Identification of novel epigenetic masterkeys in cancer

- with potential diagnostic value

Ane Brenna

A thesis for the Master's degree in Molecular Biosciences

Department of Biosciences
Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

May 2014

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

© Ane Brenna

2014

Identification of novel epigenetic masterkeys in cancer - with potential diagnostic value

Ane Brenna

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

Acknowledgements

The work presented in this thesis was carried out in the project Group of Epigenetics, Department of Cancer Prevention, Institute for Cancer Research, Norwegian Radium Hospital, Oslo University Hospital, in the period April 2013 to June 2014.

First and foremost, I would like to thank my supervisor Guro E. Lind for her excellent guidance, encouragement and support throughout this project. Her knowledge and enthusiasm regarding the field of epigenetics and cancer research has been an inspiration. I am grateful to the head of department, Professor Ragnhild A. Lothe, for her effort in creating such an outstanding academic department.

I would like to give my gratitude to my wonderful colleagues for answering my questions and helped me in the lab, especially Hilde, Hege Marie, Kim, Tone and Anne Cathrine. I would also like to thank all of you for creating such a wonderful social environment and for valuable lunch breaks filled with inspiring conversations and laughter. A special thanks goes to Gro and Heidi, my fellow master students in the group, it would not have been the same without you!

I would like to thank my family and friends, for always being there for me and for giving my life joy and meaning. Last, but not least, I would like to thank Frederik, for all the love and patience, and for always believing in me.

Oslo, May 2014

Ane Brenna

Table of content

Acknowledgements III
Summary VII
1 Introduction 1
1.1 Cancer 1
1.2 Carcinogenesis 1
1.2.1 Clonal evolution of cancer 1
1.2.2 Hallmarks of cancer 2
1.2.3 Cancer critical genes 3
1.3 Epigenetics and regulation of gene expression 7
1.3.1 Definition of epigenetics 7
1.3.2 DNA methylation 7
1.3.3 Chromatin 14
1.4 Drivers and passengers in tumor development 16
1.5 DNA methylation biomarkers 18
1.6 Epigenetics and cancer therapy 21
2 Aims 25
3 Materials and Methods 27
3.1 Materials 27
3.1.1 Cancer cell lines 27
3.2 Methods 28
3.2.1 *In vitro*, *in silico* and *in vivo* strategy for identifying DNA methylation candidate genes 28
3.2.2 Sodium bisulfite modification 31
3.2.3 Methylation specific polymerase chain reaction, MSP 32
3.2.4 Bisulfite sequencing 35
3.2.5 Quantitative methylation-specific polymerase chain reaction, qMSP 38
3.2.6 Statistics 40
4 Results 41
4.1 Identification of potential epigenetic candidate genes 41
4.2 Methylation specific PCR (MSP) 42
4.2.1 Primer design and optimization 42

4.2.2 Methylation status of cancer cell lines as assessed MSP 42
4.3 Bisulfite sequencing 45
4.4 quantitative methylation specific PCR (qMSP) 48
4.4.1 Methylation status of the entire cancer cell lines panel as assessed by qMSP ... 48
4.4.2 Comparison of qualitative and quantitative methylation analysis 50
4.5 Gene expression results from the IST database 51
5 Discussion 53
5.1 Material and Methodological considerations 53
5.1.1 Usage of cancer cell lines as *in vitro* models 53
5.1.2 Microarray results and selection of genes 53
5.1.3 Methods used for determining DNA methylation 55
5.2 *BAIAP3, CCNA1, IFFO1* and *MT3* are found to be frequently methylated in cancer cell lines 56
6 Conclusions 63
7 Future perspectives 65
Appendix 75

Summary

Cancer comprises more than 200 different diseases, with variable molecular constitutions. In spite of the molecular differences some aberrations have been shown to be common across the vast majority of cancer types, obviously providing a strong selective advantage. Common drivers of tumor development and progression could be defined as 'masterkeys'.

Epigenetic changes include any process that alters gene activity without changing the DNA sequence, and leads to modifications that can be transmitted to daughter cells. Epigenetic changes have been shown to be at least as common as genetic changes in cancer, including CpG island promoter hypermethylation of tumor-suppressor genes which is a key event in cancer development. The purpose of this project is to identify candidate genes frequently methylated across several cancer types, which may represent epigenetic 'Masterkeys' for cancer development and possibly possess biomarker potential.

The selection of potential candidate genes for methylation analysis was based on cancer cell line treatment with epigenetic drugs, gene expression microarrays and bioinformatic analysis. The selected gene candidates (n=52) were investigated in 20 cancer cell lines from 16 tissue types using methylation specific PCR (MSP). Five candidates had a methylation frequency above 50 % and were selected for further validation by bisulfite sequencing and finally investigated in a larger cell line panel (n=114 from 17 different cancer types) using quantitative MSP (qMSP).

Four candidates, *BAIAP3, CCNA1, IFFO1* and *MT3*, were found to be frequently methylated across the cell line panel, with a percentage of 46 %, 64 %, 70 % and 67 %, respectively. Future perspectives include analyzing the promoter methylation of these genes in tissue samples derived from cancer patients, as well as performing functional studies to evaluate whether they represent universal cancer drivers ('masterkeys'). If the genes are frequently methylated also in the patient material, their potential as DNA methylation biomarkers for early detection of cancer will be explored.

Abbreviations

2-HG	2-Hydroxyglutarate
5-FU	5-Fluorouracil
5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
ATP	Adenosine triphosphate
AZA	5-Azacytodine
Bp	Base pairs
CIMP	CpG Island Methylator Phenotype
CpG	Cytosine phosphate guanine
CpT	Cytosine phosphate thymine
Ct	Cycle Threshold
DNA	Deoxyribonucleic acid
ddNTP	Dideoxynucleotide triphosphate
dNTP	Deoxyribonucleotide trophosphate
DNMT	DNA methyltransferase
FDA	Food and Drug Administration
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HDM	Histone demethylase
HMT	Histone methyltransferase
HNPCC	Hereditary nonpolyposis colorectal cancer
HNSCC	Head and neck squamous cell carcinoma
IDH	Isocitrate dehydrogenase
IST	*In Silico* Transcriptomics
IVD	*In vitro*-methylated DNA
LOH	Loss of heterozygosity
MAPK	Mitogen-activated protein kinase
MBD	Methyl-binding domain

MGB	Minor groove binding
MMR	Mismatch repair
MSI	Microsatellite instability
MSP	Methylation specific polymerase chain reaction
NB	Normal blood
OGT	Oxford Gene Technology
PcG	Polycomb group
PCR	Polymerase chain reaction
PI3K	Phosphoinositide 3-kinase
PMR	Percent of methylated reference
qMSP	Quantitative methyl specific polymerase chain reaction
RefSeq	Reference sequence
RNA	Ribonucleic acid
SWI/SNF	SWItch/Sucrose NonFermentable
TALE	Transcription activator-like effector
TET	Ten eleven translocation
Tm	Melting temperature
TSA	Trichostatin A
TSS	Transcription start site
WHO	World Health Organization

Gene symbols[1]

APC	Adenomatous polyposis coli
BRCA1	Breast cancer 1
CDKN2A	Cyclin-dependent kinase inhibitor 2A
CNRIP1	Cannabinoid receptor interacting protein 1
DNMT3A	DNA (cytosine-5)-methyltransferase 3 alpha
DOT1L	DOT1-like histone H3K79 methyltransferase
EZH2	Enhancer of zeste homolog 2
FBN1	Fibrillin 1
HRAS	Harvey rat sarcoma viral oncogene homolog
IDH1	Isocitrate dehydrogenase 1
IDH2	Isocitrate dehydrogenase 2
IGF2	Insulin-like growth factor 2
INA	Internexin neuronal intermediate filament protein
KMT2A	Lysine (K)-specific methyltransferase 2A
KRAS	Kirsten rat sarcoma viral oncogene homolog
MAL	mal, T-cell differentiation protein
MGMT	O-6-methylguanine-DNA methyltransferase
MLH1	mutL homolog 1
MSH2	mutS homolog 2
MYC	v-myc avian myelocytomatosis viral oncogene homolog
NRAS	Neuroblastoma RAS viral (v-ras) oncogene homolog
RB1	Retinoblastoma 1
SEPT9	Septin 9
SNCA	Synuclein
SPG20	Spastic paraplegia 20
TP53	Tumor protein 53
VIM	Vimentin

[1] Gene symbols and full names are approved by the HUGO Gene Nomenclature Committee (http://genenames.org). Approved symbols are used throughout this thesis.

1 Introduction

1.1 Cancer

Cancer is abnormal cell division, resulting in a tumor harboring the potential of invading surrounding tissue [1]. It is viewed as over 200 separate diseases, and thought to develop as a result of a series of mutations and epigenetic defects, affecting the DNA [2].

According to the World Health Organization, WHO, there were 14.1 million new cancer cases in 2012, 8.2 million cancer deaths and a total of 32.6 million people living with a cancer diagnosis worldwide[2]. The Cancer Registry in Norway, reported around 30 000 new cases of cancer in 2012. The same year, 10 900 people died of cancer, and around 224 000 people were living with the disease. The statistics show that men are more prone to get cancer, with a lifetime risk in total of 35.5 %, compared to 28.4 % for women. The survival rate is improving in Norway by each year, and was 69.1 % for women and 67.8 % for men in 2011[3].

The WHO stated in a report published earlier this year, that life style and early detection is important when it comes to preventing cancer. They concluded that at least one third of cancer cases can be prevented[2]. Smoking, high alcohol consumption, lack of exercise, diet, older age and viruses are examples indicated as reasons for many cancer cases [3].

1.2 Carcinogenesis

1.2.1 Clonal evolution of cancer

In 1976, Nowell proposed an evolution model for cancer development, that cancer arises from a single cell, and then evolve through the repeated process of clonal expansion, acquired genetic variability and clonal selection [1]. This model has similarities with Darwinian evolution, and the principle "survival of the fittest" lead to a monoclonal cancer cell population that is adapted to survive, divide and ignore control mechanisms that applies for normal cells [4]. This model is widely accepted but there is also observed a high degree of heterogeneity in cancers, affecting key cancer pathways and driving further selective

[2] http://who.int/mediacentre/factsheets/fs297/en/
[3] http://www.kreftregisteret.no

advantage, representing one of the major challenges in targeted, as well as general, cancer therapy [5]. Interestingly, cytogenetic studies have shown that some cancers develop from a polyclonal origin, and that genetic heterogeneity can arise within the population of cancer cells by acquisition of new mutant alleles [6].

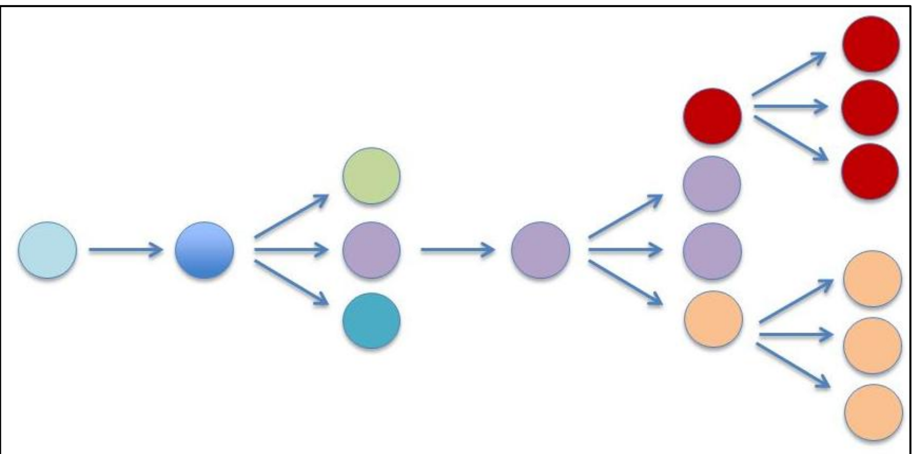


Figure 1: Clonal evolution of cancer

Cancer arises from a single cell, and sub-clones accumulate several genetic and epigenetic aberrations followed by clonal selection. The cancer cells thereby accumulate properties such as resisting apoptosis and other control mechanisms important in the cell.

The abilities a cancer need in tumorigenesis are summarized as the Hallmarks of cancer and are described in the paragraph 1.2.2 below. These abilities can be enabled through inactivation of tumors suppressor genes and DNA repair genes and activation of oncogenes, described in section 1.2.3.

1.2.2 Hallmarks of cancer

Several factors are in play during the development and evolution of a cancer [7]. In the year 2000, Hanahan and Weinberg introduced the Hallmarks of Cancer (Figure 2), indicating the abilities a cell needs to become cancerous. Originally six alterations in the cell physiology were suggested, including self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of programmed cell death (apoptosis), limitless replicative potential, sustained angiogenesis and tissue invasion and metastasis [8]. In 2011 Hanahan and Weinberg suggested four additional hallmarks known as reprogramming of energy metabolism, evading immune destruction, gene instability and mutations and tumor-

promoting inflammation [9]. Today the ten hallmarks of cancer are thought to be crucial in cancer development.

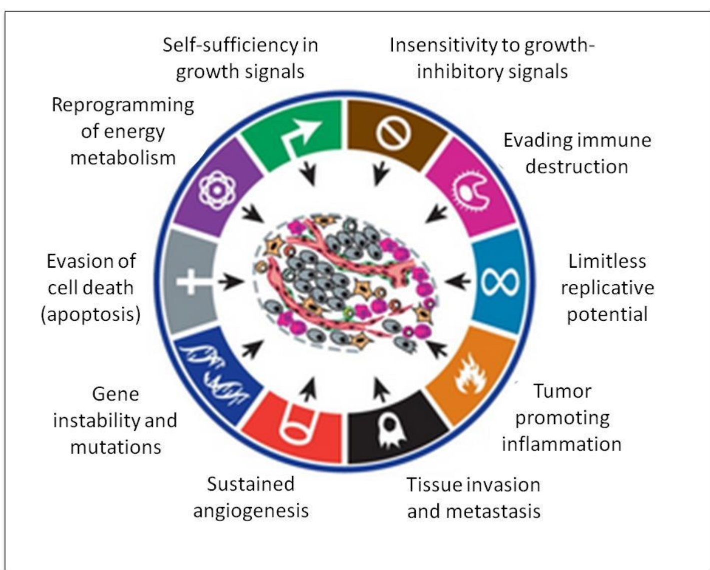


Figure 2: The ten hallmarks of cancer

These hallmarks provide a solid foundation for understanding the biology of cancer. Modified after Hanahan, 2011 [9].

1.2.3 Cancer critical genes

Most of the alterations that lead to cancer are somatic events, but a minority have been shown to be inherited [7]. Gain of function of proto-oncogenes and loss of function of tumor suppressor genes or DNA repair proteins have been intensively studied throughout the last decades, and can potentially lead to cancer (figure 3) [8].

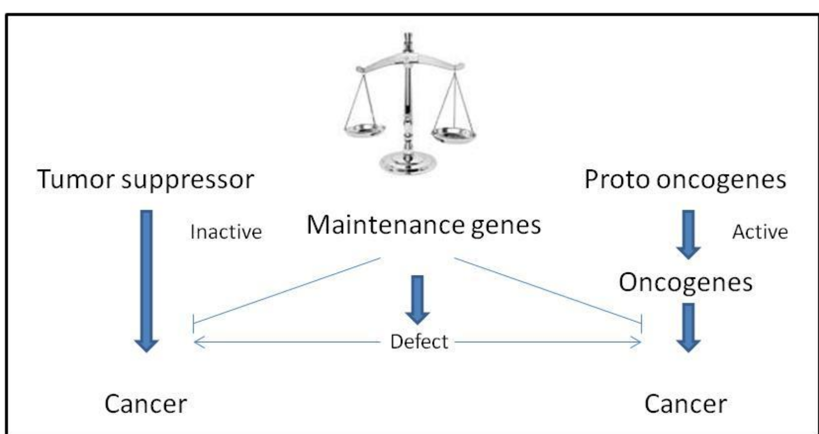


Figure 3: Cancer critical genes

The balance of cancer critical genes, called tumor suppressor genes, proto oncogenes and DNA repair genes are important for the cell. Inactive tumor suppressors and maintenance genes important in repair and activated oncogenes can lead to cancer development.

Oncogenes

Some proteins classified as proto-oncogenes have roles in controlling cell differentiation and growth, and they are important in the cell. If they are overly expressed due to mutations or translocations, they are turned into active oncogenes, and could potentially lead to cancer [7]. Oncogenes lead to gain of function, giving the cell advantage when it comes to tumor development. All genes have two coding alleles present on two identical chromosomes, with exception of the male sex chromosomes, and one alteration in one of the alleles is enough to activate the proto-oncogene into an oncogene [10]. Examples of oncogenes that are frequently activated in cancer are *KRAS* and *MYC*.

KRAS is a G-protein belonging to the Ras family together with HRAS and NRAS. It is involved in signaling, and essential for normal cell function. One point mutation is enough to activate *KRAS* into an oncogene, resulting in an overactive protein, and the consequence of this is abnormal cell growth and proliferation [11]. Oncogenic Ras proteins are reported in 30 % of human cancers, and frequently seen in colorectal cancers, lung cancers and leukemia [12-14].

MYC is an example of a transcription factor, that under certain circumstances turn into an oncoprotein, which is capable of increasing transcription of many genes, leading to cancer [15]. This is a common event in various cancer types, including in Burkitt's lymphoma, where *MYC* forms a fusion gene with an actively transcribed immunoglobulin promoter, enabling it to be transcribed at a much higher rate [16].

Tumor suppressor genes

A tumor suppressor protects the cell from becoming cancerous, and loss of function is required for tumor development [7]. The first evidence for loss-of-function of particular proteins came from studies of retinoblastoma, a cancer in the retina, in children [17]. Knudsons research of this disease, led to the "Knudsons two-hit hypothesis". According to this Knudsons hypothesis, two alterations, or hits, one in each allele, is necessary to inactivate the tumor suppressor gene [7]. Retinoblastoma can be caused either by a germinal or somatic mutation, and Knudson discovered that children born with mutation in one allele, had a much higher risk of developing cancer than children with two healthy alleles [18].

One of the best known tumor suppressors is the Retinoblastoma 1, RB1 protein [7]. It plays a major role in the regulation of the cell cycle, deciding whether the cell can go further than the R point in the G_1 phase or not [19]. The R-point is an important checkpoint, and cells passing through this point will in most cases undergo replication and division [17]. The loss of function of the RB1 protein will give the cell the opportunity to divide and replicate without control [17,19]. Most cancers find a way of inactivating RB1, either directly through a mutation, or indirectly through affecting the RB1 regulators [19]. Recently, there have also been shown that RB1 has other important roles in the cell, such as ensuring chromosome stability, induction of senescence, and regulation of apoptosis, differentiation and angiogenesis [17].

Another famous tumor suppressor is TP53, called 'the guardian of the genome' [20]. As shown in figure 4, TP53 can be activated by stress situations, such as DNA damage, hypoxia, temperature shock and spindle damage during division, and control many responses to these in the cell. TP53 controls the cell fate, halts the cell cycle if the DNA is damaged, initiate DNA repair, apoptosis and senescence. If TP53 is removed from its control post, the cell will go through cell cycle with a potential DNA damage and risk of accumulating additional aberrations. A consequence of this may be tumorigenesis [21].

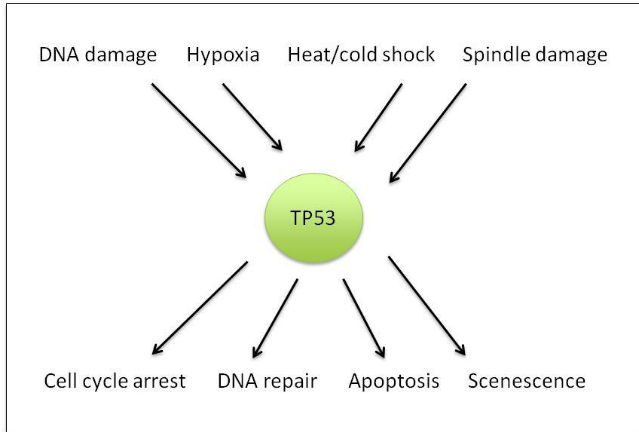


Figure 4: TP53 is the guardian of the cell and the genome.

TP53 is activated through many different stress situations in the cell, such as DNA damage, hypoxia, temperature shock and spindle damage of the division machinery. TP53 initiate many reactions to these events, such as cell cycle arrest, DNA repair, apoptosis and senescence. Modified after Purvis, 2012 [21].

Maintenance genes

Maintenance genes are responsible for correcting DNA damage and retaining genome stability. [11]. Normal reaction to DNA damage is cell cycle arrest or cell death. If the cell cycle is not halted and damage corrected, the consequence can be genome instability, resulting in tumorigenesis. TP53 is an example of a important maintenance protein in the cell, as well as the mismatch repair (MMR) genes described in the next section [22]. Mutagenic processes damaging DNA can happen in several ways, such as X-rays and UV-rays, reactive oxygen species and errors incorporated by DNA polymerase during replication. Damage due to X-rays, UV-rays and oxygen radicals are repaired by nucleotide- and base-excision repair [22]. The third and last category are errors incorporated by DNA polymerase during replication, and they are repaired by the mismatch repair system [11]. In this thesis only mismatch repair will be described in more detail.

Mismatch repair

The mismatch repair, MMR, genes encode enzymes responsible of correcting errors incorporated in the DNA by the DNA polymerase during replication. Even though the error rate is low (one per 10^5 polymerized nucleotides), it is still significant. Some of these mistakes are corrected by the proofreading capabilities of the DNA polymerase itself [11]. But since not all the mutations are corrected this way the DNA MMR genes are crucial, especially in regions with repeated sequences due to the tendency of DNA polymerase slippage. These short sequences are called microsatellites, and they are spread out in the genome, in both intronic as well as coding sequences. A defect in the MMR system leads to failure in correcting errors incorporated by the DNA polymerase in these microsatellites, resulting in microsatellite instability [23].

The mismatch repair genes can be affected both by inactivating somatic mutations or promoter silencing by DNA methylation, resulting in transcriptional silencing [23]. Some cancer critical genes, such as tumor suppressor genes, have microsatellites in their coding sequence. If the mismatch repair system is impaired, these genes can be affected, potentially leading to tumor development [11]. Epigenetic inactivation by the important mismatch repair component *MLH1* cause microsatellite instability (MSI) in approximately 10-15 % of colorectal cancers [23]. MSI can happen in several cancer types, but is more common in colorectal-, gastric- and endometrial cancers [24-26]. Hereditary nonpolyposis colorectal

cancer (HNPCC) syndrome, also called Lynch syndrome, is an autosomal dominant inherited cancer syndrome, arising from a defect in the mismatch repair genes, such as *MLH1* or *MSH2*. The syndrome is therefore characterized by MSI and leads to an early onset of cancer [23]. Lynch syndrome is relatively rare, and only 1-2 % of the MSI tumors observed are due to inherited mutations in the MMR genes [25].

1.3 Epigenetics and regulation of gene expression

1.3.1 Definition of epigenetics

The word epi means “above” or “on top” in greek, and epigenetics refers to mechanisms that regulate gene expression without altering the DNA sequence [27]. The term “epigenetics” was originally described by Conrad Waddington, as “the branch of biology which studies the casual interactions between genes and their products which bring the phenotype into being” [28]. Another definition that is commonly used is

“*cellular information, other than the DNA sequence itself, heritable during cell division”[29].*

Epigenetics is considered the key to understand differentiation of cells and tissue specific gene expression in the body [29]. Epigenetic mechanisms include DNA methylation, post-translational modifications of histones, nucleosome positioning and non-coding RNA. They all work closely together to determine the state of chromatin and thereby regulating gene expression and decision of cell fate [30,31]. Epigenetic modifications are heritable by mitotic inheritance (from mother to daughter cells). It can also be inherited in a meiotic manner (between individuals in generations), but this happens rarely [32]. The main focus of this thesis is DNA methylation, but due to the interplay between the epigenetic modifications [33], other main epigenetic mechanisms will also be mentioned.

1.3.2 DNA methylation

DNA methylation is one of the best studied covalent modification of DNA and refers to the enzymatic adding of a methyl group to the 5-position of cytosine by DNA methyltransferases (DNMTs) [30]. This methylation happens primarily to cytosines that are followed by a guanine in the genome, so-called CpG sites [34]. CpG islands are defined as short

(approximately 1 kb) CpG rich regions with a minimum CpG ratio of 0.65 and a GC content greater than 55 % [35]. CpG islands are present in about 70 % of all mammalian promoters, and the human genome contains approximately 38 000 CpG islands [35]. In general, CpGs are underrepresented in the genome because of the tendency of methylated cytosines to undergo mutations from a CpG to a TpG. Unmethylated CpGs are on the other hand conserved, and they usually flank promoter areas of housekeeping genes, in the CpG islands [35].

Three DNA enzymes, DNA methyltransferases (DNMTs), are responsible for methylation of cytosine. As figure 5 shows, DNMT1 is primarily responsible for the maintenance of methylation after replication, and DNMT3A and DNMT3B can introduce new methylation marks *de novo* [36].

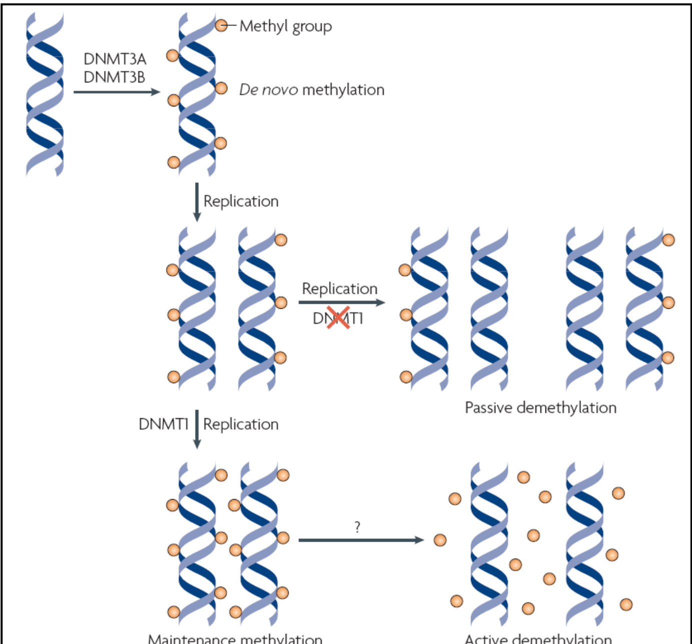


Figure 5: DNA methylation by DNA methyltransferases.
DNMT3A and DNMT3B are responsible for *de novo* methylation and DNMT1 for maintenance methylation after replication. Passive demethylation happens after replication in the absence of DNMT1. Active demethylation happens enzymatically, and until recently the enzyme catalyzing this process was not known. Now we know that the TET family enzymes have this role in the cell [37]. This is further described in the next paragraph.

DNA demethylation

The active removal of the 5-methylcytosine (5mC) mark was debated for a long time, while passive DNA demethylation was considered the consensus. DNA demethylation has now

been shown to happen in both a passive and an active way, and enzymes catalyzing active DNA demethylation have recently been identified [37]. Passive demethylation refers to the process of DNA replication in the absence of a functional DNA methylation maintenance machinery, leading to a successive depletion of DNA methylation in the genome [38]. Active demethylation is an enzymatic process, where the methyl group of 5mC is removed or modified [37].

A major breakthrough of understanding DNA demethylation was the discovery of the ten-eleven translocation (TET) family enzymes [38]. There are three known TET enzymes, TET1, TET2 and TET3 [38], and they catalyze the oxidation of 5mC to 5-hydroxymethylcytosine (5hmC) with a coupled oxidation of 2-oxoglutarate into succinate and CO_2 [39,40]. Figure 6 describes the detailed conversion of 5mC to unmethylated cytosine.

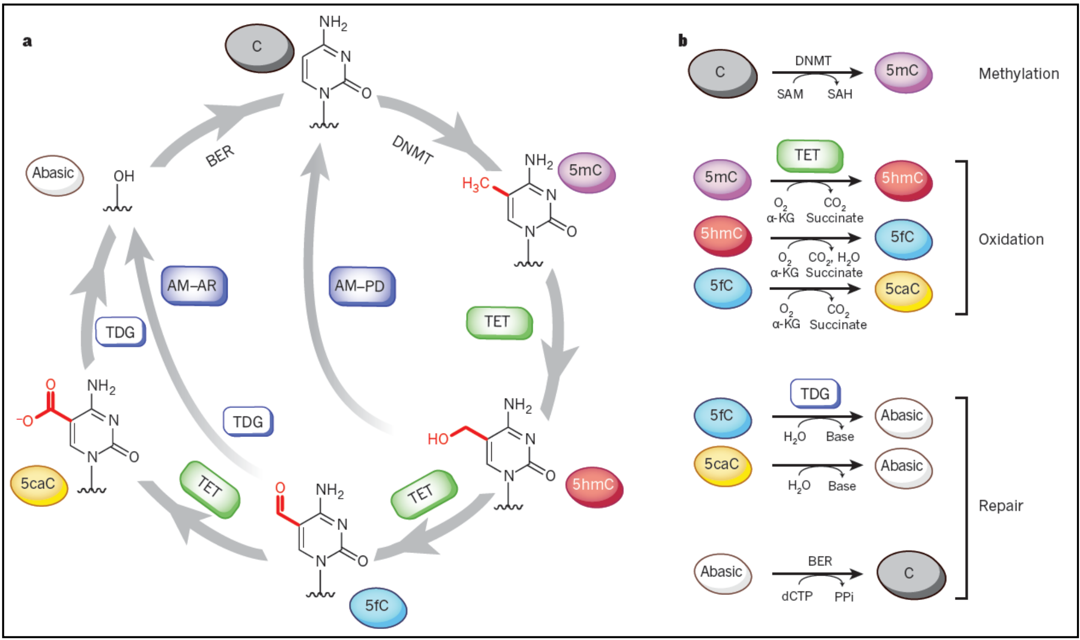


Figure 6: Demethylation of 5mC to 5hmC, and the function of TET enzymes [38].

a; The 5mC bases are introduced by DNMTs, and can be oxidized into the byproducts 5hmC, 5fC and 5caC. In the figure, 5mC are actively modified (AM) by TET to 5hmC, followed by passive dilution (PD), to unmodified cytosine. This passive dilution is replication-dependent. In the pathway of active modification (AM) followed by active restoration (AR), 5mC are firstly modified by TET to 5hmC, and then oxidized to 5fC or 5caC. 5fC or 5caC is excised in the base excision repair (BER) process, and then unmodified cytosine is the end product.

b; The individual reactions are shown. The BER pathway involved excision of the abasic site, and the nucleotide is replaced by an unmodified deoxycytidine triphosphate (dCTP) by a DNA polymerase and the nick in the strand is eventually ligated [38]. 5mC; 5-methylcytosine, 5hmC; 5-hydroxymethylcytosine, 5fC; 5-formylcytosine, 5caC; 5-carboxylcytosine, α-KG; α-ketoglutarate, SAM; *S*-adenosylmethionine, SAH; *S*-adenosylhomocysteine, BER; Base excision repair.

DNA methylation and regulation of transcription

DNA methylation plays an important role in transcriptional regulation, and active genes are generally associated with promoters containing unmethylated CpG islands, while repressed genes are generally associated with promoters containing methylated CpG islands [41]. DNA methylation is considered a long term and heritable mark, and is relatively stable compared to most histone modifications [38]. Recently it has been proposed that the position of the 5mC mark is as important as the mark itself. DNA methylation near transcription starts sites tends to block transcription while DNA methylation of CpGs present in the gene body on the other hand has no effect on gene silencing and has also been indicated to result in increased transcription [42]. It is debated whether DNA methylation is an initial silencing mechanism or if it is added to genes already silenced by other mechanisms, and it seems that the latter is most likely the case [42]. One of the evidences indicating this is a study by Ooi *et al*. where they researched *de novo* methylation in cells expressing DNMT3L (a catalytic inactive homologue of DNMT3A and DNMT3B). A complex consisting of DNMT3A2 and DNMT3L performs *de novo* methylation in these cells, and the study showed that the complex requires a nucleosome [43]. Actively transcribed genes contain transcription start sites depleted of nucleosomes, indicating that the gene already was silenced prior to methylation [42]. Further evidence, also found by Ooi *et al.* showed that *de novo* methylation did not occur on nucleosomes bearing histone marks associated with active genes [43].

DNA methylation is involved in modulating chromatin and repression of gene expression, either directly, by blocking the binding sites of necessary transcription factors or indirectly by recruitment of methyl-binding domain (MBD) proteins that further attract chromatin repressor molecules. Four such proteins have been identified, named MBD1-4 [44]. These binding domains serve as docking sites for histone deacetylases, histone methyltransferases and adenosine triphosphate (ATP)-dependent chromatin remodeling enzymes, (described in more detail in section 1.3.3). MBD proteins lead to a compact chromatin environment, and repression of transcription [35]. There are also zinc-finger proteins called Kaiso, ZBTB4 and ZBTB38 that recognize and bind to methylated DNA. These proteins exhibit repressive functions, and recruit histone deacetylases and proteins that form chromatin modifying complexes called polycomb group (PcG) proteins, commonly involved in transcriptional silencing in higher eukaryotes [44,45]. CXXC domain containing proteins are also known for binding to methylated cytosines in the CpG islands, and several proteins are known to have

these domains, including TET1 and TET3, previously mentioned to have a central role in demethylation [38].

DNA methylation, demethylation and normal development

Cell type-specific patterns of DNA methylation have been proven to be important in both normal development and disease [39]. During development of an organism, the cells undergo epigenetic and gene expression changes that give distinct cell types and tissue specificity that make up the organism [37]. The balance of methylation and demethylation are in this way extremely important for correct development and control of the cell, and malfunction of one or both of these processes can lead to disease and cancer in the organism [38]. Beside cancer there are a number of different developmental abnormalities linked to defects in the DNA methylation machinery, including brain abnormalities and Rett syndrome, a developmental malfunction [44]. In healthy cells DNA methylation contribute to genomic imprinting, X-chromosome inactivation and silencing of transposable elements [38]. DNA methylation also has an important role in repeat regions, such as centromeres, and this is important for chromosome stability [42].

The maintenance and *de novo* DNA methyltransferases are both important in setting up the DNA methylation pattern in early development, and all the three methyltransferases are required for embryonic and neonatal development [42]. Demethylation is also important during development and there are two major waves of demethylation in the genome in early stages of development. The first wave happens after merging of the sperm and the egg to a zygote, characterized as a rapid loss of 5mC is found in the paternal genome, while the maternal genome undergoes a gradual and passive form of demethylation (figure 7). A second wave of demethylation is seen in primordial germ cells, early in the development [37].

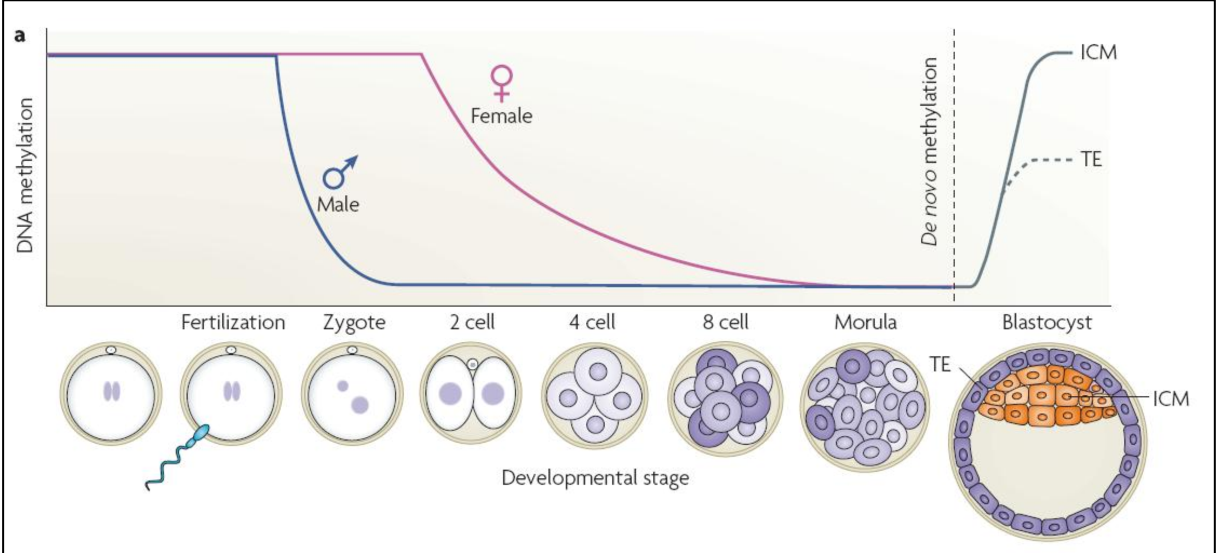


Figure 7: The first wave of demethylation during early development.

During the first wave of demethylation the paternal genome undergoes a rapid demethylated actively, while the maternal undergoes a passive gradual demethylation. ICM, inner cell mass; TE, trophoectoderm [37].

DNA methylation changes with age, and a global decrease in methylation is evident as a result of aging. This drift of methylation might be explained by replication-dependent errors in maintenance of the epigenome, and that repair of these is not uniform across all regions in the genome. Interestingly, this drift can also be linked to disease, such as cancer, and they can be accelerated by conditions such as chronic inflammation [46].

Imprinting

Imprinted genes differ from bi-allelically expressed genes by the fact that they are only expressed by one of the alleles inherited from the parents. Some imprinted genes follow a strict pattern, where only the maternal or paternal allele is expressed. Imprinted genes are important for embryonal development, and may also be involved in carcinogenesis if they are wrongly expressed [47]. One example is the loss of imprinting of insulin-like growth factor 2 (*IGF2*), observed in colorectal cancers [48]. Loss of imprinting of specific genes has also been reported in several other types of cancer, including leukemia, ovarian tumors and lung adenocarcinoma among others [47].

DNA methylation and demethylation in cancer

"The cancer epigenome" is defined as all the epigenetic alterations that occur in cancer [49]. Epigenetic abnormalities in human cancers was discovered in 1983, and it is now thought to have a equal role to genetic aberrations, opposite to earlier views [29]. There is a crosstalk between epigenetic and genetic aberrations in the sense that gene mutations can disrupt epigenetic patterns and modifications, and epigenetic modifications can result in genome instability and mutagenesis [50]. In tumorigenesis, epigenetics play a role in silencing tumor suppressor genes, activating oncogenes and generate chromosomal instability [32]. In the cancer genome there is a regional hypermethylation of specific CpG islands and a global hypomethylation [51].

Hypermethylation could potentially lead to silencing of important cancer-critical genes, such as tumor suppressor genes and genes encoding DNA repair genes [50]. It is thought that 5-10 % of normally unmethylated CpG promoters, are methylated in cancers [41]. Important genes frequently silenced by DNA methylation in cancer are *MGMT*, *BRCA1*, *MLH1* and *MSH2*. They are DNA maintenance genes, and as mentioned earlier associated with genome instability and increased frequency of point mutations in other cancer critical genes [50]. *APC* is another cancer critical gene, frequently inactivated by DNA methylation in malignant growth. APC is essential for controlling β-catenin, a protein migrating to the nucleus downstream of Wnt signaling and associating with the transcription factor Tcf/Lef, leading to increased proliferation and decreased differentiation. If β-catenin is not properly controlled, it could potentially lead to tumor growth, and APC is therefore an important protector against cancer development[11].

Global DNA hypomethylation was first observed in cancer in 1983, by Andy Feinberg and Bert Vogelstein [52,53]. Even though most CpG islands are unmethylated in somatic, healthy tissue, some CpG islands are normally methylated [29]. Feinberg and Vogelstein searched for differences between cancer cells and normal cells, and found that a substantial proportion of the CpGs methylated in normal cells were unmethylated in cancer cells. This loss of methylation was confirmed in other studies, and found to involve all tumor types studied, both benign pre-malignant tumors and malignant cancers. The studies showed that the hypomethylation occurred gradually with an increasing demethylation as the tumor progressed and it had an onset early in carcinogenesis [29]. Global hypomethylation is linked

to chromatin restructuring and the nuclear disorganization in cancer cells, and due to changes in histone-modifying enzymes and other chromatin regulators, chromosomal instability can be the result [31]. Oncogenes such as *HRAS* can become activated by hypomethylation of CpG islands, but this only happens to a few oncogenes [53]. The same is shown for testis/cancer antigenes, genes that are normally methylated and silenced in somatic cells, but demethylated and actively transcribed in testis germinal cells as well as in a range of cancers [29]. The results of both hyper and hypomethylation in the cell depicts the importance of correct regulation of both methylation and demethylation in the cell.

1.3.3 Chromatin

Chromatin can generally be observed in two different states, euchromatin which is an open state and contains actively transcribed genes, and heterochromatin containing condensed regions and inactive genes [41]. The nucleosome is the basic functional unit of chromatin, consisting of 147 base pairs of DNA, wrapped around a histone octamer consisting of H2A, H2B, H3 and H4. In addition there is also a linker histone, called H1 [41]. The histone octamer is the basis for packing of DNA in the cell, as shown in figure 8 [54].

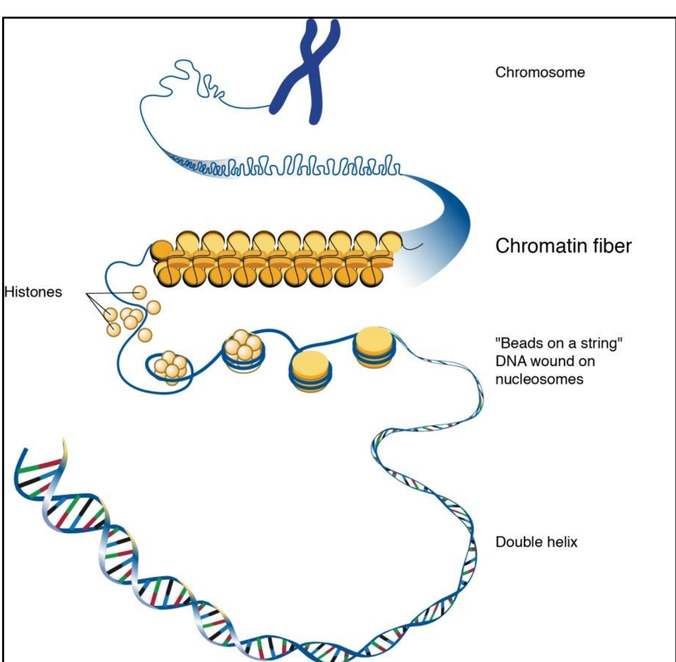


Figure 8: The packing and condensation of genes

The DNA is wrapped around nucleosomes, consisting of the histone octamer, and then further condensed into chromatin fiber[4].

[4] http://www.genome.gov/dmd/img.cfm?node=Photos/Graphics&id=85280

The state of chromatin (figure 8) is affected by DNA methylation as presented in section 1.3.2, and also modifications present on histones and chromatin remodelers [49]. It is the sum of the modifications and interplay between them, that decides if transcription is activated or repressed [33].

Histones are subjected to at several different types of post-translational modifications to their N-terminal tails [54], *i.e.* acetylation, methylation, phosphorylation, ubiquitylation and sumoylation [55]. The post-translational histone tail modifications can results in transcriptional repression or activation depending on modification type and location. Acetylation of lysines on histone tails, catalyzed by histone acetyltransferases (HATs), is an active mark associated with transcription. Methylation, catalyzed by histone methyltransferases (HMTs), can both lead to activation and repression, and examples are methylation of lysine 4 on histone 3, H3K4me, leading to active transcription, and methylation of lysine 27 on histone 3, H3K27me, leading to repression of transcription [54]. Histone deacetylases (HDACs) and demethylases (HDMs) have a role in removing the histone modifications [41].

"Reader" proteins bind to the histone modifications through special binding domains and they recruit additional chromatin modifiers and remodeling enzymes. Together they can lead to unfolding or condensation of chromatin, controlling transcription and elongation or repression of genes [41]. By attracting chromatin remodeling or chromatin modifying complexes, histone modifications have an indirect effect that may affect transcription initiation and/or elongation [56]. Chromatin remodelers, such as SWI/SNF and INO80, are ATP-dependent and capable of sliding, ejecting and inserting histone octamers [57]. Nucleosomes at the transcription start site, TSS, can be a hindrance of transcription by blocking transcription factors, and chromatin remodelers are responsible of removing these nucleosomes in actively transcribed genes [33].

Histone variants provide an additional level in the epigenetic code, and variants for all the histone family proteins, H1, H2A, H2B and H3 have been found [58]. Examples are H2A.Z and H3.3, histone variants mainly linked to open chromatin formation and active transcription. Another example is H2A.X, involved in DNA repair of double stranded breaks. Upon DNA damage, H2A.X is phosphorylated on serine 139 by a DNA-dependent protein kinase, recruiting the DNA repair machinery and chromatin remodelers. H2A.X also leads to a more relaxed chromatin structure, making the break sites more accessible [59].

1.4 Drivers and passengers in tumor development

The clonal evolution described earlier in the introduction leads to the accumulation of genetic and epigenetic aberrations in the cells during cancer development [1]. A "driver" lesion can be defined as an alteration, epigenetic or genetic, that gives the cancer cell a selective advantage in carcinogenesis. In contrast, a neutral or "passenger" lesion will give no such advantage [4]. The total number of epigenetic and genetic alterations in cancer varies significantly, as illustrated in a recent publication from Nature using high throughput sequencing data to compare genetic mutations from more than 7000 primary cancers of 30 different classes [60]. As figure 9 presents, the median number of mutations differed more than 200 times between melanomas (highest) and pilocytic astrocytoma (lowest).

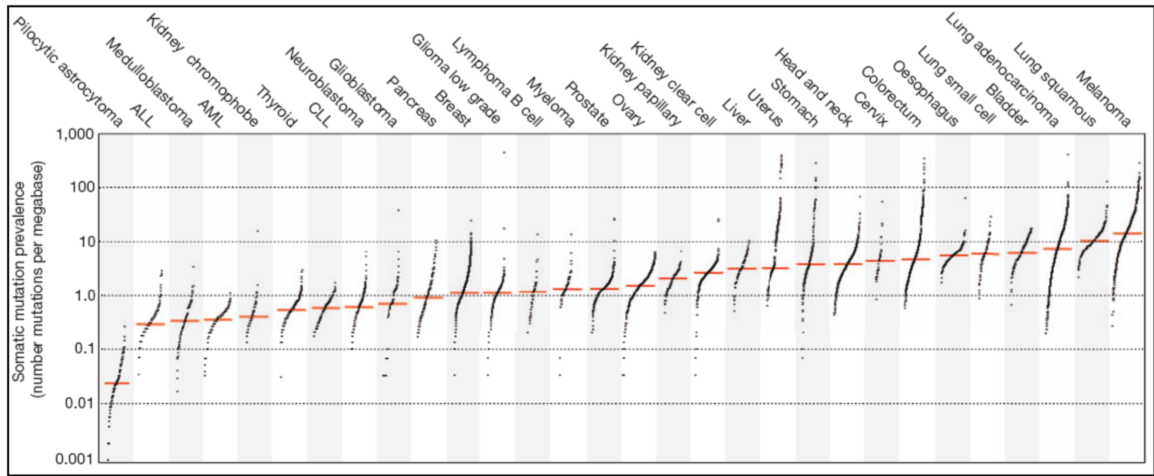


Figure 9: Somatic mutation prevalence across cancer types

The vertical axis (log scaled) shows number of mutations per megabase, while the different cancer types are shown in the horizontal axis, from lowest mutation prevalence to highest mutation prevalence. The red horizontal lines represent median number of mutations for each cancer type. The dots represent individual samples [60]. ALL, acute lymphoblastic leukaemia; AML, acute myelopid leukaemia; CLL, chronic lymphatic leukaemia.

Between two and eight drivers are enough for cancer to arise. This is relatively few compared to the amount of aberrations present in cancers, leading to the conclusion that the majority of cancer changes will be passengers [61]. Through the clonal evolution both the "drivers" and "passengers" are incorporated into the genome through different mutational processes or epigenetic modifications. In the previously mentioned Nature study, the authors suggested that the 'catalogue' (the total number and combination) of somatic mutations from a specific cancer bears "the signature of the mutational processes that have been operative". Different

mutation types or patterns, named cancer "signatures" are a product of these processes. These "signatures" can define cancer types and sub-types, but are frequently shared among cancer types. Most cancer types have more than one mutational signature, and they can occur in many combinations within the same cancer type [60].

There are several known cancer drivers, normally having an important role in the cell, and they can be classified as tumor suppressors or proto-oncogenes. Epigenetic inactivation of mismatch repair genes, such as the previously mentioned *MLH1* gene, resulting in microsatellite instable tumors is a good example of a driver that results in genetic alterations, leading to tumor growth [23]. Epigenetic inactivation of the tumor suppressor cyclin-dependent kinase inhibitor 2A (*CDKN2A*, also known as *p16INK4A*) has been seen in several cancer types. This protein has an inhibitory role in the cell cycle, and inactivation could lead to unhindered progression through the cell cycle, resulting in possible accumulation of further genetic aberrations [62]. The *APC* gene is as previously mentioned an example of a tumor suppressor gene being frequently inactivated by mutations or DNA methylation in colorectal cancers, leading to increased proliferation and decreased differentiation of the cancer cells [11]. About 90 % of colorectal patients have alterations that affect the function of APC [50].

The function of TP53 is described earlier in this introduction, and in short it is activated in response to cellular stress. Cells where TP53 is inactivated can go through cell division with DNA damage, and the loss of TP53 is a cancer driver, possibly leading to a more aggressive tumor growth [21]. Another frequent driver mutation in cancer is genetic alterations in the phosphoinositide 3-kinase (PI3K), turning it into the oncogenic form [63]. PI3K modulate several biological processes such as survival, proliferation, invasion, migration and metastasis [64]. Cancer drivers can be a target of therapy, and it will be further discussed in section 1.6.

Array-based technologies and Sanger sequencing have been the foundation for identifying the previously mentioned drivers in cancer, resulting in many important discoveries regarding the cancer genome and methylome. Now the focus has shifted to using high-throughput sequencing, such as transcriptome, exome, methylome and whole-genome sequencing [65,66]. There are some limitations regarding the classical array-based methods and Sanger sequencing, such as time spent compared to data output, and next generation sequencing techniques give promise to overcome some of these challenges [66]. But despite major advantages using next generation sequencing such as high output of data produced over a relatively short amount of time, there are also challenging perspectives, including converting

complex data to accessible and interpretable data [65]. Separating the true cancer drivers from the many passengers created in an unstable and evolving cancer genome is however still as challenging as with the more traditional methods [66].

Interestingly, from the next sequencing approach fewer recurrent mutations than expected have been identified in cancer [60]. However, several of these have been observed in genes influencing the epigenome [62]. Examples are *IDH1* and *IDH2*, isocitrate dehydrogenase 1 and 2, frequently methylated in gliomas, leukemias and several other cancers. When mutated, the IDH proteins produce a potential oncometabolite, 2-hydroxyglutarate (2-HG). This mutation has shown to alter the methylation landscape in gliomas, and is highly associated with the CpG island methylator phenotype (CIMP) [67]. Another example is mutations in DNA methyltransferase 3 alpha, *DNMT3A*, causing hypomethylation in acute myeloid leukemia, AML [62]. Mutations in genes like *IDH* and *DNMT3A* have the potential to significantly impact the epigenome, promoting cancer progression and driving tumor development.

1.5 DNA methylation biomarkers

On December 23, 1971, Nixon signed the National Cancer Act, and by this he declared war against cancer. He wanted to find a cure for the disease that was one of the leading causes of death in the US, and the rest of the world. Over 40 years after this declaration of war, cancer is still one of the leading causes of death, and the search for a cure is ongoing. The complexity of the disease, and the fact that cancer is over 200 different diseases, complicates the search for a cure. Now we might have to accept that the war may never be won, we have to coexist with cancer and direct the goal against primary prevention and control instead of a cure [68]. Biomarkers is an important weapon in primary prevention [69].

> *"Biomarkers are molecules or substances found in blood, other bodily fluids or tissues that reflect a particular biological or pathological state" [70].*

Biomarkers are used in different settings, including as diagnostic, prognostic and predictive markers. Only biomarkers used in diagnosis and early detection will be included here as they are relevant for this master project.

Diagnostic markers and early detection

A good biomarker requires high sensitivity and specificity, and ideally 100 % of both. Sensitivity refers to the proportion of sick people a particular test can detect, and specificity refers to the proportion of healthy people that test negative. High sensitivity is important since false negatives can provide a false safety for the patient, giving the cancer time to evolve further. High specificity avoid giving patients the burden of getting a cancer diagnosis without really being sick, and it is also cost-reducing since expensive additional analysis such as colonoscopy is not necessary [47].

In addition to optimal sensitivity and specificity an ideal DNA methylation biomarker should be safe for the patient, and preferably through minimal invasive approaches [71]. The biomarker assay needs to be user friendly and cost efficient to be successfully used in the clinic [72]. Early diagnosis is crucial in survival of cancer patients, and diagnostic markers could be used in early detection of cancer [58]. One example is colorectal cancer, where 5-year survival rates are up to 90 % for patients with cancer detected at a very early stage, compared to less than 10 % when the cancer has spread to distant organs [50]. Today, in the few countries that have established colorectal cancer screening programs, early diagnosis relies mainly on colonoscopy or a fecal based test [73]. Patients may have a tendency to avoid colonoscopy because of the discomfort, reducing the patient compliance, and there is also a minor risk of perforating the bowl during the procedure. Norway has not established such a screening program, and the disease may be discovered in later stages due to variability of symptoms. In the diagnosis of other cancer types, such as breast and lung cancer, imaging techniques are often used, and the downside of these techniques is the need for the neoplastic lesion to grow into a few millimeters in size before it can be detected [58]. An alternative and minimally invasive primary screening tool in detection of cancer would therefore be valuable. The biomarkers can potentially be used in screening to detect disease early, giving the possibility of starting treatment in the earliest phases of the disease, leading to better patient survival and less need for invasive treatment that could potentially lead to permanent trauma and side effects (figure 10) [47]. Many potential DNA methylation biomarkers have been identified, especially for colorectal cancer, but few of them have been implemented in the clinic. Some of the biomarkers shown to have potential will be mentioned as examples in the paragraphs below.

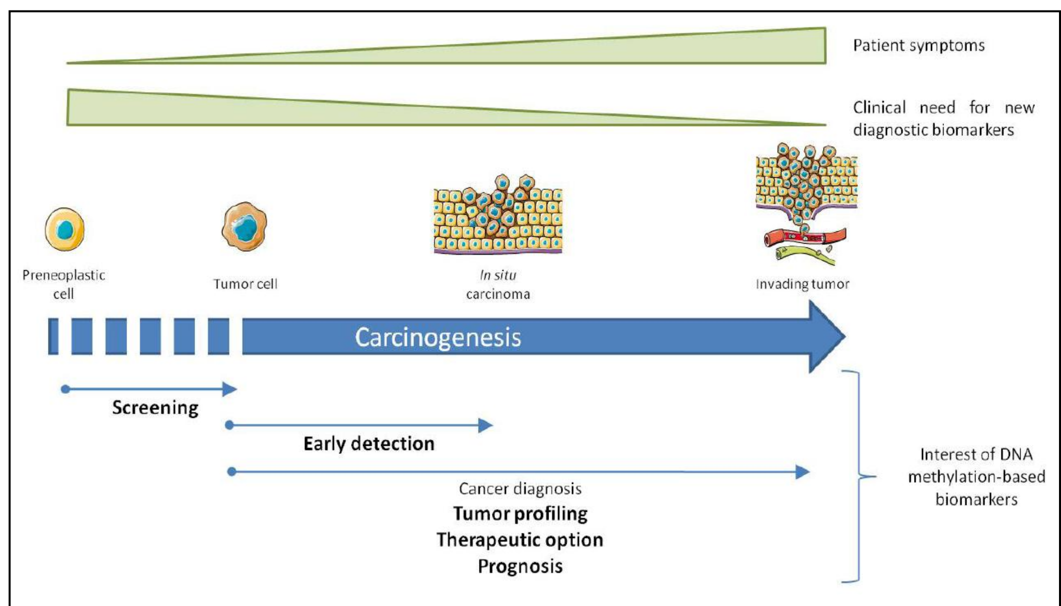


Figure 10: DNA methylation biomarkers and clinical use

This figure illustrates that the value and potential of noninvasive biomarkers is greatest early in tumor progression, because of the lack of symptoms, and therefore the high risk of overseeing the tumor. The biomarkers can potentially be used in screening or early detection, but also for prognostication, tumor profiling and even in a therapeutic setting [47].

The methylated septin 9 (*SEPT9*) biomarker obtained from plasma from colorectal cancer patients is the basis of several commercially available tests, such as ColoVantage®, Epi proColon and Abbott RealTime mS9 [74]. Several studies with large test series have examined the potential of this biomarker, and they reported a sensitivity of 60 to 70 %, with better detection of colorectal patients in later stages (stage III and IV). The test also provided around 10 % of false positives, and detection of non-malignant adenomas was low [75-77]. The low detection of non-malignant adenomas is a disadvantage with the use of *SEPT9*, since a detection of removable lesions before they turn malignant could prevent the potential development of colorectal cancer and thereby lower the incidence. In a study performed by Warren et al. the *SEPT9* test performed better than in previous studies [74]. They were able to detect 87 % of stage I and II colorectal patients, while it detected all the stage III and VI colorectal patients, leading to a sensitivity of 90 % in total. *SEPT 9* methylation was further detected in 11 out of the 94 control specimens, leading to a false positive rate of around 10 %, in accordance with previous reports [70,78].

The methylation status of vimentin (*VIM*), is forming the basis for ColoSure™, another non-invasive test used for early detection of colorectal cancer [74]. It is a commercially available,

fecal-based test with a reported sensitivity range of 53 to 83 %, and a reported specificity range of 71 to 89 % [79] and it should therefore be used in the combination with colonoscopy [50,74].

ColoGuard is a noninvasive multitarget stool test based on DNA and hemoglobin markers from Exact Sciences tested in one of the largest colorectal screening studies conducted in the Unites States. It has recently been approved by the U.S Food and Drug Administration (FDA) for use in screening[5]. The study had nearly 10 000 participants, 65 had colorectal cancer, while 757 had advanced precancerous lesions. The test reached a sensitivity of 92.3 % for detecting colorectal cancer in the patient series, and 42.4 % for detecting patients with precancerous lesions. The specificity was 86.6 % [80].

A novel epigenetic biomarker panel containing six genes (*CNRIP1*, *FBN1*, *INA*, *MAL*, *SNCA* and *SPG20)* has been developed in our department at The Radium Hospital. The biomarker panel showed a sensitivity of 94 % in colorectal cancer samples and 93 % in adenoma samples. The specificity was 98 %. The study was performed using 20 colorectal cancer cell lines and 523 human samples [81]. The results have been validated, concluding with the high potential of the novel epigenetic biomarker panel as a noninvasive test. These and six additional biomarkers were licensed to Oxford Gene Technology, OGT, in February 2012 for further development of a non-invasive test based on blood or stool samples [82].

1.6 Epigenetics and cancer therapy

The general acceptance that epigenetics play a major role in cancer development, and the reversible nature of such changes, have resulted in the search for epigenetic cancer therapy [83]. Epigenetic drug discovery is only just beginning, and there is still a long way to go before fully understanding the complexity of epigenetics and potential ways of reversing it [83]. The importance of epigenetics in cancer treatment is founded on the idea that agents modulating DNA methylation or other epigenetic marks can give an anti-tumor effect [74]. Epigenetic modifications are involved in drug resistance toward traditional drugs, such as 5-fluorouracil (5-FU), and a cocktail of both traditional and epigenetic drugs could be a strategy in fighting cancer in the future [74].

[5] https://www.exactsciences.com/

There are a number of challenges in epigenetic drug discovery. Many epigenetic targets are a part of large multi protein complexes, and it is difficult to predict the outcome of targeting one or several of these components *in vivo* [83]. The time period between a given epigenetic therapeutic agent and potentially visible effects could be extensive and some patients may not respond as efficiently to the drug, even though they have the same epigenetic alterations as the rest of the patient group [72]. In addition, the majority of studies have so far been performed in small cohorts, and the need for validation in larger groups is necessary [72].

Despite the challenges, some epigenetic drugs are already FDA approved and in clinical use [83]. When treating the cells with demethylating agents, such as DNMT inhibitors (figure 11) one might hope to reactivate essential genes, such as *CDKN2A*, *MLH1* and *RB1*, enabling the cell to re-gain control of cell proliferation, differentiation, apoptosis and other crucial homeostatic mechanisms [31].

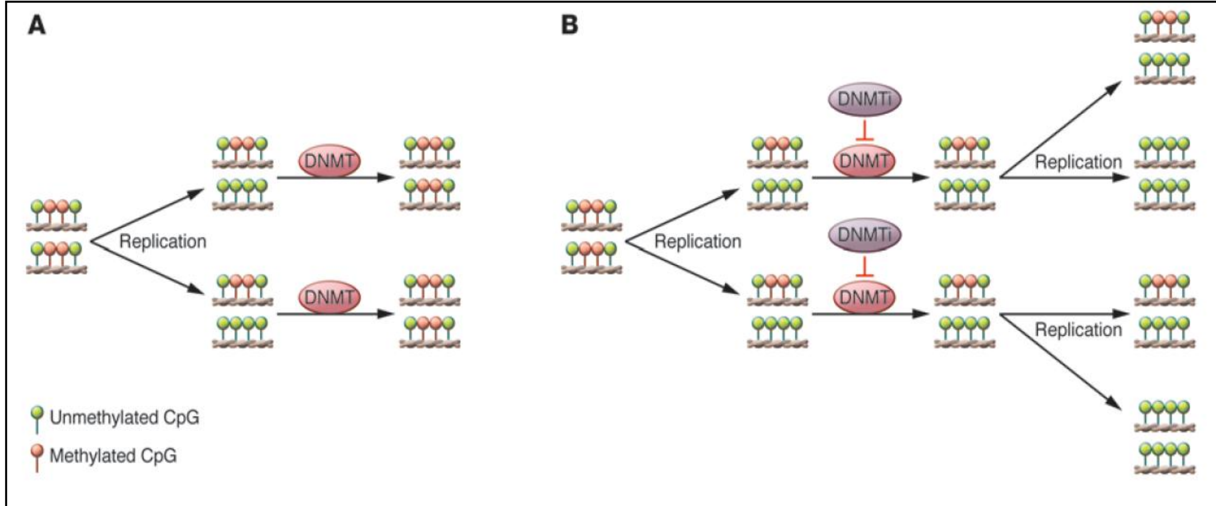


Figure 11: Function of DNMT inhibitors

DNMTs are responsible of maintaining the methylation pattern after replication. In figure A), the normal function of DNMTs is depicted. Figure B) illustrates the effect of DNMT inhibitors, leading to a passive demethylation in replicating cells [72].

Two inhibitors of DNA methylation, 5-azacytidine (AZA) and 5-aza-2`-deoxycytidine have been used to treat myelodysplastic syndrome, previously called preleukemia [27]. The azanucleotides function as a pyrimidine analog, and inhibit DNMTs [84]. Although this treatment has increased the survival rate of these patients [85], there are major disadvantages by these drugs, such as their non-specific nature potentially leading to hypomethylation of the entire genome, previously described as a potential benefit for malignant cancer development. The use of demethylating agents can also result in activation of a minority of proto-oncogenes

into oncogenes, giving the treated cells potential selective advantages when it comes to carcinogenesis and metastasis. In spite of these challenges, studies and clinical trials have shown an positive effect of these demethylating agents [31].

New demethylating approaches are emerging, and a recent innovative study has been published, using fusions of engineered transcription activator-like effector (TALE) repeat arrays and TET1 hydroxylase catalytic domain. As mentioned previously, the TET family proteins catalyze the oxidation of 5mC to 5hmC, a critical step in removal of the methyl mark. Figure 12 shows how the TALE is fused to the TET1 hydroxylase, providing a way of potentially accomplishing demethylation in a more site-specific and precise manner. The researchers demonstrated that they could modify CpG methylation at specific promoters resulting in increased gene expression [39]. By using this new technology it would be possible to avoid the general hypomethylation in the genome previously seen by general DNMT inhibitors.

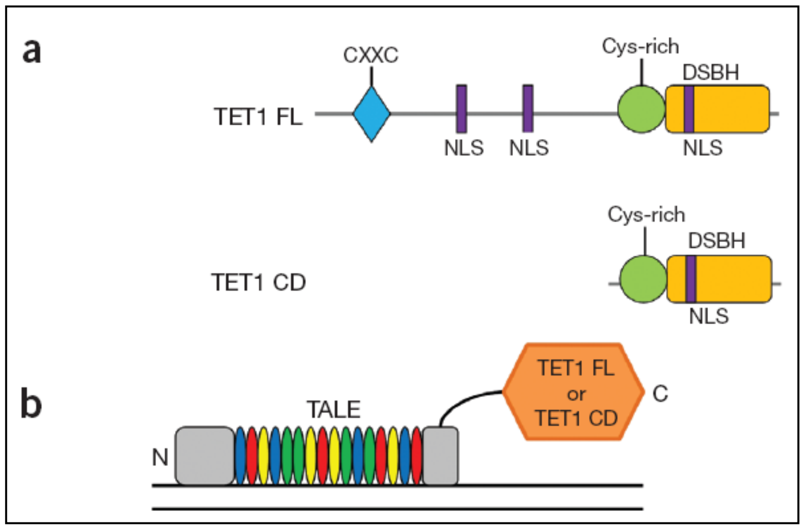


Figure 12: TALE repeat array DNA-binding domain fused with a TET1 protein

Figure a shows domain architecture of full length TET1 protein (called TET1 FL), and the catalytic domain of TET1 below (called TET1 CD). Figure b shows how TET1 FL or TET1 CD is fused with the C-terminal domain of TALE. Researchers in this study found out that TALE fused with TET1 CD resulted in significantly greater decrease of methylation in the CpG proximal to the TALE binding cite compared to the TALE fused with TET1 FL [39]. TET, ten eleven translocation; CXXC, CXXC-type zinc-binding domain; NLS, nuclear localization signal; Cys-rich, cystein-rich region; DSBH, double-stranded β-helix domain; TALE, transcription activator-like effector; FL, full length domain; CD, C-terminal domain.

In recent years, treatments targeting histone methylation and acetylation have also been used [83]. HDAC (histone deacetylation) inhibitors can be used individually, combined with DNMT inhibitors and/or together with traditional chemotherapy [72,74]. The most successful examples of HDACs inhibitors used in the clinic, which are also FDA approved, are romidepsin and vorinostat [86]. They are used to treat cutaneous T-cell lymphoma, but a disadvantage is that they may result in toxicity and only work transiently [27]. Additional HDAC inhibitors are currently being investigated in clinical and preclinical trials. HDAC inhibitors are thought to affect a range of processes, including tumor cell apoptosis, growth arrest, senescence, differentiation and inhibit angiogenesis, mechanisms crucial in fighting cancer [86].

Inhibition of *EZH2*, a component of the Polycomb machinery involved in selective methylation of the repressive histone mark H3K27, has been tested preclinically by several pharmaceutical companies. The results are promising, but several challenges remain in the development [83]. Other examples of epigenetic drugs targeting histone posttranslational modifications are inhibition of the DOT1-like histone H3K79 methyltransferase (*DOT1L)* responsible of catalyzing the active mark H3K79. DOT1L is found to be over-expressed and associated in a fusion product with the *KMT2A* gene (also known as MLL) in certain leukemias, leading to development and progression of the disease [83].

Despite promising potential of some epigenetic drugs, there is a need for more and better therapy that targets DNA methylation and histone tail modifications. Hopefully some of the many ongoing trials regarding biomarkers and epigenetic drugs will end in success [72].

2 Aims

This “Master of Science” assignment is a part of a large ongoing “Masterkeys” project at the department, outlined in figure 13 at page 29. The overall aim of the project is to:

1. Identify genes which are frequently inactivated by DNA methylation across multiple cancer types.
2. Analyze whether these genes may represent potential epigenetic drivers in cancer development.
3. Evaluate whether these or other genes from the approach represent biomarkers for cancer in general, or specific subset of cancer types.

The specific aims for the present “Master of Science” project are the following:

1. Use the processed cell line based epigenetic microarray data to generate a list of candidate genes for downstream promoter methylation analysis.
2. Design and optimize qualitative MSP assays for these candidate genes, and use them to analyze the promoter methylation status in a small panel of cell lines covering as many cancer types as possible.
3. Validate the most promising candidates using bisulfite sequencing and subsequent quantitative MSP, qMSP of all available cancer cell lines in our department (n=114).

3 Materials and Methods

3.1 Materials

3.1.1 Cancer cell lines

In the present study, DNA from an exceptionally large selection of 114 cancer cell lines from 17 different tissues has been used (Table 1). Using AmpFLSTR Identifiler PCR Amplification kit (Life Technologies) all commercially available cell lines have been authenticated. All cell lines have also tested negatively for mycoplasma infection.

Tissue	Cell line	Tissue	Cell line	Tissue	Cell line	Tissue	Cell line	Tissue	Cell line
Bile duct	EGI-1 (*) (**)	Breast	ZR-75-30	Colon	SW-620	Lung	A-549	Ovary	SK-OV-3
Bile duct	HuCCT	Colon	Caco-2	Colon	SW-948	Lung	NCI-H226	Pancreas	AsPC-1
Bile duct	KMBC	Colon	CO-115	Colon	TC-71	Lung	NCI-H23	Pancreas	BxPC-3
Bile duct	KMCH-1	Colon	COLO 205	Colon	V9P	Lung	NCI-H460	Pancreas	CFPAC-1
Bile duct	SK-ChA-1	Colon	COLO 320	Colon	WiDr	Lung	NCI-H522 (*) (**)	Pancreas	HPAF-II
Bile duct	TFK-1	Colon	DLD-1	Gall bladder	Mz-ChA-1 (*) (**)	Lymphoma	BL-41	Pancreas	PaCa-2 (*) (**)
Bladder	5637 (*) (**)	Colon	EB	Gall bladder	Mz-ChA-2	Lymphoma	K422	Pancreas	Panc-1
Bladder	HT-1197	Colon	FRI	Gastric	AGS	Lymphoma	NUDHL1	Prostate	22RV1
Bladder	HT-1376	Colon	HCC 2998	Gastric	KATO III	Lymphoma	OCILY10	Prostate	DU 145
Bladder	J82	Colon	HCT-116 (*) (**)	Gastric	NCI-N87 (*) (**)	Lymphoma	OCILY19	Prostate	LNCaP (*)
Bladder	RT4	Colon	HCT-15	Gastric	SNU-1	Lymphoma	OCILY2	Prostate	NCI-H660 (*)
Bladder	SCaBER	Colon	HT-29	Gastric	SNU-5	Lymphoma	RAJI	Prostate	PC-3
Bladder	SW 780	Colon	IS-1	Kidney	786-O	Lymphoma	ROS-50	Testis	2102Ep
Bladder	T24	Colon	IS-3	Kidney	ACHN	Lymphoma	SC-1	Testis	NCCIT
Bladder	TCCSUP (*) (**)	Colon	KM-12	Kidney	Caki-1 (*)	Lymphoma	SUDHL4 (*)	Testis	NTera2 (*)
Bladder	UM-UC-3	Colon	LOVO	Kidney	Caki-2	Lymphoma	U2932	Testis	TCAM-2
Breast	BT-20	Colon	LS-174T	Leukemia	697 (*)	MPNST	642	Testis	Tera-1
Breast	BT-474	Colon	LS-1034	Leukemia	RCH-ACV	MPNST	S1507-2	Testis	Tera-2
Breast	Hs 578T	Colon	NCI-H508	Leukemia	REH	MPNST	S462	Uterus	AN3 CA
Breast	MCF-7	Colon	RKO	Liver	Hep-G2	MPNST	STS26T (*) (**)	Uterus	HEC-1-A
Breast	SK-BR-3	Colon	SW-1116	Liver	JHH-1	Ovary	ES-2	Uterus	KLE
Breast	T-47D (*) (**)	Colon	SW-48	Liver	JHH-4	Ovary	OV-90	Uterus	RL95-2 (*) (**)
Breast	ZR-751 (*)	Colon	SW-480 (*) (**)	Liver	JHH-5 (*) (**)	Ovary	Ovcar-3		

Table 1: Cancer cell lines used in this study

Cell lines marked with (*) (n=20) were used during the early screening of the candidate genes with Methylation specific PCR, MSP method. Cell lines marked with (**) were used during validation of novel candidate genes by bisulfite sequencing (n=13) and all the cell lines in the table (n=114) were used during quantitative MSP, qMSP. The various methods are described in the next section 3.2.

3.2 Methods

3.2.1 *In vitro*, *in silico* and *in vivo* strategy for identifying DNA methylation candidate genes

The pipeline of the "Masterkeys project" as a whole and the strategy to select the best candidate genes for this Master of Science project is presented on the next page in figure 13. The figure is divided into two parts, the upper part shows methods performed in house prior to this "Master of science" project, and the lower part is an overview of the work carried out in this thesis. Both parts will be explained in more detail in the following pages. For a clearer outline of the project, the figure contains the number of genes selected from each experimental step (I to VIII). However, these will be further described in the results part (page 41 - 51).

Step I (figure 13): cDNA microarrays (Life Technologies, 1700 microarray platform and the Human Genome Survey Microarray V2.0 containing 32.878 unique 60-mer oligo-nucleotide probes) were used to investigate gene expression of cancer cell lines before and after treatment with the 5-aza-2`deoxycytidine (AZA; 1µmol/L for 72 hours) and trichostatin A (TSA; 0.5µmol/L in the last 12 of the 72 hours). AZA is a demethylating agent that impairs the DNA methylation machinery, resulting in passive DNA demethylation during cell replication. TSA inhibits class I and II mammalian histone deacetylases. Genes containing a promoter CpG-island originally silenced by DNA methylation and potentially tightly packed due to histone deacetylation could respond with an up-regulation in expression after treatment with AZA and TSA. Genes responding in this way represent candidate genes for DNA methylation biomarkers.

Step II: In order to decrease the number of genes on the initial treatment response list (step I) and simultaneously increase the likelihood of selecting good DNA methylation candidates, two different bioinformatic filters were used. The "Stringent filter" required a 2-fold up-regulation of the genes after TSA and AZA treatment. A t-test was performed and only those genes that were statistically significantly ($P \leq 0.05$) different between the treated and the untreated cell line groups were kept, resulting in altogether 830 genes. In the "exploratory filter" no fold change limit was used and all genes with higher expression after treatment were considered to be up-regulated. However, only candidates up-regulated in all the cancer tissue groups investigated (colon-, gastric- etc cancer) were kept for further analysis (n=1079).

When comparing the gene lists from the "stringent" and "exploratory filter" 468 genes were overlapping and kept as interesting candidate genes.

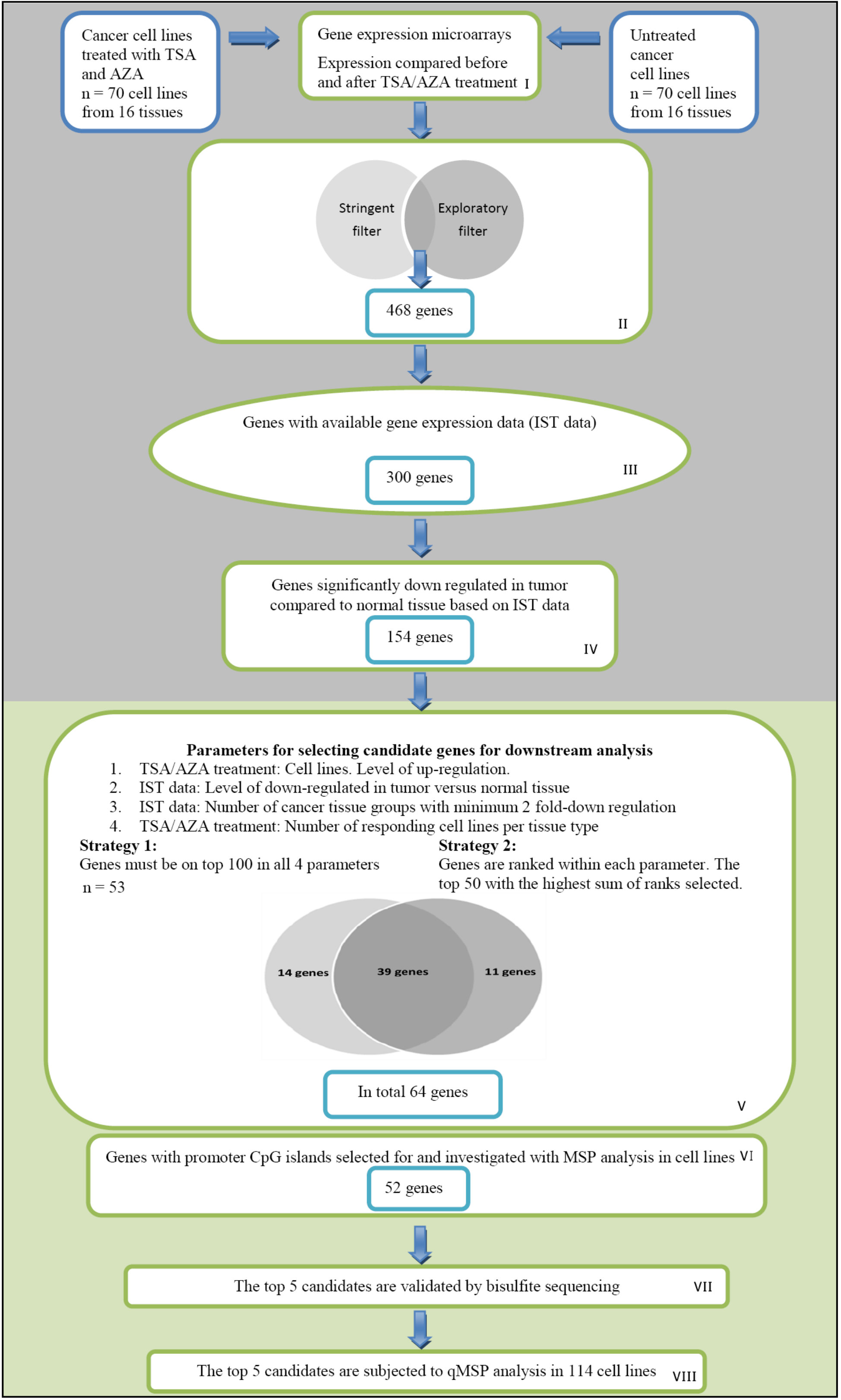


Figure 13: Outline of the "Masterkeys project" as well as the present Master of Science project

Individual steps are explained in the following text.

Step III: To further increase the likelihood of selecting the correct genes for downstream DNA methylation analysis, gene expression was also taken into account. For all genes from step II information from the *In Silico* Transcriptomics (IST) database[6] was collected. This is a large database based on data from Affymetrix arrays where the gene expression of tumor samples are collected and compared to normal samples. Only genes with expression values registered in the IST database were kept for further analysis.

Step IV: Data from the IST base was used to evaluate whether genes of interest were down-regulated more than two fold in tumor samples compared to the normal control tissues. To limit the effect of potential outliers on the selection, the median expression was chosen for this purpose. The 70 cell lines studied were divided into ten 'cancer tissue groups', according to their tissue type of origin. Unfortunately, IST data for testicular germ cell tumors, lymphoma and leukemia could not be used in this step, since data from representative controls were lacking.

Step V: In order to further narrow down the gene list from step IV, and select strong candidates for DNA methylation analysis, without setting too strict criteria and risk loss of valuable candidates, we combined two selection strategies. The parameters used in both strategies (as mentioned in figure 13) were:

- TSA/AZA treatment: Level of up regulation in cell lines.
- IST data: Level of down-regulation in tumor versus normal tissue.
- IST data: Numbers of cancer tissue groups with minimum 2 fold down-regulation.
- TSA/AZA treatment: Number of responding cell lines per tissue type.

Each of the 154 genes from step IV was ranked individually based on how they performed within each of these four parameters listed above, and in strategy 1 a requirement was set that the genes selected should be on the top 100 list for all the four parameters. In strategy 2 each gene was ranked based on how well it did according to each individual parameter. Individual ranks (for all 4 parameters) were summarized to give each gene a final score. Promising candidates were characterized by low final score, indicating that the gene was frequently ranked in the upper parts of the lists of the various parameters. The 50 genes with the lowest final score were selected for further analysis.

[6] http://www.genesapiens.org

The remaining steps (VI-VIII) will be explained in the sections that follow. The results from each step are further summarized in the Results part (pages 41-51).

3.2.2 Sodium bisulfite modification

The bisulfite modification makes it possible to separate methylated cytosines in the DNA from unmethylated cytosines. The principle behind this treatment is that the bisulfite will deaminate the unmethylated cytosines, while the methylated cytosine remains unaltered because it is protected by the methyl group. The result of the deamination by bisulfite is that the unmethylated cytosines are converted to uracil, which during further PCR is amplified as thymine. It will now be possible to separate methylated cytosines from unmethylated [87,88].

The Epitect Bisulfite Kit from Qiagen (Qiagen Co., Valencia, California, USA) was used for the bisulfite modification of the cell lines and controls in this thesis. DNA (1.3 µg) was mixed with RNase free water, bisulfite mix and DNA protection buffer. The DNA protection buffer ensures correct pH for complete and successful conversion and has an indicator color that makes this easy to visualize, by turning blue at correct conditions after mixing.

A thermo cycler (MJ Mini Personal Thermal Cycler, Bio-RAD, Hercules, CA, USA) was used for the bisulfite conversion reaction. A series of denaturation (95ºC) and incubation (60ºC) steps (n=3) ensures denaturation of DNA, followed by sulfonation and cytosine deamination. A full denaturation of the DNA is important because the reaction is single strand specific. If the DNA is not fully denatured, conversion may not be complete, potentially contributing to misleading results in the following analysis. Unmethylated cytosines that are not fully converted might be misinterpreted as methylated, leading to false positives.

The next step is desulfonation and clean-up of the bisulfite converted DNA. The QIAcube (Qiagen) is used for this purpose, and it removes bisulfite salts and chemicals. This is important to enable a successful PCR, and also ensure correct results during sequencing procedures. The DNA is eluted in 40 µl elution buffer, and the final concentration is theoretically estimated to be 32.5 µg/µl.

Several factors are important during bisulfite conversion. The DNA must be of high quality, and denaturation must be complete, as previously mentioned. The pH and temperature also needs to be optimal during the various steps. Bisulfite can oxidize in contact with oxygen, and

a free radical must therefore be present in the reaction mixture to prevent oxidative degradation. Even though the estimated rate of conversion is 99 %[7], a conversion rate of 95-98 is more likely.

3.2.3 Methylation specific polymerase chain reaction, MSP

Qualitative methylation-specific polymerase chain reaction, MSP, is a relatively simple and fast way to determine the promoter methylation status of a given gene [89]. MSP is a sensitive method, which can separate one methylated allele from 1000 unmethylated alleles [89]. Following the *in silico* analysis of the experimental pipeline, a panel of 20 cancer cell lines (marked in table 1, page 27, with *) from 16 different tissues was used as an initial screen to pinpoint frequently methylated candidate DNA methylation biomarkers worthy of analysis in the full cancer cell line panel.

Primer design

Primer design is essential for reliable MSP results, and two primer pairs are necessary, one pair annealing to methylated and one pair annealing to unmethylated DNA. The primers are designed to anneal around the transcription start of the gene, since methylation in this area potentially has a higher ability to repress the transcription. The primers cover several CpG sites, with a minimum of two CpG sites per primer. Preferably, there should be a C in a CpG present at the 3`end of the primer to increase discrimination of unmethylated and methylated DNA. Non-CpG cytosines should also be included in the primer, to ensure that potentially unconverted bisulfite treated DNA are not amplified and misinterpreted as methylated, as previously mentioned. The PCR product should preferably be as short as possible in order to increase the likelihood of also amplifying potentially challenging template, such as DNA from formalin fixed, paraffin embedded tissue which often experience fragmentation due to chemically induced double strand breaks.

The primers for the present study were designed using the Methyl Primer Express software from Life Technologies. Prior to primer design the genome browser (genome.ucsc.edu)[8] was used to download correct reference sequences, (RefSeqs), and the promoter area (1000 bp upstream and 500 bp downstream of the transcription start point) was analyzed using the CpG

---

[7] http://www.qiagen.com
[8] http://genome.ucsc.edu/

Island Searched Software[9] to determine the presence of a potential CpG island (Step VI in figure 13). If the gene had several alternative Refseqs, and the promoter area differed between them, primer pairs for all the different Refseqs were designed.

Optimization of MSP reactions

The conditions for primer pairs annealing to unmethylated and methylated sequence were optimized separately. DNA from normal blood (NB) was used as a template for primers annealing to unmethylated sequence, whereas human placenta DNA treated *in vitro* with SssI methyltransferase (IVD) was used as template for primer pairs annealing to methylated sequence. The MSP reactions were optimized regarding annealing temperature, annealing and elongation time as well as magnesium concentration (see below). A temperature gradient of 48 - 59ºC was used as a starting point for all reactions. Depending on the melting temperature of the primer pairs, temperatures in the mid-specter of this gradient were expected to be the optimal. Too low temperatures will give unspecific PCR products, whereas too high temperatures will result in weak PCR products. Magnesium is a cofactor for the DNA polymerase, and even though the manufacturer ensures that the enzyme is already optimized with the right amount of magnesium, experience has shown that additional magnesium may be necessary for optimal PCR results for specific genes. Our intention is to produce PCR bands of similar intensity from both the unmethylated and methylated reaction, which will make the visual scoring of final results easier (and probably also more correct). A magnesium gradient of 1.5, 1.7 and 2.0 mM was tested for the primer sets that were not equally efficient after the first round of temperature optimization.

Annealing and elongation lengths of 30 seconds each was our standard starting point for all reactions, as well as 35 PCR cycles. However, in some cases increasing these parameters led to better results, and for some primer pairs it was necessary to run one of the primer reactions (U or M) a couple of cycles longer, to make the resulting gel band intensities as comparable as possible.

Experimental procedure

Step VI (of figure 13): Using MSP the promoter methylation status of 52 genes was determined, in 20 different cell lines from 16 different cancer tissue types (illustrated by a * in

[9] http://cpgislands.usc.edu/

Table 1, page 27). The MSP reaction mix consisted of 0.75 µl bisulfite treated template DNA, 2 µl of each primer, forward and reverse (10 mM, Bionordika), 2.5 µl 10x Qiagen PCR buffer with 1.5 mM magnesium, 0 – 0.5 µl Qiagen Magnesium solution (25 mM), 2 µl dNTP mix (4 x 2.5 mM) and 0.2 µl Qiagen HotStar Taq Polymerase. Milli- Q water (15.55 µl) was also added, making the total reaction volume 25 µl. The DNA Engine Tetrad 2 (Bio-Rad, Cambridge, UK) was used to perform the PCR cycles. The initialization step at 95ºC for 15 minutes activates the thermo stable enzyme, followed by cycles of denaturation, annealing and elongation. Denaturation of the DNA strands are ensured by 95ºC for 30 seconds, the annealing step is at 48-59ºC for 30 seconds (depending on the melting temperature of the primer pairs), followed by elongation at 72ºC for 30 seconds. At the end the program there is a final elongation step for 7 minutes at 72ºC.

After PCR the products were separated using gel electrophoresis. A 2 % agarose gel stained with ethidium bromide 0.07 % (VWR, Pennsylvania, USA) was used. The 25 µl sample, mixed with 5 µl gel loading buffer (1 x TAE buffer and 0.1 % xylen cyanol), was loaded on the gel. The gel was run at 200V for 25 minutes. An UV trans-illuminator (Chemidoc XRS Gel Documentation System, Bio-Rad, Cambridge, UK) was used to visualize the PCR products.

Scoring results

When scoring the MSP results, the band intensity of the positive controls (NB for the unmethylated reaction and IVD for the methylated reaction) were compared with the band intensity of the samples. All samples and controls were scored according to an intensity scale from 0 to 5, where 0 represents no band and 5 is a very strong band (figure 14).

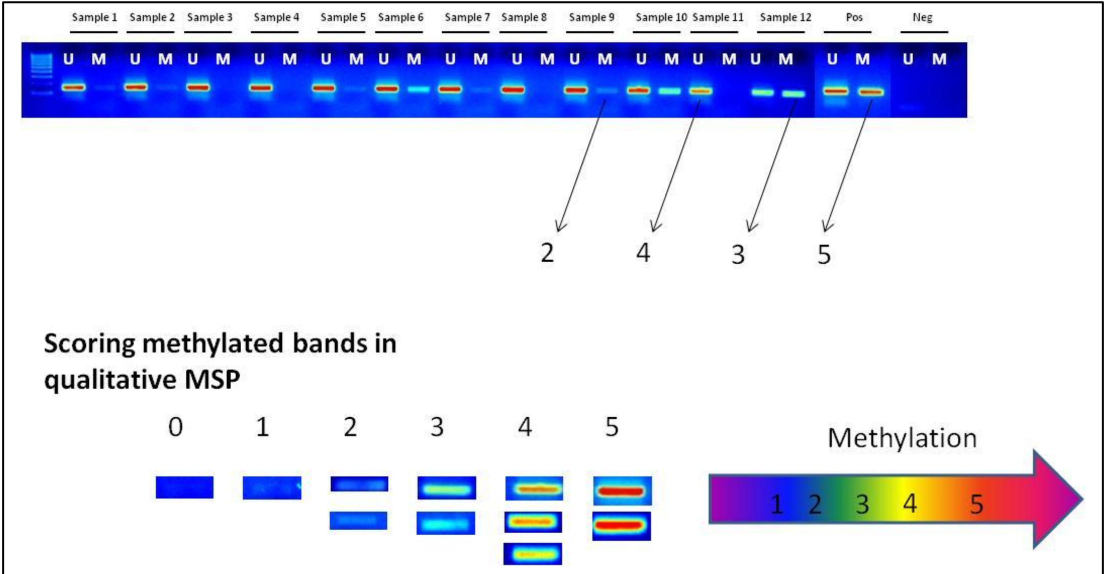


Figure 14: Scoring of bands after MSP and electrophoresis

A program in the GelDoc is used to 'translate' the bands intensities into on a colored scale from 1 to 5. The figure shows some examples of different samples and positive and negative controls. The different intensities are shown in the bottom of the figure.

A band is scored as methylated if the intensity of the sample band divided with the positive control band (IVD) is between 0.4 and 1.25. A band is scored as unmethylated if the intensity of the sample band divided with the positive control (NB) is between 0.4 and 1.25. If there are both a methylated and an unmethylated band present for the same cell line, the sample is scored as hemi methylated. The scorings were performed individually by the author and a postdoc in the group, Kim Andersen. If the scoring results should differ after two test runs, a third MSP reaction was performed. The results that appeared in two out of three runs were considered as final results.

3.2.4 Bisulfite sequencing

Bisulfite sequencing is an effective technique for detecting 5-methylcytosines, and it is widely used [90]. The method is based on traditional Sanger sequencing [91], and can detect any methylated cytosine at single base resolution in any sequence context [92]. In this method, bisulfite treated genomic DNA is used as a template. Because of the bisulfite treatment, 5-methyl cytosine will appear as cytosines, while unmethylated cytosines will appear as thymines in the resulting electropherograms.

Section VII of figure 13: The five candidates that were most frequently methylated across the 20 cancer cell lines as assessed by MSP analysis in step VI were subjected to bisulfite sequencing to both confirm the methylation status, and also to guide the design of the qMSP assay. A representative selection of 13 cell lines (labeled with ** in table 1, page 27) was used for the bisulfite sequencing, and NB and IVD were included as controls.

Primer design and optimization

Bisulfite sequencing primers are designed to avoid annealing to CpG sites, to ensure equal amplification of both methylated and unmethylated alleles. In some cases, CpG sites cannot be avoided, for instance due to a CpG island with many CpG sites. If so, primers are designed in a way that methylated and unmethylated target ideally should be amplified with equal efficiency. This should obviously be tested, and are done so routinely in the lab. Due to the bisulfite conversion, long stretches of thymines are frequently seen in the template DNA. This should preferably be avoided in the PCR, since the DNA polymerase will have problems correctly replicating more than 8 equal residues in a row due to slippage. The primers should preferably flank the MSP primers, in order to investigate the methylation status in the region of interest, and preferably have a distance to the MSP primers with about 50 bp on each side, due to the general challenge with low quality sequence reads in the beginning of a product. The primers were designed using Methyl Primer Express software (Life Technologies).

All PCR reactions were optimized according to temperature, and with magnesium if necessary. The temperature gradient was set up with temperatures between 48 and 59ºC. NB and IVD were used as a template, and the temperature that amplified both NB and IVD with equal efficiency were chosen.

Initial PCR

The initial PCR was set up in the same way and with equal amounts and concentration of template, dNTP mix, Qiagen buffer, milli-Q water and Qiagen HotStar Taq polymerase as the MSP reactions. Sequencing primers (10mM, Bionordika), 2 µl of each, were also included, and the total reaction volume was 25 µl. The PCR reaction was performed on DNA Engine Tetrad 2 (Peltier Thermal Cycler, Bio-Rad), and the PCR reaction steps were identical to the ones performed for MSP. After completion of the initial PCR, 5 µl of the products, together

with 1 µl loading buffer, were visualized on a 2 % agarose gel to confirm an amplification of the products.

Purification of PCR products

Five µl of the product after initial PCR were purified by EXOSAP-IT (GE Healthcare). The purpose of this is to rinse the PCR products for unincorporated dNTPs and primers. EXOSAP-IT contains two hydrolytic enzymes: exonuclease I and shrimp alkaline phosphatase. The PCR products were incubated at 37ºC for 15 minutes before the enzymes were inactivated at 80ºC for 15 minutes. The procedure was performed in a DNA engine Tetrad 2 (Peltier Thermal Cycler, BIO RAD).

Sequencing reaction

The sequencing reaction was completed using dGTP BigDye Terminator v3.0 Ready Reaction sequencing Kit (Life Technologies). The reaction mix contains AmpliTaq DNA polymerase and fluorescent labeled dideoxynucleoside triphosphate (ddNTPs). The reaction was set up with 2 µl purified PCR product, 0.10 µl forward or reverse primer (10 µM), 2 µl 5 x Big Dye Terminator v1.1 Sequencing Buffer (Life Technologies), 2 µl Ready Reaction mix and adjusted with RNase free water to adjust to total volume of 10 µl. The reaction was performed in a DNA engine Tetrad 2 (Peltier Thermal Cycler, BIO RAD), and the program used for the reaction included an initiation 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

The sequencing products were purified using Big Dye Xterminator (Life Technologies). The purification removes excessive dNTPs and primers. Ten µl Big Dye Xterminator solution was added to each sample, along with 45 µl Sam. The sequencing products were then vortexed in 30 minutes.

Capillary electrophoresis and analysis of sequencing elecropherograms

The sequencing products were separated according to size by capillary electrophoresis on ABI PRISM 3730 Sequencer (Life Technologies). A laser beam within the instrument excites the fluorescent ddNTP end label attached to the sequencing products, and the generation of

electropherograms was completed using the Sequencing Analysis 5.2 (Life Technologies). The resulting electropherograms were read manually, and the amount of methylation for each individual CpG site was calculated by measuring the peak of the cytosine signal, and dividing it with the sum of the cytosine the thymine signal in the same position. The resulting values ranged between 0-1, and were then multiplied by 100 to give percentage of methylation. CpG sites with methylation percentage between 0 and 20 were considered unmethylated, sites with a percentage between 20 and 80 were considered hemimethylated, and sites with a percentage above 80 were considered hypermethylated.

3.2.5 Quantitative methylation-specific polymerase chain reaction, qMSP

Quantitative methylation-specific polymerase chain reaction, qMSP, also called real-time MSP and MethyLight, is a method used for determining the amount of methylated target sequences present in a sample [93,94]. By using a standard curve created by serial dilutions of a sample of known concentration, concentrations of unknown samples can be determined by interpolating their quantity from the standard curve. In the exponential growth (log) phase of the PCR, the quantity of the PCR product is directly proportional to the amount of template nucleic acid and data should be collected during this phase. Two values are needed, 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. The PCR cycle where the threshold line is reached, is called the Ct value.

The qMSP assay includes both forward and reverse primers, and a fluorescently labeled probe binding to the methylated DNA sequence. The probe used had a fluorescent reporter dye attached to the 5`end, and a non-fluorescent quencher at the 3`end. In addition the probe also has a minor groove binding domain (MGB), attached at the 3`end. When the probe is intact the quencher reduces the fluorescence emitted by the reporter, but during primer extension Taq DNA polymerase cleaves the probe. This separates the reporter dye from the quencher. The result is fluorescence that can be captured by a laser detector.

Primer design

The primer design were performed using Methyl Primer Express software (Life Technologies) and adjusted using The Primer Express Software 3.0 (Life Technologies)

which also was used to design the probe. Both the primers and probe were designed to cover as many, and at least 10, CpG sites as possible. As with the MSP primers, they should also contain unmethylated cytosines to ensure primer specificity for bisulfite converted template. The minor binding domain of the probe raises the melting temperature (Tm) of the probe, making a shorter probe possible. This is an advantage because a shorter probe gives better discrimination. The Tm of the probe should be approximately 10ºC higher than for the primers, to ensure probe binding prior to primer annealing. The primers were purchased from BioNordika, while the probes were purchased from Life Technologies.

Experimental set-up

Step VIII (figure 13): Five genes were investigated with the qMSP reaction using 114 cell lines from 17 different tissues. Ovarian cancer cell lines were included in addition to the 16 tissues used in the initial microarray-based pipeline. Bisulfite treated NB, non-treated NB and RNase free water was used as a negative control, while bisulfite treated commercially available *in vitro* methylated DNA, IVD (Chemicon) was used as positive control. The IVD was used to make a five point standard curve from 1:5 dilutions (32.5 – 0.052ng).

Experimental procedure

The reaction was carried out in a 384-well plate. Amplifications were carried out in triplicates, and in a reaction volume of 20 µl. The reaction mix was consisting of 0.18 µl of each primer (100 µM), 0.40 µl probe (10 µM), 10 µl 1xTaqMan Universal PCR Mastermix No AmpErase UNG (with AmpliTaq Gold RNA polymerase, Life Technologies), 6.24 µl RNase free water (Sigma-Aldrich), and 3 µl bisulfite treated template DNA (approximately 32.5 ng). The EpMotion 5075 pipetting robot (Eppendorf, Hamburg, Germany) was used to automatically distribute template and master mix in the 384-well plates. The qMSP reaction was performed in a 7900HT fast Real-Time PCR system (Life Technologies). The reaction program consisted of a first denaturation step at 95ºC for 10 minutes, followed by 45 cycles of 95ºC for 15 seconds and then 60ºC for 60 seconds.

Calculation of methylation

From the qMSP analysis, the median quantity value was used to calculate the degree of methylation. All sample replicates amplified after cycle 35, were censored according to the recommendations from the manufacturer (Life Technologies). To normalize the DNA input

the ALU-C4 element was used as reference [95]. The positive control, IVD (Chemicon), was used as methylated reference, and the results were calculated as a percent of methylated reference (PMR) by dividing the normalized quantity of the samples by the normalized quantity of the IVD. This number is then multiplied by a 100 to get percentage.

$$PMR = \underbrace{(Qty_gene/Qty_ALU)_{sample}}_{\text{Normalized sample quantity}} / \underbrace{(Qty_Gene/Qty_ALU)_{IVD}}_{\text{Normalized IVD (control) quantity}} * 100$$

A threshold was set in order to dichotomize the samples into methylated and unmethylated categories. In lack of normal controls, which we usually use for determining this threshold for individual assays a threshold of PMR value of 10 was selected.

3.2.6 Statistics

The statistical analysis of the results was carried out in the SPSS 21.0 software.

Fisher`s exact test was used to establish whether or not there was an association between different tissue types and methylation percentage for the individual genes, and *P*-values were derived from two-sided tests. Independent Samples Median Test was used to compare the results obtained by qMSP and MSP. $P \leq 0.05$ was considered statistically significant for both tests.

4 Results

4.1 Identification of potential epigenetic candidate genes

Using a microarray gene expression approach based on the epigenetic treatment of 70 cancer cell lines from 16 different tissue types, and a stepwise bioinformatics pipeline (see figure 13, page 29) a list of 468 candidate DNA methylation master key genes was identified prior to this 'Master of Science' project. As many as 300 of our candidate genes had expression data registered in the IST base and 154 of these were down-regulated.

In order to narrow the list (n=154) down to a manageable number of candidates to start validating with wet lab analyses two gene lists with slightly different selection criteria were generated (see material and methods, page 29, and figure 13 step V). When both the lists were finalized we chose to examine approximately top 50 genes (53 from strategy 1 and 50 from strategy 2) from each list. There was an extensive overlap between the two lists, 39 genes were common, resulting in a final gene list of 64 genes considered to be the most promising candidates in the search for commonly methylated genes across cancer types.

Fifty-two of the 64 candidate genes had a CpG island in their promoter (figure 13, step VI). Seven of these genes had several transcription variants with different CpG islands, amounting to a total of 61 transcripts subjected to MSP analysis. In table 2 an overview of the given isoform names in this project is shown, coupled to their corresponding reference sequence.

	Given names in this project with corresponding RefSeqs
CCNA1	**CCNA1_iab** (NM_003914.3 and NM_001111045.1) and **CCNA1_ic** (NM_001111046.1 and NM_001111047.1)
BAIAP3	**BAIAP3_i1** (NM_003933.4) and **BAIAP3_i3-5** (NM_001199097.1, NM_001199098.1, NM_001199099.1, NM_001286464.1)
EFHD1	**EFHD1_i1** (NM_025202.3) and **EFHD1_i2 (NM_001243252.1)**
INPP5F	**INPPF_1** (NM_014937.3, NM_001243195.1) and **INPPF_2** (NM_001243194.1)
DAZL	**DAZL_i1** (NM_001190811.1) and **DAZL_i2** (NM_001351.3)
FHL1	**FHL1_t1,2,6** (NM_001159702,NM_001449,NM_001159703), **FHL1_t3** (NM_001159700) and **FHL1_t8** (NR_027621)
NDRG4	**NDRG4_t1** (NM_020465), **NDRG4_t2** (NM_001130487), **NDRG4_t3** (NM_022910)
GSN	**GSN_1** (NM_000177) and **GSN_2** (NM_198252 , NM_001127663, NM_001127664, NM_001127665, NM_001127666, NM_001127667)

Table 2: Genes with transcript variants and their given name in this project coupled with their corresponding RefSeqs

Several primer pairs were designed for genes with different CpG islands containing transcription variants where the promoter area was different. Given names are outlined in the table coupled with their correct reference sequences (numbers starting with NM).

4.2 Methylation specific PCR (MSP)

4.2.1 Primer design and optimization

The majority of the primer pairs designed for this study resulted in high quality data. However, a few of them were unspecific after optimization, and had to be redesigned. In total, 91 primer pairs were designed in order to thoroughly evaluate the promoter methylation status of the 52 candidate genes (61 transcript variants).

4.2.2 Methylation status of cancer cell lines as assessed MSP

The promoter methylation status of the 52 genes (61 transcript variants) after MSP in 20 cancer cell lines is summarized in figure 15. Five candidates were methylated in more than 50 % of the analyzed cell lines: *IFFO1*, *BAIAP3_i1*, *MT3* and two isoforms of the *CCNA1*, *CCNA1_iab* and *CCNA1_ic*. These were selected for validation with bisulfite sequencing and subsequent qMSP in a larger sample series. Thirteen of the transcripts evaluated by MSP were unmethylated in all cell lines, whereas 32 showed variable methylation frequencies (from 5 % to 45 %; figure 15).

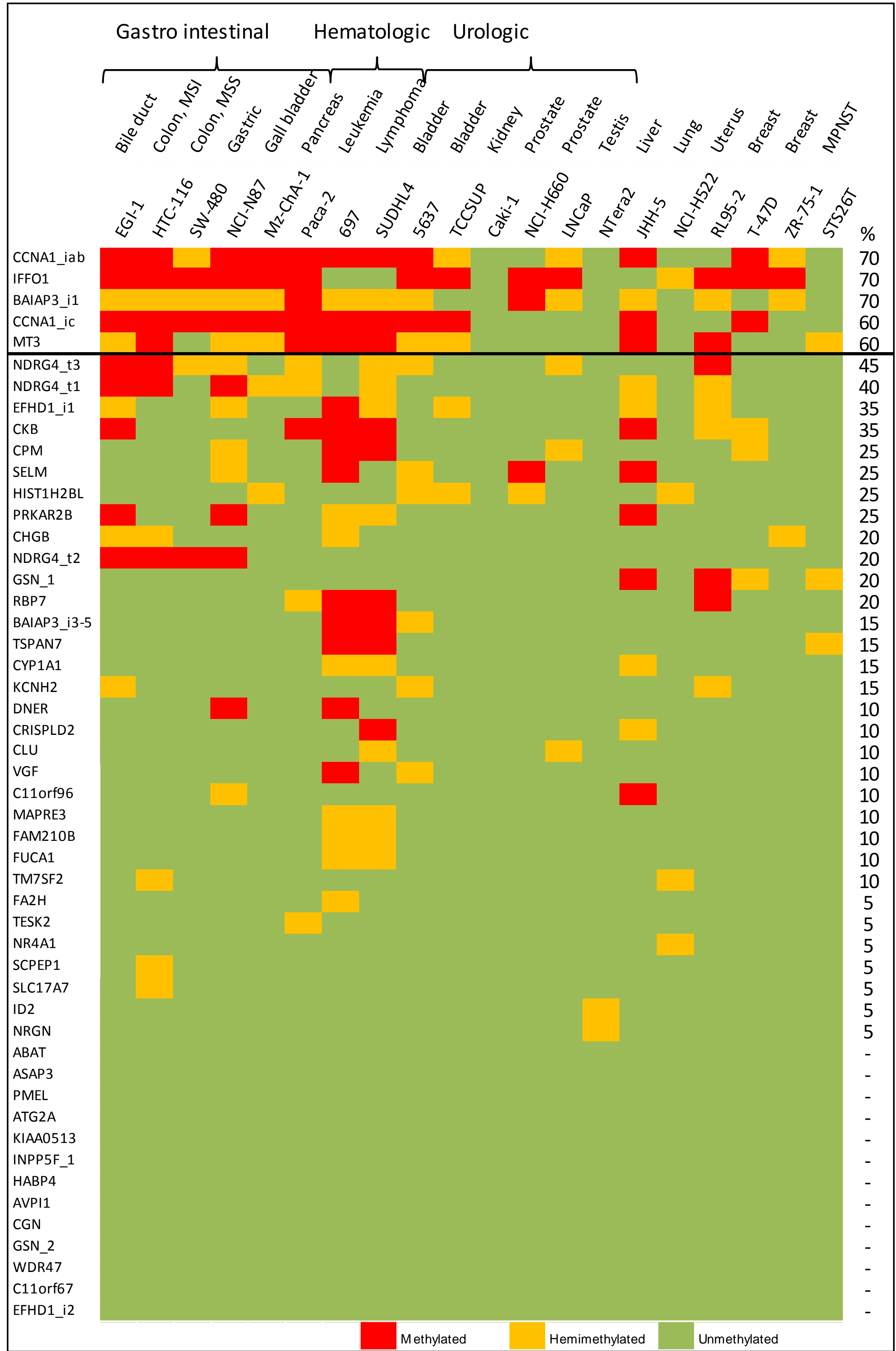


Figure 15: Methylation status as assessed by MSP

Bands scored as methylated are indicated with red color, unmethylated with green. Hemimethylated bands are indicated with orange color.

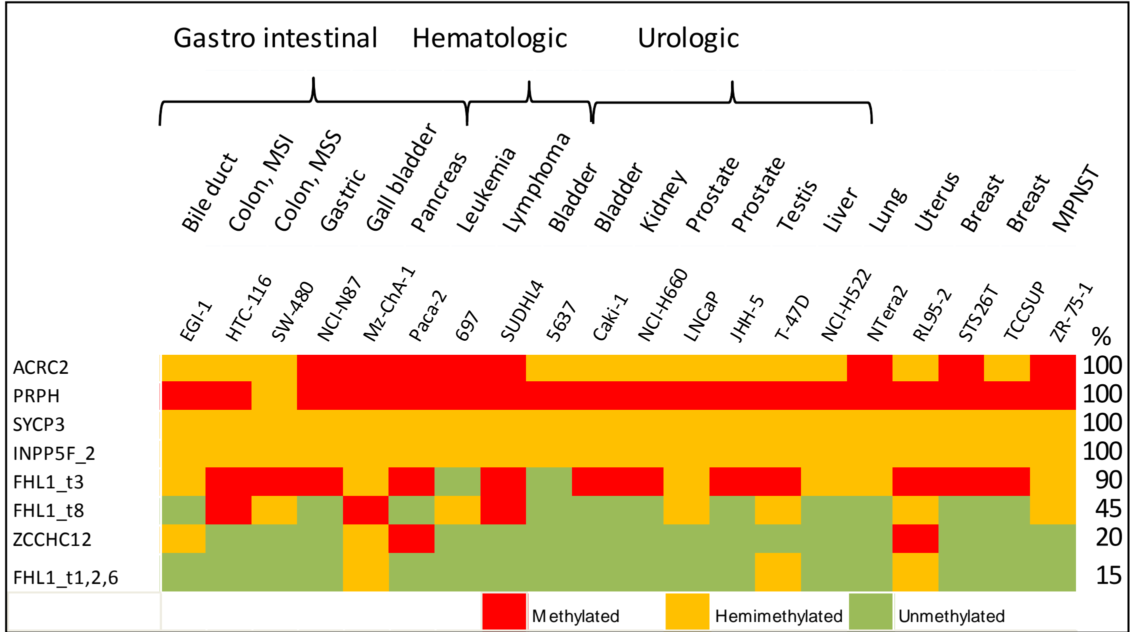


Figure 16: Methylation assays removed from further analysis due to failing quality control of the MSP analysis

The figure shows the assays that had amplified normal blood with the methylated primer pairs, indicating potential false positives. These genes were therefore not taken further in the analysis.

Six genes (including *FHL1* with three different promoter areas) were removed from both figure 15 and from further analysis because they lacked cancer specificity (the methylated primer pairs amplified normal blood; figure 16). Three of these genes, *ACRC*, *ZCCHC12* and *FHL1* are located at the X-chromosome and could potentially be mono-allelically methylated in biological material from female donors. The genes were subsequently tested in normal blood from 10 different persons, male and female. As expected, the methylated *FHL1* primer pairs only amplified DNA isolated from female blood donors. The methylated *ZCCHC12* and *ACRC2* reaction amplified DNA isolated from both male and female donors.

A representative MSP gel picture is shown in figure 17, where the methylation status of *IFFO1* is shown after gel electrophoresis.

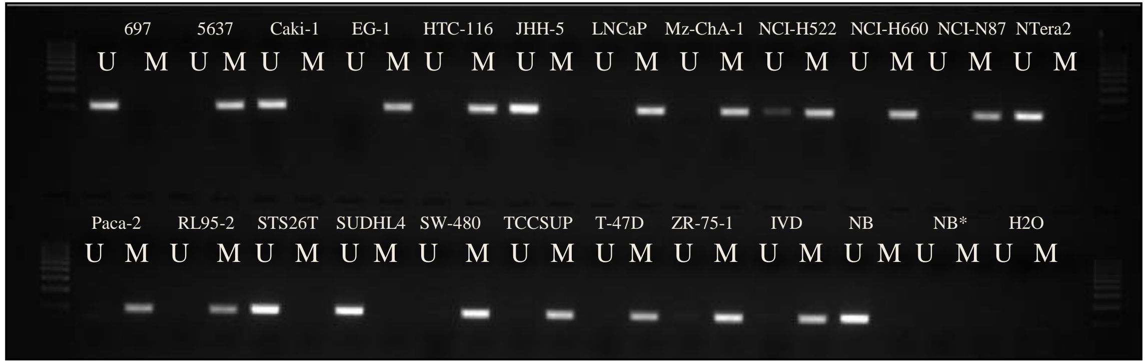


Figure 17: Picture of the methylation status of *IFFO1* after MSP reaction and gel electrophoresis

The methylated reactions are loaded in PCR lanes marked with M, and the unmethylated PCR products were loaded in lanes marked with U. The positive controls, IVD and NB, and the negative controls NB not bisulfite treated (marked with NB*) and water are marked in the figure.

Additional troublesome genes and testis cancer antigens

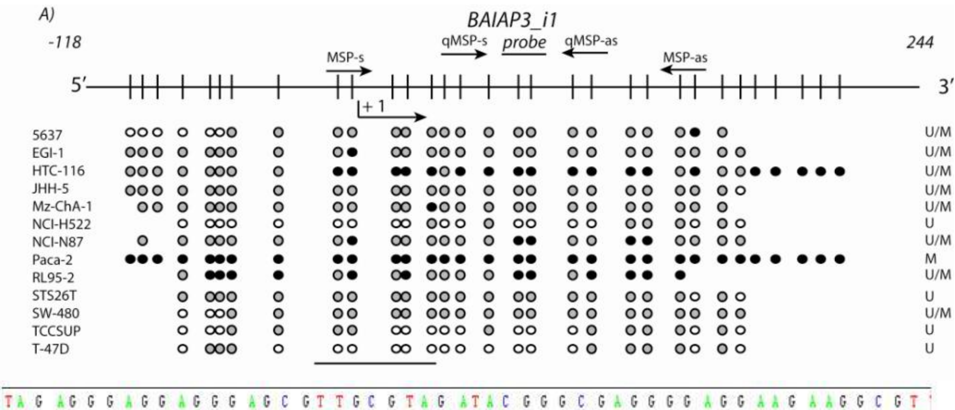
ITGA7 - 3 primer pairs designed, and none of them resulted in specific strong bands. They could not be used in a MSP reaction and were therefore not taken further in the analysis.

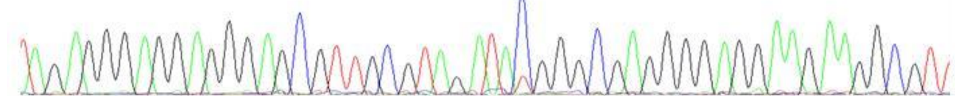
TDRD6, *SYCP3* and two transcript variants of *DAZL* - 4 primer pairs were designed for each gene and the unmethylated primer pairs resulted in unspecific PCR products or no amplification at all. After MSP reaction with only the primer pair designed for the methylated sequence it appeared that also this primer pair amplified normal blood, indicating false positives, and lack of cancer specificity. A literature search revealed findings that *TDRD6* and *DAZL*[10] are so called testis cancer antigens: methylated in all normal cells in the body except testes and frequently expressed in various cancers due to hypomethylation [96]. *DAZL* is methylated in primordial germ cells in the early embryo, and is in later stages of development unmethylated and expressed as a result of reprogramming [97]. Whether *SYCP*3 is also a testis cancer antigen is not clear, but since the behavior during primer optimization and MSP is the same as the other testis cancer antigens, this is likely.

4.3 Bisulfite sequencing

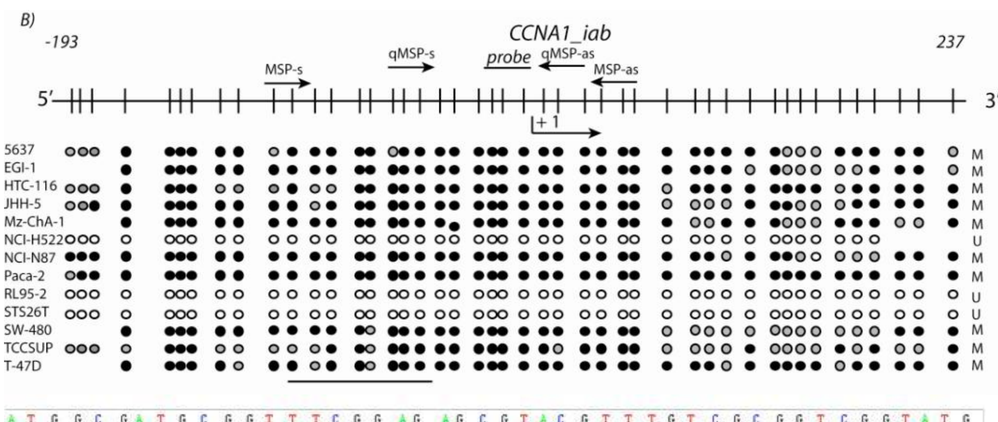
The promoter region of *BAIAP3_i1*, *CCNA1_iab*, *CCNA1_ic*, *IFFO1* and *MT3* was analyzed by direct bisulfite sequencing. These results were generally in good concordance with the results obtained by MSP, confirming that the MSP primer design was directed at a reasonable region of the promoter region (figure 18). The results were further used to direct the design of qMSP assay to regions showing consistent and frequent methylation (included in figure 18).

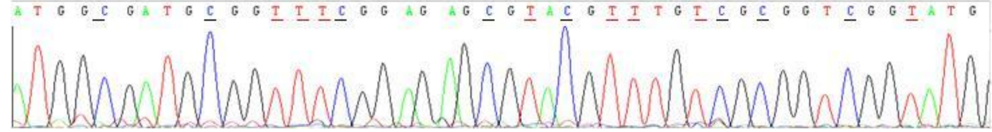
[10] http://www.ncbi.nlm.nih.gov/gene



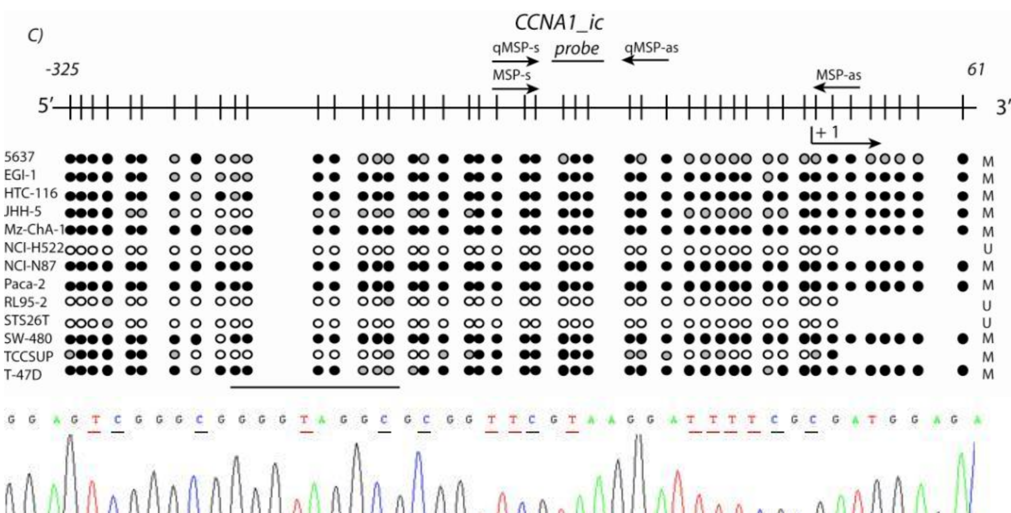


Electropherogram from the cell line HTC-116

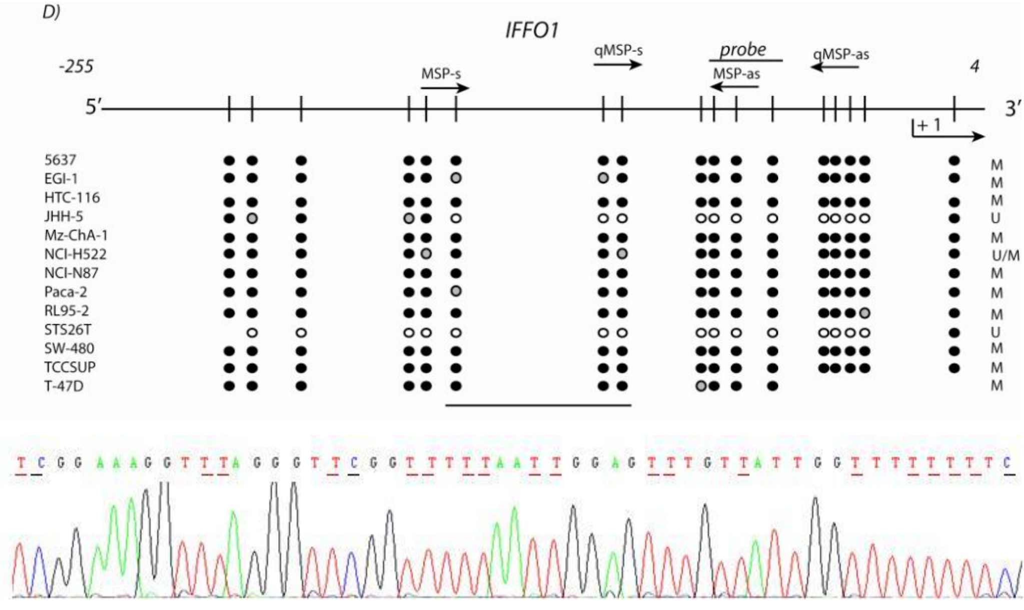




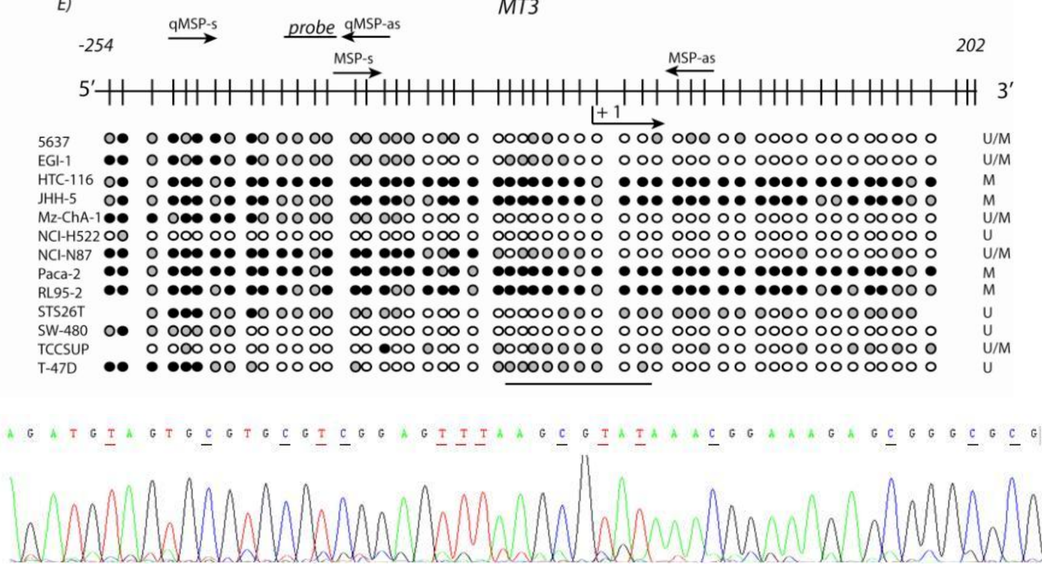
Electropherogram from the cell line 5637



Electropherogram from cell line Mz-ChA-1



Electropherogram from cell line NCI-N87



Electropherogram from cell line Paca-2

Figure 18: Methylation status of CpGs of interest in BAIAP3_i1, CCNA1_iab, CCNA1_ic, IFFO1 and MT3 as assessed by bisulfite sequencing

The figure shows parts of the CpG islands of the five genes. The vertical lines each represent one CpG site. The transcription start site is indicated by +1. For each gene, the location of primers (and probe) of the MSP and qMSp assays is indicated. Methylation status is indicated for each CpG: black circle, methylated CpG sites: grey circles, hemimethylated CpG sites: white, unmethylated CpG sites. To the right in the figures, methylation status after MSP is indicated. In the electropherogram, CpG sites are underscored by a black line whereas single cytosines in the original sequence (shown as Ts in the electropherogram) are underscored by a red line. The CpGs that are presented in the elecropherograms are marked in the figure with a long black line.

4.4 quantitative methylation specific PCR (qMSP)

4.4.1 Methylation status of the entire cancer cell lines panel as assessed by qMSP

Five candidates were indicated as interesting after investigation with MSP and bisulfite sequencing, and were further subjected to qMSP analysis in all 114 cell lines covering 16 different cancer types, summarized in figure 19. A total methylation percentage of 46 %, 64 %, 57 %, 70 % and 67 % was shown for *BAIAP3_i1*, *CCNA1_iab*, *CCNA1_ic*, *IFFO1* and *MT3*, respectively across all the cell lines analyzed (figure 19 and table 3).

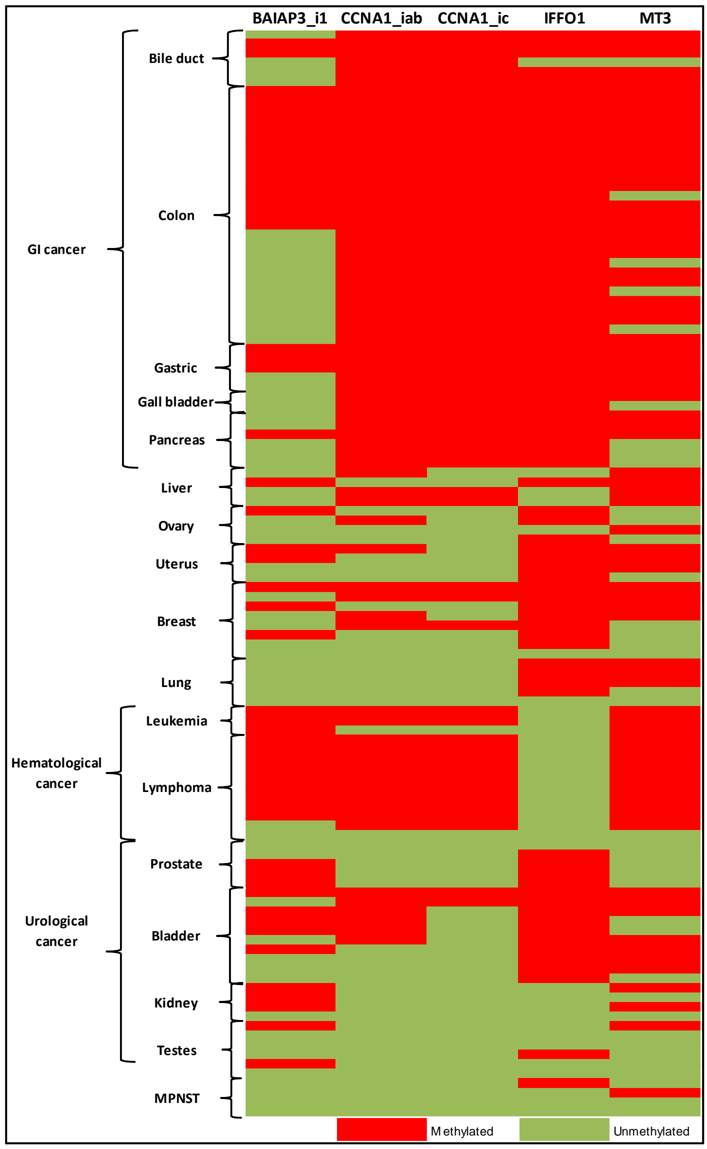


Figure 19: Methylation status of the five genes in the 114 cell lines sorted by tissue type

Cell lines with PMR value below 10 are shown in green in the figure, and cell lines with PMR value above 10 is shown in red.

The methylation frequency of *CCNA1* could be regarded as 64 %, since two transcript variants CCNA1_iab and CCNA1_ic, show a similar methylation frequency, but CCNA1_iab has methylated bands in additional cell lines compared to CCNA1_ic.

	BAIAP3_i1	CCNA1_iab	CCNA1_ic	IFFO1	MT3	# cell lines
Gastrointestinal						
Bile duct	2 (33 %)	6 (100 %)	6 (100 %)	5 (83 %)	5 (83 %)	6
Colon	15 (56 %)	27 (100 %)	27 (100 %)	27 (100 %)	23 (85 %)	27
Gastric	3 (60 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5
Gall bladder	0 (0 %)	2 (100 %)	2 (100 %)	2 (100 %)	1 (50 %)	2
Pancreas	1 (17 %)	6 (100 %)	6 (100 %)	6 (100 %)	3 (50 %)	6
Total Gastrointestinal	21 (46 %)	46 (100 %)	46 (100 %)	45 (98 %)	37 (80 %)	46
Hematological						
Leukemia	3 (100 %)	2 (67 %)	2 (67 %)	0 (0 %)	3 (100 %)	3
Lymphoma	9 (82 %)	10 (91 %)	10 (91 %)	0 (0 %)	10 (91 %)	11
Total hematological	12 (86 %)	12 (86 %)	12 (86 %)	0 (0 %)	13 (93 %)	14
Urological						
Prostate	3 (60 %)	0 (0 %)	0 (0 %)	4 (80 %)	0 (0 %)	5
Bladder	5 (50 %)	6 (60 %)	2 (20 %)	10 (100 %)	7 (70 %)	10
Kidney	3 (75 %)	0 (0 %)	0 (0 %)	0 (0 %)	2 (50 %)	4
Testes	2 (33 %)	0 (0 %)	0 (0 %)	1 (17 %)	1 (17 %)	6
Total urological	11 (58 %)	6 (32 %)	2 (11 %)	14 (74 %)	9 (47 %)	19
Liver	1 (25 %)	3 (75 %)	2 (50 %)	1 (25 %)	4 (100 %)	4
Ovary	1 (25 %)	1 (25 %)	0 (0 %)	3 (75 %)	1 (25 %)	4
Uterus	2 (50 %)	1 (25 %)	0 (0 %)	4 (100 %)	3 (75 %)	4
Breast	3 (38 %)	4 (50 %)	3 (38 %)	7 (88 %)	4 (50 %)	8
Lung	0 (0 %)	0 (0 %)	0 (0 %)	4 (80 %)	3 (60 %)	5
MPNST	0 (0 %)	0 (0 %)	0 (0 %)	1 (25 %)	1 (25 %)	4
Total	53 (46 %)	73 (64 %)	65 (57 %)	80 (70 %)	76 (67 %)	114

Table 3: Methylation status for the genes after qMSP in the 114 cell lines

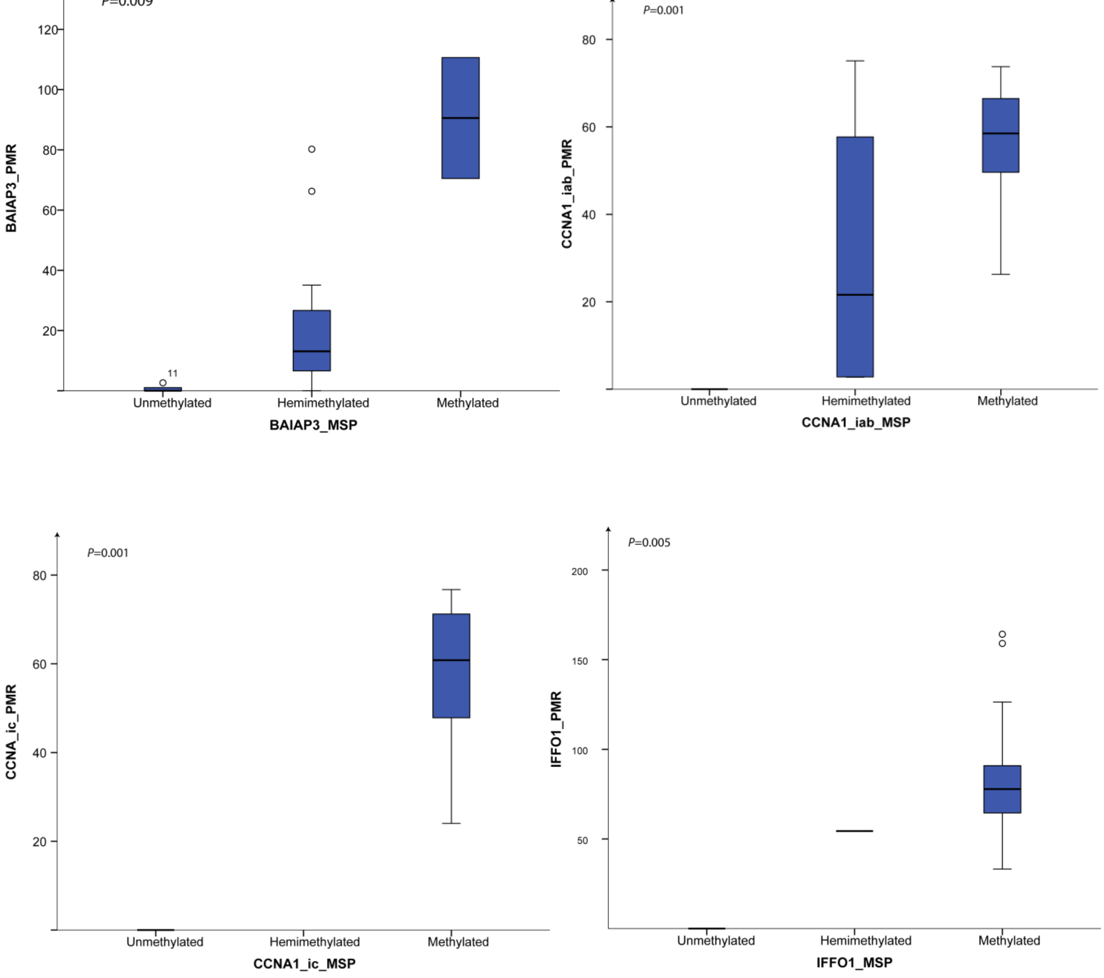
The results are summarized for each tissue group per gene, and shown in number of methylated cell line within each tissue group. The tissue groups that are closely related are also grouped together, and summarized.

In table 3 the cell lines are categorized into "meta-groups", such as gastrointestinal, hematological and urological cancers. The methylation status is different within these groups for some of the genes. For *BAIAP3* the hematological cancer cell line group is more frequently methylated (86 %) than the rest of the cell lines, arising from solid tumors (41 %; $P = 1.27 \cdot 10^{-3}$). The rest of the candidate genes, *CCNA1_iab*, *CCNA1_ic*, *IFFO1* and *MT3* were more frequently methylated in the cell lines from gastrointestinal cancer compared to the rest of the cell lines. *CCNA1_iab* and CCNA1_*ic* have 100 % methylation in the gastrointestinal cell lines compared to 40 % and 28 % methylation in the other cell line tissue groups ($P = 4.89 \cdot 10^{-13}$ and $8.64 \cdot 10^{-17}$, respectively). *IFFO1* is the candidate with the highest methylation

percentage (70 %) across all the included cell lines. But *IFFO1* is also more frequently methylated in the gastrointestinal cell line groups (98 %) compared to the rest of the cell lines (51 %; $P = 1.35 \cdot 10^{-8}$). The opposite is true for the hematological groups, where *IFFO1* is unmethylated in all 14 cell lines ($P = 4.45 \cdot 10^{-9}$). *MT3* is highly methylated in both the gastrointestinal cell line group, 80 % compared to 57 % methylation in the rest of the cell lines ($P = 1.46 \cdot 10^{-2}$), and hematological cell lines, 93 % compared to 63 % ($P = 3.26 \cdot 10^{-2}$).

4.4.2 Comparison of qualitative and quantitative methylation analysis

Results of the 20 cell lines investigated in both MSP and qMSP are compared in figure 20, and high concordance was generally found between the results produced by the two methods.



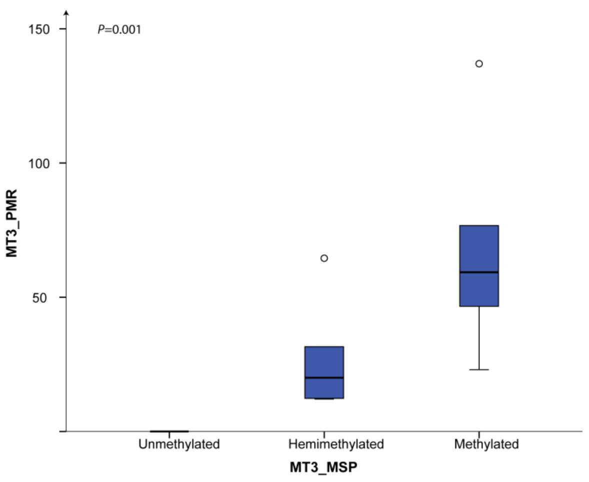


Figure 20: Comparison of methylation status in cell lines from MSP and qMSP

Box plots denote the PMR values (y-axis) determined for 20 samples by qMSP compared to the scores obtained by MSP (x-axis).

P-values after performing an Independent samples median test are shown for each gene.

The box plots comparing methylation status after assessment with MSP and qMSP show statistically significant higher PMR values for samples scored as methylated after MSP compared to those scored as hemimethylated and unmethylated.

4.5 Gene expression results from the IST database

The expression of the four most promising candidate genes from the qMSP analysis (*BAIAP3, CCNA1, IFFO1* and *MT3*) was evaluated in tissue samples using the data available from IST (figure 21). The average expression of the candidate genes in tumor samples was compared to normal samples and across cancer types. There was indeed a trend of down-regulation in tumor compared to normal, with some clear exceptions.

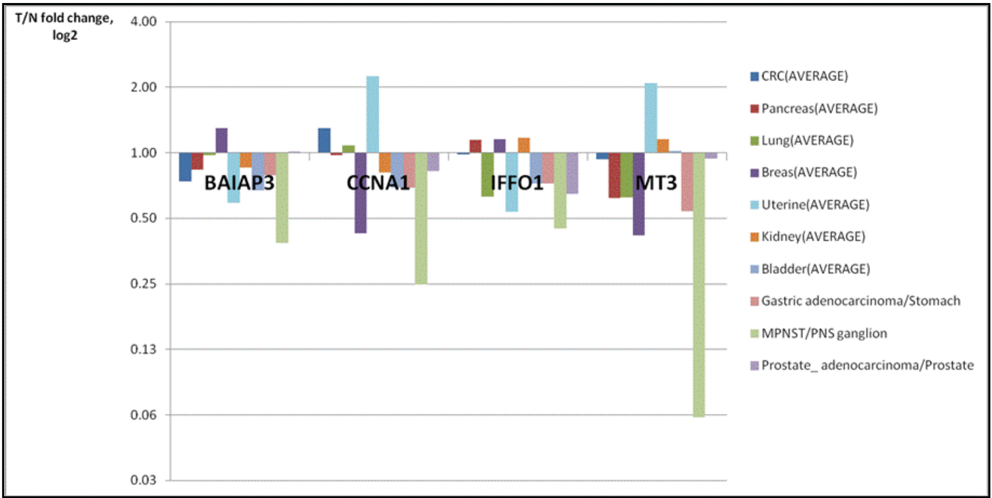


Figure 21: IST expression data of the top four candidate genes

The figure shows relative expression when tumor samples are compared to normal samples. Along the X-axis the different genes are shown with individual gene expression for the different cancer types compared to normal. Along the Y-axis fold change is plotted (log2).

5 Discussion

5.1 Material and Methodological considerations

5.1.1 Usage of cancer cell lines as *in vitro* models

Cancer cell lines are an *in vitro* model derived from human cancer cells, and they are highly valuable in research [98]. Reasons for using such cell lines are numerous. The fact that they are monoclonal and immortalized enables the retrieval of many identical copies of one cell [99]. Cell lines give the opportunity of doing more and larger studies without spending highly valued patient material. There are also a lot of reference data available for cell lines in databases online, such as DNA copy number alterations, somatic mutations, mRNA expression profiles and DNA methylation status [99].

A cancer cell line is a useful model, but it is clear that it is indeed only a model. Cell lines are not representative when it comes to normal cell behavior *in vivo*, but rather represent an isolated cell operating without all the control mechanisms that usually takes place in the body during tumor development. Validation is therefore highly necessary when it comes to findings obtained from cell lines [99], and this should be done in patient material to ensure that the initial findings are not cell line artifacts.

All the cell lines used in this study were subjected to authentication and tested for *Mycoplasma* infection. The authentication was performed to ensure the correct cell line identity. This is important due to the large number of misclassified cell lines discovered the last years, and the critical consequences of doing expensive studies in cell lines that has mistaken identity or is contaminated [100]. *Mycoplasma* contaminates about 15-35 % of cell cultures, and it can affect cell metabolism, increase sensitivity to inducers of apoptosis or inhibit cell growth and thereby affecting the gene-expression level [100].

5.1.2 Microarray results and selection of genes

During the first step in the experimental procedure to identify candidate genes silenced by epigenetic mechanisms, AZA/TSA treatment was performed in cancer cell lines followed by microarray gene expression analysis. The combination of the two drugs has synergic effects,

and up-regulation after treatment indicates that the gene investigated has been silenced by DNA methylation, lack of histone acetylation, or a combination of both, and that the transcription initially was repressed. When treating cells with demethylating drugs such as AZA and TSA there is a substantial risk of retrieving false positives. Several of the genes are up-regulated by the treatment, regardless of whether they had DNA methylation in the promoter area to begin with. This could be due to general cytotoxic effect in the cell, or be caused by inactivating methylation of other regulatory elements for the gene, such as the enhancer. This also partly explains why some of the responding genes selected after microarray gene expression analysis did not have a CpG island in the promoter region.

In order to limit the risk of selecting false positive candidate genes for methylation analysis, a conservative selection pipeline with various criteria was used including IST data. Cell lines are known to have more methylation than primary cancers, and also a different methylation pattern, perhaps due to the alterations necessary for these cells to replicate limitless in a culture dish [101,102]. The microarray analysis was performed on cell lines, and the usage of expression data from large series of cancer tissue compared to expression in normal samples as selection criteria increases the possibility of ending up with candidates that are methylated in tissue sample and not just artifacts seen in cell lines.

Two strategies in the final selection of candidate genes were used to increase the robustness of the approach. Both strategies were based on how well the genes performed within four parameters (see figure 13, page 29). In strategy 1, it was a requirement that genes selected needed to be on the top 100 list from each of the four parameters. In strategy 2 the genes were ranked within each parameter individually, and then given a final score based on the sum of these individual rankings. The top 50 genes from this strategy 2-based list were selected for further analyses. Strategy 1 could for some of the genes be strict in the sense that they could perform excellent in three out of the four parameters, but still be excluded if they fell below the top 100 list for one of the parameters. To avoid losing potentially good candidates at this stage of the selection process, we subjected all genes from strategy 1 and 2 to DNA methylation analysis.

Despite this approach, the risk of losing excellent candidate genes is definitely present. However, to investigate all genes on the list would be an enormous effort both time and resource wise, and the focus was therefore on selecting a manageable number of promising candidate genes. There are many ways to analyze a data set this size, but the identification of

several promising masterkey candidates from the present study underscores that the pipeline used here was reasonable.

A similar gene expression microarray approach as described here has in our department resulted in several excellent DNA methylation biomarkers [103-105]. There are also other methods that could have been used for such identification, including methylome sequencing of cancer cell lines and alternatively patient material. This is however expensive, and it would not be possible to include the total number of cell lines used in this study (n=114). Furthermore, high throughput sequencing approaches are challenging at the bioinformatics level. Methylation arrays could also have been an alternative to the gene expression microarray approach, giving high output at a reasonable price. However, with this approach we would have lost the valuable information regarding expression of the genes, before and after treatment with epigenetic drugs.

5.1.3 Methods used for determining DNA methylation

Bisulfite conversion of DNA is a method that dates back to the early 1990s, and since then it has become the "gold standard" when it comes to studying methylated DNA [32,87,88]. Today there are many advanced methods of studying DNA methylation, but here we chose to use MSP after bisulfite conversion in the early selection of candidate genes. Reasons for this are that MSP is relatively simple, robust and fast, enabling the screening of many genes in the search for a proper biomarker. It can successfully separate one methylated allele from 1000 unmethylated alleles, and the costs are kept low. By experience, the primer design method used here has proven to be quite good, and excellent concordance between the results obtained from MSP and bisulfite sequencing is the norm. However, a major disadvantage of MSP is its semi-quantitative nature and subsequent challenges in manually scoring the results [32]. To ensure consistency during scoring, we have developed a semi-automated system based on intensity values provided by the GelDoc after visualization of the MSP results, which is described in the methods part of this thesis (see figure 14, page 35). All the results are scored by two individuals independently to remove any "personal" bias. The second obvious and major disadvantage with the MSP method is the use of agarose gels for separating the PCR products. Casting, loading, running and taking pictures of these gels is laborious and time consuming, and works best for a small panel of samples. The quality of the primer sets is crucial to achieve correct and robust methylation results. All primer pairs

included in this study were carefully designed and optimized to ensure equal and easily comparable bands for both the methylated and unmethylated reactions. This is time consuming and frequently skipped by many labs. However, we believe that this approach pays off because of the high quality of the generated data. In the present study several primer pairs had to be redesigned and optimized multiple times.

qMSP is compared to MSP, highly quantitative, and gives more information regarding the methylation status. Bisulfite sequencing is time consuming, but valuable since it generates a detailed overview of the methylation status of every individual CpG site included in the sequenced product. This enables a qMSP assay to be design to target the most representative CpG sites in the promoter [106]. qMSP is a simple, rapid, highly sensitive and relatively easy method [47], especially when it can be standardized by using a pipetting robot in the experimental set-up. It is, however, slightly more expensive than MSP due to the fluorescence. qMSP may provide a more conservative methylation scoring than MSP due to the use of a probe in addition to the forward and reverse primer. The probe also provides better specificity by covering additional CpG sites, giving the qMSP assay a higher stringency. MSP is preferred by our group when it comes to the fast screening of many genes and subsequent selection of candidates for further analysis. In combination with subsequent bisulfite sequencing and finally qMSP, it provides a robust pipeline for analyzing DNA methylation. During the analysis of the results after qMSP, we normally use a threshold based on methylation in normal tissue compared to methylation in cancer [103,104]. In this master project, we have studied cell lines, and did not have available normal tissue for all the tissue types studied. In lack of this, we set a conservative scoring threshold at PMR 10 when dichotomizing the qMSP results into methylated and unmethylated categories.

5.2 *BAIAP3, CCNA1, IFFO1* and *MT3* are found to be frequently methylated in cancer cell lines

The present "Master of science" project is a part of the larger "Masterkeys" project that aims to identify frequently inactivated genes by DNA methylation and analyze and evaluate whether they could represent valuable cancer drivers or biomarkers. As mentioned in the introduction (page 16 to 18) a driver gives the cancer a selective advantage, enabling the cancer to grow faster and/or more aggressively. When the driver is found in several cancer types with different molecular "backgrounds" it could be considered a 'Masterkey' in cancer

development, possibly providing the cancer cells with advantages that are ‘universal’ and valuable for all cells independent of their underlying phenotype. One example of such a genetic ‘masterkey’ is the TP53 protein, mutated in at least 50 % of all cancers. The loss of function of this tumor suppressor gives several selective advantages for the cell [21]. DNA methylation is known to be tissue-specific, seen both in healthy tissue as well as in tumors. It has even been suggested that metastatic tumors with unknown origin can be traced back to the primary site by use of methylation patterns [107]. However, genes methylated across several cancer types could indicate a driver function. For a gene to be a potential driver, it needs to have an essential function in the cell, and since DNA methylation in the promoter area has a silencing effect, the preferred function in this setting would be tumor suppression. *CDKN2A* is, as mentioned in the introduction, an example of a tumor suppressor inactivated by epigenetic mechanisms in several cancer types [62], and an example of an epigenetic ‘masterkey’. The protein has an inhibitory role in the cell cycle and cancer cells that have lost this control mechanism can progress uncontrollably through the cell cycle, resulting in a driver mechanism in cancer progression.

A biomarker is used to identify a condition, such as a disease in the body. There is a constant ongoing search for cancer specific DNA methylation biomarkers, especially those that are present early in cancer development and thereby potentially useful in a screening setting [50]. DNA methylation generally happens early in the development of cancer, and may contribute in cancer progression by inactivating tumor suppressor genes and other genes important for cancer development and growth [69]. This makes DNA methylation biomarkers promising in the diagnostics.

In the present study, *BAIAP3, CCNA1, IFFO1* and *MT3* showed a relatively high methylation frequency across several cell lines from different cancer tissues making them very interesting biologically. *CCNA1_iab*, *CCNA1_ic* and *IFFO1* have a very high methylation percentage within the gastrointestinal cell lines (bile duct, colon, gastric, gall bladder and pancreas) and *MT3* also has a relatively high methylation within these cell lines groups. *BAIAP3_i1* is not as frequently methylated in the gastrointestinal cell lines as the other candidates, but it is highly methylated in the hematological cell lines. *IFFO1* show generally frequent methylation in most of the tissues, with the exception of the cell lines derived from hematological cancers, kidney, liver, testis and MPNST.

BAIAP3, BAI1-associated protein 3, located on chromosome 16 (16p13.3), encodes a protein that is a member of the secretin receptor family. It has been reported as a target for the downstream effects of TP53, preventing angiogenesis, indicating a tumor suppressor function[11]. *BAIAP3* has further been shown to be important in regulating exocytosis. There are few available articles linking BAIAP3 to cancer development, but one study by Palmer *et al*. reported that by creating a chimeric novel transcription factor called EWS-WT1 in sarcoma cell lines they managed to up-regulated expression of *BAIAP3*, resulting in enhanced neoplastic properties of the cells. This has so far not been confirmed *in vivo*, or in cell lines from other tissues [108]. The indicated anti-tumor function of *BAIAP3* and the methylation across several different types of cell lines seen in this study makes *BAIAP3* interesting for further studies.

CCNA1, Cyclin A1, is a part of the highly conserved Cyclin family and located on chromosome 13 (13q13.3). The cyclins function as regulators of the CDK kinases in the cell cycle. Indeed, CCNA1 has been shown to bind to important cell cycle regulators, such as RB1 family proteins, transcription factor E2F-1 and p21 family proteins. The resulting protein has been proposed to control CDK2 in the cell cycle at the transition from G1 phase to S phase, and G2 phase to mitosis (M) phase[12]. The cyclin A1-CDK2 complex also has a function in double stranded DNA repair following radiation damage, and cells lacking Cyclin A1 will have a possibility of proceeding through cell cycle despite doubled stranded breaks, resulting in a possible accumulation of DNA damage in the cell [109]. The expression of Cyclin A1 has further been shown to be up-regulated after UV- and γ-irradiation, while Cyclin A2 had a decreased expression, indicating different roles for the two cyclins [109]. In a cell lines study, Cyclin A1 was shown to have an increased expression due to TP53 expression, indicating that Cyclin A1 may be an effector protein of TP53 [110]. This was supported by another study that showed that TP53 activated the *CCNA1* promoter, but only in the presence of SP1 [109], a transcription factor usually found at active CpG island promoters [111]. *CCNA1* has already been reported to be hypermethylated in several cancer types, such as head and neck squamous cell carcinoma (HNSCC), cervical cancer, colorectal cancer, bladder cancer and breast cancer [112-116]. One study also showed that HNSCC patients with CCNA1 silenced by DNA methylation had a much higher risk of developing additional primary tumors [112]. A study of bladder cancer found *CCNA1* to be methylated in 61 % of the patients. They also found an

[11] http://www.genecards.org/
[12] http://www.genecards.org/

increased methylation frequency in muscle-invasive tumors, indicating a potential role of CCNA1 in the pathogenesis and metastasis of bladder cancer [115]. Use of *CCNA1* methylation status as a biomarker has also been reported to distinguish between moderate (CIN2) and severe (CIN3) dysplasia in cervical cancer [113].

We found that *CCNA1* was highly methylated across the various cancer cell lines, with a methylation percentage of 64 % and 57 % for the two transcription variants. The function of CCNA1 in DNA repair and the many studies reporting its involvement in cancer development supports our findings, and may indicate that *CCNA1* represents an epigenetic 'Masterkey'. While we found *CCNA1* to be 100 % methylated in colorectal cancer cell lines, the IST database reported an up-regulation of this gene on average in these tumors compared to their normal counterparts. Although surprising, this might be explained by the fact that *CCNA1* has many transcription variants, and that some of the non-CpG containing promoters may contribute to this up-regulation. The dual function of CCNA1, both as a cyclin in charge of driving cell cycle forward as well as a component in DNA repair, can also partly explain the increased expression seen in colorectal cancer. Cancer could get a selective advantage by up-regulating Cyclin A1 in the cell cycle context, but down-regulation of Cyclin A1 could also give a selective advantage since the cell will go through cell cycle with damage in the genome due to the lack of DNA repair. It is also important to notice that if Cyclin A1 is down-regulated, other related cyclins, such as Cyclin A2 will still drive the cell through the cell cycle, but Cyclin A2 does not have the same capacity regarding DNA repair as Cyclin A1.

IFFO1, intermediate filament family orphan 1 located on chromosome 12 (12p13.31), is a member of a family with the same name. They code for important protein components of the cytoskeleton and nuclear envelope[13]. *IFFO1* has previously been reported to be frequently methylated in ovarian cancers, and was indicated by Peter Lairds group to be the top candidate biomarker for the disease [117]. *IFFO1* had the highest methylation frequency (70 %) across the cancer cell lines in this study, and if this is confirmed in cancer tissue, *IFFO1* could have a potential as a cancer specific 'masterkey' DNA methylation biomarker.

MT3, metallothionein 3 is located on chromosome 16 (16q12.2) and functions as a metal-binding protein, limiting oxidative damage[13]. The metallothioneins constitute a family of four members and are thought to be involved in important biological functions, such as cell

[13] http://www.genecards.org/

proliferation, differentiation, invasion and apoptosis [118]. They have also been indicated to play a role in protecting against DNA damage and oxidative stress [119]. However, the role of MT3 in cancer progression is not clear, and both up-regulation and down-regulation of the gene have been reported to have an effect on tumorigenesis [119]. A study from 2013 showed that over-expression of MT3 in prostate carcinoma cells increased cell proliferation, invasion and tumorigenic activities *in vitro* and *in vivo* [118]. *MT3* is also reported to be down-regulated by DNA methylation in esophageal adenocarcinoma tumor samples (n=64) [119].

The identification of drivers is biologically interesting, providing knowledge about cancer progression in general, but it could also be useful in cancer therapy. Targeting oncogenes has previously been regarded as most promising in a therapeutic perspective, but strategies for restoring function of the inactivated tumor suppressors can also be a useful weapon in fighting tumor progression. Gene silencing due to DNA methylation is potentially reversible and new strategies for specifically targeting and reversing gene specific DNA methylation, using TALE-TET1 fusion proteins (described in the introduction, page 23), could be a potential way to treat cancers in the future [39].

'Masterkeys' could in addition to provide important knowledge about cancer development and potential targets of therapy, also be used as biomarkers in the clinic. Cancer specific markers and cancer type specific markers within metagroups or individual cancer types could be a powerful tool in diagnosis and early detection. The use of such markers in the clinic and screening is however not unproblematic and straightforward. A DNA methylation biomarker being able to detect several different cancer types in the body could not be assessed before specific diagnostic tests for each cancer type is ready for usage. To give a patient a cancer diagnosis without being able to correctly determine where the cancer is, will be unethical (and perhaps unbearable). But together with an arsenal of other DNA methylation and genetic tests in the future such "master" DNA methylation biomarkers could be a part of a diagnostic tool, saving valuable resources by discovering cancer early and therefore saving patient lives. The usage of such "master" DNA methylation biomarkers could be of high value, but only time will show if it is possible and also whether such biomarkers actually exist. Extremely high sensitivity and specificity is needed for such a marker, due to the personal consequences of misdiagnosis.

Further studies are required to determine whether any of the genes identified here have cancer driver functions when inactivated by DNA methylation and also whether they have potential

as DNA methylation cancer biomarkers. Based on results published so far and summarized above, *CCNA1* seem to be the most promising ‘Masterkey’ candidate. *BAIAP3* was found to be highly methylated in the hematological cell lines, and if the same is shown in future analyses of larger patient series, it could have the potential as a cancer type specific biomarker. The same could be the case for *MT3*, also highly methylated across the cell lines derived from hematological tumors. However, this gene was also frequently methylated in gastrointestinal tumors, undermining its cancer-type specificity. *IFFO1* is reported to be a promising DNA methylation marker for ovarian cancer, and we can also report very high methylation frequency across a range of other cancer cell lines, indicating a “masterkey” potential.

6 Conclusions

To identify genes frequently inactivated by promoter hypermethylation across the majority of cancer types, we have used an ambitious step-wise experimental approach and a panel of 114 cancer cell lines from 17 different tissue types. Four genes, *BAIAP3*, *CCNA1*, *IFFO1* and *MT3* were methylated in the majority of these cell lines. These genes could represent potential epigenetic 'masterkeys' for cancer development and will be subjected to methylation analysis in clinical patient material as well as functional studies.

7 Future perspectives

The four candidate genes (five transcripts) identified in this study will be subjected to further DNA methylation analysis by qMSP in tissue from cancer patients and healthy controls. The access to a broad range of tissues from cancer patients could be a challenge, but we currently have available material from colorectal-, gastric-, prostate-, testes- and MPNST (malignant peripheral nerve sheath tumors) cancer patients. Material from breast-, bladder-, esophagus- and ovarian- cancer patients will be available through collaborators. In addition to results generated in house we will use methylation data available through large published data sets, such as the methylome array data included in the TCGA Project.

High degree of methylation across various cancer types could indicate a cancer driver potential for these candidates. This will be investigated through functional studies, including specific demethylation of the CpG islands of these gene promoters, hopefully leading to a reactivation of the gene expression. This can be performed using the new TALE-TET technology described in the introduction (page 23). Functional effects of such a reactivation will be measured using the the xCELLigence system (ACEA/Roche), with particular focus on cell viability, proliferation, differentiation, cell invasion and migration. These cellular events could also be monitored real time using our Zeiss LSM710 Confocal Microscope. The analyses will be done in collaboration with cell biology project group in our department.

Knock-out studies, using siRNA in normal-like cell lines, followed by investigation of cell behavior is an alternative strategy to determine potential driver function. Important questions to be answered in this setting are whether the cell survives, is able to divide or undergoes cell cycle arrest and if it acquires tumor capabilities. This can also be studied using the xCELLigence system or confocal microscopy.

Reference List

1. Nowell PC: **The clonal evolution of tumor cell populations.** *Science* 1976, **194:** 23-28.

2. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG *et al*.: **Increased methylation variation in epigenetic domains across cancer types.** *Nat Genet* 2011, **43:** 768-775.

3. Peto J: **Cancer epidemiology in the last century and the next decade.** *Nature* 2001, **411:** 390-395.

4. Greaves M, Maley CC: **Clonal evolution in cancer.** *Nature* 2012, **481:** 306-313.

5. Burrell RA, McGranahan N, Bartek J, Swanton C: **The causes and consequences of genetic heterogeneity in cancer evolution.** *Nature* 2013, **501:** 338-345.

6. Heim S, Teixeira MR, Dietrich CU, Pandis N: **Cytogenetic polyclonality in tumors of the breast.** *Cancer Genet Cytogenet* 1997, **95:** 16-19.

7. Ponder BA: **Cancer genetics.** *Nature* 2001, **411:** 336-341.

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

9. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation.** *Cell* 2011, **144:** 646-674.

10. Stella GM, Luisetti M, Pozzi E, Comoglio PM: **Oncogenes in non-small-cell lung cancer: emerging connections and novel therapeutic dynamics.** *Lancet Respir Med* 2013, **1:** 251-261.

11. Weinberg RA: *The biology of cancer*. New York: Garlan Science, Taylor & Francis Group, LLC; 2007.

12. Heinemann V, Douillard JY, Ducreux M, Peeters M: **Targeted therapy in metastatic colorectal cancer -- an example of personalised medicine in action.** *Cancer Treat Rev* 2013, **39:** 592-601.

13. Minuti G, D'Incecco A, Cappuzzo F: **Targeted therapy for NSCLC with driver mutations.** *Expert Opin Biol Ther* 2013, **13:** 1401-1412.

14. Braun BS, Shannon K: **Targeting Ras in myeloid leukemias.** *Clin Cancer Res* 2008, **14:** 2249-2252.

15. Evan G: **Cancer. Taking a back door to target Myc.** *Science* 2012, **335:** 293-294.

16. Spender LC, Inman GJ: **Developments in Burkitt's lymphoma: novel cooperations in oncogenic MYC signaling.** *Cancer Manag Res* 2014, **6:** 27-38.

17. Indovina P, Marcelli E, Casini N, Rizzo V, Giordano A: **Emerging roles of RB family: new defense mechanisms against tumor progression.** *J Cell Physiol* 2013, **228:** 525-535.

18. Knudson AG, Jr.: **Mutation and cancer: statistical study of retinoblastoma.** *Proc Natl Acad Sci U S A* 1971, **68:** 820-823.

19. Dick FA, Rubin SM: **Molecular mechanisms underlying RB protein function.** *Nat Rev Mol Cell Biol* 2013, **14:** 297-306.

20. Richardson RB: **p53 mutations associated with aging-related rise in cancer incidence rates.** *Cell Cycle* 2013, **12:** 2468-2478.

21. Purvis JE, Karhohs KW, Mock C, Batchelor E, Loewer A, Lahav G: **p53 dynamics control cell fate.** *Science* 2012, **336:** 1440-1444.

22. Hoeijmakers JH: **Genome maintenance mechanisms for preventing cancer.** *Nature* 2001, **411:** 366-374.

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

24. Yeramian A, Moreno-Bueno G, Dolcet X, Catasus L, Abal M, Colas E *et al.*: **Endometrial carcinoma: molecular alterations involved in tumor development and progression.** *Oncogene* 2013, **32:** 403-413.

25. Gibson J, Lacy J, Matloff E, Robert M: **Microsatellite instability testing in colorectal carcinoma: a practical guide.** *Clin Gastroenterol Hepatol* 2014, **12:** 171-176.

26. Simpson AJ, Caballero OL, Pena SD: **Microsatellite instability as a tool for the classification of gastric cancer.** *Trends Mol Med* 2001, **7:** 76-80.

27. Cock-Rada A, Weitzman JB: **The methylation landscape of tumour metastasis.** *Biol Cell* 2013, **105:** 73-90.

28. Dupont C, Armant DR, Brenner CA: **Epigenetics: definition, mechanisms and clinical perspective.** *Semin Reprod Med* 2009, **27:** 351-357.

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

30. Lao VV, Grady WM: **Epigenetics and colorectal cancer.** *Nat Rev Gastroenterol Hepatol* 2011, **8:** 686-700.

31. Yoo CB, Jones PA: **Epigenetic therapy of cancer: past, present and future.** *Nat Rev Drug Discov* 2006, **5:** 37-50.

32. Nogueira da CA, Herceg Z: **Detection of cancer-specific epigenomic changes in biofluids: powerful tools in biomarker discovery and application.** *Mol Oncol* 2012, **6:** 704-715.

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

34. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O *et al.*: **Charting a dynamic DNA methylation landscape of the human genome.** *Nature* 2013, **500:** 477-481.

35. McCabe MT, Brandes JC, Vertino PM: **Cancer DNA methylation: molecular mechanisms and clinical implications.** *Clin Cancer Res* 2009, **15:** 3927-3937.

36. Smith ZD, Meissner A: **DNA methylation: roles in mammalian development.** *Nat Rev Genet* 2013, **14:** 204-220.

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

38. Kohli RM, Zhang Y: **TET enzymes, TDG and the dynamics of DNA demethylation.** *Nature* 2013, **502:** 472-479.

39. Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, Tsai SQ *et al.*: **Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins.** *Nat Biotechnol* 2013, **31:** 1137-1142.

40. Ponnaluri VK, Maciejewski JP, Mukherji M: **A mechanistic overview of TET-mediated 5-methylcytosine oxidation.** *Biochem Biophys Res Commun* 2013, **436:** 115-120.

41. Dawson MA, Kouzarides T: **Cancer epigenetics: from mechanism to therapy.** *Cell* 2012, **150:** 12-27.

42. Jones PA: **Functions of DNA methylation: islands, start sites, gene bodies and beyond.** *Nat Rev Genet* 2012, **13:** 484-492.

43. Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z *et al.*: **DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA.** *Nature* 2007, **448:** 714-717.

44. Piperi C, Papavassiliou AG: **Strategies for DNA methylation analysis in developmental studies.** *Dev Growth Differ* 2011, **53:** 287-299.

45. Yamamoto Y, Abe A, Emi N: **Clarifying the impact of polycomb complex component disruption in human cancers.** *Mol Cancer Res* 2014, **12:** 479-484.

46. Issa JP: **Aging and epigenetic drift: a vicious cycle.** *J Clin Invest* 2014, **124:** 24-29.

47. Delpu Y, Cordelier P, Cho WC, Torrisani J: **DNA methylation and cancer diagnosis.** *Int J Mol Sci* 2013, **14:** 15029-15058.

48. Nie ZL, Pan YQ, He BS, Gu L, Chen LP, Li R *et al.*: **Gene therapy for colorectal cancer by an oncolytic adenovirus that targets loss of the insulin-like growth factor 2 imprinting system.** *Mol Cancer* 2012, **11:** 86.

49. Ahuja N, Easwaran H, Baylin SB: **Harnessing the potential of epigenetic therapy to target solid tumors.** *J Clin Invest* 2014, **124:** 56-63.

50. Coppede F, Lopomo A, Spisni R, Migliore L: **Genetic and epigenetic biomarkers for diagnosis, prognosis and treatment of colorectal cancer.** *World J Gastroenterol* 2014, **20:** 943-956.

51. Timp W, Feinberg AP: **Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host.** *Nat Rev Cancer* 2013, **13:** 497-510.

52. Feinberg AP, Vogelstein B: **Hypomethylation distinguishes genes of some human cancers from their normal counterparts.** *Nature* 1983, **301:** 89-92.

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

54. Kouzarides T: **Chromatin modifications and their function.** *Cell* 2007, **128:** 693-705.

55. Cosgrove MS: **Writers and readers: deconvoluting the harmonic complexity of the histone code.** *Nat Struct Mol Biol* 2012, **19:** 739-740.

56. Chen T, Dent SY: **Chromatin modifiers and remodellers: regulators of cellular differentiation.** *Nat Rev Genet* 2014, **15:** 93-106.

57. Wilson BG, Roberts CW: **SWI/SNF nucleosome remodellers and cancer.** *Nat Rev Cancer* 2011, **11:** 481-492.

58. Sandoval J, Peiro-Chova L, Pallardo FV, Garcia-Gimenez JL: **Epigenetic biomarkers in laboratory diagnostics: emerging approaches and opportunities.** *Expert Rev Mol Diagn* 2013, **13:** 457-471.

59. Biterge B, Schneider R: **Histone variants: key players of chromatin.** *Cell Tissue Res* 2014.

60. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV *et al*.: **Signatures of mutational processes in human cancer.** *Nature* 2013, **500:** 415-421.

61. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW: **Cancer genome landscapes.** *Science* 2013, **339:** 1546-1558.

62. Weisenberger DJ: **Characterizing DNA methylation alterations from The Cancer Genome Atlas.** *J Clin Invest* 2014, **124:** 17-23.

63. Samuels Y, Ericson K: **Oncogenic PI3K and its role in cancer.** *Curr Opin Oncol* 2006, **18:** 77-82.

64. Liu S, Knapp S, Ahmed AA: **The structural basis of PI3K cancer mutations: from mechanism to therapy.** *Cancer Res* 2014, **74:** 641-646.

65. Gao J, Ciriello G, Sander C, Schultz N: **Collection, integration and analysis of cancer genomic profiles: from data to insight.** *Curr Opin Genet Dev* 2014, **24C:** 92-98.

66. Meyerson M, Gabriel S, Getz G: **Advances in understanding cancer genomes through second-generation sequencing.** *Nat Rev Genet* 2010, **11:** 685-696.

67. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, Yilmaz E *et al.*: **IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype.** *Nature* 2012, **483:** 479-483.

68. Erren TC, Gross JV, Koch MS, Erren M, Meyer-Rochow VB: **What do we know 40 years after Nixon declared the 'war on cancer'? On the origin, prevention and treatment of cancer.** *J Cancer Educ* 2012, **27:** 597-600.

69. Heichman KA, Warren JD: **DNA methylation biomarkers and their utility for solid cancer diagnostics.** *Clin Chem Lab Med* 2012, **50:** 1707-1721.

70. Lange CP, Laird PW: **Clinical applications of DNA methylation biomarkers in colorectal cancer.** *Epigenomics* 2013, **5:** 105-108.

71. Prensner JR, Rubin MA, Wei JT, Chinnaiyan AM: **Beyond PSA: the next generation of prostate cancer biomarkers.** *Sci Transl Med* 2012, **4:** 127rv3.

72. Treppendahl MB, Kristensen LS, Gronbaek K: **Predicting response to epigenetic therapy.** *J Clin Invest* 2014, **124:** 47-55.

73. Steele CB, Rim SH, Joseph DA, King JB, Seeff LC: **Colorectal cancer incidence and screening - United States, 2008 and 2010.** *MMWR Surveill Summ* 2013, **62 Suppl 3:** 53-60.

74. Vaiopoulos AG, Athanasoula KC, Papavassiliou AG: **Epigenetic modifications in colorectal cancer: Molecular insights and therapeutic challenges.** *Biochim Biophys Acta* 2014, **1842:** 971-980.

75. Grutzmann R, Molnar B, Pilarsky C, Habermann JK, Schlag PM, Saeger HD *et al.*: **Sensitive detection of colorectal cancer in peripheral blood by septin 9 DNA methylation assay.** *PLoS One* 2008, **3:** e3759.

76. Devos T, Tetzner R, Model F, Weiss G, Schuster M, Distler J *et al.*: **Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer.** *Clin Chem* 2009, **55:** 1337-1346.

77. Tanzer M, Balluff B, Distler J, Hale K, Leodolter A, Rocken C *et al.*: **Performance of epigenetic markers SEPT9 and ALX4 in plasma for detection of colorectal precancerous lesions.** *PLoS One* 2010, **5:** e9061.

78. Warren JD, Xiong W, Bunker AM, Vaughn CP, Furtado LV, Roberts WL *et al.*: **Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer.** *BMC Med* 2011, **9:** 133.

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

80. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP *et al.*: **Multitarget stool DNA testing for colorectal-cancer screening.** *N Engl J Med* 2014, **370:** 1287-1297.

81. 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.

82. Ahmed D, Danielsen SA, Aagesen TH, Bretthauer M, Thiis-Evensen E, Hoff G *et al.*: **A tissue-based comparative effectiveness analysis of biomarkers for early detection of colorectal tumors.** *Clin Transl Gastroenterol* 2012, **3:** e27.

83. Campbell RM, Tummino PJ: **Cancer epigenetics drug discovery and development: the challenge of hitting the mark.** *J Clin Invest* 2014, **124:** 64-69.

84. Navada SC, Steinmann J, Lubbert M, Silverman LR: **Clinical development of demethylating agents in hematology.** *J Clin Invest* 2014, **124:** 40-46.

85. Kaminskas E, Farrell AT, Wang YC, Sridhara R, Pazdur R: **FDA drug approval summary: azacitidine (5-azacytidine, Vidaza) for injectable suspension.** *Oncologist* 2005, **10:** 176-182.

86. West AC, Johnstone RW: **New and emerging HDAC inhibitors for cancer treatment.** *J Clin Invest* 2014, **124:** 30-39.

87. 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.

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

89. 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.

90. 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.

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

92. 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.

93. Heid CA, Stevens J, Livak KJ, Williams PM: **Real time quantitative PCR.** *Genome Res* 1996, **6:** 986-994.

94. 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.

95. 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.

96. Caballero OL, Chen YT: **Cancer/testis (CT) antigens: potential targets for immunotherapy.** *Cancer Sci* 2009, **100:** 2014-2021.

97. Feng S, Jacobsen SE, Reik W: **Epigenetic reprogramming in plant and animal development.** *Science* 2010, **330:** 622-627.

98. Kennedy JA, Barabe F: **Investigating human leukemogenesis: from cell lines to in vivo models of human leukemia.** *Leukemia* 2008, **22:** 2029-2040.

99. de WM, Fijneman RJ, Verheul HM, Meijer GA, Jimenez CR: **Proteomics in colorectal cancer translational research: biomarker discovery for clinical applications.** *Clin Biochem* 2013, **46:** 466-479.

100. Marx V: **Cell-line authentication demystified.** *Nat Methods* 2014, **11:** 483-488.

101. Antequera F, Boyes J, Bird A: **High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines.** *Cell* 1990, **62:** 503-514.

102. Jones PA, Wolkowicz MJ, Rideout WM, III, Gonzales FA, Marziasz CM, Coetzee GA *et al.*: **De novo methylation of the MyoD1 CpG island during the establishment of immortal cell lines.** *Proc Natl Acad Sci U S A* 1990, **87:** 6117-6121.

103. Ahmed D, Danielsen SA, Aagesen TH, Bretthauer M, Thiis-Evensen E, Hoff G *et al.*: **A tissue-based comparative effectiveness analysis of biomarkers for early detection of colorectal tumors.** *Clin Transl Gastroenterol* 2012, **3:** e27.

104. Andresen K, Boberg KM, Vedeld HM, Honne H, Hektoen M, Wadsworth CA *et al.*: **Novel target genes and a valid biomarker panel identified for cholangiocarcinoma.** *Epigenetics* 2012, **7:** 1249-1257.

105. 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.

106. Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F *et al.*: **Dynamic DNA methylation across diverse human cell lines and tissues.** *Genome Res* 2013, **23:** 555-567.

107. Fernandez AF, Assenov Y, Martin-Subero JI, Balint B, Siebert R, Taniguchi H *et al.*: **A DNA methylation fingerprint of 1628 human samples.** *Genome Res* 2012, **22:** 407-419.

108. Palmer RE, Lee SB, Wong JC, Reynolds PA, Zhang H, Truong V *et al.*: **Induction of BAIAP3 by the EWS-WT1 chimeric fusion implicates regulated exocytosis in tumorigenesis.** *Cancer Cell* 2002, **2:** 497-505.

109. Muller-Tidow C, Ji P, Diederichs S, Potratz J, Baumer N, Kohler G *et al.*: **The cyclin A1-CDK2 complex regulates DNA double-strand break repair.** *Mol Cell Biol* 2004, **24:** 8917-8928.

110. Maxwell SA, Davis GE: **Differential gene expression in p53-mediated apoptosis-resistant vs. apoptosis-sensitive tumor cell lines.** *Proc Natl Acad Sci U S A* 2000, **97:** 13009-13014.

111. Deaton AM, Bird A: **CpG islands and the regulation of transcription.** *Genes Dev* 2011, **25:** 1010-1022.

112. Rettori MM, de Carvalho AC, Longo AL, de Oliveira CZ, Kowalski LP, Carvalho AL *et al.*: **TIMP3 and CCNA1 hypermethylation in HNSCC is associated with an increased incidence of second primary tumors.** *J Transl Med* 2013, **11:** 316.

113. Lim EH, Ng SL, Li JL, Chang AR, Ng J, Ilancheran A *et al.*: **Cervical dysplasia: assessing methylation status (Methylight) of CCNA1, DAPK1, HS3ST2, PAX1 and TFPI2 to improve diagnostic accuracy.** *Gynecol Oncol* 2010, **119:** 225-231.

114. Li Y, Xu J, Ju H, Xiao Y, Chen H, Lv J *et al.*: **A network-based, integrative approach to identify genes with aberrant co-methylation in colorectal cancer.** *Mol Biosyst* 2014, **10:** 180-190.

115. Brait M, Begum S, Carvalho AL, Dasgupta S, Vettore AL, Czerniak B *et al.*: **Aberrant promoter methylation of multiple genes during pathogenesis of bladder cancer.** *Cancer Epidemiol Biomarkers Prev* 2008, **17:** 2786-2794.

116. Starlard-Davenport A, Kutanzi K, Tryndyak V, Word B, Lyn-Cook B: **Restoration of the methylation status of hypermethylated gene promoters by microRNA-29b in human breast cancer: A novel epigenetic therapeutic approach.** *J Carcinog* 2013, **12:** 15.

117. Campan M, Moffitt M, Houshdaran S, Shen H, Widschwendter M, Daxenbichler G *et al.*: **Genome-scale screen for DNA methylation-based detection markers for ovarian cancer.** *PLoS One* 2011, **6:** e28141.

118. Juang HH, Chung LC, Sung HC, Feng TH, Lee YH, Chang PL *et al.*: **Metallothionein 3: an androgen-upregulated gene enhances cell invasion and tumorigenesis of prostate carcinoma cells.** *Prostate* 2013, **73:** 1495-1506.

119. Peng D, Hu TL, Jiang A, Washington MK, Moskaluk CA, Schneider-Stock R *et al.*: **Location-specific epigenetic regulation of the metallothionein 3 gene in esophageal adenocarcinomas.** *PLoS One* 2011, **6:** e22009.

Appendix

APPENDIX I

Primer design MSP

Gene name	Methylated sense	Methylated antisense	Unmethylated sense	Unmethylated antisense	°C	Mg2+ (mM)	Fragment length	Location
ABAT	GGAGTTAGGGAGTTCGGC	CGTCTCCGATCTACGAAATC	GGAGTTAGGGAGTTTGGT	CATCTCCAATCTACAAAATC	53	1,5	121	-9 to +112
ACRC	AATTGGTTTCGGGTGTTTAC	CTTCCGAAATCAAATCTATCG	TAGAATTGGTTTTGGGTGTTTAT	CTTCCAAAATCAAATCTATCACTT	49	1,5	100	+151 to +251
ASAP3	GGTCGTTTTAGTTGAGGAGC	ACACTAAACTACTCCGCGCT	GGGTTGTTTTAGTTGAGGAGT	AACACTAAACTACTCCACACT	58	1,5	133	-52 to +81
ATG2A	GTTTGGCGTTTAGAGAAGC	AACCATCGTAACATCTCGAA	GGTTTGGTGTTTAGAGAAGT	CAACCATCATAACATCTCAAA	56	1,5	129	-3 to +126
AVPI1	TTTATCGATGATTGGCGTAC	ACTTAAATCACAACCGCGAT	TTTTTTATTGATGATTGGTGTAT	ACCACTTAAATCACAACCACAAT	53	1,5	139	-46 to +93
BAIAP3_i1	GGGAGCGTTGCGTAGATAC	CGAACCCCGAAAATCTACAT	GGAGGGAGTGTTGTGTAGATAT	CAAACCCCAAAAATCTACATTCT	58	1,5	165	-16 to +149
BAIAP3_i3-5	TTCGGTTCGGTTTAAGGTAC	GCCGACGACTAAACAACTACT	GTTTTGGTTTGGTTTAAGGTAT	ACCAACAACTAAACAACTACTAA	57	1,5	168	-121 to +47
C11orf67	TCGAGAATAATGGTTAATGGAC	ACGCTACTTCGAAATCTACG	TTTTTGAGAATAATGGTTAATGGAT	CAACACTACTTCAAAATCTACAC	53	1,5	108	-99 to +9
C11orf96	TTTTTTCGGAATTTTCGC	ATACCATTTAAAACGTACGCG	GGTTTTTTTTGGAATTTTTGT	ATACCATTTAAAACATACACACCC	M 54,6/ U 59	1,5	147	-143 to +4
CCNA1_iab	GGTGAGTAGGTCGTTATGGC	AAAACGACCGAACTTAACGA	TTAGGTGAGTAGGTTGTTATGGT	AAAACAACCAAACTTAACAATAC	49	1,5	155	-108 to +47
CCNA1_ic	TGGAGACGTAATATTGTCGC	CCTCGAAACTACGCAAACT	GATGGAGATGTAATATTGTTGT	CCTCAAAACTACACAAACTAC	53	1,5	153	-132 to +21
CGN	TTTTGTTCGAGTTCGAGTTC	TCGCTTTCCTCTCGAATAC	TTGTTTTGTTTGAGTTTGAGTTT	TCACTTTCCTCTCAAATACCCT	57	1,5	128	+43 to +171
CHGB	GTAGGTTGAGGTATTCGAAGC	AAACGAAAATTTACGCCAAT	GGAGTAGGTTGAGGTATTTGAAGT	AAACAAAAATTTACACCAATCCA	54,6	1,5	154	-113 to +41
CKB	GTTTTAGCGAGGGGATAGTTC	CGCGACTCTTAAAAAACACA	GGGTTTTAGTGAGGGGATAGTTT	CCACAACTCTTAAAAAACACAA	59	1,5	158	-151 to +7
CLU	TTATATTAGGACGGATGGGC	AACTTCCTAAACTCCGTCGA	GAGTTATATTAGGATGGATGGGT	CCTAACTTCCTAAACTCCATCAA	55	1,7	103	+190 to +293
CPM	CGTCGTTTTTTTTAGGTGC	CCGAATATTATAACCCGATCA	GGGTGTTGTTTTTTTTAGGTGT	AAACCAAATATTATAACCCAATCA	56	1,5	134	-126 to +8
CRISPLD2	TCGGGGAGTTTATAAGAGGC	AAACACAACGAACACAACGA	GTTTTGGGGAGTTTATAAGAGGT	AAACACAACAAACACAACAAACA	57	1,5	108	-41 to +67
CYP1A1	TTCGGGAAGGAGGTTATTAC	ACCCACTAAAACGCTAAACG	AATTTTGGGAAGGAGGTTATTAT	CACCCACTAAAACACTAAACATA	54,6	1,5	102	-475 to -372
DNER	GATGTTTTAGAGTCGGGGTC	ACGCCTCCTACAACTACGA	GATGTTTTAGAGTTGGGGTT	ACACCTCCTACAACTACAA	57	1,5	120	-93 to +27
EFHD1_i1	GGTAGTCGTTGAGTTTTGGC	CGCAAACTCGAAACTCTACA	GTTGGTAGTTGTTGAGTTTTGGT	CCTCACAAACTCAAAACTCTACA	58	1,5	137	-6 to +131
EFHD1_i2	GGCGGTATTTAGTAGAGGGC	CGAAAAAAAACAAACAACCG	GGTGGTATTTAGTAGAGGGT	CAAAAAAAAACAAACAACCA	52	1,5	136	-89 to +47
FA2H	TTTTCGGGATGTTAGAGGC	GCTAAACCTCGAAAAACGAA	GGGTTTTTGGGATGTTAGAGGT	CACTAAACCTCAAAAAACAAAAA	54,6	1,5	130	-5 to +125
FAM210B	TTTCGGTTTTCGAGGAGC	CTAACCCGACGAAAACCCT	GGATTTTGGTTTTTGAGGAGT	AACTAACCCAACAAAAACCCTC	58	1,5	131	-127 to +4
FHL1_t1,2,6	TTTTTAGTGGCGGGGGTAC	CCGAACTCTACCCACTACACG	TTTTTTTAGTGGTGGGGGTAT	CCAAACTCTACCCACTACACAAA	56	1,5	125	-145 to -20
FHL1_t3	GCGTGATTTTTTTCGTTTC	CGACACAACAATCGCTTAA	AGGGTGTGATTTTTTTTGTTTT	ACAACACAACAATCACTTAACC	53	1,5	172	-69 to +103
FHL1_t8	ATCGGTGTTTTTTGAGTGAC	CGAAACCCGATAACTTACAA	GTTATTGGTGTTTTTTGAGTGAT	TCAAAACCCAATAACTTACAACA	53	1,5	145	+103 to 248
FUCA1	TAGTTAGAGTGGGCGGAGTC	CGAAAACTAAACCCGACG	GTTAGTTAGAGTGGGTGGAGTT	CAAAAACTAAACCCAACACA	58	1,5	173	+9 to +182
GSN_1	TTATTTAAGGTCGGCGATTC	GAACGAACAACGACAACG	GGTTATTTAAGGTTGGTGATTT	CAAACAAACAACAACAACAC	57	1,5	139	-5 to +134
GSN_2	GGCGAGTTTGGAGTTATTC	CTACGCTCAACTAATTCCGA	GAAGGTGAGTTTGGAGTTATTT	ACTACACTCAACTAATTCCAAAC	56	1,5	170	-95 to +75
HABP4	GTTTTGGTCGTCGGTTTC	AACGACAACCACGAAACTC	GGTTTTGGTTGTTGGTTTT	CAACAACAACCACAAAACTC	56	1,5	108	+73 to +181
HIST1H2BL	AAGTTGTGCGATTGGTTTAC	TCCGTAAATAACGACAAATACC	AGGAAGTTGTGTGATTGGTTTAT	TCCATAAATAACAACAAATACCAAT	56	1,5	108	-191 to -83
ID2	ATTAATGGAAGCGTTCGTTC	AAACGACTTTTATCCGCACT	TTTATTAATGGAAGTGTTTGTTT	ACAAAACAACTTTTATCCACACT	54,6	1,7	106	-53 to +53
IFFO1	TCGGAAAGGTTTAGGGTTC	ACCAATCAAACGACTCGAC	GATTTGGAAAGGTTTAGGGTTT	CCAACCAATCAAACAACTCAAC	57	1,5	100	-143 to -43
INPP5F_1	GCGTTTTTTTTTTATCGGTC	CACGAAACGTCCTAATCG	GTGTTTTTTTTTTATTGGTT	CACAAAACATCCTAATCA	51	1,5	127	+43 to +170
INPP5F_2	TCGAGTTTAAGTTTGAGGGC	CGCATAAAAACCGCTATCTT	TAGTTGAGTTTAAGTTTGAGGGT	CACATAAAAACCACTATCTTCAA	54,6	1,5	161	-389 to -248
KCNH2	GCGCGTCGTATTAATTTTTTC	CCGAACTCCTAACTCCCG	GTGTGTTGTATTAATTTTTTT	CCAAACTCCTAACTCCCA	51	1,5	108	-22 to +86
KIAA0513	ACGTGATATTAAGCGTTGGTC	CGAACTATCAATCAACCGAA	GTTATGTGATATTAAGTGTTGGTT	ACACAAACTATCAATCAACCAAA	55	1,7	107	-160 to -53
MAPRE3	CGGTTTAGTTGTTTTGGAGAC	AACGCACAAAAACTACGAAA	GTGGTTTAGTTGTTTTGGAGAT	CAACACACAAAAACTACAAAA	54,6	1,5	113	-160 to +9

Gene name	Methylated sense	Methylated antisense	Unmethylated sense	Unmethylated antisense	°C	Mg2+ (mM)	Fragment length	Location
MT3	TGGTAGTGCGCGTATAGC	GACGCACGCACTACATCTAT	GGTTGGTAGTGTGTGTATAGT	TCCAACACACACACTACATCTAT	53	1,5	143	-93 to +50
NDRG4_t1	GCGTAGAAGGCGGAAGTTAC	CCCGCGTAAATTTAACGAAT	GTGTAGAAGGTGGAAGTTAT	CCCACATAAATTTAACAAAT	51	1,5	109	-113 to -4
NDRG4_t2	GCGATTTTTTATCGCGAC	ATTACATAACCGCGCACC	GTGATTTTTTATTGTGAT	ATTACATAACCACACACC	52	1,5	175	-39 to +136
NDRG4_t3	GGTTTTTGGTAGGATTCGC	ACTTAACGTCGCCTCCACT	TGGGTTTTTGGTAGGATTTGT	CCACTTAACATCACCTCCACT	54,6	1,5	112	-15 to +97
NR4A1	TTTTCGTCGGAATCGTATC	CCTCTTAAACGCTCCGTAAC	GTTTTTTGTTGGAATTGTATT	CTCCTCTTAAACACTCCATAAC	58,8	1,5	112	-95 to +17
NRGN	GCGGGGTTTAGGTTATATTC	ATCCTCTACGAAAACGAAAAC	GTGGGGTTTAGGTTATATTT	ATCCTCTACAAAAACAAAAAC	53	1,5	170	-35 to +135
PMEL	AACCAAACCCAATCCATACTACA	ATATACTACGTTCCATCCCGA	TTTTTATGAAGTGTTAGTGAGTT	ATATACTACATTCCATCCCAACCC	59	M 2,0 / U 1,7	147	-953 to -806
PRKAR2B	ATTATACGGAGTAGACGCGC	AACGCCTACGACGCTAAC	GTTATTATATGGAGTAGATGTGT	ACAAACACCTACAACACTAAC	57	1,5	107	-16 to +91
PRPH	AAGGCGTTTTTGTTGTTTC	CCGCTACGAAACGACTTTAT	GGAAGGTGTTTTTGTTGTTTT	AACCACTACAAAACAACTTTAT	56	1,5	137	-143 to -43
RBP7	TTGGTTTATAGGTTTCGGTTC	ATAACCCTCGAAATTATCGC	GTTTGGTTTATAGGTTTTGGTTT	ATAACCCTCAAAATTATCACTA	54,6	1,5	124	-5 to +119
SCPEP1	GTGCGGGTTTTAGTAAGTTGTC	CATCACGTAACGAAAACGAA	GTGTGGGTTTTAGTAAGTTGTT	CATCACATAACAAAAACAAA	53	1,5	118	-115 to +3
SELM	GTAGCGATTCGGAGGTTC	AATTAAACCGCTCAAACGAT	GTAGTGATTTGGAGGTTT	AATTAAACCACTCAAACAAT	53	1,5	168	+3 to +171
SLC17A7	CGTGAGTTTAGAATTTATAGGC	ACAAAACTCAACGCCGAA	AGGTGTGAGTTTAGAATTTATAGGT	ACAAAACTCAACACCAAAAAC	54,6	1,5	160	-160 to 0
SYCP3	AATTGGTTTTCGGTTATTGC	CCGTTTTCTCGTCAAAAAC	AGTAATTGGTTTTTGGTTATTGT	CCATTTTCTCATCAAAAACCCC	51	1,5	146	+87 to +233
TESK2	GGGTAGGCGTTTTAGTTTTTC	CGCCTATTTAACGAAACCA	GGGTAGGTGTTTTAGTTTTTT	CACCTATTTAACAAAACCA	54,6	1,5	104	-14 to +90
TM7SF2	AAAATTTTGGGACGAGAGC	CCAAAAAATAAACTCGACCG	TTAAAAATTTTGGGATGAGAGT	TCCCAAAAAATAAACTCAACCAC	56	1,5	100	-11 to +89
TSPAN7	CGCGTTTACGAGGTTTTTC	TACTACAAAACTCCGACGACG	TTGTGTGTTTATGAGGTTTTTT	TACTACAAAACTCCAACAACAACA	53	1,5	129	-62 to +67
VGF	CGCGTTTTTTTTTATAAGGC	CGATCGAAATCTAACGTCC	GTGTGTTTTTTTTTATAAGGT	ACAATCAAAATCTAACATCC	53	1,5	127	-43 to +84
WDR47	TCGTTTTAGTTTTGGGTTC	CGAACTCTACCCGCGAAC	TTTTTGTTTTAGTTTTGGGTTT	TTTTTGTTTTAGTTTTGGGTTT	54,6	1,5	162	-60 to +102
ZCCHC12	GTAGAGGATAAAAGTCGTCGC	GCTAAACGACCCGAAACTA	TGGTAGAGGATAAAAGTTGTTGT	CACTAAACAACCCAAAACTAC	53	1,5	114	-39 to +75

APPENDIX II

Primer design bisulfite sequencing

Gene name	Primer sense	Primer antisense	°C	Mg2+ (mM)	Fragment lenght	Location
BAIAP3_i1	TTTTTATAGAATTGGGGGTGT	CTCCTCCAAATAACCCCTAC	53	1.5	365	-118 to +244
CCNA1_iab	GTATAGTTGGAGTTGGAGGGT	TCCRAAACAACTAACAAATACA	54,6	1.5	430	-193 to +237
CCNA1_ic	TTTTTTTGGGTTATAATTTTGG	CCRCTCCTAAAAACCCTA	53	1.5	386	-325 to +61
IFFO1	TAGGGGAGATGGGTTTGTTA	AAACRAACCAAAAAAACCAATA	53	1.5	259	-255 to +4
MT3	TTTGGAGAATTTTAGAATGAAGG	CRAACTTCTCCAAACAACTAAA	53	1.5	456	-254 to +202

APPENDIX III

Primer design qMSP

Gene name	Primer sense	Primer antisense	Probe	Fragment lenght	Location
BAIAP3_i1	TCGGGTTTCGGTTTTTTGC	CCCCGAAACCTCAAACGAC	6FAM-GGCGTTAGCGGTTGG-MGB	72	+27 to +99
CCNA1_iab	TTGTCGCGGTCGGTATGG	CGACTAAACTACCCACCCGC	6FAM-CGTCGTTGATTGGTC-MGB	89	-65 to +24
CCNA1_ic	GATGGAGACGTAATATTGTCGCG	ACGCTTCCTACCGAAAAACGA	6FAM-TAGTTTCGTCGCGTTTTA-MGB	76	-134 to -58
IFFO1	TTCGGTTCGTTTTTTGATTTTTG	ACCAATACGACGCCGACG	6FAM-GTCGAGTCGTTTGATTG-MGB	78	-92 to -14
MT3	ATGGTACGTGCGCGTTTTTAC	CGCTATACGCGCACTACCAA	6FAM-CGGGATTTACGCGGGGA-MGB	97	-171 to -74

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

Gene symbol	Gene name	Location	Accession number
ABAT	4-aminobutyrate aminotransferase	16p13.2	NM_020686
ACRC	Acidic repeat containing	Xq13.1	NM_052957
ASAP3	ArfGAP with SH3 domain	1p36.13	NM_017707
ATG2A	Autophagy related 2A	11q13.1	NM_015104
AVPI1	Arginine vasopressine-induced 1	10q24.2	NM_021732
BAIAP3	BAI1-associated protein 3	16p13.3	**BAIAP3_i1** (NM_003933.4) and **BAIAP3_i3-5** (NM_001199097.1, NM_001199098.1, NM_001199099.1, NM_001286464.1)
C11orf67 (AAMDC)	Adipogenesis associated Mth938 domain containing	11q14.1	NM_024684
C11orf96	Chromosome 11 open reading frame 96	11p11.2	NM_001145033
CCNA1	Cyclin A1	13q12.3	**CCNA1_iab** (NM_003914.3 and NM_001111045.1) and **CCNA1_ic** (NM_001111046.1 and NM_001111047.1)
CGN	Cingulin	1q21	NM_020770
CHGB	Chromogranin B	20p12.3	NM_001819
CKB	Creatine kinase	14q32.32	NM_001823
CLU	Clusterin	8p21	NM_001831
CPM	Carboxypeptidase M	12q15	NM_001005502
CRISPLD2	Cysteine-rich secretory protein LCCL domain containing 2	16q24.1	NM_031476
CYP1A1	Cytochrome P450, family 1, subfamily A, polypeptide 1	15q24.1	NM_000499
DAZL	Deleted in azoospermia-like	3p24	**DAZL_i1** (NM_001190811.1) and **DAZL_i2** (NM_001351.3)
DNER	Delta/notch-like EGF repeat containing	2q36.3	NM_139072
EFHD1	EF-hand domain family, member D1	2q37.1	**EFHD1_i1** (NM_025202.3) and **EFHD1_i2 (NM_001243252.1**)
FA2H	Fatty-acid 2-hydroxylase	16q23	NM_024306
FAM210B	Family with sequence similarity 210, member B	20q13.2	NM_080821
FHL1	Four and a half LIM domains 1	Xq26.3	**FHL1_t1,2,6** (NM_001159702,NM_001449,NM_001159703), **FHL1_t3** (NM_001159700) and **FHL1_t8** (NR_027621)
FUCA1	Fucosidase	1p34	NM_000147
GSN	Gelsolin	9q34.11	**GSN_1** (NM_000177) and **GSN_2** (NM_198252 , NM_001127663, NM_001127664, NM_001127665, NM_001127666, NM_001127667)
HABP4	Hyaluronan binding protein 4	9q22.3	NM_014282
HIST1H2BL	Histone cluster 1, H2bl	6p22.1	NM_003519
ID2	Inhibitor of DNA binding 2	2p25	NM_002166
IFFO1	Intermediate filament family orphan 1	12p13.31	NM_080730
INPP5F	Inositol polyphosphate-5-phosphatase F	10q26.13	**INPPF_1** (NM_014937.3, NM_001243195.1) and **INPPF_2** (NM_001243194.1)

Gene symbol	Gene name	Location	Accession number
ITGA7	Integrin, alpha 7	12q13	NM_001144996
KCNH2	Potassium voltage-gated channel, subfamily H, member 2	7q36.1	NM_000238
KIAA0513	KIAA0513	16q24.1	NM_014732
MAPRE3	Microtubule-associated protein, RP/EB family, member 3	2p23.3	NM_012326
MT3	Metallonthionein 3	16q13	NM_005954
NDRG4	NDRG family member 4	16q21	**NDRG4_t1** (NM_020465), **NDRG4_t2** (NM_001130487), **NDRG4_t3** (NM_022910)
NR4A1	Nuclear receptor subfamily 4, member 1	12q13	NM_002135
NRGN	Neurogranine	11q24	NM_006176
PMEL	Premelanosome protein	12q13	NM_001200054
PRKAR2B	Protein kinase, cAMP-dependent, regulatory, type II, beta	7q22.3	NM_002736
PRPH	Peripherin	12q12	NM_006262
RBP7	Retinol binding protein 7	1p36.22	NM_052960
SCPEP1	Serine carboxypeptidase 1	17q22	NM_021626
SELM	Selenoprotein M	22q12.2	NM_080430
SLC17A7	Solute carrier family 17, member 7	19q13.33	NM_020309
SYCP3	Synaptonemal complex protein 3	12q23.2	NM_001177949
TDRD6	Tudor domain containing 6	6p12.3	NM_001010870
TESK2	Testis-specific kinase 2	1p32	NM_007170
TM7SF2	Transmembrane 7 superfamily member 2	11q13.1	NM_003273
TSPAN7	Tetraspanin 7	Xp11.4	NM_004615
VGF	VGF nerve growth factor inducible	7q22.3	NM_003378
WDR47	WD repeat domain 47	1p13.3	NM_014969
ZCCHC12	Zinc finger, CCHC domain containing 12	Xq24	NM_173798

Appendix V – Poster presented at the annual meeting in our Norwegian centre of excellence, Cancer Biomedicine (CCB) 2013

Ane Brenna[1,2,3], Lina Cekaite[1,2], Hilde Honne[1,2], Rolf I. Skotheim[1,2], Mette Eknæs[1,2], Stine A. Danielsen[1,2], Ragnhild A. Lothe[1,2,4], Guro E. Lind[1,2,3]

[1]Department of Cancer Prevention, Institute for Cancer Research, Oslo University Hospital HE - Norwegian Radium Hospital; [2]Centre for Cancer Biomedicine, Faculty of Medicine, University of Oslo; [3]Department of Biosciences, Faculty of Mathematics and Natural Sciences, University of Oslo; [4]Faculty of Medicine, University of Oslo

Background and aims

Cancer is a heterogenous disease with hundreds of different genetic and epigenetic alterations. However, there are some abberations common in several types of cancer. Epigenetic changes are just as frequent as mutations in cancer development, and may occur at an early stage. The masterkeys project aims to identify common epigenetic drivers and to increase knowledge of cancer development, as well as to identify relevant epigenetic markers with clinical potential.

We have used epigenetic drugs (trichostatin A (TSA) and 5-aza-2'-deoxycytidine (aza)), combined with gene expression microarray and bioinformatic analysis to identify potential markers. The markers are currently being investigated in cancer cell lines by methylation specific PCR (MSP), and promising candidates will be further investigated in tissue by quantitative MSP (qMSP).

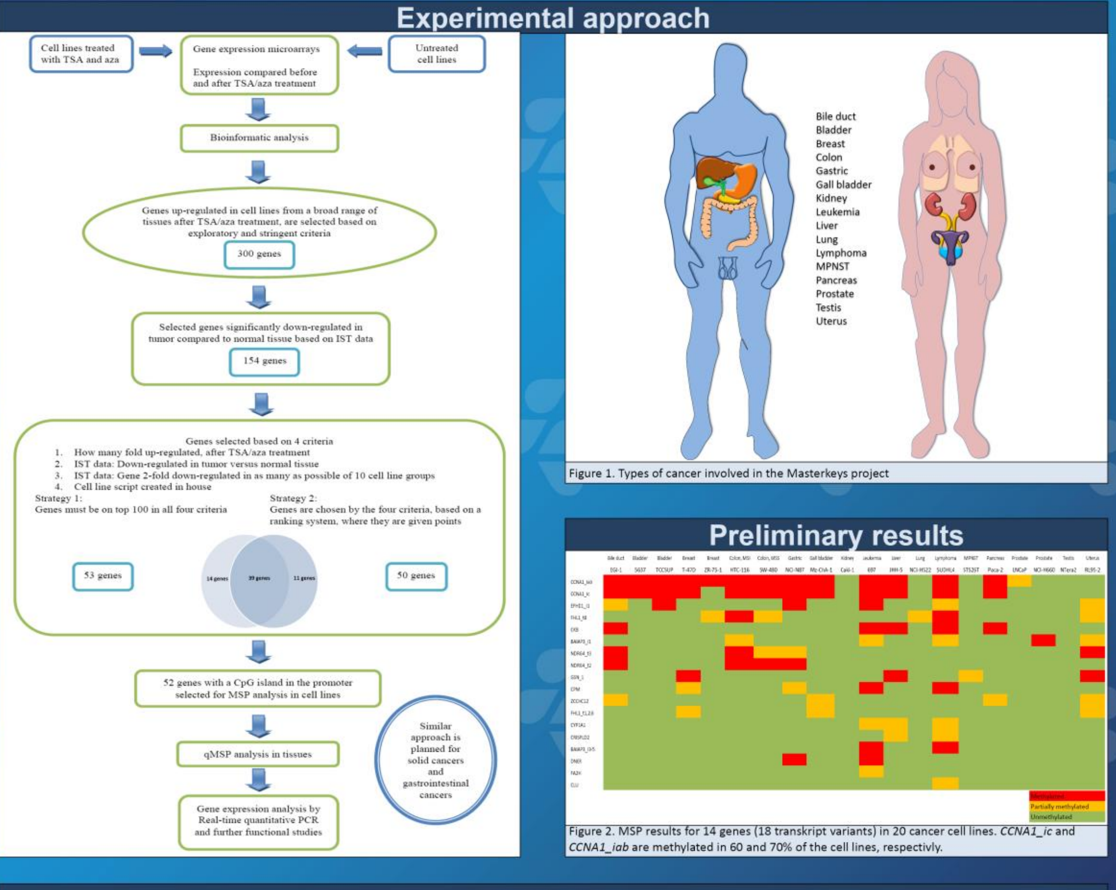


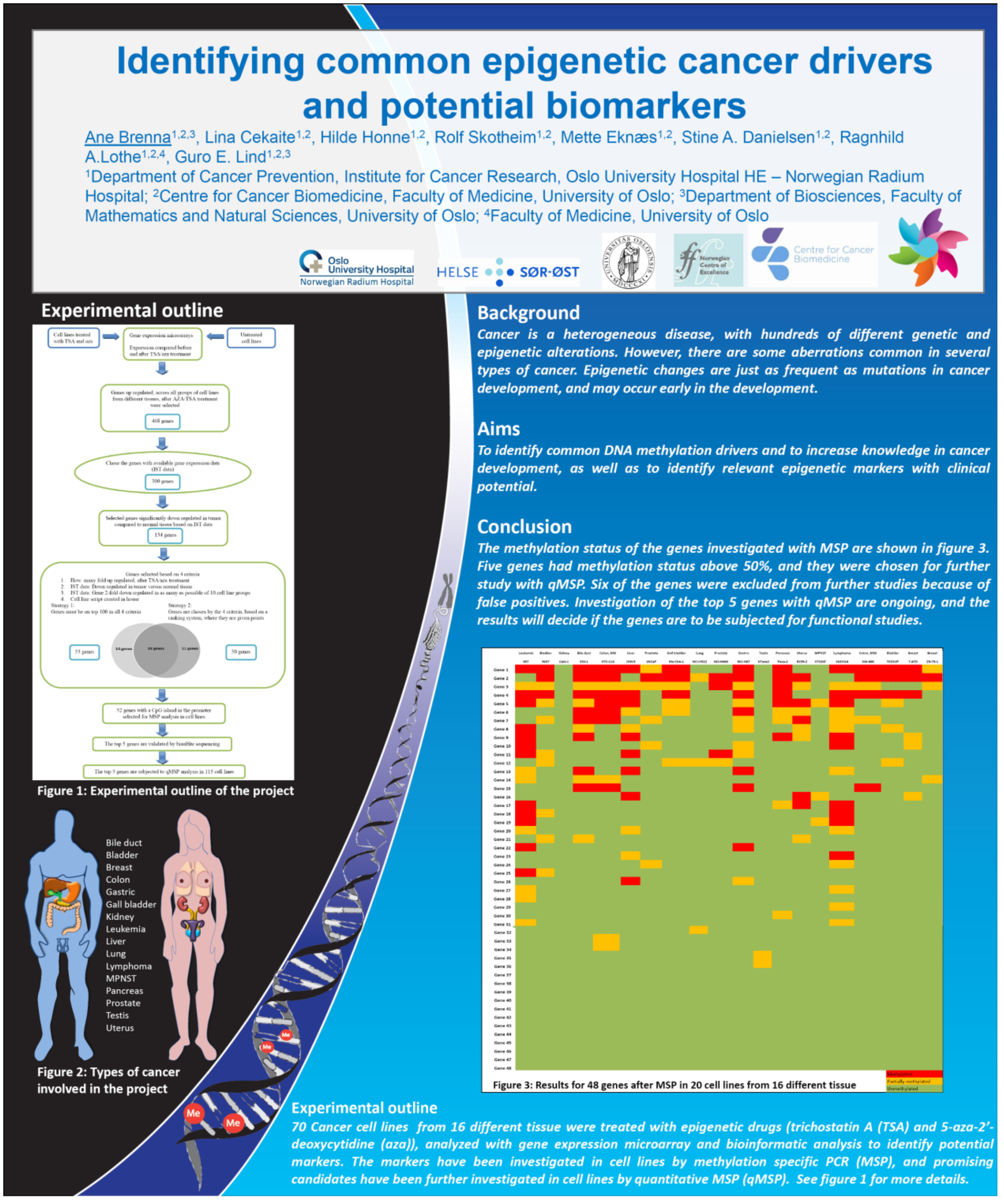
Figure 1. Types of cancer involved in the Masterkeys project

Figure 2. MSP results for 14 genes (18 transkript variants) in 20 cancer cell lines. *CCNA1_ic* and *CCNA1_iab* are methylated in 60 and 70% of the cell lines, respectivly.

Conclusion

The Masterkeys project is an ongoing project trying to identify key epigenetically regulated genes in cancer. So far 21 out of 52 genes have been examined by MSP in 20 cell lines, and two transcript variants of one gene have been found frequently methylated. Promising candidates will be validated with qMSP in tissue samples. Genes methylated across multiple cancer types or in cancer subgroups may have an important role in cancer development and a potential as epigenetic markers.

Appendix VI – Poster presented at Oslo Epigenetic Symposium, April 2014



Appendix VII – Word cloud generated by the use of tagxedo.com[14] based on words used in this master thesis

[14] http://www.tagxedo.com/