DNA methylation biomarkers for colorectal cancer detection: *CDO1*, *DCLK1*, *ZNF331*, and *ZSCAN18*

Hege Marie Vedeld

A thesis for the Master's degree in molecular biosciences, December 2011
60 study points

Department of Molecular Biosciences
Faculty of Mathematics and Natural Sciences
University of Oslo

Department of Cancer Prevention
Institute for Cancer Research
The Norwegian Radium Hospital
Oslo University Hospital

Centre for Cancer Biomedicine
Faculty of Medicine
University of Oslo

DNA methylation biomarkers for colorectal cancer detection: *CDO1*, *DCLK1*, *ZNF331*, and *ZSCAN18*

2011

DNA methylation biomarkers for colorectal cancer detection: *CDO1*, *DCLK1*, *ZNF331*, and *ZSCAN18*

Hege Marie Vedeld

http://www.duo.uio.no/

Table of contents

Acknowledgements

The work presented in this thesis was carried out in the project Group of Epigenetics, Department of Cancer Prevention, Oslo University Hospital- Radiumhospitalet, in the period August 2010 to December 2011.

First and foremost, I would like to thank my supervisor Guro E. Lind, for her great support and guiding, and for always taking time to answer my questions. Her high standards, thoroughness, and knowledge, but also her kindness and compassion have been truly inspiring. I am grateful to my co- supervisor and head of department Ragnhild Lothe, for giving me the opportunity to work with this interesting project, and to be part of a wonderful group, both scientific and socially. Thanks to everybody who have helped me in the lab and answered my questions, especially Hilde, Kim, Deeqa and Anne Cathrine. Extra thank to Deeqa and Anne Cathrine, for all our scientific conversations and for being good friends.

Special thanks to my family and friends, who mean everything to me; my brother and sister, Pål Tore and Ane Eirin, for always being there for me and for believing in me, and my two "extra sisters" Cathrine and Cathrine, for being the best friends ever. Growing up without you would definitely not be the same, and our daily "study-groups" were/are highly appreciated ;) My parents have been wonderful, and without their invaluable help this thesis would not have been possible. Thank you so much for believing in me, for encouraging me throughout my entire education, and for all support and love especially the last three years. I am truly grateful. Special thanks also go to Vilde Marie, my beautiful little- big 2 year old, for being exactly as she is.

Oslo, December 2011

Hege Marie Vedeld

Summary

Colorectal cancer is one of the most prevalent causes of cancer deaths worldwide, with an estimated 1500 deaths each year in Norway alone. Early detection of colorectal cancer may significantly reduce this number, and it is therefore of great interest to identify biomarkers that can be used in a reliable non-invasive test for early detection.

Aberrant promoter DNA methylation has great potential as diagnostic biomarkers. They are both prevalent in cancer and have been shown to be an early event in tumor development. These changes can further be detected in feces and blood, materials suitable for non-invasive testing.

The present thesis is part of an ongoing project where we have set up a step-wise experimental protocol to identify DNA methylation biomarkers for cholangiocarcinomas – a malignancy arising in the bile ducts. We and others have previously shown that gastrointestinal cancers frequently display similar epigenetic aberrations, and the main focus of this thesis was to evaluate whether the identified candidates could be used as biomarkers for early detection of colorectal cancer.

From the cholangiocarcinoma approach, 43 genes were identified as potential epigenetically deregulated genes. These genes were investigated by methylation specific PCR (MSP) in cancer cell lines (n=24). Twelve- and thirteen genes were frequently hypermethylated in the cholangiocarcinoma- and colon cancer cell lines, respectively, and were selected for further analysis in a pilot of primary tumors and normal samples from the respective malignancies.

Four genes *CDO1*, *DCLK1*, *ZNF331*, and *ZSCAN18* were found to be methylated in ≥75% of colorectal cancer samples and simultaneously weakly methylated/ unmethylated in ≥80% of the normal samples. These genes were subjected to quantitative real-time MSP (qMSP). Methylation of at least one of the four genes was observed in 62 of the 65 colorectal cancers analyzed (95% sensitivity), and in two out of the 50 normal mucosa samples (96% specificity), with a combined area under the Receiver Operating Characteristics (ROC) curve (AUC) of 0.976. The only significant association when comparing methylation status with clinicopathological features and tumor phenotype, was observed between *ZSCAN18* and microsatellite instability (MSI), indicating that aberrant methylation of the four genes is present in all tumor subtypes independent of age, gender and stage. A patent application

covering these markers has been filed, and we are currently collaborating with an industrial partner to develop a non- invasive test for early detection of colorectal cancer based on these and previously published results.

To conclude: *CDO1*, *DCLK1*, *ZNF331*, and *ZSCAN18* have been identified as novel promising DNA methylation biomarkers for early detection of colorectal cancer.

1 Introduction

1.1 Cancer as a genetic and epigenetic disease

The term cancer includes more than 200 different diseases, defined by uncontrolled cell proliferation. Cancer can arise from most of the cell types and organs within the human body [1] and demonstrates heterogeneity at the molecular, histopathological and clinical level.

Sequential accumulation of genetic and epigenetic alterations will ultimately allow a cell to overcome the multiple regulatory mechanisms that maintain homeostasis in the organ, transforming normal cells into highly malignant derivates [2]. Six hallmarks, shared by most types of human cancers, have been proposed to dictate malignant growth and include; self-sufficiency in growth signals, evasion of apoptosis, insensitivity to growth inhibitory signals, sustained angiogenesis, limitless replicative potential, and tissue invasion and metastasis [3]. Recently, Hanahan and Weinberg, the authors of the original review, suggested two novel hallmarks for cancer cells; deregulating cellular energetics and avoiding immune destruction [4]. Together, these changes create growth advantages, which through clonal evolution can lead to the outgrowth of progressively more malignant cells, and eventually cancer development [4].

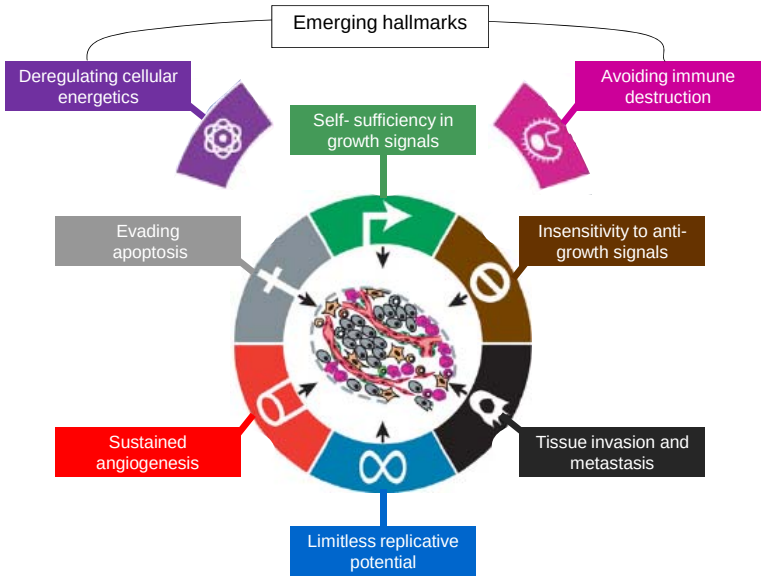


Figure 1 The hallmarks of cancer, including the six well established- and the two emerging hallmarks. These acquired capabilities are thought to be shared by most type of human cancers, and to dictate malignant growth. Modified from [4].

The cancer cell hallmarks are largely acquired through deregulation of three classes of genes: proto- oncogenes, tumor suppressor genes, and stability (also called caretaker) genes. Proto-oncogenes encode proteins that in general participate in the regulation of cell proliferation and differentiation. Dominant gain- of- function mutation or hypomethylation converts a proto-oncogene into an oncogene and may cause uncontrolled cell growth. Tumor suppressor genes normally regulate a variety of processes that maintain normal cell growth and homeostasis, *e.g.* arrest cell cycle or stimulate cell death if DNA is damaged or chromosomes are abnormal, and their silencing is a key event in tumor formation [2]. Inactivation of both parental alleles is generally required, and usually results in loss- of- function of the encoded protein [5]. The third class of genes, the stability genes encodes proteins that stabilize the genome by repairing errors in the DNA (Mismatch repair genes; MMR), and they are thus important for DNA fidelity. When their function is lost, errors will accumulate in the genome. Thirty percent of our genes encode MMR genes [6], emphasizing how important these genes are for the genome. Silencing of the stability genes and the tumor suppressor genes can occur by mutations or promoter hypermethylation, and over the past years the latter has proven to be a major contributor in the origin and progression of many cancers [7].

1.2 Epigenetics

"The major problem, I think, is chromatin… you can inherit something beyond the DNA sequence. That's where the real excitement of genetics is now"

-Watson-

The term epigentics was first used by Conrad Waddington to describe "the causal interactions between genes and their products, which bring the phenotype into being" [8]. Today epigenetics can be defined as traits heritable through meiosis or mitosis that are not dependent on the primary DNA sequence [9], and they include DNA methylation, histone modifications, and nucleosome positioning [10]. All of these epigenetic processes have an impact on chromatin organization and maintenance, and together they regulate what genetic information is accessible for the transcriptional machinery. In addition to regulating gene- and microRNA expression, epigenetic mechanisms are important for cellular processes such as suppressing the mobility of transposable elements, and for proper development and differentiation [11].

The chromatin signatures change during aging, and are altered by environmental factors, such as diet, stress, and physical exercise [9,12]. When the normal epigenetic pattern is disrupted it contributes to the development of human diseases, including neurodevelopmental (Rett- and ATRX syndrome), neurological (Fragile X syndrome, Prader-Willi and the Angelman syndromes, Parkinson's and Huntington's disease)-, and autoimmune (immunodeficiency, centromeric instability and facial anomalies; ICF see page 5, DNA methylation in cancer) disorders, as well as cancer [10,13]. DNA methylation (see below) is the main focus of this thesis, but since there is an intimate cross talk between the different types of epigenetic information [11], the other main categories will also be mentioned.

1.2.1 DNA methylation

In humans, DNA methylation occurs predominantly at the cytosine base of DNA, within CpG dinucleotides. Because of spontaneous deamination, CpG sites have been lost from the mammalian genomes during evolution [14]. Most of the CpG sites which have escaped depletion lie within repetitive sequences (for example centromeric- and transposable regions). These sites are generally methylated, which provide long- term transcriptional silencing important for maintaining genomic stability [7]. In addition, several CpG sites tend to cluster in regions called CpG Islands[1]. These Islands can be found in the 5´region of about half of all human genes, and are normally unmethylated regardless of the expression status of the associated gene [14-16]. However, some CpG island promoters become methylated during development, including one of the two X- chromosomes in females and imprinted genes, in addition to the promoter of tissue specific genes [7]. In general, DNA methylation of gene promoters is associated with loss of gene expression.

The DNA methyltransferases

DNA methylation is established and maintained by three catalytic enzymes of the DNA methyltransferase (DNMT) family; DNMT1, DNMT3A and DNMT3B which catalyze the transfer of a methyl group from S- adenosyl methionine to DNA [10]. DNMT3A/3B are *de novo* methyltransferases, thought to be responsible for establishing methylation marks in embryonic development [17], while DNMT1, showing a high preference for hemi-methylated DNA, is thought to maintain the methylation pattern in the daughter cells after DNA

[1] A region of ≥ 500 bp with a GC content of at least 55% and with a ratio of observed to expected CpG frequency of ≥0.65 (http://cpgislands.usc.edu/).

replication. However, the distinction of roles is not that clear. In addition to the 30-40 fold preference for hemi-methylated DNA, DNMT1 also possesses *de novo* activity [10]. It has further been suggested that DNMT3A/3B may methylate sites missed by DNMT1 at the replication fork [17]. Consistent with the important role of DNA methylation in critical cellular processes, deletion or partial inhibition of DNMTs results in severe cellular and developmental phenotypes and predisposition to cancer [11]. Mice deficient in DNMT3B or DNMT1 are embryonic lethal, while mice deprived of DNMT3A die within 4 weeks of birth [18].

DNA demethylation

Due to lack of observable DNA demethylases, and the involvement of 5- methylcytosine (5-meC) in long- term silencing (*i.e*. imprinting and X- chromosome inactivation), DNA methylation has been considered as a relatively stable epigenetic mechanism. However, research the past decade has revealed that the 5-meC marks are more dynamic that originally proposed, and they can in addition to being passively erased, be removed through active mechanisms. It has been known for a longer time that the paternal genome is subject to active genome- wide demethylation at two points in the reproductive development; the first time happens shortly after fertilization in the male pronucleus, while the second time occurs in the primordial germ cells (PGCs) of 11.5- 12.5 day old embryos[19]. Several methods for achieving active demethylation have been suggested, but no conclusive mechanism has yet been established [18]. Some years ago TET1 (ten eleven translocation) was shown to catalyze the hydroxylation of 5-meC to 5- hydroxymethylcytosine (5-hmeC), and it was speculated that 5- hmeC could be an intermediate component in a pathway of active demethylation [20]. Support for this hypothesis was recently demonstrated by Yu- Fei *et al.* showing that TET dioxygenases catalyze the oxidation of 5-meC and 5-hmeC to 5-caboxylcytosine (5-caC) *in vitro* and in cultured cells [21]. 5-caC can further be removed by thymidine- DNA glycosylase (TDG) - mediated base excision repair (BER), and an unmethylated cytosine could be inserted into the repaired genomic region, leading to demethylation [21]. At the same time as this article was published another group independently demonstrated that the TET proteins can generate 5-caC from 5- meC and 5-hmeC, and they mentioned the possibility of a cooperation between TET enzymes and BER in demethylating 5-meC [22].

Gene silencing by DNA methylation

DNA methylation can inhibit gene transcription by two distinct mechanisms. It has been proposed that methylated CpGs induce transcriptional repression by directly interfering with the binding of transcription factors to DNA. A number of transcription factors that cannot access their binding sites due to methylation have been identified, including AP-2, CREB, E2F, NF-κB and c- Myc [6]. However, more often, DNA methylation represses transcription by interacting with methyl CpG binding domain (MBD) proteins. The MBD proteins bind DNA and can block access to transcription factors directly, or they can recruit chromatin modifying proteins, resulting in stable transcriptional repression by formation of heterochromatin [23]. For more details see page 10, Interaction between epigenetic players to silence gene expression.

DNA methylation in cancer

DNA methylation is an essential process for maintaining homeostatic equilibrium in normal cells, which is reflected by the observation that severe diseases like Rett syndrome (a neurlological X- linked disease caused by point mutations in MeCP2) and immunodeficiency-centromeric instability and facial anomalies (ICF; see next session), develop if the normal DNA methylation pattern is disrupted. Changes in the normal pattern are believed to occur at an early stage of tumor development and accumulate throughout cancer progression [24]. Global DNA hypomethylation together with promoter specific hypermethylation are both common alterations in human cancers (Figure 2) [16,25].

Genome wide hypomethylation

Global DNA hypomethylation, defined as loss of methylation, was the first epigenetic abnormality to be identified in cancer cells [26]. The CpG sites within repetitive sequences are frequently targeted for DNA hypomethylation [27]. Methylation at these sites is important for genome integrity and disruption of the methylation pattern has been proposed to explain the increased genomic instability observed in some cancer cells [25]. One example of genomic instability is seen in patients with the immunodeficiency- centromeric instability syndrome (ICF), which in most cases is caused by an inactivating germline mutation in *DNMT3B* [28,29]. As a result, selected centromeric regions become markedly undercondensed causing centromeric instability and chromosomal rearrangements [30].

Although ICF patients are not at increased risk of developing cancer, mutations in *DNMTs* have also been reported in tumors and cancer cells, causing similar loss of DNA methylation and chromosomal instability [25,31,32].

Unique sequences including proto- oncogenes, tissue specific genes and imprinted loci are also subject to hypomethylation [27], causing their inappropriate activation. *HRAS* was the first oncogene shown to be hypomethylated, in tumors from colon and lung [33]. The tissue specific cancer/ testis antigens, which are normally methylated and silenced in normal somatic tissue are subjected to hypomethylation and subsequent reactivation in several tumor types [34]. DNA methylation normally plays a key role in genomic imprinting, where hypermethylation at one of the two parental alleles leads to monoallelic expression [10]. When it goes awry it can lead to diseases such as Beckwith- Wiederman syndrome, a pre-neoplastic disorder that is characterized by a 1000- fold higher change of Wilms tumor, and which is largely attributed to the loss of imprinting (LOI) of the *IGF2* gene [11]. LOI of *IGF2* has also been implicated in the development of other cancers, including colorectal cancer [35].

Gene specific hypermethylation

Abnormal gain of DNA methylation of promoter CpG islands is a frequent event in cancer and is associated with repressed or absent gene expression [13]. Hundreds to 1000s of genes per tumor may harbor aberrant promoter methylation, and the number of tumor suppressor genes epigenetically inactivated in a cancer cell has been shown to equal or exceed the number that is inactivated by mutations [15,36]. Most of the *de novo* methylation is likely to merely accompany carcinogenesis, *i.e.* being generated as a result of transformation [36]. However, some of the methylated genes (drivers), are presumed to have a functional role in the cancer cells. For instance, several germ-line mutated genes which cause hereditary cancers have been shown to be epigenetically silenced in the sporadic forms and are thus highly likely to promote neoplastic transformation [15]. Some examples include the tumor suppressor genes *RB*, and *VHL*, and the DNA repair genes *BRCA1,* and *MLH1* causing sporadic forms of retinoblastoma, renal cell carcinoma (RCC), breast cancer, and colorectal cancer, respectively, when hypermethylated [37,38].

In addition, gene silencing by aberrant promoter hypermethylation may affect all the molecular pathways important for transformation, and the abnormal silencing frequently

occur early in the process [11,39,40]. For example, *CDKN2A* (*p16*^{INK4A}; *p16*) which encodes a cyclin- dependent kinase (hence cell cycle) inhibitor, is found epigenetically silenced in pre-invasive stages of *e.g.* breast-, colon- and lung cancer [13]. *HIC1* (hypermethylated in cancer 1), which encodes a transcriptional repressor targeting genes involved in proliferation, tumor growth and angiogenesis, is also found hypermethylated in precancerous lesions of *i.e.* colon and breast [41]. *HIC1* silencing has been linked to upregulation of the stress sensing protein SIRT1 (a histone deacetylase), which further inhibits the transcriptional activity of TP53, as well as upregulates the Wnt pathway by silencing the Secreted frizzled-related proteins (SFRPs) [13].

The APC/ β- catenin pathway can also be disrupted by epigenetic silencing. Adenomatous polyposis coli (*APC*) is frequently disrupted by somatic mutations in sporadic colorectal cancers (approximately 80%), and hypermethylation is observed at a low frequency. However, other tumors arising in the gasterointestinal tract (GI), the liver, pancreas and stomach, frequently display *APC* promoter hypermethylation [40].

Promoter hypermethylation is clearly an important mechanism for gene silencing in tumor development. This is further supported by observations that demethylating drugs such as 5-aza-2-deoxycytidine can reactivate the affected genes and restore production of the corresponding protein [28].

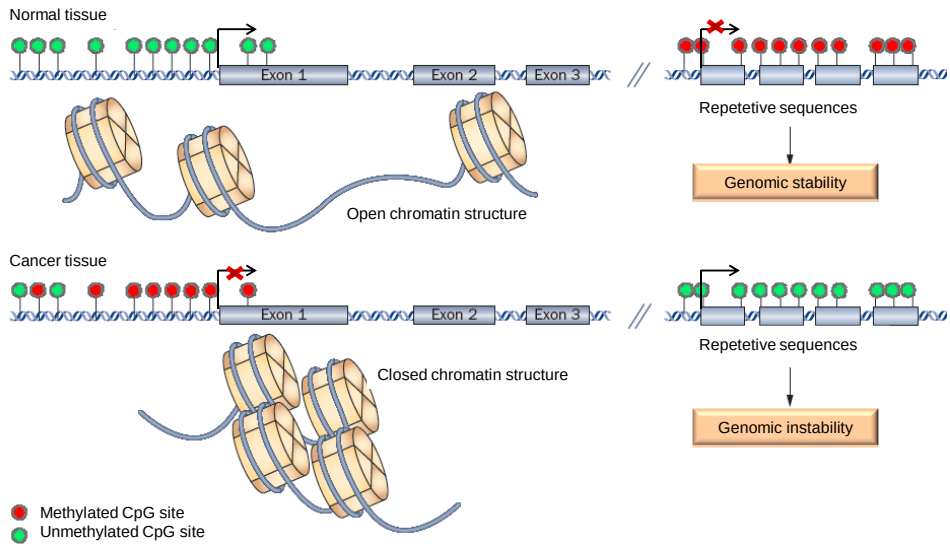


Figure 2. DNA methylation in normal versus cancerous tissue. Normal mammalian cells generally have unmethylated CpG island gene promoters, which is associated with an open chromatin structure and transcription, while the repetitive sequences are highly methylated and silenced. In cancer cells, two changes are observed: gene specific promoter hypermethylation and global hypomethylation, causing silencing of tumor suppressor genes and genomic instability, respectively. Modified after [42].

Age related methylation, field defect, and increased chance of neoplastic transformation

The variability in DNA methylation increases during aging, and include both global DNA hypomethylation and gain of methylation at specific genes. Interestingly, approximately half of the genes which display age- related methylation, as defined as type A- methylation by Toyota and Issa [16], are the same genes involved in the pathogenesis of colorectal cancer [43]. Age related methylation could be responsible for generating an epigenetic "field defect"; an area of abnormally methylated tissue that otherwise look normal, and that precedes and predisposes to the development of cancer. The DNA repair gene *MGMT* is frequently methylated in *i.e* colorectal cancer, and also in the normal mucosa of cancer patients. Shen *et al* [44] showed that patients with colorectal cancer and *MGMT* promoter methylation also had substantial *MGMT* promoter methylation in adjacent apparently normal mucosa. Fifty percent (22/44) displayed methylation in adjacent normal mucosa, compared to 6% (3/51) of those without *MGMT* methylation in the tumor. It has therefore been suggested that methylated *MGMT* may cause a "field defect", where the conditions are favorable for further alteration. The protein encoded by *MGMT* removes alkyl groups from the O^6 position of guanine, and its silencing increases the mutation frequency and thus the risk of cancer development.

1.2.2 Chromatin

Within the cell's nuclei, DNA is wrapped around a histone octamer consisting of 2* (H2A, H2B, H3, H4), forming the nucleosome which is the subunit of chromatin (Figure 3) [45]. The histone tails are subjected to post- transcriptional modifications, including *i.e.* acetylation, methylation, phosphorylation, and ubiquitylation. The sum of these modifications (site and type) constitutes a "histone code", influencing *i.e.* gene expression by affecting whether chromatin is organized into domains of open, transcriptional active heterochromatin or densely packed, transcriptional silenced euchromatin Generally, acetylation, which is catalyzed by histone acetyltransferases (HATs), is associated with active transcription while the impact of methylation, catalyzed by the histone methyltransferases (HMTs), depends on

which residues are modified, and whether the residue is mono-, di,- or tri-methylated. H3K27me3 and H3K9me3 are examples frequently found at transcriptionally repressed promoters, whereas H3K4me3 is enriched at transcriptionally active gene promoters [10].

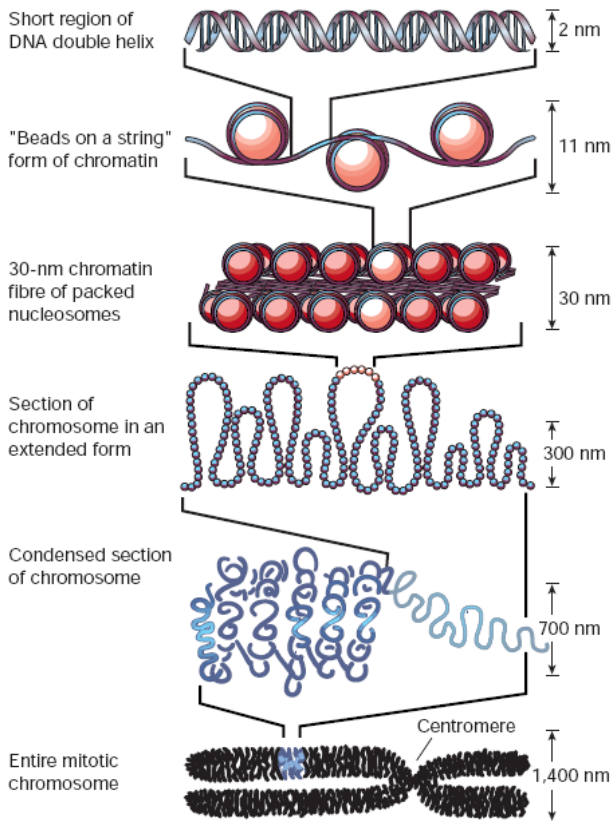


Figure 3 Packing of DNA. DNA is wrapped around a core of eight histone molecules to form the nucleosome, which is the subunit of chromatin. The histone tails can be post- translationally modified, which affect the transcription status of a gene by allowing or preventing transcription factors from binding DNA. The nucleosomes can be further packed, eventually giving rise to the visible chromosome. Figure from [45].

Chromatin packing can additionally be modified by ATP- dependent chromatin remodeling complexes as well as by the incorporation of histone variants [46]. The presence of nucleosomes hinders transcription factors and activators from binding to their respective DNA recognition sequences, which means that the polymerase cannot be engaged. Thus, for transcription to initiate, nucleosomes around the transcription start site of the gene in question need to be displaced, restructured or destabilized, which is done by different ATP-dependent chromatin remodeling complexes (such as SWI/SNF, ISWI, CDH, INO80) [10].
Incorporation of different histone variants can also regulate nucleosome positioning and gene expression [10]. One example is the histone variant H2A.Z which is an unfavorable substrate for binding of silencing proteins, thus protecting cytosines from methylation [47].

1.2.3 Interaction between epigenetic players to silence gene expression

There is a tight collaboration between DNA methylation, histone modification and nucleosomal remodeling in regulating gene activity.

As mentioned previously, DNA methylation can mediate transcriptional silencing by recruiting methyl binding proteins (MBP), which can further attract histone deacetylases (HDACs) and chromatin remodeling complexes. For instance, the methyl- CpG binding protein 2 (MECP2) can associate with the transcriptional co- repressor SIN3A, and HDACs, which subsequently induce deacetylation and gene silencing [15]. It has further been demonstrated that Brahma, a catalytic component of the SWI/ SNF chromatin remodeling complex, can also interact with MeCP2 [43], providing a link between DNA methylation, histone deacetylation and chromatin remodeling.

Histone modifications may also direct DNA methylation. The polycomb repressive complexes PRC1 and PRC2 are involved in the initiation and maintenance of transcriptional silencing, respectively, by forming and recognizing H3K27me3 [27]. EZH2, a key component of the PcG complexes, has been linked to the recruitment of DNA methyltransferases (DNMTs), suggesting a possible mechanisms where PRC mediated transcriptional silencing predispose genes to promoter CpG island methylation [48].

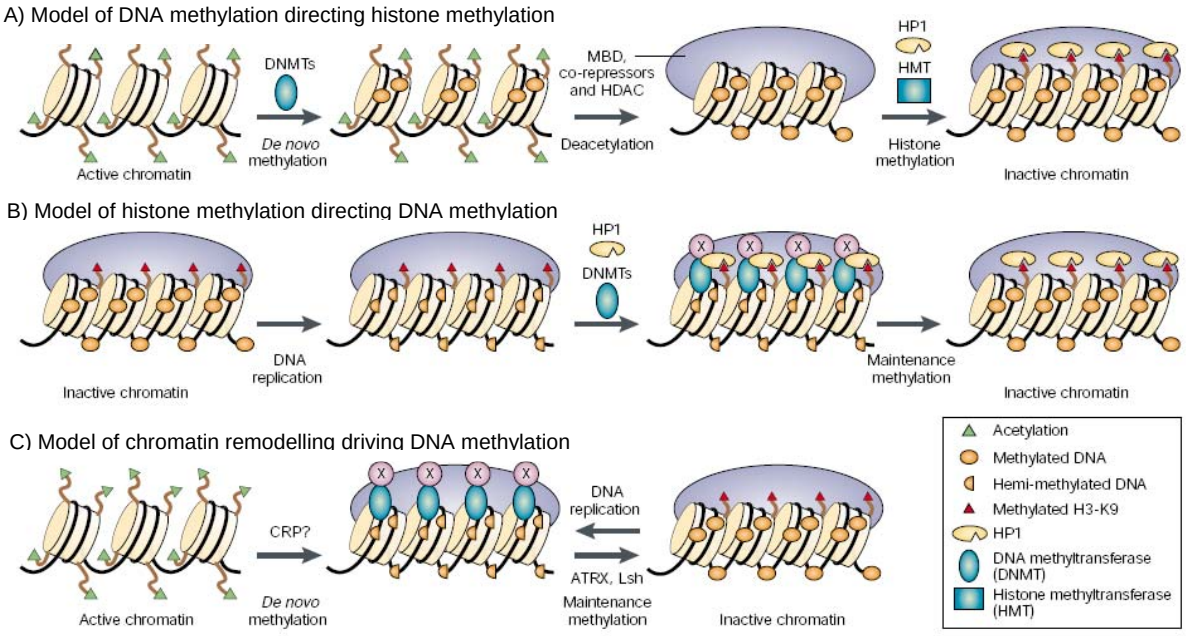


Figure 4. Links between DNA methylation, histone modifications and chromatin remodeling. Three possible models for the interaction between DNA methylation, histone modification and chromatin remodeling to silence gene expression have been suggested. A) De novo methylated DNA directs histone methylation via methyl CpG-binding proteins (MBDs). In addition to recognize and bind methylated cytosines, the MBDs interact with histone deacetylases, promoting deacetylation and further histone methylation, which stabilize the inactive chromatin state. B) After DNA replication, inactive chromatin is maintained by histone methylation (especially H3K9me3), which is recognized by heterochromatin protein 1 (HP1). HP1 recruits DNA methylatranseferases, causing DNA methylation and stabilizing of the inactive chromatin. C) The ATP-dependent chromatin-remodelling and DNA-helicase activities of proteins, such as ATRX and Lsh, might facilitate DNA methylation and histone modification by unwinding nucleosomal DNA to increase its accessibility to DNMTss, HDACs and HMTs.

Figure from [49].

In the cooperation between DNA methylation and histone modification in maintaining the aberrant silencing of hypermethylated genes in cancer, DNA methylation appears to be dominant [15,28]. Treating cancer cell lines with the histone deacetylases inhibitor Tricostatin A (TSA) is not alone sufficient to reactivate aberrantly silenced hypermethylated genes [50]. However, if the cells are first treated with demethylating drugs, such as 5-aza-2`deoxycytidine, a synergic effect is seen when TSA is additionally administered.

1.2.4 Interplay between epigenetics and genetics

"(…) it had become clear that it is impossible to understand the genetics of cancer without epigenetics."

-Issa-

Throughout the past years, it has become clear that there is a close interaction between genotype and epigenotype in cancer- associated genes. Epigenetic changes can for example precede and induce genetic mutations. The transcriptional silencing of the DNA repair gene O^6-methylguanine-DNA methyltransferase (*MGMT*) by promoter hypermethylation is observed in a wide variety of tumors, including lung and brain tumors in addition to colorectal cancers, and is associated with G:C to A:T transition mutations in *KRAS* and *TP53* [51]. *MLH1* and *BRCA1* are two other abnormally silenced DNA repair genes leading to microsatellite instability (see page 17, Instabilities in colorectal cancer) and failure to repair DNA lesions. The protein product of the tumor suppressor gene *CDKN2A*/*p16* disrupts progression through the cell cycle in times of cellular stress, and provides time to repair DNA. When *CKDN2A/p16* is silenced by promoter hypermethylation, cells will continue to divide despite of the damaged DNA, and mutations, chromosomal aberrations etc. will accumulate in the genome. Moreover, CpG methylation can directly cause genetic changes in cancer since it

is mutagenic on its own; 5- meC can undergo spontaneous hydrolytic deamination to cause C:T transitions [11].

Abnormal epigenetic events may also be caused by genetic abnormalities. Recently, a high frequency of cancer- specific mutations in genes which cause abnormal epigenome organization has been revealed in several tumor types [52]. Jones *et al* report a 57% mutation frequency in the *ARID1A* gene, which encodes a protein that participates in chromatin remodeling, in ovarian clear cell carcinoma (OCCC) [53]. By using exome sequencing Jiao and colleagues observed that the most frequently mutated genes underlying pancreatic neuroendocrine tumors (PanNETs) includes proteins implicated in chromatin remodeling [54]. Mutations in the isocitrate dehydrogenase 1 (*IDH1)* and *IDH2* are frequently found in gliomas and leukemia, and also sporadically in colorectal cancer. Mutations in these genes not only reduce the enzymatic capacity, they also change the proteins' capabilities. Mutated *IDH1* and *IDH2* use α- ketoglutarate to generate 2- hydroxyglutarate, which may alter the DNA methylation level by inhibiting the α- ketoglutarate dependent enzyme TET2 (see previous) [52]. In addition to providing a link between genetics and epigenetics in cancer development, these studies strongly imply epigenetics as a major player in tumorgenesis.

Interestingly, *cis*- acting mutations or single nucleotide polymorphisms (SNPs) of certain genes have been suggested to confer cancer susceptibility by increasing the probability of *de novo* methylation, and this can be transgenerationally inherited. In a study of a cancer affected family, Hitchins *et al* reported a dominant inheritance of a mosaic *MLH1* epimutation[2] through three generations [56]. A C-T substitution close to the transcription initiation site of the *MLH1*gene was identified, and the epimutation was erased in spermatozoa and reinstated in the somatic cells of the next generation. These findings provide evidence that individuals with a specific haplotype are genetically predisposed to promoter hypermetylation and subsequent gene silencing.

Yet another interesting connection between epigenetics and genetics was recently reported by Chang *et al* [57]. A point mutation in *BRCA1* did not impair DNA- double strand break repair, but instead caused epigenetic upregulation of miRNA-155, through the inability to recruit the histone deacetylase HDAC2. miRNA-155 has earlier been found upregulated in cancer and is implicated to have oncogenetic activity.

[2] Epigenetic silencing of a gene that is normally active, or epigenetic activation of a gene that is normally silenced [55].

"Epigenetics is a useful word if you don't know what's going on- if you do, you use something else".

-Adrian Bird-

1.3 Colorectal cancer

Approximately 1.2 million new colorectal cancer cases occurred world-wide in 2008, and with more than 600 000 annual deaths colorectal cancer is the third most common cause of cancer deaths word- wide [58]. In Norway, ~3600 people were diagnosed with colorectal cancer in 2009, which makes this the second most common cancer type in both men (after prostate cancer) and women (after breast cancer)[3]. Colorectal cancer primarily affects older individuals (median age 70 years), and the incidence is substantially higher in industrialized than in developing countries [58]. An unhealthy diet; rich in fat and red meat, and low in fiber, fruit and vegetables, in addition to obesity and physical inactivity have been associated with an increased risk of colorectal cancer [59], emphasizing the important role of life- style and environmental factors when it comes to developing this malignancy.

The vast majority of colorectal cancer cases occur sporadically, but approximately 5% have known hereditary syndromes, such as Lynch syndrome (also called hereditary nonpolyposis colorectal cancer, HNPCC) and familiar adenomatous polyposis (FAP). Patients with hereditary syndromes have an increased risk of developing colorectal cancer during their lifetime due to germline mutations in colorectal cancer critical genes, and the disease usually occurs at an early age compared with the sporadic cases. Despite the low incidence of hereditary syndromes, they are of substantial clinical and research importance and have provided valuable insight into also the sporadic forms of colorectal cancer [60].

FAP is an autosomal, dominantly inherited disease characterized by the presence of hundreds to thousands of colonic adenomas. Although only a minority of these benign lesions progress to malignancy, the high number of adenomas ensures a nearly 100% lifetime risk of developing colorectal cancer. If left untreated, the average age of colorectal cancer diagnosis for these patients is 39 years [61]. The disease is caused by a faulty gatekeeper, *APC*, which is normally involved in degrading β- catenin in the Wnt signaling pathway. Soon after it was

[3] http://www.kreftregisteret.no

discovered in FAP patients, *APC* was found to be somatically mutated also in the great majority of sporadic colorectal tumors [60]

Individuals with Lynch syndrome carry a germ line mutation in one of the DNA mismatch repair (MMR) genes, including *MSH2*, *MLH1*, and infrequently *MSH6* and *PMS2* [61]. The MMR system corrects single- base mismatches that form under DNA replication, and their inactivation give rise to tumors with microsatellite instability (MSI; for more details about MMR and MSI see page 17, Instabilities in colorectal cancer) [37,61]. Lynch patients are predisposed to various types of cancers, including a 50%- 80% life time risk for colorectal cancer.

Constitutional epimutaion[4] of *MLH1*and *MSH2* have been reported in patients with suspected Lynch syndrome, in whom no germ-line mutation of either MMS genes have been found [55,62,63]. These epimutations may arise very early in the germline or they may be inherited, but the heritability is expected to be low as they in most cases are erased by passage through the germline. It has therefore been suggested that the epimutations may be a direct consequence of a *cis*- acting mutation, as was reported by Hitchins et al (see page 11, Interplay between epigenetics and genetics). An underlying sequence variant is also most likely the cause of the *MSH2* epimutation observed by Chan *et al*, where all individuals carrying the at- risk haplotype also had the epimuation [62]. However, non- mendelian pattern have also been observed for *MLH1* [55,63]. Whether germ- line epimutations occur in humans are still controversial. See future perspectives.

1.3.1 Tumor classification and prognosis

Two staging systems are used to define the extent of invasion of colorectal cancer: The Dukes's classification and the AJCC/ UICC TNM (Tumor, Node, Metastasis) staging. Both systems are divided into four categories, Dukes: A-D, and AJCC/UICC: I- IV, with an additional O stage or carcinoma *in situ*, which are carcinomas with high grade dysplasia that have not begun to invade the colon wall. The two staging systems are highly comparable. Stage A/I tumors are confined to the intestinal mucosa and submucosa, whereas stage B/II tumors have invaded these layers and penetrated into the muscle layer. Stage C/III tumors

[4] an epimutation that is found in all tissue of the body and which may be mosaic. Evidence of transmission from the previous generation may not exist, and the epimutation may instead have occurred early in development [37].

have spread to one or several lymph nodes, and stage D/IV tumors have metastasized to distant organs, *i.e.* liver and lungs (Figure 5).

In Norway, the relative five- year survival rates for men and women are 57.6% and 61.6% respectively[5], but survival is also associated with stage at time of diagnosis. Localized disease normally has a good prognosis with a close to 85% five year relative survival in both sexes (81.6% in men and 87.8 % in women), whereas patients with distant metastasis generally have a poor prognosis with a five year relative survival close to 10 % (8.9 % in men and 9.2 % in women) (Figure 5).

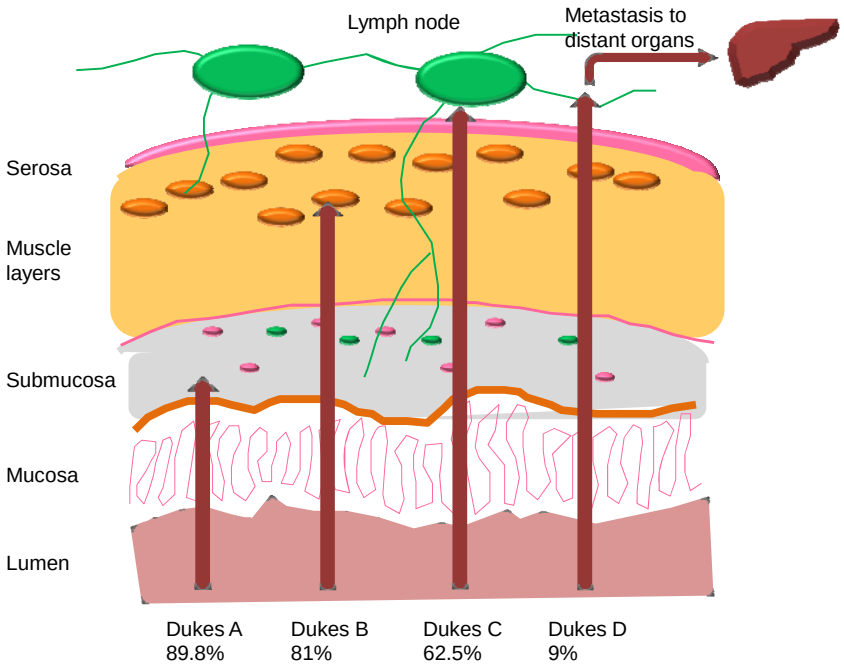


Figure 5. Tumor staging and prognosis. Survival is associated with tumor stage at time of diagnosis. Patients diagnosed with a localized disease (Dukes A and B) have a 5- year relative survival close to 85% whereas patients diagnosed with regional (Dukes C, spread to lymph nodes) and distal (Dukes D, spread to distant organs) disease have a 5- year relative survivals of approximately 60% and 10 %, respectively.

1.3.2 Histological and molecular developmental pathways

Two stepwise progression models involving histological changes and concurrent molecular changes have been proposed to explain the etiology of colon cancer from benign neoplasms to carcinoma: 1) the adenoma- carcinoma sequence and 2) the serrated pathway [6,60,64].

[5] http://www.kreftregisteret.no

The adenoma- carcinoma pathway

Most colorectal cancers develop through the adenoma to carcinoma progression sequence (Figure 6), explaining how the normal colon epithelium is transformed to cancer via distinct histological changes [65]. The dysplastic aberrant crypt foci are the earliest histological visible lesions [60], and they gradually progress through increases in size, dysplasia and the acquisition of villous (fingerlike) morphology [64]. The different steps are associated with specific genetic alterations. Inactivating mutations in the *APC* gene are frequently seen in the aberrant crypt foci and are therefore considered to be an initiating event [6,66]. As many as 80% of colorectal tumors are reported to carry a mutant *APC* gene [60], further emphasizing the important role of this tumor suppressor gene in tumor formation. Activating mutations in the *KRAS* proto-oncogene have been found in 37%- 41% of all colorectal cancers and are thought to further dictate the malignant formation by mediating adenoma growth [66]. Two distinct genetic pathways have been described to explain the changes in genome integrity during tumor development; the chromosomal instability (CIN) - and the microsatellite instability (MSI) pathways [60] see page 17, Instabilities in colorectal cancer.

The serrated pathway

The serrated pathway acknowledges that not only adenomas, but also serrated polyps can develop into colorectal cancer through a histological progression sequence with distinct genetic and epigenetic changes [67]. Because hyperplastic polyps lack cytological dysplasia, early studies concluded that they were harmless lesions without potential for malignant progression [68]. It has now become evident that serrated polyps in reality represent a heterogeneous group of lesions, comprising sessile serrated adenomas (SSA), traditional serrated adenomas (TSA) and mixed polyps in addition to the hyperplasic polyps, some of which have malignant potential [69]. In contrast to the classical adenoma- carcinoma sequence, which explains all cancers as resulting from the dysplastic aberrant crypt foci and adenomas, the serrated pathway hypothesizes that the adenomas only give rise to CIN tumors and that the sessile serrated polyps are the precursors to MSI and CpG island methylator phenotype (CIMP) tumors [68]. (For further explanation about CIMP, see below). Tumors with CIN and CIMP have been found to rarely overlap [70], further emphasizing two different mechanisms for generating molecular diversity. *BRAF* mutations and CIMP have a central role in the molecular pathogenesis of tumors evolving through the serrated pathway [67]. The vast majority of SSAs (78%- 90%) have *BRAF* mutations, which is hypothesized to be the

initiating event, and several genes are silenced by promoter hypermethylation, including *MLH1* leading to MSI [67,71]. The realization that serrated polyps can precede to cancer has important clinical implications, since such polyps initially were overlooked during colonoscopy. Several studies have found that large (≥10 mm) serrated polyps in the proximal or distal colon are strongly associated with synchronous advanced neoplasia [68].

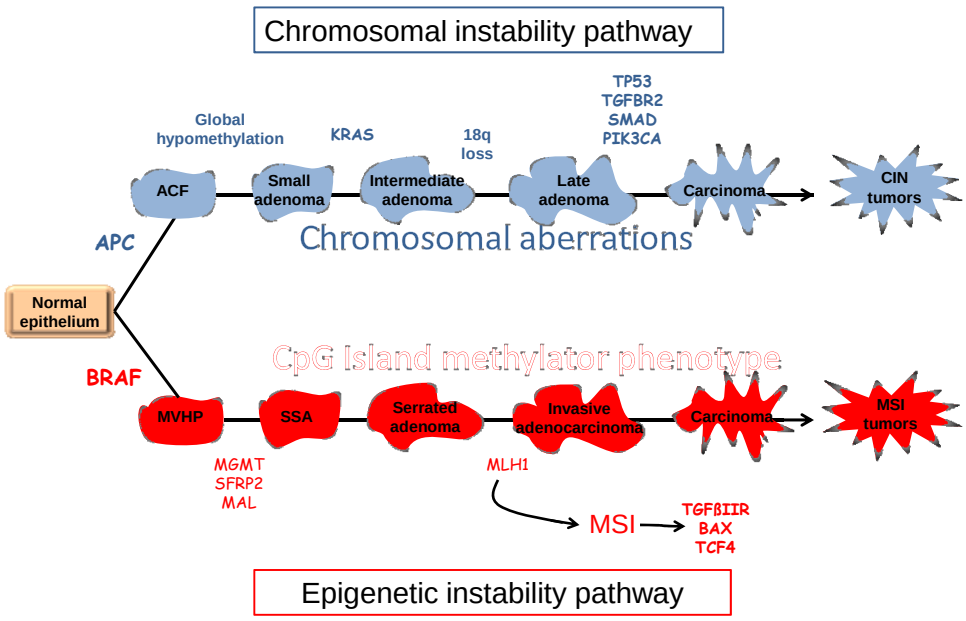


Figure 6. Histological and developmental pathways to colorectal cancer. *BRAF* mutations are hypothesized to be the initiating event in the epigenetic instability pathway, giving rise to CIMP and MSI tumors, whereas APC mutations are thought to initiate the chromosomal instability pathway. This last group of tumors displays a high degree of chromosomal aberrations whereas the MSI and CIMP tumors are generally stable on the chromosome level. Abbreviations; ACF, aberrant crypt foci; CIN, chromosomal instability; MSI, microsatellite instability; MVHP, microvesicular hyperplastic polyp; SSA, sessile serrated adenoma.

1.3.3 Instabilities in colorectal cancer

In tumor cell precursors, an increased level of instability provides the necessary genetic and epigenetic diversity for the Darwinian selection that characterizes tumor formation and progression [70]. For colorectal cancer, three molecular pathways leading to different types of instabilities have been described; CIN (chromosomal instability), MSI (microsatellite instability) and CIMP (epigenetic instability).

Chromosomal instability (CIN)

CIN is the most common genetic pathway to colorectal cancer, accounting for approximately 85% of the cases [72]. CIN cancers are characterized by substantial gains and losses of chromosomes, as well as increased loss of heterozygosity [27]. At the cytogenetic level these changes are reflected by aneuploidy (abnormal chromosome number) [60]. The causal mechanism(s) are still undefined, but the losses have been proposed to result from mitotic recombination or aberrant mitotic segregations of chromosomes [60]. Defects in chromosome cohesion, mitotic checkpoint function, centrosome copy number, kinetochore-microtubule attachment dynamics, and cell-cycle regulation have been suggested as underlying mechanisms causing CIN [73]. In accordance with the adenoma- carcinoma model, biallelic loss of the *APC* gene is considered to initiate the neoplastic process, and the increased chromosomal instability causes accumulation of mutations in ***i.e.*** *KRAS*, *TP53*, and *SMADs,* which contribute(s) to cancer progression [60].

CIN tumors are associated with distal (left) location in the colon, and the prognosis for patients diagnosed with CIN cancers are worse compared to those with MSI tumors [74].

Microsatellite instability (MSI)

Microsatellite instability is, as mentioned previously, caused by a defect in the DNA mismatch repair (MMR) system, including *MLH1*, *MSH2*, *MSH6,* and *PMS2*, resulting in increased accumulation of mutations (two to three orders of magnitude higher than in normal cells [60]. Replication slippage frequently occurs within microsatellites, repetitive units containing 1-6 nucleotides (mono-, di- tri- etc- repeats) and which are scattered around the genome. With a defect MMR-system, the insertions and deletions within microsatellites will not be corrected, resulting in abnormally long or short microsatellites which are characteristic for microsatellite unstable tumors. Rarely but significantly, microsatellites are present within the coding region of growth regulatory genes genes, including *TGFBR2*, *MSH3* and *IGF2R*, generating frameshift mutations when the MMR system is defect [2]. MSI has therefore also been referred to as the mutator phenotype.

Ten to 15% of sporadic colon cancers display microsatellite instability [51]. In addition, >90% of patients with Lynch syndrome are characterized by MSI [75]. In the majority of sporadic cases, *MLH1* is silenced by hypermethylation of the promoter region, while a

germline mutation in this or other MMR genes is responsible for the MSI phenotype in patients with Lynch syndrome. The sporadic MSI tumors are mainly located in the proximal (right) side of the colon, are associated with female gender, older age, diploid or near diploid karyotype and *BRAF* mutation [71].

CpG island methylator phenotype (CIMP)

CIMP refers to a subset of colorectal cancers and adenomas that display widespread promoter hypermethylation resulting in epigenetic inactivation of the involved genes, including tumor suppressors [16,76,77].

Hypermethylation of *MLH1* is a frequent event in the CIMP pathway, leading to microsatellite unstable tumors in a manner similar to the MSI pathway [72]. There is a strong association between CIMP and MSI. However, although most sporadic MSI associated tumors have CIMP, it is not limited to this tumor type; approximately half of all tumors with CIMP do not have methylation of *MLH1* or MSI [67]. In addition, MSI positive cases tend to have a good prognosis, while CIMP+MSI- cases have a particularly poor outcome [14]. According to the serrated pathway, CIMP actually precede the "mutator phenotype" in MSI tumor progression [2,16], with hypermethylation of *MLH1* as the underlying cause, leading to MSI. CIMP colorectal cancer is additionally characterized by distinct molecular and clinopathological features including *BRAF* and *KRAS* mutations as well as proximal location and distinct precursor lesions (serrated adenomas) [14,76].

1.4 Cholangiocarcinoma and primary sclerosing cholangitis (PSC)

Cholangiocarcinomas are malignant tumors arising from the epithelial cells (cholangiocytes) of the intra- and extrahepatic biliary tract [78] (Figure 7). It is a rare malignancy affecting 1-2 per 100 000 individuals, but the incidence is rising. In Norway 147 people (73 men and 74 women) were diagnosed with cancer in the gallbladder or the bile ducts in 2009[6].

The only curable treatment for cholangiocarcinoma is surgical resection of early stage tumors, or liver transplantation for highly selected patients [79]. Unfortunately, due to late clinical presentation, most cholangiocarcinomas are detected at an unresectable stage and are hence

[6] www.kreftregisteret.no

associated with a poor prognosis [80]. The overall survival, including resected patients is low; less than 5% are alive after 5 years [81], and 75% of patients die within one year of the diagnosis [82]. The main causes of death include cancer cachexia (a complex metabolic syndrome characterized by loss of muscle mass, including massive depletion of skeletal muscle, with or without loss of fat mass), liver failure, and recurrent sepsis due to biliary obstruction [82]. The five-year relative survival rate in Norway among persons diagnosed with cancer in the gallbladder and bile ducts were 13.6% for men and 15.5% for women in 2009[7]. The etiology of cholangiocarcinoma is usually unknown, but it is believed that chronic bile duct inflammation may lead to neoplsia of the biliary tree. Approximately 10%-30% of cholangiocarcinomas arise in patients with primary sclerosing cholangitis (PSC; see below) [83].

1.4.1 Classification of cholangiocarcinoma

Cholangiocarcinoma can be anatomically classified into (distal) extrahepatic- and intrahepatic forms of the disease, in addition to hilar- or Klatskin tumors, which are located in the perihilar ducts (Figure 7). This last group of tumors, accounting for up to 50% of all cholangiocarcinomas, includes extrahepatic tumors that arise in the perihilar region, but which may extend into the liver and thus be classified as intrahepatic lesions. The majority of cholangiocarcinomas, 80%- 90% are extrahepatic or hilar, whereas approximately 10% originate within the liver [84].

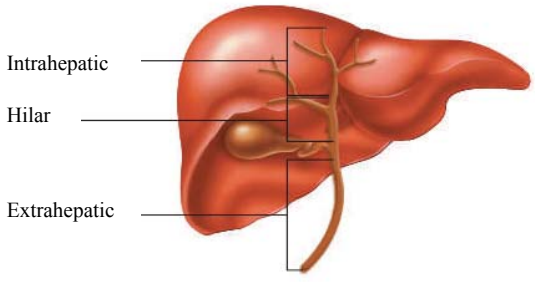


Figure 7. Anatomical classification of cholangiocarcinomas. Cholangiocarcinomas can be classified into extrahepatic and intrahepatic forms of the disease, accounting for 80%-90% and approximately 10% of the cases, respectively. Up to 50% of the cholangiocarcinomas are extrahepatic tumors which originate in the perihilar region, and which may extend into the liver and thus also be classified as intrahepatic forms of the disease. This group of lesions is referred to as Hilar lesions, or Klatskin tumors. From [84].

[7] www.kreftregisteret.no

1.4.2 Primary sclerosing cholangitis (PSC)

Primary sclerosing cholangitis (PSC), a chronic bile duct inflammation, is in the Western world identified as the major risk factor for cholangiocarcinoma [85,86]. Up to 30% of individuals diagnosed with PSC have been reported to develop neoplasia of the biliary tract [85]. PSC is a chronic liver disease characterized by inflammation and scarring of the bile ducts (Figure 8). The inflammation can obstruct the bile flow to the gut and the patients will eventually develop cirrhosis and liver failure. The mean age of a PSC patient is 30-40 years, and 70% – 80% of patients with PSC have associated inflammatory bowel disease [82,87]. The only curable option is liver transplantation. However, advanced cholangiocarcinomas often contradict liver transplantation due to high residual rate. Individuals with PSC tend to develop cholangiocarcinoma at younger age (30-50 years) compared to the general population (60-70 years), and the cancer is seldom suitable for resection [82,86]. Cholangiocarcinoma and PSC frequently result in similar cholangiographic findings which complicate the diagnosis of cholangiocarcinoma [80]. Brush cytology or biopsy may be used for the early detection of cellular atypia in patients with PSC [82].

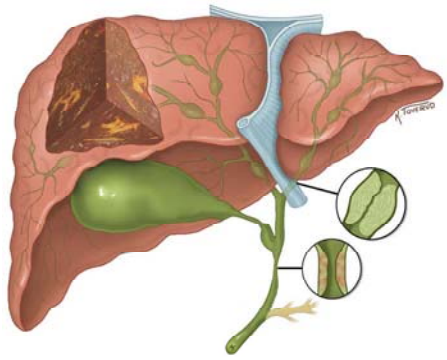


Figure 8. In PSC patients, inflammation and scarring of the bile ducts cause obstruction of the bile flow, cirrhosis and eventually liver failure. Scarring is also seen within the liver, which contradict liver resection for this patient group. Figure from the Norwegian PSC center**,** Kari C. Toverud (CMI - certified medical illustrator).

1.4.3 Diagnosis of cholangiocarcinoma

The diagnosis of cholangiocarcinoma is in general challenging. Chronic bile duct inflammation, including primary sclerosing cholangitis (PSC), is a recognized risk factor predisposing individuals to cholangiocarcinoma [79]. The clinical differentiation between patients diagnosed with PSC alone or complicated with cholangiocarcinoma is difficult, and

in up to 37% of PSC patients, cholangiocarcinoma is not detected until laparotomy in connection with intended liver transplantation or at autopsy [88]. Currently, the diagnostic approach is based on a combination of imaging modalities, including ultrasound (US), magnetic resonance imaging (MRI), computed tomography (CT) and cholangiopancreatography (MRCP), in addition to brush cytology and serum tumor markers [88]. CA 19-9 is the most commonly used tumor marker for cholangiocarcinoma. A sensitivity and specificity for detection of cholangiocarcinoma in patients with PSC is reported to be 79% and 98%, respectively, at a cutoff of 129 U/ml. In patients without PSC, the sensitivity and specificity drop to 53% and 76% - 92%, respectively, with a cutoff of >100 U/L [79]. Elevated CA 19-9 may also be associated with other malignancies such as acute bacterial cholangitis and cholestasis, and it has been reported that as many as 37% of PSC patients without cholangiocarcinoma had serum CA 19-9 > 129U/ml [88]. Needle biopsy may be necessary if many small lesions and liver cirrhosis is present. The abundant fibrotic tissue in PSC patients limits access to the actual tumor cells [84].

1.5 DNA methylation biomarkers for early cancer detection

"People keep talking about early- detection biomarkers as if they are a fact, and we only need to find them, when in reality their existence is a hypothesis that needs to be tested."

-McIntosh-

Molecular biomarkers can be defined as biological variables that correlate with biological outcome [89]. The biomarkers include *i.e.* DNA, RNA and protein and can be identified in for example tissue, blood/ plasma, feces, sperm, urine, and other bodily fluids. Cancer biomarkers can be prognostic, predictive, or diagnostic. The biomarkers with prognostic value can function as indicators for disease progression and survival and may also dictate whether further therapy is necessary. Predictive biomarkers assess the likelihood that a tumor will respond to a specific drug/treatment. It is for example well established that patients with *KRAS* mutations do not benefit from cetuximab and/or panitumumab treatment, which both are anti- EGFR monoclonal antibody drugs [90].

A biomarker suitable for early detection should be present in early- stage cancers, and also preferentially in benign precursor lesions. If the cancer is detected at an early stage the disease is most often curable, which will greatly reduce the number of cancer associated deaths. A biomarker present in high- risk benign lesions has the potential to detect the disease before it has become malignant. For colorectal cancer, colonoscopy can be used to remove these lesions, and thus reduce both the incidence and the number of deaths from the disease. In the following, existing epigenetic markers for early cancer detection as well as novel candidates will be mentioned.

1.5.1 DNA methylation markers

Today, two non- invasive DNA methylation tests for early detection of colorectal cancer are available on the marked: the Epi ProColon Early detection assay (Epigenomics) and the ColoSure test (Laboratory Corporation of America; LabCorp).

The Epi ProColon Early detection assay is based on colorectal cancer associated aberrantly methylated *SEPT9* (shed from tumor cells) in blood plasma. Several case control studies including more than 3000 persons, and also a prospect study called PRESEPT that included ~8000 healthy individuals confirmed by colonoscopy, have been used to evaluate the performance of methylated *SEPT9* as a screening biomarker. The combined sensitivity and specificity of the studies have been reported to be 70% and 91%, respectively, with a negative predication value of 99.7%[8]. Septin9 (*SEPT9*) is currently a test offered to Europeans with an average risk for developing colorectal cancer and who are without symptoms. The test was approved as a CE-marked test in October 2009. At the moment, larger clinical trials are performed and a recent press release indicated that a second generation of the test, named Epi *pro*Colon® 2.0 CE, can achieve a sensitivity of 80% and specificity of 99% with a positive predictive value (PPV) of 45%[9]. When the test performance is optimized for high sensitivity it reaches 95% with a subsequent reduction in specificity to 85%.

ColoSure is a stool based test which examines DNA shed from colon cells for aberrant cancer associated methylation of the vimentin (*VIM*) gene. The sensitivity and specificity have been reported in the range of 72-77% and 83-94%, respectively by LabCorp. Other studies have reported aberrant methylation of the *VIM* gene in fecal samples in 41%-88% of patients with

[8] http://www.aruplab.com/
[9] http://www.epigenomics.com/

colorectal cancer and 15%-45% of patients with adenomas. The specificities ranged from 82%-100% [91]. The detection rates for the general screening population have not yet been determined, and the test has not obtained approval from the U.S Food and Drug Administration (FDA).

Exact Sciences has recently developed a new stool based test for colorectal cancer screening, which is currently being tested in clinical trials. The test is named Cologuard, and consists of a panel of markers including four DNA methylation markers, two DNA mutation markers, and one human blood DNA marker. One of the methylated markers is the *VIM* gene, and if this test get FDA approval, which the company hopes to achieve in 2012, it is likely that the ColoSure test will be replaced. In a conference held in October 2010 Exact Sciences reported that Cologuard achieved 64% sensitivity for colorectal pre-cancers, 85% for cancers, and a false positive rate of 12% when the test- and validation sets were combined. Optimization of the assays is expected to improve the sensitivity.

Other DNA hypermethylation markers with potential to detect early stage colorectal cancer include *NEUROG*, which in serum have a reported sensitivity of 61% and a specificity of 91% across all tumor stages [92]; *MAL*, which is frequently methylated in both adenomas (84%) and in colorectal carcinomas (91%) and only rarely in normal mucosa (2%) [93-95], and *ITGA4* which also display a high methylation frequency in both adenoma (75%) and adenocarcinoma (92%) and in only 6% of normal colon mucosa samples [89], to mention some.

Recently, a promising six gene biomarker panel (*CNRIP1*, *FBN1*, *INA*, *MAL*, *SNCA*, and *SPG20*) was suggested by Lind *et al* [95], reporting a high methylation frequency in both adenomas (93%) and carcinomas (94%) with only 2% methylation in the normal mucosa. Although the panel has not yet been analyzed in fecal samples, the high methylation frequency in cancer tissue samples in combination with the low frequency in normal samples makes the panel a promising candidate for non- invasive early detection testing.

Few molecular biomarkers exist for cholangiocarcinoma, but *RASSF1A*, *SOCS3*, *p16INK4A, 14-3-3,* and *RUNX3* have been reported to be methylated in 27%-69%, 88%, 18%-83%, 60% and 57% of patients with cholangiocarcinoma, respectively [81,96]. Methylation of *SEMA3B* was in one study reported to occur in 100% of the individuals tested, and in none of the corresponding non- cancerous tissue samples. The study only included 15

cholangiocarcinoma patients [97]. *HOXA1* was in another study found to be hypermethylated in more than 95% of extrahepatic cholangiocarcinoma, and in none of the normal bile duct samples examined [98]. These highly sensitive and specific genes represent promising biomarkers for cholangiocarcinoma.

2 Overall project overview and the background for this master thesis

The step-wise experimental approach used in the present project to identify novel epigenetically deregulated cancer genes was first described by Lind *et al* in 2006 [99]. Colon cancer cell lines were treated with the demethylating agent 5-aza-2'deoxycytidine (AZA), and the gene expression before and after treatment was analyzed by cDNA microarrays and compared with microarray expression data from primary colorectal carcinomas and normal mucosa. CpG island containing genes upregulated after drug treatment and additionally downregulated in primary tumors compared to normal samples were subjected to downstream methylation analysis. By using this approach several novel epigenetically deregulated genes were identified for colorectal cancer [93-95,99,100].

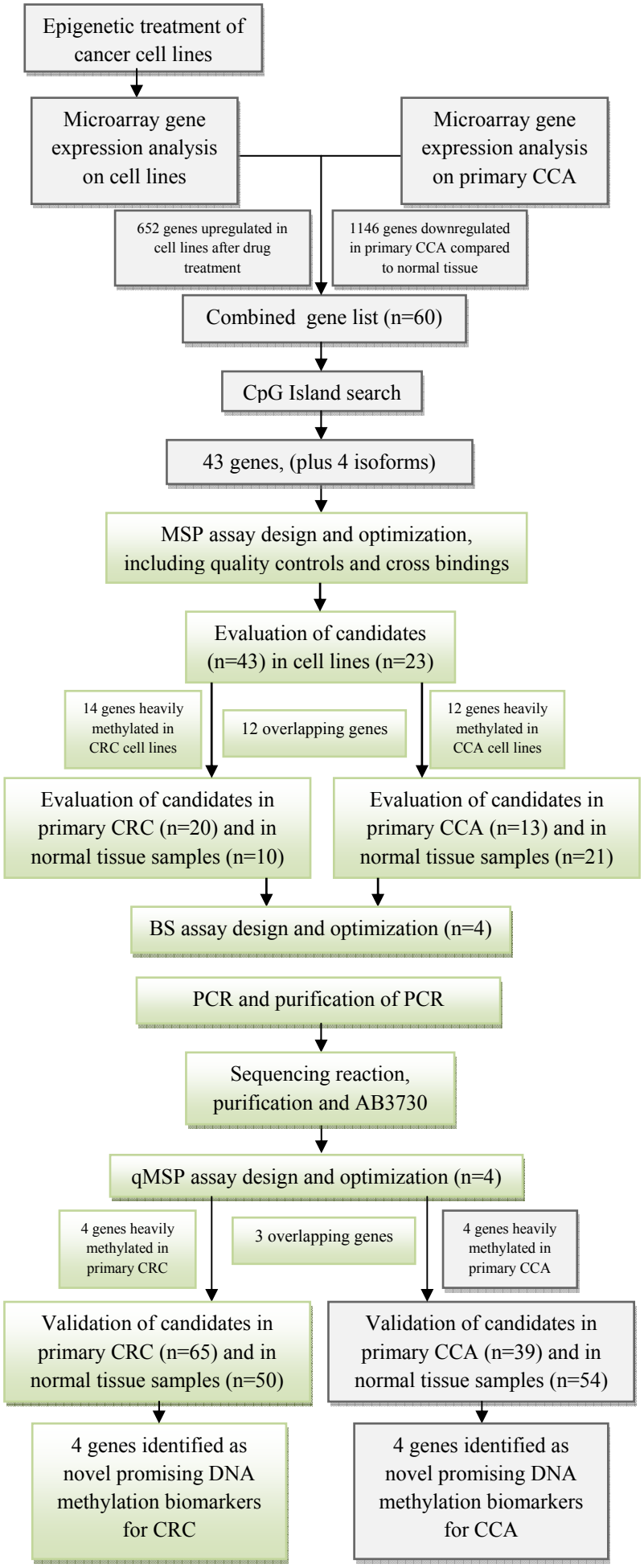
The candidate gene list analyzed in the present master thesis was generated by a PhD student in the group (Kim Andresen) in a similar manner as mentioned above. Andresen is financed by the Norwegian Centre for PSC and is primarily interested in identifying novel biomarkers for early detection of cholangiocarcinomas. In brief, six cholangiocarcinoma cell lines (see Materials) were subject to AZA/ tricostatin A (TSA) treatment, and the gene expression before and after treatment was analyzed on the AB1700 microarray platform (Applied Biosystems, Foster City, CA, USA). Six- hundred-and fifty-four genes were upregulated two or more times in at least four of the six cell lines after treatment. This dataset was combined with previously published microarray data for primary cholangiocarcinoma, where 1146 genes were found downregulated compared to normal tissue [101,102]. Sixty genes were overlapping in the two datasets. These genes were further investigated for the presence of CpG Islands in the promoter region, using the CpG Island searcher[10], leaving a final of 43 genes (in addition to four isofoms), which were subjected to downstream methylation analysis.

Tumors arising in the gastrointestinal tract have been shown to share several molecular aberrations, including genes that undergo hypermethylation [40,95]. Since the main focus in our department is on colorectal cancer we found it interesting to check if the candidate genes

[10] http://www.uscnorris.com/cpgislands

identified from the cholangicarcinoma-approach also showed promoter hypermethylation in colorectal cancer.

This project therefore includes two parts: a main part with focus on colorectal cancer and a minor, but essential part with focus on cholangiocarcinoma, which I also have been contributing on during my master. A scheme of the overall workflow is presented on the next page. The steps were I have contributed are marked in green.


Epigenetic treatment of cancer cell lines
Microarray gene expression analysis on cell lines
Microarray gene expression analysis on primary CCA
652 genes upregulated in cell lines after drug treatment
1146 genes downregulated in primary CCA compared to normal tissue
Combined gene list (n=60)
CpG Island search
43 genes, (plus 4 isoforms)
MSP assay design and optimization, including quality controls and cross bindings
Evaluation of candidates (n=43) in cell lines (n=23)
14 genes heavily methylated in CRC cell lines
12 overlapping genes
12 genes heavily methylated in CCA cell lines
Evaluation of candidates in primary CRC (n=20) and in normal tissue samples (n=10)
Evaluation of candidates in primary CCA (n=13) and in normal tissue samples (n=21)
BS assay design and optimization (n=4)
PCR and purification of PCR
Sequencing reaction, purification and AB3730
qMSP assay design and optimization (n=4)
4 genes heavily methylated in primary CRC
3 overlapping genes
4 genes heavily methylated in primary CCA
Validation of candidates in primary CRC (n=65) and in normal tissue samples (n=50)
Validation of candidates in primary CCA (n=39) and in normal tissue samples (n=54)
4 genes identified as novel promising DNA methylation biomarkers for CRC
4 genes identified as novel promising DNA methylation biomarkers for CCA

3 Aims

The overall project has two aims

1) To identify novel epigenetic biomarkers with a diagnostic potential for patients with cholangiocarcinoma alone or complicated with primary sclerosing cholangitis.

2) To test if the candidate target genes identified from the above mentioned-approach could be used as biomarkers also for early detection of colorectal cancer.

The second aim has been the main focus of this thesis.

4 Materials and methods

4.1 Materials

4.1.1 Cell lines

Thirty seven cancer cell lines were included in the present study, including six cholangiocarcinoma cell lines (EGI-1, HuCCT1, KMCH, KMBC, TFK-1, and Sk-ChA-1), two gall bladder carcinoma cell lines (Mz-ChA-1, and Mz-ChA-2), four liver carcinoma cell lines (HB8065, JHH1, JHH4, and JHH5) five pancreatic carcinoma cell lines (Panc-1, BxPC-3, AsPC-1, CFPAC-1, and PaCa-2) and twenty colon cancer cell lines: nine microsatellite instable (MSI; Co115, HCT15, HCT116, LoVo, LS174T, RKO, SW48, TC7, and TC71), and eleven microsatellite stable (MSS; ALA, Colo320, EB, FRI, HT29, IS1, IS2, IS3, LS1034, SW480, and V9P).

The cancer cell lines were purchased from the American Type Culture Collection[11] (ATCC; colon cancer cell lines), the German Collection of Microorganisms and Cell Cultures[12] (DSMZ; EGI-1 and TFK-1), the Japanese Collection of Research Bioresources[13] (JCRB; HuCCT1) or received through collaborative partners; Prof. Anne Kallioniemi, Institute of Biomedical Technology, Finland (pancreatic cancer cell lines), MD Prof. Alexander Knuth, Department of Oncology, University Hospital , Zurich (Sk-ChA-1 and gall bladder cancer cell lines) and Dr Gregory Gores, The Mayo Clinic, USA (KMCH-1 and KMBC). For detailed information about culturing conditions of the cancer cell lines, see Appendix I.

The cell lines were recently authenticated using the AmpFLSTR Identifiler (Applied Biosystems, CA, USA), which is a multiplex assay that amplifies 15 repeat loci, and the Amelogenin gender determining marker. GeneMapper (Applied biosystems) was used to analyze the samples run on an AB Prism 3730 (Applied Biosystems). All the cell lines have additionally tested negative for mycoplasma infection.

[11] http://www.lgcstandards-atcc.org/
[12] http://www.dsmz.de/
[13] http://cellbank.nibio.go.jp/cellbank_e.html

Genomic DNA from the cell lines has previously been isolated using a standard phenol/chloroform extraction protocol.

4.1.2 Tissue samples: tumors and normal tissue

Colorectal cancer samples and controls

Samples of primary colorectal cancer were collected from 65 individuals who underwent radical surgical resection at the Aker University Hospital from 2005- 2007. Normal mucosa samples (n=50) were obtained from deceased cancer- free persons (Institute of Forensic Medicine, University of Oslo).

For detailed clinico- pathological information, see Appendix II

DNA from these samples has previously been isolated using phenol/chloroform [103].

Cholangiocarcinoma samples and controls

Included in the present study were 13 primary cholangiocarcinoma samples. The samples were sourced from Rikshospitalet in the period 2008-2010 (n=9) and through a clinical collaborator (Dr Chris Wadsworth, Imperial College London, UK; n=4). Twenty-one normal peripheral liver tissue samples obtained from patients with liver diseases other than cancer were additionally included; 16 of these were diagnosed with primary sclerosing cholangitis (PSC), five had liver damage caused by alcohol, three had primary biliary cirrhosis (PBC), two were diagnosed with autoimmune hepatitis (AIH) while one person had hemakromatose and one had cryptogenic cirrhosis. In addition, dissected bile ducts were included from seven PSC patients, to a total of 28 cancer free samples.

All tissue samples were immediately after resection snap- frozen in liquid nitrogen and transferred to -80ºC. A section from each carcinoma was stained with hematoxylin and eosine and evaluated by an expert pathologist to confirm the diagnosis and estimate tumor cell content.

4.2 Methods

4.2.1 Sodium bisulfite modification

Theoretical background

The principle behind the sodium bisulfite reaction was first described in 1970 [104], but was not used to detect 5- methylcytosine until 20 years later, when Frommer and Clark published the first protocol for epigenetic analyses [105,106]. This revolutionized the field of epigenetics. With this method came the possibility to translate a chemical, non- readable code into a change in DNA sequence (unmethylated Cs are converted to Ts), which can readily be detected by numerous PCR based methods [107]. Sodium bisulfite reacts differentially with unmethylated and methylated cytosines. Under conditions with high bisulfite concentration and low pH, the bisulfite treatment will deaminate unmethylated, but not methylated, cytosines to uracils. In subsequent PCR, the uracil and thymine residues are amplified as thymine, whereas only the metylated cytosines are amplified as cytosines.

Complete denaturation of DNA is essential for this conversion, as the bisufite reaction is single strand specific. Unsuccessful denaturing may lead to insufficient conversion, which makes it difficult to discriminate between methylated and unmethylated cytosines in downstream analyzes and may even lead to false positive methylation findings when using poorly designed assays. In addition to fully denatured DNA, high DNA quality, correct pH and temperature for the various steps, and inclusion of a free radical to minimize the oxidative degradation by bisulfite help ensure effective conversion rates. It is possible to reach a 100% conversion rate although 95-98% is more likely.

Experimental procedure/ set-up

Bisulfite conversion of genomic DNA was performed using the EpiTect Bisulfite Kit from Qiagen (Qiagen Co., Valencia, California, USA) following the manufactor's standard protocol. Briefly, 1.3 µg DNA was mixed with RNase free water, a bisulfite solution and DNA protection buffer to a final volume of 140 µl. The protection buffer contains a pH indicator dye which turns from green to blue upon correct mixing, and which confirms the correct pH for complete cytosine conversion with minimal DNA fragmentation.

The reaction was performed in a thermo cycler (MJ Mini Personal Thermal Cycler, Bio-RAD, Hercules, CA, USA). The program consists of a series of denaturation (95ºC) and incubation (60ºC) steps, which ensures optimal conditions for DNA denaturation and subsequent sulfonation and cytosine deamination.

Following the conversion reaction, bisulfite converted DNA was purified using the QIAcube (Qiagen) automated pipetting system. The QIAcube provides a standardized clean- up procedure for removing bisulfite salts and chemicals that would otherwise inhibit subsequent PCR and sequencing procedures. DNA was eluted in 40µl elution buffer, giving a final DNA concentration of approximately 32.5ng/µl.

4.2.2 Methylation specific polymerase chain reaction (MSP)

Theoretical background

MSP, first described in 1996 [108], is a highly sensitive and specific technique for detecting methylated cytosines in DNA. It can distinguish as little as one methylated allele from 1000 unmethylated alleles. Bisulfite treated DNA is used as template with two primer sets designed specifically to distinguish methylated from unmethylated DNA. One primer set (containing four base types; A, T, C, and G) is designed to anneal to and amplify methylated fragments, while the other set (containing three base types; A, T, and G) is designed to anneal to and amplify unmethylated fragments. Following MSP, products can be visualized by gel electrophoresis, stained with an intercalating dye such as ethidium bromide which fluorescence under ultra violet light (Figure 9).

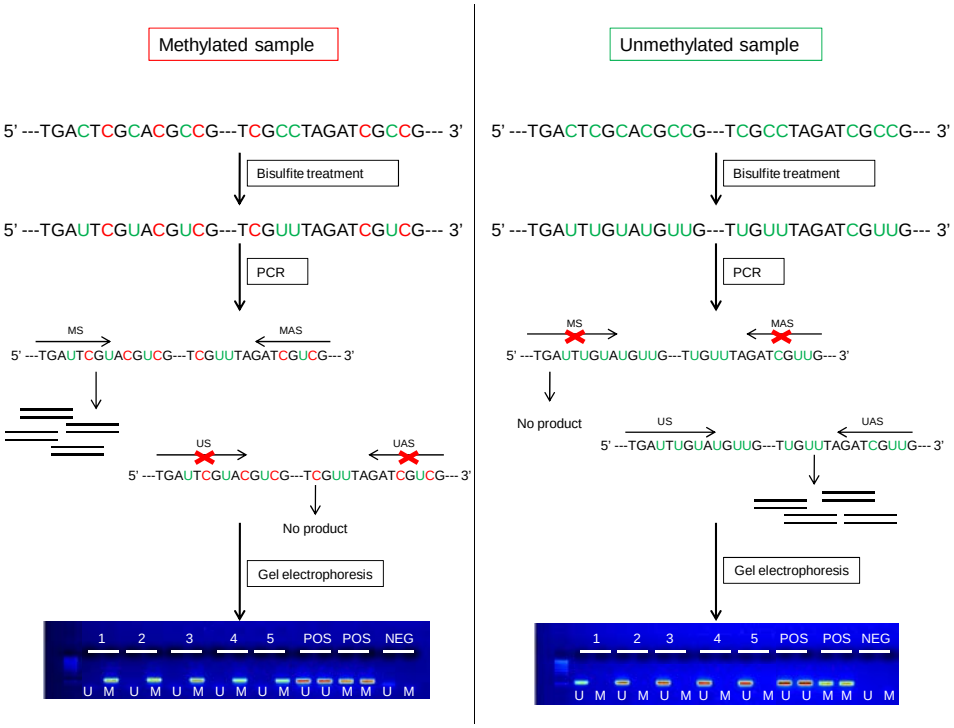


Figure 9 The principle behind methylation-specific polymerase chain reaction (MSP). Bisulfite treated DNA is used as template, with two primer sets designed specifically to distinguish methylated from unmethylated DNA. During PCR, only the methylated primer set will bind to the methylated fragments, and only primers designed towards the unmethylated fragments will bind to an amplify the unmethylated fragments. The amplified samples can be visualized by gelelectrophoresis. If both methylated and unmethylted alleles are present in a sample, a band will appear in both lanes (not shown). Abbreviations; M, methylated (fragment); MAS, methylation antisense; MS, methylation sense; NEG, negative (control); POS, positive (control); U, unmethylated (fragment) UAS, unmethyled sense; US, unmethylated sense.

Primer design is of crucial importance in MSP, as the specificity of the assay relies on the primers ability to discriminate between methylated, unmethylated, and unsuccessfully converted DNA. High quality primers should contain multiple CpG sites (≥ two in each primer), preferably with one in the very 3`end, to ensure discrimination between methylated and unmethylated sequences. Additionally, non- CpG cytosines should be included to avoid amplification of potentially unmodified DNA by the methylation specific primers, a source of false positive results. Finally, choosing the right CpG sites for methylation analysis is important. The correlation between loss of gene expression and presence of DNA promoter methylation is most likely to be observed the closer the methylation occurs to the transcription start site. Thus, ideally, primers should be designed to anneal to regions surrounding the transcription start site.

The conditions for the individual MSP reactions should be carefully optimized. Annealing temperature, magnesium concentration, and annealing- and elongation time may affect the efficiency of the PCR and the specificity of the primer sets. Magnesium functions as a cofactor for the polymerase, and is necessary for its efficient catalytic activity. However, too

high concentrations might lead to unspecific PCR products, especially in combination with low temperatures. The overall aim is to generate comparable band intensities for the unmethylated and methylated positive controls.

Experimental procedure/set-up

MSP was initially carried out to determine the promoter methylation status of 40 genes in 24 cancer cell lines. This work was shared between the Phd student working on the project and the author of the present thesis. Genes methylated in at least four of the six colon cancer cell lines (n=13) were further tested by MSP in 20 colorectal cancer tissue samples. *ASRGL1*, which was methylated in two of six cell lines was additionally included. Ten normal mucosa samples were also analyzed to evaluate whether the detected promoter methylation of the individual genes were cancer specific. With exception of *FKBP1*, which was methylated in only two of six cholangiocarcinoma cell lines, and *ASRGL1*, the same genes were additionally tested in 13 primary cholangiocarcinoma tissue samples, and 28 liver control samples from cancer free individuals. The criterion for inclusion was methylation of at least five of the six cholangiocarcinoma cell lines.

Normal blood (NB) and human placenta treated *in vitro* with *SssI* methyltransferase (*In vitro* methylated DNA; IVD) were used as positive controls for the unmetylated and methylated reactions, respectively, with Milli- Q water as negative PCR control for both reactions. In addition, to ensure primer specificity, both bisulfite treated- and non-bisulfite treated NB DNA were included in the methylated reactions, whereas IVD and non bisulfite treated NB were included in the unmethylated reactions.

MSP Primer design and optimization

The MSP primers used in this project were designed using Methyl Primer Express 1.0 (Applied Biosystems) and purchased from MedProbe (Oslo, Norway). For primer sequences, location, annealing temperature, magnesium content, and PCR program, see Appendix III.

The unmethylated and methylated primer pairs were optimized individually, with bisulfite converted NB as template for the unmethylated specific primers, and bisulfite converted IVD for the methylation-specific primers. Based on the theoretical melting temperatures (T_m) of the primer pairs, a range of temperatures for the annealing step was tested. For all primer

pairs, a magnesium gradient (1.5, 1.7, and 2.0 mM) was tested for each annealing temperature. For some primers, a higher magnesium concentration was necessary (see appendix, primer information).

Except for *SFRP1* and *TCF4*, adequate amount of MSP product was generated using standard cycling conditions (see below "MSP experimental assay"). For *SFRP1* the methylated and unmethylated gel band intensities showed high discrepancy after temperature and magnesium optimization, while both gel bands were overall very weak for *TCF4*. In these two cases an increase in annealing- and/or elongation time and /or number of cycles were tested.

MSP experimental procedure

The MSP mix consisted of 0.75µl bisulfite treated template DNA (approximately 24.4ng), 2.5µl 1xQiagen PCR buffer (containing 1.5mM $MgCl_2$), 0 – 0.5µl Qiagen $MgCl_2$ solution (25mM), 2µl of each primer (10mM, Medprobe), 2µl dNTP mix (4 x 2.5mM), 0.2µl Qiagen HotStar Taq Polymerase (Qiagen), and Milli- Q water to a total volume of 25µl. The PCR reactions were performed in a DNA Engine Tetrad 2 (Peltier Thermal Cycler, BIO RAD), and consisted of the following steps; an initialization step at 95ºC for 15 minutes to activate the enzyme, followed by 35 cycles of denaturation at 95ºC for 30 seconds, annealing at 45-60ºC for 30 seconds, elongation at 72ºC for 30 seconds and a final extension step at 72ºC for 7 minutes.

The MSP products were separated according to size by gel electrophoresis. Five µl gel loading buffer (1x TAE buffer and 0.1% xyelen cyanol) was added to the MSP products and loaded onto a 2% agarose gel stained with ethidium bromide (VWR, Pennsylvania, USA). The electrophoresis was run at 200V for 25 minutes and the products were visualized on an UV trans- illuminator (Chemidoc XRS Gel Documentation System, BIO-RAD).

Scoring of MSP results

The MSP fragments separated by gel electrophoresis were scored manually by comparing the resulting band intensities of the samples to the intensity of the positive controls (NB and IVD). For the cell lines, samples were scored as methylated when a band was present in only the methylated reaction, and as unmethylated if the band intensity for the methylated PCR reaction was very weak or absent, but the band intensity from the unmethylated PCR reaction

was acceptable. Samples in which both primer sets were amplified were considered to be hemimethylated. When scoring the tumor samples, a revised scoring system was introduced. Each band was automatically given a color by the software based on the intensity of the band, where a blue color indicated a weak signal and red a strong signal. We translated these colors into values (0-5), where 0 indicated no visible band and 5 indicated a very strong/ red band. When the positive controls were scored as 5, the samples with values ranging from 3-5 were considered methylated. For normal samples, those scored as 2 were additionally regarded as methylated. Since tumor tissue contains a heterogeneous cell population, including normal cells, tissue samples were only scored as methylated or unmethylated. The scorings were performed independently by two individuals (the author and another group member, Kim Andersen). Two rounds of analysis were performed, and in cases where the two disagreed about the scoring or the results were diverging in the two rounds, a third MSP was performed.

4.2.3 Bisulfite sequencing: validation of candidates step 1

Theoretical background

Bisulfite sequencing is considered the gold standard to detect 5- methylcytosine, and provides a reliable way of detecting any methylated cytosine at single- molecule resolution in any sequence context [109]. In this method, based on Sanger sequencing [110], genomic DNA is treated with sodium bisulfite and amplified with PCR. The primers are designed to avoid CpG sites and the target will therefore be amplified regardless of the methylation status of the internal sequence. In the final sequence, 5- methylcytosines will appear as cytosines while unmethylated cytosines will be displayed as thymines. The amplified PCR products can be sequenced directly, or alternatively be cloned into plasmid vectors before the individual clones are subject to sequencing.

Experimental design/ set-up

The cholangiocarcinoma-, gall bladder carcinoma-, and colon cancer cell lines were subject to direct bisulfite sequencing to confirm the methylation status of four genes (*CDO1*, *DCLK1*, *ZNF331*, and *ZSCAN18*), as well as to guide the design of the quantitative methylation-specific PCR assays. Normal blood and IVD were included as a methylation negative and methylation positive control, respectively.

Bisulfite sequencing assay design and optimization

Primers were designed to avoid CpG sites and to anneal to regions outside the MSP area, using Methyl Primer Express 1.0 (Applied Biosystems). Primers that amplified a region containing >9 poly(T) were additionally avoided, as it can lead to insertions or deletions due to (AmpliTaq DNA) polymerase slippage. For one primer pair, excluding CpG sites in the primer annealing area and at the same time amplifying the region of interest was not possible. In this case, to avoid preferential amplification of methylated or unmethylated DNA, the primer mix was designed to contain sequences with equal amount of both cytosine and thymine residues.

All primer pairs were optimized with respect to temperature and magnesium concentration. NB and IVD were used as templates, and the temperature and magnesium concentration which amplified NB and IVD with equal efficiency were chosen. See Appendix III for information about primer sequences and location, annealing temperature and Mg concentration.

Initial PCR

The initial PCR was performed on cancer cell lines for the four genes in question (n=14 for *CDO1*, *ZNF331*, and *ZSCAN18;* n=28 for *DCLK1*). The experimental assays were similar to the MSP reactions (see previous), with equal amount and concentration of template, Qiagen Buffer, dNTP mix, Qiagen HotStar Taq polymerase and milli- Q water. Two µl of each sequencing primer was included (10mM, Medprobe) to a final volume of 25µl. The PCR reaction was performed in a DNA Engine Tetrad 2 (Peltier Thermal Cycler, BIO RAD), and except from the annealing temperatures which varied from 49-59°C, the PCR reaction steps were identical to what has been described earlier for the MSP reactions. Five µl of the products were visualized on a 2% agarose gel.

Purification of PCR products: EXOSAP- IT

For optimal results, unincorporated dNTPs and primers from the initial PCR were removed prior to sequencing, using EXOSAP-IT (GE HEALTHCARE). One point five µl EXOSAP-IT, containing the two hydrolytic enzymes exonuclease I and shrimp alkaline phosphatase, was mixed with 10µl PCR product and incubated at 37ºC for 15 minutes before an enzyme

inactivation step at 80ºC for 15 minutes. The procedure was performed in a DNA Engine Tetrad 2 (Peltier Thermal Cycler, BIO RAD).

Sequencing reaction

The samples were sequenced using the dGTP BigDye Terminator v3.0 Ready Reaction sequencing Kit (Applied Biosystems, Foster City, CA), which contains dGTP in the deoxynucleoside triphosphate mix instead of dITP used in the standard cycle sequencing kits from Applied Biosystems. In cases where the obtained results were not satisfying, the Big Dye Terminator v1.1 Cycle sequencing Kit, containing dITP to minimize band compression, was tried with satisfying results. The Ready reaction mix additionally contains AmpliTaq DNA polymerase and fluorescent labeled dideoxynucleoside triphosphate (ddNTPs).

Each reaction consisted of 2µl purified PCR product, 0.25µl forward or reverse primer (10µM), 2µl 5 x Big Dye Terminator v1.1 Sequencing Buffer (Applied Biosystems), 2µl Ready Reaction mix and adjusted with Milli- Q water to a total volume of 10µl. The thermal cycling was performed in a DNA Engine Tetrad 2 (Peltier Thermal Cycler, BIO RAD) and involved an initial denaturation step at 96ºC for two minutes, followed by 25 cycles of denaturation for 15 seconds, annealing at 50ºC for five seconds and elongation at 60ºC for four minutes.

Purification of sequencing products: Sephadex rinsing

Before analyzing the samples by electrophoresis, unincorporated dye terminators and unused primers from the sequencing reaction were removed by centrifuging the samples through a Sephadex column, where only large molecules, likes the sequencing products, will pass relatively free through the column and be eluted with the flow of the buffer.

Sephadex G-50 Superfine Powder (GE Heakthcare) was poured onto a 96- well Multiscreen HV plate (Millipore), added 300µl Milli- Q water and left for swelling either in room temperature for at least two hours, or overnight in the refrigerator. The plate was then centrifuged at 910 rcf for five minutes, added 150µl Millli- Q water for rinsing, followed by another round of centrifugation for five minutes. The Sephadex plate was placed on a 96 well Optical Reaction Plate (Applied Biosystems), and 10µl sequencing product and 10µl Milli- Q water were added before a final round of centrifugation at 910 rcf for six minutes. A 3100

Genetic Analyzer Plate Septa (Applied Biosystems) was used to cover the plate, which was placed in a 96- well DNA Analyzer`s stacker and inserted into an ABI PRISM 3730 Sequencer (Applied Biosystems).

ABI PRISM 3730 Sequencer

In the ABI PRISM 3730 Sequencer the labeled sequencing products are separated according to size as they travel through the polymer- filled capillary array. When the DNA fragments reach the detection window, a laser beam excites the dye molecules, causing them to fluoresce. The fluorescent data is interpreted by the software and displayed as electropherograms.

Analyzing of the electropherograms

The electropherograms were analyzed manually using Sequencing Analysis 5.2 (Applied Biosystems). The approximate amount of methylation at each CpG site was calculated by dividing the peak height of the cytosine signal with the peak height of the cytosine plus the thymine signal [111], giving values ranging from 0-1. CpG sites with methylation frequencies ≤0.20 were considered unmetylated, CpG sites with methylation frequencies between 0.2 – 0.8 were considered partially methylated, while those CpG sites with values ≥0.80 were considered hypermethylated.

4.2.4 Quantitative MSP (qMSP): validation of candidates step 2

Theoretical background

Quantitative MSP or real- time MSP is a method for determining the amount of methylated target sequences present in a sample. Data is collected during the exponential growth (log) phase of the PCR, where the quantity of the PCR product is directly proportional to the amount of template nucleic acid. Two values are collected during this phase, the threshold line, and the cycle threshold (Ct). The threshold line is the level where the fluorescent intensity from the sample is greater than the background, and the Ct value is the PCR cycle at which this level is reached. When calculating the amount of starting template, absolute or relative quantification may be used. In absolute quantification a standard curve is created by serial dilutions of a sample with known concentration. The concentrations of the unknown

samples can then be accurately determined by interpolating their quantity from the standard curve. The Ct value is inversely proportional to the amount of input template, and is the value used by real- time PCR systems to perform quantization.

In addition to primers, the qMSP assay includes a fluorescent labeled probe. The TaqMan MGB probe, which is used in this project, has a fluorescent reporter dye attached to the 5' end, and a non- fluorescent quencher, in addition to a minor groove binder (MGB), attached to the 3' end (Figure 10). When the probe is intact, the proximity of the quencher reduces the fluorescence emitted by the reporter. During primer extension the Taq DNA polymerase cleaves the probe, which separates the reporter dye from the quencher, causing it to fluorescence. The fluorescence signal can be captured by a laser detector of a real-time PCR system. The MGB raises the melting temperature (Tm) of the probe, enabling the design of shorter probes which provide better discrimination. The Tm of the probe should be approximately 10°C higher than the Tm for the primers, to ensure that the probe binds prior to primer annealing and extension.

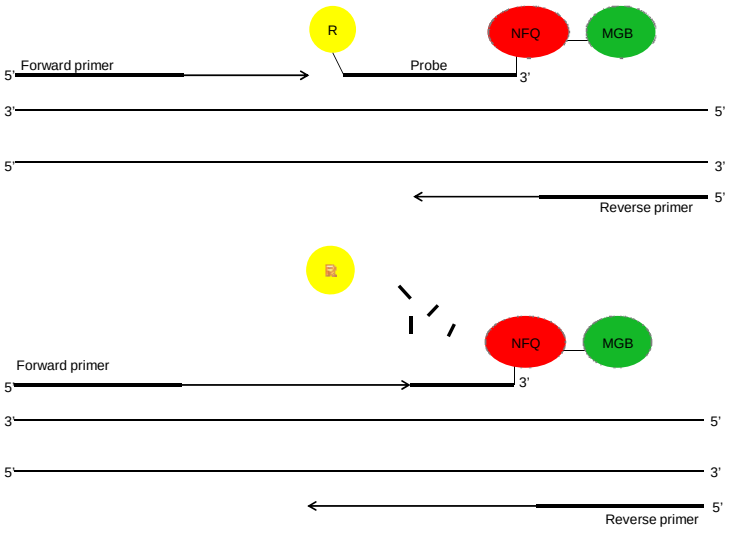


Figure 10 The TaqMan minir groove binder (MGB) assay chemistry. The TaqMan MGB probe contains a reporter dye and a quencher attached to the 5' and 3' end, respectively. When the probe is intact, the proximity of the quencher reduces the fluorescence emitted by the reporter. During primer extension the Taq DNA polymerase cleaves the probe, which separates the reporter dye from the quencher, causing it to fluorescence.

Experimental design/ set-up

The promoter methylation of four genes (*CDO1*, *DCLK1*, *ZNF331*, and *ZSCAN18*) was investigated by qMSP in a larger samples series, including 65 primary colorectal cancers and

50 normal mucosa samples. Bisulfite treated normal blood, unmodified DNA, and RNase free water (Sigma- Aldrich) were included in each reaction as negative controls, whereas commercially available bisulfite- converted *in vitro* methylated DNA (IVD; Chemicon) was used as positive control. The same bisulfite converted methylated DNA was also used to generate a standard curve from 1:5 serial dilutions (32.5ng- 0.052ng).

The Primer Express Software 3.0 (Applied Biosystems) was used to manually design primers and TaqMan MGB probes for the quantitative real- time analysis. The primers and probes for each assay were designed to preferentially cover ≥10 CpG sites, and to anneal to regions of highly methylated CpG sites (see page 37, Bisulfitesequencing: Experimental design/ set-up). The primers were purchased from MedProbe (Oslo, Norway), whereas the probes were purchased from Applied Biosystems. For primer and probe sequence information, see Appendix IV.

qMSP experimental procedure

The PCRs were carried out in a reaction volume of 20µl, consisting of 10µl 1xTaqMan Universal PCR Mastermix No AmpErase UNG (including AmpliTaq Gold DNA polymerase and passive reference; ROX, Applied Biosystems), 0.18µl of each primer (100µM), 0.40µl probe (10µM), 6.24µl RNase free water (Sigma- Aldrich), and 3µl bisulfite treated template DNA (32.5ng). Amplifications were carried in triplicates in 384-well plates in a 7900HT Fast Real- Time PCR System (Applied Biosystems), and thermal cycling was initiated with a first denaturation step at 95°C for 10 minutes, followed by 45 cycles of 95°C for 15 seconds and 60°C for 60 seconds. The EpMotion 5075 pipetting robot (Eppendorf, Hamburg, Germany), was used to automatically distribute both template and master mix to the 384- well plates, minimizing possible pipetting errors and standardizing the procedure.

Calculation of methylation (percent methylated reference)

For each sample the median quantity value was used for data analysis. In order to include samples, at least two of the three replicates had to be amplified, and a cutoff was set at cycle 35; all samples with amplification after cycle 35 were censored. To normalize for DNA input the ALU-C4 element was used as reference [112], and the positive control (IVD, Chemicon) served as a methylated reference. The qMSP results were calculated as percent of methylated

reference (PMR) by dividing the normalized quantity of the samples by the normalized quantity of the positive control (IVD Chemicon) and multiply by 100.

$$PMR = \underbrace{(Qty\_Gene/Qty\_ALU)_{sample}}_{\text{Normalized sample quantity}} / \underbrace{(Qty\_Gene/Qty\_ALU)_{IVD}}_{\text{Normalized IVD quantity}} * 100$$

To ensure high specificity for tissue samples, the thresholds for scoring samples as methylated were set according to the highest PMR value across the normal mucosa samples. Three samples had outlier PMR values (CDO1, 24.15 and 11.43; DCLK, 8.49). These were excluded when the thresholds were set.

4.2.5 Statistics

All The statistical analyses were carried out using the SPSS 18.0 software.

When comparing mean PMR values obtained by qMSP a Students T-test was used. This test was also used to evaluate a possible association between age and DNA methylation status.

For all the other calculations Pearson Chi- Square and Fisher's exact tests were used. ROC curves were generated by using the PMR values from colorectal cancer- and normal mucosa-samples.

All *P*-values derive from two-sided tests and $P \geq 0.05$ was considered statistically significant.

5 Results

5.1 Novel DNA methylation candidate genes identified by epigenetic drug treatment and microarray gene expression

From the microarray analysis and CpG Island search (see page 26, Project overview) performed before the start of this thesis a list of 43 candidate genes, in addition to four isoforms, were generated. These genes represent potentially novel epigenetically deregulated genes and were subject to downstream DNA methylation analysis. Six of the 47 candidates were excluded for further analysis after sub-optimal results in the initial quality control of the MSP assays, indicating that the primers had specificity issues that could not be resolved by redesigning the respective assay. A second round of optimization of one primer set (STXBP1) removed unspecific PCR products, but the gene was still excluded for further analysis. A final list of 40 candidates was subjected to MSP analysis. Gene symbol, gene name, chromosomal location and accession number of the analyzed genes can be found in Appendix V.

5.2 Qualitative methylation analysis by methylation-specific PCR (MSP)

5.2.1 MSP in cancer cell lines

The 40 analyzed candidate genes were grouped into three categories according to their methylation frequencies from MSP analyses of colon cancer cell lines (Figure 11). Group I (n=14) included highly methylated genes (more than 66.7%; ≥4/6), group II (n=9) included genes with intermediate methylation frequencies (between 16.7% and 50%; 1-3/6) and group III genes (n=17) were unmethylated in all colon cancer cell lines analyzed. The majority (n=12) of group I genes were also highly methylated in cholangiocarcinoma cell lines (in more than 83%: ≥5/6). These genes were the most interesting from a biomarker perspective, and were further tested in primary tumor samples and normal samples for the respective tissue types.

Figure 11. MSP results of 40 genes in 24 cancer cell lines. The genes are divided into three groups according to the methylation frequency among the colon cancer cell lines. The group to the left contains genes methylated in ≥4/6 of the colon cancer cell lines, the group in the middle contains genes methylated in 1-3/6 colon cancer cell lines, and the group to the right contains genes unmethylated in all the colon cancer cell lines. Similar methylation frequencies are seen between the colon cancer cell lines and the cholangiocarcinoma cell lines; twelve genes are frequently hypermethylated across both types of cancer cell lines. *GNG11* was not included for further studies and is therefore marked in grey. Abbreviations; CCA, cholangiocarcinoma; CRC, colorectal cancer; HCC, hepatocellular carcinoma; GBC, gallbladder carcinoma; PC, pancreatic carcinoma.

5.2.2 MSP in tissue samples

MSP in colorectal cancer- and normal tissue samples

With the exception of *GNG11* which has previously been analyzed by our group (unpublished data), all group I genes from the cell line analysis (n=13), in addition to one gene from group II (ASRGL1), were subject to MSP analysis in colorectal cancer- (n=20) and normal- tissue samples (n=10) (Figure 12). *CDO1* and *SFRP1* were methylated in 95% of the tumor samples, followed by *DCLK1* (90%), *NAP1L2* (85%), *ZNF331* (80%), and *ZSCAN18* (75%). In terms of sensitivity, these genes were the most promising. *CDO1*, *SFRP1*, *DCLK1*, and *NAP1L2* also displayed apparently high methylation frequencies among the normal mucosa samples with 80%, 100%, 80%, and 50% respectively. However, with the exception of *SFRP1* and *NAP1L2* which were excluded from further analysis, the majority of the normal samples were only weakly methylated for these genes. Hence *CDO1*, *DCLK1*, *ZNF331* and *ZSCAN18* were selected for further analysis.

BEX4, *TCF4*, *LHX6*, and *GREM1* were considerably more methylated among MSI tumors compared to the MSS tumors, and except from *BEX4* they showed low methylation frequencies in the normal mucosa.

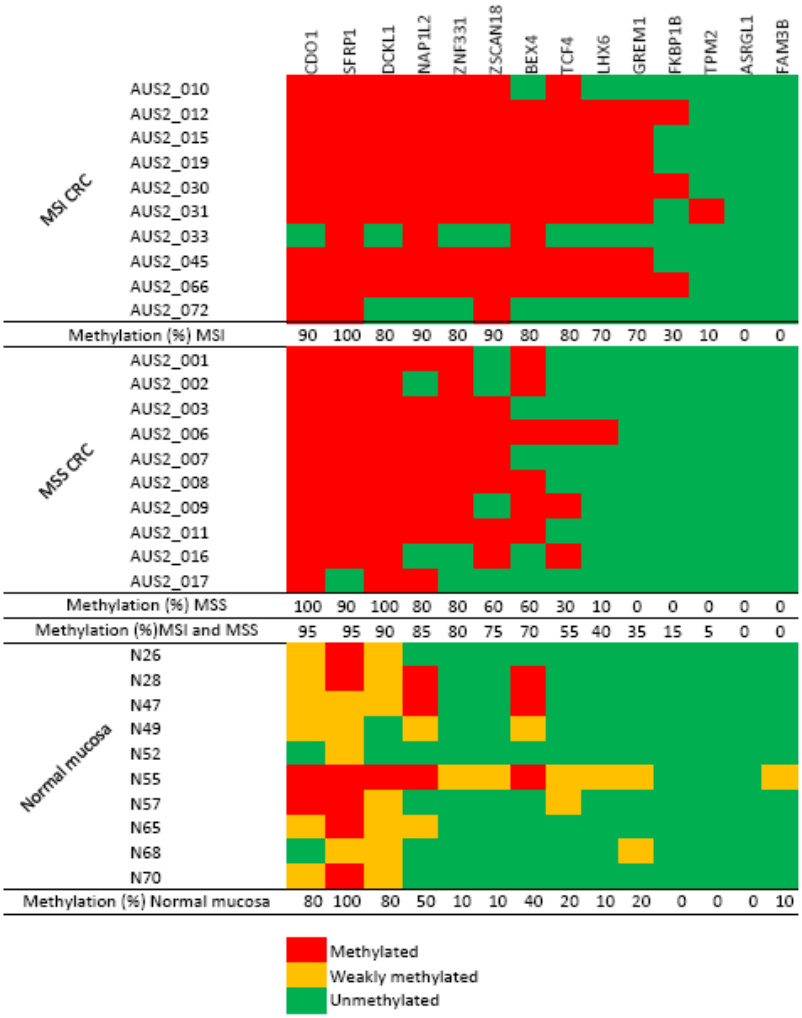


Figure 12. MSP results of 14 genes in colorectal cancer- and normal tissue samples. Six genes, CDO1, SFRP1, DCLK1, NAP1L2, ZNF331 and ZSCAN18, where methylated in ≥75% of the colorectal cancer tissue samples. Of these, CDO1, DCLK1, ZNF331 and ZSCAN18 were simultaneously unmethylated or weakly methylated in ≥80% of normal tissue samples and were selected for further analyses. Normal samples with a weak methylated band are marked in orange, whereas samples with a strong methylated band are marked in red. Abbreviations: CRC: colorectal cancer; MSI, microsatellite unstable; MSS, microsatellite stable.

MSP in cholongiocarcinoma- and cancer free tissue samples

The twelve group I genes frequently methylated also among the cholangiocarcinoma cell lines (Figure 11) were subject to MSP analyses in cholangiocarcinoma- (n=13) and cancer free liver- tissue samples (n=28). *SFRP1, DCLK1, BEX4, NAP1L2, CDO1*, and *ZSCAN18* were methylated in 85%, 83%, 69%, 69%, 62%, and 31% of the primary cholangiocarcinoma samples, respectively (Figure 13). Four of these genes (*SFRP1*, *DCLK1*, *BEX4*, and *NAP1L2*) also displayed a high methylation frequency in the cancer free samples, 86%, 100%, 33%, and 38%, respectively. Also for these genes, the intensities of the methylated bands varied, and were generally weaker than the methylated bands in the tumor samples for *SFRP1* and

DCLK1. *BEX4* and *NAP1L2* had stronger methylated bands among the normal controls, representing tumor specificity issues and were not subjected to further analyses.

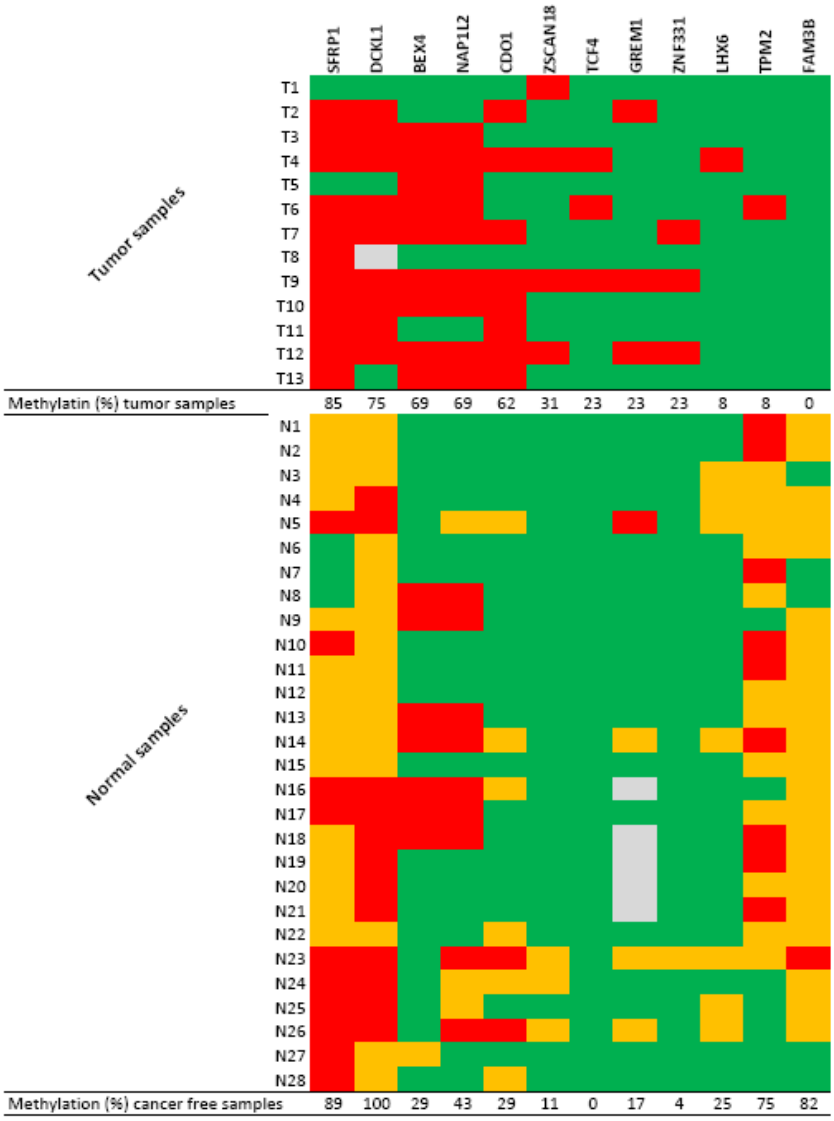
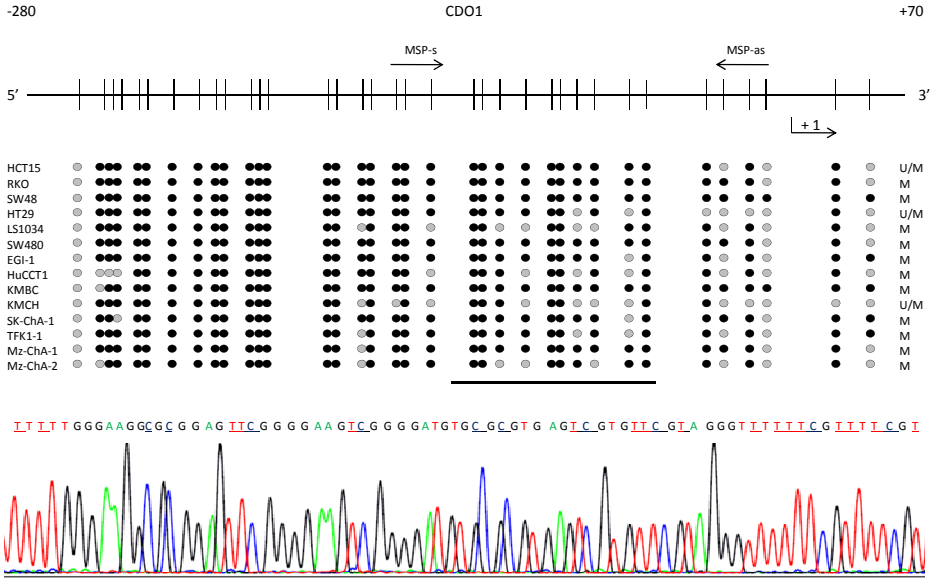


Figure 13. MSP results of 12 genes in primary cholangiocarcinoma and cancer free controls. *SFRP1* and *DCLK1* were methylated in 85% and 83% of the tumor samples respectively, followed by *BEX4* (69%), *NAP1L2* (69%), *CDO1* (62%), and *ZSCAN18* (31%). The methylation frequencies of these genes in the cancer free controls varied from 0-100%. Weakly methylated cancer free samples are marked in orange, and the strong methylated samples are marked in red.

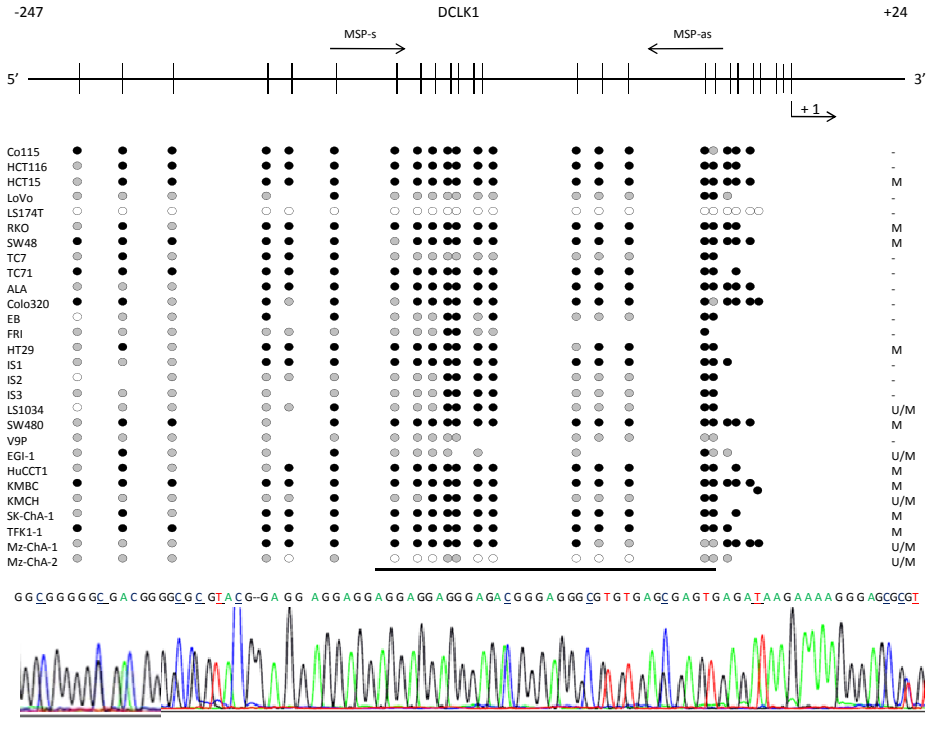
5.3 Validation of methylation status by bisulfite sequencing

The methylation status of the four most promising genes from the MSP colorectal cancer tissue analysis (*CDO1*, *DCLK1*, *ZNF331* and *ZSCAN18*) were validated by direct bisulfite sequencing of cancer cell lines. A good association between MSP status and bisulfite

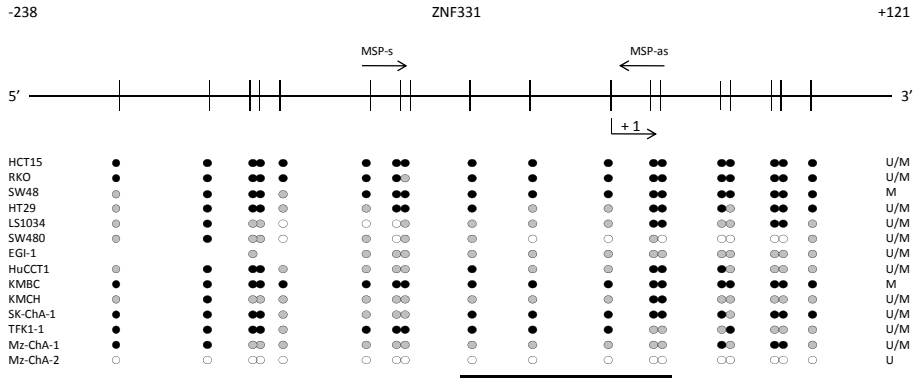
sequencing results was observed for all of the four genes in question (Figure 14). Bisulfite sequencing thus confirmed that no false positives were generated from the cell line MSP analysis (a known pitfall for this method), and further confirmed that all non- CpG sites were converted to uracil (and read as thymine), underscoring a successful bisulfite conversion.

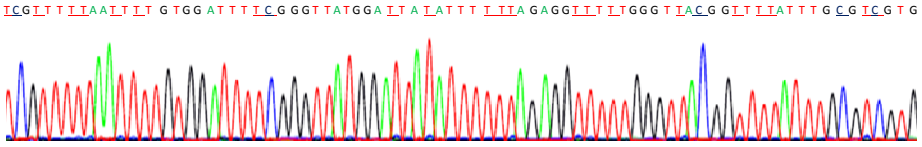


Electropherogram from the cell line SW48

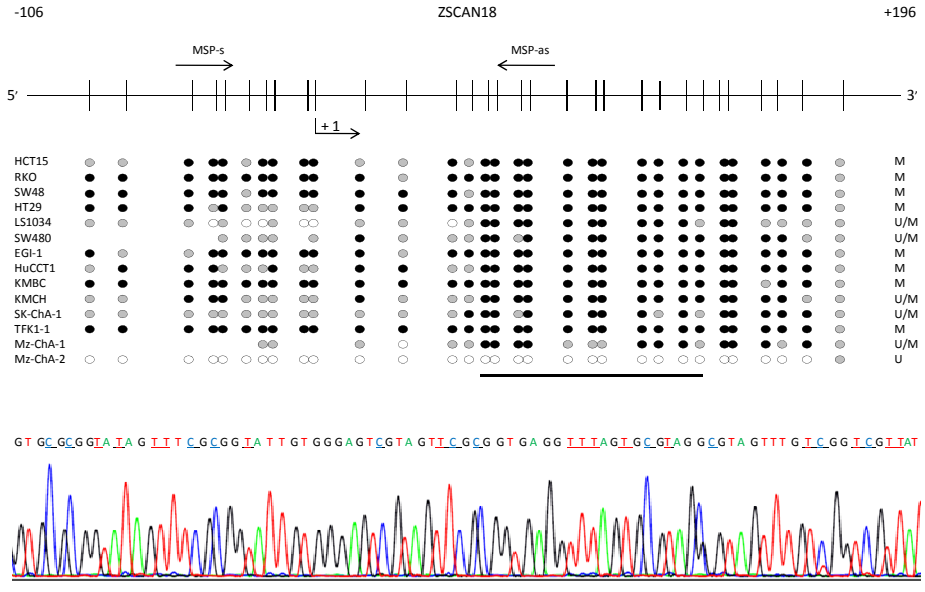


Electropherogram from the cell line Co115





Electropherogram from the cell line HCT15



Electropherogram from the cell line RKO

Figure 14. Methylation status of individual CpG sites determined by bisulfite sequencing. The upper part of each figure is a schematic presentation of the area and the CpG sites covered by the bisulfite sequencing primers. The vertical bars represent individual CpG sites, the lower arrow marked by +1 indicates the transcription start site, and the two upper arrows marked by MSP-s and MSP-as indicate the location of the methylation specific primers. For the middle part of the figures, black circles represent methylated CpG sites (the ratio of C/(C+T)≥0.8), grey circles represent partially methylated CpG sites (the ratio of C/(C+T) 0.8<0.20), and open circles represent unmethylated CpG sites (the rationof C/(C+T)<0.20). The methylation status of the individual cell lines as assessed by MSP analysis are listed to the right. In the lower part of each figure, representative electropherograms for a selected colon cancer cell line are depicted. Underlined Cs and Ts represent methylated cytosines and unmethylated cytosines before bisulfite treatment, respectively.

5.4 Quantitative methylation analysis by qMSP

5.4.1 qMSP in colorectal cancer- and normal tissue samples

Genes that were methylated in at least 75% of the tumor samples and simultaneously unmethylated/ weakly methylated in ≥ 80% of the normal samples as assessed by MSP in the pilot, were further validated by qMSP in a larger sample series. Four genes fulfilled these criteria: *CDO1*, *DCLK1*, *ZNF331*, and *ZSCAN18*. The results from the qMSP analysis are summarized in Table 1 and Figure 15. The thresholds for *CDO1*, *DCLK1*, *ZNF331* and *ZSCAN18* were set at 8, 4, 1, and 4, respectively. All samples with a PMR value equal to or above these thresholds were scored as methylation positive.

Table 1. Methylation frequencies of target genes in a validation series of primary colorectal cancers and normal mucosa samples, assessed by quantitative MSP (qMSP).

	Methylation % (ratio)	*P*-values*	Range PMR values	Median PMR values	Mean PMR values	*P*-values¥
CDO1						
CRC	91% (59/65)	**$1.96*10^{-14}$**	0-57.34	28.55	27.8	**$3.16*10^{-24}$**
Normals	4% (2/50)		0-24.15	1.32	2.36	
DCLK1						
CRC	80% (52/65)	**$8.46*10^{-15}$**	0-60.23	22.04	23.25	**$7.85*10^{-16}$**
Normals	2% (1/50)		0-8.49	0.35	0,67	
ZNF331						
CRC	66% (43/65)	**$1.45*10^{-14}$**	0-43.53	9.81	11.70	**$6.79*10^{-11}$**
Normals	0% (0/50)		0-0.68	0	0.03	
ZSCAN18						
CRC	66% (43/65)	**$1.45*10^{-14}$**	0-58.14	14.69	18.8	**$2.65*10^{-11}$**
Normals	0% (0/50)		0-3.15	0.18	0.34	

Pearson Chi- Square and Fisher's Exact tests were used to calculate the *P*- values* in the first column of *P*- values. The second column of *P*-values¥ was calculated by using Students T-test. Samples with a PMR value equal to or higher than the threshold were scored as methylation positive, whereas samples with a lower PMR value than the threshold were scored as unmethylated. Abbreviations; PMR, percent methylated reference.

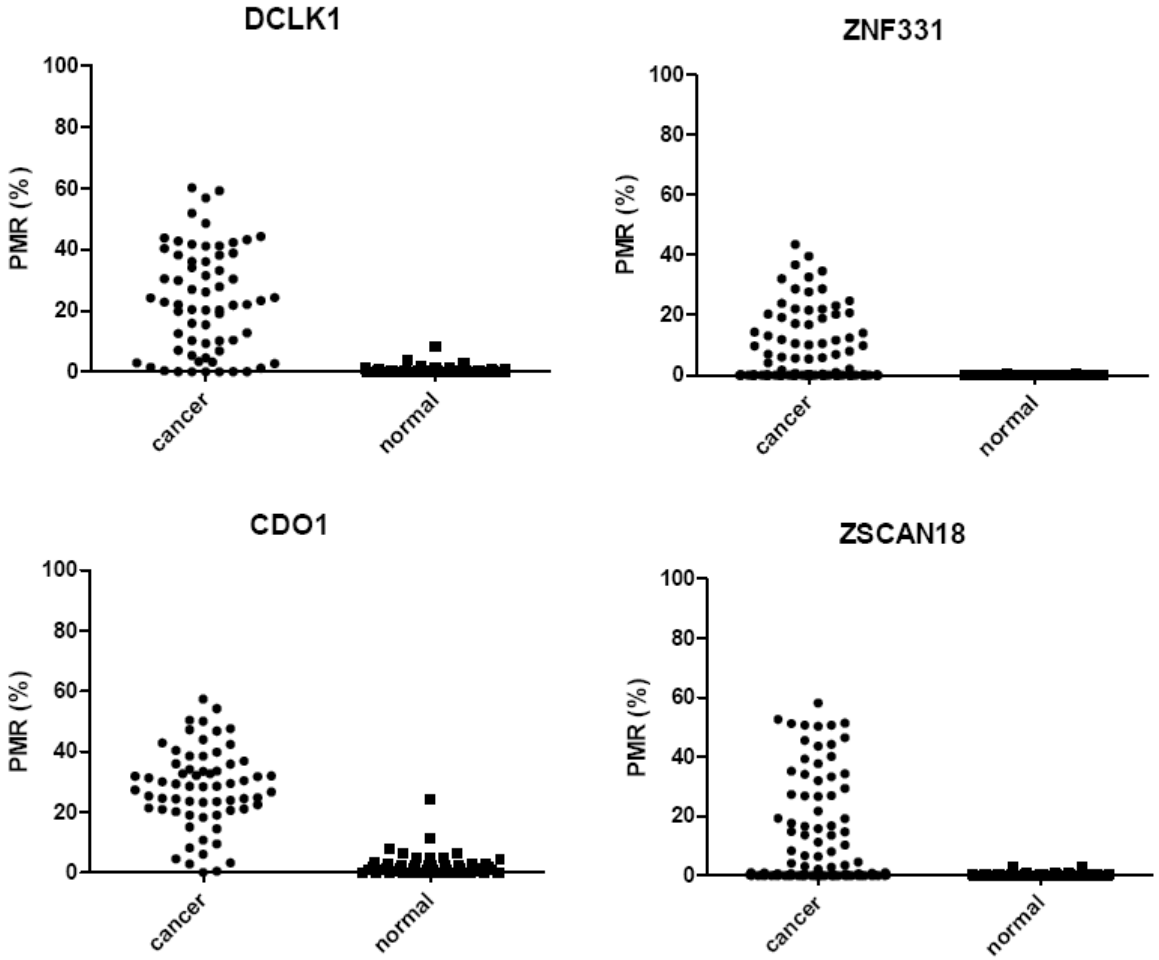


Figure 15. The percent methylated reference (PMR) values from colorectal cancer and normal mucosa samples as assessed by qMSP.

Receiver operating characteristic (ROC) curve analysis was applied to evaluate the diagnostic performance of the four genes, both individually and combined, using PMR values and tissue type (carcinoma and normal) as input. The ROC curves are visualized in Figure 16. The area under the curve (AUC) value for the individual genes varied from 0.854 (*ZSCAN18*) to 0.960 (*CDO1*), and the combined ROC curves of the four markers resulted in an AUC of 0.976. Methylation of at least one of the four genes was observed in 62 of the 65 cancers, and in 2 out of the 50 normal mucosa samples, resulting in a "panel" sensitivity and specificity of 95% and 96%, respectively.

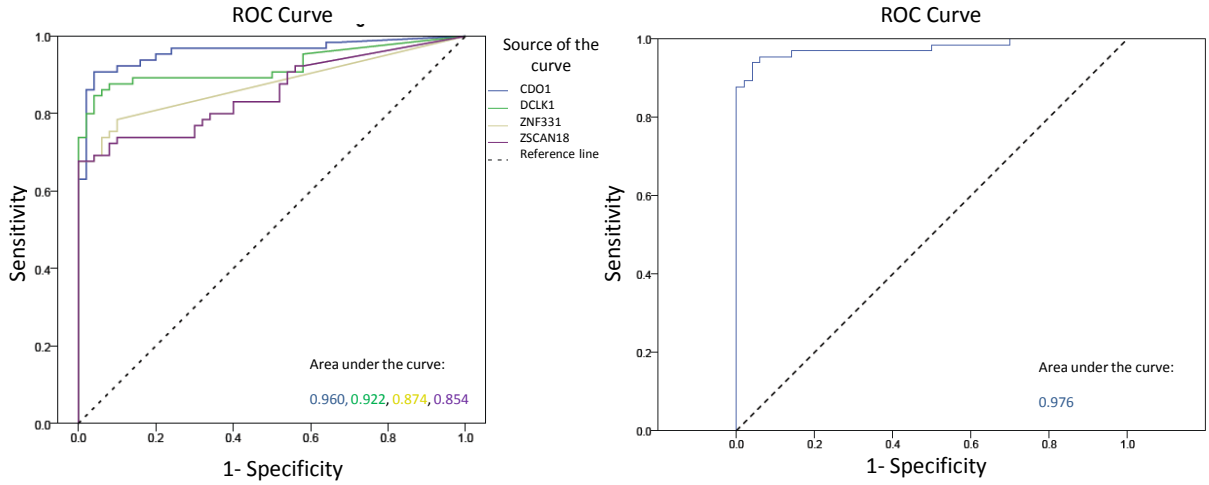


Figure 16. Receiver Operating Characteristics (ROC) curves for methylation of individual and combined biomarkers in colorectal cancer versus normal mucosa. The left figure depicts the area under the ROC curve (AUC) for the individual biomarkers, while the combined performance of the four biomarkers is depicted in the figure to the right. PMR values for cancer and normal samples are used as input.

5.5 Methylation frequencies in colorectal cancer versus cholangiocarcinoma

From the MSP analysis, the cholangiocarcinoma and colorectal cancer samples have a quite similar methylation profile, although the methylation frequencies for the individual genes are somewhat higher in the colorectal cancer samples (Table 2). Three genes (*CDO1*, *DCLK1* and *ZSCAN18*) were considered to have potential as biomarkers in both malignancies, and were subjected to quantitative methylation analysis (qMSP) also in cholangiocarcinomas (performed by Andresen, the PhD student). Two of the genes (*CDO1* and *DCLK1*) were also here, not surprisingly, more frequently methylated in colorectal cancer samples compared to the cholangiocarcinomas, with significantly higher mean PMR values. ZSCAN18 actually displayed a higher methylation frequency in cholangiocarcinomas, although the median- and mean PMR values were higher in colorectal cancer (not significant).

Table 2 Methylation frequencies of target genes in primary colorectal cancers and cholangiocarcinoma samples, assesed by quantitative MSP (qMSP).

	Methylation (%)	*P*- values	Range PMR values	Median PMR values	Mean PMR values	*P*- values
CDO1						
CRC	91% (59/65)	N.S	0-57.34	28.55	27.8	**$3.9*10^{-6}$**
CCA	85% (11/13)		0-23.68	4.99	8.68	
DCLK1						
CRC	80% (52/65)	**0.011**	0-60.23	22.04	23.25	**$1.9*10^{-10}$**
CCA	46% (6/13)		0-16.86	1.30	4.03	
ZSCAN18						
CRC	66% (43/65)	N.S	0-58.14	14.69	18.8	N.S
CCA	77% (10/13)		0-41.73	2.16	10.99	

Pearson Chi- Square and Fisher's Exact tests were used to calculate the *P*- values* in the first column of *P*- values. The second column of *P*-values[¥] was calculated by using Students T-test. Abbreviations; CCA, cholangiocarcinoma; CRC, colorectal cancer; PMR, percent methylated reference.

5.6 Associations of colorectal cancer methylation with genetic and clinico- pathological data

For the colorectal cancer sample series the following information was available: gender, age, tumor stage, location, MSI status, and *BRAF* mutation. The only significant association was seen between methylation and MSI status for *ZSCAN18*, where 92% (12/13) MSI tumors were methylation positive compared to 60% (31/52) MSS tumors ($P = 0.026$). A trend was seen between proximal location and methylation of *ZNF331* and *ZSCAN18* ($P= 0.063$), and also between the presence of *BRAF* mutation and methylation of *ZNF331*($P= 0.057$). Methylation of the individual genes was equally frequent across tumor stages, and in both sexes. Findings are summarized in Table 3. Cancer patients with promoter methylation of *CDO1* and *ZNF331* were slightly older (mean age 70 and 73, respectively) than that of cancer patients without *CDO1* and *ZNF331* methylation (59 and 64, respectively; $P=0.046$ and $P=0.012$).

Table 3 Promoter hypermethylation of biomarkers in colorectal carcinomas compared with clinico- pathological features and tumor phenotype.

	CDO1		*DCLK1*		*ZNF331*		*ZSCAN18*	
	M	**U**	**M**	**U**	**M**	**U**	**M**	**U**
Tumors								
No	59/65	6/65	52/65	13/65	43/65	22/65	43/65	22/65
Tumor phenotype								
MSI	12	1	10	3	11	2	12	1
MSS	47	5	42	10	32	20	31	21
P value	N.S		N.S		N.S		**0.026**	
BRAF ex 15								
Wild type	48	6	41	13	33	21	34	20
Mutation	11	0	11	0	10	1	9	2
P- value	N.S		N.S		0.057		N.S	
Sex								
Male	29	5	26	8	21	13	24	10
Female	30	1	26	5	22	9	19	12
P value	N.S		N.S		N.S		N.S	
Tumor site- 2 groups								
Proximal	28	1	23	6	23	6	23	6
Distal	30	5	28	7	19	16	19	16
P value	N.S		N.S		0.063		0.063	
Stage								
I	16	3	16	3	14	5	11	8
II	22	1	19	4	14	9	15	8
III	15	1	12	4	11	50	12	4
IV	6	1	5	2	4	3	5	2
P value	N.S		N.S		N.S		N.S	

Pearson Chi- Square and Fisher's Exact tests were used to calculate *P*- values.

5.7 Qualitative MSP vs. quantitative real-time MSP (qMSP)

When comparing results obtained from the qMSP and the MSP analysis from the 20 primary colorectal cancers analyzed by both methods (Table 4), full concordance (20/20; $P=7.74*10^{-6}$) was observed for three of the genes (*CDO1, DCLK1,* and *ZSCAN18*; based on the previously defined thresholds). For *ZNF331* one sample (AUS_002) was scored as methylated by MSP and as unmethylated from the qMSP analysis, resulting in an agreement in 19/20 samples ($P=1.7*10^{-4}$).

Table 4 Comparison of methylation statuses obtained from real-time qMSP and traditional qualitative MSP.

	CDO1		DCLK1		ZSCAN18		ZNF331	
	PMR=8	MSP	PMR=4	MSP	PMR=4	MSP	PMR=1	MSP
AUS2_001	32,70	M	30,37	M	0,07	U	6,92	M
AUS2_002	24,57	M	38,09	M	0,15	U	0,12	M
AUS2_003	36,84	M	19,76	M	29,36	M	8,01	M
AUS2_006	24,80	M	30,48	M	50,32	M	32,70	M
AUS2_007	57,34	M	60,23	M	52,64	M	17,23	M
AUS2_008	35,96	M	31,44	M	19,27	M	4,09	M
AUS2_009	21,08	M	26,12	M	0,38	U	10,16	M
AUS2_010	42,41	M	38,80	M	44,19	M	16,87	M
AUS2_011	29,47	M	12,50	M	6,38	M	10,59	M
AUS2_012	28,58	M	23,29	M	34,13	M	23,97	M
AUS2_015	23,50	M	20,18	M	26,66	M	20,77	M
AUS2_016	23,45	M	10,21	M	14,69	M	0,00	U
AUS2_017	24,35	M	22,04	M	0,16	U	0,07	U
AUS2_019	31,28	M	24,28	M	35,23	M	28,80	M
AUS2_030	23,21	M	21,93	M	26,95	M	22,14	M
AUS2_031	42,85	M	36,03	M	46,44	M	20,38	M
AUS2_033	0,00	U	0,10	U	0,21	U	0,00	U
AUS2_045	15,07	M	10,41	M	16,58	M	12,44	M
AUS2_066	39,75	M	38,20	M	43,66	M	39,61	M
AUS2_072	33,56	M	0,00	U	11,23	M	0,06	U

A high concordance is seen between the methylation statuses assessed from the two methods. Only one sample (marked in light grey) was scored differently between the two methods. Traditional MSP provides qualitative results (samples are scored as either unmethylated (U) or methylated (M)), whereas the PMR values for each sample is a quantitative measure.

6 Discussion

6.1 Methodological considerations

6.1.1 Treatment of cancer cell lines with 5-aza- 2'deoxycytodine (AZA) and trichostatin A (TSA)

AZA/TSA treatment of cancer cell lines followed by microarray gene expression analysis was used as the first step in the experimental procedure to identify candidate genes inactivated by epigenetic mechanisms in cancer. AZA is a cytosine analog which is incorporated into newly synthesized DNA. Once incorporated, AZA can covalently bind and deplete the cell of DNA methyltransferases [113], resulting in passive demethylation of dividing cells. TSA reversibly inhibits the mammalian histone deacetylase, thereby preventing the removal of acetyl groups from the histone tails. Since acetylation promotes an open chromatin structure, gene expression is favored. As mentioned earlier, combining the two drugs has been shown to have a synergic effect. Gene activation following AZA/TSA treatment indicates that DNA methylation has a functional role in transcription repression. However, some genes might be activated even though no CpG island is present, and it is possible that the increased gene expression is caused by *i.e*. histone (de)acetylation, cytotoxic effects of the drugs, or a secondary response. These candidates will represent false positive methylation targets. Moreover, genes with CpG island methylation may fail to be reactivated following drug treatment, leading to false negatives. In our approach, the candidate gene list from the epigenetically treated cancer cell lines is compared with gene expression analyses of normal versus tumor samples. By choosing only those genes which are upregulated in at least four of six cancer cell lines after epigenetic treatment and simultaneously dowregulated in tumor cells, the chance of selecting true epigenetically deregulated genes increases. Finally, a search for a CpG island in the promoter region of the candidates is conducted. Combining this step-wise experimental procedure with strict selection criteria, limits the number of false positives which in general is a challenge using (epi)genome wide approaches. However, of the 40 genes tested by MSP in cancer cell lines, 17 genes were unmethylated in colon cancer cell lines, and 15 in cholangiocarcinoma cell lines, indicating that we could have used even stricter selection criteria. On the other hand, stricter criteria could result in the loss of potential biomarkers.

6.1.2 Cancer cell lines as *in vitro* models for research

In the present project we have used cancer cell lines as *in vitro* models in the search for candidate DNA methylated genes. These are permanently established cell cultures, grown under controlled conditions, which proliferate indefinitely. Although they resemble their tissue of origin, and are invaluable tools in cancer research, the relevance of cell lines for the *in vivo* situation has been debated. Some studies have reported a higher prevalence of CpG island promoter methylation in cancer cell lines compared to the primary malignancies they represent [114], while others report a similar overall distribution [115,116]. In this study we observed that the methylation frequencies obtained from the cell line analysis was a good indication for the methylation status in cancer tissue samples.

A potential problem when working with cultured cells is cross contamination, which naturally negatively influences the quality of the research. Passing cells over numerous generations may also give rise to different subtypes, *i.e.* faster growing cells which eventually outnumber the others, and the new cell line population may possess different genetic and/or epigenetic characteristics, not representative for the original tumor. Authentication is therefore extremely important, and this is routinely done in our lab. All cell lines are also routinely checked for mycoplasma infection.

6.1.3 Qualitative MSP versus quantitative real- time MSP (qMSP)

Traditional MSP, which measures the amount of end product after amplification, was applied for initial screening of the candidate genes. The samples are separated by gel electrophoresis and scored as either methylated or unmethylated based on the intensity of the gel bands compared to positive controls. Although a qualitative method, the intensity of the gel bands can to some degree be used to indicate the amount of starting material, thus giving semi-quantitative results. The method is sensitive, easy to perform, and cost-effective, but the quality of the results is highly dependent on the quality of the primers used. Inclusion of positive and various negative controls, ensures that potentially false positives or negatives generated by the lack of primer specificity are detected. In this study, primer sets with unspecific results were redesigned. In cases where the performance did not improve, the primer pairs were not included in the subsequent DNA methylation analysis. One exception was made for *SFRP1*, where the designed primers displayed a weak crossbinding (methylated primers amplified bisulfite treated NB). Normally, we would have excluded this gene for

further analysis, but since the gene is known from the literature, and also known to have some specificity issues [117], it was included.

Although a well established method, the gel electrophoresis, manually scoring of the MSP as well as the dependence on primer design may contribute to reproducibility problems across laboratories, and the method is therefore not the first choice when it comes to routine clinical testing [118]. Using a quantitative method such as real- time MSP (qMSP or MethyLight), which can easily be standardized and accurately determine the relative prevalence of a particular pattern of DNA methylation [119] increases the reproducibility. qMSP measures the amount of methylation in the exponential phase, where the intensity of the signal is directly proportional to amount of template. To normalize for input DNA, ALU-C4 is used as a reference. ALU-C4 is a short interspersed nucleotide element present in ~1million copies per haploid genome, and copy number changes caused by *e.g.* chromosomal alterations will not influence its distribution remarkably as opposed to single copy genes [112]. ALU-C4 can therefore be used as a universal reference also across various cancer types.

The **sensitivity** of a test/biomarker is a measure for the proportion of people who have the disease who actually test positive for it, while the **specificity** refers to the proportion of healthy people who actually test negative for the disease.

In this thesis the qMSP threshold values for each gene were set according to the highest PMR values for the normal mucosa samples, aiming for the highest possible specificity. A high specificity is important to ensure a low number of false positives in a future non- invasive test (see page 63, Early detection biomarkers for colorectal cancer, for more details about the topic). Both the sensitivity and specificity depend on the thresholds. A high threshold will increase the specificity but the sensitivity may drop, and lowering the threshold could potentially decrease the specificity, but the sensitivity will increase. For *CDO1*, two of the normal samples had outlier PMR values, and if these had been included to ensure a specificity of 100%, the sensitivity would have dropped considerably. We therefore decided to disregard these samples when setting the threshold at 8, resulting in a sensitivity and specificity of 91% and 96%, respectively. One normal sample also displayed a higher PMR value for *DCLK1* than the rest of the normal mucosa samples, and as for *CDO1* we decided to disregard this

sample when setting the threshold at 4, resulting in a sensitivity of 80% and a specificity of 98%. Interestingly, the sample with outlier PMR for *DCLK1* was one of samples with outlier PMR for *CDO1*. This sample has also in previous analyses from our lab shown relatively high methylation frequencies across several genes. These analyzed normal controls are autopsy material from colorectal cancer free individuals and we cannot exclude that they have other types of cancer (*i.e*. hematological cancer) that could influence the results.

When the results from the MSP and qMSP assays were compared, an almost perfect association was seen across the tumor samples, except for a single sample which was scored as methylated from the MSP analysis, but as unmethylated from qMSP. Although the qMSP is a highly sensitive method, the inclusion of a probe makes it more "conservative" than traditional MSP, which may explain this discrepancy. Nevertheless, the high overall concordance reflects the sensitivity and robustness of traditional MSP, and also the quality of the MSP and the conservative scorings conducted in our lab.

6.1.4 Bisulfite sequencing

The amplified PCR product can be sequenced directly, as was done in this project. This gives an average methylation value for each CpG site based on the methylation status across all the cells in the analyzed sample. Alternatively, the PCR products can be cloned into plasmid vectors and individually subjected to sequencing. When sequencing tumor samples (containing a hetrogenous cell population, including normal cells), cloning before sequencing should be preferred over direct sequencing due to technical challenges. This is because the normal cells are expected to contain unmethylated alleles, which will have a different molecular weight than the methylated alleles and lead to "messy" electropherograms that are hard to interpret. Cloning will solve this problem, and thereby give a more detailed methylation profile. It is also often difficult to assess methylation levels lower than 15%-20% when using direct sequencing, due to the variable sequencing background signal. Although cloning is the gold standard for sequencing, it is labor intensive as it requires 10- 12 clones per gene to be able to say something certain about the methylation frequency. In this project, cancer cell lines were used as template (which represent close to monoclonal cell populations), and direct bisulfite sequencing was therefore used. The overall goal was 1) to verify the methylation status as assessed by MSP (since this method can be prone to false positives), 2) to confirm the success of bisulfite conversion and 3) to use the results to guide

the design of qMSP primers. The methylation status of various CpG sites within a CpG island may be heterogeneous, and bisulfite sequencing is a valuable tool to ensure that a representative region of the promoter is selected for MSP and/or qMSP analyses. This is exemplified by Mori *et al* [120], and Lind *et al* [93,94], who reported hypermethylation of the *MAL* gene in 6% and 80% of colorectal carcinomas, respectively. The former study analyzed a region located approximately 200bp upstream of the transcription start site, while the latter analyzed a region close to the transcription start site. By bisulfite sequencing both regions Lind *et al* showed that the CpG sites covered by the Mori primer set underestimated the promoter methylation status of *MAL* since the assay did not cover a representative region of the promoter.

For bisulfite sequencing purposes, primers are designed to avoid CpG sites and optimized to avoid the preferential amplification of either methylated or unmethylated fragments. Despite of this, we observed that the primers used to amplify *ZNF331*, amplified normal blood (unmethylated) at a higher efficiency than IVD (methylated). Four primer pairs were tested, all with the same outcome. After bisulfite treatment methylated DNA will contain a higher G+C content than unmethylated sequences. It is possible that one population of sequences may amplify preferentially, giving rise to PCR bias and an inaccurate estimate of methylation [121]. Possibly, secondary structures in the methylated template could be competing with primer binding, resulting in the observed preference for the unmethylated template. Since the two strands are no longer complimentary after bisulfite treatment, the chemistry between them may vary considerable depending on the amount of cytosines in non- CpG sites. One possibility could therefore be to design primers to the opposite strand. However, the observed preference for unmethylated template is not necessarily a big problem because 1) the sequencing results are only semi- quantitative, and 2) although the CpG site methylation displayed in the electropherogram is measured accurately, giving values ranging from 0 to 1, the final results are divided into three broad categories. This minimizes the risk of mis-interpreting the data.

6.2 Epigenetic changes in colorectal cancer and cholangiocarcinoma: differences and similarities?

Cholangiocarcinoma is an uncommon disease, while colorectal cancer is the third most common malignancy word wide. Naturally, more research has been done on colorectal cancer,

where a great amount of hypermethylated genes have been identified. Even though cholangiocarcinoma first lately has gained increased research focus, and much fewer genes have been subjected to DNA methylation analysis, several genes, including *APC*, *CDKN2A/p16*, *MGMT*, *MLH1*,and *p73*, and have been found to harbor promoter DNA hypermethylation with similar frequencies in both cholangiocarcinoma and colorectal cancer [81,89,96,122].

Table 5 Genes methylated in both cholangiocarcinoma and colorectal cancer with comparable frequencies.

	Methylation (%)	
	Cholangiocarcinoma	**Colorectal cancer**
APC	27-46	21-28 (36)
CDKN2A/p14	25-38	12-88 (39-86)
CDKN2A/p16	18-83	28-63 (36-64)
MGMT	33-49	40-53 (39-43)
MLH1	8-46	0 (39)
p73	36-49	63
RASSF1A	27-69	21-81
RUNX3	57	34

Methylation (%) obtained from[89] (colorectal cancer) and [81,96,122] (cholangiocarcinoma).Values in parenthesis include MSI tumors.

The decision to test candidate genes from the cholangiocarcinoma approach also in colorectal cancers was based on similar observations in our group, where six genes were found to be methylated with comparable frequencies across cell lines from various cancers arising in the gastrointestinal tract [95]. The qualitative methylation analyses (MSP) performed in the present project clearly support this observation, although colorectal cancer tissue samples displayed a somewhat higher methylation frequency than cholangiocarcinoma. When comparing the results obtained for the quantitative analysis (qMSP), *ZSCAN18* were more frequently methylated in cholangiocarcinoma tissue samples (77% compared to 66% of the colorectal cancer samples). This could be due to variations introduced by the small number of cholangiocarcinomas (n=13) included in the present study, or the lower threshold sat for this malignancy. The individual PMR values for all the three genes analyzed by qMSP in both malignancies (*CDO1*, *DCLK1* and *ZSCAN18*) were considerably higher in the colorectal cancer tissues.

Even though the function of the gene is not important when considering its suitability as a biomarker, it is interesting from a biological perspective. *CDO1* (Cysteine dioxygenase type 1), which was found to be the most frequently methylated gene in both colorectal cancer and

cholangiocarcinoma, is a mammalian non-heme iron enzyme that initiates a number of significant metabolic pathways associated with pyruvate and sulfurate compounds including taurine[14]. Taurine is involved in several crucial physiological processes, such as cardiovascular function, and development and function of skeletal muscle, and it is also a major constituent of bile. *CDO1* is highly expressed in the liver, where it has an essential role in maintaining the hepatic concentration of intracellular free cysteine within a proper narrow range.

Promoter methylation of *CDO1* has been shown to be a strong predictor for distant metastasis in estrogen receptor- and lymph node positive breast cancer patients [123]. *CDO1* has further been suggested to be epigenetically silenced in the colon cancer cell lines HCT116 and SW480, and to be reactivated after AZA treatment [124]. It has also been hypothesized that epigenetic silencing or deletion of the chromosomal region where *CDO1* is located contributes to colorectal tumorigenesis [125]. These studies support our findings, in that methylation of *CDO1* is epigenetically regulated and may play a role in colorectal cancer development.

Promoter methylation of *DCLK1* (Doublecortin-like kinase 1), which is a serine-threonine kinase primarily expressed in the central nervous system, was found to be significantly more prevalent in colorectal cancer- compared to cholangiocarcinoma samples. The protein encoded by the gene contains two N- terminal doublecortin domains, which bind microtubules and regulate microtubule polymerization[15]. The *DCLK1*-derived microtubule-associated proteins (MAPs) are crucial for proliferation and survival of neuroblasts and are highly expressed in neuroblastomas as well as in gliomas [126]. In this study, DCLK1 was identified as hypermethylated in both colon cancer and cholangiocarcinoma.

Both *ZNF331* (zink finger protein 331) and *ZSCAN18* (zink finger and SCAN domain containing 18) belong to the krueppel C_2H_2- type zink finger protein family[16], which have been shown to interact with nucleic acids and to have diverse functions, including a transcription factor role. Hypermethylation of *ZSCAN18* has previously been reported in 32% of primary renal cell carcinoma [127]. In the same study, RNAi knockdown of the *ZSCAN18* transcript resulted in an anchorage- independent growth advantage. *ZNF331* was originally

[14] http://smd.stanford.edu/
[15] http://www.ncbi.nlm.nih.gov/gene/9201
[16] http://smd.stanford.edu/

referred to as RITA (rearranged in thyroid adenomas), after it was discovered as a candidate gene in the 19q13 breakpoint found in follicular adenomas (OMIM#606043). The gene has lately been identified as an imprinted gene [128].

6.3 Early detection biomarkers for colorectal tumors

Colorectal cancer is one of the most prevalent causes of cancer deaths worldwide, with an estimated 1500 deaths each year in Norway alone. If the cancer is diagnosed and treated at the localized stage (I), surgery alone could result in complete remission in the majority of cases. However, today approximately half of the tumors are detected after spreading to the lymph nodes (stage III) or distant organs (stage IV) where the five year relative survival has decreased to 63% and 9%, respectively. Early detection of colorectal cancer as well as high risk adenomas may therefore significantly reduce the number of colorectal cancer associated deaths. Colorectal cancer is especially suitable for early detection since it is developed through defined precursor lesions and the development takes from 5-35 years.

A reduced mortality is seen in randomized clinical trials where sigmoidoscopy or fecal occult blood testing (FOBT) have been used as screening tools [129]. In the USA, increased colonoscopy among the general population is thought to explain the 30% decrease in colorectal cancer death rates noted between 1990 and 2006, as well as the reduced incidence. However, although colonoscopy is currently the most reliable screening tool for colorectal cancer, detecting more than 95% of the carcinomas and polyps, it is invasive, costly, provides a minor risk for complications, and is further hampered by low compliance [130]. FOBT screening has been shown to reduce colorectal cancer associated deaths with approximately 16% [129]. It is a non- invasive, easy and inexpensive test, but it has the limitation of a low sensitivity and specificity (generally between 30% - 80% and 91- 98%, respectively), especially for advanced adenoma (11%-56% sensitivity) [129]. The need for biannual testing has also been shown to decrease patient compliance over time and thus the effect of screening. It is therefore of great interest to identify biomarkers that can be used in a reliable non-invasive test for early detection of colorectal tumors.

Promoter DNA methylation has great potential as diagnostic indicators for disease risk. It often precedes morphological aberrations and has been shown to be an early event in tumor progression. DNA methylation is further common in the majority of cancer types, and its

stability makes it suitable for clinical testing [28]. Furthermore, only one test is required per gene to establish the methylation status as opposed to mutational analysis where several possible inactivation mutations need to be analyzed per gene [118].

The **positive predictive value (PPV)** refers to the proportion of patients with a positive test who in fact has the underlying condition that is tested for. The **negative predictive value (NPV)** refers to the proportion of individuals with a negative test who do not have the disease.

The sensitivity and specificity are two important measurements for evaluating the performance of a biomarker. Another frequently used measurement is the positive predictive value (PPV). In contrast to the sensitivity and specificity measurements, the PPV takes the disease prevalence into account, and a high value means that few people will be misdiagnosed (low false positive rate), which is important in a population based screening program where the majority of individuals are expected to be healthy. This is clearly demonstrated by Feng and Longton, and Mor, who calculated the PPV of a short lived test called OvaSure to be 7% and 99%, respectively, based on different assumptions about the prevalence of ovarian cancer [131]. When screening for a serious disease in a high-risk group, the threshold for a positive test should be chosen in such a way as to provide good sensitivity, even if the false positive rate is high, because the benefit of detecting the disease outweigh the possible high cost/ worries caused by following up the false positives. On the other hand, when screening for a disease whose prevalence is low, a high specificity and low false positive rate is required. If the PPV is low in such a group, almost all of the positive cases will be false positives, resulting in many unnecessary, risky follow-up examinations and/or treatments.

A high degree of false positives is a general problem with new biomarkers. For colorectal cancer, where the age- adjusted incidence in Norway is approximately 25 per 100.000 people per year, a false negative rate of 12%, which is reported for Cologuard will give a low positive predictive value, meaning that a great proportion of healthy individuals will be wrongly diagnosed. These people may suffer unnecessary worries going through a colonoscopy, which is also very expensive. The Epi *pro*Colon® 2.0 CE has a reported PPV of 45%, which is pretty good, and which will greatly reduce the proportion of individuals sent to colonoscopy compared to colonoscopy as the main screening method. However, the test is

still going through clinical trials, and with a sensitivity and specificity of 80% and 99%, respectively, there is still room for improvements.

In addition to *SEPT9*, included in the Epi *pro*Colon, *VIM* is as mentioned earlier a promising biomarkers for early detection, but the neither sensitivity nor specificity is optimal (72%-77% and 83%-94%, respectively). At this point they are not good enough to replace colonoscopy as a screening test for high risk persons. The test for *VIM* is based on qualitative MSP, which is a good method in a research setting, but which may be prone to reproducibility problems in routine clinical testing laboratories [118].

A **Receiver Operating Characteristic (ROC) curve** is a statistical method to assess the diagnostic accuracy presented as a graphical plot of the sensitivity versus the 1- specificity (the true positive rate vs. the false positives rate), by calculating the **area under the curve (AUC).**

[132]

The marker panel recently published by Lind *et al* [95] is promising for early detection of colorectal tumors. The combined biomarker panel has a high performance (AUC= 0.984 for cancers and 0.968 for adenomas compared with normal tissue) and provides more robustness than a single marker when considering a future non- invasive test. Multiple parallel pathways to colorectal cancer development have been suggested, giving rise to highly different subtypes (*i.e.* MSI, CIN and CIMP tumors). Detecting all these "tumor variants" in a screening setting is a challenge, but a combination of markers, as Lind suggests, would most likely provide a higher sensitivity and specificity for early detection. Indeed, this panel is positive in tumors independent of the patient's gender and age, as well as of tumor stage, phenotype and location. At this point the sensitivity and specificity measurements from non-invasive sample material are not completed, but the initial methylation frequencies reported for four of the genes in tissue samples are higher than what was reported for *VIM* [95]. Interestingly, four of the cancer samples that were negative for the original six-marker panel from Linds analyses were methylated for one or more of the markers identified in this thesis, *CDO1*, *DCLK1*, *ZSCAN18*, and *ZNF331*. The novel biomarkers were recently included in a patent

application[17] and we are currently collaborating with an industrial partner to develop a non-invasive test for early detection of colorectal cancer based on a combined biomarker panel. A blood- or fecal based test may lead to higher compliance in colorectal cancer screening, and thus earlier detection of the disease which further reduce the overall cancer burden and mortality rates.

6.4 Biomarkers for cholangiocarcinoma

Early detection of cholangiocarcinoma is a major clinical challenge, especially in patients with primary sclerosing cholangitis (PSC), and the tumors are often detected too late to give the patient a therapeutic benefit [80]. Currently, the diagnosis is based on imaging analyses, brush cytology, histology and tumor markers in serum, but none of these techniques are optimal when it comes to sensitivity and specificity. CEA and cancer antigen 19-9, which may be used, have poor sensitivity and specificity and no tumor-specific markers have so far been identified for cholangiocarcinoma [83]. The identification of epigenetic biomarkers for early detection of cholangiocarcinoma can improve the diagnostic accuracy and thus increase the survival. Bile, blood and possibly fecal samples are potential additional sources of DNA for non-invasive testing.

In this study, four genes were found to be frequently methylated in cholangiocarcinoma tissue samples (*CDO1*, *DCLK1*, *SFRP1,* and *ZSCAN18*). The same genes have been subjected to qMSP in a larger samples series containing both fresh frozen- (tumors, n=13; cancer free samples, n=21) and formalin- fixed, paraffin- embedded (FFPE) archival samples (carcinomas, n=26; cancer free samples, n=33) by another group member (Andresen). The AUC for the combined marker panel was found to be 0.996 for the fresh frozen sample set, and 0.904 for the archival material (Andresen *et al*, unpublished). If this panel can be detected by *e.g.* brush cytology, which is a minimal invasive procedure, it could help produce a more accurate diagnosis, hopefully at an earlier stage. The biomarkers are negative in normal tissue as well as tissue samples from PSC patients, which makes them promising in diagnosing cholangiocarcinoma complicated with PSC.

[17] *Lothe RA, Ahmed D, Andresen K, Skotheim R, and Lind GE. Methods and biomarkers for detection of gastrointestinal cancers. US provisional application filed 61/451,198, INVEN-31899/US-1/PRO.*

7 Conclusions

By using cholangiocarcinoma cell lines, patient material and a step-wise epigenomic approach, four genes *CDO1*, *DCLK1, SFRP1,* and *ZSCAN18* were identified as methylated targets and potential biomarkers for cholangiocarcinomas. Genes from this experimental strategy were further analyzed in colorectal cancers and normal mucosa samples, and *CDO1*, *DCLK1*, *ZNF331* and *ZSCAN18* were identified as promising biomarkers for early detection of this disease. A patent application has been filed based on the present results, and we are currently in collaboration with an industrial partner to develop a non- invasive test for early detection of colorectal cancer.

8 Future and ongoing perspectives

The results obtained in the present study will be verified in a large clinically independent sample series, including also adenomas. At the moment the four markers identified in the present project (*CDO1*, *DCLK1*, *ZNF331*, and *ZSCAN18*), in addition to the six biomarkers previously identified in our lab (*CNRIP1*, *FBN1*, *INA*, *SNCA*, *MAL,* and *SPG20*) are undergoing DNA methylation analyses in a pilot series of blood (plasma and serum) samples from patients with colorectal cancer, people with adenoma, and healthy individuals.

The four genes (*CDO1*, *DCLK1*, *ZSCAN18,* and *SFRP1*) identified as frequently methylated in cholangiocarcinoma samples are currently being analyzed in biliary brush samples from cholangiocarcinoma patients and PSC patients in order to see if they are suitable for minimally invasive early detection of cholangiocarcinoma.

Finally, it has been suggested that as much as 20%-30% of all colorectal cancers may have a heritably cause, but only approximately 5% have a known hereditary syndrome [61]. Epimutations, either constitutional or germline, may explain some of these cases, where no underlying sequence variant has been identified. By using global DNA methylation analysis (arrays from Illumina), tumors and blood samples from early- and late- onset patients will be analyzed. The aim of the study is to identify novel constitutional epimutations among the early-onset patient group as well as to identify epigenetic differences between late onset and early onset colorectal cancer in general. Potential novel epimutations will be subjected to functional analyses.

Reference List

1. Stratton MR, Campbell PJ, Futreal PA: **The cancer genome.** *Nature* 2009, **458:** 719-724.

2. William M: **Cancer, Overview.** In *Encyclopedia of Gastroenterology*. Edited by Editor-in-Chief:-á-á Leonard Johnson. New York: Elsevier; 2004:256-264.

3. Hanahan D, Weinberg RA: **The hallmarks of cancer.** *Cell* 2000, **100:** 57-70.

4. Hanahan D, Weinberg RA: **Hallmarks of Cancer: The Next Generation.** *Cell* 2011, 646-674.

5. Stratton MR: **Exploring the genomes of cancer cells: progress and promise.** *Science* 2011, **331:** 1553-1558.

6. Grady WM, Carethers JM: **Genomic and epigenetic instability in colorectal cancer pathogenesis.** *Gastroenterology* 2008, **135:** 1079-1099.

7. Sharma S, Kelly TK, Jones PA: **Epigenetics in cancer.** *Carcinogenesis* 2010, **31:** 27-36.

8. Waddington C. The Epigenotype. 1942.
Ref Type: Generic

9. **Making a mark.** *Nat Biotechnol* 2010, **28:** 1031.

10. Portela A, Esteller M: **Epigenetic modifications and human disease.** *Nat Biotechnol* 2010, **28:** 1057-1068.

11. Sawan C, Vaissiere T, Murr R, Herceg Z: **Epigenetic drivers and genetic passengers on the road to cancer.** *Mutat Res* 2008, **642:** 1-13.

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

13. Jones PA, Baylin SB: **The epigenomics of cancer.** *Cell* 2007, **128:** 683-692.

14. Issa JP: **CpG island methylator phenotype in cancer.** *Nat Rev Cancer* 2004, **4:** 988-993.

15. Jones PA, Baylin SB: **The fundamental role of epigenetic events in cancer.** *Nat Rev Genet* 2002, **3:** 415-428.

16. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP: **CpG island methylator phenotype in colorectal cancer.** *Proc Natl Acad Sci U S A* 1999, **96:** 8681-8686.

17. Jones PA, Liang G: **Rethinking how DNA methylation patterns are maintained.** *Nat Rev Genet* 2009, **10:** 805-811.

18. Wu SC, Zhang Y: **Active DNA demethylation: many roads lead to Rome.** *Nat Rev Mol Cell Biol* 2010, **11:** 607-620.

19. Gehring M, Reik W, Henikoff S: **DNA demethylation by DNA repair.** *Trends Genet* 2009, **25:** 82-90.

20. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y *et al.*: **Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.** *Science* 2009, **324:** 930-935.

21. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q *et al.*: **Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA.** *Science* 2011, **333:** 1303-1307.

22. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA *et al.*: **Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.** *Science* 2011, **333:** 1300-1303.

23. Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N *et al.*: **Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription.** *Nat Genet* 1998, **19:** 187-191.

24. Carmona FJ, Esteller M: **Moving closer to a prognostic DNA methylation signature in colon cancer.** *Clin Cancer Res* 2011, **17:** 1215-1217.

25. Estecio MR, Issa JP: **Dissecting DNA hypermethylation in cancer.** *FEBS Lett* 2010.

26. Feinberg AP, Tycko B: **The history of cancer epigenetics.** *Nat Rev Cancer* 2004, **4:** 143-153.

27. van EM, Derks S, Smits KM, Meijer GA, Herman JG: **Colorectal Cancer Epigenetics: Complex Simplicity.** *J Clin Oncol* 2011.

28. Herman JG, Baylin SB: **Gene silencing in cancer in association with promoter hypermethylation.** *N Engl J Med* 2003, **349:** 2042-2054.

29. Kondo Y, Issa JP: **Epigenetic changes in colorectal cancer.** *Cancer Metastasis Rev* 2004, **23:** 29-39.

30. Brun ME, Lana E, Rivals I, Lefranc G, Sarda P, Claustres M *et al.*: **Heterochromatic genes undergo epigenetic changes and escape silencing in immunodeficiency, centromeric instability, facial anomalies (ICF) syndrome.** *PLoS One* 2011, **6:** e19464.

31. Chik F, Szyf M, Rabbani SA: **Role of epigenetics in cancer initiation and progression.** *Adv Exp Med Biol* 2011, **720:** 91-104.

32. Karpf AR, Matsui S: **Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells.** *Cancer Res* 2005, **65:** 8635-8639.

33. Feinberg AP, Vogelstein B: **Hypomethylation of ras oncogenes in primary human cancers.** *Biochem Biophys Res Commun* 1983, **111:** 47-54.

34. Zendman AJ, Ruiter DJ, Van Muijen GN: **Cancer/testis-associated genes: identification, expression profile, and putative function.** *J Cell Physiol* 2003, **194:** 272-288.

35. Cui H, Cruz-Correa M, Giardiello FM, Hutcheon DF, Kafonek DR, Brandenburg S *et al*.: **Loss of IGF2 imprinting: a potential marker of colorectal cancer risk.** *Science* 2003, **299:** 1753-1755.

36. Ushijima T, Asada K: **Aberrant DNA methylation in contrast with mutations.** *Cancer Sci* 2010, **101:** 300-305.

37. Dobrovic A, Kristensen LS: **DNA methylation, epimutations and cancer predisposition.** *Int J Biochem Cell Biol* 2009, **41:** 34-39.

38. Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, Liu S *et al*.: **Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma.** *Proc Natl Acad Sci U S A* 1994, **91:** 9700-9704.

39. Ahlquist T, Lind GE, Costa VL, Meling GI, Vatn M, Hoff GS *et al*.: **Gene methylation profiles of normal mucosa, and benign and malignant colorectal tumors identify early onset markers.** *Mol Cancer* 2008, **7:** 94.

40. Esteller M, Corn PG, Baylin SB, Herman JG: **A gene hypermethylation profile of human cancer.** *Cancer Res* 2001, **61:** 3225-3229.

41. Baylin SB, Ohm JE: **Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction?** *Nat Rev Cancer* 2006, **6:** 107-116.

42. Lao VV, Grady WM: **Epigenetics and colorectal cancer.** *Nat Rev Gastroenterol Hepatol* 2011.

43. Harikrishnan KN, Chow MZ, Baker EK, Pal S, Bassal S, Brasacchio D *et al*.: **Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing.** *Nat Genet* 2005, **37:** 254-264.

44. Shen L, Kondo Y, Rosner GL, Xiao L, Hernandez NS, Vilaythong J *et al*.: **MGMT promoter methylation and field defect in sporadic colorectal cancer.** *J Natl Cancer Inst* 2005, **97:** 1330-1338.

45. Felsenfeld G, Groudine M: **Controlling the double helix.** *Nature* 2003, **421:** 448-453.

46. Iacobuzio-Donahue CA: **Epigenetic changes in cancer.** *Annu Rev Pathol* 2009, **4:** 229-249.

47. Kamakaka RT, Biggins S: **Histone variants: deviants?** *Genes Dev* 2005, **19:** 295-310.

48. Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L *et al*.: **A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing.** *Nat Genet* 2007, **39:** 237-242.

49. Li E: **Chromatin modification and epigenetic reprogramming in mammalian development.** *Nat Rev Genet* 2002, **3:** 662-673.

50. Cameron EE, Bachman KE, Myohanen S, Herman JG, Baylin SB: **Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer.** *Nat Genet* 1999, **21:** 103-107.

51. Toyota M, Suzuki H: **Epigenetic drivers of genetic alterations.** *Adv Genet* 2010, **70:** 309-323.

52. Baylin SB, Jones PA: **A decade of exploring the cancer epigenome - biological and translational implications.** *Nat Rev Cancer* 2011, **11:** 726-734.

53. Jones S, Wang TL, Shih I, Mao TL, Nakayama K, Roden R *et al*.: **Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma.** *Science* 2010, **330:** 228-231.

54. Jiao Y, Shi C, Edil BH, de Wilde RF, Klimstra DS, Maitra A *et al*.: **DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors.** *Science* 2011, **331:** 1199-1203.

55. Suter CM, Martin DI, Ward RL: **Germline epimutation of MLH1 in individuals with multiple cancers.** *Nat Genet* 2004, **36:** 497-501.

56. Hitchins MP, Rapkins RW, Kwok CT, Srivastava S, Wong JJ, Khachigian LM *et al*.: **Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR.** *Cancer Cell* 2011, **20:** 200-213.

57. Chang S, Wang RH, Akagi K, Kim KA, Martin BK, Cavallone L *et al*.: **Tumor suppressor BRCA1 epigenetically controls oncogenic microRNA-155.** *Nat Med* 2011, **17:** 1275-1282.

58. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM: **Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008.** *Int J Cancer* 2010, **127:** 2893-2917.

59. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D: **Global cancer statistics.** *CA Cancer J Clin* 2011, **61:** 69-90.

60. Kinzler KW, Vogelstein B: **Lessons from hereditary colorectal cancer.** *Cell* 1996, **87:** 159-170.

61. Jasperson KW, Tuohy TM, Neklason DW, Burt RW: **Hereditary and familial colon cancer.** *Gastroenterology* 2010, **138:** 2044-2058.

62. Chan TL, Yuen ST, Kong CK, Chan YW, Chan AS, Ng WF *et al*.: **Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer.** *Nat Genet* 2006, **38:** 1178-1183.

63. Hitchins MP, Wong JJ, Suthers G, Suter CM, Martin DI, Hawkins NJ *et al*.: **Inheritance of a cancer-associated MLH1 germ-line epimutation.** *N Engl J Med* 2007, **356:** 697-705.

64. Fearon ER, Vogelstein B: **A genetic model for colorectal tumorigenesis.** *Cell* 1990, **61:** 759-767.

65. Muto T, Bussey HJ, Morson BC: **The evolution of cancer of the colon and rectum.** *Cancer* 1975, **36:** 2251-2270.

66. Grady WM, Markowitz SD: **Genetic and epigenetic alterations in colon cancer.** *Annu Rev Genomics Hum Genet* 2002, **3:** 101-128.

67. Leggett B, Whitehall V: **Role of the serrated pathway in colorectal cancer pathogenesis.** *Gastroenterology* 2010, **138:** 2088-2100.

68. Schreiner MA, Weiss DG, Lieberman DA: **Proximal and large hyperplastic and nondysplastic serrated polyps detected by colonoscopy are associated with neoplasia.** *Gastroenterology* 2010, **139:** 1497-1502.

69. Terdiman JP, McQuaid KR: **Surveillance guidelines should be updated to recognize the importance of serrated polyps.** *Gastroenterology* 2010, **139:** 1444-1447.

70. Issa JP: **Colon cancer: it's CIN or CIMP.** *Clin Cancer Res* 2008, **14:** 5939-5940.

71. O'Brien MJ, Yang S, Mack C, Xu H, Huang CS, Mulcahy E *et al*.: **Comparison of microsatellite instability, CpG island methylation phenotype, BRAF and KRAS status in serrated polyps and traditional adenomas indicates separate pathways to distinct colorectal carcinoma end points.** *Am J Surg Pathol* 2006, **30:** 1491-1501.

72. Ahnen DJ: **The American College of Gastroenterology Emily Couric Lecture--the adenoma-carcinoma sequence revisited: has the era of genetic tailoring finally arrived?** *Am J Gastroenterol* 2011, **106:** 190-198.

73. Thompson SL, Bakhoum SF, Compton DA: **Mechanisms of chromosomal instability.** *Curr Biol* 2010, **20:** R285-R295.

74. Popat S, Hubner R, Houlston RS: **Systematic review of microsatellite instability and colorectal cancer prognosis.** *J Clin Oncol* 2005, **23:** 609-618.

75. Matsuzaki K, Deng G, Tanaka H, Kakar S, Miura S, Kim YS: **The relationship between global methylation level, loss of heterozygosity, and microsatellite instability in sporadic colorectal cancer.** *Clin Cancer Res* 2005, **11:** 8564-8569.

76. Kim JH, Shin SH, Kwon HJ, Cho NY, Kang GH: **Prognostic implications of CpG island hypermethylator phenotype in colorectal cancers.** *Virchows Arch* 2009, **455:** 485-494.

77. Samowitz WS, Albertsen H, Herrick J, Levin TR, Sweeney C, Murtaugh MA *et al*.: **Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer.** *Gastroenterology* 2005, **129:** 837-845.

78. Malhi H, Gores GJ: **Cholangiocarcinoma: modern advances in understanding a deadly old disease.** *J Hepatol* 2006, **45:** 856-867.

79. Blechacz B, Gores GJ: **Cholangiocarcinoma: advances in pathogenesis, diagnosis, and treatment.** *Hepatology* 2008, **48:** 308-321.

80. Lankisch TO, Metzger J, Negm AA, Vosskuhl K, Schiffer E, Siwy J *et al*.: **Bile proteomic profiles differentiate cholangiocarcinoma from primary sclerosing cholangitis and choledocholithiasis.** *Hepatology* 2011, **53:** 875-884.

81. Isomoto H: **Epigenetic alterations associated with cholangiocarcinoma (review).** *Oncol Rep* 2009, **22:** 227-232.

82. Mosconi S, Beretta GD, Labianca R, Zampino MG, Gatta G, Heinemann V: **Cholangiocarcinoma.** *Crit Rev Oncol Hematol* 2009, **69:** 259-270.

83. Mary Lee K: **Cholangiocarcinoma.** In *Encyclopedia of Gastroenterology*. Edited by Editor-in-Chief:-á-á Leonard Johnson. New York: Elsevier; 2004:301-304.

84. Patel T: **Cholangiocarcinoma.** *Nat Clin Pract Gastroenterol Hepatol* 2006, **3:** 33-42.

85. Taniai M, Higuchi H, Burgart LJ, Gores GJ: **p16INK4a promoter mutations are frequent in primary sclerosing cholangitis (PSC) and PSC-associated cholangiocarcinoma.** *Gastroenterology* 2002, **123:** 1090-1098.

86. Aljiffry M, Walsh MJ, Molinari M: **Advances in diagnosis, treatment and palliation of cholangiocarcinoma: 1990-2009.** *World J Gastroenterol* 2009, **15:** 4240-4262.

87. Khan SA, Thomas HC, Davidson BR, Taylor-Robinson SD: **Cholangiocarcinoma.** *Lancet* 2005, **366:** 1303-1314.

88. Boberg KM, Lind GE: **Primary sclerosing cholangitis and malignancy.** *Best Pract Res Clin Gastroenterol* 2011, **25:** 753-764.

89. Kim MS, Lee J, Sidransky D: **DNA methylation markers in colorectal cancer.** *Cancer Metastasis Rev* 2010, **29:** 181-206.

90. La Thangue NB, Kerr DJ: **Predictive biomarkers: a paradigm shift towards personalized cancer medicine.** *Nat Rev Clin Oncol* 2011, **8:** 587-596.

91. Ned RM, Melillo S, Marrone M: **Fecal DNA testing for Colorectal Cancer Screening: the ColoSure test.** *PLoS Curr* 2011, **3:** RRN1220.

92. Herbst A, Rahmig K, Stieber P, Philipp A, Jung A, Ofner A *et al*.: **Methylation of NEUROG1 in Serum Is a Sensitive Marker for the Detection of Early Colorectal Cancer.** *Am J Gastroenterol* 2011.

93. Lind GE, Ahlquist T, Kolberg M, Berg M, Eknaes M, Alonso MA *et al*.: **Hypermethylated MAL gene - a silent marker of early colon tumorigenesis.** *J Transl Med* 2008, **6:** 13.

94. Lind GE, Ahlquist T, Lothe RA: **DNA hypermethylation of MAL: a promising diagnostic biomarker for colorectal tumors.** *Gastroenterology* 2007, **132:** 1631-1632.

95. Lind GE, Danielsen SA, Ahlquist T, Merok MA, Andresen K, Skotheim RI *et al*.: **Identification of an epigenetic biomarker panel with high sensitivity and specificity for colorectal cancer and adenomas.** *Mol Cancer* 2011, **10:** 85.

96. Sandhu DS, Shire AM, Roberts LR: **Epigenetic DNA hypermethylation in cholangiocarcinoma: potential roles in pathogenesis, diagnosis and identification of treatment targets.** *Liver Int* 2008, **28:** 12-27.

97. Tischoff I, Markwarth A, Witzigmann H, Uhlmann D, Hauss J, Mirmohammadsadegh A *et al*.: **Allele loss and epigenetic inactivation of 3p21.3 in malignant liver tumors.** *Int J Cancer* 2005, **115:** 684-689.

98. Kim BH, Cho NY, Shin SH, Kwon HJ, Jang JJ, Kang GH: **CpG island hypermethylation and repetitive DNA hypomethylation in premalignant lesion of extrahepatic cholangiocarcinoma.** *Virchows Arch* 2009, **455:** 343-351.

99. Lind GE, Kleivi K, Meling GI, Teixeira MR, Thiis-Evensen E, Rognum TO *et al*.: **ADAMTS1, CRABP1, and NR3C1 identified as epigenetically deregulated genes in colorectal tumorigenesis.** *Cell Oncol* 2006, **28:** 259-272.

100. Lind GE, Raiborg C, Danielsen SA, Rognum TO, Thiis-Evensen E, Hoff G *et al*.: **SPG20, a novel biomarker for early detection of colorectal cancer, encodes a regulator of cytokinesis.** *Oncogene* 2011, **30:** 3967-3978.

101. Miller G, Socci ND, Dhall D, D'Angelica M, DeMatteo RP, Allen PJ *et al*.: **Genome wide analysis and clinical correlation of chromosomal and transcriptional mutations in cancers of the biliary tract.** *J Exp Clin Cancer Res* 2009, **28:** 62.

102. Obama K, Ura K, Li M, Katagiri T, Tsunoda T, Nomura A *et al*.: **Genome-wide analysis of gene expression in human intrahepatic cholangiocarcinoma.** *Hepatology* 2005, **41:** 1339-1348.

103. Danielsen SA, Lind GE, Bjornslett M, Meling GI, Rognum TO, Heim S *et al*.: **Novel mutations of the suppressor gene PTEN in colorectal carcinomas stratified by microsatellite instability- and TP53 mutation- status.** *Hum Mutat* 2008, **29:** E252-E262.

104. Shapiro R, Weisgras JM: **Bisulfite-catalyzed transamination of cytosine and cytidine.** *Biochem Biophys Res Commun* 1970, **40:** 839-843.

105. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW *et al*.: **A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.** *Proc Natl Acad Sci U S A* 1992, **89:** 1827-1831.

106. Clark SJ, Harrison J, Paul CL, Frommer M: **High sensitivity mapping of methylated cytosines.** *Nucleic Acids Res* 1994, **22:** 2990-2997.

107. Laird PW: **Principles and challenges of genome-wide DNA methylation analysis.** *Nat Rev Genet* 2010, **11:** 191-203.

108. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB: **Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands.** *Proc Natl Acad Sci U S A* 1996, **93:** 9821-9826.

109. Warnecke PM, Stirzaker C, Song J, Grunau C, Melki JR, Clark SJ: **Identification and resolution of artifacts in bisulfite sequencing.** *Methods* 2002, **27:** 101-107.

110. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors.** *Proc Natl Acad Sci U S A* 1977, **74:** 5463-5467.

111. Melki JR, Vincent PC, Clark SJ: **Concurrent DNA hypermethylation of multiple genes in acute myeloid leukemia.** *Cancer Res* 1999, **59:** 3730-3740.

112. Weisenberger DJ, Campan M, Long TI, Kim M, Woods C, Fiala E *et al*.: **Analysis of repetitive element DNA methylation by MethyLight.** *Nucleic Acids Res* 2005, **33:** 6823-6836.

113. Hagemann S, Heil O, Lyko F, Brueckner B: **Azacytidine and decitabine induce gene-specific and non-random DNA demethylation in human cancer cell lines.** *PLoS One* 2011, **6:** e17388.

114. Smiraglia DJ, Rush LJ, Fruhwald MC, Dai Z, Held WA, Costello JF *et al*.: **Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies.** *Hum Mol Genet* 2001, **10:** 1413-1419.

115. Ueki T, Walter KM, Skinner H, Jaffee E, Hruban RH, Goggins M: **Aberrant CpG island methylation in cancer cell lines arises in the primary cancers from which they were derived.** *Oncogene* 2002, **21:** 2114-2117.

116. Lind GE, Thorstensen L, Lovig T, Meling GI, Hamelin R, Rognum TO *et al*.: **A CpG island hypermethylation profile of primary colorectal carcinomas and colon cancer cell lines.** *Mol Cancer* 2004, **3:** 28.

117. Suzuki H, Gabrielson E, Chen W, Anbazhagan R, van EM, Weijenberg MP *et al*.: **A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer.** *Nat Genet* 2002, **31:** 141-149.

118. Lorincz AT: **The Promise and the Problems of Epigenetics Biomarkers in Cancer.** *Expert Opin Med Diagn* 2011, **5:** 375-379.

119. Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, Shibata D *et al*.: **MethyLight: a high-throughput assay to measure DNA methylation.** *Nucleic Acids Res* 2000, **28:** E32.

120. Mori Y, Cai K, Cheng Y, Wang S, Paun B, Hamilton JP *et al*.: **A genome-wide search identifies epigenetic silencing of somatostatin, tachykinin-1, and 5 other genes in colon cancer.** *Gastroenterology* 2006, **131:** 797-808.

121. Warnecke PM, Stirzaker C, Melki JR, Millar DS, Paul CL, Clark SJ: **Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA.** *Nucleic Acids Res* 1997, **25:** 4422-4426.

122. Tischoff I, Wittekind C, Tannapfel A: **Role of epigenetic alterations in cholangiocarcinoma.** *J Hepatobiliary Pancreat Surg* 2006, **13:** 274-279.

123. Dietrich D, Krispin M, Dietrich J, Fassbender A, Lewin J, Harbeck N *et al*.: **CDO1 promoter methylation is a biomarker for outcome prediction of anthracycline treated, estrogen receptor-positive, lymph node-positive breast cancer patients.** *BMC Cancer* 2010, **10:** 247.

124. Mossman D, Scott RJ: **Long term transcriptional reactivation of epigenetically silenced genes in colorectal cancer cells requires DNA hypomethylation and histone acetylation.** *PLoS One* 2011, **6:** e23127.

125. Staub E, Grone J, Mennerich D, Ropcke S, Klamann I, Hinzmann B *et al*.: **A genome-wide map of aberrantly expressed chromosomal islands in colorectal cancer.** *Mol Cancer* 2006, **5:** 37.

126. Verissimo CS, Molenaar JJ, Fitzsimons CP, Vreugdenhil E: **Neuroblastoma therapy: what is in the pipeline?** *Endocr Relat Cancer* 2011, **18:** R213-R231.

127. Morris MR, Ricketts CJ, Gentle D, McRonald F, Carli N, Khalili H *et al*.: **Genome-wide methylation analysis identifies epigenetically inactivated candidate tumour suppressor genes in renal cell carcinoma.** *Oncogene* 2011, **30:** 1390-1401.

128. Pollard KS, Serre D, Wang X, Tao H, Grundberg E, Hudson TJ *et al*.: **A genome-wide approach to identifying novel-imprinted genes.** *Hum Genet* 2008, **122:** 625-634.

129. Bretthauer M: **Colorectal cancer screening.** *J Intern Med* 2011, **270:** 87-98.

130. Calderwood AH, Schroy PC, III: **Colorectal cancer: Increasing colorectal cancer screening--miles to go.** *Nat Rev Gastroenterol Hepatol* 2011, **8:** 421-422.

131. Buchen L: **Cancer: Missing the mark.** *Nature* 2011, **471:** 428-432.

132. Linden A: **Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis.** *J Eval Clin Pract* 2006, **12:** 132-139.

9 Appendix

Appendix I: Culturing conditions of cancer cell lines

Cell line	Medium	Additives
ALA	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
Co115	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
Colo320	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
EB	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
FRI	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
HCT116	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
HCT15	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
HT29	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
IS1	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
IS2	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
IS3	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
LoVo	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
LS1034	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
LS174T	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
RKO	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
SW48	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
SW480	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
TC7	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
TC71	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*

V9P	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (D-MEM/F12) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
EGI-1	Dulbecco's Modified Eagle Medium (DMEM) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
HuCCT1	Roswell Park Memorial Institute (RPMI) 1640 medium *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
KMCH	Dulbecco's Modified Eagle Medium (DMEM) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
KMBC	Dulbecco's Modified Eagle Medium (DMEM) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum* and Horse Serum*
SK-ChA-1	Minimum Essential Medium (MEM) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
TFK-1	Roswell Park Memorial Institute (RPMI) 1640 medium *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
Mz-ChA-1	Roswell Park Memorial Institute (RPMI) 1640 medium *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
Mz-ChA-2	Minimum Essential Medium (MEM) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
HB8065	Minimum Essential Medium (MEM) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
JHH-1	William's Medium E *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
JHH-4	Minimum Essential Medium (MEM) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
JHH-5	William's Medium E *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
AsPc-1	Roswell Park Memorial Institute (RPMI) 1640 medium *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
BxBc-3	Roswell Park Memorial Institute (RPMI) 1640 medium *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
CFPAC-1	Iscove's Modified Dulbecco's Medium (IMDM) #	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
HPAFII	Minimum Essential Medium (MEM) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*
PaCa-2	Dulbecco's Modified Eagle Medium (DMEM) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum* and MEM Non Essential Amino Acids*
Panc-1	Dulbecco's Modified Eagle Medium (DMEM) *	Penicillin-Streptomycin-Glutamine*, Fetal Bovine Serum*

Medium and additives were added according to requirements for each cell line.

* Gibco, Invitrogen, Carlsbad, CA, USA

ATCC, Manassas, VA, USA

Appendix II: Clinico- pathological information about colorectal tissue samples

Sample name	Age	Dukes	Localization	Sex	MSI status	BRAF_ex15
AUS2_001	71	1	Distal	woman	MSS	W.T
AUS2_002	75	1	Distal	man	MSS	W.T
AUS2_003	79	2	Proximal	woman	MSS	W.T
AUS2_006	62	1	Proximal	man	MSS	W.T
AUS2_007	87	1	Distal	woman	MSI	W.T
AUS2_008	39	1	Distal	woman	MSS	W.T
AUS2_009	62	1	Distal	woman	MSS	W.T
AUS2_010	64	2	Proximal	man	MSI	W.T
AUS2_011	67	3	Distal	woman	MSS	W.T
AUS2_012	84	2	Proximal	woman	MSI	BRAFV600E
AUS2_013	59	2	Proximal	woman	MSS	W.T
AUS2_015	66	4	Proximal	woman	MSI	BRAFV600E
AUS2_016	44	1	Distal	man	MSS	W.T
AUS2_017	74	2	Distal	woman	MSS	W.T
AUS2_018	79	3	Distal	man	MSS	W.T
AUS2_019	71	3	Proximal	man	MSI	W.T
AUS2_020	42	4	Proximal	man	MSS	BRAFV600E
AUS2_021	77	1	Distal	man	MSS	W.T
AUS2_024	78	3	Proximal	man	MSS	W.T
AUS2_025	71	1	Distal	man	MSS	W.T
AUS2_026	62	2	Proximal	woman	MSS	W.T
AUS2_027	59	3	Proximal	woman	MSS	BRAFV600E
AUS2_028	81	2	Distal	woman	MSS	W.T
AUS2_029	66	2	Distal	woman	MSS	W.T
AUS2_030	84	2	Proximal	woman	MSI	BRAFV600E
AUS2_031	67	2	Distal	woman	MSI	BRAFV600E
AUS2_032	69	1		man	MSI	BRAFV600E
AUS2_033	64	1	Proximal	woman	MSI	W.T
AUS2_035	81	1	Distal	woman	MSS	W.T
AUS2_036	35	3	Distal	man	MSS	W.T
AUS2_037	70	2	Proximal	man	MSI	W.T
AUS2_038	82	2	Distal	woman	MSS	BRAFV600E
AUS2_039	54	1	Distal	man	MSS	W.T
AUS2_040	69	4	Distal	man	MSS	W.T
AUS2_041	78	3	Proximal	man	MSS	W.T
AUS2_042	84	2	Proximal	man	MSS	W.T
AUS2_043	58	3	Distal	man	MSS	W.T
AUS2_044	79	3	Proximal	man	MSS	W.T
AUS2_045	75	1	Proximal	man	MSI	BRAFV600E

AUS2_047	47	1	Distal	woman	MSS	W.T
AUS2_048	84	2	Proximal	woman	MSS	W.T
AUS2_049	34	4	Distal	man	MSS	W.T
AUS2_050	66	2	Proximal	woman	MSS	W.T
AUS2_051	79	2	Distal	man	MSS	W.T
AUS2_052	65	2	Distal	man	MSS	W.T
AUS2_053	83	3	Distal	man	MSS	W.T
AUS2_054	83	3	Proximal	man	MSS	W.T
AUS2_055	51	2	Proximal	woman	MSS	W.T
AUS2_056	83	3	Distal	man	MSS	BRAFV600E
AUS2_058	63	3	Distal	man	MSS	W.T
AUS2_059	83	2	Distal	woman	MSS	W.T
AUS2_060	71	1	Distal	man	MSS	W.T
AUS2_061	74	4	Proximal	woman	MSS	W.T
AUS2_063	67	4	Distal	woman	MSS	W.T
AUS2_065	91	2	Proximal	woman	MSS	W.T
AUS2_066	84	2	Proximal	woman	MSI	BRAFV600E
AUS2_067	81	2	Distal	man	MSS	W.T
AUS2_068	75	2	Proximal	woman	MSS	W.T
AUS2_069	59	1	Distal	man	MSS	W.T
AUS2_070	58	3	Distal	man	MSS	W.T
AUS2_071	85	1	Distal	man	MSS	W.T
AUS2_072	64	3	Proximal	man	MSI	W.T
AUS2_073	64	4	Proximal	woman	MSS	W.T
AUS2_074	73	1	Proximal	man	MSS	W.T
AUS2_075	90	3	Distal	woman	MSS	W.T
N1	44		Distal	man		
N2	54		Proximal	man		
N3	33		Proximal	man		
N9	63		Proximal	woman		
N10	72		Distal	woman		
N11	40		Distal	woman		
N12	40		Proximal	man		
N13	38		Distal	man		
N14	38		Proximal	man		
N15	82		Distal	woman		
N16	54		Proximal	man		
N17	60		Distal	woman		
N18	75		Proximal	woman		
N19	39		Distal	woman		
N20	48		Proximal	woman		
N21	54		Distal	man		
N22	86		Proximal	woman		
N23	39		Proximal	man		
N24	40		Distal	man		
N26	74		Distal	man		

N27	51	Distal	man
N28	79	Distal	woman
N30	24	Proximal	man
N36	85	Distal	man
N38	22	Distal	woman
N3	29	Distal	woman
N42	46	Proximal	woman
N43	35	Distal	man
N47	62	Proximal	woman
N49	73	Distal	woman
N50	61	Proximal	man
N52	62	Proximal	man
N53	59	Proximal	woman
N55	78	Proximal	woman
N57	81	Proximal	man
N58	38	Distal	man
N60	53	Distal	woman
N61	55	Distal	man
N63	54	Distal	man
N64	28	Distal	man
N65	62	Distal	man
N66	60	Proximal	man
N67	55	Proximal	woman
N68	72	Proximal	man
N70_re	85	Distal	man
N70_co	85	Proximal	man
N71_re	57	Distal	man
N71_co	57	Proximal	man
N72_re	66	Distal	man
N72_co	66	Proximal	man

Appendix III: MSP- and BS- primer sequences, fragment size, location, and PCR conditions

Primer set	Sense primer	Antisense primer	Frg. Size, bp	Fragment location	An. Temp	$MgCl_2$	El. Time (sec)
ASRGL1_MSP_M	GAGATAGGTGCGCGTTAGTC	ACAACGATTCTACGCCTACG	91	+60 to +151	57	1.5	30
ASRGL1_MSP_U	AGTGAGATAGGTGTGTGTTAGTT	ACAACAATTCTACACCTACACAC	94	+57 to +151	57	1.5	30
ATF3_MSP_M	AGCGAGTACGTATATTTGGC	AAAACGAAACCGAAAACG	174	-221 to -47	53	1.5	30
ATF3_MSP_U	AGTAGTGAGTATGTATATTTGGT	ACCAAAACAAAACCAAAAACA	180	-224 to -44	53	1.5	30
BEX4_MSP_M	AGGGGTTGATTCGAAAGTTTC	TCTAACGCCAAAACGAAACA	132	-90 to +42	55	1.5	30
BEX4_MSP_U	GATAGGGGTTGATTTGAAAGTTTT	AACTCTAACACCAAAACAAAACA	138	-93 to +45	55	1.5	30
CALCOCO1_MSP_M	TACGTTTTTTAGGATGTCGC	CTTTTACCGCTACGTACTCG	116	-118 to -2	55	1.5	30
CALCOCO1_MSP_U	AATTATGTTTTTTAGGATGTTGT	CCCTTTTACCACTACATACTCAA	121	-121 to 0	55	1.5	30
CDO1_MSP_M	TTGGGACGTCGGAGATAAC	GACCCTCGAAAAAAAAACGA	145	-153 to -8	53	1.5	30
CDO1_MSP_U	TTTTTGGGATGTTGGAGATAAT	AACCCTCAAAAAAAAAACAAAAC	148	-156 to -8	53	1.5	30
CDO1_BS	TTTTTTTTGTTTAYGTTTTA	ACAAATCAAATTCAAATCT	350	-280 to +70	49	1.7	30
CLU_MSP_M	TTTTTTTTATTGGAAGCGTC	AAAAAATACCGCGAAAAAC	165	-147 to +18	52	2.4	30
CLU_MSP_U	GGTTTTTTTTTATTGGAAGTGTT	CCAAAAAATACCACAAAAAACA	170	-150 to +20	52	2.4	30
CRISPLD2_MSP_M	TTCGTTTATTCGGCGTTC	ACTCAACGTACCGCCTCTT	172	-178 to -6	52	1.5	30
CRISPLD2_MSP_U	TTTTTTGTTTATTTGGTGTTT	AAAACTCAACATACCACCTCTT	178	-181 to -3	52	1.5	30
CSRP1_MSP_M	ACGTGTAAGACGTTTTTCGC	AACCCGACGATACTACCCTC	147	-126 to +21	55	1.7	30
CSRP1_MSP_U	GTATGTGTAAGATGTTTTTTGT	AACCCAACAATACTACCCTCCT	149	-128 to +21	56	1.5	30
CTGF_MSP_M	TCGGAGCGTATAAAAGTTTC	CTATCGACCGAAACGACTAC	122	-34 to +88	56	2.5	30
CTGF_MSP_U	GTTTGGAGTGTATAAAAGTTTT	CTATCAACCAAAACAACTACCA	124	-36 to +88	56	2.5	30
DCLK1_MSP_M	GCGTTTTGTTAAGAAGGGC	ACGCGCTCCCTTTTCTTAT	108	-127 to -19	53	1.5	30
DCLK1_MSP_U	GTGTTTTGTTAAGAAGGGT	ACACACTCCCTTTTCTTAT	108	-127 to -19	53	1.5	30
DCLK1_BS	AAGATTATTTGTGGGGATTAGG	AACCTCTCTCTCCAAAAAAAAA	271	-247 to +24	57	1.5	30
DUSP5_MSP_M	GAGTGAGTTTTTTAGCGAAGC	ATAAATACCGTCCGTAACGC	198	-192 to +6	52	1.5	30

DUSP5_MSP_U	GAGTGAGTTTTTTAGTGAAGT	ATAAATACCATCCATAACAC	198	-192 to +6	52	1.5	30
EGR2_MSP_M	TATATGGGTAGCGACGTTAC	TCGCCGAACTATTAATCAATTA	104	-108 to -4	52	2	30
EGR2_MSP_U	TTATATATGGGTAGTGATGTTAT	CCCTCACCAAACTATTAATCAATTA	110	-111 to -1	52	1.5	30
FAM3B_MSP_M	GGGGAACGGGTTTATTTTTC	GCGACCAATCGAACAAAT	137	-120 to +17	53	1.5	30
FAM3B_MSP_U	GGGGAATGGGTTTATTTTTT	ACAACCAATCAAACAAAT	137	-120 to +17	53	1.5	30
FHL1_MSP_M	TCGTGTAGTGGGTAGAGTTC	CTCCGCCGAACGATAAAT	165	-160 to +5	57	1.5	30
FHL1_MSP_U	TTTTTGTGTAGTGGGTAGAGTTT	CCCCTCCACCAAACAATAAAT	171	-163 to +8	57	1.5	30
FKBP1B_MSP_M	GGTTCGTTAATAGTCGGGC	CTAAAATCGAAACCTACGCG	126	-158 to -32	55	2	30
FKBP1B_MSP_U	TTAGGTTTGTTAATAGTTGGGT	ACTAAAATCAAAACCTACACAAA	130	-161 to -31	52	1.5	30
GNG11_MSP_M	TCGGATGTGATTTGGAAAC	CGCGAAAAACGACTAAACT	112	-48 to +64	56	1.5	30
GNG11_MSP_U	ATTTGGATGTGATTTGGAAAT	CCCACAAAAAACAACTAAACT	116	-50 to +66	56	1.5	30
GPR124_MSP_M	GGGTTTAGGTTTGGTCGC	CCGCTCCGTACCATAAATAA	119	-124 to -5	55	2.5	30
GPR124_MSP_U	AGAGGGTTTAGGTTTGGTTGT	CCACCACTCCATACCATAAATAA	125	-127 to -2	55	1.5	30
GREM1_MSP_M	AGTAGATAAAGAGGCGAGGC	AAATACCGACGACAAAACG	172	-198 to -26	53	1.5	30
GREM1_MSP_U	GGGAGTAGATAAAGAGGTGAGGT	AAATACCAACAACAAAACACAA	175	-201 to -26	53	1.5	30
HABP4_MSP_M	CGTGACGTGATAGTAGTCGGTC	CTATCCGACCCCTACCGAC	149	-115 to +34	58	1.5	30
HABP4_MSP_U	GTGTGATGTGATAGTAGTTGGTT	CCTATCCAACCCCTACCAAC	151	-116 to +35	59	1.5	30
ID3_MSP_M	TTCGGAGGAGTTGTGGTTC	CGCTAATACCGAAAAAAAACG	173	-32 to +141	55	1.5	30
ID3_MSP_U	GATTTTGGAGGAGTTGTGGTTT	CACTAATACCAAAAAAAAACAAAC	176	-35 to +141	55	1.5	30
INPP5A_MSP_M	TTAGCGGATTTAATGGTTGC	TAACCGAAACTCCGACCTC	113	-20 to +93	50	1.5	30
INPP5A_MSP_U	TTAGTGGATTTAATGGTTGT	TAACCAAAACTCCAACCTC	113	-20 to +93	50	1.5	30
ITPR1_MSP_M	ATTTAGGGTTTAGTTCGGGC	ACACTTTAAAACGACTCCGAA	148	-146 to +2	55	2.5	30
ITPR1_MSP_U	TTTATTTAGGGTTTAGTTTGGGT	ACTACACTTTAAAACAACTCCAAA	154	-149 to +5	55	2	30
LHX6_MSP_M	TGCGGTTGTGGTTTTTTTC	CCGAAACGACGTTCTCAT	100	-69 to +31	54	1.5	30
LHX6_MSP_U	TATTGTGGTTGTGGTTTTTTTT	ACACCAAAACAACATTCTCAT	106	-72 to +34	54	1.5	30
LMCD1_MSP_M	GGTAGTCGGCGTTTAGTTTC	CGCAACTAAACCGCTTTAAT	165	-176 to -11	55	1.5	30
LMCD1_MSP_U	TAGGGTAGTTGGTGTTTAGTTTT	AAACACAACTAAACCACTTTAAT	171	-179 to -8	55	1.5	30
MLLT11_MSP_M	TTTTTCGGGTTAGTTTTGC	AACCGAACGAATTTCGTAAT	110	-118 to -8	51	1.8	30
MLLT11_MSP_U	GGGTTTTTTGGGTTAGTTTTGT	CCAAACCAAACAAATTTCATAAT	116	-121 to -5	52	1.5	30
MT1F_MSP_M	GTTTAGGGGATTTTGCGTTC	ACAACCGACCGCTACTTTAA	147	-110 to +37	55	1.5	30

MT1F_MSP_U	GTTTAGGGGATTTTGTGTTT	ACAACCAACCACTACTTTAA	147	-110 to +37	55	1.5	30
MT1X_MSP_M	GGTTTACGGGTTGTTGTATTC	AAAAACCGACGACTCTCTTT	129	-136 to -7	55	1.5	30
MT1X_MSP_U	GGGTTTATGGGTTGTTGTATTT	CAAAAACCAACAACTCTCTTT	131	-137 to -6	55	1.5	30
MT2A_MSP_M	GTGTGTAGAGTCGGGTGC	AAAACCGAAACGAATACAAAA	132	-108 to -240	55	1.5	30
MT2A_MSP_U	GTGTGTAGAGTTGGGTGT	AAAACCAAAACAAATACAAAA	132	-108 to -240	55	1.5	30
NAP1L2_MSP_M	GCGTAATTATATTGCGGTATC	TACGTTAACCGATCCTACAA	116	+8 to +124	56	1.5	30
NAP1L2_MSP_U	GTTGTGTAATTATATTGTGGTATT	AACTACATTAACCAATCCTACAA	122	+5 to +127	56	1.5	30
NR4A3_MSP_M	TTTTCGTATACGCGGAATC	TCGACACGTCATTTATACCAC	142	-126 to +16	52	1.5	30
NR4A3_MSP_U	TTTTTTTTGTATATGTGGAATT	CTCTCAACACATCATTTATACCAC	148	-129 to +19	52	1.5	30
PDE2A_MSP_M	ATTAGGCGAAGTTGTCGC	CGACTCGTCCGACTTAAAA	161	+10 to +171	53	1.8	30
PDE2A_MSP_U	GGATTAGGTGAAGTTGTTGT	AACAACTCATCCAACTTAAAA	165	+8 to +173	53	1.8	30
REEP1_MSP_M	GGACGCGTTCGTTTTTAGTC	AACCGCGACACGTTCTAAC	149	-162 to -13	55	2.5	30
REEP1_MSP_U	GTAGGATGTGTTTGTTTTTAGTT	AACCACAACACATTCTAACAAC	152	-165 to -13	55	2.5	30
RNASE4_MSP_M	TAAATTTCGGACGAGTTTTC	TCGCGAAACAATTTATATTTC	101	-143 to -42	53	2.5	30
RNASE4_MSP_U	GTTTAAATTTTGGATGAGTTTTT	CCATCACAAAACAATTTATATTTC	107	-146 to -39	53	1.5	30
SFRP1_MSP_M	TAGTAAATCGAATTCGTTCGC	TACGCGAAACTCCTACGAC	141	-138 to +3	45	1.5	30
SFRP1_MSP_U	TTTTAGTAAATTGAATTTGTTTGT	TACACAAAACTCCTACAACCAA	144	-141 to +3	45	1.5	30
SLC46A3_MSP_M	GTTGAGTGGTTGTTCGGTC	CCCGACTCTCCTACGATTAA	151	-152 to -1	57	1.5	30
SLC46A3_MSP_U	GTGTTGAGTGGTTGTTTGGTT	TACCCAACTCTCCTACAATTAA	155	-154 to +1	58	1.5	30
SYT11_MSP_M	CGTTTTGGAATTATAGCGC	TTCCGAATAATCCTCGAAA	158	-222 to -64	50	1.8	30
SYT11_MSP_U	TTTTGTTTTGGAATTATAGTGT	CTCTTCCAAATAATCCTCAAAA	164	-225 to -61	50	1.8	30
TCF4_MSP_M*	GAATTTGTAATTTCGTGCGTTTC	AAAAAAAACTCTCCGTACACCG	258	+322 to +580	57	1.5	60
TCF4_MSP_U*	TGAATTTGTAATTTTGTGTGTTTTG	AAAAAAAACTCTCCATACACCACC	259	+321 to +580	57	1.5	60
TPM2_MSP_M	ATCGTCGGGGTTTTTTTAGTC	AACAAAAACACGACCCGAC	152	-156 to -4	61	1.5	30
TPM2_MSP_U	GTATTGTTGGGGTTTTTTTAGTT	AAACAAAAACACAACCCAACC	155	-158 to -3	61	1.5	30
ZNF331_MSP_M	GGTAGGACGTTTTTAGGGTC	ATACAACTCTACACGACGCA	143	-120 to +23	55	1.7	30
ZNF331_MSP_U	TAAGGTAGGATGTTTTTAGGGTT	AACATACAACTCTACACAACACA	143	-120 to +23	55	1.5	30
ZNF331_BS	TTTTTGGGGTATGGTTTTATTA	TCCTCATTAAACTATACCCCAA	359	-238 to +121	50	1.5	30
ZSCAN18_MSP_M	GTTTAAAATGACGTAGGCGTC	AATACCGCGAAACTATACCG	131	-52 to +79	55	1.8	30
ZSCAN18MSP_U	GGTGTTTAAAATGATGTAGGTGTT	ACAATACCACAAAACTATACCAC	131	-55 to +79	55	1.5	30

ZSCAN18_BS	TTTTGGTTGTTAGGGGTTTATT	ACCCACCTACTACRCAACTAC	302	-106 to +196	59	1.5	30

* Primer sequences and amplification conditions were obtained from Kim et al., 2008.

Appendix IV: qMSP primer sequences, fragment size and location

Gene name	Sense primer	Antisense primer	probe	Frg. Size, bp	Fragment location
CDO1_qMSP	CGAATTATAGCGGCGGAGGT	AAATCGCGTAAACTCCGCG	6FAM-CGTTAGGTCGGGCGGT-MGB	101	-329 to -153
DCLK1_qMSP	GCGCGTACGCGGAGG	CGACGACGAACGCGCT	6FAM-GTGTGAGCGAGTGAGAT-MGB	86	-96 to -10
ZNF331_qMSP	TTAGAGAAGTTTCGACGTAGTTGGAA	CGACTCCATTTACGCCGTATAAA	6FAM-CGTTCGTTAGTGTTTTTTAG-MGB	80	-7 to +73
ZSCAN18_qMSP	CGCGGTATAGTTTCGCGGTAT	CGCGATAACGACCGACAAA	6FAM-CGTAGTTCGCGGTGAGG-MGB	84	+57 to +141

Appendix V: Gene symbol, gene name, chromosomal location, and accession number of the analyzed genes

Gene symbol*	Gene name*	Location*	Accesion number#
ASRGL1	asparaginase like 1	11q12.3	NM_001083926
ATF3	activating transcription factor 3	1q32.3	NM_001040619
BEX4	brain expressed, X-linked 4	Xq22.1-q22.3	NM_001080425
CALCOCO1	calcium binding and coiled-coil domain 1	12q13.13	NM_020898
CDO1	cysteine dioxygenase, type I	5q23.2	NM_001801
CLU	clusterin	8p21-p12	NM_001831
CRISPLD2	cysteine-rich secretory protein LCCL domain containing 2	16q24.1	NM_031476
CSRP1	cysteine and glycine-rich protein 1	1q32	NM_004078
CTGF	connective tissue growth factor	6q23.2	NM_001901
DCLK1	doublecortin-like kinase 1	13q13.3	NM_004734
DPYSL3	dihydropyrimidinase-like 3	5q32	NM_001387
DUSP5	dual specificity phosphatase 5	10q25	NM_004419
EGR2	early growth response 2	10q21.1	NM_001136177
FAM3B	family with sequence similarity 3, member B	21q22.3	NM_058186
FHL1	four and a half LIM domains 1	Xq26.3	NM_001449
FKBP1B	FK506 binding protein 1B, 12.6 kDa	2p23.3	NM_054033
GPR124	G protein-coupled receptor 124	8p11.22	NM_032777
GREM1	gremlin 1	15q13.3	NM_013372
HABP4	hyaluronan binding protein 4	9q22.3-q31	NM_014282
ID3	inhibitor of DNA binding 3, dominant negative helix-loop-helix protein	1p36.13-p36.12	NM_002167
INPP5A	inositol polyphosphate-5-phosphatase, 40kDa	10q26.3	NM_005539
ITPR1	inositol 1,4,5-trisphosphate receptor, type 1	3p26.1	NM_002222
LHX6	LIM homeobox 6	9q33.2	NM_014368
LMCD1	LIM and cysteine-rich domains 1	3p26-p24	NM_014583
MLLT11	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 11	1q21	NM_006818
MT1F	metallothionein 1F	16q13	NM_005949
MT1X	metallothionein 1X	16q13	NM_005952
MT2A	metallothionein 2A	16q13	NM_005953
NAP1L2	nucleosome assembly protein 1-like 2	Xq13	NM_021963
NR4A3	nuclear receptor subfamily 4, group A, member 3	9q22	NM_173198
PDE2A	phosphodiesterase 2A, cGMP-stimulated	11q13.1-q14.1	NM_001143839
REEP1	receptor accessory protein 1	2p11.2	NM_001164732
RNASE4	ribonuclease, RNase A family, 4	14q11	NM_002937

SFRP1	secreted frizzled-related protein 1	8p11.21	NM_003012
SLC46A3	solute carrier family 46, member 3	13q12.3	NM_181785
SYT11	otagmin XI	1q22	NM_152280
TCF4	transcription factor 4	18q21.1	NM_001083962
TPM2	tropomyosin 2 (beta)	9p13	NM_001145822
ZNF331	zinc finger protein 331	19q13	NM_018555
ZSCAN18	zinc finger and SCAN domain containing 18	19q13.43	NM_023926

* Gene Symbol, full Gene Name, and chromosome location are in accordance with the approved guidelines from the HUGO Nomenclature Committee at the European Bioinformatics Institue, http://www.genenames.org

Sequence Accesion Numbersare from the USCS Genome Browser http://genome.ucsc.edu/ and represent sequences used for primer design.