already described procedure. For this purification, 3 reactions for each sample were combined in one 1.5 ml tube resulting in 4 tubes for every capture experiment. Every purified sample was eluted in 25µl TE buffer and the eluted DNA sample separated into 4 tubes was combined into one 1.5 ml collection tube resulting in 100µl volume. The sample concentration was measured by NanoDrop technology. The expected concentration of the eluted DNA should be more than 200ng/µl to ensure a successful performed enrichment experiment. To estimate the relative fold-enrichment, a quantitative real-time PCR on LM-PCR amplified samples was performed. Therefore a SYBR Green master mix was generated for each of the 4 NimbleGen Sequence Capture (NSC-0237, NSC-0247, NSC-0268, NSC-0272) control locus assay. For a single NSC assay 5.9µl water were combined with 0.3µl of 2µM NSC assay forward primer, 0.3µl of 2µM NSC assay reverse primer, 7.5µl of SYBR Green master mix (2X) and with 1µl of a 5ng/µl template solution for non-captured sample, captured sample, positive control (genomic DNA) and no-template-control (NTC). Using the 7900HT Fast Real-Time PCR system the following cycling conditions were applied for qPCR, a pre-incubation step at 95°C for 5 minutes was followed by 40 cycles of amplification with 10 seconds at 95°C, 1 minute at 60°C. To analyze the generated qPCR data we first determined the average Ct-value for all replicate reactions and calculated the fold-enrichment for a NSC control locus by raising the PCR Efficiency (E) for the specific assay to the power of the delta-Ct value measured for the corresponding control locus, or $E^{\text{delta-Ct}}$, whereas delta-Ct is the difference between the average Ct-value of the non-captured sample to the average Ct-value of the captured sample. The efficiency for every NSC assays was determined by applying a linear regression analysis to amplification data from template standard dilution series. Therefore the slope of the standard curve is used to calculate the efficiency with $E = 10^{(-1/\text{slope})}$. The calculated qPCR efficiencies for the NSC assays are 1.84 for NSC-0237, 1.80 for NSC-0247, 1.78 for NSC-0268 and 1.93 for NSC-0272. After determining the fold-enrichment, the captured and amplified DNA can be used directly for 454 Library preparation and sequencing or equivalent for sequencing with Illumina Genome Analyzer.

2.6.15 SECTIONING OF PARAFFIN EMBEDDED TISSUE

MATERIAL:

RNase-free water (DEPC treated)

DEPC (*Sigma Aldrich*)

PROTOCOL:

Colon tissues for *in-situ* hybridization (ISH) were obtained from surgical operation and paraffin embedded. For paraffin tissue sectioning, a manual and semi-automated rotary microtome was used. Fresh deionized water was used for the water bath of the microtome and warmed to 40°C. The paraffin tissue block was placed on ice before it was inserted into the microtome and trimmed to an appropriate surface to get full sections of the tissue. Sections of 6µM were cut and the wax ribbon was transferred to the water bath. Using a fine paint brush the relaxed and wrinkle free wax band was floated to an adhesive microscope glass slide (SuperFrost Plus slides) and they were dried on 37°C plate for 1-2 and stored afterwards at 4°C in an RNase-free environment.

2.6.16 *IN-SITU* HYBRIDIZATION

MATERIAL:

Xylene		(*Roth*)
100% ethanol		(*VWR*)
96% ethanol		(*VWR*)
70% ethanol		(*VWR*)
1X PBS	4.3 mM Na_2HPO_4	(*Merck*)
	1.47 mM KH_2PO_4	(*Merck*)
	2.7 mM KCl	(*Merck*)
	137 mM NaCl (pH 7.4)	(*Merck*)
	in RNase-free water and adjust to a final pH of 7.4.	
2X Proteinase-K buffer	100 mM Tris-Cl (pH 7.5)	(*Merck*)
	200 mM NaCl	(*Merck*)
	2 mM EDTA	(*Roth*)
	1% SDS	(*Roth*)
	in RNase-free water	
Proteinase K (15µg/ml)		(*Applichem*)
Hybridization buffer		(*Roche*)
5'-DIG-labeled LNA probes		(*Exiqon*)

Formamide		(*Roth*)
5X SSC	0.75M NaCl,	(*Merck*)
	0.075M sodium citrate	(*Merck*)
	in RNase-free water	
1X SSC		
0.2X SSC		
Blocking buffer		(*Roche*)
PBST	4.3 mM Na_2HPO_4	(*Merck*)
	1.47 mM KH_2PO_4	(*Merck*)
	2.7 mM KCl	(*Merck*)
	137 mM NaCl (pH 7.4)	(*Merck*)
	0.1% Tween-20	(*Applichem*)
	in RNase-free water and adjust to a final pH of 7.4.	
NBT 4-Nitro blue tetrazolium chloride solution		(*Roche*)
BCIP solution		(*Roche*)
Anti-digoxigenin-alkaline phosphatase labeled antibody		(*Roche*)
AP buffer	100mM Tris	(*Merck*)
	100mM NaCl	(*Merck*)
	in RNase-free water	
AP substrate	135µl NBT 4-Nitro blue tetrazolium chloride	
	105µl BCIP	
	in AP buffer	
KTBT buffer	50mM Tris-HCL	(*Merck*)
	150mM NaCl	(*Merck*)
	10mM KCL	(*Merck*)
	in RNase-free water	
Fluoromount-G medium		(*SouthernBiotech*)

PROTOCOL:

Detection of microRNAs by ISH in colon FFPE sections was performed using 5'-DIG-labeled LNA probes. As a positive control we used a single-DIG-labeled probe against snRNA U6. The 6µm sections were placed in a Coplin jar and deparaffinized in Xylene, 3 times for 5 minutes and afterwards hydrated through serial dilutions of ethanol to PBS. Therefore the slides were immersed 2 times in

100% ethanol followed by 5 minutes in 100%, 96% and 70% ethanol before they were placed for 2-5 minutes in Coplin jar with 1X PBS. Then, the slides were treated with proteinase K (15µg/ml) diluted in proteinase-K buffer for 10 minutes at 37°C. Afterwards the slides were washed in 1X PBS and dehydrated in 70% ethanol for 1 minute, 96% ethanol for 1 minute and in 100% ethanol for 1 minute. The slides were air dried for 15 minutes and the colon tissue section was framed with a hydrophobic barrier using DAKO-Pen. For hybridization we denatured the 5'-DIG-labeled LNA probe at 90°C for 4 minutes and applied 50nM of the probe diluted in Roche hybridization buffer to the colon tissue section and covered tissue sections with plastic coverslips. We hybridized in a humid chamber (5X SSC, 50% formamide) for 2h at the estimated hybridization temperature of 20°C below the calculated Tm of the individual LNA probe. After hybridization, the slides were rinsed in 5X SSC at room temperature, washed were washed in SSC solutions at increasing stringency, 5 minutes in 5X SSC at hybridization temperature, twice in 1X SSC and twice in 0.2X SSC at hybridization temperature followed by a single wash step in 0.2X SSC and 1X PBS for 1 minute at room temperature. The sections were blocked for 15 minutes in blocking buffer and incubated with anti-digoxigenin-alkaline phosphatase labeled antibody at 1:8000 dilution in blocking solution for 1h at room temperature. After washing, the slides 3 times in PBST for 3 minutes were incubated with AP substrate diluted in AP buffer overnight at 30°C and incubated in KTBT buffer to stop the reaction. Finally, slides were mounted in Fluoromount-G medium.

2.6.17 PREPARATION OF 3′-END DIG LABELED OLIGONUCLEOTIDE DETECTION PROBES

MATERIAL:

DIG Oligonucleotide 3′-End Labeling Kit (# 03353575910)	(*Roche*)
5X reaction buffer	(*Roche*)
$CoCl_2$ solution	(*Roche*)
DIG-ddUTP solution	(*Roche*)
Terminal transferase	(*Roche*)
0.2M EDTA (pH 8.0)	(*Roth*)
Anti-digoxigenin-alkaline phosphatase labeled antibody	(*Roche*)
DIG Wash and Block Buffer Set (#11585762001)	(*Roche*)
Blocking buffer	(*Roche*)
Washing buffer	(*Roche*)
Detection buffer	(*Roche*)
CDP-Star, ready-to-use Kit (#12041677001)	(*Roche*)

PROTOCOL:

For the nonradioactive DIG oligonucleotide 3′-end labeling the DIG Oligonucleotide 3′-End Labeling Kit, 2nd Generation was used. This protocol was optimized for a standard labeling reaction of up to 100 pmol (1µg of a 30-mer) oligonucleotide. For the labeling reaction 100pmol oligonucleotide were

brought to a final volume of 10µl. For the control reaction, 5µl control oligonucleotide was added to 5µl sterile double distilled water. The reaction vials were placed on ice and completed with 4µl of a 5X reaction buffer, 4µl $CoCl_2$ solution to a final concentration of 5mM, 1µl DIG-ddUTP solution to a final concentration of 0.05mM and with 1µl terminal transferase to a final concentration 20U/µl. The solution was mixed, centrifuged briefly, incubated at 37°C for 15 minutes, and then cooled on ice. To stop the reaction, 2µl of 0.2M EDTA (pH 8.0) was added. To determine the yielded DIG-labeling efficiency, a series of dilutions of the DIG-labeled oligonucleotide was generated. The dilution series had final concentrations of 100fmol/µl, 30fmol/µl, 10fmol/µl, 3fmol/µl and 1fmol/µl as well as a negative control. A similar dilution series was performed from a supplied control oligonucleotide. To a positively charged nylon membrane, 1µl of every solution was spotted. The nucleic acid was fixed to the membrane by cross-linking with UV-light. For the chemiluminescent detection procedure the CDP-Star, ready-to-use Kit was used. Before detection, the membrane was incubated with an anti-digoxigenin antibody conjugated with alkaline phosphatase (AP). Then, the antibody was diluted in blocking buffer in a 1:10000 dilution and the spotted membrane was incubated with 10ml of this solution for 30 minutes at room temperature. After incubating the antibody, the membrane was washed in several washing steps using the DIG Wash and Block Buffer Set. The spotted membrane was washed 2 times for 15 minutes in 100 ml washing buffer. To equilibrate the membrane it was immersed in 20 ml detection buffer for 2-5 minutes before being incubated with 1 ml CDP-Star working solution in a development folder. Therefore CDP-Star stock solution was diluted 1:100 in detection buffer. The luminescence was exposed to an X-ray film for 15-25 minutes.

2.6.18 NORTHERN BLOT ANALYSES FOR UNKNOWN MICRORNAS

MATERIAL:

Ultra-pure ethidium bromide		(*Roth*)
15% denaturing polyacrylamide gel		(*SequaGel*)
DIG Easy Hyb System		(*Roche*)
Hybridization buffer		(*Roche*)
DIG-labeled DNA probe		
5X SSC	0.75M NaCl	(*Merck*)
	0.075M sodium citrate	(*Merck*)
	in RNase-free water	
Washing buffer	2X SSC	
	0.1% SDS	(*Roth*)
	in RNase-free water	

Stringent washing buffer	0.5X SSC 0.1% SDS in RNase-free water	(*Roth*)

DIG Wash and Block Buffer Set (#11585762001)	(*Roche*)
Blocking buffer	(*Roche*)
Washing buffer	(*Roche*)
Detection buffer	(*Roche*)
Anti-digoxigenin-alkaline phosphatase labeled antibody	(*Roche*)
CDP-Star, ready-to-use Kit (#12041677001)	(*Roche*)

PROTOCOL:

Northern blot analysis for potential novel microRNAs was performed on total RNA of cell lines. In this case, 20µg of total RNA were extracted from colon cancer cell lines SW480 and SW620 according to the described protocol of section **2.6.2.** Twenty micrograms of the total RNA were heated at 95°C for 3 minutes and loaded onto a 15% denaturing polyacrylamide gel. The gels were stained with ethidium bromide to verify the amount and quality of the RNA before transferring to a plus-positively charged nylon membrane. The RNA was blotted for 4h at 4°C. When the RNA was transferred to the membrane, it was immobilized through UV cross linkage and stored at 4°C for later hybridization procedures. To hybridize the DIG-labeled probe to the blotted RNA the DIG Easy Hyb System was used. Then, the hybridization buffer was pre-warmed to the appropriate hybridization temperature of 45°C. To equilibrate the membrane, it was immersed for 30 minutes in the preheated hybridization buffer. Three hundred nanograms of the DIG-labeled DNA probe were diluted in 3 ml hybridization solution and carefully mixed to avoid air bubble formation. Before incubating to the membrane, the DNA probe mixture was denatured for 10 minutes at 95°C. After removing the pre-hybridization solution from the membrane, it was immediately incubated with the prepared hybridization mixture for approximately 6-10h at 45°C. For post hybridization washes, the membrane was washed 2 times for 5 minutes in washing buffer containing 2X SSC and 0.1% SDS at room temperature followed by a two stringent washing steps with 0.5X SSC and 0.1% SDS for 5 minutes at hybridization temperature. After stringency washes, the membrane was rinsed for approximately 5 minutes in washing buffer supplied by the DIG Wash and Block Buffer Set at room temperature. The following steps of blocking and detection were performed at room temperature. So the membrane was blocked for 30 minutes in 100 ml blocking buffer and afterwards incubated for 30 minutes in 20 ml of antibody solution with an Antibody concentration of 1:10000. As an antibody, the Anti-Digoxigenin-AP antibody was used. After incubating the antibody the membrane was washed 2 times for 15 minutes in 100 ml washing buffer and then equilibrated in 20 ml detection buffer for approximately 5 minutes before the detection reaction was initiated by applying 1ml CDP-Star working solution to the membrane. Next, CDP-Star stock solution was diluted 1:100 in detection buffer. For luminescence exposure, the incubated membrane was applied to an X-Ray film in development folder. The developing time could vary between for 15-25 minutes.

2.6.19 RESTRICTION ENZYME DIGEST

MATERIAL:

5-10 units of restriction enzyme	(*New England Biolabs*)
10X reaction buffer	(*New England Biolabs*)
10X BSA if necessary	(*New England Biolabs*)

PROTOCOL:

The enzymatic technique of restriction digest is characterized by the use of restriction endonucleases. Those restriction enzymes can be divided into three classes. All of them are able to cleave DNA molecules but only class II is the group that is most interesting for molecular biology approaches. The members of this group are able to recognize and cleave the DNA at specific and exact definable nucleotide sequences. Two main applications were used. For an analytical restriction digest the following protocol was used as a standard: 2.5µg of plasmid DNA were combined with 5-10 units of the corresponding restriction enzyme, 2µl of 10X reaction buffer, and if necessary, 2µl of 10X BSA. The reaction was brought up to a final volume of 20µl by adding water. It was incubated at 37°C overnight. For a double or multiple digest it was essential to coordinate the activity of all enzymes in the applied reaction buffer. An analytical restriction digest was applied to verify the correctness of the plasmid. For vector generation a preparative restriction digest was used. Therefore up to 20µg of the corresponding plasmid DNA were combined with approximately 20 units of each restriction enzyme, 5µl of 10X reaction buffer and if necessary with 5µl of BSA. The reaction was made up to final volume of 50µl by adding the acquired volume of water and it was incubated overnight at 37°C. The restriction digest was essential for later ligation steps to prepare the required vector.

2.6.20 LIGATION

MATERIAL:

10X ligase buffer	(*New England Biolabs*)
T4 DNA ligase	(*New England Biolabs*)

PROTOCOL:

For the process of DNA ligation ATP is essential to generate phosphodiester bonds between the 3' hydroxyl ends of a nucleotide with the 5' phosphatase end of another nucleotide. The ligation reaction was performed by the enzyme T4 DNA ligase. For a ligation reaction with cohesive ends of vector and DNA fragments, 1µl of T4 DNA ligase and 2µl of 10X ligase buffer were combined with DNA fragments and 5 times more vector DNA in a 20µl reaction volume and incubated at 16°C overnight.

2.6.21 PLASMID CONSTRUCTS GENERATED FOR GFP-REPORTER ASSAYS

To generate microRNA target gene constructs for GFP-reporter assays the complete 3'UTRs of the *miR-1* target genes *PTPLAD1* (protein tyrosine phosphatase-like A domain containing 1), *VEGFA* (vascular endothelial growth factor A), *PDGFA* (platelet-derived growth factor alpha polypeptide) and *HIF1A* (hypoxia inducible factor 1, alpha subunit) were directly amplified from human genomic DNA. For molecular cloning of 3'UTR target genes, we used pEGFP-C1 vector which contains a multiple clowning site with restrictions sites for different enzymes, such as XhoI and HindIII (Figure **2.1**). The vector pdsRED2-N1 was used as a marker for transfection efficiencies *in vitro* (Figure **2.2**). To amplify the 3'UTRs from *PTPLAD1*, *VEGFA*, *PDGFA* and *HIF1A* we used primers bearing restriction sites for XhoI and EcoRI or HindIII (*PTPLAD1*: XhoI and HindIII, *VEGFA*: XhoI and HindIII, *PDGFA*: XhoI and EcoRI, *HIF1A*: XhoI and HindIII) listed in Table **2.1**. After restriction digest of the generated amplicons and linearization of pEGFP-C1 vector with corresponding restriction enzymes the double stranded oligonucleotides were inserted into pEGFP-C1 using T4 DNA ligase (Section **2.6.20**). The generated plasmids were transformed into *E. coli* cells and the amplified plasmids were isolated using QIAprep Miniprep Kit (Section **2.7.2**). After checking the correctness of the plasmids by restriction enzyme digest (Section **2.6.19**) and agarose gel electrophoresis (Section **2.6.1**) they were purified (Section **2.6.1**) and co-transfected into GM2929 cells to produce high copy numbers of the correct plasmid reporter constructs which were isolated by EndoFree Plasmid Purification Kit (Section **2.7.2**). The created pEGFP-C1- reporter constructs for PTPLAD1, VEGFA, PDGFA and HIF1A were ready to be used for transfection in human cell cultures (Section **2.8.2**).

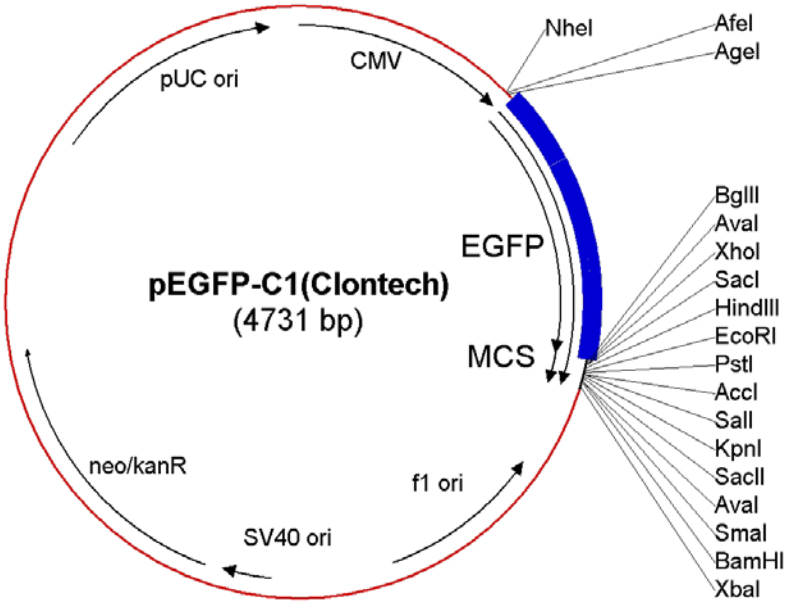

Figure 2.1: Vector map of pEGFP-C1 (Clonthech).

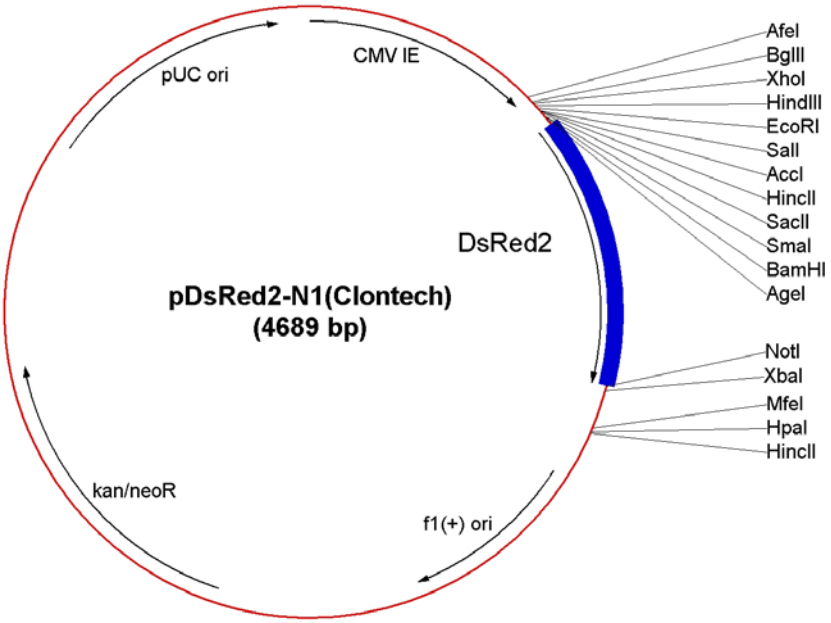

Figure 2.2: Vector map of pdsRED2-N1 (Clonthech).

2.7 MICROBIOLOGICAL METHODS

2.7.1 PREPARATION AND TRANSFORMATION OF HEAT-SHOCK COMPETENT *E. COLI*

MATERIAL:

LB (Luria-Bertani) medium	10g/l tryptone	(*Merck*)
	5g/l yeast extract	(*Merck*)
	10g/l NaCl	(*Merck*)
	in ddH_2O	
LB agarose	14g/l agar in LB medium	(*Merck*)
TB buffer	10mM $CaCl_2$	(*Merck*)
	10mM Pipes-NaOH	(*Merck*)
	15mM KCl_2	(*Merck*)
	55mM $MnCl_2$	(*Merck*)
	in ddH_2O and adjust to pH 6.7	
20mM $MgSO_4$		(*Merck*)
DMSO		(*Applichem*)

PROTOCOL:

The protocol of Inoue et al. 1990 [173] was applied to prepare heat-shock competent *E. coli* cells. Therefore a flask with 100ml of LB (Luria-Bertani) medium without antibiotics, but containing 20mM of $MgSO_4$, was inoculated with cells from a frozen stock of competent DH5α *E. coli* cells. Alternatively, it is possible to plate the desired *E. coli* cells onto a LB agar plate without selective antibiotics and incubate them at 37°C overnight, and then use a colony to inoculate the 100ml of LB plus 20mM $MgSO_4$ which is then incubated and shaken vigorously overnight at 23°C. From this starter culture, 2ml were used to inoculate a flask with 200ml of LB medium containing 20mM $MgSO_4$. This main-culture of *E. coli* cells was incubated overnight, vigorously shaking, at 23°C. When the culture was grown to OD_{600} = 0.6 the flask was put on ice for 10 minutes and the culture was split to 4 separate 50 ml falcon tubes and the cells were centrifuged down at 2500xg for 10 minutes at 4°C. The cell pellets were resuspended and combined in 2X 32ml of ice-cold TB buffer and incubated on ice for 10 minutes before they were centrifuged down, as described before. The cell pellets were resuspended and combined an entire volume of 16ml ice-cooled TB buffer that contained 7% DMSO. The cell suspension was mixed well and incubated on ice for 10 minutes. Afterwards the cells were dispensed in 100µl aliquots and immediately flash-frozen by immersion in liquid N_2. The aliquots of competent cells were stored at -80°C.

2.7.2 PLASMID PREPARATION

MATERIAL:

LB (Luria-Bertani) medium	10g/l tryptone	(*Merck*)
	5g/l yeast extract	(*Merck*)
	10g/l NaCl	(*Merck*)
	in ddH_2O	
QIAprep Miniprep Kit (# 27104)		(*Qiagen*)
Preparation buffer P1		(*Qiagen*)
Lysis buffer P2		(*Qiagen*)
Neutralization buffer N3		(*Qiagen*)
QIAprep spin column		(*Qiagen*)
Binding buffer BP		(*Qiagen*)
Wash buffer PE		(*Qiagen*)
Elution buffer EB		(*Qiagen*)
LB agarose	14g/l agar in LB medium	(*Merck*)
EndoFree Plasmid Purification Kit (# 12362)		(*Qiagen*)
Preparation buffer P1		(*Qiagen*)

Preparation buffer P2	(*Qiagen*)
Preparation buffer P3	(*Qiagen*)
QIAfilter Maxi Cartridge	(*Qiagen*)
Elution buffer ER	(*Qiagen*)
QIAGEN-tip 500	(*Qiagen*)
Equilibration buffer QBT	(*Qiagen*)
Wash buffer QC	(*Qiagen*)
Elution buffer QN	(*Qiagen*)
70% ethanol	(*VWR*)
TE buffer	(*Qiagen*)

PROTOCOL:

For plasmid preparation, the QIAprep Miniprep Kit from Qiagen was applied. In this case, 1-5ml of an overnight culture of *E. coli* in LB medium was used to purify up to 20µg of high-copy plasmid DNA. As a first step, the cells were centrifuged down and the supernatant was discarded. Then, the bacteria cell pellet was resuspended in 250µl of P1 Buffer and transferred to a 1.5ml collection tube, where 250µl of the lysis Buffer P2 were added. The sample was mixed thoroughly by inverting the tube 4-6 times and incubated at room temperature for 1 minute. To precipitate the proteins 350µl of Buffer N3 were added and mixed immediately and thoroughly by inverting the tube 4-6 times. The sample was centrifuged for 10 min at 10000xg to for a compact pellet of cell debris on the bottom of the collection tube. The supernatant was transferred to the QIAprep spin column by decanting or pipetting and centrifuged for 1 minute at 10000xg. The flow-through was discarded and the column was washed with 500µl Buffer BP and again centrifuged for 1 minute at 10000xg. After discarding the flow-through, the column was washed with 750µl Buffer PE, centrifuged for 1 minute at 10000xg and the flow-through was discarded. To remove residual washing buffer the column was centrifuged for an additional minute and the plasmid DNA was eluted with 50ml EB Buffer. The concentration of the DNA was determined by NanoDrop technology.

For downstream applications such as transfection experiments, an EndoFree Plasmid Purification Kit was applied, which allowed the clean up of ultrapure supercoiled plasmid DNA. Therefore a single colony from a selective agarose LB plate was picked to inoculate a starter culture of 2-5ml LB medium containing the corresponding selective antibiotic. The starter culture was incubated for 8 h at 37°C with vigorous shaking. The starter culture was diluted 1:750 into selective LB medium in a final volume of 100ml. The culture was incubated at 37°C for approximately 12-16 h with vigorous shaking. Then, the cells were centrifuged at 6000xg for 15 minutes at 4°C and the bacterial pellet was resuspended in 10ml P1 Buffer. To lysate the cells, 10ml of Buffer P2 were added and mixed vigorously by inverting the tube 4-6 times before it was incubated for 5 minutes at room temperature. For protein precipitation 10ml of chilled Buffer P3 were added and the tube was mixed immediately by inverting the tube 4-6 times. Afterwards the lysate was applied to the prepared QIAfilter Maxi Cartridge and incubated for 10 minutes at room temperature before the cap was removed and the plunger was inserted into the barrel of the filter cartridge. After 10 minutes of incubation, the cell lysate was filtered into a 50ml falcon tube

and 2.5ml of ER Buffer were added to the filtered lysate. The sample was mixed by inverting the tube 10 times and incubated for 30 minutes on ice. During incubation the QIAGEN-tip 500 was equilibrated using 10ml QBT Buffer to the column. When the buffer passed the column by gravity, the filtered lysate was applied to the QIAGEN-tip and was also allowed to move through the filter by gravity flow. Afterwards the column was washed two times with 30ml QC Buffer and the plasmid DNA was eluted applying 15ml QN Buffer to the QIAGEN-tip. To precipitate the DNA 10.5ml (0.7 volumes) room temperature isopropanol were added and the sample was mixed and centrifuged immediately at 15000xg for 30 minutes at 4°C. Afterwards the supernatant was carefully removed and the DNA pellet was washed with 5ml of endotoxin-free room temperature 70% ethanol and centrifuged for 10 minutes at 15000xg. After air drying for 5-10 minutes, the pellet was dissolved in an appropriate volume, such as 100-200µl TE buffer. The yield of the plasmid DNA was determined using NanoDrop technology.

2.7.3 TRANSFORMATION HEAT-SHOCK COMPETENT *E. COLI*

MATERIAL:

LB (Luria-Bertani) medium	10g/l tryptone	(*Merck*)
	5g/l yeast extract	(*Merck*)
	10g/l NaCl	(*Merck*)
	in ddH_2O	
LB agarose	14g/l agar in LB medium	(*Merck*)
Antibiotic stock solution (50g/l)		(*Sigma Aldrich*)

PROTOCOL:

For transformation of heat-shock competent *E. coli* cells they were thawed on ice for approximately 5 minutes and 50µl of cell suspension was combined with 1-5µl of plasmid DNA that showed an approximate concentration of 50-100ng/µl. Cells and DNA were incubated for 30 minutes on ice and heat-pulsed without agitation at 42°C for 30 seconds. Afterwards the heat-shock cells were transferred to ice and 800µl of LB medium without antibiotics were added before the samples were incubated at 37°C for 60 minutes. The cells were centrifuged down at 3800xg for 5 minutes and 50-150µl were plated onto a LB-agar plate with antibiotics, for example ampicillin or kanamycin in a 50µg/ml concentration, and incubated at 37°C overnight.

2.8 CELL BIOLOGICAL METHODS

2.8.1 PREPARATION AND CULTURING OF CELLS

MATERIAL:

DMEM (Dulbeco's Modified Eagle Medium)	(*Biochrom AG*)
PBS	(*Biochrom AG*)
FCS (Fetal calf serum)	(*Biochrom AG*)
Penicillin/Streptomycin	(*Biochrom AG*)
L-glutamine	(*Biochrom AG*)
Trypsin	(*Biochrom AG*)

Complete medium	DMEM	
	100ml/l FCS	
	10ml/l L-glutamine	
	10^5U/l Penicillin/Streptomycin	
Freezing medium	100ml/l DMSO	(*Applichem*)
	200mM FCS	(*Biochrom AG*)

PROTOCOL:

Cells were grown in DMEM at standard culture conditions of 5% CO_2 and 90% humidity at 37°C in air. Under optimal conditions cells were subcultured twice a week. To passage, cells were washed in prewarmed PBS and incubated for 5 minutes with 5µl/cm^2 of trypsin and removed from the cell culture flask surface. Trypsinization was stopped by resuspending the cells in 6ml of prewarmed medium and one third of the cell solution was transferred to a new T-75 flask and filled up with fresh and prewarmed complete medium to a final volume of 12ml.

For freezing of cells they were grown to confluence in a monolayer and washed with PBS before they were trypsinized. Cells were removed from the culture flask surface and 10ml of medium were added. The cell solution was transferred to a 15ml falcon tube and centrifuged at 500xg for 5 minutes at 4°C. The supernatant was discarded and the pellet was resuspended in 1ml of ice-cold freezing medium and the 1ml suspension transferred into two cryovials. The cells were frozen slowly in a styrofoam at -80°C and after two days they were stored permanently in a liquid nitrogen container.

While thawing the cells it is necessary to remove the toxic DMSO containing medium as fast as possible from the cells. Therefore cells were taken out from the liquid nitrogen container and thawed by adding stepwise 10 ml of prewarmed fresh medium containing 200ml/l FCS. The cell suspension was transferred to a 15ml falcon tube and cells were centrifuged for 5 minutes at 500xg and the cell pellet was resuspended in 5 ml prewarmed medium containing 150ml/l FCS. Then, the cells were

transferred to a T-25 flask and grown until the cells attach to the bottom of the flask. Then, the cells were split at a ratio 1:2 and slowly adapted to complete medium containing 100ml/l FCS.

2.8.2 TRANSFECTION OF CELLS

MATERIAL:

20µM microRNA miRIDIAN mimics stock solution	(*Dharmacon*)
Hiperfect transfection reagent	(*Qiagen*)
pEGFP-C1-reporter construct	
pEGFP-C1 vector	(*Clontech*)
pdsRED2-N1 vector	(*Clontech*)
Lipofectamine reagent	(*Invitrogen*)
Optimem	(*Invitrogen*)

PROTOCOL:

In general, transfection conditions were optimized for every cell line and experiment. For microRNA and corresponding mRNA quantification SW480 and SW620 cell lines were transfected with 100nM microRNA mimics using Hiperfect transfection reagent. For every quantification experiment $1.2x10^5$ cells were seeded in a 12-well plate in 1100µl complete medium 24h before transfection and incubated at standard culture conditions of 5% CO_2 in air at 37°C. For transfection 100nM of microRNA mimics were diluted in 100µl DMEM without serum and antibiotics and 6µl of Hiperfect transfection reagent were added. The solution was mixed by vortexing and incubated at room temperature for 10 minutes to allow the formation of transfection complexes. Then, the solution was added drop-wise onto the cells and incubated at standard culture conditions.

For cell viability assays SW480 and SW620 cells were seeded into 96-well plates at concentrations of 5000 cells/well in 150µl complete medium and incubated 24h at standard culture conditions. For transfection 12.5nM of microRNA mimics was diluted in 50µl DMEM without serum and antibiotics and 0.75µl of Hiperfect transfection reagent were added. The solution was incubated at room temperature for 10 minutes before it was added drop-wise onto the cells and incubated at standard culture conditions.

For wound healing experiments $1x10^6$ SW480 and SW620 cells were seeded in a 12-well plate in 1000µl complete medium and incubated 24h at standard culture conditions before transfection. For transfection 600nM microRNA mimics were diluted in 100µl DMEM without serum and antibiotics and 12µl of Hiperfect transfection reagent were added. The solution was mixed by vortexing and incubated at room temperature for 10 minutes it was added drop-wise onto the cells and incubated at standard culture conditions.

Transfection conditions for the GFP-reporter assays were optimized for HEK-293 cells. Transfection was performed in 12-well plates. Therefore $5x10^4$ HEK-293 cells were seeded in 800µl DMEM without antibiotics per well and grown at standard culture conditions for 24h. Cells were transfected with 300nM microRNA mimics and 12.5ng of each plasmid (pEGFP-C1-reporter construct, pEGFP-C1 vector, pdsRED2-N1 vector) using 2.5µl of Lipofectamine reagent in 200µl Optimem. The solution was mixed by vortexing and incubated for 25 minutes at room temperature. The transfection complex was transferred onto cells by drop-wise pipetting and incubated at standard culture conditions.

2.8.3 CELL VIABILITY ASSAY

MATERIAL:

AlamarBlue reagent	(*Invitrogen*)
1µM Camptothecin stock solution	(*Thermofisher Scientific*)

PROTOCOL:

To evaluate the viability of microRNA transfected and control cells we used the vital dye alamarBlue. Therefore colon cancer cells SW480 and SW620 were seeded into quadruplicate wells of a 96-well plate at concentration of 5000 cells/100µl at standard culture conditions and transfected after 24h with microRNA mimics and control microRNA mimics (Section **2.8.2**). For camptothecin treatment cells were treated for 24h and 48h with 0.06µM camptothecin after 24h of transfection. The alamarBlue reagent was directly added in $1/10^{th}$ volume to the cells in culture medium. The reduced alamarBlue dye was measured in various time intervals (24h, 48h, 72h) by fluorescence measurement after 4h of incubation with alamarBlue at 37°C. Therefore we applied a fluorescence excitation wavelength of 570nm and read the fluorescence emission at 585nm with a fluorescence spectrophotometer.

2.8.4 WOUND HEALING ASSAY

For wound healing experiments, SW480 and SW620 cells were seeded at $1x10^6$ in 1000µl medium in 12-well plates and cultured at standard growth conditions of 5% CO_2 in air over night at 37°C. The following day, cells were transfected with microRNA mimics using Hiperfect reagent (Section **2.8.2**). After 24h, wounds were generated through the monolayer using a 1ml pipette tip and wound closure caused by migration of cells was assessed and photographed, using an inverted microscope, 24h and 48h after wounding.

2.8.5 FACS ANALYSIS FOR GFP-REPORTER ASSAY

MATERIAL:

PBS		
FACS buffer	1X PBS	(*Gibco*)
	10% FCS	(*Hyclone*)

PROTOCOL:

For quantification and sorting of GFP containing HEK-293 cells fluorescence-activated cell sorting (FACS) analysis was applied. Therefore cells were washed in PBS and harvested 24h and 48h after transfection (Section **2.8.2**). Cells were resuspended in 200µl FACS buffer and by manual scratching with a clean pipette tip removed from the surface of the 12-well plate and transferred into test tubes. If cells are more adherent to the 12-well plate surface 50µl of trypsin could be used before resuspending in 200µl of FACS buffer. The cell suspension was analyzed and sorted at the BD FACSAria II flow cytometry cell sorter.

3 RESULTS

3.1 ANALYSIS AND IDENTIFICATION OF MICRORNAS AS BIOMARKER CANDIDATES IN COLON CANCER

3.1.1 MICRORNA EXPRESSION IN COLON CANCER PATIENTS

In order to examine the genome-wide microRNA expression in human colorectal cancer (CRC) we used next generation sequencing (NGS) technologies. Analyses using several bioinformatics algorithms characterized the expression levels of different mature microRNAs and identified a range of significantly deregulated microRNAs for the colorectal tumor tissue as well as for the liver metastasis tissue compared to the non-neoplastic tissue. Altogether, we analyzed tissues of eight human colorectal cancer patients, which had reached the tumor-node-metastasis stage IV of CRC disease. This indicates an advanced cancer stage that has metastasized to distant sites, for example liver, and which is usually incurable. Investigating matching metastases tissues delivered the possibility to search for causes that enable malignant tumors to invade surrounding tissues and to spread into different areas of the body to generate metastases.

The sample set encompassed four male and four female patients from the age of 44 until 80 years with proximal and distal localization of CRC. The set also included microsatellite stable (MSS) and instable (MSI) tumors (Table **3.1**, Table **A.1**). The selection of this quite heterogeneous sample set provided the possibility to investigate the whole range of CRC diseases.

Table 3.1: Clinical information for colon cancer patients P1-P8 investigated by next generation technology.

	Patient	***Sex***	***Age***	***Organ***	***MS-Status***
P1	Tumor	M	70	Sigmoid	MSS
	Metastases			Liver	
P2	Tumor	M	74	Coecum	MSS
	Metastases			Liver	
P3	Tumor	M	44	Colon asc.	MSI
	Metastases			Liver	
P4	Tumor	F	49	Rectum/Sigmoid	MSS
	Metastases			Liver	
P5	Tumor	F	80	Colon	MSI
	Metastases			Lymph node	
P6	Tumor	F	66	Colon asc.	MSS
	Metastases			Liver	
P7	Tumor	F	76	Colon asc.	MSS
	Metastases			Lymph node	
P8	Tumor	M	74	Coecum	MSS
	Metastases			Liver	

To profile the smallRNA fractions of 8 colon cancer patients to identify constantly up- or down-regulated microRNAs, we applied Illumina high throughput sequencing technologies and received in average 33 million reads for each sample of which 10 million could be aligned uniquely and of which more than 71% were located on microRNA regions according to miRBase v16. The obtained raw data were filtered by different parameters, such as quality and read, length. Adaptor sequences and non-human sequences were excluded and the remaining reads were aligned to known microRNA precursor sequences annotated in miRBase and used for further profiling analyses. Sequencing reads were compiled into a data table representing the smallRNA sequencing statistic for all colon cancer patients (Table **3.2**). A difference in the number of achieved unique reads between patients 1-3 and 4-8 was recognizable. This might be due to the fact that the samples were prepared with distance of time and chemicals used for library preparation and cluster generation. In addition sequencing procedures were improved during this period of time.

Table 3.2: SmallRNA sequencing statistic for eight colon cancer patients, showing the detailed number of all obtained reads, unique aligned reads and reads that could be matched to the miRBase database for the colon cancer patients P1-P8 (N = normal, T = tumor, M = metastases).

Sample ID	*All*	*Unique aligned*	*Aligned in MirBase15*	*Aligned in MirBase15 %*
P1N	8,318,722	1,595,578	1,238,277	77.61
P1T	34,210,702	2,458,732	1,216,090	49.46
P1M	29,395,966	4,207,844	2,895,932	68.82
P2N	5,525,567	1,943,571	1,569,145	80.74
P2T	7,718,045	2,273,725	1,668,554	73.38
P2M	12,799,354	2,101,331	1,322,240	62.92
P3N	14,402,298	2,787,552	2,029,651	72.81
P3T	15,656,268	2,646,769	1,556,585	58.81
P3M	15,045,131	3,197,741	2,383,248	74.53
P4N	31,916,936	12,477,233	10,464,573	83.87
P4T	22,440,569	9,242,312	7,730,374	83.64
P4M	36,746,364	13,624,244	9,690,250	71.13
P5N	35,830,076	13,831,345	11,816,231	85.43
P5T	34,976,086	11,236,863	9,180,294	81.70
P5M	35,911,669	11,139,157	8,406,136	75.46
P6N	23,205,273	3,453,283	2,203,214	63.80
P6T	32,536,119	7,188,326	4,741,135	65.96
P6M	34,242,222	13,061,905	8,944,651	68.48
P7N	36,926,851	10,016,831	6,386,987	63.76
P7T	36,724,529	11,280,167	7,606,005	67.43
P7M	34,806,226	8,901,319	4,723,212	53.06
P8N	36,146,363	12,533,493	9,495,127	75.76
P8T	38,642,318	13,543,510	10,790,075	79.67
P8M	37,993,244	12,775,038	7,531,147	58.95

Considering deep sequencing data for all colon cancer patients we detected an absolute number of 559 microRNAs, representing at least one read in each tissue (Figure **3.1**). An intersection analysis determined between normal, tumor and metastases tissue, resulted in 34 tumor specific and 29 metastases specific microRNAs compared to normal tissue, which showed 19 specific microRNAs.

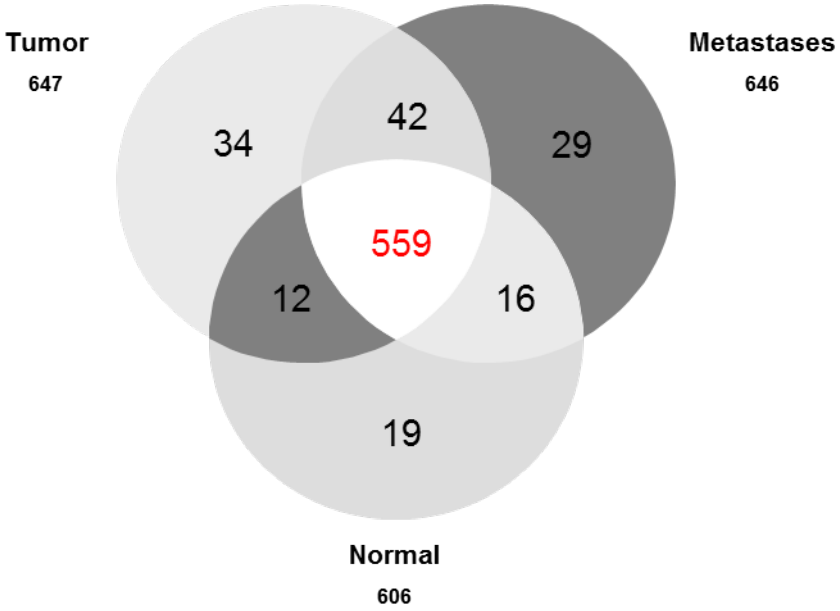

Figure 3.1: Venn diagram representing the absolute number of detected microRNAs over all tissues in CRC patients. The intersection analysis determined by NGS technology delivered a number of specific microRNAs for each tissue (Normal (N) = 19, Tumor (T) = 34, Metastases (M) = 29) and detectable over all tissues (559).

Investigating the differentially down-regulated and up-regulated fraction of microRNAs separately, we identified for the down-regulated microRNAs 13 tumor specific and 16 metastases specific microRNAs and 29 microRNAs that were found as an overlap between both malign tissues. Interestingly, a higher number of specific deregulated microRNAs were found for the metastases, this aids the assumption that the metastasis is a progression of tumor tissue. Compared to the tumor tissue we also found 8 microRNAs with a significantly decreased expression specific for the metastases. We identified one microRNA, *miR-559*, which is constantly and significantly down-regulated over the malignant tissues compared to the normal tissue (Figure **3.2**). This candidate also showed a stronger down-regulation in the metastases compared to the tumor tissue. Interestingly, *miR-559* is known to cooperatively regulate, with *miR-548d-3p*, the translation of *ERBB2* (v-erb-b2 erythroblastic leukemia viral oncogene homolog 2) [174]. Out of these lists of overlapping deregulated microRNAs we selected candidates for further analyses such as *miR-1*, *miR-135b*, *miR-129*, *miR-215*, *miR-497* and *miR-493*, which were relevant for both malign tissues. In detail, we found *miR-1*, *miR-129*, *miR-215* and *miR-497* in the fraction of the down-regulated microRNAs for the tumor/normal and metastasis/normal comparison and in the fraction of the up-regulated microRNAs we determined *miR-135b* and *miR-493* for the tumor/normal and metastasis/normal comparison.

In regard to the up-regulated fraction of differentially expressed microRNAs, we also observed a higher number of deregulated metastases specific microRNAs (T/N = 19 and M/N = 29). For the metastases to tumor comparison we identified four metastases specific microRNAs. The overlap between T/N and M/N delivered 17 deregulated microRNAs that were present in both tissues. For the total overlap analysis we could not find a candidate that was significantly differentially expressed (p-value ≤ 0.05) over all comparisons (Figure **3.3**).

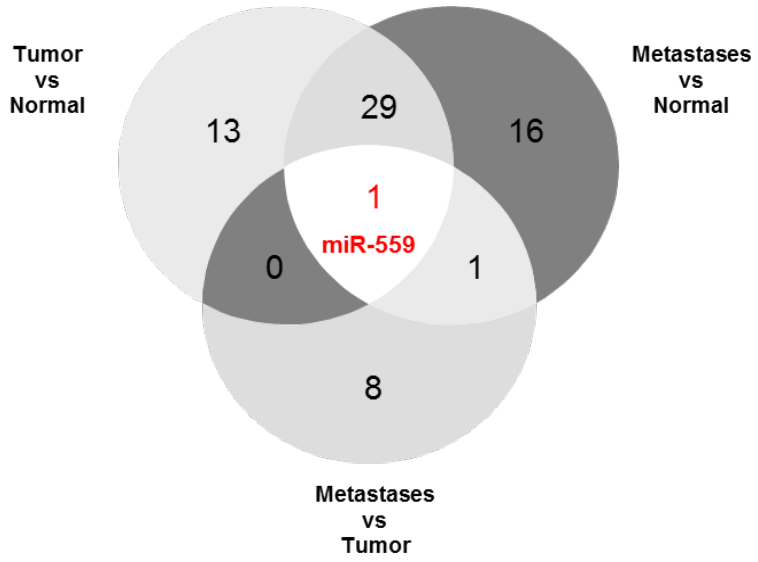

Figure 3.2: Intersection analysis for the significantly down-regulated microRNAs (p-value ≤ 0.05) for all comparisons (N/T, N/M and T/M) (t-test). T/N showed 13 specific up-regulated microRNAs and M/N 16 microRNAs. For M/T we determined 8 microRNAs. An overall comparison delivered *miR-559* as a significant down-regulated microRNA present in all tissues.

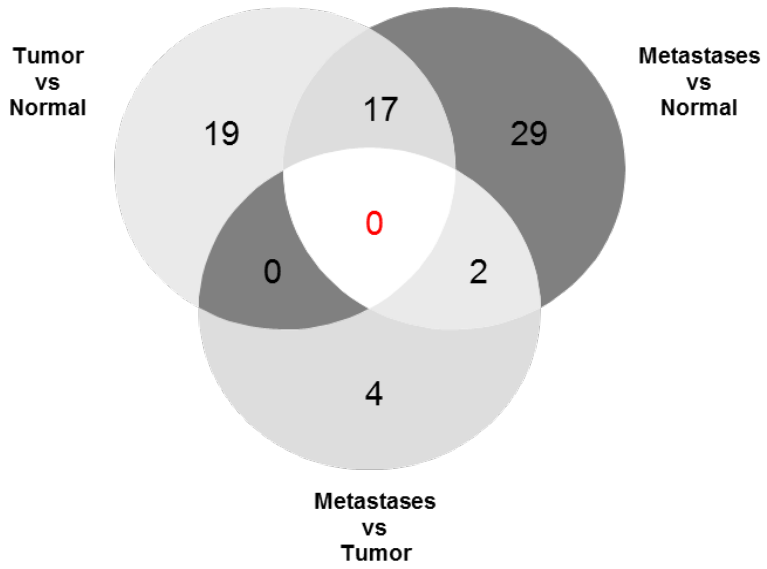

Figure 3.3: Intersection analysis for the significantly up-regulated microRNAs (p-value ≤ 0.05) for all comparisons (N/T, N/M and T/M; N = normal, T = tumor, M = metastasis) (t-test). T/N showed 19 specific up-regulated microRNAs and M/N 29 microRNAs with a significantly increased expression in the malignant tissue. For M/T we determined 4 up-regulated microRNAs in the metastases.

In general we found many significantly up- or down-regulated microRNAs in our patients for single tumor or metastasis tissues, but regarding the significance of the regulation which is determined by a p-value ≤ 0.05 and the same regulation in the three comparisons T/N, M/N and M/T we determined no or at least one candidate (*miR-559*) which comply the conditions set out (Figure **3.2**, Figure **3.3**).

Sequencing the entity of colon cancer patients we identified several significantly deregulated microRNAs in tumor and metastases tissue over all patients and determined the average regulation of the detected microRNAs, which is represented by the median log2ratio. All sequencing reads that passed the filtering procedure and aligned to precursor sequences of known microRNAs described in miRBase were used to identify significantly deregulated microRNAs in colon cancer. The significance level was defined by a p-value threshold ≤ 0.05. As a selection the top 25 microRNA candidates for tumor and metastases matched to normal tissue were displayed in plots (Figure **3.4**, Figure **3.5**, Table **A.1**, Table **A.2**).

Some of the differentially expressed microRNAs showed a consistent deregulation in malign tissues over all patients. *MiR-135b*, already known in the literature [110] showed a significant 21-fold up-regulation in colon cancer samples (p-value = 0.0001) as well as in metastasis samples where it is even 28.5-fold higher expressed (p-value = 0.001) than in normal tissues. For *miR-1-1* and *miR-1-2*, which are localized on different chromosomes, we observed a distinctive down-regulation in both malignant tissues with a more than 90-fold down-regulation in tumor and a 95-fold down-regulation for *miR-1-1* and a 93-fold down-regulation for *miR-1-2* in metastases tissue. *MiR-133*, which is clustered on the same chromosomal locus as *miR-1* was down-regulated as well. The genome locus of *miR-133a-2* which is clustered together with *miR-1-1* on chromosome 20 showed a 1.4-fold down-regulation in tumor and a 1.33-fold decreased expression in metastases, whereas the p-value for both expression values were not significant. *MiR-133a-1*, which is localized in a cluster with *miR-1-2* on chromosome 18, is not present in our NGS data set. Other interesting candidates such as *miR-129*, *miR-215*, *miR-497* or *miR-145* were determined as significantly down-regulated in tumor and metastases. Beside microRNAs that showed a considerable reduction in expression we also identified additional candidates that displayed a significantly overexpression in both malign tissues, for example, *miR-21*, *miR-31*, *miR-183* and *miR-493*. The top 25 significantly up- and down-regulated microRNAs for each malign tissues compared to the benign tissue were plotted as box plots (Figure **3.4**, Figure **3.5**).

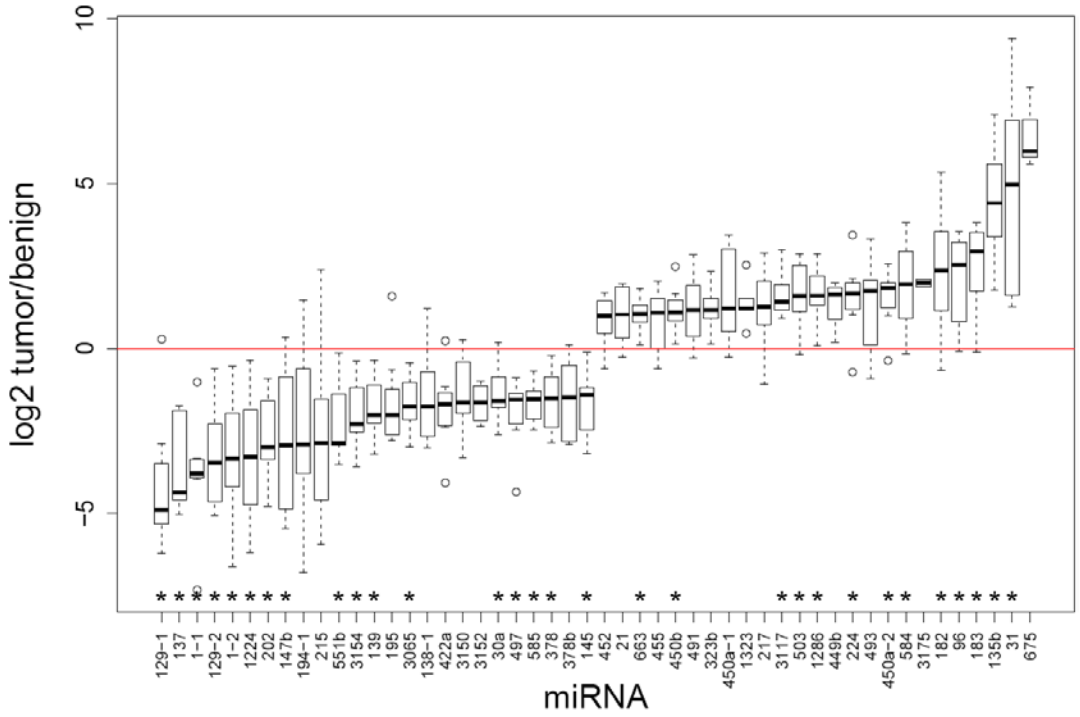

Figure 3.4: Box plot for the top 25 significantly up- and down-regulated microRNAs, comparing tumor versus normal colon samples, sorted by median log2ratio. Analysis of 16 colorectal cancer samples using NGS technology delivered a range of tumor specific deregulated microRNAs. All depicted microRNAs sufficed a p-value threshold ≤ 0.05. A star indicates samples with a p-value of below ≤ 0.01.

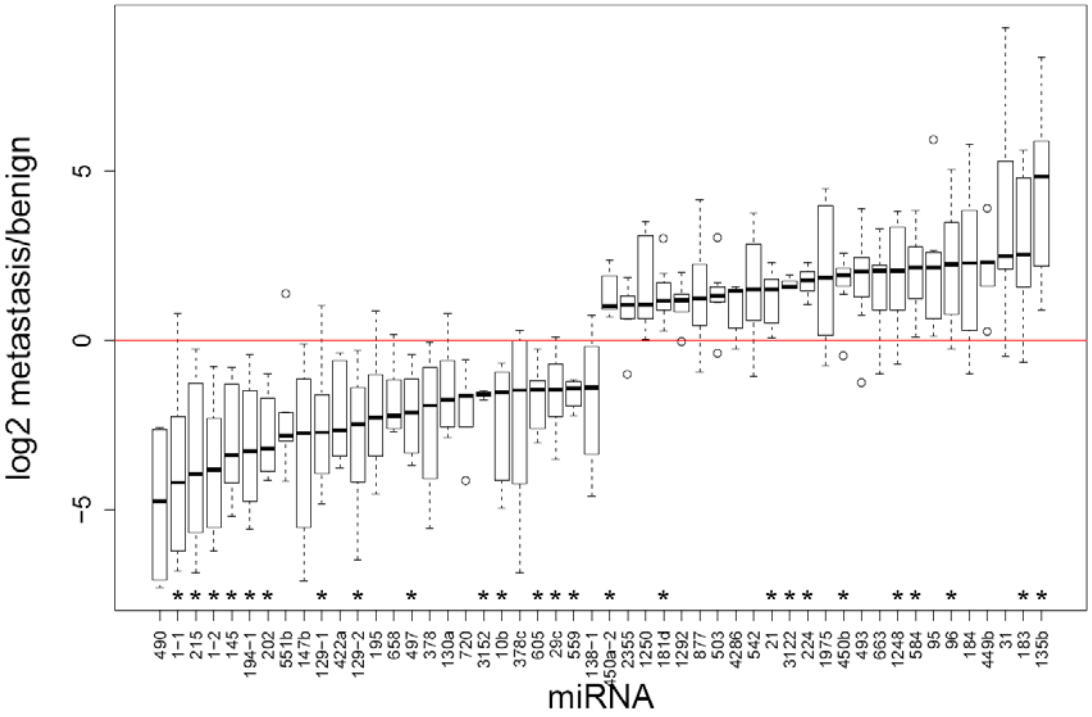

Figure 3.5: Box plot for the top 25 significantly up- and down-regulated microRNAs comparing metastasis versus normal colon samples, sorted by median log2ratio. Analysis of 16 colorectal cancer samples using NGS technology delivered a range of tumor specific deregulated microRNAs. All depicted microRNAs sufficed a p-value threshold ≤ 0.05. A star indicates samples with a p-value of below ≤ 0.01.

An altered expression can be caused by various reasons, such as epigenetic variations. Since we had access to the methylated DNA immunoprecipitation sequencing (MeDIP-Seq) data of 14 different colon cancer patients we wondered which microRNA might be de-regulated due to a differential DNA methylation event (Grimm et al., in preparation 2012). In cooperation with Grimm et al. the MeDIP protocol developed by Weber et al. 2005 [175] was applied to estimate the DNA methylation status of 14 CRCs. The sample set of 14 colon cancer patients consisted of only primary tumor and benign tissue for every patient. Searching for conformities in expression and methylation profiles we compared the data from the NGS expression profiling with the data from the MeDIP analysis and found *miR-129* to be hyper-methylated in tumor tissue for the 14 colon cancer patients (Figure **3.6**). This microRNA is positioned on two different chromosomes. MicroRNA *miR-129-1* is located on chromosome 7: 127847925-127847996 with a genome size of 72 bases. The genome localization for the precursor sequence of *miR-129-2* is located on chromosome 11:43602944-43603033 and has a genome size of 90 bases. Interestingly, a CpG-island was associated to the genome localization of *miR-129-2* with 670 bases in length (chr11:43602546-43603215). CpG-islands are genomic regions that are characterized by a cytosine base followed by a guanine base (a CpG), which is relatively rare in the vertebrate genome. In CpG-islands the CpGs are overrepresented compared to the entire genome and the cytosine bases tend to be methylated. This methylation aids DNA proofreading after duplication and influences regulation of gene expression. A methylation of microRNA regions caused by tumor progression, as we observed for *miR-129*, could lead to an inhibition of expression. For this specific microRNA we observed a corresponding down-regulation of the microRNA *miR-129* in the primary tumor of colon cancer patients, determined by NGS technology. Here we detected a 91-fold down-regulation for *miR-129-2* (p-value = 0.0005) and a 29-fold down-regulation for *mir-129-1* (p-value = 0.0027) (Table **A.1**). In addition we observed a more than 5.6-fold decreased expression (p-value = 0.005) in metastases tissue for *miR-129-2* and a 6.57-fold down-regulation for *miR-129-1* (p-value = 0.0062) (Table **A.2**). To validate those findings and to check if we could also detect this decreased expression for the actual mature microRNA sequence *miR-129*, we performed qPCR using TaqMan technology specific for mature microRNAs. Therefore we checked 13 out of the 14 patients used for the MeDIP analysis. Using the delta delta Ct method we could confirm the accepted down-regulation in the tumor tissue for 11 candidates compared to the normal colonic mucosa. Only patient PIV and PIX were outliers with an up-regulation of *miR-129*. However, for patient PIX we observed only a relatively weak hyper-methylation in the tumor for this patient, which could explain the increased expression. In conclusion, the comparison of microRNA expression profiles and methylation patterns in colon cancer patients, leads to the assumption that a hyper-methylation in tumor tissues could repress expression of microRNAs (Figure **3.6**).

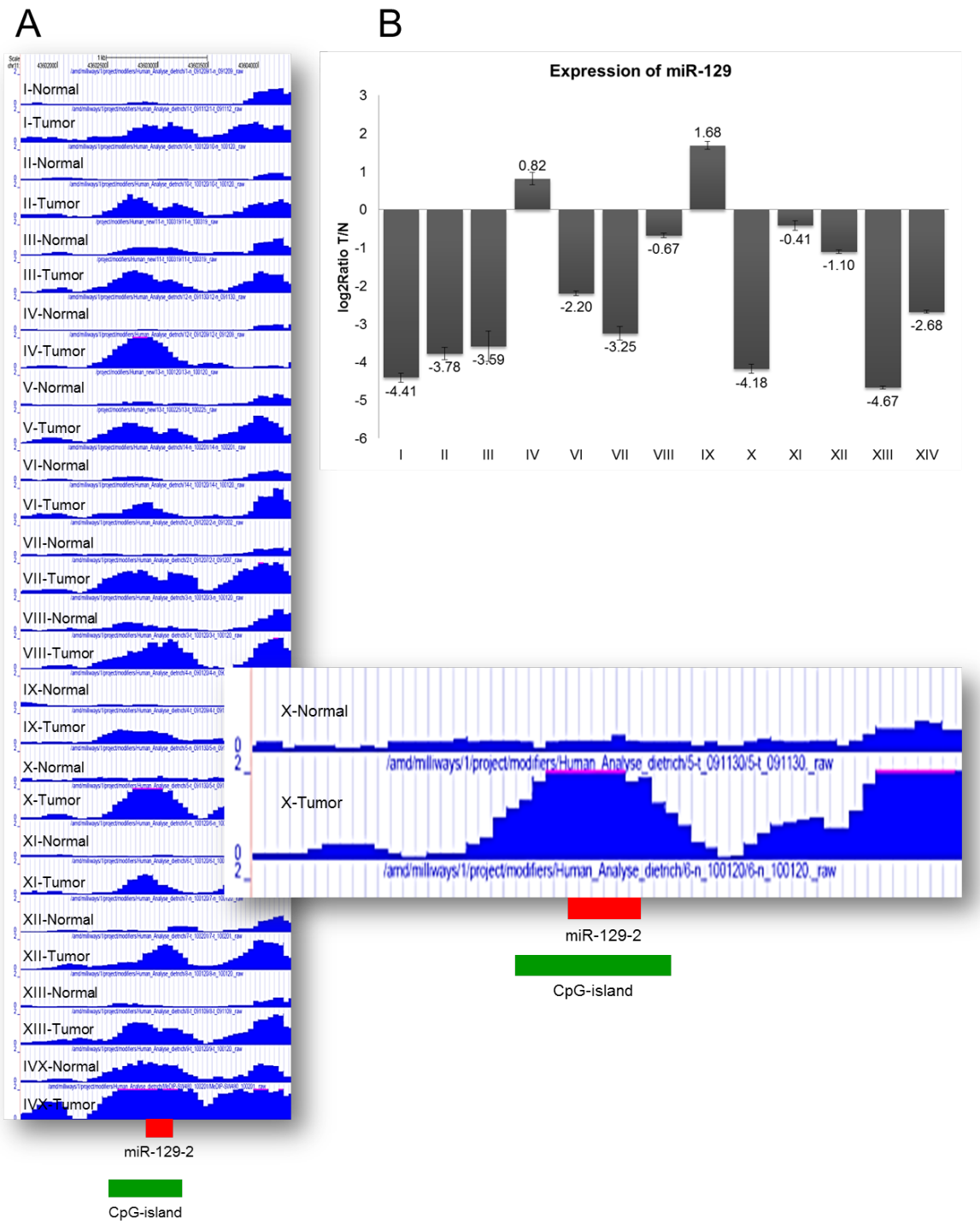

Figure 3.6: Methylation pattern and validation of *miR-129* in 14 colon cancer patients. (**A**) We observed a hyper-methylation in the primary colon tumor of 14 patients. Enlarged, the differential methylation of *miR-129* for patient X, which showed a higher grade of methylation in the tumor compared to the normal tissue. DNA-methylation patterns were determined by MeDIP-seq in 14 CRC patients (Grimm et al., in preparation 2012). (**B**) Expression level of mature *miR-129* for 13 colon cancer patients determined with real-time PCR. For 11 patients *miR-129* was determined as down-regulated. The expression values were calculated using the delta delta Ct method, using RNU44 as an internal control gene. The log2ratio of tumor versus normal was plotted for each patient.

In contrast to *miR-129*, we found a hypo-methylation in the tumor tissue for at least 7 of the 14 CRC patients in the genomic region of *miR-493* (PIII, PIV, PV, PVI, PVIII, PX, PXII) analyzed by the MeDIP protocol (Figure **3.7**). In addition, we determined the expression patterns *for miR-493* on the basis of NGS technology in our sample set of 8 colon cancer patients. As expected, we found a 3.35-fold median up-regulation in tumor with a p-value of 0.0343 for *miR-493* (Table **A.1**).

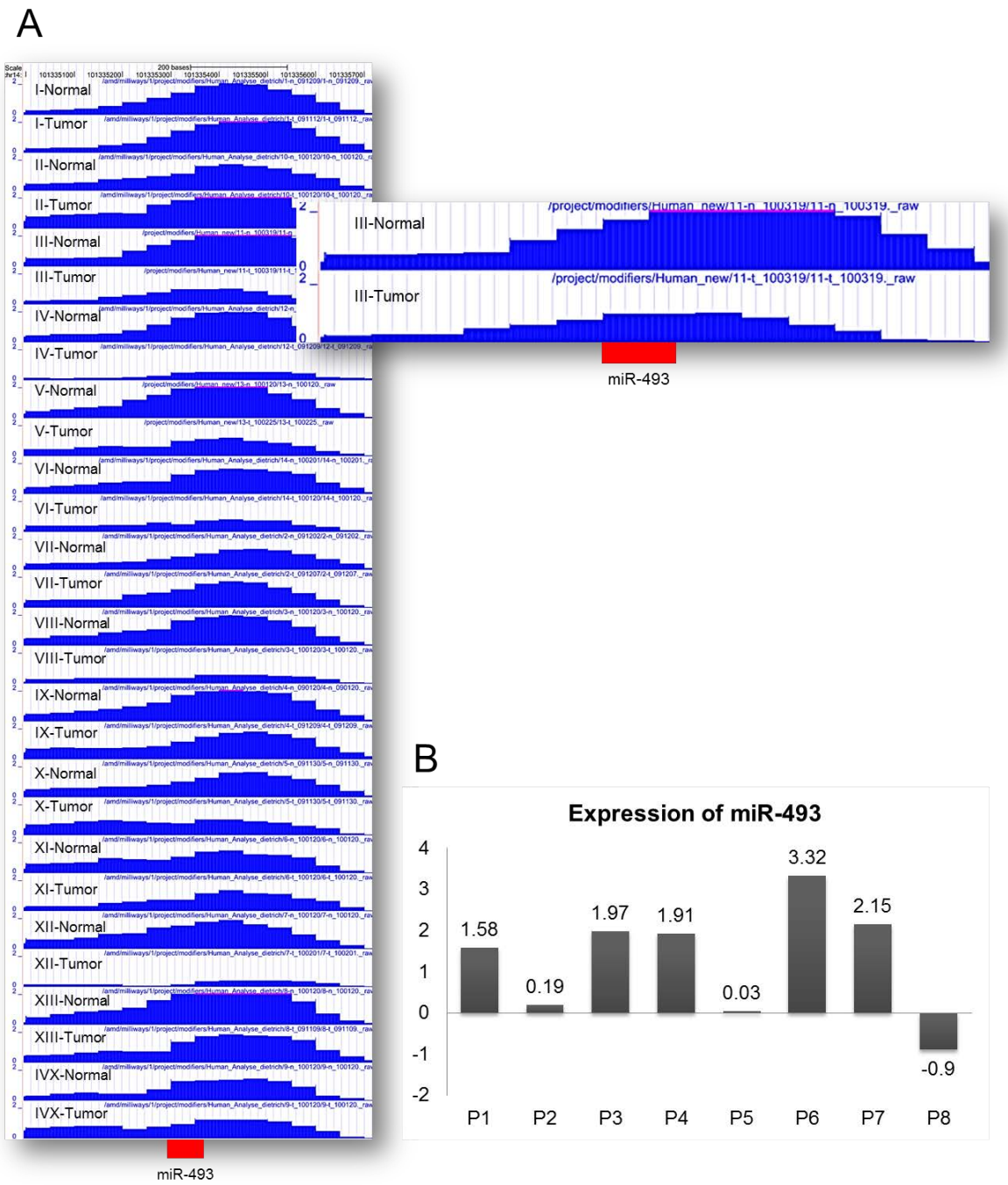

Figure 3.7: Methylation pattern and Validation of *miR-493* in 14 colon patients. (**A**) We observed a hypo-methylation in the primary colon tumor of 7 patients. Enlarged, the differential methylation of *miR-493* for patient III, which showed a lower grade of methylation in the tumor compared to the normal tissue. DNA-methylation patterns were determined by MeDIP-seq in 14 CRC patients (Grimm et al., in preparation 2012). (**B**) Expression level of *miR-493* for 8 colon cancer patients determined with next generation sequencing. For 5 patients *miR-493* was determined as up-regulated. The log2ratio of tumor versus normal was plotted for every patient.

Another microRNA, *miR-215*, showed a strong down-regulation in tumor (fold change = 7.3, p-value = 0.02) and an even lower expression in metastases (fold change = 15.43, p-value = 0.004) (Table **A.1,** Table **A.2**). *MiR-215* is known to be down-regulated in colon cancer [163] and was recently discovered by Karaayvaz et al. to act as a potential prognostic marker in stage II and III of colon cancer patients [164].

Beside microRNAs that are tumor or metastases specific relative to normal tissue, microRNAs that play a critical role for the progression of colorectal cancer into metastasis are of interest. Therefore we compared the expression of microRNAs in the tumor with those in the metastases tissue of the colon cancer patients and looked for significantly deregulated microRNAs. Only a small number of significantly deregulated microRNAs could be identified for the metastases with a p-value less or equal to 0.05. Based on significance criteria we determined six microRNA candidates that were up-regulated in the metastases, *miR-184*, *miR-204*, *miR-7-2*, *miR-653*, *miR-3195*, *let-7a-3*, with the most significantly up-regulation for *miR-184* (5-fold, log2ratio = 2.34, p-value of 0.01). For significant down-regulated microRNAs in the metastases relative to the tumor a higher number of eleven microRNAs could be determined. Beside *miR-1266* (log2ratio = -1.66, p-value = 0.006), that showed the strongest decrease in expression, ten other microRNAs showed a significant down-regulation, such as, *miR-3139*, *miR-20a*, *miR-217*, *miR-16-1*, *miR-559*, *miR-17*, *miR-32*, *miR-18b*, *miR-1278*, *miR-33b* (Figure **3.8**, Table **A.3**).

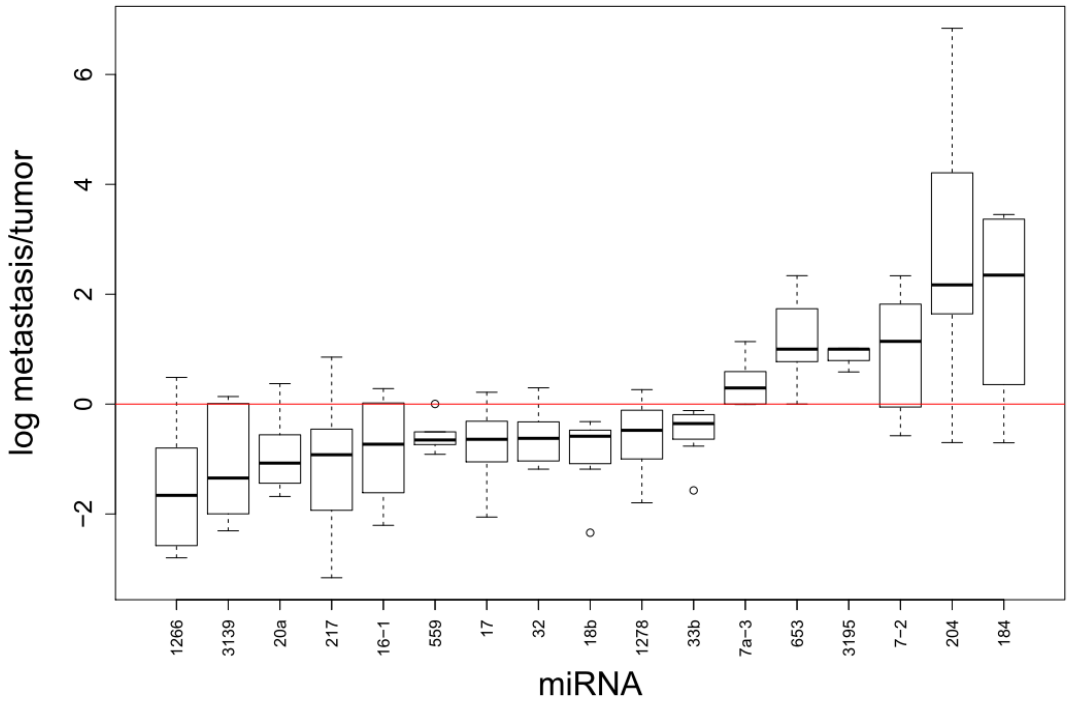

Figure 3.8: Box plot for all significant up- and down-regulated microRNA candidates for metastases versus tumor colon samples, sorted by median log2ratio (t-test).

3.1.2 MICRORNA EXPRESSION IN COLON CANCER CELLS AND COMPARISON TO PATIENT MATERIAL

For a comprehensive investigation of colon cancer and for subsequent functional assays we studied SW480 and SW620 cell lines which were derived from primary colon tumor and metastases tissue, respectively. Sequencing the eight colon cancer patients with normal, tumor and metastases tissue delivered a total number of 559 microRNAs detected with at least one read and described in the miRBase database. Considering the sequencing runs for the cell lines, in addition to the patient's material, resulted in an even higher number of 724 detected microRNAs.

Analyzing deep sequencing results for the primary colon cancer cell line SW480 and its corresponding metastasis cell line SW620, delivered a number of metastasis cell specific differentially expressed microRNAs, such as *miR-1323*, *miR-372*, *miR-371*, *miR-599* and *miR-34c*, as top 5 candidates for the down-regulated microRNAs in the metastases cell line SW620. Interestingly, *miR-559* could also be found in the metastasis cell line as a significantly deregulated microRNA. This reflects our observation for the colon cancer patients. The top 5 microRNAs that showed an increased expression level in SW620 are *miR-142*, *miR-1252*, *miR-31*, *miR-375* and *miR-466* (Figure **3.9**, Table **A.4**).

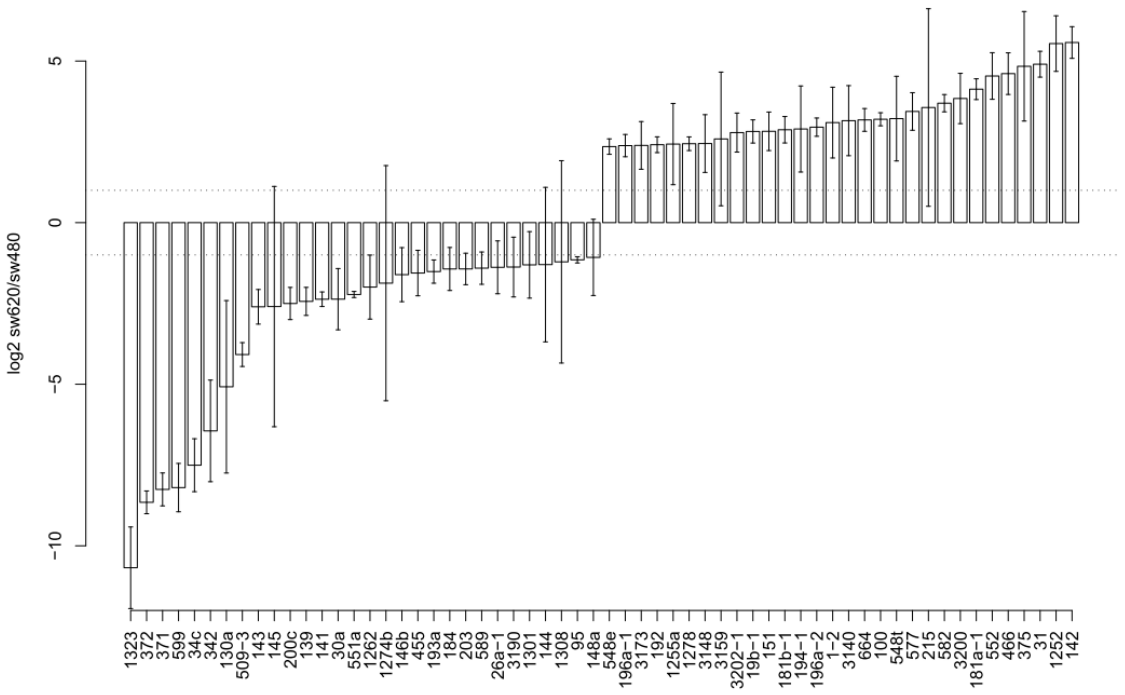

Figure 3.9: Box plot of significantly top 30 up- and down-regulated microRNAs comparing the colon cancer cell lines SW480 and its metastasis cell line SW620, sorted by median log2ratio. The standard deviation was determined by two replicates of the deep sequencing experiment for both cell lines.

Comparing the concordance of results for the colon cancer patients (M/T) and cell line material we only found *miR-3159* as significantly up-regulated in SW620 and in metastases as common feature. Interestingly, the only overlap we identified between cell lines and CRC patients, comparing all significantly down-regulated microRNAs, was *miR-559*, which was mentioned before in the context of an intersection analyses applied on the sample set of CRC patients. To estimate the reproducibility of our next generation sequencing results and, accordingly, to determine the reliability of our data, we analyzed the smallRNA fraction of SW480 and SW620 colon cancer cells in two technical replicates. The correlation analysis resulted in a strong reproducibility for the technical replicate pairs with a Pearson correlation coefficient of 0.97 for SW480 and 0.86 for SW620 (Table **3.3**). This strong equivalence between the technical replicates of the NGS experiments indicates the high reproducibility of our results.

Table 3.3: Correlation coefficients for technical replicates of SW480 and SW620 of sequencing run I and II. For the duplicate experiments of SW480 deep sequencing we determined a correlation of 0.97 and for the SW620 a correlation coefficient of 0.86.

Sample/ Sequence run	***SW480/ I***	***SW620/ I***
SW480/ II	0,97	0,52
SW620/ II	0,52	0,86

3.1.3 VALIDATION OF DEEP SEQUENCING DATA USING TAQMAN ASSAYS FOR MICRORNAS

To validate the generated NGS data, we used TaqMan assays specific for mature microRNAs. Altogether we validated 10 different microRNAs for 16 CRC patients. As a first set we verified the expression of 10 microRNAs with four of them up-regulated (*miR-31*, *miR-135b*, *miR-183*, *miR-96*) and six down-regulated (*miR-1*, *miR-145*, *miR-133a*, *miR-133b*, *miR-150*, *miR-375*) on a set of three colon cancer patients (P1-3). Using the delta delta Ct method we normalized the expression levels of every microRNA in tumor and metastases against a stable internal control gene, the small nuclear RNA *RNU44* and compared them to the expression level in the normal tissue (Figure **3.10**, Figure **3.11**). Here, we could validate the expression for nearly all microRNAs in tumor and metastases, except one outlier in the tumor tissue of P1 for *miR-96* and P2 for *miR-135b*, *miR-183* and *miR-96* in the metastases. Altogether we performed 60 assays for this sample set (3 patients, 2 malign tissues versus normal tissue, 10 microRNAs) and 54 of them delivered a confirmation of our NGS data.

For a larger patient sample set we narrowed down the number of microRNAs to *miR-1*, *miR-145*, *miR-31*, *miR-135b* and checked the expression level on 16 different CRC patients for the tumor tissue on 8 CRCs and in addition for the metastasis tissue (Figure **3.12**, Figure **3.13**). Those microRNAs were present in the top 25 of significantly deregulated microRNAs determined by deep sequencing. In total we performed 64 assays, whereas 58 assays showed a concordance to our NGS data.

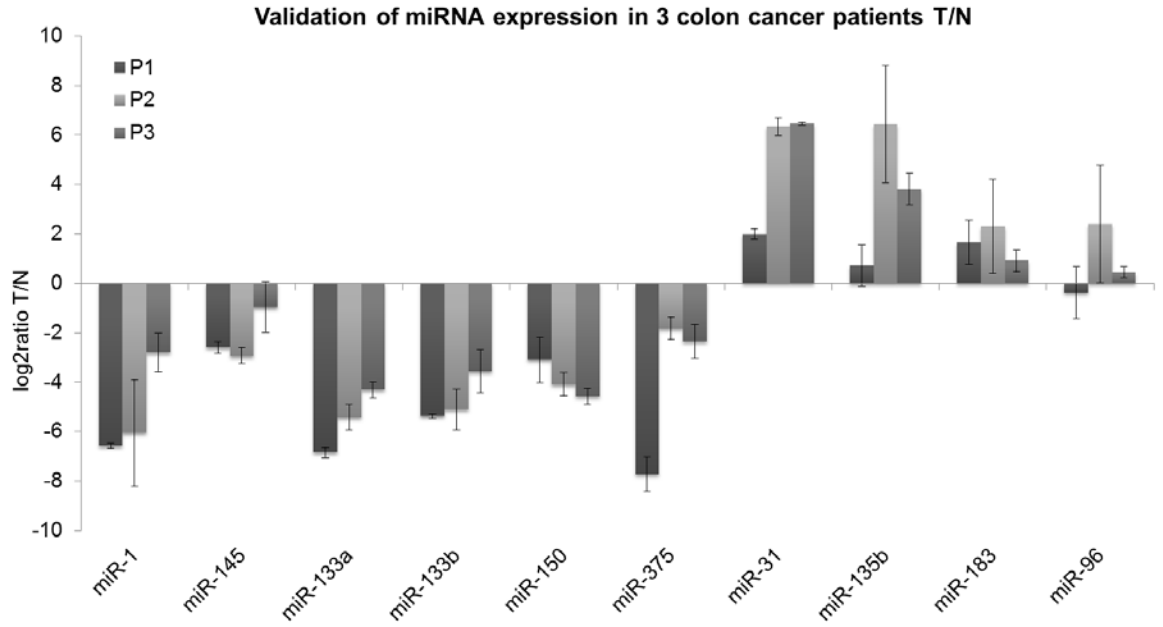

Figure 3.10: Validation of microRNA expression patterns generated by NGS technology, using TaqMan assays for mature microRNAs. Three CRC patients were analyzed on a set of 10 microRNAs (*miR-31*, *miR-135b*, *miR-183*, *miR-96*, determined as up-regulated and *miR-1*, *miR-145*, *miR-133a*, *miR-133b*, *miR-150*, *miR-375*, determined as down-regulated). The expression values were calculated using the delta delta Ct method, using *RNU44* as an internal control gene. The log2ratio of tumor versus normal were plotted for every patient (P1-P3).

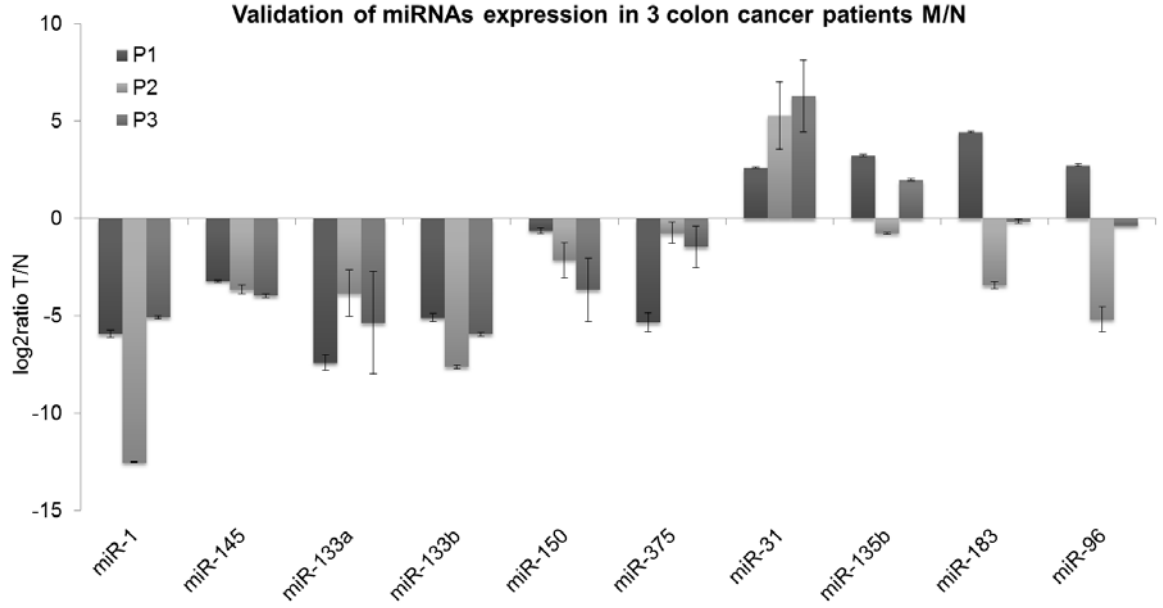

Figure 3.11: Validation of microRNA expression patterns generated by NGS technology, using TaqMan assays for mature microRNAs. Three CRC patients were analyzed on a set of 10 microRNAs (*miR-31*, *miR-135b*, *miR-183*, *miR-96*, determined as up-regulated and *miR-1*, *miR-145*, *miR-133a*, *miR-133b*, *miR-150*, *miR-375*, determined as down-regulated). The expression values were calculated using the delta delta Ct method, using *RNU44* as an internal control gene. The log2ratio of metastases versus normal were plotted for every patient (P1-P3).

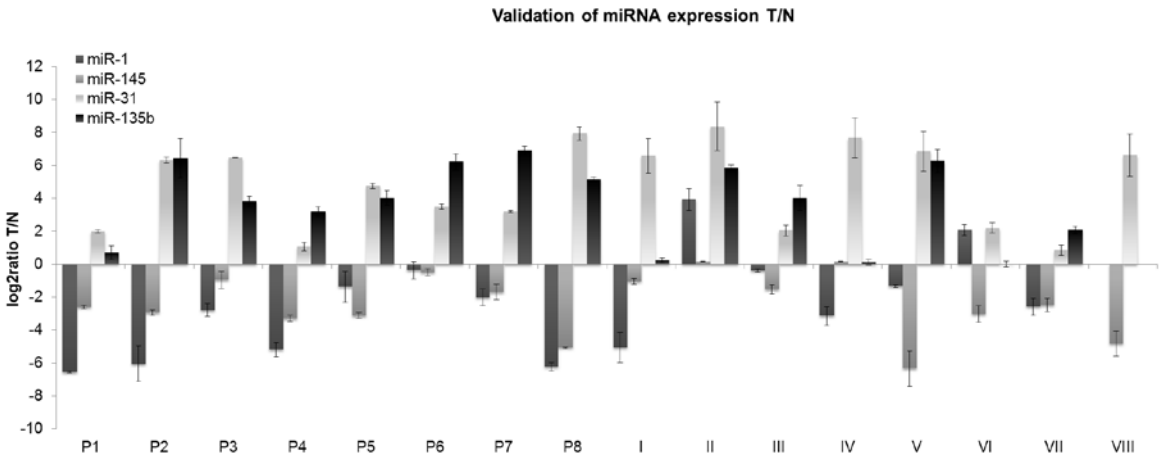

Figure 3.12: Validation of microRNA expression pattern generated by NGS technology, using TaqMan assays for mature microRNAs. For a set of 4 microRNAs (*miR-1*, *miR-145* predicted as down-regulated and *miR-135b*, *miR-31* predicted as up-regulated) we tested the expression in tumor tissue on a sample set of 16 CRC patients. The log2ratio of every patient was calculated using delta delta Ct method using normalization towards normal tissue and the small nuclear RNA *RNU44*.

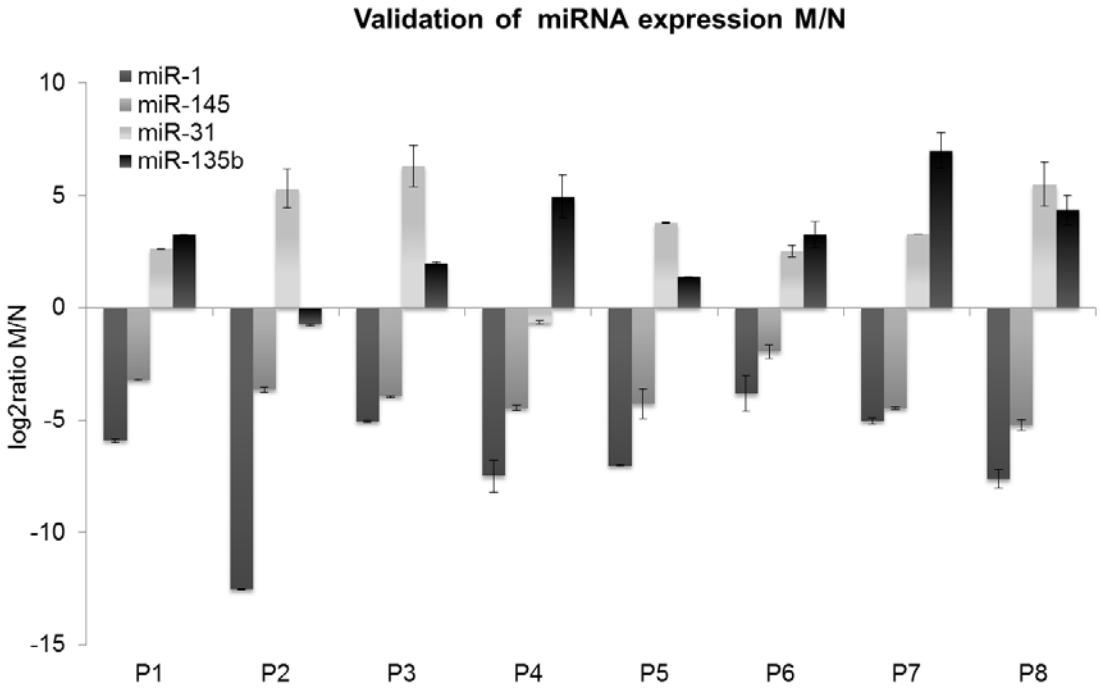

Figure 3.13: Validation of microRNA expression pattern generated by NGS technology, using TaqMan assays for mature microRNAs. For a set of 4 microRNAs (*miR-1*, *miR-145* predicted as down-regulated and *miR-135b*, *miR-31* predicted as up-regulated) we tested the expression in metastasis tissue on a sample set of 8 CRC patients. The log2ratio of every patient was calculated using delta delta Ct method using normalization towards normal tissue and the small nuclear RNA *RNU44*.

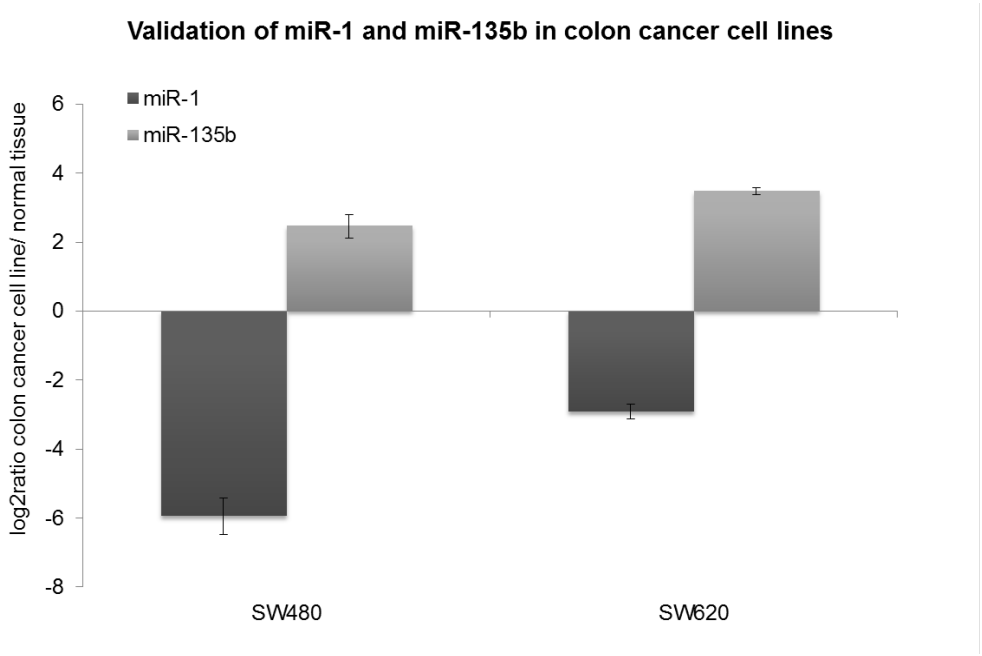

Figure 3.14: Validation of microRNA expression pattern generated by NGS technology, using TaqMan assays for mature microRNAs. We validated the expression of *miR-1* and *miR-135b* for colon cancer cell line SW480 and SW620. The log2ratio of both cell lines was calculated using the delta delta Ct method and normalization was performed against normal tissue of the median of all colon cancer patients and the small nuclear RNA *RNU44*.

In addition we analyzed the expression of *miR-1* and *miR-135b* in colon cancer cell lines SW480 and SW620 (Figure **3.14**). For normalization purposes a corresponding normal colon cell line was not available, therefore we decided to validate the obtained expression values for *miR-1* and *miR-135* to the median of expression of all normal tissues of our colon cancer patients. For both microRNAs we could validate the predicted expression in the primary colon cancer cell line (SW480) and its corresponding metastases cell line (SW620).

Taken together, the validation experiments showed a strong concordance to our sequencing data, when comparing expression values determined by NGS technology and real-time PCR. To visualize the concordance of the NGS and TaqMan data, we compared the estimated log2ratio expression values of our microRNAs on each platform and visualized the data in a scatter plot. For comparison, the data of 10 different microRNAs and 8 patients, determined by both technologies were taken into consideration for the comparison of tumor versus normal and metastases versus normal tissue. The regression analyses resulted in Pearson correlation coefficient of 0.89 for both data sets (Figure **3.15**).

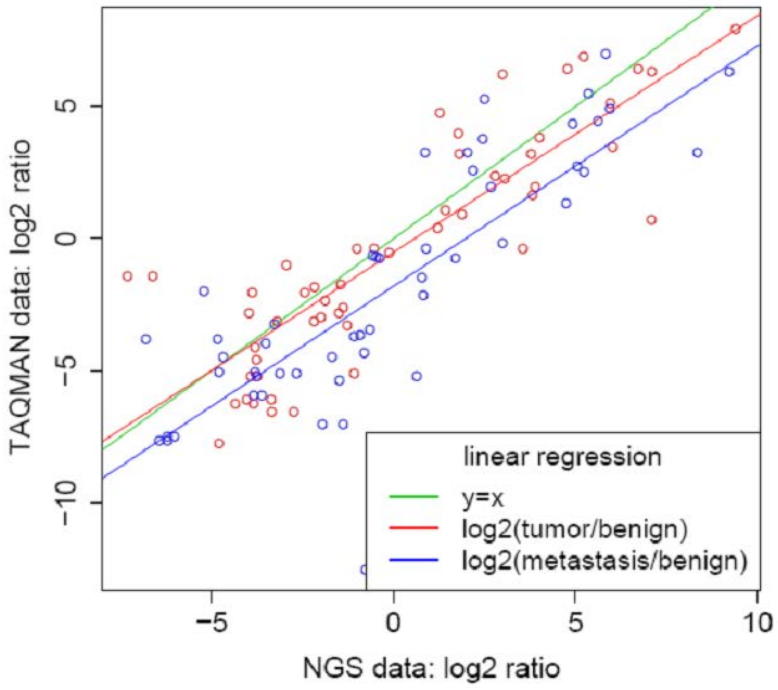

Figure 3.15: Comparison of NGS and TaqMan data. Depicted are the log2ratios for 10 different microRNAs (*miR-31*, *miR-135b*, *miR-183*, *miR-96*, *miR-1, miR-145*, *miR-133a*, *miR-133b*, *miR-150*, *miR-375*) in up to 8 patients as determined by both technologies for the comparisons tumor (red) and metastasis (blue) versus normal. The regression analyses resulted in Pearson correlation coefficients of 0.89 for both comparisons. The identity line (y=x) is depicted in green.

3.1.4 SOM (SELF-ORGANIZING MAP) –ANALYSIS

In collaboration with Hans Binder and Henry Wirth (Interdisziplinäres Zentrum für Bioinformatik (IZBI), Universität Leipzig) we arranged our multidimensional high-throughput sequencing data in SOMs (Self-Organzing-Maps). SOM mapping reduces the dimension of expression data from thousands of genes to a handful of metagenes in which each metagene acts as representative of one minicluster of co-regulated single genes [176]. This mathematical procedure offers the opportunity to classify the samples according to their similarity of microRNA expression patterns. Therefore 400 metagenes, based on different microRNA expression patterns, were arranged in a two-dimensional mosaic matrix that represents a specific profile for every sample. MicroRNAs with an increased expression level were shown as red data points and microRNAs with a decreased expression were shown as blue data points in the matrix. Green indicates a non-deregulated microRNA. The more metagenes included in this matrix the more sensitive the classification of different samples, and this could lead to precise distribution of different tissues (Figure **3.16**). Using SOM-analysis it was possible to separate normal and malign tissues by the expression profile of the 400 metagenes with two outliers of the normal tissue reaching into the area of the malign tissues. The tumor and metastases samples merged together, which indicates the similarity of those tissues based on the microRNA expression. A ‘super-SOM-plot’ represents the data for the 8 colon cancer patients and 3 tissues applying the expression profiles of the detected microRNAs (Figure **3.17**).

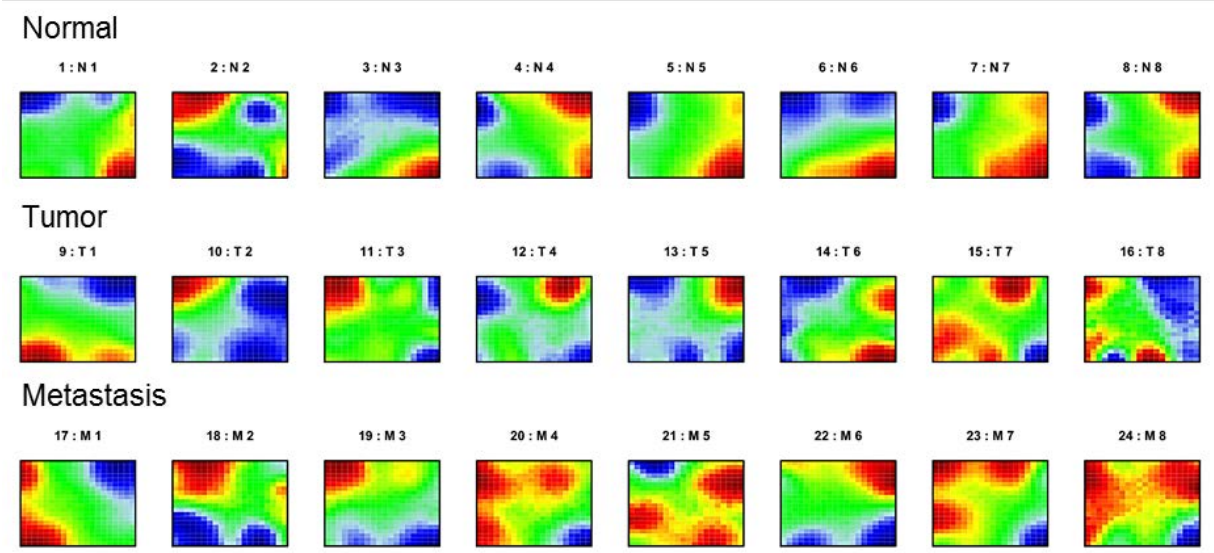

Figure 3.16: Minicluster of 400 metagenes, for every patient and every tissue (normal (N) = red, tumor (T) = green, metastases (M) = blue). The samples were classified according to their similarity of the metagene expression pattern. The expression values were color coded and arranged in a two-dimensional mosaic matrix, where red indicates a high expression, blue a reduced expression and green a non-deregulated microRNA.

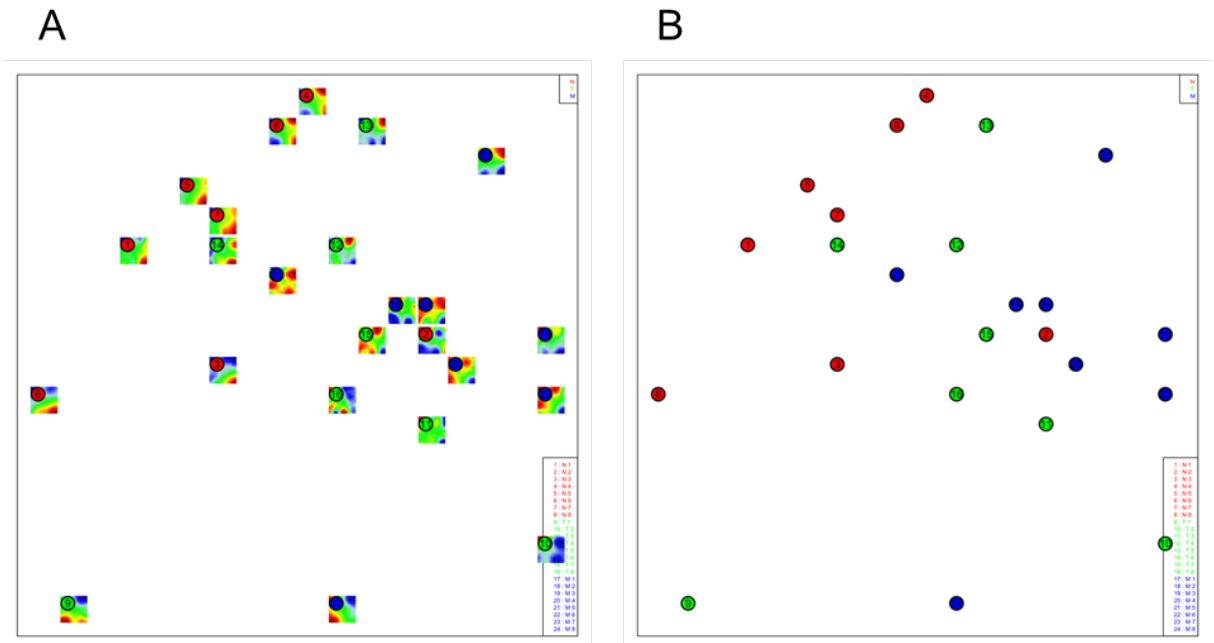

Figure 3.17: SOMs (Self-Organzing-Maps) for microRNA expression data for 8 colon cancer patients in normal ('N' = red), tumor ('T' = green) and metastases ('M' = blue) tissue. (**A**) The samples were arranged in the plot according to their expression pattern of 400 metagenes. (**B**) Simplified visualization of SOM analysis for microRNA expression values of 8 colon cancer patients in normal, tumor and metastases.

3.1.5 HIERARCHICAL CLUSTERING OF COLON CANCER PATIENTS

As an additional mathematic tool for NGS data mining and visualization we used hierarchical clustering, which enables the partitioning of a high-throughput data set in cluster of different hierarchies that merge together at certain distances, based upon their similarity relation.

To investigate whether it is possible to separate the different tissues on the basis of their microRNA expression pattern we applied hierarchical clustering analysis. We performed the analysis in three different design modes: Normal versus tumor, normal versus metastasis and tumor versus metastasis. We included all microRNAs with a p-value below 0.01 for tumor/normal and metastasis/normal comparisons (Figure **3.18**, Figure **3.19**). Regarding this criterion we incorporated 30 microRNAs as tumor specific and 25 microRNAs as metastases specific into the clustering analyses. For the comparison of both malign tissues we applied all significantly deregulated microRNAs, characterized by a p-value below 0.05 (Figure **3.20**). Presented in heatmaps, the NGS data show a distinct separation of normal and malign tissues. The separation of the malign tissues is not possible.

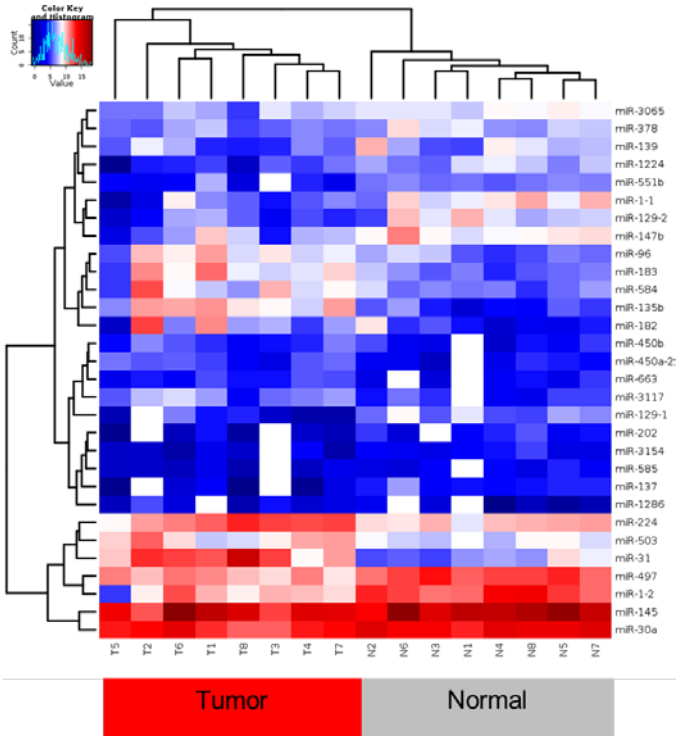

Figure 3.18: Hierarchical Clustering of normal and tumor tissues. Tissues of eight patients were clustered based on significantly deregulated microRNAs (p-value< 0.01) as provided by NGS technology. The normalized counts are color coded with red for higher values and blue for lower values.

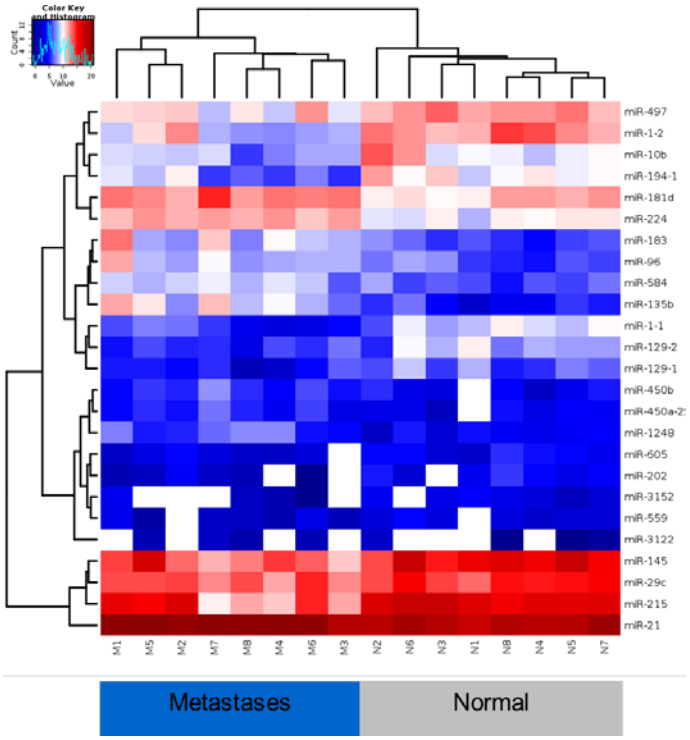

Figure 3.19: Hierarchical Clustering of normal and metastasis tissue for colon cancer patients. Tissues of eight patients were clustered based on significantly deregulated microRNAs (p-value<0.01) as provided by NGS technology. The normalized counts are color coded with red for higher values and blue for lower values.

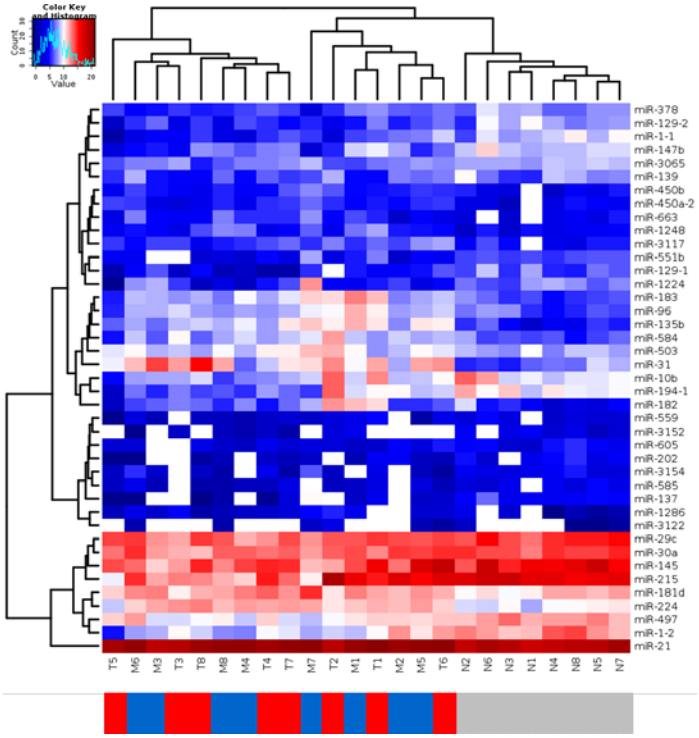

Figure 3.20: Hierarchical Clustering of normal, tumor and metastases tissue for colon cancer patients. The bar represents the tissue types with normal shown in grey, tumor in red and metastasis in blue. Tissues of eight patients were clustered based on their normalized read count for all significantly deregulated microRNAs (p-value<0.05) as provided by NGS technology. The normalized counts are color coded with red for higher values and blue for lower values.

3.1.6 PRINCIPAL COMPONENT ANALYSES (PCA) OF COLON CANCER PATIENTS

As an additional unsupervised analysis tool beside hierarchical clustering and SOM-analysis we applied a principal component analysis (PCA) and investigated the possibility to separate the different tissues based on their microRNA expression patterns. Like SOM analysis, PCA is a mathematical procedure that is used to organize and simplify comprehensive data sets. It condenses the expression profile of each sample into one single feature with the aim to establish similarity relations between the samples in terms of mutual distances in the coordinate system segmented by the principal components [176].

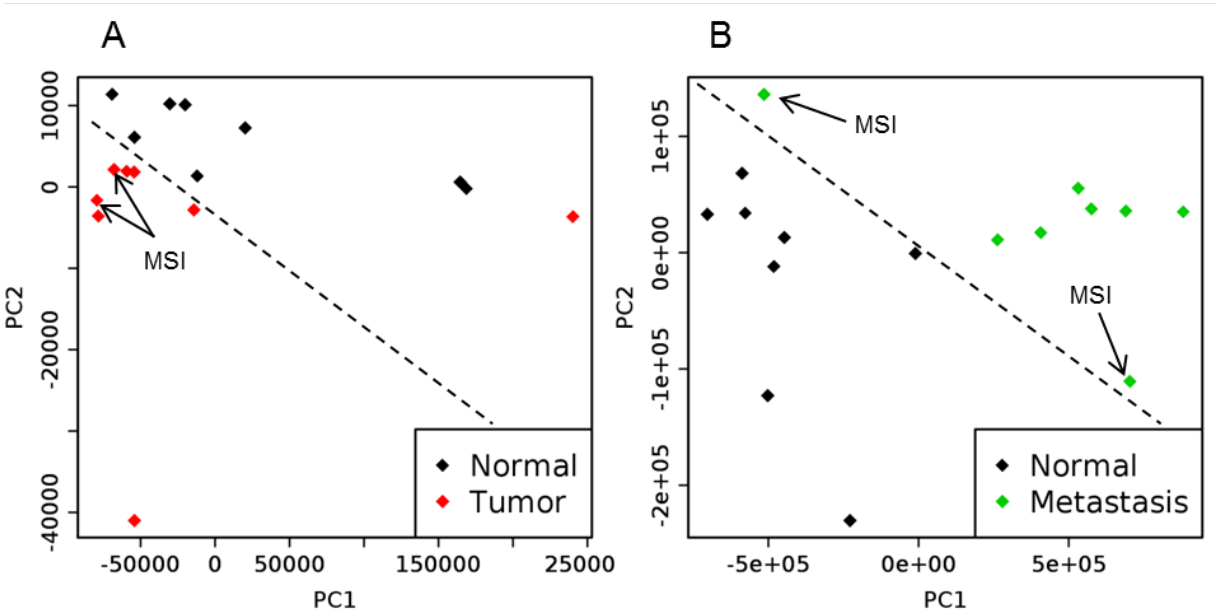

Figure 3.21: Principal Component Analysis (PCA) of tumor, metastasis and matched normal tissues of eight CRC patients using all detected microRNAs. Read counts per microRNA were normalized by the sequencing depth per tissue. (**A**) PCA plot of tumor (red) and normal (black) samples. Arrows (red dots) indicate microsatellite instable (MSI) CRC samples. (**B**) PCA plot of metastases (green) and normal (black) tissue samples. Arrows (green dots) indicate microsatellite instable CRC samples.

Using this method on the NGS data set for the eight colon cancer patients a complete partitioning of normal and malignant tissues was possible when we used all detected microRNAs. Interestingly, for the metastasis tissue we observed a separation of two patients that were characterized as microsatellite instable (MSI). For the tumor and normal comparison we could not observe a separation of the MSI samples. They were grouped together with the rest of the CRC patients that were determined as microsatellite stable (Figure **3.21**). For the combined PCA analysis of normal, tumor and metastases we detected a separation of malign and benign tissues as well. A separation of tumor and metastasis did not occur. They merge together in one area of the PCA plot (Figure **3.22**).

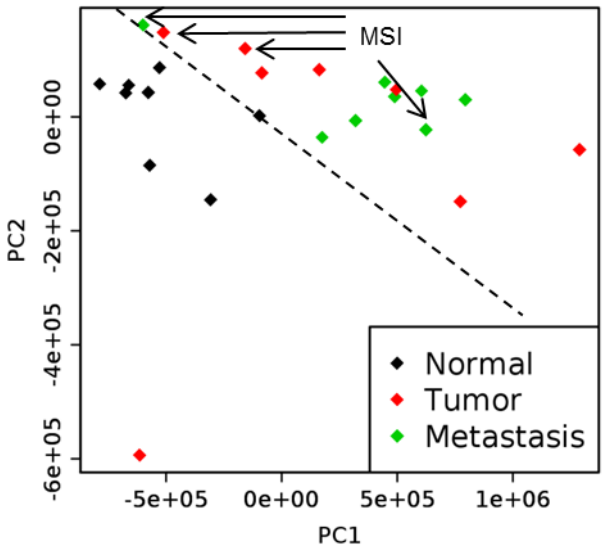

Figure 3.22: Principal Component Analysis (PCA) of tumor, metastasis and matched normal tissues of eight CRC patients. Read counts per microRNA were normalized by the sequencing depth per tissue. PCA plot for all tissues, normal (black), tumor (red) and metastases (green). Arrows (red dots) indicate microsatellite instable (MSI) tumor samples and arrows (green dots) specify the microsatellite instable metastasis samples.

3.2 TUMOR TISSUE SCREEN FOR MICRORNAS

On the basis of the NGS data, we selected interesting significantly deregulated microRNA candidates for further analyses. We wanted to know whether these microRNAs are specific for colon cancer or if they act as general tumor suppressors or oncogenes. To investigate their capability to act as biomarkers in colon cancer and other cancer types, we performed a cancer screen on 18 different tumor entities, including adipose tissue, brain, breast, colon, endometrium, kidney, liver, lung, lymph node, muscle, ovary, pancreas, prostate, stomach, testis, thyroid gland, melanoma and spleen. However spleen and melanoma provide only tumor samples or normal samples and therefore they were excluded from further analyses. For the remaining 16 tumor tissues, we used the corresponding normal tissue from a healthy control and investigated the expression of selected microRNAs (Table **3.4**). Altogether we tested six microRNAs (*miR-1*, *miR-135b*, *miR-129*, *miR-493*, *miR-215* and *miR-497*) in 330 tissue samples with TaqMan assays. The candidates were selected by different criteria such as differential expression, epigenetic modifications and literature entries (Table **3.5**).

Table 3.4: Composition of tissue samples screen including 330 different samples.

tissue	***# tumor samples***	***# normal samples***
adipose tissue	21	8
brain	20	6
breast	30	13
colon	10	5
endometrium	14	4
kidney	21	7

liver	10	5
lung	12	8
lymph node	13	3
muscle	11	5
ovary	15	1
pancreas	14	6
prostate	8	4
stomach	11	3
testis	10	4
thyroid gland	14	4
melanoma	9	-
spleen	-	1
total	**330**	

Table 3.5: Selected significantly deregulated microRNAs for the tumor tissue screen. The expression of the candidates was determined by NGS technology.

microRNAs	***mean fold-change (log2 ratio) Tumor***	***mean fold-change (log2 ratio) Metastases***	***chromosone localization***
miR-1-1	-3,80	4,20	Chr 20: 61151513-61151583 [+]
miR-1-2	-3,33	-3,82	Chr 18: 19408965-19409049 [-]
miR-129-1	-4,89	-2,72	Chr 7: 127847925-127847996 [+]
miR-129-2	-3,45	-2,49	Chr 11: 43602944-43603033 [+]
miR-215	-2,87	-3,95	Chr 1: 220291195-220291304 [-]
miR-497	-1,55	-2,13	Chr 17: 6921230-6921341 [-]
miR-135b	4,40	4,89	Chr 1: 205417430-205417526 [-]
miR-493	1,75	2,03	Chr 14: 101335397-101335485 [+]

MiR-1 and *miR-135b* were chosen for screening experiments because they showed the most prominently altered expression in primary colon cancer samples and metastases tissue. We applied the TaqMan assay technologies for the mature microRNA fraction and calculated the expression value of every microRNA by the delta delta Ct method. We normalized against a stable internal control gene (*RNU44*) and compared the expression to the normal tissue. The normal and tumor samples were not matched samples and therefore we combined all samples of one tumor type and calculated the median regulation over tumor and the corresponding normal tissue. A significantly up- or down-regulation was determined by a fold change of 1.5 ($-0.58 \geq \log2\text{ratio} \geq 0.58$).

We observed a constant down-regulation for *miR-1* over nearly all cancer tissues (Table **A.5**, Figure **3.23**). For muscle ($2^{-(\Delta\Delta Ct)}$ value = 0.05), ovary ($2^{-(\Delta\Delta Ct)}$ value = 0.01) and prostate ($2^{-(\Delta\Delta Ct)}$ value = 0.02) tissue we observed the most significant down-regulation in tumor. As expected, we also found a down-regulation of *miR-1* (fold change = 13.9) in the sample set of colon cancer patients.

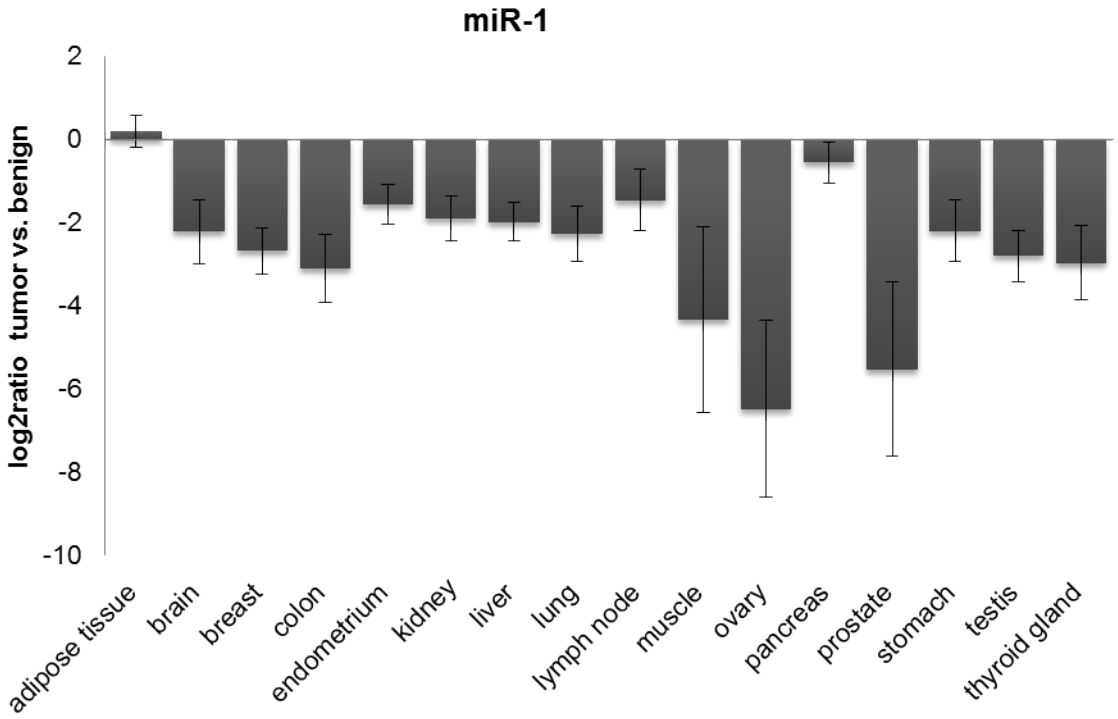

Figure 3.23: Expression of *miR-1* in different cancer entities using the TaqMan technology. We estimated the expression values of microRNAs using the delta delta Ct method with normalization against a stable internal control gene (*RNU44*) and as a second normalization relative to the normal tissues.

In comparison, the expression analysis of *miR-135b* delivered a variable pattern across all cancer tissues investigated (Table **A.6**, Figure **3.24**). In colon cancer we could observe an up-regulation (9.24-fold), as expected. In endometrium and stomach we estimated a significant up-regulation for the tumor tissues as well (endometrium = 15.45-fold, stomach = 5.71-fold). In four other tissues (adipose tissue, brain, lung, ovary) we found an increased expression level of *miR-135b*. In the majority of the investigated tissues, such as breast, kidney, liver, lymph node, muscle, pancreas, prostate, testis and thyroid gland, *miR-135b* showed a down-regulation.

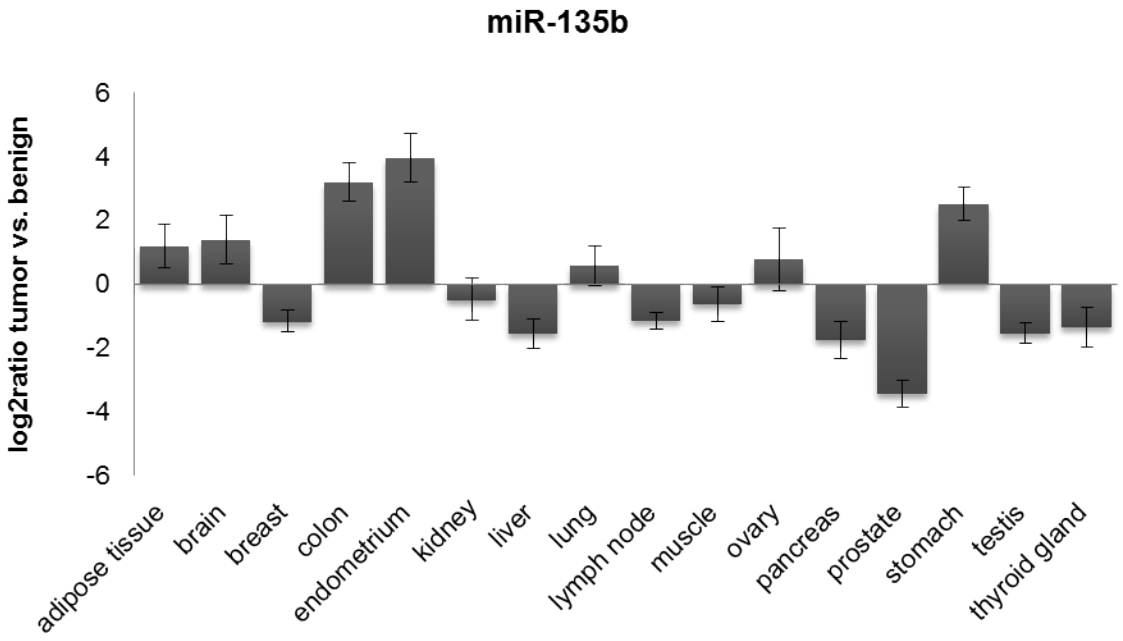

Figure 3.24: Expression of *miR-135b* in different cancer entities using the TaqMan technology. We estimated the expression values of microRNAs using the delta delta Ct method with normalization against a stable internal control gene (*RNU44*) and as a second normalization relative to the normal tissues.

Two candidates, *miR-129* and *miR-493*, were selected because we had found their chromosomal locus being differentially methylated. They exhibited specific methylation patterns in colon cancer patients using the MeDIP protocol published by Weber et al. [175]. The observed methylation status of each microRNA of interest was compared to the expression level generated by NGS technology of this specific microRNA. *MiR-129* is a candidate that showed a constant hyper-methylation in the tumor tissue of the colon cancer patients and it was represented in our NGS data-set as one of the top 25 down-regulated microRNA candidates. As expected, this microRNA showed a strong down-regulation in the colon cancer sample set ($2^{-(\Delta\Delta Ct)}$ value = 0.07) and over all other tumor tissues investigated in the screen. For brain ($2^{-(\Delta\Delta Ct)}$ value = 0.01), muscle ($2^{-(\Delta\Delta Ct)}$ value = 0.02) and testis ($2^{-(\Delta\Delta Ct)}$ value = 0.01) the expression in the tumor was even more reduced (Figure **3.25**, Table **A.7**).

In contrast to *miR-129*, we observed a hypo-methylation for *miR-493* in tumor tissues. Based on the NGS data we could only detect a 3.35-fold up-regulation in the tumor tissue and a 4.08-fold up-regulation in the metastases tissue (Table **A.1**, Table **A.2**). The majority of cancer tissues showed a down-regulation for *miR-493*. This microRNA is significantly up-regulated only in malignant adipose and brain tissue (adipose tissue = 3.21-fold, brain = 2.29-fold) (Figure **3.26**, Table **A.8**).

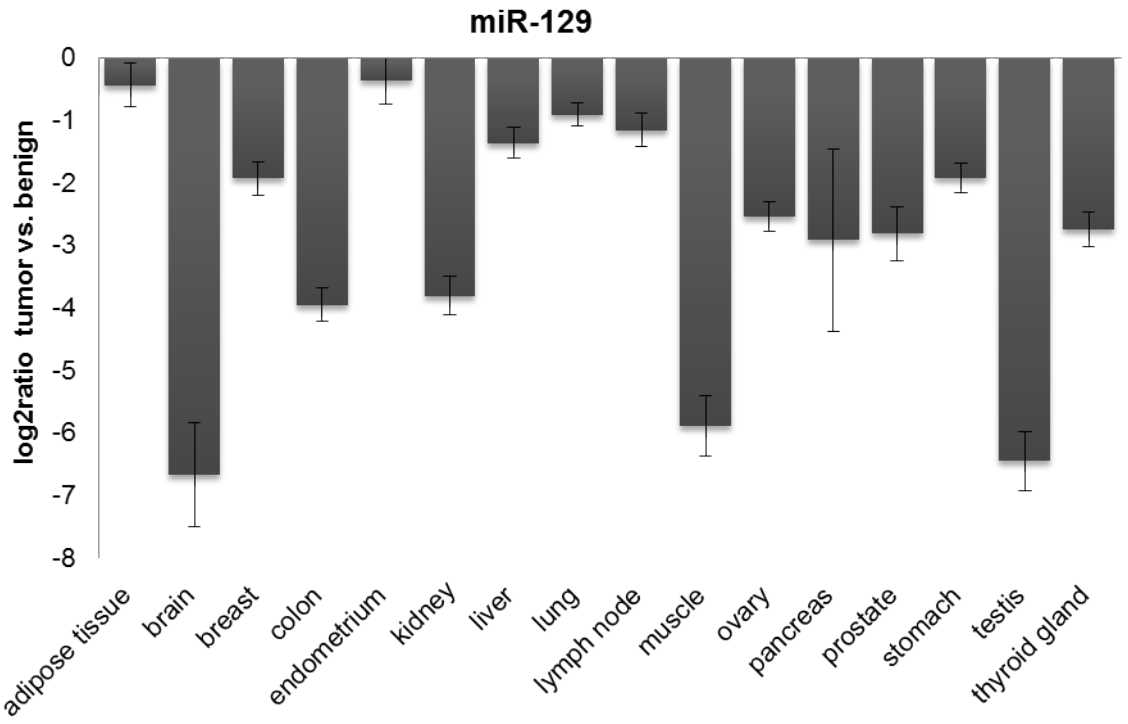

Figure 3.25: Expression of *miR-129* in different cancer entities using the TaqMan technology. We estimated the expression values of microRNAs using the delta delta Ct method with normalization against a stable internal control gene (*RNU44*) and as a second normalization relative to the normal tissues.

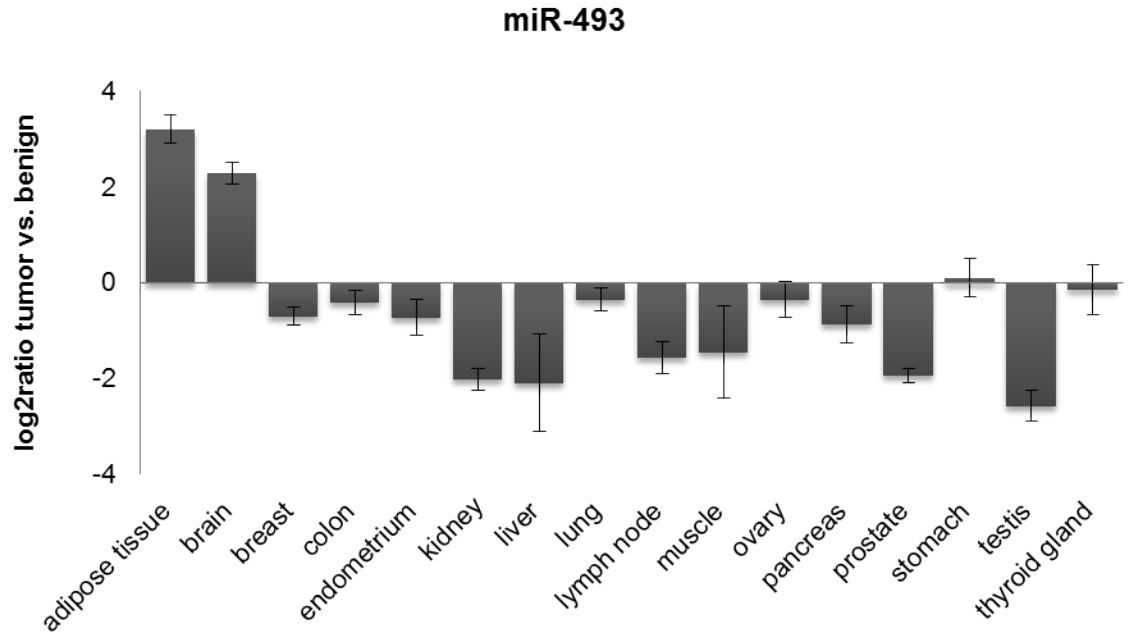

Figure 3.26: Expression of *miR-493* in different cancer entities using the TaqMan technology. We estimated the expression values of microRNAs using the delta delta Ct method with normalization against a stable internal control gene (*RNU44*) and as a second normalization relative to the normal tissues.

Two additional candidates, *miR-215* and *miR-497* have been included for further analyses. Our NGS investigations identified a significant reduction of *miR-215* in tumor (fold change = 2.93) and an even more decreased expression in metastasis tissues (fold change = 4.37) (Table **A.1**, Table **A.2**). For *miR-497*, we also detected a strong down-regulation in tumor and metastasis through our NGS studies. Screening the large tumor tissue sample set elucidated a down-regulation for *miR-215* in 14 out of 16 tissues, most notably in muscle tumor ($2^{-(\Delta\Delta Ct)}$ value = 0.09) (Figure **3.27**, Table **A.9**) Furthermore, a constant down-regulation for *miR-497* in nearly all tumor tissues was observed, especially in ovary ($2^{-(\Delta\Delta Ct)}$ value = 0.04) and breast tumor ($2^{-(\Delta\Delta Ct)}$ value = 0.11) (Figure **3.28**, Table **A.10**).

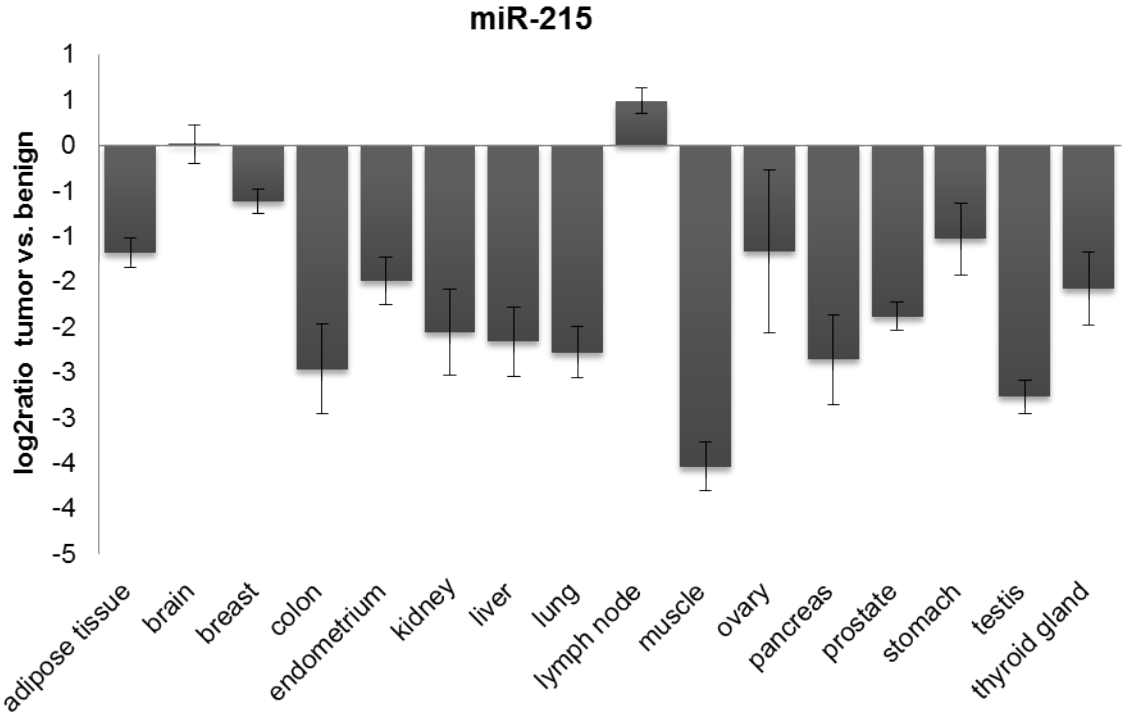

Figure 3.27: Expression of *miR-215* in different cancer entities using the TaqMan technology. We estimated the expression values of microRNAs using the delta delta Ct method with normalization against a stable internal control gene (*RNU44*) and as a second normalization relative to the normal tissues.

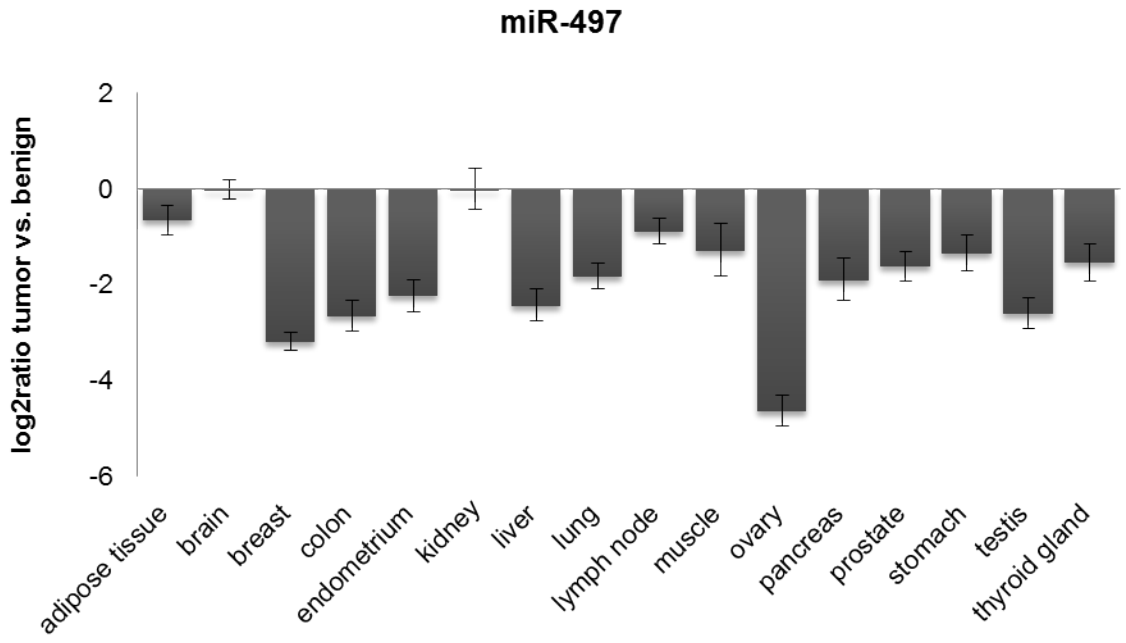

Figure 3.28: Expression of *miR-497* in different cancer entities using the TaqMan technology. We estimated the expression values of microRNAs using the delta delta Ct method with normalization against a stable internal control gene (*RNU44*) and as a second normalization relative to the normal tissues.

3.2.1 PRINCIPAL COMPONENT ANALYSES (PCA)

In order to estimate the power of the selected microRNA candidates (*miR-1*, *miR-135b, miR-129*, *miR-493*, *miR-215*, *miR-497*) to act as biomarkers with the capability to distinguish between tumor and benign tissues, we performed principal component analyses (PCA). Applying PCA on our expression data for the six microRNA candidates and 16 cancer tissues we could observe a separation of tumor and normal tissues. Here, the two-dimensional distribution of the different samples is less significant, considering the specific expression patterns of the selected molecular markers (Figure **3.29**). The PCA analyses were performed on individual tissues and for brain, prostate and stomach and we achieved a complete separation for benign and malignant tissues (Figure **3.30**). For colon tissue we determined a nearly complete separation except of one outlier in the tumor tissue.

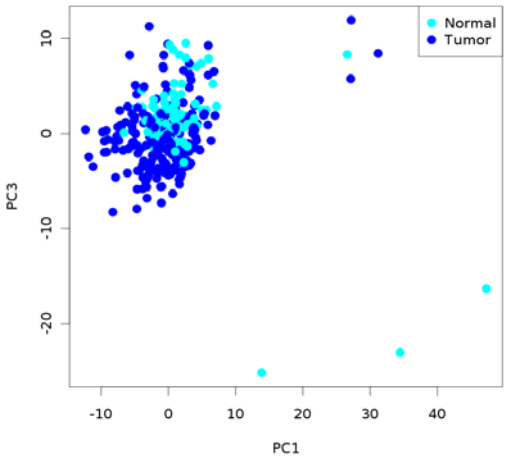

Figure 3.29: Principal Component Analysis (PCA) of the 330 tumor (dark blue) and corresponding normal tissues (light blue) contained in the large microRNA screen using TaqMan expression values of the microRNAs *miR-1*, *miR-135b*, *miR-129*, miR-493, *miR-215* and *miR-497*. PCA plots were generated for the union of all tumor and normal samples.

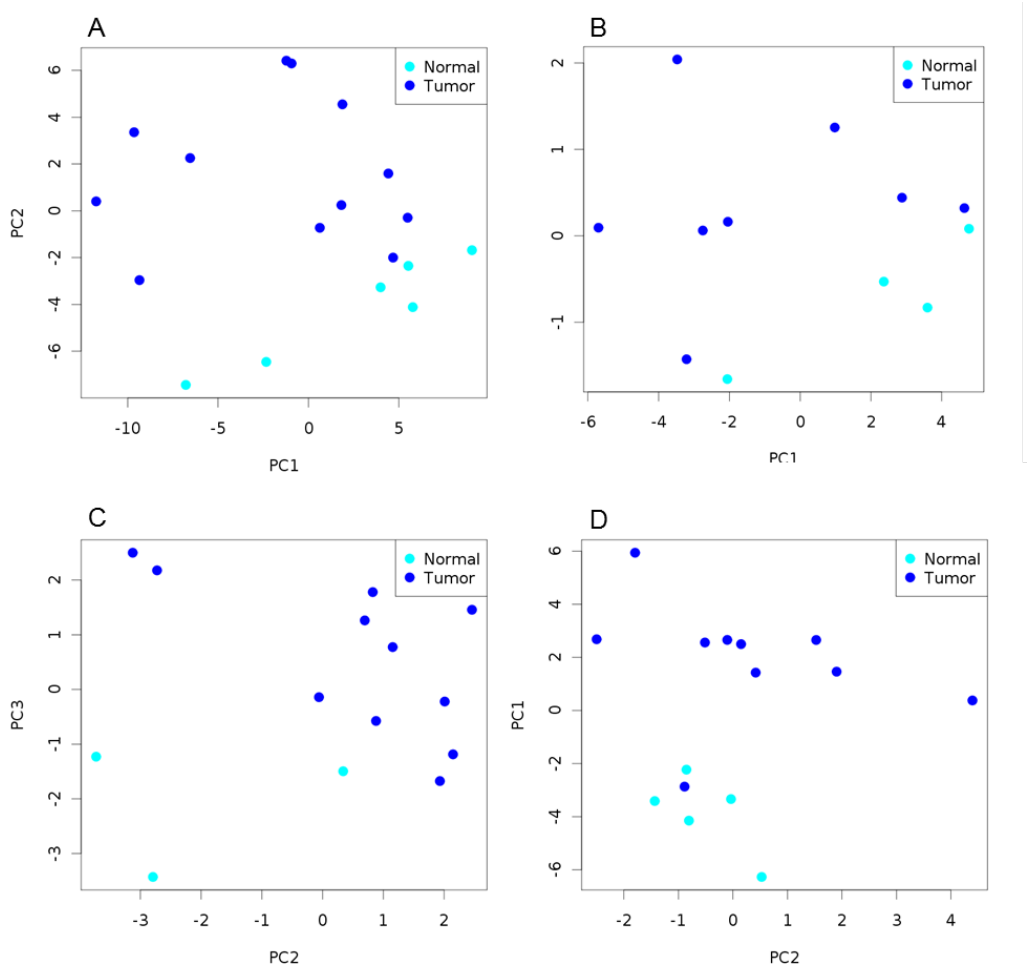

Figure 3.30: Principal Component Analysis (PCA) of tumor (dark blue) and corresponding normal tissues (light blue) using TaqMan expression values of the microRNAs *miR-1*, *miR-135b*, *miR-129*, *miR-493*, *miR-215* and *miR-497*. (**A**) PCA plot for brain tissues. (**B**) PCA plot for prostate tissues. (**C**) PCA plot for stomach samples and (**D**) PCA plot for colon samples.

3.2.2 HIERARCHICAL CLUSTER ANALYSES FOR MICRORNA CANDIDATES

Since we were interested in the possibility of the selected microRNA candidates (*miR-1*, *miR-135b*, *miR-129*, *miR-493*, *miR-215*, *miR-497*) to function as colon specific biomarkers we applied hierarchical clustering analyses. The subsequent heatmaps symbolize the expression values for selected microRNA candidates calculated by the delta Ct method with normalization against the internal control gene *RNU44*. A reduction of the number of considered microRNAs for clustering analysis still remained in a significant differentiation of normal and malignant colon tissues (Figure **3.31**). The initial clustering analysis was performed on the set of colon cancer patients used for the tumor screen (Table **A.11**).

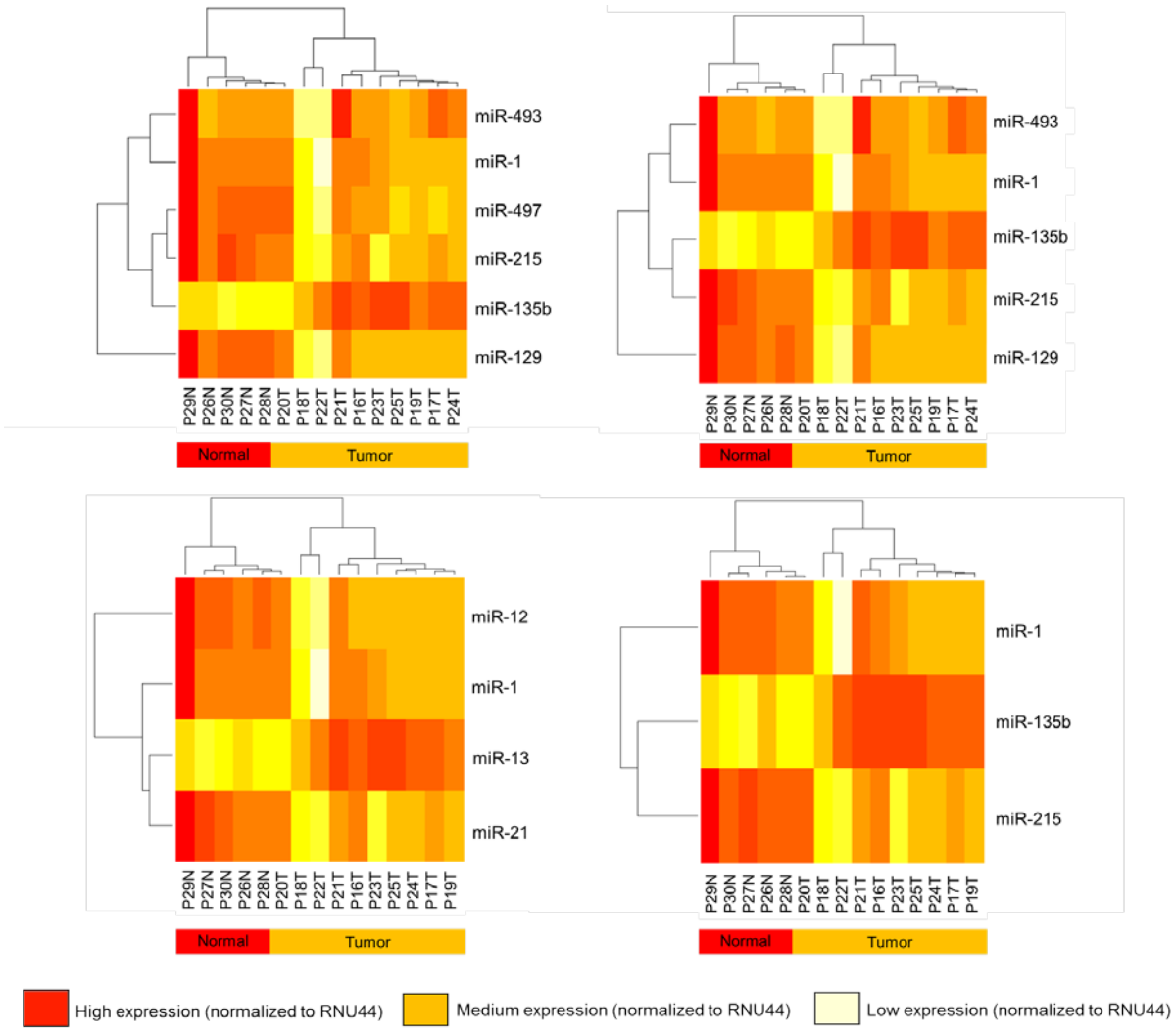

Figure 3.31: Hierarchical cluster analysis of tumor and normal colon samples based on the expression levels of *miR-1*, *miR-135b*, *miR-129*, *miR-493*, *miR-215*, *miR-497* in the tumor screen set. Calculations were done by the delta Ct method including normalization to the internal control gene *RNU44*.

Since it was possible to differentiate malignant and tumor tissue on the basis of six and even fewer microRNA candidates, we further narrowed down the number of potential biomarkers and enlarged the set of colon cancer patients to include for clustering analysis samples from the tumor screen and the

eight colon cancer patients we investigated in previous studies (Table **A.11**) and applied delta Ct values for every sample generated by TaqMan technology for selected microRNAs, *miR-1* and *miR-135b*. Since *miR-1* and *miR-135b* were the most significantly deregulated microRNAs in colorectal tumors, we asked whether it is possible to distinguish tumor and normal tissues from each other on the basis of their expression profile. The cluster analysis revealed a nearly complete separation of tumor and normal tissues (Figure **3.32**).

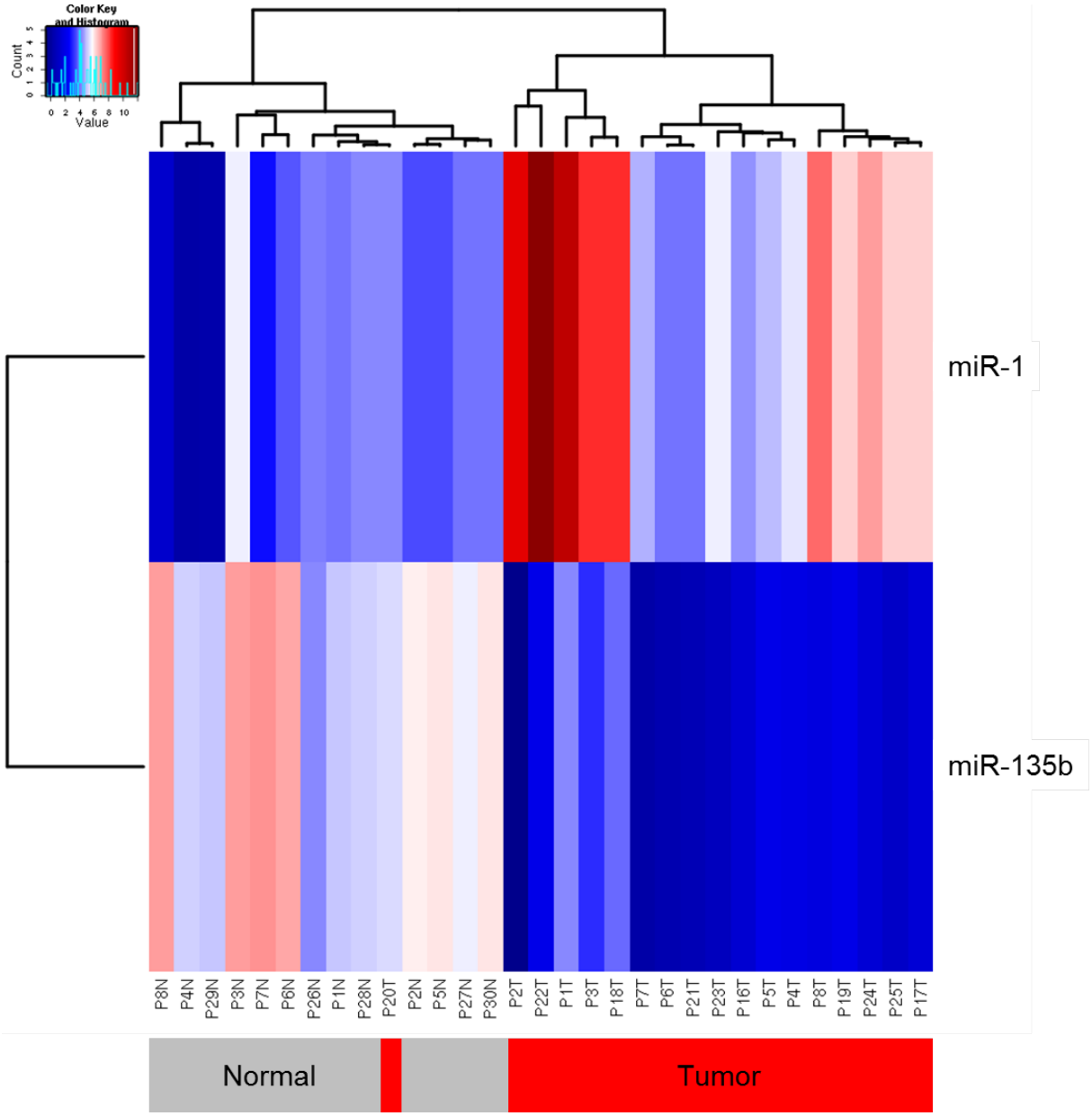

Figure 3.32: Hierarchical cluster analysis based on the expression levels of *miR-1* and *miR-135b* for the unity of tumor and normal colon samples investigated by TaqMan technology. Data were calculated by the delta Ct method including normalization to the internal control gene *RNU44*.

3.3 *MiR-1* AS A TUMOR SUPPRESSOR IN CRC

Our data of microRNA expression in colon cancer revealed that the level of *miR-1* was significantly reduced in primary tumor and metastases. In addition we found *miR-1* constantly down-regulated over diverse tumor entities. This aids the assumption of a tumor suppressor-like role for *miR-1* in cancer, which we deciphered in further functional analyses.

3.3.1 EXPRESSION OF *MIR-1* IN CRC AND REPRESENTATIVE CELL LINES

By sequencing smallRNA fractions of CRC we found that the level of *miR-1* was reduced in colon cancer and metastasis tissue when compared to the matched normal samples. *MiR-1* showed a down-regulation not only in CRC patients but also in other cancer tissues, for example, brain, breast, endometrium, kidney, liver, lung, lymphatic system, muscle, ovary, pancreas, prostate, stomach, testis and thyroid gland (Figure **3.23**, Table **A.5**). To prove the assumption that *miR-1* can act as tumor suppressor we performed functional assays, where we estimated the influence of *miR-1* overexpression towards cell viability and wound healing behavior of colon cancer cells. First we examined the expression level of *miR-1* in SW480 and SW620 colorectal cancer cell lines. Applying quantitative real-time PCR to representative cell lines, such as SW480 and SW620, we detected a decreased level of *miR-1* as compared to the normal tissue of all colon cancer patients. We determined a 61-fold down-regulation in the primary colon cancer cell line SW480 and a 7-fold reduced expression in the metastases cell line SW620 (Figure **3.14**, Figure **3.33**). The even more reduced level of *miR-1* in SW480 was also detectable when we compared expression values acquired through next generation sequencing. This corresponds clearly to the qPCR data. We therefore were interested in the functional impact of *miR-1* on the cellular homeostasis.

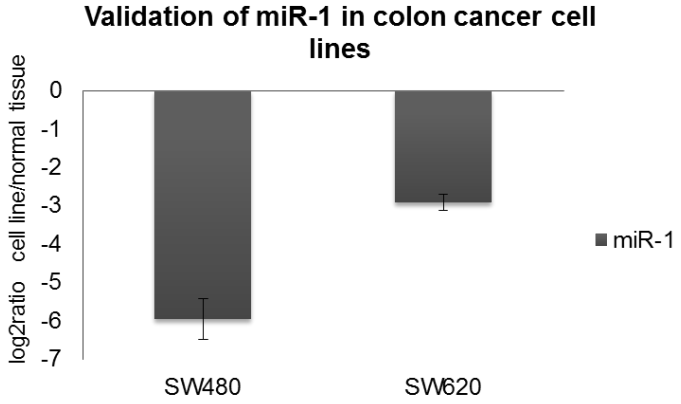

Figure 3.33: Expression level of *miR-1* in colon cancer cell lines SW480 and SW620 using quantitative real-time PCR. The obtained Ct-values were normalized against normal tissue of CRC patients. For SW480 we observed a 61-fold and for SW620 a 7-fold reduced expression.

3.3.2 EFFECT OF *MIR-1* ON CELL VIABILITY

After having established SW480 and SW620 cell lines as good model systems to study the function of *miR-1*, we checked whether *miR-1* has an influence on the viability of colon cancer cells. We transfected SW480 and SW620 cells with *miR-1* mimics determined the *miR-1* expression level by TaqMan assays and monitored the metabolism of the cells using the dye alamarBlue. For both cell lines we found decreased cell viability with increased *miR-1* levels. We observed the most noticeable effect on the SW480 cells, after 72h of transfection compared to the non-transfected cells. Measuring the fluorescence intensity of the metabolizing cells, we detected decreased cell viability after 48h and 72h compared to the non-transfected cells. We detected a decrease of approximately 25% after 48 hours and 30% after 72 hours for SW480 as well as a decrease of 25% after 48 hours and 20% after 72 hours for SW620 cells after *miR-1* transfection (Figure **3.34**).

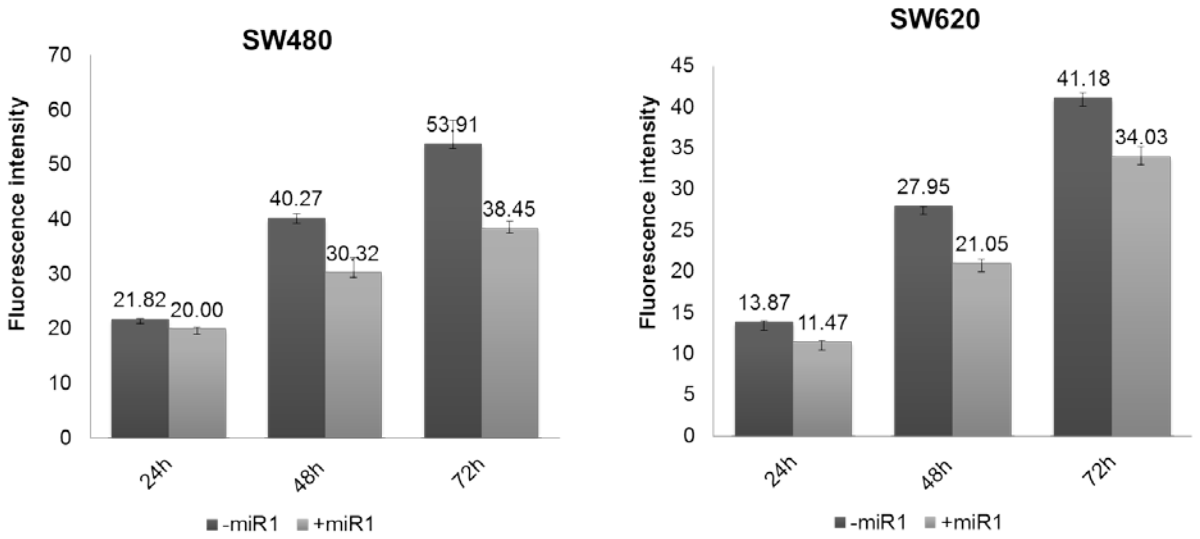

Figure 3.34: Cell viability assay with alamarBlue to test the effect of *miR-1* in the primary colon cancer cell line SW480 and the corresponding metastases cell line SW620. Cells were transfected with *miR-1* mimics and measured cell viability using a spectrophotometer after 24h, 48h and 72h.

3.3.3 SCRATCH AND WOUND HEALING EXPERIMENTS

We also examined the motility of colon cancer cells (SW480, SW620) at 0 and 24hours using a wound healing assay. Scratching a confluent monolayer of cells with a tip generated a gap void of cells 'the wound'. We then documented and compared the migration of the cells to close the wound in transfected and non-transfected cells. For the treated cells we could observe a much slower migration rate towards the wound than for the non-treated cells. This was detectable for both cell lines after 24h of transfection and demonstrates that there is a marked effect on the migration rate of colon cancer cells depending on their *miR-1* expression level (Figure **3.35**).

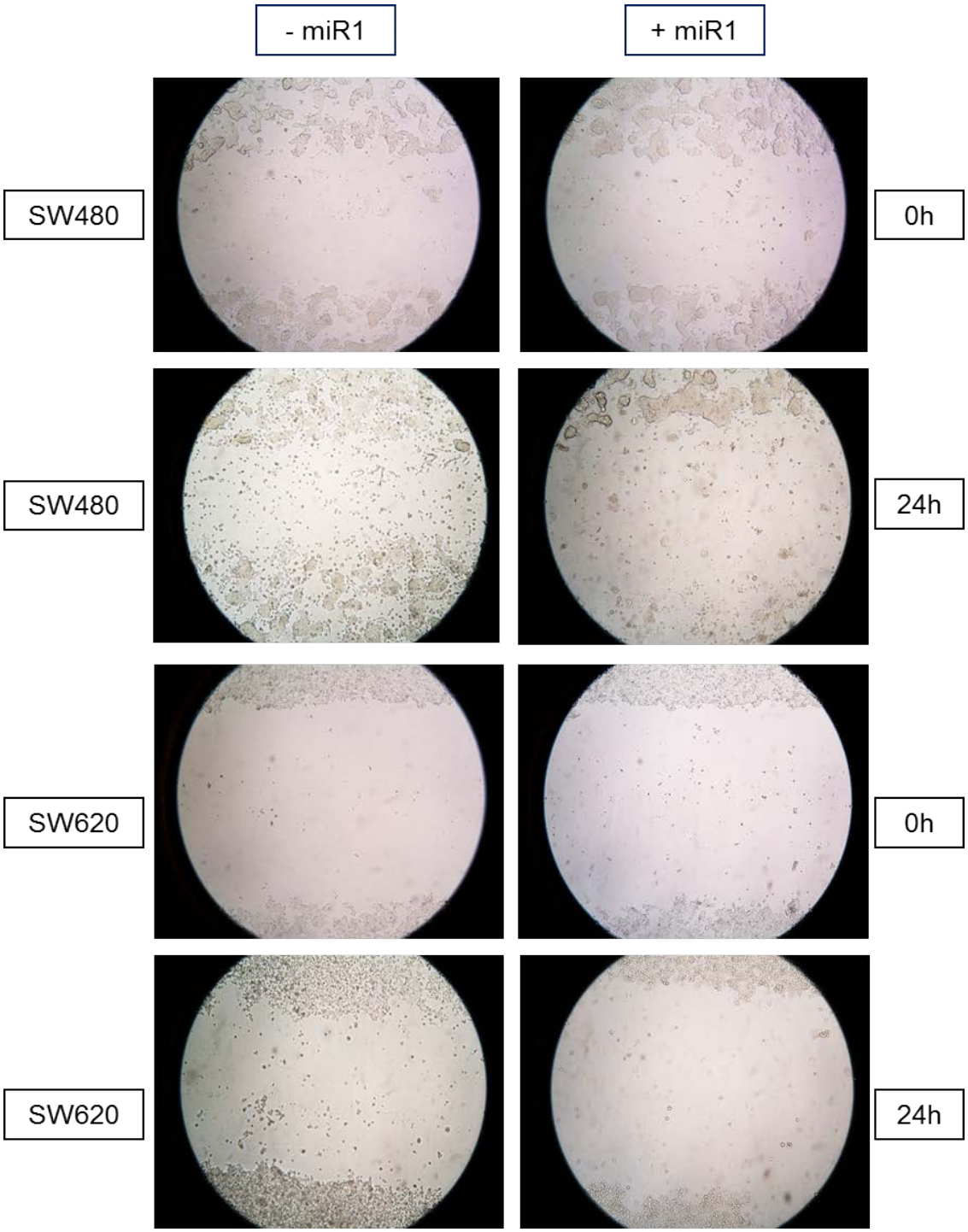

Figure 3.35: Scratch and wound healing assay of *miR-1* transfected SW480 and SW620 cells after 24h of transfection with *miR-1* mimics, a uniform scratch was generated through each confluent cell layer and 'wound' closure was documented after 24h using a phase-contrast microscope (n=2).

3.3.4 ANALYSIS OF *MIR-1* TARGETS

Each microRNA can target different genes and regulate the protein expression on a transcriptional and post-transcriptional level. To identify genes which are targeted by *miR-1*, we selected candidates using different target prediction algorithms, such as TargetScan and PicTar. Using different target prediction

tools delivered diverse pools of mRNA target genes based on the algorithms they apply to scan for conserved binding sites in the 3'UTR of mRNAs and for perfect complementarity of the microRNA 'seed' region (2-8 nucleotides in length). On the basis of these prediction tools, we selected *PTPLAD1* (protein tyrosine phosphatase-like A domain containing 1) as the most promising candidate, which was presented as top target gene in each prediction database with three conserved binding sites in its 3'UTR for *miR-1*. Additional target genes were chosen, such as *VEGFA* (vascular endothelial growth factor A) and *PDGFA* (platelet-derived growth factor alpha polypeptide), based upon their role as therapeutic targets in cancer treatment. The target genes *PDGFA* and *VEGFA* are targets for effective anti-cancer treatment using approved target agents, such as bevacizumab [177,178]. In contrast to *PTPLAD1*, both of them possess only one binding motif in their 3'UTR for *miR-1*. Interestingly, the three selected candidates, *PTPLAD1*, *PDGFA* and *VEGFA*, are predicted to interact with *HIF1A* (hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)). Additionally, the 3'UTR of *HIF1A* revealed to be a target element for *miR-1*, based on bioinformatic analysis.

Since we expected a negative correlation between microRNA and gene expression we analyzed the endogenous expression level of the specific target genes in the human primary colon tumors. Therefore we performed whole transcriptome sequencing using Illumina sequencing by synthesis technology (Table **3.6**).

Table 3.6: MessengerRNA sequencing statistic for colon cancer patients P3, P4, P5, P7 (the RNA for the runs labeled with 1 was previously applied to the Ribominus Transcriptome Isolation Kit and for cDNA synthesis the RNA was reverse transcribed into cDNA using random hexamer primers before the sequencing procedure; the RNA for the runs labeled with 2 was not applied to the Ribominus Transcriptome Isolation Kit and the reverse transcription was performed using poly-dT primer before sequencing).

Run	Sample	# raw reads	# unique mapped reads
1	P3-Normal	59,900,420	13,053,767
1	P3-Tumor	60,656,428	12,170,871
1	P3-Metastases	63,080,304	11,197,763
1	P4-Normal	63,741,116	17,198,033
2	P4-Normal	32,294,448	1,413,461
1	P4-Tumor	29,748,452	5,091,495
2	P4-Tumor	34,785,688	5,242,675
1	P4-Metastases	45,487,456	14,106,164
2	P4-Metastases	31,714,848	4,283,291
1	P5-Normal	89,940,968	42,278,118
1	P5-Tumor	56,228,600	23,773,207
1	P5-Metastases	68,114,084	9,493,795
1	P7-Normal	20,241,412	9,707,686
1	P7-Tumor	73,901,972	21,478,952
1	P7-Metastases	65,852,464	20,620,425

The differential expression of the transcriptome was determined for four different colon cancer patients using the Cufflinks software. Data for transcriptome sequencing were generated by mapping adaptor-clipped reads to the human genome version GRCh37 (hg19) using transcriptome models taken from ENSEMBL v64 with TopHat [179]. Applying the Cufflinks software bundle [180] delivered differentially expressed mRNA transcripts for each tissue of every patient. In detail, transcript models from all tissues were merged for four patients (P3, P4, P5, P7) using cuffcompare and differential expression was determined using cuffdiff with upper quartile normalization. For selected *miR-1* targets, such as *PTPLAD1*, *VEGFA* and *PDGFA*, we determined an up-regulation at least in one patient in both malignant tissues (Figure **3.36**). Because of the constant down-regulation of *miR-1*, we expected an up-regulation for the selected target genes on the mRNA level. At least for *VEGFA* and *PTPLAD1* we could observe a constant up-regulation in tumor tissue. Going into detail, we could detect an up-regulation in both malignant tissues for *PTPLAD1* especially in patients P4, P5 and P7; for *VEGFA* in patients P3 and P4; and for *PDGFA* at least in P1 of the four patients. In addition, we checked the expression by an independent method using qPCR with gene specific primers where we also included patient 1. For patient 5 and 8 not enough RNA was available. We detected an up-regulation for *PTPLAD1* and *VEGFA*. Going into detail and interpreting the single data for every patient, we investigated the significance of the regulation for the selected targets. For *PTPLAD1* we observed a notable increase of expression in P1, P3, P4 and P6 in tumor tissue, which correlates to the next generation sequencing data. In the metastasis tissue, the corresponding data were ambivalent. The target gene *VEGFA* was detected as an up-regulated gene in patients P1, P3, P4 and P6 either by next generation sequencing or qPCR, whereas the results for P3 and P4 corresponded in both methods for tumor and metastases. Checking the data by qPCR delivered global concordance to the estimated regulation in the tumor. The regulation of *PDGFA* in the metastasis was more versatile over the CRC patients. Here, an increased expression in patient P3 could be observed (Figure **3.36**).

Considering the connection of miR-1 regulation in all CRC patients, we could not observe a direct dependence of the miR-1 expression and regulation of the selected target genes in tumor or metastasis. Only for the target genes *VEGFA* and *PTPLAD1* we could observe a tendency of correlation in some of the CRC patients, for example P3 and P6 (Table **3.7**, Figure **3.36**, Figure **3.37**).

Table 3.7: Correlation of *miR-1* regulation and target gene regulation for *PTPLAD1*, *VEGFA* and *PDGFA* in tumor (T) and metastasis (M) for eight colon cancer patients determined by next generation sequencing and qPCR (↓ = down-regulation, ↑ = up-regulation, n/a = not annotated).

	Patient 1		***Patient 2***		***Patient 3***		***Patient 4***		***Patient 5***		***Patient 6***		***Patient 7***		***Patient 8***	
	T	M	T	M	T	M	T	M	T	M	T	M	T	M	T	M
miR-1	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
PTPLAD1	↑	↓	n/a	n/a	↑	↑	↑	↑	↑	n/a	↑	↑	↓	↓	n/a	n/a
VEGFA	↑	↑	n/a	n/a	↑	↑	↑	↑	↑	↓	↑	↑	↑	↓	n/a	n/a
PDGFA	↓	↓	n/a	n/a	↓	↑	↓	↓	↓	↓	↑	↑	↓	↓	n/a	n/a

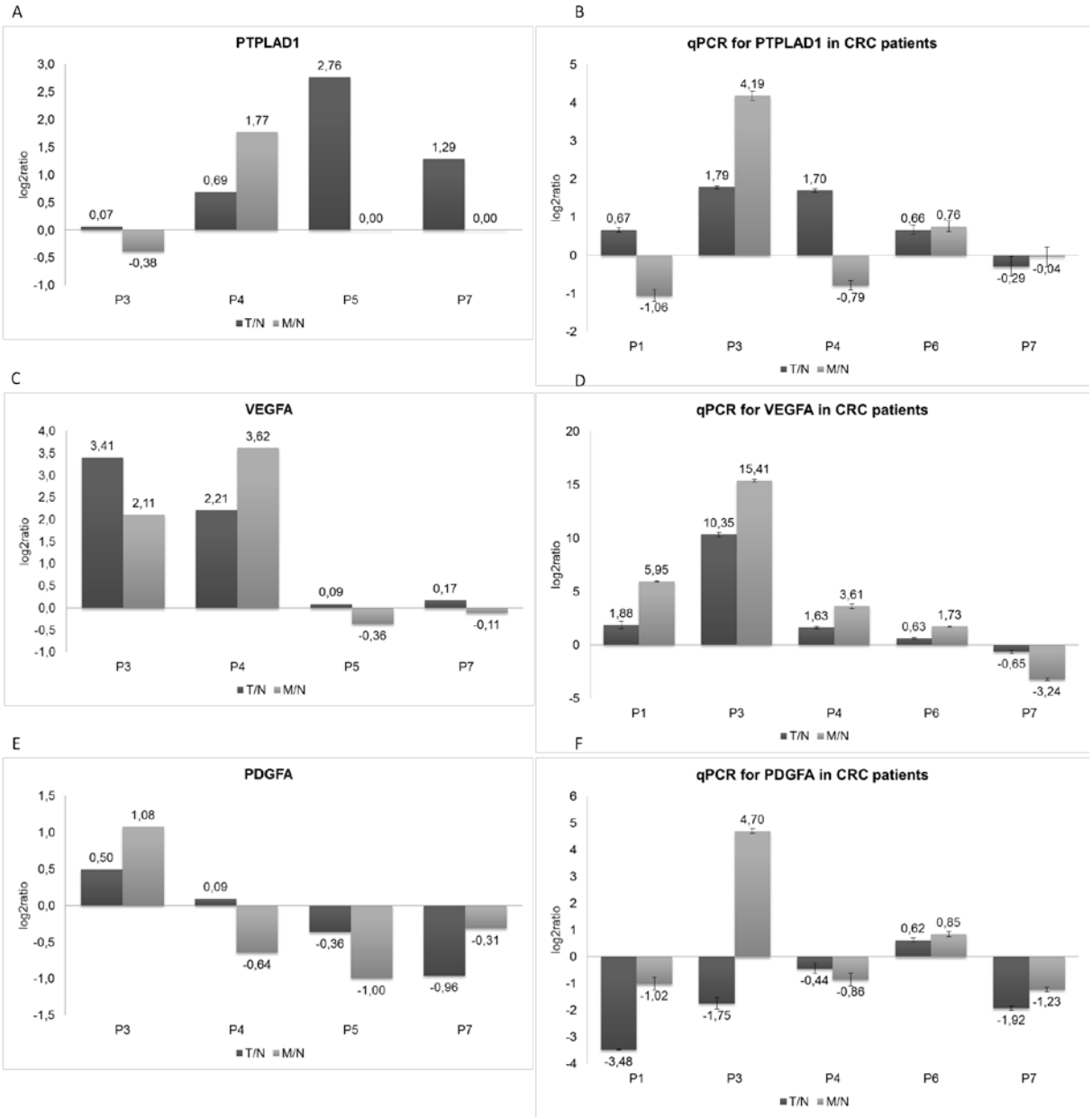

Figure 3.36: Expression values for mRNA of *PTPLAD1*, *VEGFA* and *PDGFA* for tumor and metastasis in four colon cancer patients (P3, P4, P5, P7) were represented as log2ratios. Next generation sequencing data (**A**) for *PTPLAD1* (**C**) for *VEGFA* and (**E**) for *PDGFA* were shown. In addition, qPCR data were calculated by the delta delta Ct method with normalization towards *HPRT* and normal tissue (**B**) for *PTPLAD1*. (**D**) for *VEGFA* and (**G**) for *PDGFA*.

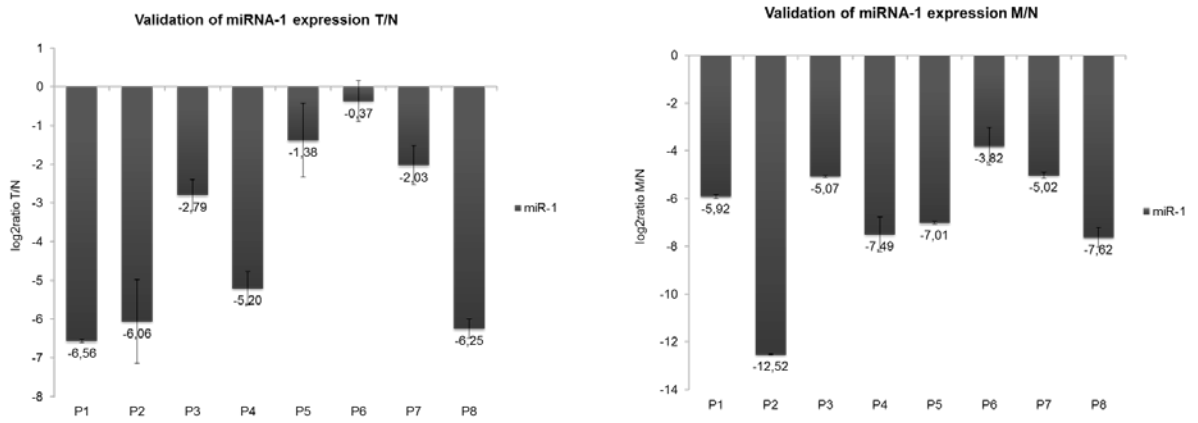

Figure 3.37: Expression of *miR-1* in tumor (T) and metastasis (M) tissue normalized to matched normal (N) samples of 8 colon cancer patients determined by TaqMan technology. The expression was calculated by delta delta Ct method with normalization against *RNU44* and normal tissue.

Looking for *miR-1* target genes that showed a significant up-regulation in our next generation sequencing data and plays a critical role in cancer progression we determined *ERBB4* (Receptor tyrosine-protein kinase erbB-4) (Figure **3.38**). This target gene is known to play a critical role in human tumor progression, among other things, in colon cancer [181] and may be therapeutically beneficial in cancer [182].

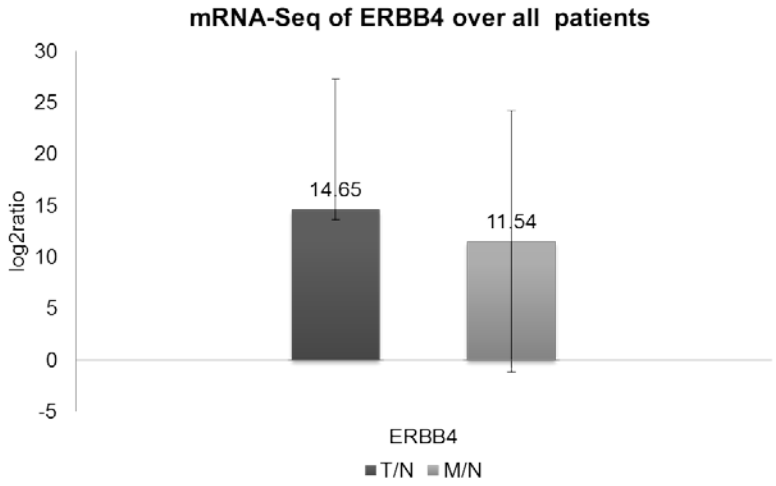

Figure 3.38: Expression values of mRNA next generation sequencing for *ERBB4*, determined for four colon cancer patients. The expression was calculated by Cufflinks software and the average log2ratio of all patients is represented for tumor versus normal and metastases versus normal tissue.

In general, correlation analyses as shown here cannot be used as proof that a specific microRNA regulates gene expression, they provide only indications. In addition, the regulation networks in tumor cells are complex, so if one does not find good correlations, it does not mean that the gene is not a target, it might be just differentially regulated by other mechanisms. To investigate if *miR-1* regulates

PTPLAD1, *VEGFA* and *PDGFA,* we performed microRNA-overexpression and luciferase experiments in SW480 cells. Therefore we transfected this colon cancer cell line with *miR-1* mimics and determined the mRNA levels of the targets. The overexpression of mature *miR-1* was detected by quantitative real-time PCR (Figure **3.39**).

Applying this technology on the transfected cells, we observed a significant up-regulation of *miR-1* in the cell line SW480 after 24h and 48h. To determine the expression level of the mRNA target elements, we used qPCR with gene specific primers. We were able to observe a significant decreased expression for *PTPLAD1* and *VEGFA* related to the control (number of experiments n=3). For the target gene *PDGFA* no reduction in expression was visible (Figure **3.40**). These results indicate that *miR-1* regulates *PTPLAD1* and *VEGFA*. Most microRNA-based regulations occur in the 3'prime untranslated region (3'UTR) of the respective genes. A microRNA silencing of these genes would be possible through the three conserved binding sites for *miR-1* in the 3'UTR of *PTPLAD1*. *VEGFA* contains at least one binding site for *miR-1*. To prove the assumption of microRNA silencing for these target genes, we generated GFP-vector constructs that included the 3'UTR of *PTPLAD1*, *VEGFA*, as well as for *PDGFA* and *HIF1A*. Then, the GFP-reporter construct was co-transfected with *miR-1* mimics into HEK-293 cells and the obtained fluorescence intensity in the presence and absence of *miR-1* was analyzed using the fluorescence activated cell sorting (FACS) technology. Only for *PTPLAD1* we could observe a strong and significant reduction in the GFP signal after 48h of transfection, verified in three biological replicates (Figure **3.41**). After 24h of transfection we could detect only a weak difference between the *miR-1* treated and non-treated cells. Here we a measured the geometric mean of the fluorescence signal (Geo. Mean FL1-H) of 43.85 for the *miR-1* transfected cells and 47.55 for the control sample. The measured GFP signal for the treated cells after 48h was 60.05 and for the non-treated 87.2, which resulted in a 31% reduction of the signal intensity after 48h of overexpression of *miR-1*. This means that *miR-1* directly regulates *PTPLAD1* over its 3'UTR (Figure **3.41**). For *PDGFA* we could not measure a constant effect after overexpression of *miR-1*. This might be due to the fact that it contains only one *miR-1* target element in their 3'UTR sequence, although the binding site is strongly conserved. Actually, this conservation in the binding motif let us anticipate that there is a possible interaction and inhibition between *miR-1* and *PDGFA*. The GFP-reporter assay for *HIF1A* showed no inhibitory effect of *miR-1* with the 3'UTR of *HIF1A* (Figure **3.42**). For the target gene *VEGFA*, we could not generate a functional GFP reporter construct and therefore it was not possible to investigate the interaction between *miR-1* and *VEGFA*, even though we expected a similar result as we received for *PDGFA*, because both contain only one conserved binding site.

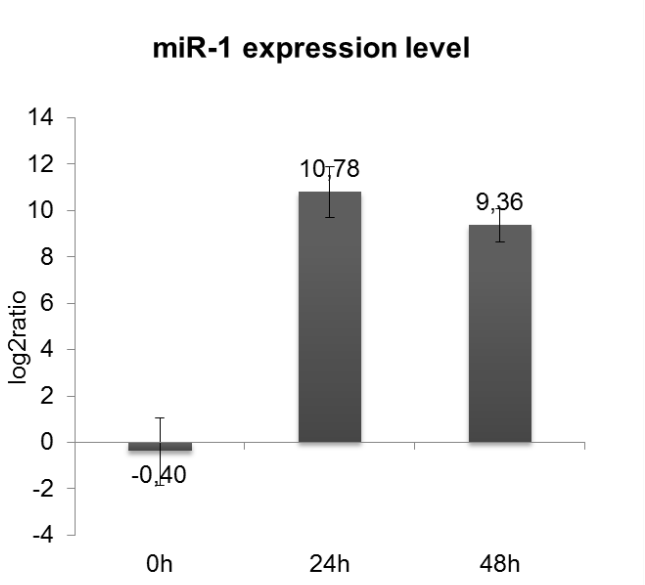

Figure 3.39: Expression level of mature *miR-1* after transfection in SW480 colon cancer cells. The expression was detected by quantitative real-time PCR against the mature microRNA and was normalized to an internal control *RNU44*.

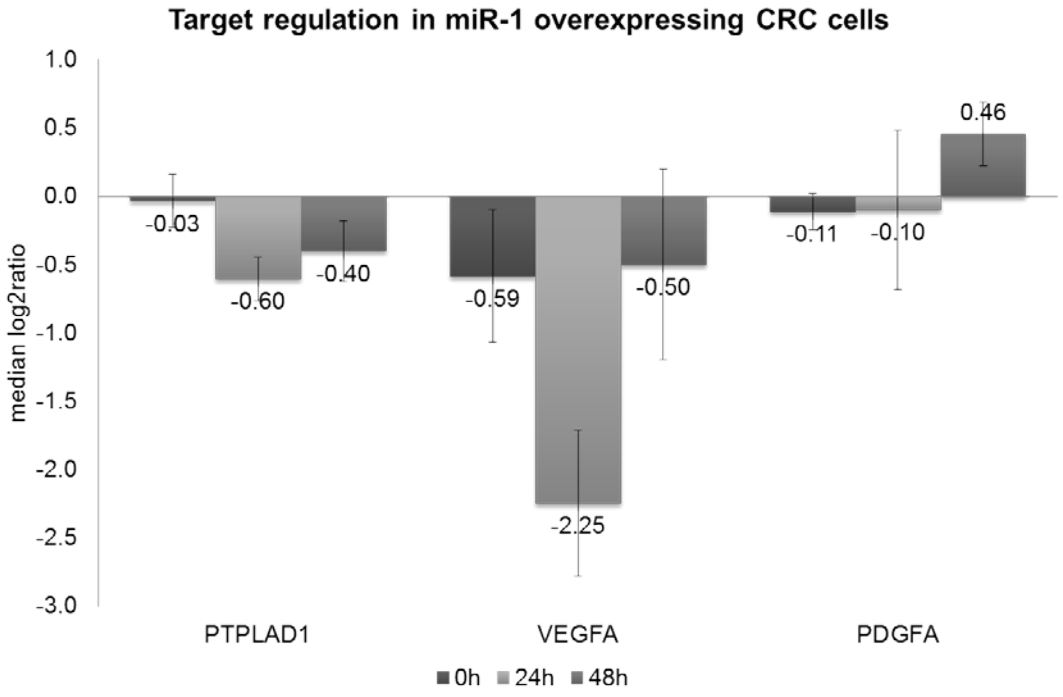

Figure 3.40: Detection of *miR-1* target regulation. *MiR-1* was overexpressed in the colon cancer cell line SW480. The expression levels of *PTPLAD1*, *VEGFA* and *PDGFA* were measured using quantitative real-time PCR after 0h, 24 and 48h of transfection and normalized to the transfection control.

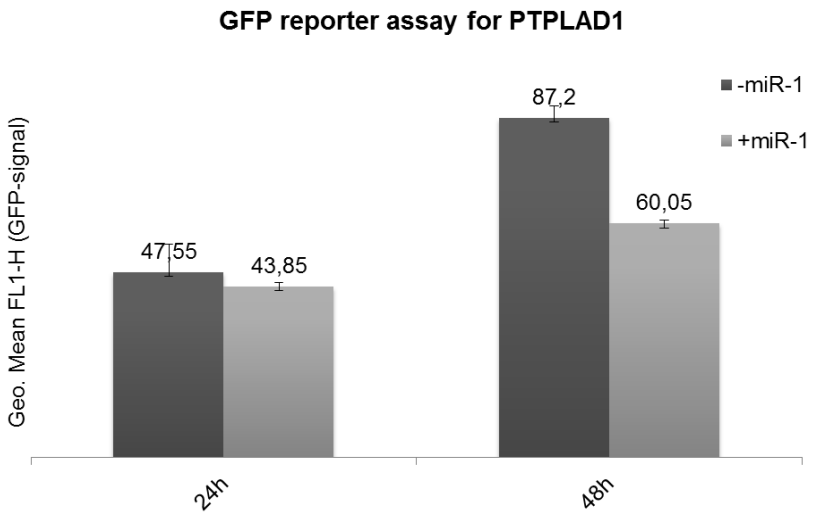

Figure 3.41: GFP reporter assay for *PTPLAD1* in the presence and absence of *miR-1*. The GFP reporter construct includes the 3'UTR of *PTPLAD1* and was co-transfected with *miR-1* mimics into HEK-293 cells. The fluorescence was measured after 24h and 48h using FACS analysis.

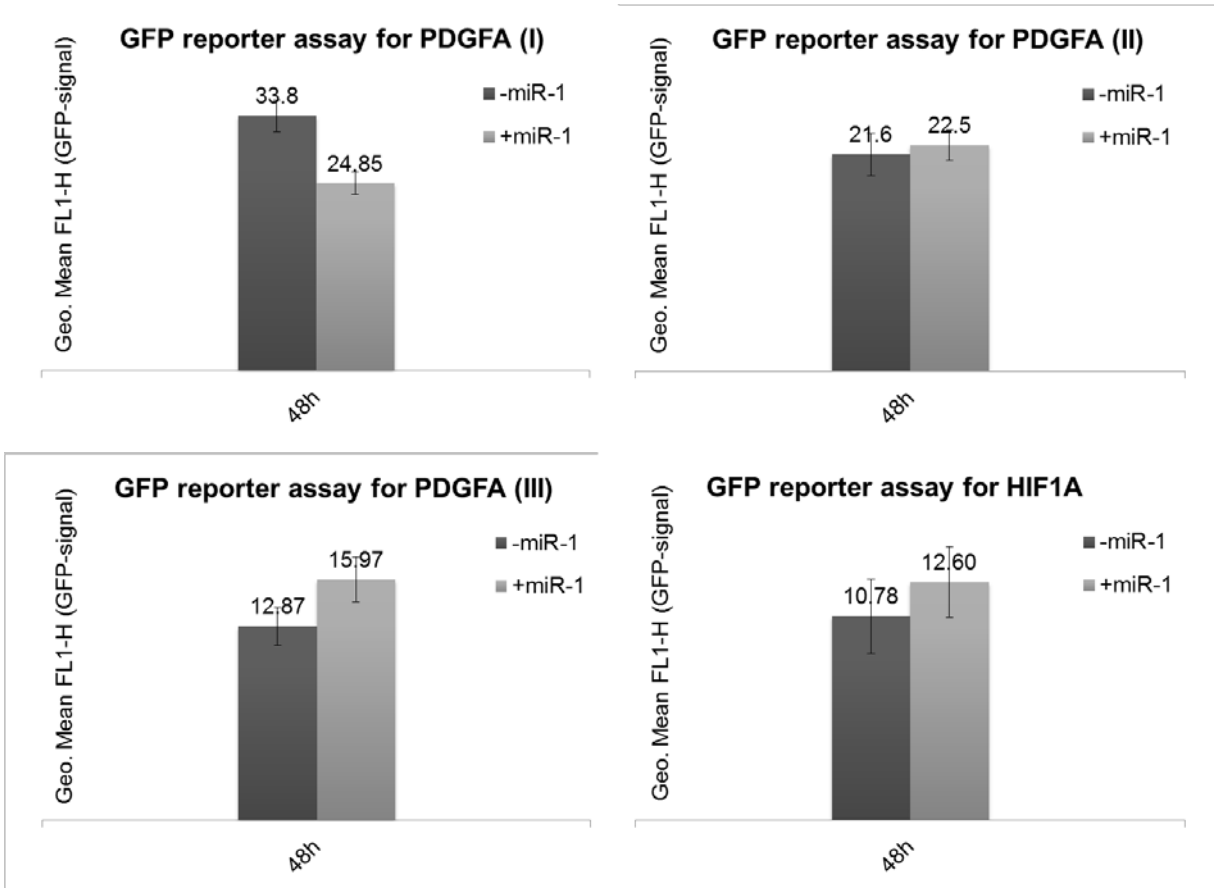

Figure 3.42: Single Experiments of GFP reporter assay for *PDGFA* and *HIF1A* in the presence of *miR-1*. The GFP-reporter constructs include the 3'UTR of *PDGFA* or *HIF1A* and were co-transfected with *miR-1* mimics into HEK-293 cells. The fluorescence was measured after 48h using FACS analysis.

3.3.5 *MiR-1* IN ASSOCIATION TO CANCER DRUG TREATMENT

A major challenging aspect in cancer therapy is resistance towards chemotherapeutics. Therefore, sensitization of cancer cells towards chemotherapy is desirable. Recent studies showed an effect of microRNA expression to the response of CRC cells to different chemotherapeutics [163,183-188]. In addition we found references that *miR-1* can modulate the impact of chemotherapeutics in lung cancer cells [94]. Nasser et al., found that *miR-1* acts synergistically with doxorubicin in lung cancer cells. These findings lead us to the investigation of chemosensitivity of CRC cells towards camptothecin in the presence of *miR-1* overexpression. We transfected SW480 and SW620 cells with *miR-1* mimics and treated the cells with 0.06µM camptothecin, a topoisomerase inhibitor, in the presence and absence of *miR-1*. We measured the cell viability with an alamarBlue assay after 24h and 48h of treatment. We determined the chemo sensitivity of the camptothecin treated cells with the vital dye alamarBlue and compared the emitted fluorescence signal of the metabolizing cells towards the non-treated *miR-1* transfected cells. Previous experiments showed that an overexpression of *miR-1* resulted in a reduced level of cell viability (Figure **3.34**). Furthermore, the colon cancer cell lines transfected with *miR-1* showed a higher sensitivity towards the chemotherapeutic agent camptothecin. Compared to the non-transfected cells we detected a reduced viability of 17% and 8% after 24 hours camptothecin treatment in SW480 and SW620 cells, respectively. A combination of camptothecin treatment and expression of *miR-1* further delivered a reduced cell viability of 28% and 25%, respectively. The effect was even stronger after 48 hours: Camptothecin treatment alone showed a cell viability reduction of 43% and 49% in SW480 and SW620, respectively, as compared to the controls. A combination therapy with *miR-1* further decreased the viability by 55% and 57%. The effect of *miR-1* overexpression by its own was shown in section **3.3.2**. Here we determined a reduction of cell viability of 25% in both cell lines after 48h of *miR-1* transfection. Thus, in both cell lines camptothecin and *miR-1* most like function synergistically, further demonstrating a tumor suppressor like function of *miR-1* (Figure **3.43**).

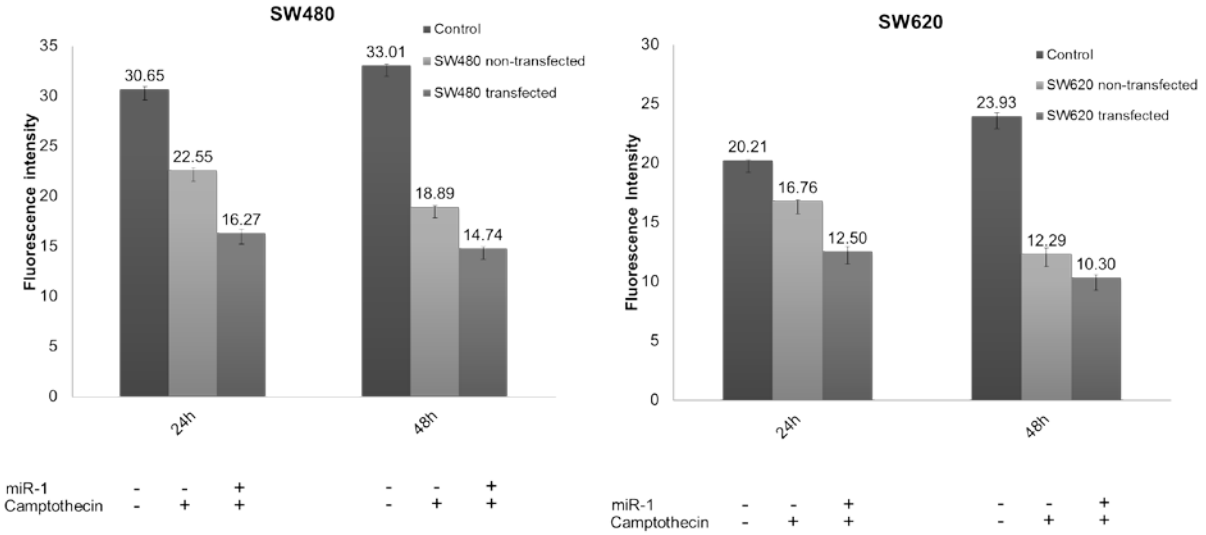

Figure 3.43: AlamarBlue cell viability assay with in SW480 and SW620 cells before and after camptothecin treatment and *miR-1* transfection. Cell viability was measured after 24h and 48h of drug treatment using a spectrophotometer.

3.3.6 *IN-SITU* DETECTION OF *MIR-1* IN COLON TISSUE

To analyze and visualize the localization of *miR-1* expression in colon tissues, we applied locked nucleic acid (LNA) modified probes for *in situ* hybridization (ISH) against the mature *miR-1* on paraffin fixed tissue sections. The use of LNA-probes is a standard method to improve the stability of the microRNAs and enables a higher sensitivity of microRNA detection [189,190]. In those modified probes the ribose ring is locked in a stable conformation achieved through a connection between the 2'-oxygen (O) and 4'-carbon (C) atom by a methylene bridge. The applied miRCURY LNA detection probes were double digoxigenin (DIG)-labeled, which should increase the specific signal compared to a single 5'-DIG labeled probe [191]. To get more advanced insight to various microRNA expression patterns in colon tissue, we investigated an additional microRNA, *miR-129*. This microRNA is similar to *miR-1*, down-regulated in the tumor. In order to evaluate the general protocol conditions and to optimize the sensitivity and performance, we first optimized the ISH protocol for the control small nuclear RNA U6. The staining for RNA U6 showed a general intense blue nuclear staining over the entire colon tissue section. An improved U6 staining does not necessarily indicate for an optimized mature microRNA staining. Therefore *in situ* conditions for a specific detection of *miR-1* and *miR-129* were adapted to the common protocol by adjusting different parameters such as hybridization temperature and proteinase K digestion. Formalin-fixed-paraffin embedded (FFPE) tissue samples used for ISH were generated from serial sections of human colon. With microscopic observation and 100x magnification, the DIG-labeled probes of *miR-1* and *miR-129* stained the cytoplasm of different cell types in the colon tissue, such as epithelial cells in the crypts of the colon mucosa and cells of the lymph follicle, which are located in the submucosa. For *miR-1* we could also detect a weak signal in the cytoplasm of cells in the lamina propria and muscularis mucosae. We also received a signal for cells that belong to the primary colon tumor. However this signal seems to be weaker than for the normal epithelial layer (Figure **3.44**). Unfortunately the difference is not very strong and further optimizations of the protocol are required to investigate the appearance of *miR-1* and *miR-129* in normal and tumor tissue areas. The malignant area of the consecutive sections was determined by histological observations together with Prof. Dr. M. Hummel (Dept. of Pathology, Charité).

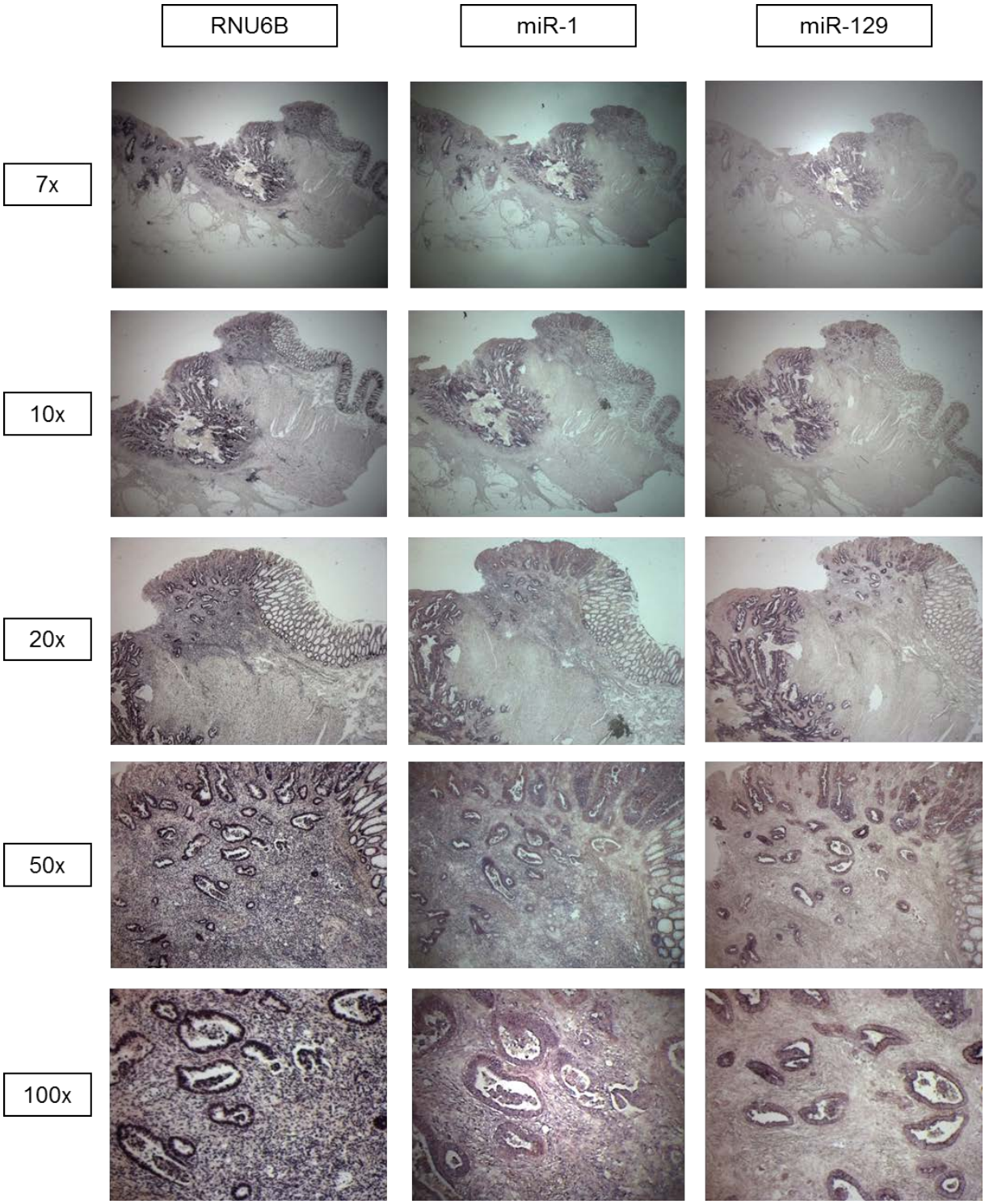

Figure 3.44: In-situ hybridization of *miR-1*, *miR-129* and *RNU6B* as control were performed on FFPE colon tissue sections. Different magnifications (7x, 10x, 20x, 50x and 100x) were documented by microscope technology.

3.3.7 MODELING OF THE DEREGULATED *MIR-1* AND IMPACT ON THE EXPRESSION OF CANCER ASSOCIATED GENES

As shown previously, *miR-1* and *miR-135b* in combination suffice to distinguish normal and malignant tissues. In addition, *miR-1* is constantly down-regulated in many different tumor entities, thus indicating a statistically significant co-incidence between *miR-1* down-regulation and tumor development. Using the PyBioS modeling [192-194] and simulation software, a prognosis of gene expression based on the deregulation of *miR-1* was tested. This molecular interaction modeling system contains more than 2718 genes. In addition, the model integrates information about microRNAs and their regulated targets [195]. Information for the microRNA interactions and activities were extracted from the miRBase database. For *in silico* simulation, the model was set to a basic state and ensuing from this point the model was initiated by mRNA expression data generated by next generation transcriptome sequencing experiments applied for individual colon cancer patients. Therefore we used the expression data of normal, tumor and metastasis tissues of 4 different colon cancer patients (P3, P4, P5, P7), all of them arose from the set of colon cancer patients we already had sequenced for smallRNA analyses (Table **A.11**). For each tissue, the basic state was modified based on their corresponding mRNA expression data, defined as N_0, T_0 and M_0 state, and an overexpression and reduction of *miR-1* expression level was simulated (initialized state). Relative to the basic state and the initialized state N_0, T_0 and M_0, the degrees of variances for different *mir-1* concentrations were calculated for the tumor and metastasis tissues, determined as T_1 or M_1, visualized for each component of the *in silico* model (Figure **3.45**).

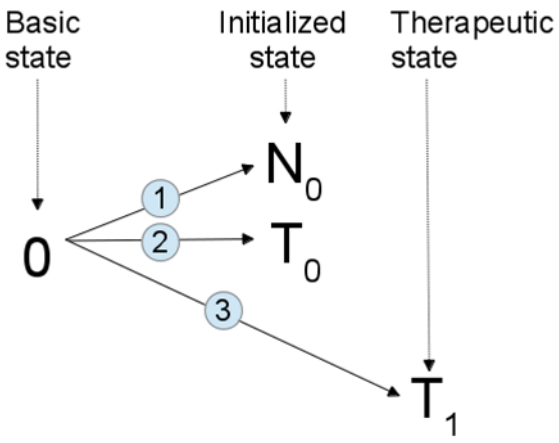

Figure 3.45: Schematics of the modeling approach. The scheme shows the modification of normal (N) and tumor (T) tissue from their basic state into an initialized state applying mRNA expression data for individual colon cancer patients.

The simulation results were determined for each patient and for tumor and metastasis tissue separately. Analyzing the results, a scheme was developed to summarize the changes in gene expression after changing *miR-1* expression in the modeling system. Therefore we divided the

alterations into five different groups, each of them was subdivided by the direction of change (‘plus’ and ‘minus’): ‘desired’, ‘flip effect’, ‘aggravating’, ‘side effect’ and ‘weak effect’ (Figure **3.46**).

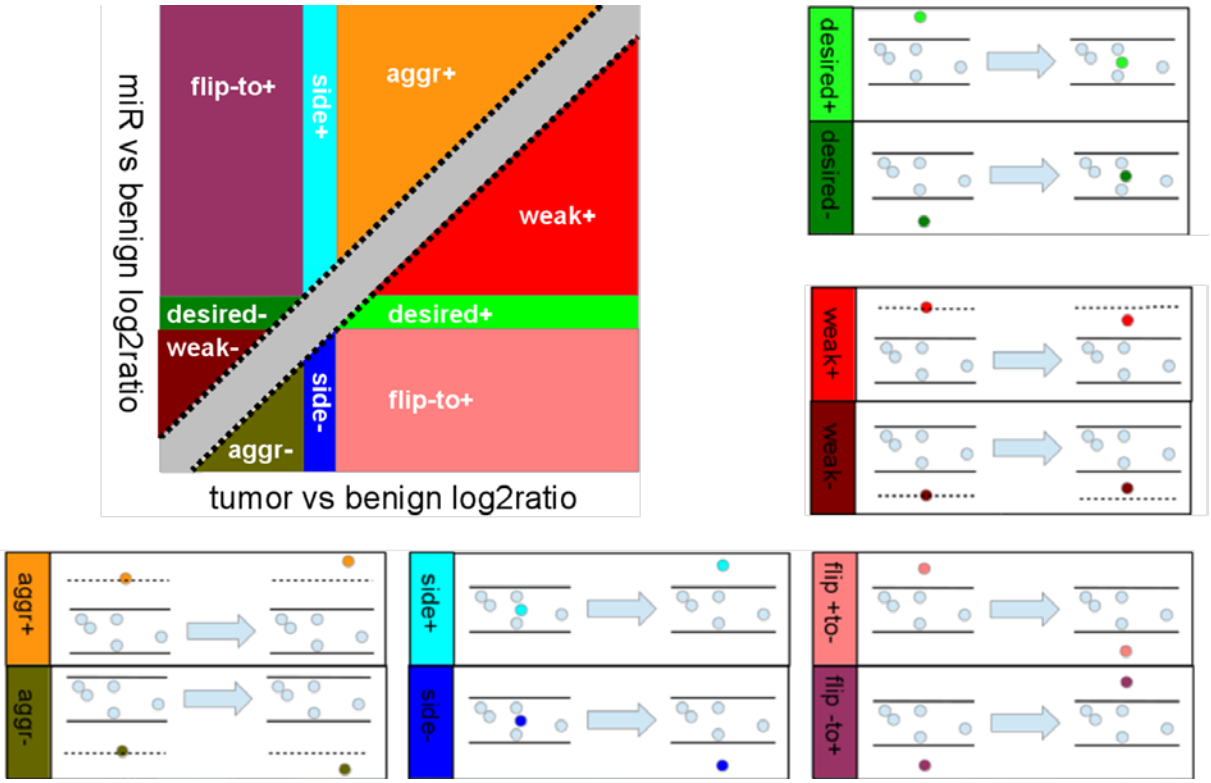

Figure 3.46: Simulation of modeling effects. Log2ratios were calculated from the modeling output for tumor vs. normal (T-N) and any *miR-1* treatment vs. normal (miR-N). The predicted component concentration changes were divided into five different groups (or 10 different sub-groups) depending on the changes between the T-N and *miR-1* log2ratios: Desired (desired therapeutic effect, component concentration levels back to ‘normal’), weak effect (weak effect in changing the component concentration but tendency to ‘normal’), side effect (negative therapeutic effect), aggravated (component concentration aggravated in the same direction), flip (component concentration status flips from up- to down-regulated or vice versa). Each group can be divided into two sub-groups ‘plus’ and ‘minus’ which means an up- or down-regulation of a component determined by a log2ratio cutoff greater than 0.58 or smaller than -0.58, respectively.

Genes whose expression after *miR-1* treatment returned to no significant deregulation were defined as ‘desired’. The log2ratio of those genes lay between 0.58 (fold change = 1.5) and -0.58 (fold change = -1.5). If the expression of genes changed beyond that defined level after altering the *miR-1* concentration they were determined as ‘side effect minus’ or ‘side effect plus’. Genes with no significant changes in their expression after a *miR-1* increase or decrease were specified as ‘weak effect’ either with a tendency to a more decreased expression (‘weak minus’) or with a more increased expression (‘weak plus’). This classification is determined by a log2ratio cutoff greater than 0.58 or smaller than -0.58. Gene expressions that became more pronounced were defined as ‘aggravate plus’

or ‘aggravate minus’. If gene expression changed to the contrary, it was calculated as ‘flip plus to minus’ or ‘flip minus to plus’ (Figure **3.46**).

Within this *in silico* model we first simulated a 1.01-, 1.1-, 1.2-, 1.4-, 1.6-, 1.8-, 2-, 4-, 10-, 20- and 100fold down-regulation of *miR-1* and determined the frequency of genes that changed their expression compared to the initiated model. Interestingly, we found for the majority of genes an amplification of their over-expressed state. In particular, the number of genes that were calculated as ‘aggravated plus’ for tumor and metastasis tissues of all patients was strongly increased by further decreasing the *miR-1* level. Approximately 20 to 40% of the genes turned into this ‘aggravated plus’ state after a 1.2-fold decrease of *miR-1* expression level. Many genes also showed a ‘flip minus to plus’ in their expression (Figure **3.47**).

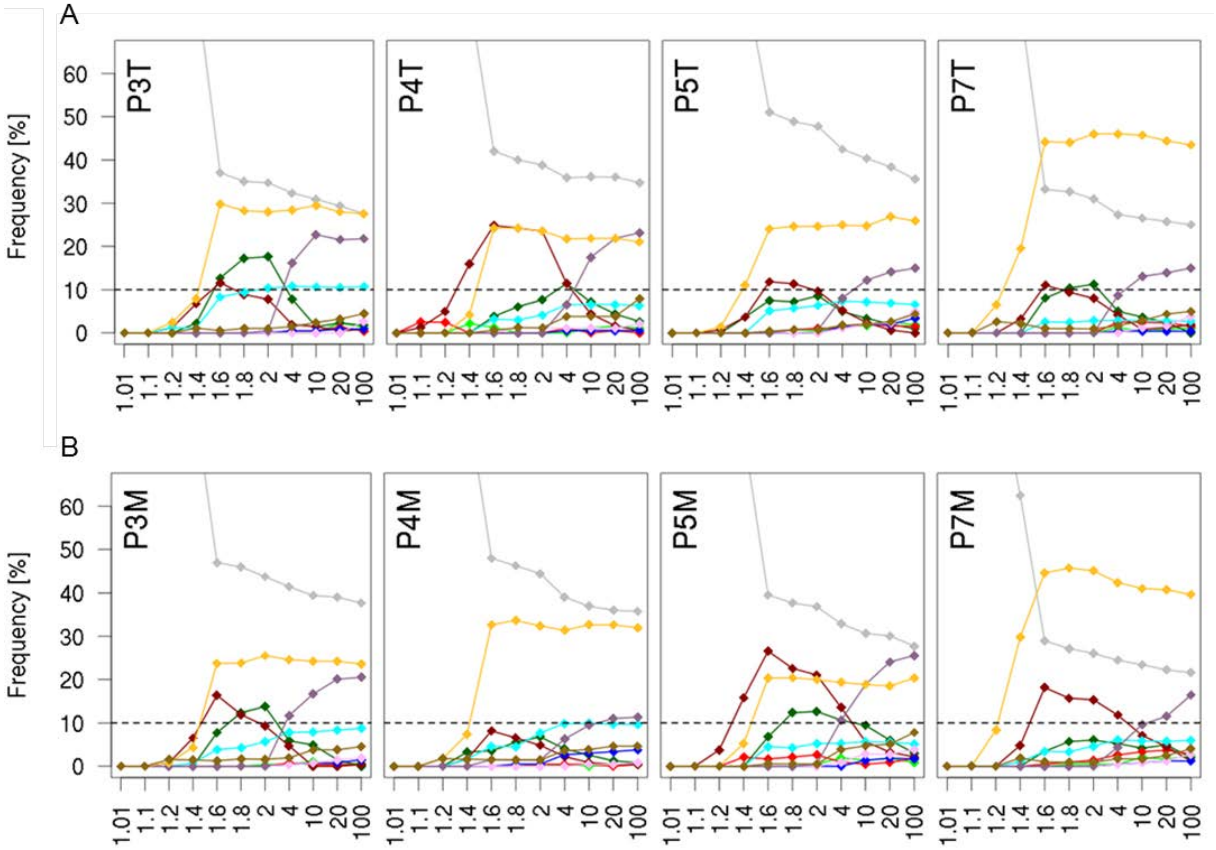

Figure 3.47: Simulation of gene regulation in dependence of *miR-1* down-regulation (1.01-, 1.1-, 1.2-, 1.4-, 1.6-, 1.8-, 2-, 4-, 10-, 20- and 100-fold) in (**A**) tumor (T) and (**B**) metastasis (M) tissues for each patient (P3, P4, P5 and P7). The frequency of genes is plotted for different alteration groups, such as, ‘desired plus’, ‘desired minus’, ‘weak plus’, ‘weak minus’, ‘aggravate plus’, ‘aggravate minus’, ‘flip plus to minus’, ‘flip minus to plus’, ‘side effect plus’ and ‘side effect minus’, regarding different *miR-1* concentrations.

To investigate the potential therapeutic effect of *miR-1* overexpression we simulated an increase in *miR-1* levels by changing the initial state of *miR-1* concentration for tumor (T_0) and metastasis (M_0) up

to concentration changes of 1.01- to 100-fold overexpression relative to the normal (N_0) tissues. Overall 20 to 30% of the genes were grouped into 'aggravated minus' after a 1.2 to 1.4-fold increase of *miR-1* expression in both tissues, except P4M. We could also determine many genes that were grouped into the 'weak plus' group after 1.6-fold increase of *miR-1* over all patients and tissues. Especially for P7 approximately 40% of the genes changed into this group. In general, an optimal therapeutic effect was determined between 1.6 and 1.8fold increase of *miR-1*, albeit at a very low level. A higher frequency of genes could be also determined for the 'desired plus' group for P3, P4, and P5 in tumor and metastasis tissues (Figure **3.48**).

Extracting only cancer relevant genes from the cancer gene census list [196], tumor suppressor and oncogenes, we found approximately 30 to 50% of the genes in the 'aggravate plus' group after a decreased concentration of *miR-1* for each patient in tumor and metastasis. Many genes were also determined as 'side plus' (Figure **3.49**). This corresponds to the observation we made for the entirety of genes.

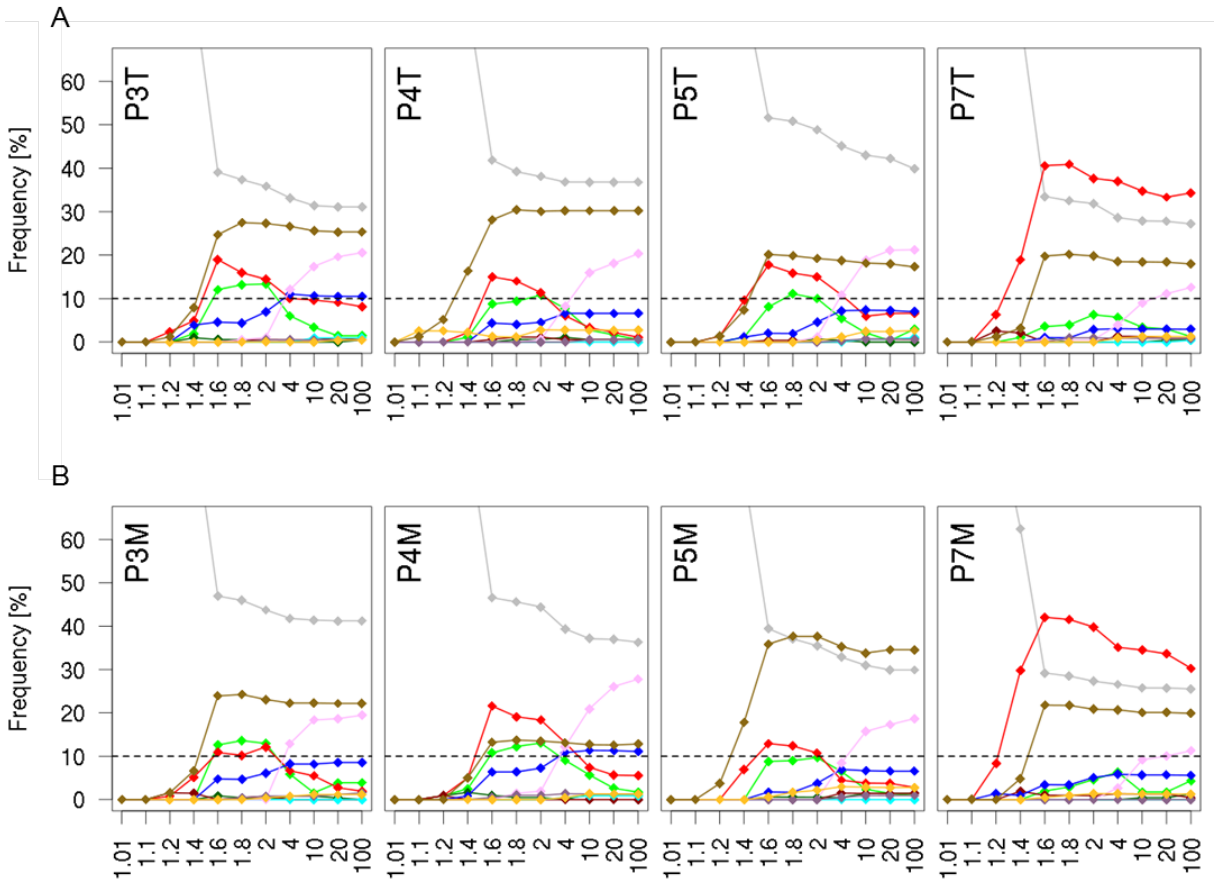

Figure 3.48: Simulation of gene regulation in dependence of *miR-1* up-regulation (1.01-, 1.1-, 1.2-, 1.4-, 1.6-, 1.8-, 2-, 4-, 10-, 20- and 100-fold) in (**A**) tumor (T) and (**B**) metastasis (M) tissues for each patient (P3, P4, P5 and P7). The frequency of genes is plotted for different alteration groups, such as, 'desired plus', 'desired minus', 'weak plus', 'weak minus', 'aggravate plus', 'aggravate minus', 'flip plus to minus', 'flip minus to plus', 'side effect plus' and 'side effect minus', regarding different *miR-1* concentrations.

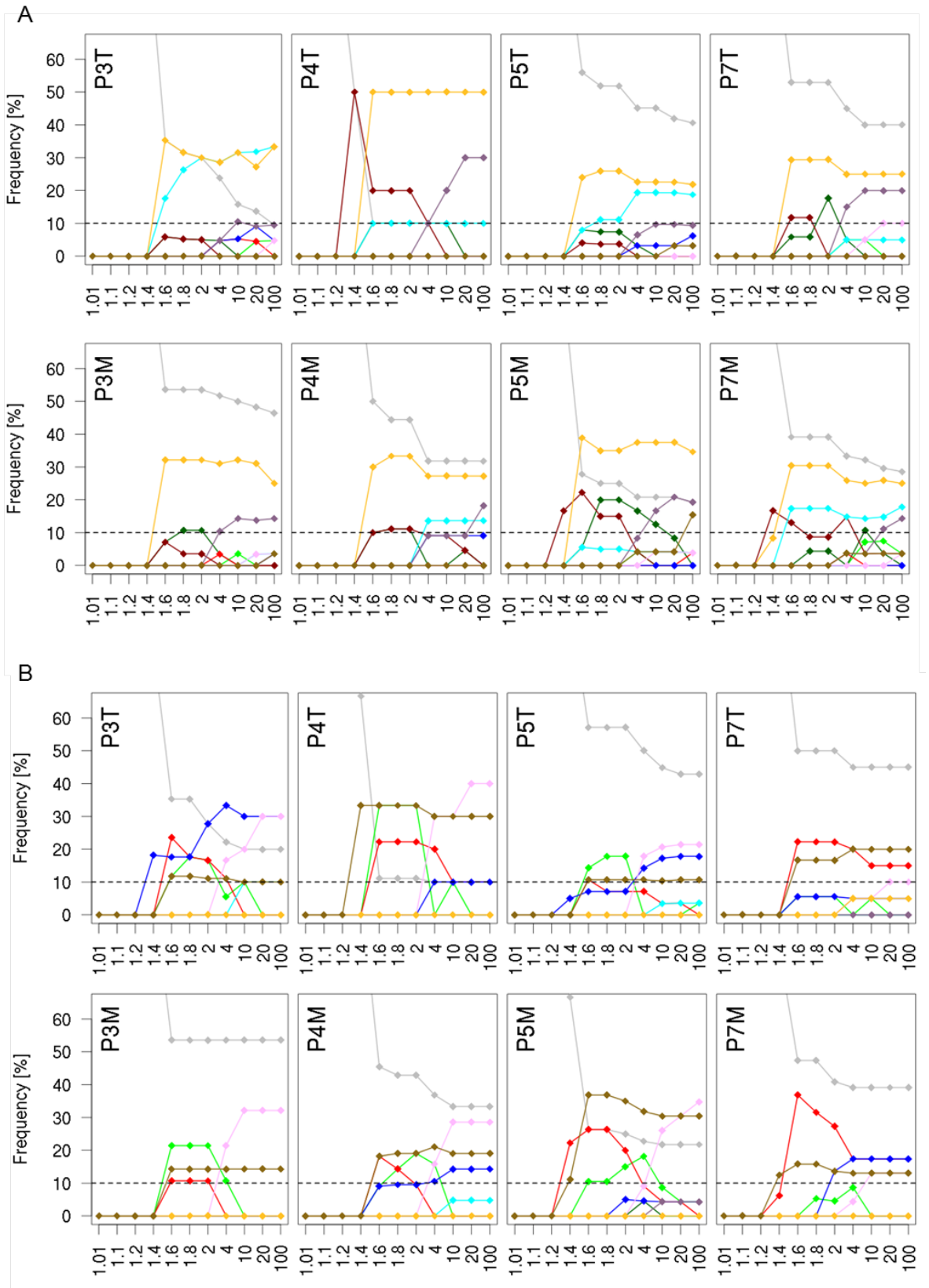

Figure 3.49: Simulation of regulation for cancer relevant genes in dependence of (**A**) *miR-1* down- and (**B**) up-regulation (1.01-, 1.1-, 1.2-, 1.4-, 1.6-, 1.8-, 2-, 4-, 10-, 20- and 100-fold) in tumor (T) and metastasis (M) tissues for every patient (P3, P4, P5 and P7). The frequency of genes is plotted for different alteration groups, such as, 'desired plus', 'desired minus', 'weak plus', 'weak minus', 'aggravate plus', 'aggravate minus', 'flip plus to minus', 'flip minus to plus', 'side effect plus' and 'side effect minus', regarding different *miR-1* concentrations.

The analysis of cancer relevant genes for an overexpression of *miR-1* delivered a higher frequency of genes grouped into 'desired plus' and 'weak plus' resulting in positive alterations in both tissues. However, we also determined a fraction of genes affected by *miR-1* de-regulation for the groups 'flip plus to minus', 'aggravate minus', 'weak plus' and 'side minus' (Figure **3.49**).

3.4 ANALYSES OF UNKNOWN MICRORNAS IN COLON CANCER CELLS

Next generation sequencing not only enables the sequencing of millions of RNA, microRNA and DNA fragments from the same samples, it also makes it possible to discover novel microRNAs. For the discovery and validation of novel microRNAs, we chose colon cancer cell lines (SW480 and SW620), which can provide a vast amount of RNA for further studies such as northern blot analyses to validate potential novel microRNA candidates. For microRNA profiling by deep sequencing in general, the obtained raw sequences were first filtered for the usable reads, excluding adapter sequences and non-determinable sequences. The reads which align to known microRNA precursor sequences, annotated in miRBase, and other small RNAs were compiled for profiling analyses. The remaining unknown sequences that do not map to known microRNAs are used to search for potential novel microRNAs. The first analytical step in the discovery of novel microRNAs includes the mapping of RNA reads towards the genome. The second step is to investigate their secondary RNA structure to form a potential microRNA precursor. The algorithm which was used to analyze novel microRNAs was adapted from the miRDeep algorithm which was published by Rajewski and Friedländer in 2008 [197]. The prediction of the algorithm is based on the biogenesis of microRNAs, which are processed by Dicer. This processing step generates defined products while cleaving precursor microRNAs. The processing of hairpin creates three products: the mature microRNA sequence, the star sequence and the loop structure. The sequenced RNA reads were mapped to the genome and the candidate microRNA hairpin structures were determined. It is expected that the majority of reads correspond to the mature microRNA, resulting in a characteristic picture of a mountain and a flat side (Figure **3.50**). Based on different features, such as a likely microRNA precursor structure, read position and frequency, as well as energetic stability, an algorithm score was determined for each potential candidate. This scoring can achieve a maximum value of 1, which represents a 100% percent probability for a novel microRNA. For the two cell lines, SW480 and SW620, we could identify 39 potential novel mRNAs. The classification of those candidates was scored from 0.51 to 1, with 20% that showed a score of more than 0.9. Out of these potential new microRNAs, we chose five different candidates for validation experiments with northern blots to confirm their expression. Different features were taken into account for their selection. Three of the microRNAs were selected by their scoring, which ranged from 0.99 to 1. Furthermore, two of those novel candidates on chr19:44650704-44650771 and chr4:90872081-90872164 showed a relatively high scoring and a high expression in at least one cell line (Table **A.12**).

As an internal control, we used the small nuclear RNA U6, which should be expressed ubiquitously. For all 5 candidates, we visualized the RNA read distribution via the UCSC genome browser (hg18)

and estimated the frequency of the microRNA Dicer products along their potential precursor sequence. We also checked their genome localization to determine their exact localization in predominantly intronic regions. For both data sets (set1 = SW480 and set2 = SW620) we received reads for all of the five candidates over the microRNA precursor sequence (Figure **3.50**).

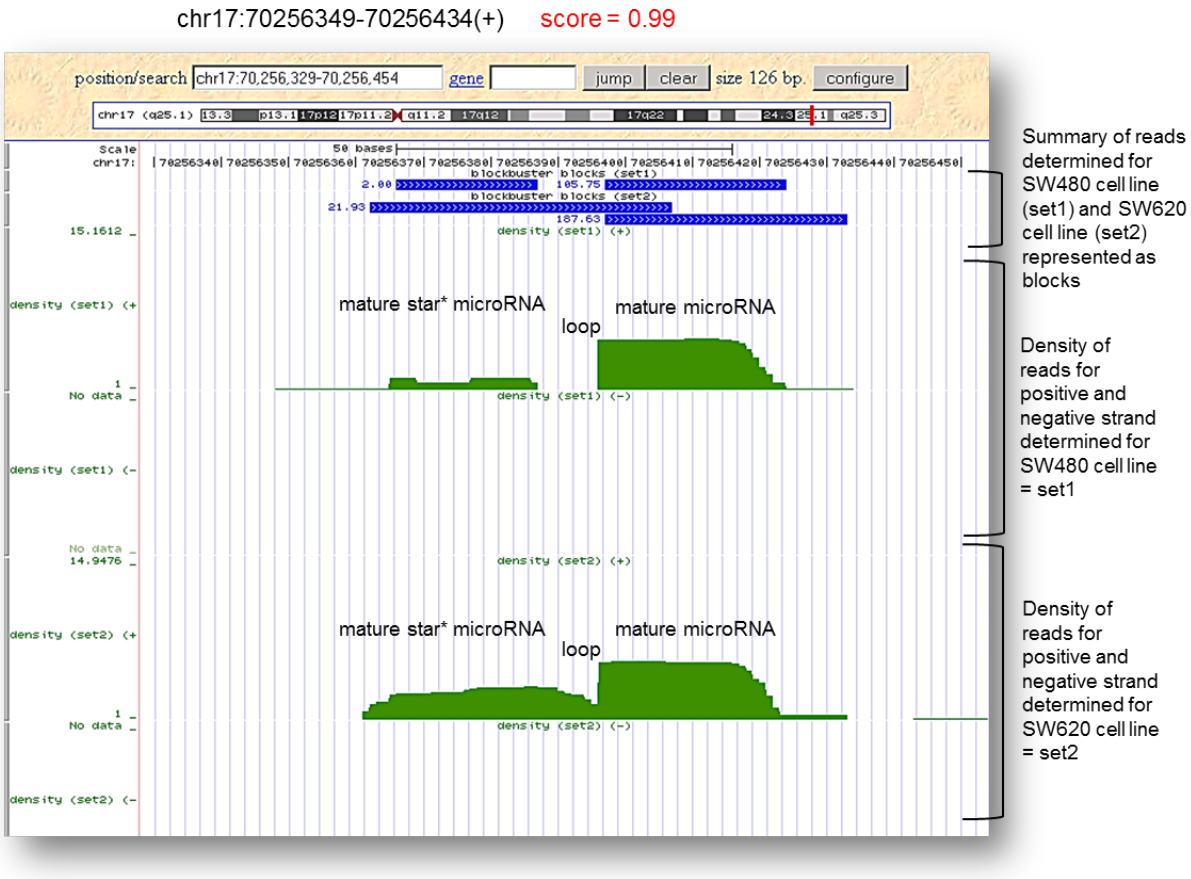

Figure 3.50: Visualization of RNA read density and distribution for a novel microRNA candidate located on chromosome 17 for cell lines (set1 = SW480 and set2 = SW620), density of reads is represented in green and summary of reads is represented as blue blocks for every set.

As a proof-of-principle study we performed northern blot analyses for all of the selected five microRNA candidates. Due to technical difficulties we could validate only one of the candidates on chromosome 17 (chr17:70256349-70256434(+)) (Figure **3.51**). Northern Blotting revealed bands which corresponded to the mature microRNA product at approximately 30bp and for primary and precursor microRNA sequences between 70 and 100bp. Interestingly, for the microRNA candidate located on chromosome 17, we could determine a higher expression in SW620 cell lines using the NGS technology. The northern blot analysis showed a similar result. Here we detected a stronger signal for the microRNA in the lane of the SW620 RNA (Figure **3.51**).

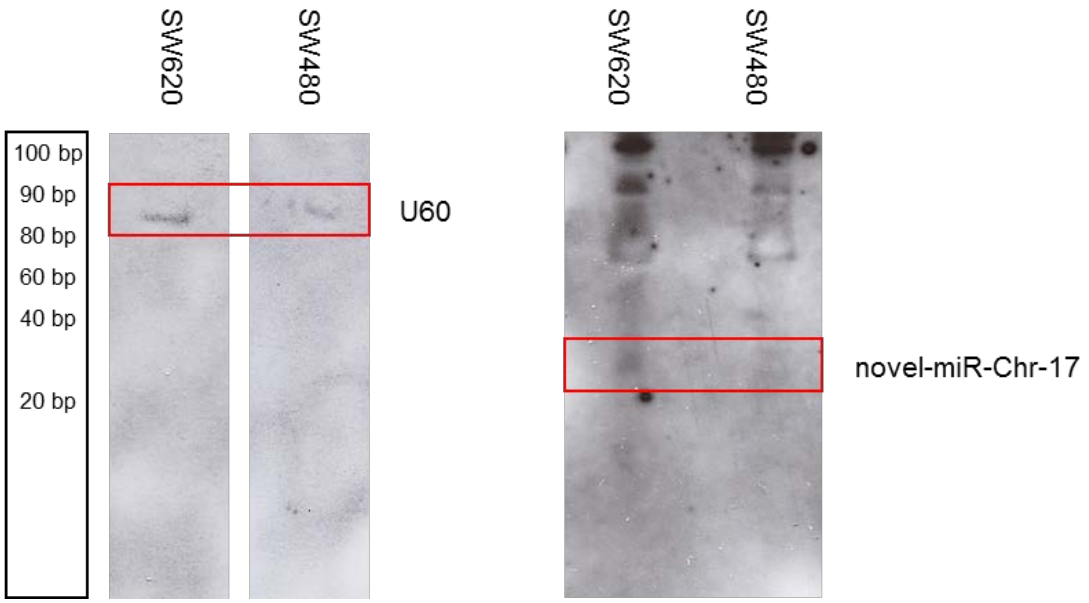

Figure 3.51: Northern blot analysis for a novel microRNA candidate located on chromosome 17 (chr17:70256349-70256434) with digoxegenin (DIG)-labeled probes. As loading control U60 RNA was used.
Figure 3.8:

To prove the secondary structure of this potential candidate we applied an online available RNA structure prediction tool 'CentroidHomfold' provided by www.ncRNA.org (Figure **3.52**). The sequence of the novel microRNA on chromosome 17 (chr17:70256349-70256434) was uploaded into the CentroidHomfold software and an RNAfold output was shown in several ways, as input sequence with standard base-pair symbolizations and as a secondary structure graph, where each base-pair is color coded with a heat color gradient, which represents a base-pairing probability via hydrogen bonds from 0 to 1. A hair-pin structure is recognizable for this novel microRNA, with a consensus sequence in the stem, which represents mature and mature star* sequences with a very high probability of binding through hydrogen bonds and a loop structure between those parts. The loop is segmented in smaller loop structures, but the probability of the binding in this area is not very high, which led to the assumption that this segment could function as the terminal loop-structure of a typical microRNA. Another indication is the 3'-end overhang at the end of the microRNA precursor structure, which is known to be a typical Dicer processing product. These results led to the assumption that this potential smallRNA, which was identified by NGS technology, might be indeed a novel microRNA.

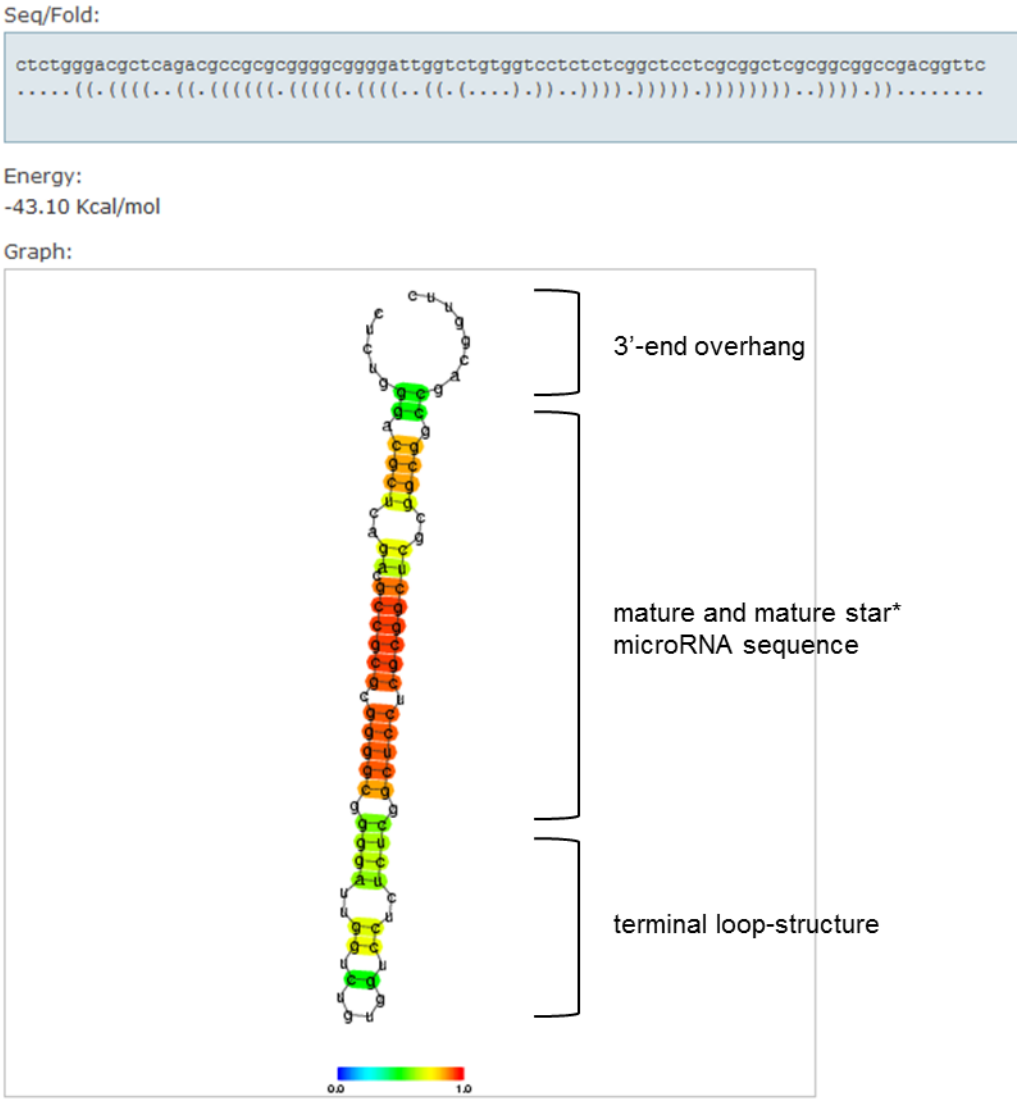

Figure 3.52: MicroRNA structure prediction for the novel microRNA localized on chromosome 17 (chr17:70256349-70256434). The base-pairs of the predicted sequence are color coded between the values 0 and 1, representing the base-pairing probability (www.ncRNA.org).

3.5 IDENTIFICATION OF MUTATIONS IN MICRORNA REGIONS

Single-nucleotide polymorphisms (SNPs) can be present in microRNA-binding sites and may contribute to cancer susceptibility while modifying the binding capacity of microRNAs to their mRNA targets. MicroRNAs contain a 7-8 nucleotide segment in length, the so called 'seed' region, which is highly complementary to a corresponding region located in the 3'UTR of specific mRNA targets. This region is essential for target specific binding and therefore a SNP in this region would affect microRNA-target interactions and, consequently, expression of those genes [198,199]. For that reason, we searched for SNPs located in microRNAs regions, especially in the seed section of those non-coding RNAs. We sequenced the normal and tumor tissues of six different colon cancer patients (P9N, P9T; P13N, P13T; P15aN, P15aT; P31N, P31T, P32N, P32T; P33N, P33T (Table **A.11**) and determined unknown and known SNPs in both tissues. In addition, we examined the single-nucleotide variations in

representative colon cancer cell lines, SW480 and SW620 and compared the occurrence of SNPs to the patients material. We found 65 different SNPs localized in microRNA regions in the six colon cancer patients. Out of these 65 different variations, 27 SNPs were already annotated in hg18, and 38 were not annotated. The 38 unknown SNPs were located in 28 different microRNAs (Table **A.13**). In the cell lines, we could only find 14 of those SNPs present in the patients, eight of them already known and six not yet annotated. The eight annotated SNPs reside in *miR-612* (rs12803915), *miR-492* (rs2289030), *miR-300* (rs12894467), *miR-663* (rs7266947, rs28670321), *miR-577* (rs34115976), *miR-146a* (rs2910164) and *miR-1304* (rs2155248). The locations of the six unknown SNPs included *miR-1324* and *miR-663b*, with four unknown SNPs present in *miR-663b* (chr2:132731110-132731110, chr2:132731082-132731082, chr2:132731072-132731072, chr2:132731050-132731050) and two in *miR-1324* (chr3:75762663-75762663, chr3:75762648-75762648) (Table **3.8**). These unknown variations localized in *miR-663b* and *miR-1324* were also present in cell line material. None of the identified SNPs were predicted as somatic mutations. They all could be found in normal and tumor tissue of the investigated colon cancer patients.

Table 3.8: List of unknown colonic SNPs identified in microRNA regions in cell lines and CRC patients. The presented variations were determined as germline mutations, present in normal and tumor tissue of colon cancer patients.

Gene name	***Location***	***Reference***	***Mutation***
miR-663b	chr2:132731110-132731110	C	T
miR-663b	chr2:132731082-132731082	A	C
miR-663b	chr2:132731072-132731072	G	C
miR-663b	chr2:132731050-132731050	C	T
miR-1324	chr3:75762663-75762663	A	G
miR-1324	chr3:75762648-75762648	C	T

3.6 MUTATION ANALYSES IN CRC PATIENTS APPLYING RE-SEQUENCING TECHNOLOGIES

For mutation analyses we used two different enrichment approaches on two different CRC sample sets, both of them included MSI and MSS patients (Table **3.9**). For the first sample set, which consists of two patients (patient 1 and patient 2) with a high grade adenocarcinoma (G3) and a proximal localization of the tumor, we sequenced the complete exome of more than 135000 exons applying the microarray-based-genomic selection technology developed by Roche NimbleGen in 2007 and the 454 sequencing technology (Table **A.14**). Patient 1 showed a microsatellite instable tumor, whereas, the microsatellite status of patient 2 was determined to be stable. Both patients provided only normal and tumor tissues. The second sample set, which consisted of two patients which achieved the tumor-node-stage IV of CRC is characterized by a more advanced cancer that metastasized into distant sites. This provided us the possibility to analyze metastasis tissues as well. Here we used the multiplex exon capture approach, established by Shendure and Church et al in 2007 [60], based on the application of molecular inversion probes. The two metastasized patients could be further separated into microsatellite stable (patient 3) and microsatellite instable (patient 4). For this sample set we

captured 54105 different exons and 14407 genes with 55460 targeting probes and sequenced the enriched fraction using the Illumina sequencing technology (Table **A.14**).

Table 3.9: Colorectal cancer patients and summary of clinical informationand enrichment technology selected for mutation analyses (MS = microsatellite, MSI = microsatellite instable; MSS = microsatellite stable).

	Sample set 1		***Sample set 2***	
Patient	patient 1	patient 2	patient 3	patient 4
Age	59	65	70	80
Gender	male	male	male	female
Histological grading	G3	G3	-	-
Staging	-	-	TMN IV	TMN IV
Localization	proximal CRC	proximal CRC	Sigmoid	Colon
MS status	MSI	MSS	MSS	MSI
enrichment technology	microarray-based genomic selection method		multiplex exon capture method	

The identification of single-nucleotide variations in coding regions was one cause for applying re-sequencing technologies to selected genomic regions, such as the whole exome on the sample sets of CRC patients. We received a higher number of SNVs for the MSI patients than for the MSS patients. We found for the first sample set 12767 variants for the MSI CRC patient in tumor tissue and 10622 variants in the MSS patient. Out of these SNVs a range of 2404 (MSI patient) and 1288 (MSS patient) variations showed a low prevalence, determined by a minor allele frequency of less than 1% or by the fact that they had not yet been annotated in the dbSNP or the 1000 Genomes Project. As somatic mutations we determined 1304 variations in the MSI and 198 somatic variations for the MSS CRC. To predict the probability and functional consequence of a damaging effect in proteins affected by coding somatic variations, it is possible to apply functional prediction algorithms, such as, Polyphen and MutationTaster. These mathematical tools provide the possibility to select for genes with an amino acid change that contributes destructive mutations. Using this approach we identified, unknown somatic SNPs with a predicted damaging effect in receptor proteins, in kinases and in genes of the repair system. For sample set1 we determined 359 variants that were predicted to alter functions for the MSI patient and 45 for the MSS patient (Table **3.10**). Consequently we found a general higher mutation level in the MSI compared to the MSS CRC and an approximately eight times higher coding somatic mutation rate with the ability to alter gene function for the MSI colon cancer patient, using the array based enrichment technology and 454 sequencing [40].

To confirm this observation we also analyzed enrichment data based on a multiplex exon capture method for the second sample set of CRC patients (patient 3 and 4). In addition this set enabled us to search for metastases specific somatic mutations. Applying the multiplex exon capture re-sequencing technology to capture 54105 different exons, we found 2227 variations for the MSS patient (patient 3) in coding regions in the tumor and 46 additional variations only present in the metastasis sample. Investigating low prevalence SNVs we detected 416 tumor and additional 42 metastases specific

variations and determined 111 to be somatic tumor specific variations and 15 additional to be somatic metastases specific. Analyzing the MSI CRC patient we identified 3226 variations in coding regions in the tumor and 127 additional variations only present in the metastases. As low prevalence SNVs we identified 1094 tumor and 124 additional metastases specific variations and determined 359 to be somatic variations in the tumor and 58 additional variations to be somatic in in the metastasis. Searching for variants in MSI and MSS colon cancers we identified a rate 3 times higher for somatic mutations in the MSI compared to the MSS CRC, using the exon capture re-sequencing technology. For the metastases specific somatic mutations we observed an even higher increased level of approximately 4 times more mutations. Using different enrichment and sequencing technologies, we were able to observe a higher number of somatic mutations for the MSI patients than for the MSS patients. Hence, the increased mutation rate in the MSI CRC was a reproducible observation (Table **3.10**).

Table 3.10: Determined mutations in two sample sets for MSI and MSS colorectal cancers. For the mutation analyses of both sample sets, different enrichment technologies were applied, microarray-based genomic selection technology for sample set 1 and multiplex exon capture method for sample set 2 (MS = microsatellite, MSI = microsatellite instable; MSS = microsatellite stable). Metastasis counts depict the mutations not found in the corresponding tumor tissue.

	Sample set 1		***Sample set 2***			
patient	patient 1	patient 2	patient 3		patient 4	
tissue	tumor	tumor	tumor	metastasis	tumor	metastasis
MS status	MSI	MSS	MSS		MSI	
variants in genic regions	49,484	46,247	2,237	49	3,254	130
variants in coding regions	12,767	10,622	2,227	46	3,226	127
unknown variants with low prevalence	2,404	1,288	416	42	1,094	124
somatic variants	1,304	198	111	15	359	58
variants predicted to alter functions	359	45	65	10	190	38

4 DISCUSSION

4.1 ANALYSIS AND IDENTIFICATION OF MICRORNAS AS BIOMARKER CANDIDATES IN COLON CANCER AND IDENTIFICATION OF UNKNOWN MICRORNAS

In the past decade, the development of next generation sequencing (NGS) technologies made it possible to sequence DNA in a massively and parallel manner and it enabled us to address different genetic questions from different fields of molecular genetics. Profiling microRNAs by next generation sequencing allows investigation in their entirety including the discovery of novel microRNAs and other smallRNA species. Compared to microarray applications the deep sequencing technology is not restricted by a certain array content or by the abundance of previously discovered microRNAs. Therefore we applied specific bioinformatics algorithm to detect novel microRNAs in colon cancer cell lines. Altogether we identified 39 potential microRNA candidates which were assessed by a classification score. Of those 39 candidates we selected 5 potential microRNAs which were analyzed by northern blot analyzes and at least one candidate could be validated by this method. This microRNA exhibited a classification score of 0.99, indicating for the strong probability to represent a new, previously unidentified, microRNA. The genomic localization (chr17:70256349-70256434(+)) of this novel microRNA candidate delivered an exonic localization. Normally microRNAs are found within intron regions of protein-coding genes, but in some cases they are located in exons of host genes [200]. Applying an RNA-fold algorithm on the precursor sequence of the selected novel microRNA candidate delivered a proper hair-pin structure. Taken together, those findings strengthen the integrity of the novel microRNA candidate. However, compared to the expression level of annotated microRNAs in our data, this candidate exhibited a modest expression (fold change = 0.51), perhaps indicating that it might not form substantial functions in these cell types. The potential biological function should be investigated in further experiments.

The ability to discover novel microRNAs is only one advantage of the next generation sequencing technology. For the purpose of this work we were concentrated on known microRNA candidates to identify biomarkers in colorectal cancer. Sequencing the smallRNA fractions is one application which can help to identify a range of oncogenic and tumor-suppressing microRNAs that are expressed genome-wide. However, grouping cancer-associated microRNAs only into oncogenes and tumor suppressor genes might be too simple. MicroRNAs have a more global effect on protein output, where the concept of ‘large-scale fine tuning’ might be the best definition for their role as posttranscriptional regulators. Recent investigations demonstrate not only the ability of microRNAs to inhibit gene expression and protein synthesis, they also show the potential to up-regulate protein translation while interacting with their 5’UTR. This is in particular the case for ribosomal proteins [201]. Apparently, microRNAs bear a pleitropic effect on cell regulation and understanding the comprehensive function of microRNAs in the tumorigenesis of cancer offers the valuable opportunity to identify novel molecular biomarkers for cancer prevention and treatment. Studies highlight the promising diagnostic and prognostic potential of microRNAs in colorectal cancer [147]. In addition, identification of mRNA targets associated with different colorectal carcinogenesis pathways furthermore provides the potential for

novel therapeutic targets. For improving the potential therapeutic value of microRNAs it could be one approach to combine them with chemotherapeutic drugs, and to study their potential synergistic effect with microRNAs.

Preforming genome-wide analyses of microRNA expression on a set of colon cancer patients and colon cancer cell lines has delivered a range of interesting microRNA candidates, deregulated in tumor and metastases. Interesting candidates, such as *miR-1*, *miR-129*, *miR-215*, *miR-145*, *miR-135b*, *miR-183*, *miR-31*, *miR-96*, *miR-21* or *miR-497* are identified as tumor and metastases specific microRNAs that showed significantly down- or up-regulation in both malignant tissues. Validation experiments performed by TaqMan technology delivered a solid concordance for the NGS generated data, where we achieved conformity of 0.89 pearson's correlation.

MicroRNAs *miR-1* and *miR-135b*, both were determined as top deregulated candidates in tumor and metastases, become the focus of our attention. While *miR-1* is significantly down-regulated in tumor tissues and cell lines we found *miR-135b* is up-regulated in colon cancer. *MiR-135b* has been already described in the context of colorectal cancer (CRC) by Nagel et al. [110], Bandres et al.[162] , Wang et al [202] and Necela et al. [203] . They described a novel genetic mechanism of *APC* (adenomatous polyposis coli) regulation, controlling the Wnt signaling pathway which depends on the expression levels of *miR-135b* and *miR-135a*. An up-regulation of this microRNA showed a negative correlation to the *APC* expression *in vitro*. Using GFP-sensor vector system in HeLa cells for *miR-135a&b* and in stably expressing *miR-135b* HEK-293 cells they identified a direct regulation of the expression of *APC*. A corresponding reduction in *APC* mRNA levels was detected by qRT-PCR experiments for *miR-135a&b*. Co-transfection experiments with double mutants for *miR-135a&b* binding sites revealed the direct suppression of *APC* by *miR-135a&b*. Using a functional genetic screening approach, they showed also a potent effect of *miR-135a&b* on the Wnt pathway activity and determined *miR-135a&b* as highly expressed in colorectal adenomas and carcinomas, with reduced levels of *APC* which leads to the assumption that alterations in the *miR-135a&b* expression contributes to colorectal cancer pathogenesis. Less is known about the contribution of *miR-135b* to the progression of metastases. Since *miR-135b* showed an even more increased expression in metastases tissue of our CRC patients as compared to the primary tumor, it aids the assumption that *miR-135b* could play a critical role in further stages of colon cancer. The gene *SMAD5* (mothers against decapentaplegic homolog 5) is validated as a target of *mir-135b* [204] and can activate downstream *TGF-β* (transforming growth factor beta) gene transcription which could be an indication for the potential role of *miR-135b* in more advanced colon cancer stages, such as metastasis progression. The TGF-β protein controls proliferation, cellular differentiation and plays a critical role in immunity as a cytokine and in oncogenesis as an antiproliferative factor [205,206]. In cancer cells TGF-β signaling pathway is influenced and causes immunosuppression and angiogenesis, which results in a more invasive cancer [207]. This could be a possible explanation for the correlation of an increased *miR-135b* level in CRC metastasis as we detect.

Not only *miR-135b* showed a considerable increased expression in malignant tissues, we also determined and validated microRNA candidates, such as *miR-31*, *miR-183* and *miR-96*. For those candidates literature entries have demonstrated an association between their up-regulation and CRC [162].

An altered expression of microRNAs can be caused by various mechanisms, including epigenetic modifications. Therefore we analyzed microRNA genes that showed a significantly differential expression with a corresponding epigenetic modification of the genomic location. To estimate the methylation status of colon cancer patients and cell lines we used methylated DNA immunoprecipitation sequencing (MeDIP-Seq) [175] data of 14 different colon cancer patients and colon cancer cell lines SW480 and SW620 generated by Grimm et al. (Grimm et al., in preparation 2012) and analyzed the changes in microRNA expression in parallel. Two microRNA candidates exposed an aberrant methylation in tumor, *miR-129* which showed a hyper-methylation in tumor tissue and *miR-493* which represented a hypo-methylation in colon cancer. The first connection between aberrant DNA methylation and the regulation of microRNA expression was described in bladder cancer cells by Saito et al. [208] in 2006. In 2009 Bandres et al [209] depicted the epigenetic regulation of microRNAs in colorectal cancer using methylation specific PCR (MSP) and bisulfide sequencing for the promoter region of *miR-129* after they had observed a down-regulation for this specific microRNA in CRC patients. We analyzed the epigenetic repression of *miR-129* and could identify a constant hyper-methylation for one of the two genomic variations of *miR-129* in the tumor tissue. *MiR-129-2* is located at chromosome 11p11.2 and it is embedded in a CpG island. Our next generation sequencing analyses of *miR-129* elucidated a strong down-regulation in tumor and a much lower expression level in metastases of colon cancer patients. By screening different cancer tissues we could additionally observe a significant decreased expression in another sample set of colon cancer patients matched to normal tissues. *MiR-129* is also described as down-regulated in other tumor tissues, such as breast cancer [210], gastric cancer [211], endometrial cancer [212] or bladder cancer [213]. In terms of the constant down-regulation of *miR-129* across all tumor tissues which we could also observe in the screening experiment we could provide *miR-129* as tumor marker candidate not only for colon cancer. These findings provide another evidence for the validity of our microRNA sequencing data and pave the way for the reliability of other significant microRNA candidates that we have identified. For example, *miR-493* showed a significant hypo-methylation in the tumor tissue of our colon cancer patient cohort. This matched to the up-regulation we detected in the tumor and metastases tissue. In metastases the expression was even higher than in the tumor tissue, tempting to speculate, that the altered DNA methylation of the *miR-493* could be involved in the development of colon cancer into distant sites. Nevertheless, to analyze which microRNAs might be involved in the metastatic process, further analyses are required. Less is known about the possible function of *miR-493*. Stilling et al. [214] determined an up-regulation of mature *miR-493* in pituitary carcinomas based on microarray data, real-time PCR and in-situ hybridization. Altogether the specific microRNA methylation in CRC could be beneficial for diagnosis, prognosis and it could function as a potential drug target for chromatin modifying therapeutics leading to the activation of tumor suppressor microRNAs such as *miR-129*.

For two additional interesting microRNAs we could observe a constant down-regulation over 16 additional tumor samples based on the results of the TaqMan microRNA screening experiment. Both *miR-215* and *miR-497* showed a down-regulation in our next generation sequencing experiments. *MiR-215* is already described in the context of colon cancer by Georges et al. [215] and Braun et al. [216] in 2008. They described the correlation between *miR-215* (which is arranged in a cluster with *miR-192* and *miR-194*) and a cell cycle arrest at G2-phase due to a TP53-dependent up-regulation of *TP21* in colon cancer cells. They showed that *TP53* activation induces an up-regulation of the homologous microRNAs *miR-192* and *miR-215*. Using microarray profiling and reporter assays they determined that *miR-192/215* induces cell cycle arrest via downstream target inhibition, including a number of G1 and G2 checkpoints. The results presented a role for *miR-192/215* in cell proliferation and support the assumption that *miR-192/215* function as tumor suppressors. Beside those findings Song et al. in 2010 [163] found that protein levels of *DHFR* (Dihydrofolate reductase) and *TS* (thymidylate synthase), both important targets for antifolate- and fluoropyrimidine-based chemotherapies in cancer treatment, were suppressed by *miR-215* in colon cancer cell lines. They investigated the role of *miR-215* in the chemoresistance to *DHFR* inhibitor methotrexate (MTX) and *TS* inhibitor Tomudex (TDX). Interestingly, they determined that an ectopic expression of *miR-215* and a down-regulation of DHFR and TS proteins resulted in a decreased sensitivity towards MTX and TDX, whereas small-interfering RNAs (siRNAs) against *DHFR* or *TS* had the opposite effect and showed an increasing sensitivity to MTX and TDX. Taken together, the findings for *miR-215* suggest a significant role of *miR-215* in the mechanism of tumor chemoresistance. For *miR-497* Guo et al. [217] was the first group who showed a significant down-regulation in gastric cancer cells. Interestingly, the group of Zhu et al. [218] investigated the role of *miR-497*, which is located in a cluster with *miR-195* on chromosome 17, in multidrug-resistant human gastric cancer (SGC7901/vincristine) and lung cancer (A549/cisplatin) cell lines, where they showed that *miR-497* is down-regulated. Applying *in vitro* drug sensitivity assays and luciferase reporter assays they proved that *miR-497* modulates multi-drug resistance in both cell lines by repressing BCL2 (B-cell CLL/lymphoma 2) protein expression. The negative regulation of apoptosis pathways is one of the major mechanism in drug resistance of cancer cells [219]. Based on this clinical relevant function of *miR-215* and *miR-497* we decided to include them in our screening experiment. According to literature entries we could determine a constant down-regulation for both microRNAs in all cancer entities investigated. In general we observed a strong concordance of several published microRNA data to the microRNA data we obtained by different technologies such as next generation sequencing or quantitative real-time PCR. This indicates for the validity of our data.

The identification of metastatic-specific microRNAs was difficult, mainly because the pattern between tumor and metastasis tissues were quite similar. In addition to the tissue samples we also investigated SW480 and SW620 cell lines. These cell lines were derived from the tumor and metastasis of one patient, respectively. The consensus between the tissue data and the cell line data was very poor. Regarding the fact that the patients material is less homogeneously because it consists of different cell types, whereas the cell lines are clearly more uniform could deliver an explanation to this observation. The direct comparison of significant deregulated candidates delivered one microRNA candidate, *miR-106b*, which showed an up-regulation in the metastasis tissues compared to the primary tumor

tissues and an increased expression in SW620 compared to SW480. This microRNA is described in the context of prostate cancer, brain cancer, gastric cancer, breast cancer, hepatocellular carcinoma and colorectal cancer [220-227]. Here it is known to be frequently up-regulated in various types of human cancers. As a direct target of *miR-106b* the tumor suppressor gene *RB1* (retinoblastoma 1) was identified in human laryngeal carcinoma cells and it was shown that a reduction of *miR-106b* induced cell cycle arrest in those cells [228]. It was also shown that there is a significant correlation between *miR-106b*, the gene *SIX1* (SIX homeobox 1), and an activated TGF-β signaling in human breast cancers which could lead to a shift of TGF-β from tumor suppressive role to a tumor promoting function [220].

Comparing the expression of significant deregulated microRNAs for the three tissues (normal, tumor, metastases), we identified a total of 29 microRNAs whose expression was down-regulated and 17 microRNAs whose expression was up-regulated in tumor and metastases compared to the normal tissue. The overlap between normal, tumor and metastases tissue included our selected microRNA candidates *miR-1*, *miR-135b*, *miR-129*, *miR-215*, *miR-497* and *miR-493*. An interesting microRNA that becomes delineated during the overlap analysis was *miR-559*, which is 2.66-fold down-regulated in metastasis tissues versus normal tissues and 1.57-fold down-regulated in metastases versus tumor (Table **A.2**, Table **A.3**). This microRNA targets the proto-oncogene *ERBB2* (v-erb-b2 erythroblastic leukemia viral oncogene homolog 2) [174]. The *ERBB2* or *HER2* gene (human epidermal growth factor receptor 2) encodes a member of the epidermal growth factor (*EGF*) receptor family of receptor tyrosine kinases. The receptors of this family are involved in signal transduction pathways leading to cell growth and differentiation. The overexpression of *ERBB2* is associated to the onset of certain types of cancer including 20-30% of breast and ovarian carcinomas [229,230,231]. The human epidermal growth factor receptor 2 is an example for an approved candidate for targeted cancer therapies [177,178]. Therefore we checked the regulation of *ERBB2* in our colon cancer patients by NGS technology and quantitative real-time PCR. The data delivered quite diverse results. For both methods we could not detect a constant up-regulation in tumor tissue as expected. Nevertheless, due to the fact that we found *miR-559* constantly and significantly down-regulated in our colon cancer samples led us to the assumption that *miR-559* could be a promising candidate for further studies and possibly for the development of targeted therapies in colon cancer treatment.

To assess the possibility of selected microRNA candidates to function as biomarkers for clinical applications was one of the major tasks of this work. Due to the fact that the alterations we observed in the microRNA expression for the clinical samples were less concordantly with those of the cell lines we gave priority to the deregulated microRNAs of the cancer patients in further SOM (Self-Organizing Map) and PCA (Principal Component Analyses) analyses. To detect significant differentially expressed microRNAs as determined by next generation sequencing we performed PCA analyses and hierarchical clustering and achieved a distinct separation of malignant and benign tissue. Applying SOM analyses over all colon cancer patients showed a separation of malignant and benign tissue. This indicates for the possibility of the microRNAs to function as considerable tumor markers based on variations in expression levels which occur during tumor progression and metastasize.

Further PCA and hierarchical cluster analyses on a smaller set of microRNAs (*miR-1*, *miR-135b*, *miR-129*, *miR-493*, *miR-215* and *miR-497*) on different cancer entities and for brain, breast, colon, kidney, muscle, prostate, stomach, testis and thyroid gland were able to separate tumor from normal tissue. We could distinguish 9 different cancer entities, including brain, breast, colon, kidney, muscle, prostate, stomach, testis and thyroid gland. For colon, brain, prostate and stomach we achieved a completely separation of the two different tissues. Even two microRNAs alone, *miR-1* and *miR-135b*, achieved a nearly complete partitioning of tumor and normal tissue for colon cancer. Altogether, our data suggest and support the relevance of the selected microRNA candidates to function as biomarkers for cancer diagnosis and they provide the potential to deliver novel therapeutic targets.

Driven by the motivation to find de-regulated microRNAs that might have a central carcinogenic role in colon cancer and different tumor types we found a constant decrease in all tumor tissues for down-regulated microRNAs, such as *miR-1*, *miR-129*, *miR-215* and *miR-497*, whereas for *miR-135b* and *miR-493* the patterns were more heterogeneous between the tumor tissues. A general decrease of microRNAs expression was described as a common phenomenon in tumor and thus might play a causal role in tumor progression [232]. This reduced microRNA expression might be reasonably suspected as a consequence of an altered microRNA processing machinery which also promotes oncogenic transformation [233].

4.2 ROLE OF *MIR-1* IN COLON CANCER

During this work we investigated the complete set of microRNAs by high-throughput next generation sequencing to identify new diagnostic, prognostic therapeutic molecules. We selected *miR-1* as top drug candidate based on our findings that *miR-1* showed a significant down-regulation in colon cancer tumor and metastasis tissues. For *miR-1* we observed a distinctive down-regulation in both malignant tissues with a more than 3.79-fold down-regulation for *miR-1-1* (p-value = 0.0004) and a 3.33-fold down-regulation for *miR-1-2* (p-value = 0.001) in tumor. In metastasis tissues we determined a 4.19-fold down-regulation for *miR-1-1* (p-value = 0.004) and a 3.81-fold down-regulation for *miR-1-2* (p-value = 0.0009). In addition it showed a constant down-regulation in 16 additional tumor entities including adipose tissue, brain, breast, colon, endometrium, kidney, liver, lung, lymph node, muscle, ovary, pancreas, prostate, stomach, testis, thyroid gland. On average we observed a 2.3-fold-down-regualtion in the tumor tissues compared to the corresponding normal tissues. In functional assays we could also proof a tumor suppressive role of *miR-1* in colon cancer cell lines and we determined a synergistic effect on colon cancer cells with camptothecin cancer drug. *MiR-1* is arranged in a cluster with *miR-133* and both have been previously described in adult cardiac and skeletal muscle tissues [234]. *MiR-1* inhibits cell cycle progression of myoblasts and promotes their proliferation and differentiation by repressing the *HDAC4* (histone deacetylase 4) and *SRF* (serum response factor) protein [235]. In rhabdomyosarcoma *miR-1* suppresses tumor growth by targeting the oncogene *c-Met* (met proto-oncogene (hepatocyte growth factor receptor)) and it is also known as down-regulated in primary lung cancer tissues and cell lines [236, 94]. Here overexpressing of *miR-1* induced apoptosis in

the lung cancer cells A549 in response to the anticancer drug doxorubicin by an enhanced activation of caspase 3 and 7 and by a depletion of anti-apoptotic *MCL1* (myeloid cell leukemia sequence 1 (BCL2-related)). These findings and our results call attention to the potential therapeutic regulatory role of *miR-1* in colon cancer.

Assuming that *miR-1* could also have a triggering effect on colon cancer, we applied *in silico* analysis to address the question and used comprehensive systems biology gene expression model, the PyBioS software developed by Wierling et al., to simulate an interaction of *miR-1* with direct and indirect target genes. We simulated *in silico* a reduction or overexpression of *miR-1* in a model for primary tumors and metastases and compared the results with gene expression analyses from primary tumor (or metastasis) and normal tissues acquired from the same patients after a *miR-1* treatment. The *in silico* PyBioS model system is based on microRNA data and target information collected from various databases and studies.

To initialize the model we used data from an mRNA sequencing experiment on the same colon cancer patients we have had generated the microRNA expression pattern for. We then changed the *miR-1* levels and detected the relative change in gene expression before and after changing the quantitative levels of *miR-1*. As expected, a reduction of *miR-1* showed a considerable effect on the expression of direct and indirect downstream targets. The possible effects of altered *miR-1* levels on gene expression were grouped into five classes (ten subclasses), ‘plus’ and ‘minus’ of ‘desired’, ‘aggravating’, ‘side effect’, ‘weak effect’ and ‘flip to’. In addition we could observe ‘no effect’ on the gene regulation after *miR-1* treatment. We counted the frequency of genes that belong to each class. As expected, a simulated down-regulation of *miR-1* (1.01-, 1.1-, 1.2-, 1.4-, 1.6-, 1.8-, 2-, 4-, 10-, 20- and 100fold) led to a significant number of up-regulated genes. Interestingly, a high number of genes were grouped into an ‘aggravated plus’ state indicating a continuation of the carcinogenic process in tumor and metastasis patients. The simulation of *miR-1* as a therapeutic agent by computing an up-regulation of *miR-1* expression (1.01- to 100-fold) delivered more genes with reduced gene expression than increased expression calculated for the groups ‘aggravated minus’, ‘weak plus’ and ‘desired plus’. The direction of change for gene expression in those groups showed a similar pattern, from the tumor expression level to a more reduced ‘normal’ expression. In general we could observe a beneficial effect for all patients after an 1.4 to 1.6-fold increase of *miR-1*. The beneficial effects stayed constant from a 1.6-fold increase of *miR-1* upwards. Since we used individual RNA expression data of colon cancer patients to initialize the modeling processes for each tissue within the model we could determine and investigate effects of *miR-1* treatment individually. In theory, this approach could be scaled up for a large number of patients. In addition, the combination of several microRNAs which might have a synergistic effect can potentiate the therapeutic outcome. Investigating the anti-cancer effects of *miR-1* in functional experiments for colon cancer patients would not be possible for routine clinical use. Therefore we assume that an *in silico* modeling approach in combination with mRNA and microRNA sequencing data might be an effective approach for an individualized cancer therapy and for the selection of a suitable treatment.

A major aspect of investigating *miR-1* in colon cancer is its regulatory role regarding oncogenic target genes and their tumorigenic potential. Not only based on *miR-1* expression data, but also based on functional overexpression and knock down experiments we were able to show a tumor suppressive function of *miR-1*, for example, growth and migration inhibition by alamarBlue cell viability assay and wound healing assay. Looking for interesting target genes that are associated to this observation we found that *miR-1* targets important growth factors, such as *VEGFA* (vascular endothelial growth factor A) and *PDGFA* (platelet-derived growth factor alpha polypeptide), which are capable to stimulate cellular growth. In addition, both genes are targets for effective anti-cancer treatment using approved target agents, such as bevacizumab [177,178]. For *VEGFA* we were able to observe a general up-regulation in colon cancer patients by NGS technology and real-time PCR. It is well established that *VEGFA* is highly expressed in many types of cancer and is fundamentally involved in tumor angiogenesis [178]. To investigate the regulatory effect of *miR-1* on the expression of *VEGFA*, we overexpressed *miR-1* in colon cancer cells and determined a significant down-regulation of *VEGFA* on the mRNA level. Beside the angiogenic factors *VEGFA* and *PDGFA*, *PTPLAD1* (protein tyrosine phosphatase-like A domain containing 1), alias B-ind1, is an additional target gene of *miR-1* and controls the transcriptional activity of *NF-KB* (tumor necrosis factor receptor superfamily, member 11a, NFKB activator) [237]. It is also predicted to interact with *HIF1A* (hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)) [238]. Additionally, the 3'UTR of *HIF1A* revealed to be a target element of *miR-1*, based on bioinformatical analysis (microrna.org, 2011). These analyses propose that *miR-1* could have an effect on tumor angiogenesis. To test this hypothesis, we up-regulated *miR-1* in colon cancer cells and looked for effects on the migration rate of the cells. Indeed, we could observe an increased migration of the treated cells compared to those cells without an elevated level of *miR-1*. In *miR-1* overexpressing cells we also checked the expression of the mentioned *miR-1* targets and identified a reduction in the mRNA expression level of *PTPLAD1* and *VEGFA*. Using GFP-reporter assay we proved an interaction for the 3'UTR microRNA-binding sites of *PTPLAD1* with *miR-1* and observed an inhibition of gene expression. Interestingly, Taguwa et al. [239,240] showed in 2008 and 2009 that there is a possible binding of *PTPLAD1* to the ubiquitous molecular chaperone *HSP90* (heat shock protein 90kDa alpha (cytosolic), class A member 1) and *HSP90* is a target for cancer treatment. A frequently used *HSP90* inhibitor for cancer therapeutics is geldanamycin. A drug combination of *HSP90* inhibitor and topoisomerase II inhibitor, such as doxorubicin, was proven to show a synergistic effect on cell death in cancer cells [241,242]. In view of the fact that the topoisomerase II inhibitor doxorubicin was already described in the context of *miR-1* in lung cancer cells, where an overexpression of *miR-1* caused an increased sensitization of those cancer cells towards the chemotherapeutic treatment [94], we investigated *miR-1* overexpression in colon cancer cells in combination with an topoisomerase inhibitor. Indeed, we recognized a remarkable synergistic effect in the reduction of cell viability for SW480 and SW620 colon cancer cell lines after 48h. Our data suggest a model to explain this synergistic effect, which could be caused by a binding of *miR-1* to the 3'UTR of *PTPLAD1*.This causes an inhibition of *PTPLAD1* expression. Taguwa et al.[240] proposed a potential co-chaperone-like activity of *PTPLAD1* while binding to *HSP90* and therefore regulating its chaperone activity. An observed inhibition of *PTPLAD1*, like in our data set, could lead to an decreased activity of *HSP90*, which is known to stabilize a range of growth factor

genes, such as *EGFR* [243]. In addition, camptothecin treatment stabilizes the topoisomerase I binding to the DNA and leads to an open unknotted conformation which causes DNA damage and apoptosis (Figure **4.1**).

Summarizing our results we propose *miR-1* as a novel tumor suppressor which could play an important role in cancer treatment, where a reduced amount of toxic chemotherapeutic can be administer to cancer patients and therefore might minimize the chemotherapeutic side effects. Nevertheless, further investigations are necessary to proof this assumption.

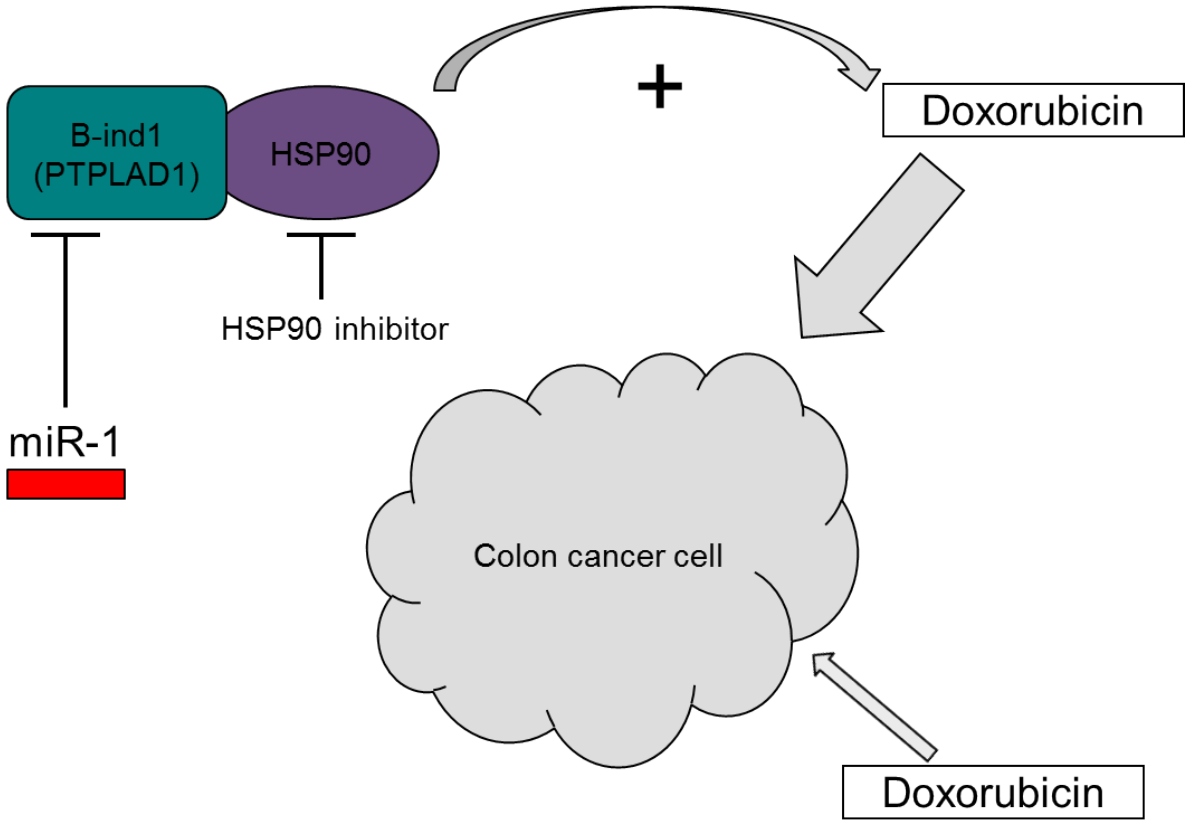

Figure 4.1: Model of synergy effect for *PTPLAD1* inhibition through *miR-1* translational repression or *HSP90* inhibitor in combination with a doxorubicin treatment in colon cancer cells. A *miR-1* overexpression in colon cancer cells reduces *PTPLAD1* expression level, which binds to *HSP90* and regulating its chaperone activity. This lead to a higher sensitization of a colon cancer cells towards a doxorubicin therapy as a single treatment of doxorubicin.

Taken together, the central carcinogenic role of microRNAs, such as *miR-1*, in the pathomechanisms of cancer can be regarded as something indisputable. In addition, microRNAs provide the possibility to act as new therapeutic molecules in cancer treatment. To develop new strategies for individualized cancer treatment the 'third-generation-sequencing' technologies could be an encouraging tool. Further improvements of those technologies provide a less expensive, time saving and more accurate possibility for the assessment of new biomarkers.

4.3 MUTATION ANALYSES IN COLON CANCER PATIENTS

Mutational events characterize the transition from normal colon epithelium to an adenoma and carcinoma. Analyzing mutations in colon cancer could reveal the possibility to discover functional and perhaps clinically relevant mutations either genes, microRNAs, microRNA-binding sites or in coding regions. Therefore we investigated polymorphisms in non-coding and coding regions in colon cancer patients using next generation sequencing technologies and enrichment applications. For mutation analyses we applied two different enrichment approaches, a microarray-based-genomic selection technology and a multiplex exon capturing approach, on two different CRC sample sets, whereas both of them included MSI (microsatellite instable) and MSS (microsatellite stable) patients. Considering the mutation rate of MSI and MSS patients we could predict a three to eight times higher mutation degree for the MSI patients. MSI tumors are characterized by a deficiency in the mismatch repair (MMR) system. These findings demonstrate the possibility of resequencing technologies to classify microsatellite associated colon cancers and in parallel identify the underlying mutation [40] and applying functional prediction algorithms, such as Polyphen and MutationTaster we predicted the damaging effect of an amino acid changes.

To estimate the biological relevance and the association of those detected mutations to a certain subtype of colorectal cancer and treatment outcome further analyses are essential. Nevertheless, targeted re-sequencing approaches represent an encouraging tool for individualized cancer therapies. Recent studies also addressed the association of microRNA polymorphisms to human cancers [198,199,244]. Functional microRNA-polymorpishms were described to inhibit microRNA-mediated regulation of drug target genes which affect drug resistance [245]. Mishra et al. determined a loss of *mir-24* microRNA function caused by a SNP in the microRNA binding site of *DHFR* (dihydrofolate reductase), which led to an over-expression of the target gene and an increased drug resistance towards methotrexate. During our investigations on 6 colon cancer patients and colonic cell lines we could identify at least 65 different microRNA-polymorphisms in the sequence of the microRNAs, while we investigated matched tumor and normal tissues samples. None of the predicted SNPs were somatic, which makes it impossible to find targets for cancer therapy on the basis of this sample set. Nevertheless, the investigation of microRNA-SNPs provides the opportunity to find cancer associated target genes that affect the drug behavior of patients belonging to a specific colorectal cancer subtype.

5 SUMMARY

Colorectal cancer (CRC) is the third most common malignant neoplasm and a major cause of cancer mortality worldwide. It is known that microRNAs play a critical role in oncogenic signaling pathways, including oncogenesis, progression, invasion, metastasis and angiogenesis. Previous studies of microRNA expression patterns in CRC elucidated a strong association between expression levels of microRNAs and the tumor stage as well as the survival prognosis for cancer patients. To investigate the microRNA expression patterns in a global scale we performed microRNA expression analyses with Illumina next generation sequencing technologies. We sequenced smallRNAs of eight colon cancer patients, each with matching normal, tumor and metastasis tissues from the same patient and analyzed in addition the microRNA expression in colon cancer cell lines SW480 and their corresponding metastasis cell line SW620. We found new and already known microRNAs, and validated different expression patterns using TaqMan assay technologies. We found that an identified signature of two microRNAs, *miR-1* and *miR-135b*, suffice in combination to completely discriminate tumor from normal colorectal mucosa. We extended our screen to 330 additional tumor and normal tissue samples and were able to confirm these microRNAs as potential biomarkers in tumors. The most promising candidate we identified, *miR-1*, showed a significantly reduced expression in primary colon tumor and metastasis and was constantly down-regulated over 16 additional tumor entities. This microRNA is reported to target and inhibit *c-Met* (met proto-oncogene (hepatocyte growth factor receptor)) in rhabdomyosarcoma and it is also known as down-regulated in primary lung cancer tissues and cell lines. Based on these findings and on the results of our expression analyses we could demonstrate an effect of *miR-1* overexpression on cell viability and on wound healing behavior in colon cancer cells. The functional analyses showed a tumor suppressive activity of *miR-1* in the advancement of CRC. Using target prediction algorithms, such as TargetScan and PicTar, we identified *PTPLAD1* (protein tyrosine phosphatase-like A domain containing 1) as target and showed a direct regulation through *miR-1* through GFP reporter assay. Additionally, the influence on the chemosensitivity of CRC cells towards camptothecin in response to a *miR-1* overexpression was investigated, and a synergistic effect was shown in colon cancer cell lines. Within further studies we used PyBioS, a systems-biology simulation of cellular cancer models, to examine *miR-1* function and the effects of a *miR-1* depletion as well as overexpression as a potential therapeutic option. For our colon cancer patients we showed that a combination of deep sequencing data and a systems biological modeling of microRNA drug treatment could be a promising tool to realize new strategies for personalized cancer therapy, where a patient-specific prognosis could be possible.

In this study not only known microRNAs were analyzed using next generation sequencing. In addition, we were able to identify and validate not yet annotated microRNAs. Using Northern Blot analyses we proofed the existence of a novel microRNA localized on chromosome 17 (chr17:70256349-70256434) with a predicted secondary structure that underlines the integrity of this novel microRNA.

An overview of the analyses performed in colorectal cancer for this study to identify new biomarkers for cancer therapy is schematically summarized in Figure **5.1**.

As an additional characteristic for colon cancer progression we investigated mutations in microRNA regions, whereas the identified unknown variations localized in *miR-663b* and *miR-1324* were determined as somatic mutations. Based on the whole exome re-sequencing data we analyzed mutation rates in microsatellite instable (MSI) and stable (MSS) colon cancer patients. Applying two different re-sequencing technologies, microarray-based-genomic selection technology and a multiplex exon capturing approach, we observed a higher mutation rate in MSI colon cancer patients.

In conclusion, this work emphasized the critical role of microRNAs in the pathogenesis of colorectal cancer and underlines the significance of microRNAs in the development of new strategies for individualized cancer treatment. This approach could facilitate the identification of novel diagnostic and prognostic biomarkers, but also for the design of new therapeutic molecules.

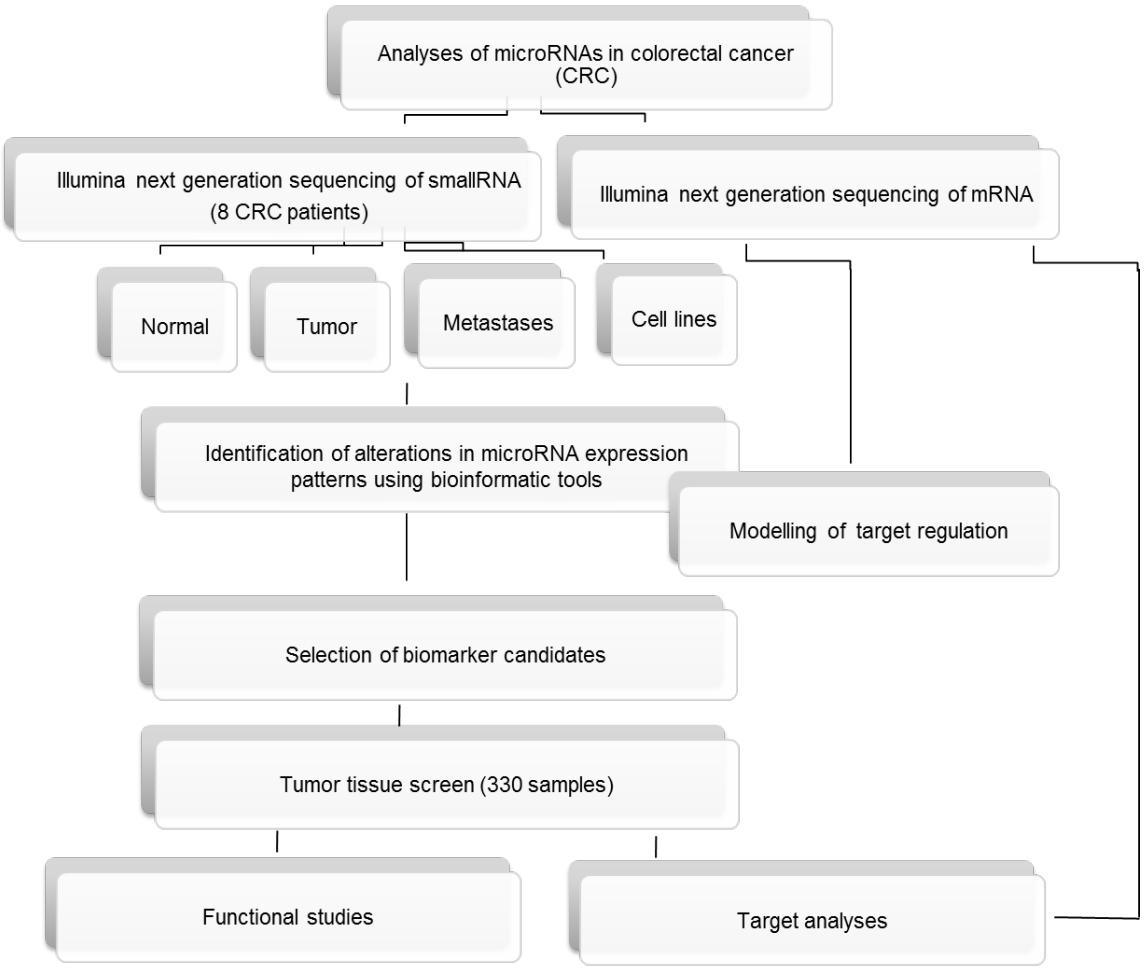

Figure 5.1: Schematic overview of the design of experiments of microRNAs analyses in colon cancer for an identification of potential biomarker and targets for cancer therapy.

6 ZUSAMMENFASSUNG (GERMAN SUMMARY)

Kolorektale Karzinome gehören zu den dritthäufigsten malignen Neoplasmen und zählen weltweit zu der häufigsten Todesursache.

Es ist bekannt, dass microRNAs eine entscheidende Rolle in Signalwegen, wie der Onkogenese, Tumorprogression und -invasion, sowie der Metastasierung und Angiogenese spielen. Frühere Studien von microRNA-Expressionsmustern im kolorektalen Karzinom zeigten einen engen Zusammenhang zwischen den Expressionsintesitäten von microRNAs und dem Tumorstadium sowie der Überlebensprognose von Krebspatienten. Um microRNA-Expressionsmuster im globalen Maßstab zu untersuchen, führten wir microRNA-Expressionsanalysen unter Anwendung der Illumina „Next-Generation“ Sequenzierungstechnologie (NGS) durch (Abbildung **5.1**). Wir sequenzierten die „smallRNA“ Fraktion von acht Darmkrebspatienten, dabei von jedem Patienten jeweils Normal-, Tumor- und Metastasengewebe. Zusätzlich wurde die microRNA-Expression in der Darmkrebszelllinie SW480 und ihrer entsprechenden Metastasenzelllinie SW620 von uns analysiert. Hierbei fanden wir neue als auch bereits bekannte microRNAs und validierten verschiedene Expressionsmuster unter Verwendung der „TaqMan Assay“ Technologie. Eine von uns identifizierte Expressionssignatur von zwei microRNAs, *miR-1* und *miR-135b*, genügte, um eine vollständige Auftrennung des Tumorgewebes von der normalen kolorektalen Schleimhaut zu erreichen. Wir erweiterten unsere Untersuchungen auf 330 zusätzliche Tumor- und Normalgewebeproben und waren in der Lage, die erwähnten microRNAs als potentielle Biomarker für Tumore zu bestätigen. Der vielversprechendste von uns identifizierte Kandidat, *miR-1*, zeigte eine signifikant reduzierte Expression im primären Darmkarzinom und in Metastasen, ebenso war *miR-1* in 16 zusätzlichen Tumorentitäten konstant herunterreguliert. Diese microRNA interagiert und inhibiert das Onkogen *c-Met* (met proto-oncogene (hepatocyte growth factor receptor)) im Rhabdomyosarkom, und ist ebenfalls im primären Lungenkarzinomgewebe als auch in Lungenkarzinomzelllinien herunterreguliert. Basierend auf diesen Forschungsergebnissen und anhand der von uns aus Expressionsanalysen erhaltenen Resultate, konnten wir einen Effekt auf die Zellviabilität und das Wundheilungsverhalten von *miR-1*-überexpremierenden Darmkarzinomzellen demonstrieren. Funktionelle Analysen zeigten eine tumorsuppressive Aktivität von *miR-1* im Hinblick auf die fortschreitende Entwicklung eines kolorektalen Karzinoms.

Unter Anwendung von Vorhersagealgorithmen, wie TargetScan und PicTar, identifizierten wir verschiedene *miR-1* Targetgene und selektionierten *PTPLAD1* (protein tyrosine phosphatase-like A domain containing 1), *VEGFA* (vascular endothelial growth factor A), *PDGFA* (platelet-derived growth factor alpha polypeptide) und *HIF1A* (hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)) für weiterführende GFP-Reporterassays, in deren Verlauf wir eine direkte Regulation von *PTPLAD1* durch *miR-1* beweisen konnten. Darüber hinaus untersuchten wir den Einfluss auf die Chemosensitivität in kolorektalen Karzinomzellen gegenüber Champtothecin als Antwort auf eine *miR-1* Überexpression und konnten einen synergistischen Effekt in Darmkrebszelllinien nachweisen. In weiterführenden Analysen verwendeten wir PyBioS, ein

systembiologisches Simulationsprogramm, das auf zellulären Krebsmodellen basiert, um *miR-1* als mögliche Therapieoption zu untersuchen. Für die von uns untersuchten Darmkrebspatienten konnten wir zeigen, dass eine Kombination von Sequenzierungsdaten und einer systembiologischen Modellierung der microRNA-bezogenen Arzneimitteltherapie ein vielversprechendes Hilfsmittel darstellt, um neue Strategien in der personalisierten Krebstherapie zu realisieren. Die Erstellung patientenspezifischer Prognosen könnte dadurch erreicht werden.

In dieser Studie konnten nicht nur bereits bekannte microRNAs mit Hilfe der Next-Generation-Sequenzierungstechnologie analysiert, sondern auch unbekannte microRNAs identifiziert und validiert werden. Unter Anwendung von Northern Blot Analysen war es möglich, die Existenz einer neuen microRNA zu beweisen, die auf Chromosom 17 (chr17:70256349-70256434) lokalisiert ist, und deren prognostizierte Sekundär-Struktur ebenfalls die Integrität der neu identifizierten microRNA unterstreicht.

Als eine weitere mögliche Ursache der Darmkrebsentwicklung untersuchten wir Mutationen in microRNA-Regionen, wobei die identifizierten und noch nicht annotierten Variationen, lokalisiert in *miR-663b* und *miR-1324*, als somatische Mutationen bestimmt wurden. Im Rahmen der Mutationsanalysen konnten wir auch zeigen, dass die Mutationsrate in mikrosatelliten-instabilen (MSI) Darmtumoren im Gegensatz zu mikrosatelliten-stabilen (MSS) signifikant erhöht sind.

Zusammenfassend lässt sich sagen, dass die vorliegende Arbeit die entscheidende Rolle von microRNAs in der Pathogenese des kolorektalen Karzinoms hervorhebt, und die Signifikanz von microRNAs in der Entwicklung neuer Strategien für die individualisierte Krebstherapie unterstreicht, in der microRNAs sowohl als neue diagnostische und prognostische, aber auch als potentielle neue therapeutische Wirkstoffe fungieren können.

BIBLIOGRAPHY

1. Bretthauer, M. Colorectal cancer screening. *J Intern Med* **270**, 87-98 (2011).
2. Figueredo, A., Coombes, M.E. & Mukherjee, S. Adjuvant therapy for completely resected stage II colon cancer. *Cochrane Database Syst Rev*, CD005390 (2008).
3. Markowitz, S.D., Dawson, D.M., Willis, J. & Willson, J.K. Focus on colon cancer. *Cancer Cell* **1**, 233-6 (2002).
4. Huxley, R.R. *et al.* The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. *Int J Cancer* **125**, 171-80 (2009).
5. Potter, J.D. Colorectal cancer: molecules and populations. *J Natl Cancer Inst* **91**, 916-32 (1999).
6. Hawk, E.T. & Levin, B. Colorectal cancer prevention. *J Clin Oncol* **23**, 378-91 (2005).
7. Poynter, J.N. *et al.* Statins and the risk of colorectal cancer. *N Engl J Med* **352**, 2184-92 (2005).
8. Worthley, D.L. & Leggett, B.A. Colorectal cancer: molecular features and clinical opportunities. *Clin Biochem Rev* **31**, 31-8.
9. Fearon, E.R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-67 (1990).
10. Lengauer, C., Kinzler, K.W. & Vogelstein, B. DNA methylation and genetic instability in colorectal cancer cells. *Proc Natl Acad Sci U S A* **94**, 2545-50 (1997).
11. Bellam, N. & Pasche, B. Tgf-beta signaling alterations and colon cancer. *Cancer Treat Res* **155**, 85-103.
12. Rowan, A. *et al.* Refining molecular analysis in the pathways of colorectal carcinogenesis. *Clin Gastroenterol Hepatol* **3**, 1115-23 (2005).
13. Lynch, H.T. & de la Chapelle, A. Hereditary colorectal cancer. *N Engl J Med* **348**, 919-32 (2003).
14. Rustgi, A.K. The genetics of hereditary colon cancer. *Genes Dev* **21**, 2525-38 (2007).
15. Galiatsatos, P. & Foulkes, W.D. Familial adenomatous polyposis. *Am J Gastroenterol* **101**, 385-98 (2006).
16. Sillars-Hardebol, A.H. *et al.* Identification of key genes for carcinogenic pathways associated with colorectal adenoma-to-carcinoma progression. *Tumour Biol* **31**, 89-96.
17. Kinzler, K.W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159-70 (1996).
18. Polakis, P. The many ways of Wnt in cancer. *Curr Opin Genet Dev* **17**, 45-51 (2007).
19. Segditsas, S. & Tomlinson, I. Colorectal cancer and genetic alterations in the Wnt pathway. *Oncogene* **25**, 7531-7 (2006).
20. Aoki, K. & Taketo, M.M. Adenomatous polyposis coli (APC): a multi-functional tumor suppressor gene. *J Cell Sci* **120**, 3327-35 (2007).
21. Brocardo, M. & Henderson, B.R. APC shuttling to the membrane, nucleus and beyond. *Trends Cell Biol* **18**, 587-96 (2008).

22. Kennell, J. & Cadigan, K.M. APC and beta-catenin degradation. *Adv Exp Med Biol* **656**, 1-12 (2009).
23. Goss, K.H. & Groden, J. Biology of the adenomatous polyposis coli tumor suppressor. *J Clin Oncol* **18**, 1967-79 (2000).
24. Baker, S.J. *et al.* p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res* **50**, 7717-22 (1990).
25. Vazquez, A., Bond, E.E., Levine, A.J. & Bond, G.L. The genetics of the p53 pathway, apoptosis and cancer therapy. *Nat Rev Drug Discov* **7**, 979-87 (2008).
26. Yuen, S.T. *et al.* Similarity of the phenotypic patterns associated with BRAF and KRAS mutations in colorectal neoplasia. *Cancer Res* **62**, 6451-5 (2002).
27. Saif, M.W. & Shah, M. K-ras mutations in colorectal cancer: a practice changing discovery. *Clin Adv Hematol Oncol* **7**, 45-53, 64 (2009).
28. Issa, J.P. CpG island methylator phenotype in cancer. *Nat Rev Cancer* **4**, 988-93 (2004).
29. Kondo, Y. & Issa, J.P. Epigenetic changes in colorectal cancer. *Cancer Metastasis Rev* **23**, 29-39 (2004).
30. Kim, M.S., Lee, J. & Sidransky, D. DNA methylation markers in colorectal cancer. *Cancer Metastasis Rev* **29**, 181-206.
31. Weisenberger, D.J. *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* **38**, 787-93 (2006).
32. Boland, C.R. & Goel, A. Microsatellite instability in colorectal cancer. *Gastroenterology* **138**, 2073-2087 e3.
33. Vilar, E. & Gruber, S.B. Microsatellite instability in colorectal cancer-the stable evidence. *Nat Rev Clin Oncol* **7**, 153-62.
34. Shah, S.N., Hile, S.E. & Eckert, K.A. Defective mismatch repair, microsatellite mutation bias, and variability in clinical cancer phenotypes. *Cancer Res* **70**, 431-5.
35. Poulogiannis, G., Frayling, I.M. & Arends, M.J. DNA mismatch repair deficiency in sporadic colorectal cancer and Lynch syndrome. *Histopathology* **56**, 167-79.
36. Auclair, J. *et al.* Intensity-dependent constitutional MLH1 promoter methylation leads to early onset of colorectal cancer by affecting both alleles. *Genes Chromosomes Cancer* **50**, 178-85.
37. Bellizzi, A.M. & Frankel, W.L. Colorectal cancer due to deficiency in DNA mismatch repair function: a review. *Adv Anat Pathol* **16**, 405-17 (2009).
38. Guastadisegni, C., Colafranceschi, M., Ottini, L. & Dogliotti, E. Microsatellite instability as a marker of prognosis and response to therapy: a meta-analysis of colorectal cancer survival data. *Eur J Cancer* **46**, 2788-98.
39. Sheffer, M. *et al.* Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. *Proc Natl Acad Sci U S A* **106**, 7131-6 (2009).
40. Timmermann, B. *et al.* Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis. *PLoS One* **5**, e15661 (2010).

41. Maxam, A.M. & Gilbert, W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A* **74**, 560-4 (1977).
42. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-7 (1977).
43. Smith, L.M. *et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674-9 (1986).
44. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-45 (2004).
45. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
46. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-80 (2005).
47. Venter, J.C., Levy, S., Stockwell, T., Remington, K. & Halpern, A. Massive parallelism, randomness and genomic advances. *Nat Genet* **33 Suppl**, 219-27 (2003).
48. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728-32 (2005).
49. Bentley, D.R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-9 (2008).
50. Branton, D. *et al.* The potential and challenges of nanopore sequencing. *Nat Biotechnol* **26**, 1146-53 (2008).
51. Gupta, P.K. Single-molecule DNA sequencing technologies for future genomics research. *Trends Biotechnol* **26**, 602-11 (2008).
52. Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* **4**, 265-70 (2009).
53. Metzker, M.L. Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46 (2010).
54. Tewhey, R. *et al.* Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat Biotechnol* **27**, 1025-31 (2009).
55. Albert, T.J. *et al.* Direct selection of human genomic loci by microarray hybridization. *Nat Methods* **4**, 903-5 (2007).
56. Hodges, E. *et al.* Genome-wide in situ exon capture for selective resequencing. *Nat Genet* **39**, 1522-7 (2007).
57. Okou, D.T. *et al.* Microarray-based genomic selection for high-throughput resequencing. *Nat Methods* **4**, 907-9 (2007).
58. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* **27**, 182-9 (2009).
59. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* **106**, 19096-101 (2009).
60. Porreca, G.J. *et al.* Multiplex amplification of large sets of human exons. *Nat Methods* **4**, 931-6 (2007).
61. Krishnakumar, S. *et al.* A comprehensive assay for targeted multiplex amplification of human DNA sequences. *Proc Natl Acad Sci U S A* **105**, 9296-301 (2008).

62. Sassen, S., Miska, E.A. & Caldas, C. MicroRNA: implications for cancer. *Virchows Arch* **452**, 1-10 (2008).
63. Lee, R.C., Feinbaum, R.L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-54 (1993).
64. Bentwich, I. A postulated role for microRNA in cellular differentiation. *FASEB J* **19**, 875-9 (2005).
65. Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20 (2005).
66. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-9 (2003).
67. Yi, R., Qin, Y., Macara, I.G. & Cullen, B.R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**, 3011-6 (2003).
68. Lee, Y., Jeon, K., Lee, J.T., Kim, S. & Kim, V.N. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* **21**, 4663-70 (2002).
69. Chendrimada, T.P. *et al.* TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740-4 (2005).
70. Jackson, R.J. & Standart, N. How do microRNAs regulate gene expression? *Sci STKE* **2007**, re1 (2007).
71. Lytle, J.R., Yario, T.A. & Steitz, J.A. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc Natl Acad Sci U S A* **104**, 9667-72 (2007).
72. Calin, G.A. & Croce, C.M. MicroRNA-cancer connection: the beginning of a new tale. *Cancer Res* **66**, 7390-4 (2006).
73. Calin, G.A. *et al.* Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* **99**, 15524-9 (2002).
74. Calin, G.A. *et al.* Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A* **101**, 2999-3004 (2004).
75. Ota, A. *et al.* Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. *Cancer Res* **64**, 3087-95 (2004).
76. Cimmino, A. *et al.* miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A* **102**, 13944-9 (2005).
77. Esquela-Kerscher, A. & Slack, F.J. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* **6**, 259-69 (2006).
78. Iorio, M.V. *et al.* MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* **65**, 7065-70 (2005).
79. Michael, M.Z., SM, O.C., van Holst Pellekaan, N.G., Young, G.P. & James, R.J. Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol Cancer Res* **1**, 882-91 (2003).
80. Johnson, S.M. *et al.* RAS is regulated by the let-7 microRNA family. *Cell* **120**, 635-47 (2005).
81. O'Donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V. & Mendell, J.T. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* **435**, 839-43 (2005).

82. Bloomston, M. *et al.* MicroRNA expression patterns to differentiate pancreatic adenocarcinoma from normal pancreas and chronic pancreatitis. *JAMA* **297**, 1901-8 (2007).
83. Fulci, V. *et al.* Quantitative technologies establish a novel microRNA profile of chronic lymphocytic leukemia. *Blood* **109**, 4944-51 (2007).
84. Landgraf, P. *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401-14 (2007).
85. Meng, F. *et al.* MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. *Gastroenterology* **133**, 647-58 (2007).
86. Roldo, C. *et al.* MicroRNA expression abnormalities in pancreatic endocrine and acinar tumors are associated with distinctive pathologic features and clinical behavior. *J Clin Oncol* **24**, 4677-84 (2006).
87. Volinia, S. *et al.* A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A* **103**, 2257-61 (2006).
88. Wang, T. *et al.* A micro-RNA signature associated with race, tumor size, and target gene activity in human uterine leiomyomas. *Genes Chromosomes Cancer* **46**, 336-47 (2007).
89. Metzler, M., Wilda, M., Busch, K., Viehmann, S. & Borkhardt, A. High expression of precursor microRNA-155/BIC RNA in children with Burkitt lymphoma. *Genes Chromosomes Cancer* **39**, 167-9 (2004).
90. Kluiver, J. *et al.* BIC and miR-155 are highly expressed in Hodgkin, primary mediastinal and diffuse large B cell lymphomas. *J Pathol* **207**, 243-9 (2005).
91. Calin, G.A. *et al.* A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med* **353**, 1793-801 (2005).
92. Yanaihara, N. *et al.* Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9**, 189-98 (2006).
93. O'Connell, R.M., Taganov, K.D., Boldin, M.P., Cheng, G. & Baltimore, D. MicroRNA-155 is induced during the macrophage inflammatory response. *Proc Natl Acad Sci U S A* **104**, 1604-9 (2007).
94. Nasser, M.W. *et al.* Down-regulation of micro-RNA-1 (miR-1) in lung cancer. Suppression of tumorigenic property of lung cancer cells and their sensitization to doxorubicin-induced apoptosis by miR-1. *J Biol Chem* **283**, 33394-405 (2008).
95. Tang, F., Hajkova, P., Barton, S.C., Lao, K. & Surani, M.A. MicroRNA expression profiling of single whole embryonic stem cells. *Nucleic Acids Res* **34**, e9 (2006).
96. Lim, L.P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769-73 (2005).
97. Orom, U.A., Kauppinen, S. & Lund, A.H. LNA-modified oligonucleotides mediate specific inhibition of microRNA function. *Gene* **372**, 137-41 (2006).
98. Hutvagner, G. & Zamore, P.D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056-60 (2002).
99. Krutzfeldt, J. *et al.* Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438**, 685-9 (2005).
100. Blower, P.E. *et al.* MicroRNA expression profiles for the NCI-60 cancer cell panel. *Mol Cancer Ther* **6**, 1483-91 (2007).

101. Fidler, I.J. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nat Rev Cancer* **3**, 453-8 (2003).
102. Raver-Shapira, N. *et al.* Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. *Mol Cell* **26**, 731-43 (2007).
103. Asangani, I.A. *et al.* MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene* **27**, 2128-36 (2008).
104. Christoffersen, N.R., Silahtaroglu, A., Orom, U.A., Kauppinen, S. & Lund, A.H. miR-200b mediates post-transcriptional repression of ZFHX1B. *RNA* **13**, 1172-8 (2007).
105. Hurteau, G.J., Carlson, J.A., Spivack, S.D. & Brock, G.J. Overexpression of the microRNA hsa-miR-200c leads to reduced expression of transcription factor 8 and increased expression of E-cadherin. *Cancer Res* **67**, 7972-6 (2007).
106. Gregory, P.A. *et al.* The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* **10**, 593-601 (2008).
107. Korpal, M., Lee, E.S., Hu, G. & Kang, Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J Biol Chem* **283**, 14910-4 (2008).
108. Bracken, C.P. *et al.* A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. *Cancer Res* **68**, 7846-54 (2008).
109. Burk, U. *et al.* A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep* **9**, 582-9 (2008).
110. Nagel, R. *et al.* Regulation of the adenomatous polyposis coli gene by the miR-135 family in colorectal cancer. *Cancer Res* **68**, 5795-802 (2008).
111. Dews, M. *et al.* Augmentation of tumor angiogenesis by a Myc-activated microRNA cluster. *Nat Genet* **38**, 1060-5 (2006).
112. La Rocca, G. *et al.* Mechanism of growth inhibition by MicroRNA 145: the role of the IGF-I receptor signaling pathway. *J Cell Physiol* **220**, 485-91 (2009).
113. Shi, B. *et al.* Micro RNA 145 targets the insulin receptor substrate-1 and inhibits the growth of colon cancer cells. *J Biol Chem* **282**, 32582-90 (2007).
114. Pechlivanis, S. *et al.* Insulin pathway related genes and risk of colorectal cancer: INSR promoter polymorphism shows a protective effect. *Endocr Relat Cancer* **14**, 733-40 (2007).
115. Ciardiello, F. & Tortora, G. EGFR antagonists in cancer treatment. *N Engl J Med* **358**, 1160-74 (2008).
116. Tsang, W.P. & Kwok, T.T. The miR-18a* microRNA functions as a potential tumor suppressor by targeting on K-Ras. *Carcinogenesis* **30**, 953-9 (2009).
117. Akao, Y., Nakagawa, Y. & Naoe, T. let-7 microRNA functions as a potential growth suppressor in human colon cancer cells. *Biol Pharm Bull* **29**, 903-6 (2006).
118. Chen, X. *et al.* Role of miR-143 targeting KRAS in colorectal tumorigenesis. *Oncogene* **28**, 1385-92 (2009).

119. Guo, C. *et al.* The noncoding RNA, miR-126, suppresses the growth of neoplastic cells by targeting phosphatidylinositol 3-kinase signaling and is frequently lost in colon cancers. *Genes Chromosomes Cancer* **47**, 939-46 (2008).
120. Krichevsky, A.M. & Gabriely, G. miR-21: a small multi-faceted RNA. *J Cell Mol Med* **13**, 39-53 (2009).
121. Schetter, A.J. *et al.* MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA* **299**, 425-36 (2008).
122. Slaby, O. *et al.* Altered expression of miR-21, miR-31, miR-143 and miR-145 is related to clinicopathologic features of colorectal cancer. *Oncology* **72**, 397-402 (2007).
123. Hermeking, H. p53 enters the microRNA world. *Cancer Cell* **12**, 414-8 (2007).
124. Chang, T.C. *et al.* Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol Cell* **26**, 745-52 (2007).
125. Bommer, G.T. *et al.* p53-mediated activation of miRNA34 candidate tumor-suppressor genes. *Curr Biol* **17**, 1298-307 (2007).
126. Yamakuchi, M., Ferlito, M. & Lowenstein, C.J. miR-34a repression of SIRT1 regulates apoptosis. *Proc Natl Acad Sci U S A* **105**, 13421-6 (2008).
127. Tazawa, H., Tsuchiya, N., Izumiya, M. & Nakagama, H. Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. *Proc Natl Acad Sci U S A* **104**, 15472-7 (2007).
128. Lodygin, D. *et al.* Inactivation of miR-34a by aberrant CpG methylation in multiple types of cancer. *Cell Cycle* **7**, 2591-600 (2008).
129. Gabriely, G. *et al.* MicroRNA 21 promotes glioma invasion by targeting matrix metalloproteinase regulators. *Mol Cell Biol* **28**, 5369-80 (2008).
130. Shell, S. *et al.* Let-7 expression defines two differentiation stages of cancer. *Proc Natl Acad Sci U S A* **104**, 11400-5 (2007).
131. Garzon, R., Fabbri, M., Cimmino, A., Calin, G.A. & Croce, C.M. MicroRNA expression and function in cancer. *Trends Mol Med* **12**, 580-7 (2006).
132. Hardcastle, J.D. *et al.* Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* **348**, 1472-7 (1996).
133. Huang, C.S., Lal, S.K. & Farraye, F.A. Colorectal cancer screening in average risk individuals. *Cancer Causes Control* **16**, 171-88 (2005).
134. Mandel, J.S. *et al.* Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. *N Engl J Med* **328**, 1365-71 (1993).
135. Loktionov, A. *et al.* Quantitation of DNA from exfoliated colonocytes isolated from human stool surface as a novel noninvasive screening test for colorectal cancer. *Clin Cancer Res* **4**, 337-42 (1998).
136. Vogelstein, B. *et al.* Genetic alterations during colorectal-tumor development. *N Engl J Med* **319**, 525-32 (1988).
137. Green, D.R. & Kroemer, G. Cytoplasmic functions of the tumour suppressor p53. *Nature* **458**, 1127-30 (2009).
138. Iacopetta, B. TP53 mutation in colorectal cancer. *Hum Mutat* **21**, 271-6 (2003).

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

140. Dietmaier, W. *et al.* Diagnostic microsatellite instability: definition and correlation with mismatch repair protein expression. *Cancer Res* **57**, 4749-56 (1997).

141. Ribic, C.M. *et al.* Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. *N Engl J Med* **349**, 247-57 (2003).

142. Esteller, M., Levine, R., Baylin, S.B., Ellenson, L.H. & Herman, J.G. MLH1 promoter hypermethylation is associated with the microsatellite instability phenotype in sporadic endometrial carcinomas. *Oncogene* **17**, 2413-7 (1998).

143. Boynton, K.A., Summerhayes, I.C., Ahlquist, D.A. & Shuber, A.P. DNA integrity as a potential marker for stool-based detection of colorectal cancer. *Clin Chem* **49**, 1058-65 (2003).

144. Imperiale, T.F., Ransohoff, D.F., Itzkowitz, S.H., Turnbull, B.A. & Ross, M.E. Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *N Engl J Med* **351**, 2704-14 (2004).

145. Duffy, M.J. Carcinoembryonic antigen as a marker for colorectal cancer: is it clinically useful? *Clin Chem* **47**, 624-30 (2001).

146. Ward, D.G. *et al.* Identification of serum biomarkers for colon cancer by proteomic analysis. *Br J Cancer* **94**, 1898-905 (2006).

147. Markou, A. *et al.* Prognostic value of mature microRNA-21 and microRNA-205 overexpression in non-small cell lung cancer by quantitative real-time RT-PCR. *Clin Chem* **54**, 1696-704 (2008).

148. Akao, Y., Nakagawa, Y. & Naoe, T. MicroRNA-143 and -145 in colon cancer. *DNA Cell Biol* **26**, 311-20 (2007).

149. Faber, C., Kirchner, T. & Hlubek, F. The impact of microRNAs on colorectal cancer. *Virchows Arch* **454**, 359-67 (2009).

150. Yang, L., Belaguli, N. & Berger, D.H. MicroRNA and colorectal cancer. *World J Surg* **33**, 638-46 (2009).

151. Mitchell, P.S. *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* **105**, 10513-8 (2008).

152. Kosaka, N., Iguchi, H. & Ochiya, T. Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis. *Cancer Sci* **101**, 2087-92.

153. Ng, E.K. *et al.* Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. *Gut* **58**, 1375-81 (2009).

154. Huang, Z. *et al.* Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. *Int J Cancer* **127**, 118-26.

155. Ahmed, F.E. *et al.* Diagnostic microRNA markers for screening sporadic human colon cancer and active ulcerative colitis in stool and tissue. *Cancer Genomics Proteomics* **6**, 281-95 (2009).

156. Diaz, R. *et al.* Deregulated expression of miR-106a predicts survival in human colon cancer patients. *Genes Chromosomes Cancer* **47**, 794-802 (2008).

157. Koga, Y. *et al.* MicroRNA expression profiling of exfoliated colonocytes isolated from feces for colorectal cancer screening. *Cancer Prev Res (Phila)* **3**, 1435-42.
158. Kulda, V. *et al.* Relevance of miR-21 and miR-143 expression in tissue samples of colorectal carcinoma and its liver metastases. *Cancer Genet Cytogenet* **200**, 154-60 (2010).
159. Schepeler, T. *et al.* Diagnostic and prognostic microRNAs in stage II colon cancer. *Cancer Res* **68**, 6416-24 (2008).
160. Slaby, O., Svoboda, M., Michalek, J. & Vyzula, R. MicroRNAs in colorectal cancer: translation of molecular biology into clinical application. *Mol Cancer* **8**, 102 (2009).
161. Wang, C.J. *et al.* Clinicopathological significance of microRNA-31, -143 and -145 expression in colorectal cancer. *Dis Markers* **26**, 27-34 (2009).
162. Bandres, E. *et al.* Identification by Real-time PCR of 13 mature microRNAs differentially expressed in colorectal cancer and non-tumoral tissues. *Mol Cancer* **5**, 29 (2006).
163. Song, B. *et al.* Molecular mechanism of chemoresistance by miR-215 in osteosarcoma and colon cancer cells. *Mol Cancer* **9**, 96.
164. Karaayvaz, M. *et al.* Prognostic significance of miR-215 in colon cancer. *Clin Colorectal Cancer* **10**, 340-7 (2011).
165. Nakajima, G. *et al.* Non-coding MicroRNAs hsa-let-7g and hsa-miR-181b are Associated with Chemoresponse to S-1 in Colon Cancer. *Cancer Genomics Proteomics* **3**, 317-324 (2006).
166. Motoyama, K. *et al.* Over- and under-expressed microRNAs in human colorectal cancer. *Int J Oncol* **34**, 1069-75 (2009).
167. Ahlquist, D.A. Molecular detection of colorectal neoplasia. *Gastroenterology* **138**, 2127-39.
168. Cummins, J.M. *et al.* The colorectal microRNAome. *Proc Natl Acad Sci U S A* **103**, 3687-92 (2006).
169. Xi, Y. *et al.* Prognostic Values of microRNAs in Colorectal Cancer. *Biomark Insights* **2**, 113-121 (2006).
170. Saiki, R.K. *et al.* Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350-4 (1985).
171. Blin, N. & Stafford, D.W. A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res* **3**, 2303-8 (1976).
172. Schmittgen, T.D. & Livak, K.J. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101-8 (2008).
173. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of Escherichia coli with plasmids. *Gene* **96**, 23-8 (1990).
174. Chen, H. *et al.* Preliminary validation of ERBB2 expression regulated by miR-548d-3p and miR-559. *Biochem Biophys Res Commun* **385**, 596-600 (2009).
175. Weber, M. *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* **37**, 853-62 (2005).
176. Wirth, H., Loffler, M., von Bergen, M. & Binder, H. Expression cartography of human tissues using self organizing maps. *BMC Bioinformatics* **12**, 306 (2011).
177. Di Cosimo, S. & Baselga, J. Targeted therapies in breast cancer: where are we now? *Eur J Cancer* **44**, 2781-90 (2008).

178. Ferrara, N., Hillan, K.J., Gerber, H.P. & Novotny, W. Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nat Rev Drug Discov* **3**, 391-400 (2004).
179. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11 (2009).
180. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**, 2325-9 (2011).
181. Frey, M.R., Hilliard, V.C., Mullane, M.T. & Polk, D.B. ErbB4 promotes cyclooxygenase-2 expression and cell survival in colon epithelial cells. *Lab Invest* **90**, 1415-24 (2010).
182. Paatero, I. & Elenius, K. ErbB4 and its isoforms: patentable drug targets? *Recent Pat DNA Gene Seq* **2**, 27-33 (2008).
183. Akao, Y. *et al.* Dysregulation of microRNA-34a expression causes drug-resistance to 5-FU in human colon cancer DLD-1 cells. *Cancer Lett* **300**, 197-204 (2011).
184. Borralho, P.M. *et al.* MicroRNA-143 reduces viability and increases sensitivity to 5-fluorouracil in HCT116 human colorectal cancer cells. *FEBS J* **276**, 6689-700 (2009).
185. Li, J., Zhang, Y., Zhao, J., Kong, F. & Chen, Y. Overexpression of miR-22 reverses paclitaxel-induced chemoresistance through activation of PTEN signaling in p53-mutated colon cancer cells. *Mol Cell Biochem* **357**, 31-8 (2011).
186. Song, B. *et al.* miR-192 Regulates dihydrofolate reductase and cellular proliferation through the p53-microRNA circuit. *Clin Cancer Res* **14**, 8080-6 (2008).
187. Song, B. *et al.* Mechanism of chemoresistance mediated by miR-140 in human osteosarcoma and colon cancer cells. *Oncogene* **28**, 4065-74 (2009).
188. Wang, C.J., Stratmann, J., Zhou, Z.G. & Sun, X.F. Suppression of microRNA-31 increases sensitivity to 5-FU at an early stage, and affects cell migration and invasion in HCT-116 colon cancer cells. *BMC Cancer* **10**, 616 (2010).
189. Schmidt, K.S. *et al.* Application of locked nucleic acids to improve aptamer in vivo stability and targeting function. *Nucleic Acids Res* **32**, 5757-65 (2004).
190. Valoczi, A. *et al.* Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes. *Nucleic Acids Res* **32**, e175 (2004).
191. Jorgensen, S., Baker, A., Moller, S. & Nielsen, B.S. Robust one-day in situ hybridization protocol for detection of microRNAs in paraffin samples using LNA probes. *Methods* **52**, 375-81 (2010).
192. Klipp E., L.W., Wierling C., Kowald A., Lehrach H., Herwig R. *Systems Biology - A Textbook*, (Wiley-VCH, Weinheim, 2009).
193. Wierling, C., Herwig, R. & Lehrach, H. Resources, standards and tools for systems biology. *Brief Funct Genomic Proteomic* **6**, 240-51 (2007).
194. Wierling, C. *et al.* Prediction in the face of uncertainty: A Monte Carlo-based approach for systems biology of cancer treatment. *Mutat Res* **746**, 163-70 (2012).
195. Li, J., Pandey, V., Kessler, T., Lehrach, H. & Wierling, C. Modeling of miRNA and drug action in the EGFR signaling pathway. *PLoS One* **7**, e30140 (2012).
196. Futreal, P.A. *et al.* A census of human cancer genes. *Nat Rev Cancer* **4**, 177-83 (2004).

197. Friedlander, M.R. *et al.* Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol* **26**, 407-15 (2008).
198. Landi, D. *et al.* Polymorphisms within micro-RNA-binding sites and risk of sporadic colorectal cancer. *Carcinogenesis* **29**, 579-84 (2008).
199. Yu, Z. *et al.* Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers. *Nucleic Acids Res* **35**, 4535-41 (2007).
200. Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L. & Bradley, A. Identification of mammalian microRNA host genes and transcription units. *Genome Res* **14**, 1902-10 (2004).
201. Orom, U.A., Nielsen, F.C. & Lund, A.H. MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell* **30**, 460-71 (2008).
202. Wang, Y.X. *et al.* Initial study of microRNA expression profiles of colonic cancer without lymph node metastasis. *J Dig Dis* **11**, 50-4.
203. Necela, B.M., Carr, J.M., Asmann, Y.W. & Thompson, E.A. Differential expression of microRNAs in tumors from chronically inflamed or genetic (APC(Min/+)) models of colon cancer. *PLoS One* **6**, e18501.
204. Li, Z. *et al.* A microRNA signature for a BMP2-induced osteoblast lineage commitment program. *Proc Natl Acad Sci U S A* **105**, 13906-11 (2008).
205. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
206. Herpin, A., Lelong, C. & Favrel, P. Transforming growth factor-beta-related proteins: an ancestral and widespread superfamily of cytokines in metazoans. *Dev Comp Immunol* **28**, 461-85 (2004).
207. Blobe, G.C., Schiemann, W.P. & Lodish, H.F. Role of transforming growth factor beta in human disease. *N Engl J Med* **342**, 1350-8 (2000).
208. Saito, Y. & Jones, P.A. Epigenetic activation of tumor suppressor microRNAs in human cancer cells. *Cell Cycle* **5**, 2220-2 (2006).
209. Bandres, E. *et al.* Epigenetic regulation of microRNA expression in colorectal cancer. *Int J Cancer* **125**, 2737-43 (2009).
210. Wang, Q.Y., Tang, J., Zhou, C.X. & Zhao, Q. [The down-regulation of miR-129 in breast cancer and its effect on breast cancer migration and motility]. *Sheng Li Xue Bao* **64**, 403-11 (2012).
211. Tsai, K.W. *et al.* Epigenetic regulation of miR-34b and miR-129 expression in gastric cancer. *Int J Cancer* **129**, 2600-10 (2011).
212. Huang, Y.W. *et al.* Epigenetic repression of microRNA-129-2 leads to overexpression of SOX4 oncogene in endometrial cancer. *Cancer Res* **69**, 9038-46 (2009).
213. Dyrskjot, L. *et al.* Genomic profiling of microRNAs in bladder cancer: miR-129 is associated with poor outcome and promotes cell death in vitro. *Cancer Res* **69**, 4851-60 (2009).
214. Stilling, G. *et al.* MicroRNA expression in ACTH-producing pituitary tumors: up-regulation of microRNA-122 and -493 in pituitary carcinomas. *Endocrine* **38**, 67-75.
215. Georges, S.A. *et al.* Coordinated regulation of cell cycle transcripts by p53-Inducible microRNAs, miR-192 and miR-215. *Cancer Res* **68**, 10105-12 (2008).

216. Braun, C.J. *et al.* p53-Responsive micrornas 192 and 215 are capable of inducing cell cycle arrest. *Cancer Res* **68**, 10094-104 (2008).
217. Guo, J. *et al.* Differential expression of microRNA species in human gastric cancer versus non-tumorous tissues. *J Gastroenterol Hepatol* **24**, 652-7 (2009).
218. Zhu, W. *et al.* miR-497 modulates multidrug resistance of human cancer cell lines by targeting BCL2. *Med Oncol*.
219. Rabik, C.A. & Dolan, M.E. Molecular mechanisms of resistance and toxicity associated with platinating agents. *Cancer Treat Rev* **33**, 9-23 (2007).
220. Smith, A.L. *et al.* The miR-106b-25 cluster targets Smad7, activates TGF-beta signaling, and induces EMT and tumor initiating cell characteristics downstream of Six1 in human breast cancer. *Oncogene* (2012).
221. Nishida, N. *et al.* Microarray analysis of colorectal cancer stromal tissue reveals upregulation of two oncogenic miRNA clusters. *Clin Cancer Res* **18**, 3054-70 (2012).
222. Kim, K. *et al.* Induction of the transcriptional repressor ZBTB4 in prostate cancer cells by drug-induced targeting of microRNA-17-92/106b-25 clusters. *Mol Cancer Ther* **11**, 1852-62 (2012).
223. Zhang, W., Edwards, A., Fan, W., Flemington, E.K. & Zhang, K. miRNA-mRNA correlation-network modules in human prostate cancer and the differences between primary and metastatic tumor subtypes. *PLoS One* **7**, e40130 (2012).
224. Wang, X. *et al.* Upregulation of miR-20a and miR-106b is involved in the acquisition of malignancy of pediatric brainstem gliomas. *Oncol Rep* **28**, 1293-300 (2012).
225. Yao, Y. *et al.* MicroRNA profiling of human gastric cancer. *Mol Med Report* **2**, 963-70 (2009).
226. Li, B. *et al.* Down-regulation of microRNA 106b is involved in p21-mediated cell cycle arrest in response to radiation in prostate cancer cells. *Prostate* **71**, 567-74 (2011).
227. Hudson, R.S. *et al.* MicroRNA-106b-25 cluster expression is associated with early disease recurrence and targets caspase-7 and focal adhesion in human prostate cancer. *Oncogene* (2012).
228. Cai, K., Wang, Y. & Bao, X. MiR-106b promotes cell proliferation via targeting RB in laryngeal carcinoma. *J Exp Clin Cancer Res* **30**, 73 (2011).
229. Sahin, A.A. Biologic and clinical significance of HER-2/neu (cerbB-2) in breast cancer. *Adv Anat Pathol* **7**, 158-66 (2000).
230. Pauletti, G., Godolphin, W., Press, M.F. & Slamon, D.J. Detection and quantitation of HER-2/neu gene amplification in human breast cancer archival material using fluorescence in situ hybridization. *Oncogene* **13**, 63-72 (1996).
231. Pegram, M.D. *et al.* The effect of HER-2/neu overexpression on chemotherapeutic drug sensitivity in human breast and ovarian cancer cells. *Oncogene* **15**, 537-47 (1997).
232. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-8 (2005).
233. Kumar, M.S., Lu, J., Mercer, K.L., Golub, T.R. & Jacks, T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat Genet* **39**, 673-7 (2007).
234. Chen, J.F. *et al.* The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet* **38**, 228-33 (2006).

235. van Rooij, E. & Olson, E.N. MicroRNAs: powerful new regulators of heart disease and provocative therapeutic targets. *J Clin Invest* **117**, 2369-76 (2007).
236. Yan, D. *et al.* MicroRNA-1/206 targets c-Met and inhibits rhabdomyosarcoma development. *J Biol Chem* **284**, 29596-604 (2009).
237. Courilleau, D. *et al.* B-ind1, a novel mediator of Rac1 signaling cloned from sodium butyrate-treated fibroblasts. *J Biol Chem* **275**, 17344-8 (2000).
238. Benita, Y. *et al.* An integrative genomics approach identifies Hypoxia Inducible Factor-1 (HIF-1)-target genes that form the core response to hypoxia. *Nucleic Acids Res* **37**, 4587-602 (2009).
239. Taguwa, S. *et al.* Human butyrate-induced transcript 1 interacts with hepatitis C virus NS5A and regulates viral replication. *J Virol* **82**, 2631-41 (2008).
240. Taguwa, S. *et al.* Cochaperone activity of human butyrate-induced transcript 1 facilitates hepatitis C virus replication through an Hsp90-dependent pathway. *J Virol* **83**, 10427-36 (2009).
241. Munster, P.N., Basso, A., Solit, D., Norton, L. & Rosen, N. Modulation of Hsp90 function by ansamycins sensitizes breast cancer cells to chemotherapy-induced apoptosis in an RB- and schedule-dependent manner. See: E. A. Sausville, Combining cytotoxics and 17-allylamino, 17-demethoxygeldanamycin: sequence and tumor biology matters, Clin. Cancer Res., 7: 2155-2158, 2001. *Clin Cancer Res* **7**, 2228-36 (2001).
242. Barker, C.R. *et al.* Inhibition of Hsp90 acts synergistically with topoisomerase II poisons to increase the apoptotic killing of cells due to an increase in topoisomerase II mediated DNA damage. *Nucleic Acids Res* **34**, 1148-57 (2006).
243. Sawai, A. *et al.* Inhibition of Hsp90 down-regulates mutant epidermal growth factor receptor (EGFR) expression and sensitizes EGFR mutant tumors to paclitaxel. *Cancer Res* **68**, 589-96 (2008).
244. Landi, D., Gemignani, F., Barale, R. & Landi, S. A catalog of polymorphisms falling in microRNA-binding regions of cancer genes. *DNA Cell Biol* **27**, 35-43 (2008).
245. Mishra, P.J., Humeniuk, R., Longo-Sorbello, G.S., Banerjee, D. & Bertino, J.R. A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance. *Proc Natl Acad Sci U S A* **104**, 13513-8 (2007).

American Cancer Society: www.cancer.org

CentroidHomfold: www.ncRNA.org

Illumina: www.illumina.com

Integrated DNA Technologies (IDT): eu.idtdna.com/site

Microrna.org: www.microrna.org/microrna/home.do

miRBase: www.mirbase.org

World Health Organization (WHO): www.who.int/en

LIST OF PUBLICATIONS

Röhr C, Kerick M, Fischer A, Kühn A, Kashofer K, Timmermann B, Meinel T, Drichel D, Boerno S, Nowka A, Krobitsch S., Becker T, Wunderlich A, Barmeyer C, Viertler C, Zatloukal K, Wierling C, Lehrach H, Schweiger MR, 2013. High-throughput miRNA and mRNA sequencing of paired colorectal normal, tumor and metastasis tissues and bioinformatic modeling of miRNA-1 therapeutic applications. PLoS ONE 8(7): e67461.

Schweiger MR, Hussong M, Röhr C, Lehrach H: Genomics and epigenomics of colorectal cancer. *WIREs Syst Biol Med* 2012. doi: 10.1002/wsbm.1206

Börno ST, Fischer A, Kerick M, Fälth M, Laible M, Brase JC, Kuner R, Dahl A, Grimm C, Sayanjali B, Isau M, Röhr C, Wunderlich A, Timmermann B, Claus R, Plass C, Graefen M, Simon R, Demichelis F, Rubin MA, Sauter G, Schlomm T, Sültmann H, Lehrach H, Schweiger MR: Genome-wide DNA methylation events in TMPRSS2:ERG fusion negative prostate cancers implicate an EZH2 dependent mechanism with microRNA-26a hypermethylation. *Cancer Discov*. 2012 Aug **28**.

Hallen L, Klein H, Stoschek C, Wehrmeyer S, Nonhoff U, Ralser M, Wilde J, Röhr C, Schweiger MR, Zatloukal K, Vingron M, Lehrach H, Konthur Z, Krobitsch S: The KRAB-containing zinc-finger transcriptional regulator ZBRK1 activates SCA2 gene transcription through direct interaction with its gene product, ataxin-2. *Hum Mol Genet.* 2011 Jan 1; **20 (1)**: 104-14.

Timmermann B, Kerick M, Roehr C, Fischer A, Isau M, Boerno ST, Wunderlich A, Barmeyer C, Seemann P, Koenig J, Lappe M, Kuss AW, Garshasbi M, Bertram L, Trappe K, Werber M, Herrmann BG, Zatloukal K, Lehrach H, Schweiger MR: Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis. *PLoS One.* 2010 Dec 22; **5 (12)**:e15661.

Ziegler G, Harhausen D, Schepers C, Hoffmann O, Röhr C, Prinz V, König J, Lehrach H, Nietfeld W, Trendelenburg G: TLR2 has a detrimental role in mouse transient focal cerebral ischemia. *Biochem Biophys Res Commun.* 2007 Aug 3; **359 (3)**: 574-9.

APPENDIX

Table A.1: Significant top 25 up- and down-regulated microRNAs comparing tumor versus normal colon samples, sorted by median log2ratio. Analysis of 8 colorectal cancer patients using NGS technologies. All listed microRNAs sufficed a p-value threshold ≤ 0.05.

ID	*P1*	*P2*	*P3*	*P4*	*P5*	*P6*	*P7*	*P8*	*median*	*p-values*	*fold change*
	log2ratio										
down-regulated in T/N											
miR-129-1	-5.30	NA	-4.09	-4.89	-5.31	-2.88	-6.21	0.29	-4.89	0.00270	29.65
miR-137	-1.74	NA	NA	-4.58	-4.39	-5.04	-1.89	-4.34	-4.37	0.00164	20.63
miR-1-1	-3.34	-3.37	-3.96	-3.75	-7.32	-1.01	-3.89	-3.84	-3.80	0.00042	13.90
miR-129-2	-4.32	-1.60	-5.07	-2.96	-4.98	-3.33	-3.58	-0.62	-3.45	0.00057	10.96
miR-1-2	-2.74	-4.03	-1.50	-3.92	-6.62	-0.53	-2.43	-4.34	-3.33	0.00179	10.06
miR-1224	-4.24	-2.94	-0.36	-3.64	-6.18	-1.96	-1.76	-5.25	-3.29	0.00189	9.78
miR-202	-0.91	NA	NA	-2.79	-3.20	-1.58	-3.37	-4.79	-2.99	0.00436	7.97
miR-147b	0.34	-3.81	-4.98	-1.53	-5.47	-4.73	-2.06	-0.22	-2.93	0.00950	7.64
miR-194-1	0.21	1.47	-4.20	-2.62	-6.79	-1.41	-3.37	-3.19	-2.90	0.02981	7.49
miR-215	-3.16	2.38	-5.36	-1.77	-5.93	-1.28	-3.80	-2.57	-2.87	0.02249	7.30
miR-551b	-0.14	-2.86	NA	-1.07	-1.72	-3.00	-3.50	-2.90	-2.86	0.00337	7.28
miR-3154	-1.91	-2.28	NA	-0.47	-0.39	-3.58	-2.57	-2.51	-2.28	0.00428	4.85
miR-139	-2.00	-2.05	-1.32	-2.49	-0.92	-0.36	-2.03	-3.21	-2.01	0.00079	4.04
miR-195	1.57	-2.80	-2.72	-1.82	-0.65	-2.09	-2.53	-1.93	-2.01	0.01643	4.02
miR-3065	-1.94	-2.21	-0.43	-1.59	-2.09	-1.07	-1.00	-2.99	-1.76	0.00067	3.40
miR-138-1	-1.64	-1.86	1.22	-2.58	-3.01	-1.54	-2.73	0.11	-1.75	0.02309	3.37
miR-422a	-1.68	NA	-1.17	-1.51	-4.07	-2.28	-2.39	0.24	-1.68	0.01020	3.21
miR-3150	-0.32	-1.76	-3.32	-1.53	-2.10	-1.85	-0.50	0.25	-1.65	0.01105	3.13
miR-3152	NA	-1.28	-2.00	-0.98	NA	NA	-2.37	NA	-1.64	0.01391	3.12
miR-30a	-1.43	-1.47	-2.60	-1.70	-0.30	0.18	-1.69	-1.85	-1.58	0.00346	2.99
miR-497	-2.13	-1.59	-4.35	-1.24	-0.88	-1.51	-2.46	-1.47	-1.55	0.00140	2.93
miR-585	NA	-1.28	NA	-2.47	-2.13	-1.58	-0.68	-1.49	-1.54	0.00158	2.90
miR-378	-2.11	-1.10	-2.66	-0.21	-1.01	-2.85	-1.92	-0.72	-1.51	0.00231	2.85
miR-378b	0.06	NA	-2.91	-1.10	-2.81	-2.79	-1.49	0.12	-1.49	0.02089	2.80
miR-145	-1.39	-2.00	-2.95	-1.27	-3.19	-0.12	-1.45	-1.09	-1.42	0.00213	2.68
up-regulated in T/N											
miR-675	NA	7.91	NA	5.98	5.59	NA	NA	NA	5.98	0.01198	62.92
miR-31	3.89	7.10	6.72	1.43	1.27	6.02	1.80	9.39	4.96	0.00332	31.02
miR-135b	7.08	4.78	4.01	3.78	1.78	3.01	5.21	5.97	4.40	0.00014	21.08
miR-183	3.81	3.07	1.88	3.38	-0.10	1.59	2.80	3.65	2.93	0.00104	7.63
miR-96	3.55	2.80	1.21	2.99	-0.10	0.42	2.29	3.45	2.54	0.00404	5.82
miR-182	5.34	2.93	1.32	2.64	-0.65	0.96	2.11	4.15	2.37	0.00948	5.18
miR-3175	NA	NA	NA	1.87	NA	NA	NA	2.10	1.98	0.03761	3.95
miR-584	0.26	3.80	3.25	1.56	-0.16	2.23	1.67	2.65	1.95	0.00568	3.87
miR-450a-2	NA	1.84	1.00	2.12	2.56	1.84	1.49	-0.36	1.84	0.00593	3.57

miR-493	1.58	0.19	1.97	1.91	0.03	3.32	2.15	-0.90	1.75	0.03431	3.35
miR-224	2.11	1.02	1.84	1.83	-0.71	1.48	1.34	3.44	1.66	0.00715	3.15
miR-449b	NA	NA	1.58	NA	0.20	NA	1.68	2.00	1.63	0.04206	3.10
miR-1286	NA	1.60	1.32	2.87	2.20	NA	1.60	0.10	1.60	0.00793	3.03
miR-503	NA	2.86	1.10	2.84	1.59	1.12	2.21	-0.18	1.59	0.00706	3.02
miR-3117	NA	2.98	0.92	2.41	1.31	1.45	1.41	1.04	1.41	0.00125	2.65
miR-217	-1.10	1.31	0.32	2.88	2.37	1.72	1.08	1.24	1.27	0.02545	2.41
miR-1323	NA	2.53	NA	1.20	0.46	NA	1.21	1.51	1.21	0.01457	2.32
miR-450a-1	1.26	NA	NA	3.45	3.00	-0.26	1.15	0.51	1.20	0.04886	2.30
miR-323b	NA	0.91	1.17	0.13	1.52	NA	2.34	NA	1.17	0.02840	2.25
miR-491	2.85	1.72	0.15	1.40	-0.30	0.58	0.91	2.13	1.16	0.01525	2.23
miR-450b	NA	0.96	0.74	2.47	1.25	1.66	1.09	0.15	1.09	0.00518	2.12
miR-455	-0.61	1.08	1.38	1.07	2.03	1.66	0.13	-0.15	1.07	0.04080	2.10
miR-663	NA	1.31	1.32	1.82	0.78	NA	0.11	0.78	1.04	0.00833	2.06
miR-21	1.96	0.24	-0.27	1.22	1.79	0.83	0.39	1.97	1.02	0.01206	2.03
miR-452	1.68	0.93	0.44	1.38	-0.61	1.06	0.48	1.50	1.00	0.01394	2.00

Table A.2: Significant top 25 up- and down-regulated microRNAs comparing metastases versus normal colon samples, sorted by median log2ratio. Analysis of 8 colorectal cancer patients using NGS technologies. All listed microRNAs sufficed a p-value threshold ≤ 0.05.

ID	*P1*	*P2*	*P3*	*P4*	*P5*	*P6*	*P7*	*P8*	*median*	*p-values*	*fold change*
	log2ratio										
down-regulated in Metastases/Normal											
miR-490	NA	-2.58	NA	-6.86	-2.66	NA	NA	-7.32	-4.76	0.03307	27.12
miR-1-1	-3.60	0.78	-3.12	-6.02	-1.38	-6.80	-4.79	-6.43	-4.20	0.00409	18.34
miR-215	-1.58	-0.26	-5.82	-5.53	-0.94	-3.95	-6.85	-3.95	-3.95	0.00413	15.43
miR-1-2	-3.84	-0.76	-2.66	-6.22	-1.96	-4.82	-3.80	-6.22	-3.82	0.00092	14.12
miR-145	-3.27	-0.91	-3.52	-1.69	-0.80	-5.20	-4.68	-3.77	-3.39	0.00149	10.51
miR-194-1	-0.42	-1.86	-5.57	-4.79	-1.13	-3.51	-4.74	-3.02	-3.26	0.00212	9.61
miR-202	-3.88	-0.99	NA	NA	-1.72	-3.56	-2.82	-4.13	-3.19	0.00264	9.13
miR-551b	-2.11	-4.16	NA	-2.98	-2.14	-2.98	1.39	-2.83	-2.83	0.01411	7.09
miR-147b	-0.11	-1.02	-3.50	-1.98	-4.14	-6.91	-7.12	-1.28	-2.74	0.01051	6.67
miR-129-1	-4.27	-1.76	1.04	-3.12	-2.31	-4.83	-1.48	-3.61	-2.72	0.00622	6.57
miR-422a	-2.65	NA	-0.37	-3.42	-0.59	NA	NA	-3.77	-2.65	0.03843	6.30
miR-129-2	-6.48	-0.31	-1.02	-1.99	-1.75	-4.89	-2.98	-3.46	-2.49	0.00557	5.60
miR-195	0.87	-0.18	-2.04	-4.54	-2.52	-1.85	-4.09	-2.75	-2.28	0.01265	4.86
miR-658	-2.71	NA	-2.32	-1.17	0.16	NA	-2.60	-2.15	-2.24	0.01038	4.72
miR-497	-2.12	-0.43	-3.68	-3.30	-2.14	-0.70	-3.34	-1.58	-2.13	0.00157	4.37
miR-378	-2.27	-0.05	-2.99	-0.48	-1.57	-5.56	-5.19	-1.11	-1.92	0.01310	3.79
miR-130a	-1.98	0.79	0.00	-2.65	-1.19	-2.45	-2.87	-1.55	-1.77	0.01451	3.40
miR-720	NA	NA	NA	-4.15	-1.65	-2.56	-1.60	-0.58	-1.65	0.02446	3.13
miR-3152	-1.48	NA	NA	-1.76	NA	NA	NA	-1.58	-1.58	0.00247	3.00

miR-10b	-1.75	-4.97	-0.69	-1.16	-0.75	-4.68	-1.35	-3.59	-1.55	0.00690	2.93
miR-378c	0.00	0.01	-3.04	-1.65	-1.31	-6.86	-5.44	0.30	-1.48	0.04877	2.79
miR-605	-1.29	-0.25	NA	-2.79	-1.47	-2.41	-1.11	-3.02	-1.47	0.00354	2.77
miR-29c	-0.32	0.10	-1.40	-2.80	-1.54	-1.73	-3.53	-1.08	-1.47	0.00826	2.77
miR-559	NA	NA	-1.59	-1.17	-2.23	-1.93	-1.23	-1.22	-1.41	0.00032	2.66
miR-138-1	-4.61	0.75	-0.17	-1.77	-3.14	-0.16	-3.60	-1.00	-1.39	0.03822	2.61
up-regulated in Metastases/Normal											
miR-135b	8.35	1.71	2.68	5.95	4.76	0.88	5.82	4.91	4.83	0.00156	28.49
miR-183	5.62	-0.64	3.00	5.40	2.06	1.37	4.19	1.80	2.53	0.00702	5.77
miR-31	2.18	2.50	9.23	-0.48	2.45	5.24	2.02	5.35	2.48	0.01125	5.58
miR-449b	NA	1.59	2.32	NA	0.26	NA	3.91	2.31	2.31	0.02461	4.95
miR-184	3.05	-0.99	4.61	0.83	1.54	5.79	3.02	-0.23	2.28	0.03269	4.86
miR-96	5.04	0.65	0.89	3.52	1.98	-0.26	3.44	2.49	2.24	0.00909	4.72
miR-95	2.57	0.11	1.92	0.82	2.37	5.93	2.64	0.45	2.15	0.01435	4.43
miR-584	1.81	0.65	0.09	3.16	2.09	2.35	2.21	3.83	2.15	0.00221	4.42
miR-1248	1.19	3.01	2.17	3.80	0.57	-0.69	1.94	3.67	2.05	0.00931	4.15
miR-663	NA	-0.99	2.17	1.93	0.89	NA	2.22	3.29	2.05	0.04652	4.14
miR-493	2.64	-1.25	0.73	1.88	2.18	3.88	2.26	1.84	2.03	0.01258	4.08
miR-450b	NA	-0.45	1.87	2.56	1.92	2.11	2.16	1.36	1.92	0.00463	3.78
miR-1975	3.57	-0.26	-0.75	2.11	0.54	1.60	4.40	4.47	1.85	0.03027	3.61
miR-224	2.05	1.90	2.02	2.30	1.58	1.35	1.06	1.63	1.77	0.00001	3.40
miR-3122	NA	NA	NA	NA	1.94	NA	1.57	1.58	1.58	0.00487	3.00
miR-21	1.46	1.55	0.06	2.31	1.76	0.51	0.52	1.84	1.51	0.00285	2.84
miR-542	3.77	1.01	0.41	3.64	0.78	-1.08	2.05	2.00	1.50	0.03001	2.84
miR-4286	1.52	NA	0.00	1.46	1.49	-0.24	0.73	1.58	1.46	0.01939	2.76
miR-503	NA	-0.38	1.30	1.15	1.10	1.71	3.05	1.43	1.30	0.01276	2.46
miR-877	4.16	0.18	1.27	1.19	0.68	2.50	1.99	-0.93	1.23	0.03944	2.34
miR-1292	NA	NA	2.00	1.18	0.84	NA	1.36	-0.03	1.18	0.03254	2.26
miR-181d	1.44	1.12	3.02	1.06	0.73	1.22	1.98	0.27	1.17	0.00249	2.25
miR-1250	NA	0.01	2.80	0.91	1.07	3.50	3.38	0.36	1.07	0.02126	2.10
miR-2355	-0.99	1.86	0.96	0.61	1.45	1.15	1.13	0.65	1.05	0.02488	2.06
miR-450a-2	NA	1.01	2.00	0.68	1.01	1.81	2.37	0.80	1.01	0.00154	2.01

Table A.3: All significant up- and down-regulated microRNAs comparing metastases versus tumor colon samples, sorted by median log2ratio. Analysis of 8 colorectal cancer patients using NGS technology delivered a range of metastases specific deregulated microRNAs. All listed microRNAs sufficed a p-value threshold ≤ 0.05.

ID	*P1*	*P2*	*P3*	*P4*	*P5*	*P6*	*P7*	*P8*	*median*	*p-values*	*fold change*
	log2ratio										
down-regulated in Metastases/Tumor											
miR-1266	-1.87	-1.45	-0.58	0.49	-2.58	-2.80	-1.01	-2.57	-1.66	0.00644	3.16
miR-3139	NA	0.14	NA	-1.00	-2.00	0.01	-1.69	-2.31	-1.35	0.04291	2.54
miR-20a	-1.30	-1.68	-1.58	-0.53	-0.58	-0.97	0.37	-1.18	-1.07	0.00573	2.11

miR-217	-1.11	-2.45	-0.74	-1.42	-0.30	0.86	-3.16	-0.61	-0.92	0.03990	1.89
miR-16-1	-0.27	0.28	-2.21	0.26	-1.76	-0.22	-1.19	-1.47	-0.73	0.04549	1.66
miR-559	-0.74	NA	NA	-0.74	0.00	-0.51	-0.91	-0.57	-0.65	0.00676	1.57
miR-17	-0.69	-2.06	-0.94	-0.44	-0.18	-1.16	0.22	-0.59	-0.64	0.01949	1.56
miR-32	-0.78	-1.18	-0.32	-0.91	-1.16	0.30	-0.33	-0.46	-0.62	0.01116	1.54
miR-18b	-2.34	-1.18	NA	-0.58	-0.32	-0.45	-0.50	-0.98	-0.58	0.01411	1.50
miR-1278	-0.26	-1.80	-0.58	-1.41	0.04	-0.43	0.26	-0.53	-0.48	0.04809	1.39
miR-33b	-0.77	-0.12	-0.17	-0.51	-0.21	-0.48	-1.57	-0.23	-0.35	0.02061	1.28
up-regulated in Metastases/Tumor											
miR-184	3.34	-0.13	3.39	1.69	3.00	3.45	0.84	-0.70	2.35	0.01695	5.09
miR-204	5.59	1.36	2.29	1.93	2.83	2.05	6.84	-0.70	2.17	0.01347	4.50
miR-7-2	0.38	2.14	-0.49	1.06	1.23	-0.58	2.34	1.50	1.14	0.04478	2.21
miR-653	1.00	0.72	NA	0.00	1.46	2.01	0.82	2.34	1.00	0.00769	2.00
miR-3195	0.58	NA	1.00	NA	NA	NA	NA	1.02	1.00	0.02546	2.00
let-7a-3	0.70	1.14	0.45	0.14	0.49	-0.01	0.00	0.00	0.29	0.04146	1.23

Table A.4: Significant top 30 up- and down-regulated microRNAs comparing the colon cancer cell lines SW480 and its metastases cell line SW620, sorted by median log2ratio. Deep sequencing data for two experiments using NGS technology.

ID	***log2ratio***	***Stdev***	***fold change***
down-regulated SW620/SW480			
miR-1323	-11.60	1.26	3103.02
miR-372	-9.58	0.35	762.85
miR-371	-9.18	0.51	578.29
miR-599	-9.12	0.75	556.36
miR-34c	-8.43	0.82	344.10
miR-342	-7.37	1.57	164.88
miR-130a	-6.00	2.67	64.01
miR-509-3	-5.00	0.37	32.06
miR-143	-3.53	0.54	11.51
miR-145	-3.52	3.72	11.45
miR-200c	-3.42	0.50	10.73
miR-139	-3.36	0.43	10.25
miR-141	-3.29	0.23	9.79
miR-30a	-3.29	0.95	9.77
miR-551a	-3.14	0.09	8.84
miR-1262	-2.92	0.99	7.54
miR-1274b	-2.79	3.64	6.93
miR-146b	-2.53	0.84	5.78
miR-455	-2.48	0.70	5.58
miR-193a	-2.43	0.36	5.41
miR-184	-2.35	0.67	5.12
miR-203	-2.35	0.49	5.11

miR-589	-2.33	0.50	5.02
miR-26a-1	-2.30	0.82	4.94
miR-3190	-2.29	0.93	4.91
miR-1301	-2.23	1.03	4.68
miR-144	-2.22	2.39	4.66
miR-1308	-2.13	3.13	4.39
miR-95	-2.07	0.10	4.20
miR-148a	-2.00	1.18	3.99
up-regulated SW620/SW480			
miR-142	4.65	0.49	25.16
miR-1252	4.62	0.86	24.59
miR-31	3.98	0.40	15.80
miR-375	3.92	1.69	15.11
miR-466	3.69	0.64	12.91
miR-552	3.62	0.72	12.27
miR-181a-1	3.21	0.32	9.25
miR-3200	2.92	0.78	7.57
miR-582	2.78	0.27	6.85
miR-215	2.64	3.06	6.24
miR-577	2.52	0.58	5.74
miR-548t	2.30	1.31	4.92
miR-100	2.28	0.20	4.85
miR-664	2.26	0.35	4.78
miR-3140	2.23	1.08	4.71
miR-1-2	2.18	1.10	4.52
miR-196a-2	2.03	0.28	4.10
miR-194-1	1.98	1.33	3.94
miR-181b-1	1.95	0.41	3.87
miR-151	1.91	0.60	3.75
miR-19b-1	1.90	0.36	3.74
miR-3202-1	1.87	0.60	3.65
miR-3159	1.67	2.07	3.18
miR-3148	1.53	0.89	2.88
miR-1278	1.52	0.21	2.87
miR-1255a	1.51	1.26	2.85
miR-192	1.49	0.24	2.81
miR-3173	1.47	0.74	2.77
miR-196a-1	1.46	0.35	2.76
miR-548e	1.43	0.24	2.70

Table A.5: Expression of miR-1 in tumor tissue screen using TaqMan technology. The expression values of miR-1 were estimated for every tissue using the delta delta Ct method. Normalization was performed against a stable internal control RNU44 and expression values were compared to the corresponding normal tissue ($2^{-(\Delta\Delta Ct)}$) = fold change).

	$2^{-(\Delta\Delta Ct)}$ value	log2
adipose tissue tumor	1.14	0.19
brain tumor	0.22	-2.21
breast tumor	0.16	-2.68
colon tumor	0.12	-3.09
endometrium tumor	0.34	-1.56
kidney tumor	0.27	-1.90
liver tumor	0.25	-1.98
lung tumor	0.21	-2.27
lymph node tumor	0.36	-1.46
muscle tumor	0.05	-4.34
ovary tumor	0.01	-6.47
pancreas tumor	0.68	-0.55
prostate tumor	0.02	-5.52
stomach tumor	0.22	-2.20
testis tumor	0.14	-2.80
thyroid gland tumor	0.13	-2.97

Table A.6: Expression of miR-135b in tumor tissue screen using TaqMan technology. The expression values of miR-135b were estimated for every tissue using the delta delta Ct method. Normalization was performed against a stable internal control RNU44 and expression values were compared to the corresponding normal tissue ($2^{-(\Delta\Delta Ct)}$ = fold change).

	$2^{-(\Delta\Delta Ct)}$ value	log2
adipose tissue tumor	2.27	1.18
brain tumor	2.59	1.38
breast tumor	0.44	-1.17
colon tumor	9.24	3.21
endometrium tumor	15.45	3.95
kidney tumor	0.71	-0.49
liver tumor	0.34	-1.56
lung tumor	1.48	0.57
lymph node tumor	0.45	-1.15
muscle tumor	0.64	-0.64
ovary tumor	1.71	0.78
pancreas tumor	0.29	-1.76
prostate tumor	0.09	-3.45
stomach tumor	5.71	2.51
testis tumor	0.35	-1.53
thyroid gland tumor	0.39	-1.36

Table A.7: Expression of miR-129 in tumor tissue screen using TaqMan technology. The expression values of miR-129 were estimated for every tissue using the delta delta Ct method. Normalization was performed against a stable internal control RNU44 and expression values were compared to the corresponding normal tissue ($2^{-(\Delta\Delta Ct)}$ = fold change).

	$2^{-(\Delta\Delta Ct)}$ value	log2
adipose tissue tumor	0.74	-0.44
brain tumor	0.01	-6.67
breast tumor	0.26	-1.93
colon tumor	0.07	-3.94
endometrium tumor	0.78	-0.37
kidney tumor	0.07	-3.80
liver tumor	0.39	-1.36
lung tumor	0.53	-0.91
lymph node tumor	0.45	-1.16
muscle tumor	0.02	-5.89
ovary tumor	0.17	-2.54
pancreas tumor	0.13	-2.91
prostate tumor	0.14	-2.81
stomach tumor	0.26	-1.92
testis tumor	0.01	-6.44
thyroid gland tumor	0.15	-2.75

Table A.8: Expression of miR-493 in tumor tissue screen using TaqMan technology. The expression values of miR-493 were estimated for every tissue using the delta delta Ct method. Normalization was performed against a stable internal control RNU44 and expression values were compared to the corresponding normal tissue ($2^{-(\Delta\Delta Ct)}$ = fold change).

	$2^{-(\Delta\Delta Ct)}$ value	log2
adipose tissue tumor	9.24	3.21
brain tumor	4.88	2.29
breast tumor	0.61	-0.71
colon tumor	0.75	-0.42
endometrium tumor	0.60	-0.74
kidney tumor	0.24	-2.03
liver tumor	0.23	-2.10
lung tumor	0.78	-0.35
lymph node tumor	0.34	-1.57
muscle tumor	0.36	-1.46
ovary tumor	0.78	-0.36
pancreas tumor	0.55	-0.88
prostate tumor	0.26	-1.95
stomach tumor	1.07	0.10
testis tumor	0.17	-2.59
thyroid gland tumor	0.90	-0.15

Table A.9: Expression of miR-215 in tumor tissue screen using TaqMan technology. The expression values of miR-215 were estimated for every tissue using the delta delta Ct method. Normalization was performed against a stable internal control RNU44 and expression values were compared to the corresponding normal tissue ($2^{-(\Delta\Delta Ct)}$ = fold change).

	$2^{-(\Delta\Delta Ct)}$ value	log2
adipose tissue tumor	0.44	-1.18
brain tumor	1.01	0.01
breast tumor	0.65	-0.62
colon tumor	0.18	-2.46
endometrium tumor	0.35	-1.50
kidney tumor	0.24	-2.06
liver tumor	0.22	-2.16
lung tumor	0.21	-2.28
lymph node tumor	1.40	0.49
muscle tumor	0.09	-3.54
ovary tumor	0.44	-1.17
pancreas tumor	0.20	-2.36
prostate tumor	0.27	-1.88
stomach tumor	0.49	-1.03
testis tumor	0.15	-2.77
thyroid gland tumor	0.34	-1.57

Table A.10: Expression of miR-497 in tumor tissue screen using TaqMan technology. The expression values of miR-497 were estimated for every tissue using the delta delta Ct method. Normalization was performed against a stable internal control RNU44 and expression values were compared to the corresponding normal tissue ($2^{-(\Delta\Delta Ct)}$ = fold change).

	$2^{-(\Delta\Delta Ct)}$ value	log2
adipose tissue tumor	0.64	-0.65
brain tumor	0.98	-0.02
breast tumor	0.11	-3.18
colon tumor	0.16	-2.66
endometrium tumor	0.21	-2.24
kidney tumor	1.00	0.00
liver tumor	0.19	-2.43
lung tumor	0.28	-1.82
lymph node tumor	0.54	-0.88
muscle tumor	0.41	-1.28
ovary tumor	0.04	-4.64
pancreas tumor	0.27	-1.90
prostate tumor	0.33	-1.62
stomach tumor	0.40	-1.34
testis tumor	0.16	-2.60
thyroid gland tumor	0.34	-1.54

Table A.11: Summary of clinical information for the all patients investigated in this study (N = normal, T = tumor, M = metastasis, NGS = next generation sequencing, MS = microsatellite).

ID	*Patient*		*Performed analysis*	*Sex*	*Age*	*Organ*	*MS-Status*
P1N	P1	Normal	NGS, qPCR	M	70	Sigmoid	MSS
P1T		Tumor	NGS, qPCR				
P1M		Metastasis	NGS, qPCR			Liver	
P2N	P2	Normal	NGS, qPCR	M	74	Coecum	MSS
P2T		Tumor	NGS, qPCR				
P2M		Metastasis	NGS, qPCR			Liver	
P3N	P3	Normal	NGS, qPCR	M	44	Colon asc.	MSI
P3T		Tumor	NGS, qPCR				
P3M		Metastasis	NGS, qPCR			Liver	
P4N	P4	Normal	NGS, qPCR	F	49	Rectum/ Sigmoid	MSS
P4T		Tumor	NGS, qPCR				
P4M		Metastasis	NGS, qPCR			Liver	
P5N	P5	Normal	NGS, qPCR	F	80	Colon	MSI
P5T		Tumor	NGS, qPCR				
P5M		Metastasis	NGS, qPCR			Lymph node	
P6N	P6	Normal	NGS, qPCR	F	66	Colon asc.	MSS
P6T		Tumor	NGS, qPCR				
P6M		Metastasis	NGS, qPCR			Liver	
P7N	P7	Normal	NGS, qPCR	F	76	Colon asc.	MSS
P7T		Tumor	NGS, qPCR				
P7M		Metastasis	NGS, qPCR			Lymph node	
P8N	P8	Normal	NGS, qPCR	M	74	Coecum	MSS
P8T		Tumor	NGS, qPCR				
P8M		Metastasis	NGS, qPCR			Liver	
P9N	I	Normal	MeDIP, qPCR, NGS	F	72	Rectum	MSS
P9T		Adenocarcinoma	MeDIP, qPCR, NGS				
P10N	II	Normal	MeDIP, qPCR	M	73	Colon	MSS
P10T		Tubular Adenocarcinoma	MeDIP, qPCR				
P11N	III	Normal	MeDIP, qPCR	M	85	Rectum	MSS
P11T		Tubular Adenocarcinoma	MeDIP, qPCR				
P12N	IV	Normal	MeDIP, qPCR	F	45	Colon	MSI
P12T		Mucinious Adenocarcinoma	MeDIP, qPCR				
P13N	V	Normal	MeDIP, qPCR, NGS	M	71	Rectum	MSS
P13T		Adenocarcinoma	MeDIP, qPCR, NGS				
P14N	VI	Normal	MeDIP, qPCR	M	52	Sigmoid	MSS
P14T		Tubular Adenocarcinoma	MeDIP, qPCR				
P15N	VII	Normal	MeDIP, qPCR	F	82	Rectum	MSS
P15T		Tubular Adenocarcinoma	MeDIP, qPCR				
P15aN	VIII	Normal	MeDIP, qPCR, NGS	M	50	Rectum	MSS
P15aT		Tubular Adenocarcinoma	MeDIP, qPCR, NGS				
P16T	-	Tumor	qPCR (tumor screen)	-	-	Colon	-

P17T	-	Tumor	qPCR (tumor screen)	-	-	Colon	-
P18T	-	Tumor	qPCR (tumor screen)	-	-	Colon	-
P19T	-	Tumor	qPCR (tumor screen)	-	-	Colon	-
P20T	-	Tumor	qPCR (tumor screen)	-	-	Colon	-
P21T	-	Tumor	qPCR (tumor screen)	-	-	Colon	-
P22T	-	Tumor	qPCR (tumor screen)	-	-	Colon	-
P23T	-	Tumor	qPCR (tumor screen)	-	-	Colon	-
P24T	-	Tumor	qPCR (tumor screen)	-	-	Colon	-
P25T	-	Tumor	qPCR (tumor screen)	-	-	Colon	-
P26N	-	Normal	qPCR (tumor screen)	-	-	Colon	-
P27N	-	Normal	qPCR (tumor screen)	-	-	Colon	-
P28N	-	Normal	qPCR (tumor screen)	-	-	Colon	-
P29N	-	Normal	qPCR (tumor screen)	-	-	Colon	-
P30N	-	Normal	qPCR (tumor screen)	-	-	Colon	-
P31N		Normal	NGS	-	-	Colon	-
P31T		Adenocarcinoma	NGS	-	-	Colon	-
P32N		Normal	NGS	-	-	Colon	-
P32T		Adenocarcinoma	NGS	-	-	Colon	-
P33N		Normal	NGS	-	-	Colon	-
P33T		Adenocarcinoma	NGS	-	-	Colon	-

Table A.12: Potential novel microRNAs identified in colon cancer cell lines (SW480, SW620) using next generation sequencing technology. The determined expression level in both cell lines is represented as well as the microRNA classification score. Five selected microRNAs, highlighted in red, were chosen for further northern blot analyzes.

Location	***Normalized expression (SW480)***	***Normalized expression (SW620)***	***FoldChange (Set1/set2)***	***Annotation***	***microRNA classification***
	Set1	Set2			
chr1:237594712-237594778(-)	39.85	100.13	0.40	NA	0.51
chr6:28709855-28709910(-)	38.22	128.41	0.30	NA	0.68
chr5:74099189-74099224(-)	0	9263.07	0.00	NA	0.55
chr1:33570591-33570662(-)	235.76	226.93	1.04	NA	0.56
chr12:88270272-88270401(-)	42.34	344.04	0.12	NA	0.6
chr12:112309970-112310026(-)	0	212.51	NA	NA	0.96
chr5:74099186-74099307(-)	9371.76	0	NA	NA	0.55
chr5:172151616-172151656(+)	21.34	125.57	0.17	NA	0.6
chr10:115923860-115923927(-)	409.02	333.28	1.23	NA	0.85
chr15:78921389-78921449(+)	0.00	125.88	NA	NA	0.87
chr1:1093116-1093179(-)	0.00	284.75	NA	NA	0.77
chr6:36698200-36698275(-)	**298.62**	**0**	**NA**	**NA**	**0.99**
chr13:49468564-49468649(-)	**113.39**	**46.31**	**2.45**	**NA**	**1**
chr18:28674580-28674638(-)	159.99	196.14	0.82	NA	0.82
chr3:153006625-153006655(-)	32.89	108.99	0.30	NA	0.65

chr17:44045104-44045156(+)	58.71	588.75	0.10	NA	0.8
chr19:44650704-44650771(+)	**1029.65**	**462.25**	**2.23**	**NA**	**0.72**
chr17:76714262-76714364(+)	43.00	219.61	0.20	NA	1
chr17:53515778-53515889(+)	63.17	101.37	0.62	NA	0.56
chr7:145325367-145325455(-)	0.00	110.37	NA	NA	0.66
chr2:207356215-207356272(+)	43.33	140.35	0.31	NA	0.81
chr14:19881213-19881412(-)	196.36	330.67	0.59	NA	0.53
chr2:2556921-2556996(+)	4211.01	1298.18	3.24	NA	0.52
chr3:31178211-31178270(-)	0	112.25	NA	NA	1
chr2:140694407-140694519(-)	140.84	698.57	0.20	NA	0.53
chr16:359152-359180(-)	15586.66	25981.90	0.60	NA	0.91
chr11:74792736-74792874(+)	141.21	14.53	9.72	NA	0.89
chr5:93932003-93932084(+)	22.49	2410.78	0.01	NA	0.52
chr21:43999497-43999621(+)	108.69	83.83	1.30	NA	0.87
chr2:225051201-225051297(-)	75.50	965.42	0.08	NA	0.81
chr8:41637119-41637184(+)	114.53	154.67	0.74	NA	0.96
chr9:19268549-19268592(+)	60.30	116.31	0.52	NA	0.52
chr14:19881067-19881186(-)	40.69	330.67	0.12	NA	0.53
chr17:40581621-40581693(+)	123.89	151.02	0.82	NA	0.59
chr17:70256349-70256434(+)	**107.75**	**209.57**	**0.51**	**NA**	**0.99**
chr4:90872081-90872164(+)	**1708.34**	**5275.19**	**0.32**	**NA**	**0.55**
chr17:33043487-33043536(-)	66.59	113.99	0.58	NA	0.53
chr5:11083170-11083235(-)	59.51	120.21	0.50	NA	0.84
chr14:65007580-65007645(-)	288.28	794.43	0.36	NA	0.76
Control (U6)					
chr16:2144983-2145107(-)	408.88				

Table A.13: List of Mutations determined in microRNA regions for tumor (T) and normal (N) tissue of six colon cancer patients.

Gene name	***Location***	***Reference***	***Mutation***	***annotated in dbSNP database or unknown***
Patient P9N				
hsa-mir-146a	chr5:159844996-159844996	C	G	dbSNP: rs2910164
hsa-mir-147b	chr15:43512544-43512544	A	G	unknown
hsa-mir-182	chr7:129197463-129197463	C	T	unknown
hsa-mir-196a-2	chr12:52671866-52671866	C	T	dbSNP: rs11614913
hsa-mir-27a	chr19:13808292-13808292	T	C	dbSNP: rs895819
hsa-mir-450b	chrX:133501893-133501893	G	T	unknown
hsa-mir-453	chr14:100592309-100592309	T	C	dbSNP: rs56103835
hsa-mir-492	chr12:93752417-93752417	G	C	dbSNP: rs2289030
hsa-mir-499	chr20:33041863-33041863	G	A	unknown
hsa-mir-532	chrX:49654575-49654575	A	G	dbSNP: rs456617
hsa-mir-618	chr12:79853667-79853667	A	C	dbSNP: rs2682818

hsa-mir-646	chr20:58316929-58316929	T	C	dbSNP: rs6513496
hsa-mir-656	chr14:100602846-100602846	C	T	dbSNP: rs58834075
hsa-mir-663	chr20:26136824-26136824	T	C	dbSNP: rs28670321
hsa-mir-663b	chr2:132731024-132731024	G	T	unknown
hsa-mir-663b	chr2:132731026-132731026	C	T	unknown
hsa-mir-663b	chr2:132731050-132731050	C	T	unknown
hsa-mir-663b	chr2:132731072-132731072	G	C	unknown
hsa-mir-663b	chr2:132731082-132731082	A	C	unknown
hsa-mir-663b	chr2:132731089-132731089	C	T	unknown
hsa-mir-663b	chr2:132731110-132731110	C	T	unknown
Patient P9T				
hsa-mir-1302-3	chr2:114057098-114057098	C	T	dbSNP: rs7589328
hsa-mir-1304	chr11:93106514-93106514	G	T	dbSNP: rs2155248
hsa-mir-1324	chr3:75762663-75762663	A	G	unknown
hsa-mir-1324	chr3:75762683-75762683	C	T	unknown
hsa-mir-146a	chr5:159844996-159844996	C	G	dbSNP: rs2910164
hsa-mir-182	chr7:129197463-129197463	C	T	unknown
hsa-mir-196a-2	chr12:52671866-52671866	C	T	dbSNP: rs11614913
hsa-mir-27a	chr19:13808292-13808292	T	C	dbSNP: rs895819
hsa-mir-300	chr14:100577480-100577480	C	T	dbSNP: rs12894467
hsa-mir-302b	chr4:113789099-113789099	A	G	unknown
hsa-mir-453	chr14:100592309-100592309	T	C	dbSNP: rs56103835
hsa-mir-532	chrX:49654575-49654575	A	G	dbSNP: rs456617
hsa-mir-595	chr7:158018264-158018264	C	T	dbSNP: rs4909237
hsa-mir-604	chr10:29874004-29874004	A	G	dbSNP: rs2368393
hsa-mir-604	chr10:29874009-29874009	G	A	dbSNP: rs2368392
hsa-mir-618	chr12:79853667-79853667	A	C	dbSNP: rs2682818
hsa-mir-532	chrX:49654575-49654575	A	G	dbSNP: rs456617
hsa-mir-646	chr20:58316929-58316929	T	C	dbSNP: rs6513496
hsa-mir-656	chr14:100602846-100602846	C	T	dbSNP: rs58834075
hsa-mir-663	chr20:26136831-26136831	G	T	unknown
hsa-mir-663b	chr2:132731082-132731082	A	C	unknown
Patient P13N				
hsa-mir-1304	chr11:93106514-93106514	G	T	dbSNP: rs2155248
hsa-mir-1307	chr10:105144079-105144079	A	G	dbSNP: rs7911488
hsa-mir-1324	chr3:75762634-75762634	C	T	dbSNP: rs7614638
hsa-mir-146a	chr5:159844996-159844996	C	G	dbSNP: rs2910164
hsa-mir-300	chr14:100577480-100577480	C	T	dbSNP: rs12894467
hsa-mir-412	chr14:100601607-100601607	A	G	dbSNP: rs61992671
hsa-mir-532	chrX:49654575-49654575	A	G	dbSNP: rs456617
hsa-mir-577	chr4:115797446-115797446	C	G	dbSNP: rs34115976
hsa-mir-618	chr12:79853667-79853667	A	C	dbSNP: rs2682818
hsa-mir-532	chrX:49654575-49654575	A	G	dbSNP: rs456617
hsa-mir-663	chr20:26136824-26136824	T	C	dbSNP: rs28670321

hsa-mir-663	chr20:26136912-26136912	A	C	dbSNP: rs7266947
hsa-mir-663b	chr2:132731082-132731082	A	C	unknown
hsa-mir-663b	chr2:132731110-132731110	C	T	unknown
Patient P13T				
hsa-mir-1304	chr11:93106514-93106514	G	T	dbSNP: rs2155248
hsa-mir-1307	chr10:105144079-105144079	A	G	dbSNP: rs7911488
hsa-mir-1324	chr3:75762634-75762634	C	T	dbSNP: rs7614638
hsa-mir-146a	chr5:159844996-159844996	C	G	dbSNP: rs2910164
hsa-mir-300	chr14:100577480-100577480	C	T	dbSNP: rs12894467
hsa-mir-412	chr14:100601607-100601607	A	G	dbSNP: rs61992671
hsa-mir-543	chr14:100568105-100568105	T	G	dbSNP: rs72172875
hsa-mir-577	chr4:115797446-115797446	C	G	dbSNP: rs34115976
hsa-mir-618	chr12:79853667-79853667	A	C	dbSNP: rs2682818
hsa-mir-663	chr20:26136824-26136824	T	C	dbSNP: rs28670321
hsa-mir-663	chr20:26136912-26136912	A	C	dbSNP: rs7266947
hsa-mir-663b	chr2:132731072-132731072	G	C	unknown
hsa-mir-663b	chr2:132731082-132731082	A	C	unknown
Patient P15aN				
hsa-mir-1294	chr5:153706962-153706962	A	G	dbSNP: rs13186787
hsa-mir-1304	chr11:93106514-93106514	G	T	dbSNP: rs2155248
hsa-mir-1307	chr10:105144079-105144079	A	G	dbSNP: rs7911488
hsa-mir-1324	chr3:75762648-75762648	C	T	unknown
hsa-mir-1324	chr3:75762663-75762663	A	G	unknown
hsa-mir-1324	chr3:75762683-75762683	C	T	unknown
hsa-mir-412	chr14:100601607-100601607	A	G	dbSNP: rs61992671
hsa-mir-449b	chr5:54502301-54502301	A	G	dbSNP: rs10061133
hsa-mir-518d	chr19:58930020-58930020	C	T	unknown
hsa-mir-553	chr1:100519431-100519431	A	C	unknown
hsa-mir-597	chr8:9636686-9636686	G	A	unknown
hsa-mir-618	chr12:79853667-79853667	A	C	dbSNP: rs2682818
hsa-mir-663	chr20:26136824-26136824	T	C	dbSNP: rs28670321
hsa-mir-663	chr20:26136912-26136912	A	C	dbSNP: rs7266947
hsa-mir-663b	chr2:132731050-132731050	C	T	unknown
hsa-mir-663b	chr2:132731082-132731082	A	C	unknown
hsa-mir-933	chr2:175740622-175740622	G	A	unknown
Patient P15aT				
hsa-mir-1265	chr10:14518624-14518624	T	C	dbSNP: rs11259096
hsa-mir-1294	chr5:153706962-153706962	A	G	dbSNP: rs13186787
hsa-mir-1304	chr11:93106514-93106514	G	T	dbSNP: rs2155248
hsa-mir-1307	chr10:105144079-105144079	A	G	dbSNP: rs7911488
hsa-mir-1324	chr3:75762634-75762634	C	T	dbSNP: rs7614638
hsa-mir-1324	chr3:75762648-75762648	C	T	unknown
hsa-mir-1324	chr3:75762663-75762663	A	G	unknown
hsa-mir-1324	chr3:75762683-75762683	C	T	unknown

hsa-mir-196a-2	chr12:52671866-52671866	C	T	dbSNP: rs11614913
hsa-mir-300	chr14:100577480-100577480	C	T	dbSNP: rs12894467
hsa-mir-196a-2	chr12:52671866-52671866	C	T	dbSNP: rs11614913
hsa-mir-412	chr14:100601607-100601607	A	G	dbSNP: rs61992671
hsa-mir-449b	chr5:54502301-54502301	A	G	dbSNP: rs10061133
hsa-mir-518d	chr19:58930020-58930020	C	T	unknown
hsa-mir-519d	chr19:58908461-58908461	T	C	unknown
hsa-mir-532	chrX:49654572-49654572	A	G	unknown
hsa-mir-532	chrX:49654575-49654575	A	G	dbSNP: rs456617
hsa-mir-597	chr8:9636686-9636686	G	A	unknown
hsa-mir-618	chr12:79853667-79853667	A	C	dbSNP: rs2682818
hsa-mir-635	chr17:63932236-63932236	C	A	unknown
hsa-mir-663	chr20:26136912-26136912	A	C	dbSNP: rs7266947
hsa-mir-663b	chr2:132731082-132731082	A	C	unknown
hsa-mir-663b	chr2:132731089-132731089	C	T	unknown
hsa-mir-663b	chr2:132731110-132731110	C	T	unknown
hsa-mir-933	chr2:175740622-175740622	G	A	unknown
Patient P31N				
hsa-mir-124-2	chr8:65454319-65454319	A	C	unknown
hsa-mir-125b-2	chr21:16884457-16884457	A	C	unknown
hsa-mir-1290	chr1:19096158-19096158	A	G	unknown
hsa-mir-1304	chr11:93106514-93106514	G	T	dbSNP: rs2155248
hsa-mir-1324	chr3:75762663-75762663	A	G	unknown
hsa-mir-16-1	chr13:49521111-49521111	T	C	unknown
hsa-mir-300	chr14:100577480-100577480	C	T	dbSNP: rs12894467
hsa-mir-520f	chr19:58877304-58877304	G	A	unknown
hsa-mir-608	chr10:102724768-102724768	C	G	dbSNP: rs4919510
hsa-mir-618	chr12:79853667-79853667	A	C	dbSNP: rs2682818
hsa-mir-646	chr20:58316929-58316929	T	C	dbSNP: rs6513496
hsa-mir-663	chr20:26136824-26136824	T	C	dbSNP: rs28670321
hsa-mir-663	chr20:26136912-26136912	A	C	dbSNP: rs7266947
hsa-mir-663b	chr2:132731050-132731050	C	T	unknown
hsa-mir-663b	chr2:132731082-132731082	A	C	unknown
hsa-mir-663b	chr2:132731110-132731110	C	T	unknown
Patient P31T				
hsa-mir-1290	chr1:19096158-19096158	A	G	unknown
hsa-mir-1304	chr11:93106514-93106514	G	T	dbSNP: rs2155248
hsa-mir-1324	chr3:75762634-75762634	C	T	dbSNP: rs7614638
hsa-mir-146a	chr5:159844996-159844996	C	G	dbSNP: rs2910164
hsa-mir-16-1	chr13:49521111-49521111	T	C	unknown
hsa-mir-27a	chr19:13808292-13808292	T	C	dbSNP: rs895819
hsa-mir-300	chr14:100577480-100577480	C	T	dbSNP: rs12894467
hsa-mir-449b	chr5:54502291-54502291	G	A	unknown
hsa-mir-520f	chr19:58877304-58877304	G	A	unknown

hsa-mir-608	chr10:102724768-102724768	C	G	dbSNP: rs4919510
hsa-mir-618	chr12:79853667-79853667	A	C	dbSNP: rs2682818
hsa-mir-646	chr20:58316929-58316929	T	C	dbSNP: rs6513496
hsa-mir-663	chr20:26136824-26136824	T	C	dbSNP: rs28670321
hsa-mir-663	chr20:26136912-26136912	A	C	dbSNP: rs7266947
hsa-mir-663b	chr2:132731082-132731082	A	C	unknown
hsa-mir-663b	chr2:132731110-132731110	C	T	unknown
Patient P32N				
hsa-mir-125a	chr19:56888396-56888396	C	G	unknown
hsa-mir-1304	chr11:93106514-93106514	G	T	dbSNP: rs2155248
hsa-mir-1307	chr10:105144079-105144079	A	G	dbSNP: rs7911488
hsa-mir-146a	chr5:159844996-159844996	C	G	dbSNP: rs2910164
hsa-mir-412	chr14:100601607-100601607	A	G	dbSNP: rs61992671
hsa-mir-485	chr14:100591562-100591562	C	T	unknown
hsa-mir-596	chr8:1752832-1752832	T	C	unknown
hsa-mir-604	chr10:29874004-29874004	A	G	dbSNP: rs2368393
hsa-mir-604	chr10:29874009-29874009	G	A	dbSNP: rs2368392
hsa-mir-618	chr12:79853667-79853667	A	C	dbSNP: rs2682818
hsa-mir-663	chr20:26136824-26136824	T	C	dbSNP: rs28670321
hsa-mir-663b	chr2:132731050-132731050	C	T	unknown
hsa-mir-663b	chr2:132731072-132731072	G	C	unknown
hsa-mir-663b	chr2:132731082-132731082	A	C	unknown
hsa-mir-663b	chr2:132731089-132731089	C	T	unknown
hsa-mir-663b	chr2:132731110-132731110	C	T	unknown
Patient P32T				
hsa-mir-1302-3	chr2:114057098-114057098	C	T	dbSNP: rs7589328
hsa-mir-1304	chr11:93106514-93106514	G	T	dbSNP: rs2155248
hsa-mir-1307	chr10:105144079-105144079	A	G	dbSNP: rs7911488
hsa-mir-1324	chr3:75762648-75762648	C	T	unknown
hsa-mir-1324	chr3:75762663-75762663	A	G	unknown
hsa-mir-146a	chr5:159844996-159844996	C	G	dbSNP: rs2910164
hsa-mir-27a	chr19:13808292-13808292	T	C	dbSNP: rs895819
hsa-mir-300	chr14:100577480-100577480	C	T	dbSNP: rs12894467
hsa-mir-596	chr8:1752832-1752832	T	C	unknown
hsa-mir-604	chr10:29874004-29874004	A	G	dbSNP: rs2368393
hsa-mir-604	chr10:29874009-29874009	G	A	dbSNP: rs2368392
hsa-mir-618	chr12:79853667-79853667	A	C	dbSNP: rs2682818
hsa-mir-663	chr20:26136824-26136824	T	C	dbSNP: rs28670321
hsa-mir-663	chr20:26136912-26136912	A	C	dbSNP: rs7266947
hsa-mir-663b	chr2:132731050-132731050	C	T	unknown
hsa-mir-663b	chr2:132731072-132731072	G	C	unknown
hsa-mir-663b	chr2:132731082-132731082	A	C	unknown
hsa-mir-663b	chr2:132731110-132731110	C	T	unknown
hsa-mir-933	chr2:175740622-175740622	G	A	unknown

Patient P33N				
hsa-mir-1208	chr8:129231615-129231615	G	T	unknown
hsa-mir-1304	chr11:93106514-93106514	G	T	dbSNP: rs2155248
hsa-mir-1307	chr10:105144079-105144079	A	G	dbSNP: rs7911488
hsa-mir-1324	chr3:75762634-75762634	C	T	dbSNP: rs7614638
hsa-mir-1324	chr3:75762683-75762683	C	T	unknown
hsa-mir-300	chr14:100577480-100577480	C	T	dbSNP: rs12894467
hsa-mir-412	chr14:100601607-100601607	A	G	dbSNP: rs61992671
hsa-mir-492	chr12:93752417-93752417	G	C	dbSNP: rs2289030
hsa-mir-570	chr3:196911485-196911485	T	C	unknown
hsa-mir-597	chr8:9636686-9636686	G	A	unknown
hsa-mir-604	chr10:29874004-29874004	A	G	dbSNP: rs2368393
hsa-mir-604	chr10:29874009-29874009	G	A	dbSNP: rs2368392
hsa-mir-618	chr12:79853667-79853667	A	C	dbSNP: rs2682818
hsa-mir-646	chr20:58316929-58316929	T	C	dbSNP: rs6513496
hsa-mir-663	chr20:26136824-26136824	T	C	dbSNP: rs28670321
hsa-mir-663	chr20:26136912-26136912	A	C	dbSNP: rs7266947
hsa-mir-663b	chr2:132731050-132731050	C	T	unknown
hsa-mir-663b	chr2:132731082-132731082	A	C	unknown
hsa-mir-663b	chr2:132731089-132731089	C	T	unknown
hsa-mir-663b	chr2:132731110-132731110	C	T	unknown
Patient P33T				
hsa-mir-1208	chr8:129231615-129231615	G	T	unknown
hsa-mir-1304	chr11:93106514-93106514	G	T	dbSNP: rs2155248
hsa-mir-1307	chr10:105144079-105144079	A	G	dbSNP: rs7911488
hsa-mir-1324	chr3:75762634-75762634	C	T	dbSNP: rs7614638
hsa-mir-1324	chr3:75762653-75762653	C	T	unknown
hsa-mir-1324	chr3:75762683-75762683	C	T	unknown
hsa-mir-146a	chr5:159844996-159844996	C	G	dbSNP: rs2910164
hsa-mir-412	chr14:100601607-100601607	A	G	dbSNP: rs61992671
hsa-mir-492	chr12:93752417-93752417	G	C	dbSNP: rs2289030
hsa-mir-570	chr3:196911485-196911485	T	C	unknown
hsa-mir-597	chr8:9636686-9636686	G	A	unknown
hsa-mir-604	chr10:29874004-29874004	A	G	dbSNP: rs2368393
hsa-mir-612	chr11:64968555-64968555	G	A	dbSNP: rs12803915
hsa-mir-613	chr12:12808925-12808925	T	G	unknown
hsa-mir-618	chr12:79853667-79853667	A	C	dbSNP: rs2682818
hsa-mir-646	chr20:58316929-58316929	T	C	dbSNP: rs6513496
hsa-mir-646	chr20:58317000-58317000	T	G	unknown
hsa-mir-663	chr20:26136824-26136824	T	C	dbSNP: rs28670321
hsa-mir-663	chr20:26136905-26136905	C	T	unknown
hsa-mir-663	chr20:26136912-26136912	A	C	dbSNP: rs7266947
hsa-mir-663b	chr2:132731072-132731072	G	C	unknown
hsa-mir-663b	chr2:132731082-132731082	A	C	unknown

Table A.14: Sequencing statistics for CRC patients analyzed.

sample		*All reads*	*All aligned reads*	*Unique aligned reads*	*NGS technology*
Patient 1	normal tissue	5,678,446	5,659,707	5,180,233	454 Roche
	tumor tissue	5,729,335	5,569,487	5,285,822	454 Roche
Patient 2	normal tissue	2,496,558	2,425,905	2,304,598	454 Roche
	tumor tissue	4,781,983	4,624,656	4,367,855	454 Roche
Patient 3	normal tissue	11,048,717	3,624,645	3,348,967	Illumina
	tumor tissue	10,095,338	3,537,675	3,157,577	Illumina
	metastasis tissue	18,327,918	4,864,471	4,116,569	Illumina
Patient 4	normal tissue	9,941,706	3,887,314	3,469,402	Illumina
	tumor tissue	9,770,946	3,726,702	3,324,821	Illumina
	metastasis tissue	11,597,294	4,310,802	3,880,557	Illumina

SELBSTSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, Christina Röhr, dass ich die vorliegende Dissertationsschrift mit dem Titel ‘Analyses of regulatory mechanisms and identification of microRNAs as biomarkers in colorectal cancer’ selbstständig und unter Verwendung keiner anderen als der von mir angegebenen Quellen und Hilfsmittel verfasst habe. Ferner erkläre ich, dass ich bisher weder an der Freien Universität Berlin noch anderweitig versucht habe, eine Dissertation einzureichen oder mich einer Doktorprüfung zu unterziehen.

Datum Unterschrift