EMPLOYING EPIGENETIC MARKS TO DETECT CANCER. STUDIES ON NASOPHARYNGEAL CARCINOMA AND LUNG CANCER

Imran Nawaz

Stockholm 2015
All previously published papers were reproduced with permission from the publishers.

Published and printed by Karolinska University Press
Box 200, SE-171 77 Stockholm, Sweden
© Imran Nawaz, 2015
To my parents and family
ABSTRACT

Tumor suppressor genes (TSGs) or oncogenes aberrantly methylated in transcription control regions during early carcinogenesis are potential tools for early detection of cancer. We have identified suitable genes and explored assays based on their methylation status aiming for early detection of cancer: nasopharyngeal carcinoma (NPC) and non-small cell lung cancer (NSCLC).

We established and developed further the “multiplex methylation specific-PCR (MMSP)” assay designed to detect the tumor-specific methylation status of several NPC-related genes (paper I). It provided information about the methylation status of multiple genes simultaneously in a single PCR with small amounts of tumor DNA derived from nasopharyngeal swabs. It was shown to be applicable with DNA from as few as 10 cells. The detection rate of NPC from nasopharyngeal swabs was 98%. The false positive rate was zero.

We employed the MMSP assay on NPC tumors from two other regions (Morocco and Italy) and compared our results with those on Chinese NPC patients from paper I (paper II). We also did a pilot study using sera from Italian and Chinese NPC. We updated the panel of MMSP markers and modified the assay to improve its applicability to NPC from different geographical locations. We could detect at least any one methylation marker gene in 97% of the EBNA1 positive samples with a specificity of 94%, while the results on sera were less informative than using swabs.

We used an established NotI microarray method to identify gene losses (by deletion or methylation) in chromosome 3 of NPC tumors (paper III). This chromosome is known to contain TSGs involved in many cancer types. Ten candidate TSGs were found. Among them, the CpG rich area in the promoter region of Integrin α9 (ITGA9) was confirmed to be hypermethylated in NPC by bisulfite cloned sequencing, bisulfite pyrosequencing and methylation specific PCR. ITGA9 was downregulated in NPC clinical samples and 5-aza-2’-deoxycytidine restored the expression of ITGA9 in NPC derived cell lines. The functional role of ITGA9 downregulation in NPC should be explored further.

We developed an MMSP assay for analysis of the methylation status of multiple potential TSGs in NSCLC samples (paper IV). Thirty-eight potential TSGs were selected, based on literature search, genome-wide CpG methylation and expression microarrays performed on NSCLC tissues and matched control tissues. After evaluation by methylation specific PCR (MSP) six of these genes were selected for inclusion into the MMSP assay. Subsequently, 70 NSCLC DNA samples with matched controls and 24 non-cancerous DNA samples were screened with this assay. With a cut off of methylation of at least any two of these marker genes 87% of the cancer samples were detected with a specificity of 94%. Early stage I or II NSCLC showed a 100% specificity and 86% sensitivity.

Key words: NPC, NSCLC, Methylation, TSG, EBV, ITGA9, MMSP
LIST OF PUBLICATIONS


III. Nawaz I*, Hu LF#, Du ZM*, Moumad K, Ignatyev I*, Pavlova TV, Kashuba V, Zabarovsky ER, Ernberg I#. Integrin α9 gene promoter is hypermethylated and downregulated in nasopharyngeal carcinoma. (Manuscript)


* These authors contributed equally as first authors.
# These senior authors contributed equally to this work.
# TABLE OF CONTENTS

1 INTRODUCTION .................................................................................................................. 1
   1.1 General introduction to cancer ......................................................................................... 1
      1.1.1 Cancer etiology ............................................................................................................ 1
      1.1.2 Cancer control strategies ............................................................................................. 2
      1.1.3 Hallmarks of cancer ..................................................................................................... 2
      1.1.4 Multistage nature of cancer initiation and development ............................................. 3
   1.2 Epigenetic mechanisms of gene regulation and their role in tumorigenesis ............. 3
      1.2.1 DNA methylation ........................................................................................................ 6
      1.2.2 Histone modification .................................................................................................. 11
      1.2.3 Non-coding RNAs ..................................................................................................... 14
      1.2.4 Chromatin remodeling ............................................................................................... 15
      1.2.5 Cross-talk between epigenetic mechanisms and tumorigenesis ............................... 16
   1.3 Introduction to nasopharyngeal carcinoma ................................................................. 18
      1.3.1 Anatomy and histopathology ....................................................................................... 18
      1.3.2 Epidemiology and etiology ......................................................................................... 18
      1.3.3 Clinical presentation and diagnosis ............................................................................ 19
      1.3.4 Treatment and prognosis ........................................................................................... 20
      1.3.5 Epigenetic aspect of nasopharyngeal carcinoma ..................................................... 20
   1.4 Introduction to lung cancer ............................................................................................. 24
      1.4.1 Epigenetic aspect of non-small cell lung cancer ....................................................... 25
2 AIMS OF THIS STUDY .......................................................................................................... 27
3 RESULTS AND DISCUSSION ............................................................................................... 28
   3.1 Paper I .............................................................................................................................. 28
   3.2 Paper II ............................................................................................................................ 29
   3.3 Paper III .......................................................................................................................... 30
   3.4 Paper IV .......................................................................................................................... 32
4 METHODS .............................................................................................................................. 34
   4.1 Extraction of DNA .......................................................................................................... 34
   4.2 Bisulfite treatment of DNA ............................................................................................. 34
   4.3 Methylation specific PCR ............................................................................................... 36
   4.4 Multiplex methylation specific PCR .............................................................................. 37
   4.5 Cell lines or tumor tissues ............................................................................................... 37
   4.6 Treatment of cells with 5-aza-2'-deoxycytidine ............................................................. 38
   4.7 Bisulfite sequencing of cloned PCR products ............................................................... 38
   4.8 Pyrosequencing of bisulfite converted DNA ................................................................. 38
   4.9 Immunostaining .............................................................................................................. 39
   4.10 NotI microarray ............................................................................................................. 39
   4.11 Methylation microarray ............................................................................................... 39
   4.12 Expression microarray .................................................................................................. 40
   4.13 Statistical analysis ........................................................................................................ 40
5 CONCLUDING REMARKS .................................................................................................... 41
6 ACKNOWLEDGMENTS ......................................................................................................... 43
7 REFERENCES .......................................................................................................................... 47
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-aza-C</td>
<td>5-aza-2’-deoxycytidine</td>
</tr>
<tr>
<td>5hmC</td>
<td>5-hydroxymethylcytosine</td>
</tr>
<tr>
<td>5mC</td>
<td>5-methylcytosine</td>
</tr>
<tr>
<td>CALCA</td>
<td>calcitonin-related polypeptide alpha</td>
</tr>
<tr>
<td>CGI</td>
<td>CpG island</td>
</tr>
<tr>
<td>COBRA</td>
<td>combined bisulfite restriction analysis</td>
</tr>
<tr>
<td>CpG</td>
<td>cytosine-guanine (dinucleotide)</td>
</tr>
<tr>
<td>CTCF</td>
<td>CCCTC-binding zinc finger factor</td>
</tr>
<tr>
<td>DAPK</td>
<td>death-associated protein kinase</td>
</tr>
<tr>
<td>DLEC1</td>
<td>deleted in lung and esophageal cancer 1</td>
</tr>
<tr>
<td>DNMT</td>
<td>DNA methyltransferase</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleotide triphosphate</td>
</tr>
<tr>
<td>EA</td>
<td>early antigen</td>
</tr>
<tr>
<td>EBNA</td>
<td>Epstein-Barr nuclear antigen</td>
</tr>
<tr>
<td>EBV</td>
<td>Epstein-Barr virus</td>
</tr>
<tr>
<td>GAPDH</td>
<td>glyceraldehyde-3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>HDAC</td>
<td>histone deacetylase</td>
</tr>
<tr>
<td>HOXA9</td>
<td>homeobox A9</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatographic</td>
</tr>
<tr>
<td>ITGA9</td>
<td>Integrin α9</td>
</tr>
<tr>
<td>KAT</td>
<td>lysine acetyltransferase</td>
</tr>
<tr>
<td>LMP</td>
<td>latent membrane protein</td>
</tr>
<tr>
<td>IncRNA</td>
<td>long ncRNA</td>
</tr>
<tr>
<td>MBD</td>
<td>methyl-CpG binding domain</td>
</tr>
<tr>
<td>MeDIP</td>
<td>methyl-DNA immunoprecipitation</td>
</tr>
<tr>
<td>miRNAs</td>
<td>micro RNAs</td>
</tr>
<tr>
<td>MMSP</td>
<td>multiplex methylation specific PCR</td>
</tr>
<tr>
<td>MRI</td>
<td>magnetic resonance imaging</td>
</tr>
<tr>
<td>MS-DGGE</td>
<td>methylation-specific denaturing gradient gel electrophoresis</td>
</tr>
<tr>
<td>MS-DHPLC</td>
<td>methylation-specific denaturing HPLC</td>
</tr>
<tr>
<td>MS-HRM</td>
<td>methylation-sensitive high resolution melting</td>
</tr>
<tr>
<td>Ms-SnuPE</td>
<td>methylation-sensitive single nucleotide primer extension</td>
</tr>
<tr>
<td>MS-SSCA</td>
<td>methylation- specific single-strand conformation analysis</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

1.1 General introduction to cancer

Cancer is one of the leading causes of death globally.\(^1\) It claimed 8.2 million human lives in 2012 while an estimated 14 million new cases of cancer were diagnosed in the same year.\(^2\) It has been estimated that the number of new cases may rise by about 70% over the next 20 years.\(^1\) The most common cancer related deaths globally are lung, liver, stomach, colorectal, breast and esophageal cancer.\(^2\) There are dramatic differences in the incidence of specific cancer types between different geographic regions.

Tumorigenesis is a multistep process that reflects genetic alterations driving the progressive transformation of normal human cells into highly malignant tumor cells. The resulting abnormal cells grow rapidly, do not respect their usual boundaries and thus may invade adjoining parts of the body and spread to other distant body organs. This spreading process of cancer is known as metastasis. Metastases are the main cause of deaths from cancer irrespective of the part the body from where cancer actually originated.

1.1.1 Cancer etiology

Cancer is thought to be initiated from a single cell. A normal cell undergoes multiple stages on its way to a cancer cell. This is also characteristic of the progression from a pre-cancerous lesion into a malignant tumor.

Tumorigenesis is the outcome of interplay between the genetic make-up of individual and different types of external carcinogens. These external carcinogenic factors can be categorized into four main types: physical carcinogens (e.g. ultraviolet and ionizing radiation); chemical carcinogens (e.g. asbestos, components of tobacco smoke, food and water contaminants, e.g. aflatoxin, arsenine); biological carcinogens (e.g. certain viruses, bacteria or parasites) and general life style factors (e.g. diet). Use of tobacco, alcohol consumption, unhealthy diet, obesity and physically inactive life are some common leading cancer risk factors worldwide.\(^1\) Aging is another important factor in cancer development. The risk of cancer development increases with age.\(^1\) This depends on the time needed for the multistep tumor progression, but also due to
an accumulation of risk exposures with age combined with less effective cellular repair mechanisms and immune defense at an older age.

1.1.2 Cancer control strategies

Reduction in global cancer incidence and mortality is only possible by implementing evidence-based approaches for cancer prevention, early detection and better treatment. Early diagnosis of cancer is the first step to better cancer management and its treatment. The majority of cancer patients are diagnosed at an advanced stage of the disease. Late diagnosis of cancer is a major reason for cancer related deaths worldwide. In case of late diagnosis, the chances of survival are lower because the disease is already at an advanced stage and the risk of metastatic spread is larger. Another reason for the high cancer mortality is lack of efficient treatments, particularly for advanced stages. Early detection of cancer is important for efficient treatment and thus is one key to reduce cancer related deaths. Early detection of cancer is also very relevant in the absence of cancer screening programs in low-resource settings.¹

1.1.3 Hallmarks of cancer

It was suggested that the genetic changes, that are common to most types of cancers, may be grouped into ten essential alterations in cell physiology that collectively dictate malignant growth, designated as the hallmarks of cancer.³,⁴ These hallmarks are presented schematically in figure 1. When acquired as physiological capabilities they make tumor cells resistant to defense mechanisms of the normal cell and the host.³,⁴

![Figure 1. The hallmarks of cancer. Cited from Hanahan and Weinberg 2011.⁴](image)
1.1.4 Multistage nature of cancer initiation and development

It has been suggested that tumor development proceeds like a Darwinian evolutionary process during which a sequence of genetic changes, each which results in the selective survival advantage and sequential acquisition of tumor properties by the cell – a process commonly referred to as tumor progression. Four to seven such changes in the genome are considered the minimum requirement to develop many types of cancers. The model for multistage development of colorectal cancer was the first and so far still the best well defined, suggested by Fearon and Vogelstein in 1990 (Figure 2).³

![Figure 2. Fearon and Vogelstein's model for the initiation and development of colorectal carcinoma.](image)

According to this model the colorectal carcinoma develops through a multistep process in which genetic and epigenetic alterations accumulate in a sequential (linear) order and result in the loss of tumor suppressor genes (TSGs) and aberrant activation of oncogenes affecting multiple pathogenic pathways.⁵, ⁶ The resulting deregulated expression of TSGs and oncogenes were thought to be behind the continued clonal selection and heterogeneity of the tumor cells.⁵ In line with this model the earlier changes during the course of cancer development could be a potential useful screening tool for colorectal carcinoma.⁷

1.2 Epigenetic mechanisms of gene regulation and their role in tumorigenesis

Historically, the word “epigenetics” was first coined by Conrad Waddington in 1942 to describe events that were not justifiable by genetic principles. According to him epigenetics is the branch of biology which deals with the study of causal connections between genes and their products that give rise to the phenotype.⁸ The increase in
molecular understanding of genome functions has demanded a refinement of the definition of epigenetics. Now epigenetics refers to heritable alterations of phenotype that are independent of DNA sequence, but still reversible.\textsuperscript{9-11} In principle cells in a multicellular organism share the same genotype but during cellular differentiation attain different phenotypes, distinct profiles of gene expression and cellular functions. This was first proposed by Waddington in 1957 (Figure 3).\textsuperscript{12} Thus epigenetics is a bridge between genotype and phenotype as they can be brought about by exogenous, environmental factors and thus may modify the phenotype by its bridging between genome and the environment.\textsuperscript{11, 13}

\textbf{Figure 3. Waddington's classical epigenetic landscape.} Cited from Waddington 1957.\textsuperscript{12}

Epigenetic alterations play a major role in a vast array of biological functions during normal development and growth.\textsuperscript{14-19} Epigenetic mechanisms are associated with the suppression of the large number of transposable and retroviral elements in the mammalian genome.\textsuperscript{20-22} Epigenetic aberrations also contribute to pathological conditions,\textsuperscript{23} such as cancer\textsuperscript{24, 25} and disorders of the immune,\textsuperscript{26} endocrine,\textsuperscript{27} and nervous systems.\textsuperscript{28}

Epigenetic alterations can act as an alternative to mutations in the process of tumorigenesis. Now it has been well established that epigenetic alterations are essential not only for the progression but also in the initiation process preceding the first mutations (Figure 4).\textsuperscript{29}
Figure 4. Genetic versus epigenetic models of cancer. (A) The classical monoclonal genetic model of cancer. (B) The epigenetic progenitor model of cancer. Cited from Feinberg et al. 2006.

Epigenetic mechanisms of gene regulation involve semi-reversible covalent chemical modifications of DNA bases and the proteins with which DNA is associated in the nucleus of the cell. These mechanisms include DNA methylation, histone modification, but also noncoding RNAs and chromatin remodeling associated gene silencing. Some of the aberrantly expressed genes involved in epigenetic modifications in tumor cells are listed in Table 1.

Table 1. Some epigenetic modifier genes in human cancer.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Change</th>
<th>Cancer</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGF2</td>
<td>Increased</td>
<td>Colorectal, gastric and breast cancers</td>
<td>31-33</td>
</tr>
<tr>
<td>HDACs</td>
<td>Increased</td>
<td>Several cancers</td>
<td>34-42</td>
</tr>
<tr>
<td>EZH2</td>
<td>Increased</td>
<td>Prostate and breast cancers</td>
<td>43, 44</td>
</tr>
<tr>
<td>KATs</td>
<td>Decreased</td>
<td>Several cancers</td>
<td>37</td>
</tr>
<tr>
<td>SIRT1 and SIRT3</td>
<td>Increased</td>
<td>Prostate and breast cancers</td>
<td>45, 46</td>
</tr>
<tr>
<td>KDM5C</td>
<td>Increased</td>
<td>Breast cancer</td>
<td>47</td>
</tr>
<tr>
<td>SMYD3</td>
<td>Increased</td>
<td>Liver, colon and breast cancers</td>
<td>48</td>
</tr>
<tr>
<td>EHMT1</td>
<td>Decreased</td>
<td>Medulloblastoma</td>
<td>49</td>
</tr>
<tr>
<td>DNMT1 and DNMT3B</td>
<td>Increased</td>
<td>Pancreas, liver, bladder and breast cancers</td>
<td>50-54</td>
</tr>
<tr>
<td>AID</td>
<td>Increased</td>
<td>Leukaemia</td>
<td>55</td>
</tr>
</tbody>
</table>

AID: activation-induced cytidine deaminase; DNMT: DNA methyltransferase; EHMT1: euchromatic histone-lysine N-methyltransferase 1; EZH2: enhancer of zeste homologue 2; KAT: lysine acetyltransferase; HDAC, histone deacetylase; IGF2: insulin-like growth factor 2; KDM5C: lysine-specific demethylase 5C; SIRT: sirtuin; SMYD3: SET and MYND domain-containing 3.
1.2.1 DNA methylation

DNA methylation is the most extensively studied epigenetic mark. In mammals it occurs primarily by the covalent addition of a methyl group on the 5-carbon of cytosine residues in cytosine–guanine (CpG) dinucleotides that may result in the suppression of gene expression (Figure 5).

![Figure 5. Cytosine methylation.](image)

DNA methylation plays a significant role during normal mammalian development by affecting a vast range of biological functions, such as genomic imprinting, cell differentiation, progression through cell division checkpoints, DNA repair, X-chromosome inactivation. It is also associated with the prevention of chromosomal instability by suppression of massive number of transposable elements and inactivation of retroviral elements in the mammalian genome. Aberrant DNA methylation contributes to pathological conditions, such as cancer.
1.2.1.1 CpG rich regions in the genome

Regions rich in of CpG dinucleotides are called CpG islands (CGIs). There is no standard definition of CGI but usually the regions of DNA that contain more than 50% GC content with an observed to expected CpG ratio of greater than 60% (which is significantly higher than the average in non-CGI sequence, due to the general CpG suppression). They are several hundred bases long and there are around 27 thousand CGIs within the human genome. In humans, 50-70% genes are associated with at least one CGI. CGIs are primarily located close to promoter regions or transcription start sites, in enhancer regions, or in repetitive sequences. CpGs are generally free of methylation in normal cells, however some CGIs can be methylated during differentiation and in some tissues which results in the transcriptional silencing of their associated genes. Tissue specific methylated DNA regions that are up to 2kb away from CGIs have also been discovered and are known as CpG shores. The location of different CpG contexts within the human genome and their most common methylation status in normal cell genomes are illustrated in figure 6.

![Figure 6. Distribution of CpG dinucleotides in the human genome. The location of CpG dinucleotides within the genome and their methylation status in normal human cells.](image)

1.2.1.2 DNA methyltransferases

DNA methyltransferases (DNMTs) are the enzymes involved in the methylation process. DNMTs catalyze the methylation of mammalian genomic DNA together with some co-factors, such as the polycomb proteins. S-adenosyl-methionine (SAM) serves as a methyl donor during this process. Five types of DNMTs have been reported in
mammals i.e. DNMT1, DNMT2, DNMT3A, DNMT3B and DNMT3L. A schematic representation of these DNMTs is shown in figure 7.

DNMT1 is the most abundant DNA methyltransferase. It is involved in the maintenance of DNA methylation during DNA replication. It is found abundantly at the replication fork and methylates newly synthesized daughter strand DNA. \(^{75, 76}\) DNMT2 is the smallest DNMT and does not have the regulatory N-terminal region. Instead of DNA, DNMT2 methylates RNA. \(^{77, 78}\) DNMT3A and DNMT3B function as *de novo* methyltransferases. \(^{79, 80}\) DNMT3L lacks a DNA binding domain in its C-terminal domain and thus lacks any DNA methylation capacity but functions as a regulatory factor for DNMT3A and DNMT3B by increasing their potential to bind to the methyl group donor, SAM. \(^{80, 81}\) The precise role of DNMT3L still needs to be determined.

![Figure 7. Schematic representation of human DNMTs.](image)

**Figure 7. Schematic representation of human DNMTs.** The N-terminal contains motifs of interaction with proteins or DNA. The C-terminal contains the catalytic methyltransferase domains. PHD: Plant homology domain. Adopted and modified from Subramaniam et al. 2014. \(^{76}\)

### 1.2.1.3 Demethylation of DNA:

DNA methylation is a reversible process. It is still controversial how active demethylation in specific sites can be achieved. Most likely it is via the enzymatic removal of the methyl group from the fifth carbon of the cytosine residue. \(^{82}\) The ten-
eleven translocation (TET) dioxygenase family of enzymes oxidizes the methyl group on 5-methylcytosine (5mC) and converts it to 5-hydroxymethylcytosine (5hmC). The further series of oxidation reactions of 5hmC by TETs ultimately leads to the base pair excision repair completing the demethylation of 5mC (Figure 8).  

![Diagram of TET dioxygenase reactions](image)

**Figure 8. Modification of 5-methyl cytosine (5mC).** 5mC is oxidized into 5hmC with the catalytic effect of TETs in the presence of alpha-ketoglutarate (α KG), O₂, ATP and Fe²⁺. Similar oxidation reactions then further modify 5hmC into 5fC and 5CaC. Adopted from Liyanage et al. 2014.

1.2.1.4 DNA methylation and gene expression

DNA methylation can result in down regulation of a gene by direct interference with transcription; recruitment of repressive protein complexes; and/or by interacting with histone modifying mechanisms (Figure 9). DNA methylation may also play an indirect role in chromatin remodeling through core histone modification and linker histone occupancy and thereby making transcriptional machinery less accessible to the methylated gene. The interdependent nature of DNA methylation and chromatin compaction has been reported. Methylation of CpGs in gene body does not affect transcription as the RNA polymerases can read through 5mCs.

Some proteins, such as methyl-CpG-binding domain (MBD) have increased binding affinity to methylated DNA. MBD family of proteins act as insulators for transcription factor binding and includes MeCP2, MBD1, MBD2, MBD3 and MBD4. MeCP2 has been reported to bind to methylated DNA and repress transcription of genes such as Sst, Oprk1, and Mef2c by preventing the binding of transcription activator CREB1 to its target site. Transcriptional activation of 85% of such genes has been shown in MeCP2 null mice.

One example of DNA methylation based interference with transcription is the impairment of the binding preference of DNA methylation sensitive transcription factor Gli-type transcription factor YY1 to methylated over unmethylated target sequences. Gli-type transcription factor YY1 controls the paternal expression
of Peg3. In in vitro experiments, the methylation of a key CpG site present in the YY1-target region in Peg3 can prevent the binding of YY1. It has also been reported that repression of transcription is not the only mechanism through with DNA methylation can exert its affect. But more interestingly some of the transcription factors, like zinc finger family transcription factor Kaiso and ZFP57, have been reported to have a possible preference for methylated promoter sites and thus in these cases DNA methylation can also enhance transcription. In short, although DNA methylation is mostly a repressive mark, in some cases it can also positively regulate the transcription.

Figure 9. DNA methylation and gene expression. (A) Transcription factors that are sensitive to methylated DNA (like YY1) would not facilitate the transcription if the template is methylated; (B) MBD proteins (like MeCP2) can bind on methylated DNA and prevent the binding of transcription factors to DNA; (C) methylated DNA can be compacted by the action of chromatin remodeling enzymes and thereby hampering the access of transcription factors to the DNA. Cited from Zhang and Pradhan 2011.

1.2.1.5 DNA methylation and tumorigenesis

Aberrant alterations in DNA methylation (global hypomethylation and CGI hypermethylation) were the first epigenetic events discovered in cancer cells. The role of DNA methylation in cancer is now well established. Aberrant hypermethylation of CGIs in cancer results in the downregulation of TSGs whereas the global genome hypomethylation leads to genomic instability and re-expression of genes that are silenced in the normal cells. Kalari and Pfeifer coined the concept
of “driver” and “passenger” DNA methylation modifications in cancer. DNA
hypermethylation of TSG promoters was referred to as a driver event because of their
obvious carcinogenic effects. Whereas DNA hypermethylation without any effect on
cell transformation was designated a passenger hypermethylation events. These
aberrant changes in DNA methylation during carcinogenesis have been attributed to
abnormality and/or imbalance of DNA demethylating TETs and DNA methylating
DNMTs. The aberrant hypermethylation of promoter CpG-rich regions of TSGs
in a variety of malignancies is attributed to DNMT3A, DNMT3B and DNMT3L. DNMT1, 3A, and 3B are overexpressed in several human cancers, such as colon,
prostate, breast, liver and in leukemia.

1.2.2 Histone modification

DNA is wrapped around a core of histone proteins. A series of epigenetic
modifications of histones often can change the structure of chromatin and thus
interfere with the access of various cellular factors to DNA (Figure 10). So far
more than 100 distinct histone modifications have been identified. The most
important histone modifications are acetylation and methylation of the histones. Other
histone modification mechanisms include serine/threonine phosphorylation, ADP
ribosylation, glycosylation, ubiquitination, sumoylation etc. The mechanisms and
the effects of certain histone modifications still need to be established.

1.2.2.1 Histone acetylation

The presence of acetyl groups in histones was first reported in 1963. Hyperacetylated histones were then reported to be associated with highly transcribed
genes that suggested that histone acetylation could be involved in facilitating gene
transcription. Acetylation occurs on lysine resides on N-terminal tail of the core
histone proteins and is catalyzed by lysine acetyltransferases (KATs), whereas
histone deacetylases (HDACs) remove the acetyl group from the histone lysine.
Acetylation and deacetylation of histone lysines influence chromatin remodeling and
in this way regulate the access of transcription machinery to the DNA. Histone
acetylation has also been suggested to play some role in DNA replication and
repair in case of double-strand breaks. Aberrant histone acetylation or deacetylation may lead to up- or down-regulation of
gene expression in cancer. The overexpression of HDACs has been reported in various cancers.\textsuperscript{39, 126, 127} The KATs have also been reported to play an important role in tumor development e.g. in leukemia.\textsuperscript{128, 129}

![Diagram of Chromatin Structure](image)

**Figure 10. Chromatin structure of active and inactive promoter.** (A) A transcriptionally active chromatin is structurally open and unmethylated at its cytosines, trimethylated at H3K4 and hyperacetylated at histones H3 and H4 at multiple sites. (B) Methylated cytosines bind MBDs that attract HDACs that in turn remove acetyl groups (Ac) from H3 and H4. Trimethylation of H3 at lysine 9 (H3K9me3) results in closed chromatin that is a repressive state. Methylated cytosines serve as docking sites for MBD proteins such as MeCP2.\textsuperscript{25, 130}

1.2.2.2 Histone methylation

Histone methylation may either activate or deactivate transcription. This depends upon the amino acid methylated. Unlike acetylation, methylation of a lysine does not neutralize its positive charge that indicates an indirect effect of lysine methylation on nucleosome dynamics and transcription.\textsuperscript{131, 132} Histone methylation can also occur on arginines,\textsuperscript{133} but the effects of this modification is still unclear. Methylated lysine 4 and 36 (H3K4me and H3K36me) are considered to be activation marks whereas methylated lysine 9 and 27 (H3K9 and H3K27) are thought to be repression marks of transcription (Figure 10). The active histone methylation marks (H3K4me3 and H3K36me3) are considered to have an indirect regulatory function. It has been suggested that H3K4me3 may exert its activation effect by inhibition of H3K27me3 methylation.\textsuperscript{134, 135} The other active histone mark i.e. H3K36me3 has been reported to
regulate the histone deacetylation by the catalytic activity of Rpd3S lysine deacetylase complex as it has been reported that a recruited Rpd3S remains inactive in the absence of H3K36me3.\textsuperscript{136} Rpd3S complex prevents histone turnover by deacetylating histones.\textsuperscript{137} This is why H3K36me3 is considered to be involved in nucleosome stabilization. Another mechanism through which H3K36me3 stabilizes nucleosomes is chromatin remodeling, as it enhances the affinity of the repressive Isw1b chromatin remodeling complex.\textsuperscript{138, 139} The inactive histone marks (methylated H3K9 and H3K27) are associated with a compact chromatin structure known as heterochromatin formation and the silencing of polycombs respectively. H3K9me and H3K27me are also suggested to be involved in nucleosome stabilization. Methylation of H3K9 and H3K27 increases the affinity of certain protein modules for histone residues. H3K9me and H3K27me bind chromodomains of HP1 and polycomb respectively. A methylated lysine residue can be associated with one, two or three methyl groups. The increase in the number of methyl groups on these residues affects nucleosome stabilization.\textsuperscript{140, 141}

Several factors are supposed to be involved in the number of methyl groups attached to a specific lysine residue. Differences in the degree of accessibility and exposure time of a histone methyltransferase enzyme to a given lysine residue affects its methylated state. For example H2BK123ub1 promotes H3K4me3 and H3K79me3,\textsuperscript{142} because histone ubiquitynation modifies the chromatin conformation and increases the access of H3K4 to its modifying enzyme. Similarly, histone methyltransferases can modify a given lysine in distinct state of methylation depending upon the stability of the nucleosome. The more stable the nucleosome, the more methyl groups can be attached to a specific lysine by modifying enzymes.\textsuperscript{143, 144}

\textbf{1.2.2.2.1 Histone methylation and cancer}

Widespread changes in histone methylation patterns have been observed in cancer. Any aberrant change in the active histone methylation can result in the activation of genes that could lead to tumor development e.g. it has been reported that chromosomal translocations of H3K4 methylating enzyme MLL plays an important role in the development of leukemia.\textsuperscript{145} Altered methylation patterns of H3K9 and H3K27 result in aberrant gene silencing in various cancers.\textsuperscript{146, 147} Deregulation of histone methyltransferases (HMTs) leads to aberrant silencing of TSGs. For example, overexpression of H3K27 methylating EZH2 has been reported in breast and prostate cancer.\textsuperscript{147} The H3K9 methylating enzyme G9a is overexpressed in liver cancer.\textsuperscript{126, 148}
1.2.3 Non-coding RNAs

Non-coding RNAs (ncRNAs) are transcribed from the DNA but are not translated into proteins. These are functionally active in another way than transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) and can repress or activate gene expression. ncRNAs are involved in the modulation of various biological processes such as cell division, differentiation and apoptosis and are also involved in the development of various diseases.\(^{100, 149}\) Non-coding RNAs can be classified on the basis of their size into long ncRNAs (lncRNAs) and small ncRNAs (sncRNAs).

1.2.3.1 Long non-coding RNAs

Long non-coding RNAs are longer than 200 nucleotides\(^ {149}\) and are without any open reading frame (ORF).\(^ {150}\) lcnRNAs are involved in various biological processes such as cell differentiation, organ development, and disease progression.\(^ {151}\) They can regulate the gene expression at the epigenetic, transcriptional and post-transcriptional levels.\(^ {152}\) They can interfere in the gene transcription by direct interference with the transcription process, in chromatin remodeling by recruiting chromatin modifying enzymes, or by forming a triplex structure and silencing the targeted rRNA by promoter methylation recruiting the DNA methylating enzyme DNMT3b.\(^ {153}\)

Aberrant expression of lncRNAs is reported in various cancers with both oncogenic and tumor suppressive functions.\(^ {154}\) HOX antisense intergenic RNA (HOTAIR) was the first lncRNA identified to be involved in tumorigenesis. HOTAIR is deregulated in a number of cancers.\(^ {155-158}\) Its expression in breast cancer was reported to be upregulated which leads to enhanced repressive activity of polycomb repressor complex 2 (PRC2).\(^ {155}\) Activated PRC2 causes modifications in histone structure and down regulates the expression of target genes.\(^ {159}\) MALAT1 is another such oncogenic lncRNA that is upregulated in several cancers.\(^ {160-164}\)

1.2.3.2 Small non-coding RNAs

Small non-coding RNAs (sncRNAs) may range between \(~18-27\) nucleotides. Micro RNAs (miRNAs) are an abundant class of sncRNAs that are also the most well studied. In the updated miRBase database, 35,828 mature miRNAs were recently registered across all the species (http://mirbase.org, Release 21, accessed June 2014).\(^ {165}\) miRNAs are \(~22\) nucleotides long and are transcribed by RNA polymerase II
pri-miRNA which are then processed to form the RNA-induced silencing complex (RISC). In the RISC complex, miRNAs act as a probe and cleaves the mRNAs that are partially complementary to the miRNA. This results in transcriptional silencing of the targeted gene.\textsuperscript{166, 167} It is estimated that miRNAs target \textasciitilde 50\% of human genes,\textsuperscript{168} and are involved in almost every biological process. These are also associated to the pathogenesis of several diseases.\textsuperscript{168-170}

miRNAs were first reported to be associated with tumorigenesis in 2002 where miRNA-15 and miRNA-16 were reported to be deleted or down-regulated in 68\% B-cell chronic lymphocytic leukemia cases.\textsuperscript{171} miRNAs can be tumor suppressors and oncogenes. Aberrant expression of miRNAs has been associated with the development of a number of human cancers.\textsuperscript{172-174} For example, a cluster made by miRNA-15a, miRNA-16-1 and DLEU2 regulates B cell proliferation by targeting multiple oncogenes and their deletion results in the development of chronic lymphocytic leukemia.\textsuperscript{175, 176} miRNA-34 in coordination with p53 suppresses prostate cancer.\textsuperscript{177}

1.2.4 Chromatin remodeling

Epigenetic modifications play an important role in biological processes by affecting the chromatin structure and its organization.\textsuperscript{114} Approximately two meters long DNA present in each human cell is compacted 10,000- to 20,000-fold as a nucleosomal fiber (chromatin) and in higher-order chromatin structures in the cell nucleus that is only \textasciitilde 6 \textmu m in diameter. This is comparable to put 40 km long thin thread into a tennis ball.\textsuperscript{178, 179} The packaging of genome in its three-dimensional spatial structure also affects gene expression. Through this mechanism, an enhancer element that could be located at a long distance (cis interactions) or even on different chromosomes (trans interactions) could affect gene expression.

A key factor that organizes the genome in its 3D spatial structure is CCCTC-binding zinc finger factor CTCF (Figure 11).\textsuperscript{88} The CTCF mediates both cis and trans interactions by forming multiple distinct sections of chromosome, i.e. transcriptional harbors.\textsuperscript{88} CTCF can modulate the chromatin structure through different mechanisms.\textsuperscript{180} For example, CTCF associated transcriptional harbors are enriched in active histone methylation signatures (H3K4 and H3K36) which are otherwise enriched with inactive histone methylation marks (H3K9, H3K20, and H3K27).\textsuperscript{88} It has been
reported that chromosomes are dynamically packaged at larger than a mega base scale but still the actively transcribed genes are made accessible to the transcriptional machinery by the establishment of chromosomal boundaries.\textsuperscript{181} Chromatin remodeling is important in the regulation of gene expression. It is also associated with cancer development.\textsuperscript{182}

![Figure 11. Three-dimensional chromatin interaction.](image)

**1.2.5 Cross-talk between epigenetic mechanisms and tumorigenesis**

Different epigenetic mechanisms extensively cross-talk to each other to maintain and regulate the complex transcriptional network in cells. For example, the epigenetic processes of DNA methylation/demethylation and chromatin modifications are reported to interact with each other.\textsuperscript{183} The go-betweens during such interactions are proteins that bind to the proteins involved in these epigenetic modifications. For instance, histone H3 trimethylated at lysine 4 (H3K4me3) binds to recruit tumor suppressor inhibitor of growth protein 1b (Ing1). Ing1 interacts with growth arrest and DNA damage protein 45a (GADD45a).\textsuperscript{184} The function of GADD45 in turn is to facilitate DNA demethylation of specific genes by recruiting DNA repair enzymes.\textsuperscript{185} The H3K4me3 also recruits KATs that results in an increased level of histone acetylation and reduced level of 5mC in the nucleosome.\textsuperscript{186, 187} Another example of the cross-talk between different epigenetic mechanisms could be the frequent silencing of miRNAs expression by promoter hypermethylation in various cancers.\textsuperscript{188}
For example miR-34b/c has been reported to be hypermethylated in lung cancer\textsuperscript{190},
and chronic lymphocytic leukemia.\textsuperscript{192}

Figure 12. A model for the cross-talk between different epigenetic mechanisms in normal versus cancer cells. In normal cells, low-level of double-stranded RNAs are transcribed from hypermethylated repetitive DNA. This results in recruitment of the repressive epigenetic machinery (RNAi, DNMTs, MBDs). CpG islands (CGIs), in normal cells, are hypomethylated and the associated potential TSGs are transcribed. Binding of CTCF inhibits the methylation of these CGIs. In cancer cells, genome-wide hypomethylation results in increased genomic instability as it disrupts normal transcription of TSGs by generation of ncRNA, recruitment of the RNAi machinery, transcriptional interference and aberrant re-expression of BORIS that displaces CTCF from CGI which results in its hypermethylation by DNMTs.

In cancer aberrant changes can occur in all the epigenetic mechanism of gene regulation. For example, the genome-wide hypomethylation that occurs particularly at repetitive DNA sequences could result in the production of antisense transcripts and non-coding RNAs (ncRNAs) and the recruitment of the RNA interference (RNAi) machinery (Figure 12). This results in the establishment of repressive epigenetic marks, such as DNA methylation, histone modifications and chromatin remodeling at these sites. All these epigenetic changes interact with each other and result in the cancer initiation and progression.
1.3 Introduction to nasopharyngeal carcinoma

1.3.1 Anatomy and histopathology

Nasopharyngeal carcinoma (NPC) is a tumor that arises from mucosal epithelial cells covering the nasopharyngeal surface, mostly within the lateral nasopharyngeal recess, a space behind the nose.\(^{193}\) It is significantly different from other cancers of head and neck area in its epidemiologic features, etiology, pathogenesis, histological types, treatment and prognosis.\(^ {194}\) The most recent histological classification of NPC by World Health Organization (WHO) was done in 2005. According to this classification, NPC is divided into three major types i.e. squamous cell carcinoma (SCC), non-keratinizing carcinoma (differentiated or undifferentiated) and basaloid squamous cell carcinoma. Squamous cell carcinoma is highly differentiated and retains its epithelial growth pattern and keratin filaments. Non-keratinizing differentiated carcinoma lacks keratin filaments but retains its epithelial shape and growth pattern and in non-keratinizing undifferentiated carcinoma, the cells do not produce keratin filaments and also lack their distinctive epithelial growth pattern. Basaloid squamous cell carcinoma type of NPC contains a basaloid neoplasm in the regions of squamous differentiation. Only a few cases of basaloid squamous cell carcinoma have been reported which makes it a clinically less important.\(^ {195}\)

1.3.2 Epidemiology and etiology

NPC has a notable geographical and ethnic distribution.\(^ {196}\) It is an uncommon malignancy in most countries of the world. In non-endemic areas, such as United States and Europe, its incidence and age-adjusted incidence for both male and female is less than one per 100,000 persons. Intermediate incidences ranging between 5-7/100,000 is reported in several Southeast Asian and North African countries such as Indonesia, Malaysia, Thailand, Vietnam Kenya, Greenland, Morocco and Sudan. In endemic regions (Southern China) the annual average incidence rate is around 20/100,000.\(^ {197-199}\) Its incidence rate has also been reported to be exceeding 30/100,000 in Sarawak, Malaysia.\(^ {200}\) Differences in the incidence of NPC are also reported in different ethnic groups living in the same geographical location.\(^ {199, 201}\) NPC is more prevalent in males as compared to females.\(^ {2}\) Its incidence also increases with the increase in age. However a bimodal distribution of NPC is also observed in
Mediterranean populations and such juvenile cases of NPC accounts for 20% of the all NPC cases in this population.

The incidence of NPC types is not uniform in different geographical regions. Squamous cell carcinoma is prevalent in low-incidence regions such as Europe, where non-keratinizing carcinoma is reported to be on a rise. Non-keratinizing undifferentiated carcinoma is the most common type of NPC in regions of high-incidence regions such as Southern China and accounts for more than 97% of all NPC. One of the unique features of non-keratinizing undifferentiated carcinoma is its close to 100% association with Epstein-Barr virus (EBV). Familial NPC has been linked with susceptibility loci on chromosome 3p 21, 4p15.1-q12 and 5p 13. Consumption of certain food items such as salted fish and preserved food containing volatile nitrosamines, especially during childhood, are also reported to play some role in the etiology of NPC. These findings suggest the role of genetic, ethnic and environmental factors in the etiology of NPC.

1.3.3 Clinical presentation and diagnosis

The early diagnosis of NPC is challenging as compared to most other cancers because of its non-specific early signs and hidden anatomical site. Early symptoms may not be frequently recognizable by the physicians. Patients often show a spectrum of symptoms. A majority of the patients present with cervical lymphadenopathy. Other symptoms may include neck mass, nasal and aural dysfunction, headache, diplopia, facial numbness, weight loss and trismus. Clinical examination includes endoscopy, computed tomography (CT) scan and magnetic resonance imaging (MRI). Histological examination of nasopharyngeal biopsy is the classical standard for NPC diagnosis. However, the invasiveness of this method makes it less suitable as an early screening method in NPC high-risk population. Measurement of EBV viral load in serum or plasma has been shown to be useful for diagnosis and monitoring in patients affected by NPC. For this purpose, the analysis of EBV-specific antibodies such as serum titers of IgA antibodies to viral capsid antigen (VCA), early antigen (EA), nuclear antigen (EBNA) and EBV-specific DNAs are being used for screening and monitoring the disease in NPC high-risk population. Nevertheless, the sensitivity (Sn) and specificity (Sp) of such assays are not satisfactory.
1.3.4 Treatment and prognosis

NPC is a radiosensitive tumor and radiotherapy is the standard treatment for almost all NPC patients without distant metastasis\textsuperscript{223, 224}. Although more than 70% overall survival rates can be expected for stages I and II, treatment outcomes are unsatisfactory for advanced-stage NPC when radiotherapy is offered alone\textsuperscript{223-226}. The vast majority of NPC patients (75-90\%) are diagnosed with a loco-regionally advanced disease, commonly with cervical nodal metastasis\textsuperscript{196}. Because of chemosensitive nature of NPC, concurrent chemo-radiotherapy is the standard treatment for these patients\textsuperscript{227, 228}. The presence of EBV only in tumor cells provides the possibility of targeting EBV-specific therapy for NPC. Efforts are being made to develop EBV-targeting immunotherapeutic or gene therapeutic strategies\textsuperscript{229-233}. In spite of all these efforts no significant improvements could be made in the early diagnosis of NPC. Although NPC is sensitive to chemo-radiotherapy the survival rates remained disappointing\textsuperscript{234}. A majority of the NPC patients are still diagnosed late at stage III and IV with a loco-regionally advanced disease when the five-year survival rate is only 40-60\%\textsuperscript{235-238}. The five-year survival rates for NPC patients could be more than 90\% if the disease could be diagnosed at stage I\textsuperscript{215}. Metastasis and loco-regional recurrence are the two major reasons of failure in NPC treatment\textsuperscript{239, 240}.

1.3.5 Epigenetic aspect of nasopharyngeal carcinoma

A large number of identified genes are silenced in NPC by epigenetic mechanisms, primarily by aberrant promoter hypermethylation\textsuperscript{241-244}. Several key TSGs, such as p53, mutated in more than 50\% of all tumors, and retinoblastoma (Rb) are usually wild-type in NPC\textsuperscript{241, 245-247}. Deregulated expression of p53 in NPC has been linked with methylation dependent silencing of a TSG named UCHL1\textsuperscript{248}. Hypermethylation of several other important TSGs such as P16, RASSF1A\textsuperscript{249, 250}, RARβ2, DAPK, DLEC1 have also been reported in NPC\textsuperscript{243, 249-253}. Hypermethylation of several TSGs involved in fundamental cellular pathways such as apoptosis, DNA repair, tumor invasion and metastasis has been reported in NPC\textsuperscript{254}. Aberrantly methylated genes reported in NPC are summarized in Table 2.
### Table 2. List of aberrantly methylated genes in NPC.

<table>
<thead>
<tr>
<th>Cancer-related process</th>
<th>Gene ID</th>
<th>Chromosomal location</th>
<th>Function</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell cycle</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDKN2B</td>
<td>9p21</td>
<td></td>
<td>Cyclin-dependent kinase inhibitor for CDK4 and CDK6, a cell growth regulator of cell cycle G1 progression</td>
<td>243, 255-257</td>
</tr>
<tr>
<td>p16</td>
<td>9p21</td>
<td></td>
<td>Cell cycle regulation</td>
<td>243, 255-258</td>
</tr>
<tr>
<td>CHFR</td>
<td>12q24.33</td>
<td></td>
<td>Mitotic checkpoint regulator early in G2-M transition</td>
<td>243, 259</td>
</tr>
<tr>
<td>BRD7</td>
<td>16q12</td>
<td></td>
<td>Transcriptional regulation, inhibits G1-S transition</td>
<td>260</td>
</tr>
<tr>
<td>FHT</td>
<td>3P14.2</td>
<td></td>
<td>Cell-cycle regulation, G1-S phase checkpoint, DNA-damage response, nucleotide and nucleic acid metabolism</td>
<td>261</td>
</tr>
<tr>
<td>GADD45G</td>
<td>9q22</td>
<td></td>
<td>Inhibits G1-S and G2-M transition, apoptosis</td>
<td>262</td>
</tr>
<tr>
<td>DLEC1</td>
<td>3p22-21.3</td>
<td></td>
<td>G1 cell cycle arrest</td>
<td>263</td>
</tr>
<tr>
<td>ZMYND10</td>
<td>3P21.3</td>
<td></td>
<td>Cell cycle</td>
<td>264</td>
</tr>
<tr>
<td>MIPOL1</td>
<td>14q13.1</td>
<td></td>
<td>Negative regulator of G1 progression</td>
<td>265</td>
</tr>
<tr>
<td>PRDM2</td>
<td>1p36.21</td>
<td></td>
<td>G2-M cell cycle arrest</td>
<td>266</td>
</tr>
<tr>
<td>LTF</td>
<td>3p21.3</td>
<td></td>
<td>Cell cycle regulation</td>
<td>267</td>
</tr>
<tr>
<td>CCNA1</td>
<td>13q12.3-q13</td>
<td></td>
<td>An important regulator of the cell cycle required for S phase and passage through G2</td>
<td>268</td>
</tr>
<tr>
<td>PTPRG</td>
<td>3p14-21</td>
<td></td>
<td>Cell cycle regulator via inhibition of pRB phosphorylation through down-regulation of cyclin D1</td>
<td>269</td>
</tr>
<tr>
<td>TP73</td>
<td>1p36.3</td>
<td></td>
<td>Cell cycle, DNA damage response, apoptosis, transcription factor</td>
<td>270</td>
</tr>
<tr>
<td><strong>Apoptosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAPK</td>
<td>9p34.1</td>
<td></td>
<td>Positive mediator of gamma-interferon induced apoptosis</td>
<td>271</td>
</tr>
<tr>
<td>CASP8</td>
<td>2q33-q34</td>
<td></td>
<td>Apoptosis</td>
<td>272</td>
</tr>
<tr>
<td>GSTP1</td>
<td>11q13</td>
<td></td>
<td>Apoptosis, metabolism, energy pathways</td>
<td>273</td>
</tr>
<tr>
<td>CMTM3</td>
<td>16q21</td>
<td></td>
<td>Induces apoptosis with caspase-3 activation</td>
<td>274</td>
</tr>
<tr>
<td>CMTM5</td>
<td>14q11.2</td>
<td></td>
<td>Induces apoptosis with activation of caspase 3, 8 and 9, synergistic effects with TNF-α</td>
<td>275</td>
</tr>
<tr>
<td>ZNF382</td>
<td>19q13.12</td>
<td></td>
<td>Key regulator of cell proliferation, differentiation, and apoptosis, repress NF-kB and AP-1 signaling</td>
<td>276</td>
</tr>
<tr>
<td>TNFRSF11B</td>
<td>8q24</td>
<td></td>
<td>Induces apoptosis, inhibits tumor growth specifically in bones</td>
<td>277</td>
</tr>
<tr>
<td>PLA2G16</td>
<td>11q12.3</td>
<td></td>
<td>Proapoptotic function through the inhibition of PP2A</td>
<td>278</td>
</tr>
<tr>
<td><strong>Invasion and metastasis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDH1</td>
<td>16q22.1</td>
<td></td>
<td>Calcium-dependent adhesion and cell migration</td>
<td>279</td>
</tr>
<tr>
<td>CDH13</td>
<td>16q23.3</td>
<td></td>
<td>Calcium-dependent adhesion and cell migration</td>
<td>280</td>
</tr>
<tr>
<td>PCDH10</td>
<td>4q28.3</td>
<td></td>
<td>Calcium-dependent adhesion and cell migration</td>
<td>281</td>
</tr>
<tr>
<td>CDH4</td>
<td>20q13.3</td>
<td></td>
<td>Calcium-dependent adhesion and cell migration</td>
<td>282</td>
</tr>
<tr>
<td>OPCML</td>
<td>11q25</td>
<td></td>
<td>Cell adhesion, cell-cell recognition</td>
<td>283</td>
</tr>
<tr>
<td>TFP12</td>
<td>7q22</td>
<td></td>
<td>Serine protease inhibitor</td>
<td>284</td>
</tr>
<tr>
<td>MMP19</td>
<td>12q14</td>
<td></td>
<td>Extra cellular matrix</td>
<td>285</td>
</tr>
<tr>
<td>THBS1</td>
<td>15q15</td>
<td></td>
<td>An adhesive glycoprotein, mediates cell-to-cell and cell-to-matrix interactions</td>
<td>286</td>
</tr>
<tr>
<td>Cx43</td>
<td>20q11</td>
<td></td>
<td>Gap junction and intercellular communication</td>
<td>287</td>
</tr>
<tr>
<td>Cancer-related process</td>
<td>Gene ID</td>
<td>Chromosomal location</td>
<td>Function</td>
<td>Refs</td>
</tr>
<tr>
<td>------------------------</td>
<td>---------</td>
<td>----------------------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>Cancer-related process</td>
<td>CADM1</td>
<td>11q23</td>
<td>Cell adhesion molecule, mediate cell-cell interaction</td>
<td>280, 281</td>
</tr>
<tr>
<td>DNA repair</td>
<td>MGMT</td>
<td>10q26</td>
<td>Repair alkylated guanine</td>
<td>251</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>ARF/P14</td>
<td>9p21</td>
<td>Stabilizes p53, interacts with MDM2</td>
<td>243, 255</td>
</tr>
<tr>
<td>Angiogenesis</td>
<td>EDNRB</td>
<td>13q22</td>
<td>Negative regulator of ET/ETAR pathway</td>
<td>297, 298</td>
</tr>
<tr>
<td>Tissue development and differentiation</td>
<td>Myocd</td>
<td>17p11.2</td>
<td>Transcription factor, involved in smooth muscle cell differentiation</td>
<td>305</td>
</tr>
<tr>
<td>Others</td>
<td>SCGB3A1</td>
<td>5q35</td>
<td>Involved in epithelial cell differentiation, cell-cycle reentry regulator, suppresses tumor cell migration and invasion, induces apoptosis</td>
<td>306</td>
</tr>
<tr>
<td>Others</td>
<td>LOX</td>
<td>5q23.2</td>
<td>Inhibits clonogenicity and cell growth</td>
<td>290</td>
</tr>
<tr>
<td>Others</td>
<td>ADAMTS18</td>
<td>16q23.1</td>
<td>Cell adhesion modulator, inhibits growth factor independent cell proliferation</td>
<td>282, 283</td>
</tr>
<tr>
<td>Others</td>
<td>THY1</td>
<td>11q23.3</td>
<td>Regulates cytoskeletal organization, focal adhesion and migration by modulating the activity of p190 RhoGAP and Rho GTPase</td>
<td>284</td>
</tr>
<tr>
<td>Others</td>
<td>PCDH8</td>
<td>13q21.1</td>
<td>Cell colony formation and cell migration</td>
<td>285</td>
</tr>
<tr>
<td>Others</td>
<td>THY1</td>
<td>11q23.3</td>
<td>Regulations cytoskeletal organization, focal adhesion and migration by modulating the activity of p190 RhoGAP and Rho GTPase</td>
<td>284</td>
</tr>
<tr>
<td>Others</td>
<td>PCDF8</td>
<td>13q21.1</td>
<td>Cell colony formation and cell migration</td>
<td>285</td>
</tr>
<tr>
<td>Others</td>
<td>PCDH8</td>
<td>13q21.1</td>
<td>Cell colony formation and cell migration</td>
<td>285</td>
</tr>
<tr>
<td>Others</td>
<td>PCDF8</td>
<td>13q21.1</td>
<td>Cell colony formation and cell migration</td>
<td>285</td>
</tr>
<tr>
<td>Others</td>
<td>LOX</td>
<td>5q23.2</td>
<td>Inhibits clonogenicity and cell growth</td>
<td>290</td>
</tr>
<tr>
<td>Others</td>
<td>ADAMTS18</td>
<td>16q23.1</td>
<td>Cell adhesion modulator, inhibits growth factor independent cell proliferation</td>
<td>282, 283</td>
</tr>
<tr>
<td>Others</td>
<td>THY1</td>
<td>11q23.3</td>
<td>Regulates cytoskeletal organization, focal adhesion and migration by modulating the activity of p190 RhoGAP and Rho GTPase</td>
<td>284</td>
</tr>
<tr>
<td>Others</td>
<td>PCDF8</td>
<td>13q21.1</td>
<td>Cell colony formation and cell migration</td>
<td>285</td>
</tr>
<tr>
<td>Others</td>
<td>LOX</td>
<td>5q23.2</td>
<td>Inhibits clonogenicity and cell growth</td>
<td>290</td>
</tr>
<tr>
<td>Others</td>
<td>ADAMTS18</td>
<td>16q23.1</td>
<td>Cell adhesion modulator, inhibits growth factor independent cell proliferation</td>
<td>282, 283</td>
</tr>
<tr>
<td>Others</td>
<td>THY1</td>
<td>11q23.3</td>
<td>Regulates cytoskeletal organization, focal adhesion and migration by modulating the activity of p190 RhoGAP and Rho GTPase</td>
<td>284</td>
</tr>
<tr>
<td>Others</td>
<td>PCDF8</td>
<td>13q21.1</td>
<td>Cell colony formation and cell migration</td>
<td>285</td>
</tr>
<tr>
<td>Others</td>
<td>LOX</td>
<td>5q23.2</td>
<td>Inhibits clonogenicity and cell growth</td>
<td>290</td>
</tr>
<tr>
<td>Others</td>
<td>ADAMTS18</td>
<td>16q23.1</td>
<td>Cell adhesion modulator, inhibits growth factor independent cell proliferation</td>
<td>282, 283</td>
</tr>
<tr>
<td>Others</td>
<td>THY1</td>
<td>11q23.3</td>
<td>Regulates cytoskeletal organization, focal adhesion and migration by modulating the activity of p190 RhoGAP and Rho GTPase</td>
<td>284</td>
</tr>
<tr>
<td>Others</td>
<td>PCDF8</td>
<td>13q21.1</td>
<td>Cell colony formation and cell migration</td>
<td>285</td>
</tr>
<tr>
<td>Others</td>
<td>LOX</td>
<td>5q23.2</td>
<td>Inhibits clonogenicity and cell growth</td>
<td>290</td>
</tr>
<tr>
<td>Others</td>
<td>ADAMTS18</td>
<td>16q23.1</td>
<td>Cell adhesion modulator, inhibits growth factor independent cell proliferation</td>
<td>282, 283</td>
</tr>
<tr>
<td>Others</td>
<td>THY1</td>
<td>11q23.3</td>
<td>Regulates cytoskeletal organization, focal adhesion and migration by modulating the activity of p190 RhoGAP and Rho GTPase</td>
<td>284</td>
</tr>
<tr>
<td>Others</td>
<td>PCDF8</td>
<td>13q21.1</td>
<td>Cell colony formation and cell migration</td>
<td>285</td>
</tr>
<tr>
<td>Others</td>
<td>LOX</td>
<td>5q23.2</td>
<td>Inhibits clonogenicity and cell growth</td>
<td>290</td>
</tr>
<tr>
<td>Others</td>
<td>ADAMTS18</td>
<td>16q23.1</td>
<td>Cell adhesion modulator, inhibits growth factor independent cell proliferation</td>
<td>282, 283</td>
</tr>
<tr>
<td>Others</td>
<td>THY1</td>
<td>11q23.3</td>
<td>Regulates cytoskeletal organization, focal adhesion and migration by modulating the activity of p190 RhoGAP and Rho GTPase</td>
<td>284</td>
</tr>
<tr>
<td>Others</td>
<td>PCDF8</td>
<td>13q21.1</td>
<td>Cell colony formation and cell migration</td>
<td>285</td>
</tr>
<tr>
<td>Others</td>
<td>LOX</td>
<td>5q23.2</td>
<td>Inhibits clonogenicity and cell growth</td>
<td>290</td>
</tr>
<tr>
<td>Others</td>
<td>ADAMTS18</td>
<td>16q23.1</td>
<td>Cell adhesion modulator, inhibits growth factor independent cell proliferation</td>
<td>282, 283</td>
</tr>
<tr>
<td>Others</td>
<td>THY1</td>
<td>11q23.3</td>
<td>Regulates cytoskeletal organization, focal adhesion and migration by modulating the activity of p190 RhoGAP and Rho GTPase</td>
<td>284</td>
</tr>
<tr>
<td>Others</td>
<td>PCDF8</td>
<td>13q21.1</td>
<td>Cell colony formation and cell migration</td>
<td>285</td>
</tr>
<tr>
<td>Others</td>
<td>LOX</td>
<td>5q23.2</td>
<td>Inhibits clonogenicity and cell growth</td>
<td>290</td>
</tr>
<tr>
<td>Others</td>
<td>ADAMTS18</td>
<td>16q23.1</td>
<td>Cell adhesion modulator, inhibits growth factor independent cell proliferation</td>
<td>282, 283</td>
</tr>
<tr>
<td>Others</td>
<td>THY1</td>
<td>11q23.3</td>
<td>Regulates cytoskeletal organization, focal adhesion and migration by modulating the activity of p190 RhoGAP and Rho GTPase</td>
<td>284</td>
</tr>
<tr>
<td>Others</td>
<td>PCDF8</td>
<td>13q21.1</td>
<td>Cell colony formation and cell migration</td>
<td>285</td>
</tr>
<tr>
<td>Others</td>
<td>LOX</td>
<td>5q23.2</td>
<td>Inhibits clonogenicity and cell growth</td>
<td>290</td>
</tr>
<tr>
<td>Others</td>
<td>ADAMTS18</td>
<td>16q23.1</td>
<td>Cell adhesion modulator, inhibits growth factor independent cell proliferation</td>
<td>282, 283</td>
</tr>
<tr>
<td>Others</td>
<td>THY1</td>
<td>11q23.3</td>
<td>Regulates cytoskeletal organization, focal adhesion and migration by modulating the activity of p190 RhoGAP and Rho GTPase</td>
<td>284</td>
</tr>
<tr>
<td>Others</td>
<td>PCDF8</td>
<td>13q21.1</td>
<td>Cell colony formation and cell migration</td>
<td>285</td>
</tr>
<tr>
<td>Others</td>
<td>LOX</td>
<td>5q23.2</td>
<td>Inhibits clonogenicity and cell growth</td>
<td>290</td>
</tr>
<tr>
<td>Cancer-related process</td>
<td>Gene ID</td>
<td>Chromosomal location</td>
<td>Function</td>
<td>Refs</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------</td>
<td>----------------------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>NR4A3</td>
<td>1p34.3</td>
<td>Interaction partner of the mitochondrial ATP synthase subunit OSCP/ATP5O protein, a stress-responsive gene</td>
<td>307</td>
</tr>
<tr>
<td></td>
<td>LARS2</td>
<td>3p21.3</td>
<td>Essential roles in group I intron RNA splicing and protein synthesis within the mitochondria, indirectly required for mitochondrial genome maintenance</td>
<td>308</td>
</tr>
<tr>
<td></td>
<td>CRYAB</td>
<td>11q23.1</td>
<td>An important nuclear role in maintaining genomic integrity</td>
<td>309</td>
</tr>
<tr>
<td></td>
<td>CYB5R2</td>
<td>11p15.4</td>
<td>Carcinogen metabolizing enzyme</td>
<td>310, 311</td>
</tr>
</tbody>
</table>

CDKN2B: Cyclin-dependent kinase inhibitor 2B; ADAMTS18: ADAM metallopeptidase with thrombospondin type 1 motif, 18; ADAMTS9: A disintegrin-like and metallopeptidase with thrombospondin type 1 motif 9; ARF: Alternate open reading frame; BRD7: Bromodomain containing 7; CADM1: Cell adhesion molecule 1; CASP8: Caspase 8; CCNA1: Cyclin A1; CDH1: Cadherin 1; CDH13: Cadherin 13; CDH4: Cadherin 4; CHFR: Checkpoint with forkhead and ring finger domains; CMTM3: CKLF like MARVEL transmembrane domain-containing member 3; CMTM5: CKLF like MARVEL transmembrane domain-containing member 5; CRYAB: Crystallin, alpha B; Cx43: Connexin 43; CYB5R2: Cytochrome b5 reductase 2; FBLN2: Fibulin 2; FHIT: Fragile histidine triad gene; GADD45G: Growth arrest and DNA-damage inducible, gamma; GSTP1: Glutathione S-transferase pi 1; LOX: Lysyl oxidase; LTF: Lactoferrin; MIPOL1: Mirror-image polydactyly 1; MLH1: MutL homolog 1; MMP19: Matrix metalloproteinase-19; Myocd: Myocardin; NR4A3: Nuclear receptor subfamily 4, group A, member 3; OPCML: Opioid binding protein/cell adhesion molecule like; PCDH10: Protocadherin 10; PCDH8: Protocadherin 8; PLA2G16: Phospholipase A2, group XVI; PRDM2: PR domain containing 2, with ZNF domain; PTPRG: Receptor-type tyrosine-protein phosphatase gamma; RARRES1: Retinoic acid receptor responder (tazarotene induced) 1; RARβ2: Retinoic acid receptor beta 2; RASAL1: RAS protein activator like 1; RASFF2A: Ras association domain family member 2A; RASSF1A: Ras association domain family member 1A; RB1: Retinoblastoma protein 1; RBP7: Retinol binding protein 7; SCGB3A1: Secretoglobin, family 3A, member 1; SFN: Stratifin; TFPI2: Tissue factor pathway inhibitor-2; THBS1: Thrombomodulin 1; THY1: Thy-1 cell surface antigen; TNFRSF11B: Tumor necrosis factor receptor superfamily, member 11b; TP73: Tumor protein p73; UCHL1: Ubiquitin carboxyl-terminal esterase L1; WIF1: Wnt inhibitory factor-1; ZMYND10: Zinc finger, MYND-type containing 10; ZNF382: Zinc finger protein 382.

1.3.5.1 Epigenetic regulation of EBV gene expression

In NPC, DNA methylation associated gene silencing is not limited to the inactivation of cellular TSGs. It has been shown that DNA methylation also modifies the Wp and Cp promoters in the EBV genome that leads to the silencing of several EBV genes (nuclear proteins EBNA2, 3A, 3B and 3C). This also plays an important role in the establishment of EBV latency type in NPC.\textsuperscript{312, 313} The expression of EBV LMP1 oncogene in NPC is associated with its promoter methylation status.\textsuperscript{314} Epigenetic regulation of cellular and EBV genes plays a significant role in NPC tumorigenesis, although not yet fully understood.
1.3.5.2  Role of EBV in DNA epigenetic regulation of genes in NPC

EBV-associated NPC originates from the monoclonal proliferation of EBV-infected epithelial cells.\textsuperscript{247, 315} EBV infection is reported to induce genome-wide increase in methylation in EBV-associated NPC.\textsuperscript{254, 316} In comparison with EBV-negative NPC, unique epigenetic profiles with much higher frequencies of TSG promoter hypermethylation has been shown in EBV-positive NPC.\textsuperscript{243} EBV may act as an epigenetic driver for the NPC tumorigenesis;\textsuperscript{243, 247, 317} EBV oncoprotein LMP1 has been shown to drive DNA hypermethylation by activating cellular DNA methyltransferases through c-jun NH\textsubscript{2}-terminal kinase signaling and subsequently induce hypermethylation of several cellular TSGs.\textsuperscript{318, 319} LMP1 has also been suggested to play a role in hypermethylation of several EBV and cellular TSGs by upregulating DNMT1 and polycomb group (PcG) protein Bmi-I.\textsuperscript{320} The histone modifications mediated by PcG also has a role in the hypermethylation of certain TSGs.\textsuperscript{321}

1.3.5.3  Detection of NPC using hypermethylated TSGs

Most studies on detection of NPC based on epigenetic marks have employed analysis of promoter methylation of TSG panels.\textsuperscript{322} In an earlier work,\textsuperscript{323} DNA samples either from NPC paraffin fixed samples or nasopharyngeal brushings were analyzed for ten TSGs by using MSP method. Of these CADM1, CDH13, DAPK1, DLC1 were frequently methylated in individuals with elevated EBV IgA and viral load. A combination of CHFR, p16, RASSF1A, RIZ1 and WIF1 as methylation markers was reported to detect 98\% of the NPC samples.\textsuperscript{323} Other studies have also reported differential methylation of genes such as CDH1,\textsuperscript{251, 252, 256, 324, 325} p16,\textsuperscript{251, 256, 306, 324} RAR\beta\textsubscript{2},\textsuperscript{251, 253, 305} RASSF1A\textsuperscript{251, 253, 256, 306, 324} to be useful biomarker for the detection of NPC (Table 2).

1.4  Introduction to lung cancer

Lung cancer is the leading cause of cancer-related deaths worldwide. Its incidence and mortality is increasing steadily.\textsuperscript{326} It accounts for nearly 13\% (1.6 million cases) of all cancer cases and 18\% (1.4 million) cancer deaths each year on a global scale.\textsuperscript{327} In China, lung cancer is the most common type of cancer detected in the male population (350 000 cases, 21.7\% of all cancers) and is the second most common type
of cancer in the female population (170,000 cases, 14.3%) after breast.\textsuperscript{328} It has shown a dramatic increase of 4.7 times over the past three decades in China.\textsuperscript{329}

Lung cancer is divided into non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC).\textsuperscript{330} NSCLC accounts for about 80\% of the global lung cancer cases.\textsuperscript{331} Histologically, NSCLC is classified into adenocarcinoma, squamous cell carcinoma, and large cell carcinoma.\textsuperscript{330} NSCLC shows a slower growth and spread pattern as compared with SCLC.\textsuperscript{332} The higher mortality rate of NSCLC patients can be attributed to its late diagnosis at an advanced stage of the disease.\textsuperscript{333} The reasons for late diagnosis are the late appearance of symptoms.\textsuperscript{332}

1.4.1 Epigenetic aspect of non-small cell lung cancer

In lung cancer, epigenetic alterations are more frequent than genetic mutations.\textsuperscript{334} Dysfunctional epigenetic mechanisms, such as DNA methylation, histone modifications and ncRNA regulation are common events in lung cancer which can lead to development of lung cancer by activating oncogenes and silencing TSGs.\textsuperscript{335,336} This makes epigenetic alterations potentially interesting biomarkers for early diagnosis in cancer.\textsuperscript{337-339}

DNMTs are upregulated in NSCLC.\textsuperscript{340-342} Promoter hypermethylation of genes involved in key cellular functions such as cell cycle regulation, proliferation, apoptosis, cellular adhesion, motility, DNA repair and genome stability is frequently reported.\textsuperscript{343} The aberrant methylation of several genes involved in important cellular processes such as DAPK, RASSF1A, p16, MGMT, DAPK, RASSF1A, PAX5β, GATA5 and CXCL14 has been associated with lung cancer tumorigenesis.\textsuperscript{344-348} The aberrant promoter DNA methylation of p16 and/or MGMT genes was detectable in sputum of all the SCC subtype of NSCLC patients up to three years before their clinical diagnosis.\textsuperscript{349} Short stature homeobox 2 (SHOX2) hypermethylation has also been suggested to be a potential marker for diagnosis of lung cancer.\textsuperscript{350}

Genome-wide hypomethylation has also been reported during NSCLC development.\textsuperscript{351} It has been associated with genomic instability\textsuperscript{352} activation of the proto-oncogenes\textsuperscript{353},\textsuperscript{354} and loss of imprinting.\textsuperscript{355} In lung cancer, genomic hypomethylation has been reported at nuclear elements, long terminal repeat elements, segmental duplicates, and sub-telomeric regions.\textsuperscript{356}
Histone modifications have been reported in lung cancer where HDACs are overexpressed.\textsuperscript{42, 357, 358} Hyperacetylation at H4K5/H4K8, hypoacetylation at H4K12/H4K16 and loss of trimethylation at H4K20 are frequent in lung cancer.\textsuperscript{359}

Expression of miRNAs is also deregulated in lung cancer.\textsuperscript{360, 361} For example, as compared to normal cells, miR-196a and miR-200b are reported to be 23 times and 37 times overexpressed in lung cancer cells respectively.\textsuperscript{360} IncRNAs such as metastasis-associated lung adenocarcinoma transcript 1 (MALAT-1) and HOTAIR have also been observed to be upregulated in NSCLC.\textsuperscript{362-366} MALAT-1 is involved in silencing of genes that regulate cell migration, invasion, and metastasis.\textsuperscript{367-369} HOTAIR is involved in the regulation of HOXD expression.\textsuperscript{370}

Cross-talk between different epigenetic mechanisms has also been observed in lung cancer. For example, hypermethylation of miR-124a,\textsuperscript{371} miR-34b/c,\textsuperscript{190, 191} miR-9-3,\textsuperscript{372} and miR-193a\textsuperscript{372} has been reported in lung cancer. Moreover, miR-29 targets the transcripts of DNMT3a and DNMT3b and restores their normal DNA methylation patterns.\textsuperscript{373}
2 AIMS OF THIS STUDY

1. To evaluate MMSP assays for simultaneous detection of methylations status of multiple genes. Evaluate its feasibility for identifying NPC using nasopharyngeal swabs as source of DNA (Paper I).

2. Update the MMSP assay with a panel of markers suitable for NPC detection in different NPC endemic areas and to evaluate the feasibility of MMSP for detection of aberrant DNA promoter hypermethylation in samples from different geographical regions (Paper II).

3. To search for putative TSGs in chromosome 3 in NPC by Not I arrays and analyze if such candidate genes are regulated by promoter methylation (Paper III).

4. To select specific and sensitive markers genes, which are hypermethylated and downregulated, for development of an MMSP assay for lung cancer (NSCLC). Evaluate the feasibility of MMSP for simultaneous detection of several aberrantly methylated genes in samples from NSCLC (Paper IV).

5. To try several different approaches for identification of candidate marker genes for the MMSP assays: high throughput expression and methylation microarrays, NotI microarray and literature searches (all papers).
3 RESULTS AND DISCUSSION

3.1 Paper I

Development of a non-invasive method, multiplex methylation specific PCR (MMSP), aiming for early detection of nasopharyngeal carcinoma

DNAs from seven NPC cell lines including a xenograft were used in this study to validate the reproducibility of the MMSP assay. C15 is an African-origin NPC xenograft which is EBV-positive and expresses LMP1. C666-1 is an EBV-positive and LMP1 non-expressing NPC cell line. The other five NPC-derived cell lines used in this study i.e. CNE1, CNE2, HONE1, SUNE1 and HK1, are all EBV negative. The EBV-gene EBNA1 was used as a marker for the presence of EBV in the MMSP assay. It can be detected in the C666-1 cell line and the C15 xenograft. Unmethylated-LMP1 can only be detected in the LMP1 expressing xenograft, C15. None of the other cell lines showed signals for EBNA1 or LMP1. All cell lines and the xenograft showed hypermethylation of at least one of the two TSGs, included in the MMSP assay, i.e. RASSF1A and DAPK

DNAs from 49 NPC biopsies and their corresponding nasopharynx swabs were included in this study to examine the reliability of nasopharyngeal swabs as the source of DNA for the MMSP assay. DNA from nasopharyngeal biopsies and corresponding swabs from 20 non-cancerous volunteers were also included as controls. All DNAs from NPC and control swab samples showed detectable signals for β-Actin gene in our MMSP analysis, indicating enough DNA yield as well as efficient bisulfite conversion of DNA. EBV-encoded EBNA1 was detected in 100% (49/49) of the NPC tumors and 98% (48 of 49) of their paired swab samples. All controls were negative for EBNA1. Unmethylated LMP1 (U-LMP1) was detected in 63% (31/49) of NPC tumors and in 55% (27/49) of their corresponding swab samples, suggesting expression of LMP1 in these samples. Our U-LMP1 result in NPC samples is consistent with the reported expression level of LMP1 in NPC. The sensitivity (Sn) for detection of hypermethylated of RASSF1A and DAPK were 80% (39/49) and 67% (33/49) in the biopsy samples, respectively, while 55% (27/49) and 57% (29/49) respectively in the DNA from their matched swabs.

Among DNA samples from NPC swabs 18% (9/49) were positive for all four MMSP
markers, 35% for three, 31% for two, and 14% for one marker only. Around 43% of the nasopharyngeal swabs showed a pattern in the MMSP assay which matched perfectly with the corresponding biopsies. If we include the presence of EBNA1 plus at least any one of the other three markers as a diagnostic criterion for NPC, we reached a Sn of 98% in the swab samples. No EBV or methylated markers were found in DNAs from swabs of the controls, suggesting 100% specificity (Sp). This result supports the idea to develop MMSP to a screening tool in high-risk populations. Since patterns and timing of methylation status of specific genes are also associated with defined biological behaviors, the MMSP method carries the potential to predict individual tumor behavior, given that the appropriate genes are included in the assay. EBV load and methylation levels of some specific genes have been shown to detect disease relapse after treatment.

Considering the flexibility of the MMSP panel, allowing replacement of tumor-specific markers it will be possible to extend the usage of MMSP assay for diagnosis of other cancers using body fluids, such as urine to detect prostate cancer or bladder cancer, cervical cancer by cervical swab and lung cancer by sputum or saliva etc. (Paper IV).

### 3.2 Paper II

**Detection of nasopharyngeal carcinoma (NPC) from different geographic regions using multiplex methylation specific PCR (MMSP) biomarker assay**

We set out to evaluate and explore combinations of several genes and to try to improve our MMSP assay for NPC (Paper I), so that it can be used on samples from different geographical locations. The DNAs from 44 NPC and 18 normal biopsy samples from Morocco were screened using single gene MSP for 12 markers, including those used in paper I, and in addition eight new markers i.e. ITGA9, p16, WNT7A, CHFR, CYB5R2, WiFi, RIZ1 and FSTL1. Of the Moroccan NPC samples, 36 (82%) were positive while eight samples (18%) were negative for EBNA1. This differs from the pattern in Chinese NPC (Paper I). Among marker genes used in Paper I, RASSF1A was positive in 29/44 and thus was the most sensitive marker (66%) after EBNA1. DAPK showed the lowest sensitivity 11/44 (25%) and specificity 13/18 (72%) in Moroccan NPC. These results demonstrated the need of replacement of DAPK with a more informative marker with higher Sn and Sp. Of the
eight additional genes that were tested in the DNA samples from Moroccan NPC and controls ITGA9: 22/44 (50%) (Paper III) and p16: 20/44 (45%) were the two marker genes with highest Sn showing 100% Sp. On the basis of MSP results a modified MMSP assay was developed which contained EBNA1, LMP1, RASSF1A as before while DAPK was replaced with ITGA9 and p16. The house keeping reference gene β-ACTIN was also replaced with GAPDH to make the developed MMSP assay better suitable for both biopsy and serum samples.

This modified MMSP was used to screen 64 NPC biopsies from three different geographical regions (Morocco 44, Italy 15 and China 5) and 20 non-cancerous nasopharyngeal biopsies from two different geographical regions (Morocco 18 and China 2). At least any one modified MMSP marker could be detected in 91% NPC biopsy samples with 90% specificity. We could also identify three out of eight EBV negative samples with at least one cellular marker. In 16 NPC serum samples from two different geographical regions (Italy 11 and China 5) we could detect at least any one MMSP marker in 88% samples but the specificity was as low as 50%.

We did not observe any marked differences in the methylation patterns among samples from different geographical locations. This MMSP panel seems to work on samples from different geographical locations.

3.3 Paper III

Integrin α9 gene promoter is hypermethylated and downregulated in nasopharyngeal carcinoma

Changes in chromosome 3, which harbors several TSGs, are common in many tumors. NotI microarray (NMA) was performed to screen for potentially methylated genes in chromosome 3 in NPC using genomic DNA from three NPC tumor biopsies, two control nasopharyngeal epithelial tissues, three NPC cell lines (CNE1, TWO3 and C666-1) and one control nasopharyngeal epithelium derived cell line NP69. Ten genes (ALDH1L1, BCL6, EPHB3, FGD5, FGF12, ITGA9, NUDT16P, RBSP3, WNT7A and ZIC4) were identified with the NMA showing reduced signal in NPC samples as compared to control samples suggesting methylation or deletion of these genes in NPC. Correlation between hypermethylation and downregulation of these genes was investigated by performing MSP and Q-PCR.
After further analysis only two genes i.e. ITGA9 and WNT7A qualified for further validation for their downregulation by promoter hypermethylation. This shows the limitations for such screening using NMA.

Downregulation of ITGA9 and WNT7A was observed in NPC samples as compared to normal samples using immunostaining performed on three cases of NPC tumor cells and adjacent normal nasopharyngeal epithelium.

Treating CNE1 and TWO3 with demethylating agent 5-aza-2′-deoxycytidine (5-aza-C) restored ITGA9 and WNT7A expression which supported a hypermethylation of the promoters of ITGA9 and WNT7A. This was further confirmed by performing MSP and Q-PCR after the 5-aza-C treatment.

The methylation status of the CpG rich region in the ITGA9 gene promoter was investigated by bisulfite sequencing in nine NPC biopsies and six control nasopharyngeal epithelia biopsy. This region containing 11 CpG sites and was partially methylated in all NPC samples, whereas only a few CpG sites were partially methylated in the control samples. Similarly methylation status of the CpG rich region in the WNT7A gene promoter was investigated by bisulfite sequencing in four NPC biopsies and three control nasopharyngeal epithelia biopsies. This region contained 84 CpG sites. The WNT7A promoter was partially methylated in the four NPC samples. There was virtually no methylation in the two control samples while one control sample showed some methylation at several CpG sites. Taken together these results suggested that ITGA9 would be a better candidate as a hypermethylated marker gene in NPC than WNT7A.

The hypermethylated status of the ITGA9 gene promoter was quantitatively validated using pyrosequencing of bisulfite converted DNA. Two pyrosequencing assays targeted seven CpGs that were located in the same region that had been before sequenced by cloning of bisulfite converted DNAs. Five biopsies and two control nasopharyngeal epithelia biopsies were sequenced with this method. The majority of the CpGs in ITGA9 gene promoter were partially methylated in all the NPC samples, while they were unmethylated in the control samples.

ITGA9 expression in four control epithelial tissues and three NPC samples was analyzed by using Q-PCR. The average expression level of ITGA9 was 4.0±2.6 in
NPC vs. 19.9±10.2 in the controls. ITGA9 expression was downregulated in NPC tumor samples as compared to control nasopharyngeal epithelium.

Methylation specific PCR was used to screen DNA from 36 EBV positive NPC biopsy samples and 18 non-cancerous control samples from Morocco. ITGA9 methylation was detected by MSP in 56% of NPC DNA samples with 100% specificity. WNT7A methylation was detected by MSP in 69% of the samples but was less specific (83%). Although WNT7A shows higher sensitivity the low level of specificity makes it less attractive to include in screening methods like MMSP. However ITGA9 also is not optimal with a relatively low sensitivity. Nevertheless we employed ITGA9 as one marker in the MMSP assay for NPC in paper II.

Our results suggest that ITGA9 might be a TSG in NPC and promoter hypermethylation one mechanism for downregulation. Aberrant methylation and/or downregulation of ITGA9 has been reported in multiple human cancers.\textsuperscript{378-382} We found that ITGA9 is methylated and downregulated in Chinese NPC patients.

Integrin α9 forms a heterodimer with β1 chain to form a single integrin, α9β1.\textsuperscript{383} This receptor plays an important role in different signal transduction pathways controlling cellular proliferation and differentiation. Integrin α9β1 mediates cell migration in glioblastoma.\textsuperscript{384} As ITGA9 protein executes significant functions related to tumor cell biology, in depth functional studies are relevant to understand its functional role in the development of NPC.

Our results suggest that ITGA9 aberrant epigenetic deregulation of the ITGA9 gene could be useful for future diagnostic approaches in NPC. We explored the possibility of using hypermethylation of ITGA9 as a diagnostic mark for NPC and for this purpose we have included ITGA9 into our assay for early detection of nasopharyngeal carcinoma (Paper II).

### 3.4 Paper IV

**Development of a multiplex methylation specific PCR suitable for (early) detection of non-small cell lung cancer**

Methylation and expression microarrays were performed on three NSCLC and matched distant non-cancerous tissue pairs to identify hypermethylated and downregulated
genes. From these array results 22 genes were identified. The most informative 11 of the 22 markers with 27 additional methylation regulated TSGs 38 (11+27) identified in the literature were pre-screened on a small group of DNA samples from lung cancer and matched control tissues by MSP. The six markers with the highest Sn and Sp identified from this screening were HOXA9, TBX5, PITX2, CALCA, RASSF1A and DLEC1. These six markers were employed in an MMSP assay testing 70 NSC lung cancer with matched controls and 24 additional normal lung tissue samples from patients with benign pulmonary lesion.

Among these six genes, HOXA9 showed the highest Sn i.e. 87% (61/70) whereas DLEC1 was lowest with a Sn of 37%. RASSF1A and DLEC1 provided the highest Sp as 99% (69/70) whereas the TBX5 showed the lowest Sp 86%. All DNAs from 24 normal lung tissue samples from patients with benign pulmonary lesion were negative using these six methylation markers. Methylation of RASSF1A was significantly different between treated and untreated patients. HOXA9 showed a significant difference between smokers and non-smokers.

Among the 70 NSCLC tissue samples, 61 samples were methylated in at least two up to six markers in different combinatorial patterns. The Sn and Sp for at least any two markers were 87% and 94% respectively. An increase in the number of minimal markers increased the specificity but the sensitivity reduced e.g. the Sn and Sp for at least any three markers were 76% and 97% respectively. All six markers were methylated in 21% samples with 100% specificity. Thus, the Sn and Sp revealed by at least any two of the marker genes reached levels which are fit for NSCLC detection.

More importantly with this criterion we could detect 86% (24/28) of the stage I and II samples with a Sp of 100% (28/28). MMSP did not detect methylation in any of the markers in any of the 24 normal lung tissue samples from patients with benign pulmonary lesion whereas its detection rate for any one marker for cancer sample was 99% (69/70). Although this is a satisfactory outcome of the evaluation of this MMSP, one might still consider to combine it with detection of “diagnostic” mutations.
4 METHODS

For details of protocols see the individual papers. This chapter contains a brief discussion of methods used in this thesis in the context of available other methods.

4.1 Extraction of DNA

Several methods were employed but conventional phenol-chloroform method provides high yields of pure DNA, and performs better than commercial kits particularly when working with small biopsies (which is usually the case for NPC).

4.2 Bisulfite treatment of DNA

Techniques for analyzing the methylation status of DNA may provide information from genome-wide methylation level to methylation of single residues in specific genes (Figure 13). Each one of such techniques has its own limitations that may affect the interpretation of data. For example, high performance liquid chromatographic (HPLC) or array based methods can be used to analyze DNA methylation at global or genomic levels while other techniques, such as pyrosequencing and MSP, can be used to analyze methylation of a specific small set of CpGs. Currently several high-throughput methods of genome-wide methylation analysis have been developed and are widely used. They require some validation by gene- or locus-specific methods.

![Classification of techniques for DNA methylation studies](image)

Figure 13. Classification of techniques for DNA methylation studies. Cited from Shen and Waterland RA 2007.
DNA methylation analysis methods can be broadly grouped on the basis of different techniques used for pretreatment of DNA such as affinity enrichment of methylated DNA, chemical modification with sodium bisulfite, and digestion with methylation-sensitive restriction enzymes. In techniques based on affinity enrichment of methylated DNA, such as methyl-DNA immunoprecipitation (MeDIP) and methyl CpG binding protein capture (MBDCap), DNA is enriched with anti-5mC antibodies or MBD proteins for DNA methylation analysis. In methods like HPLC, enzymatic hydrolysis of the DNA is done before analysis. In other procedures, such as, restriction landmark genomic scanning (RLGS), DNA is pretreated with enzymatic digestion. But the majority of techniques such as bisulfite-based genomic sequencing, MSP, bisulfite pyrosequencing, Illumina® and combined bisulfite restriction analysis (COBRA), depend on the pre-treatment of genomic DNA with sodium bisulfite, originally described in 1970.

Bisulfite modification of DNA converts unmethylated cytosine residues to uracil with no effect on methylated cytosines. It is the golden standard for detection of the methylation status of CpGs in a sequence-specific context. Table 3 enlists some of the techniques in which bisulfite treated DNA is used.

**Table 3.** Some techniques for DNA methylation analysis based on bisulfite converted DNA.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Application</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisulfite genomic sequencing</td>
<td>Qualitative analysis</td>
<td>391-393</td>
</tr>
<tr>
<td>MSP</td>
<td>Qualitative analysis</td>
<td>394, 395</td>
</tr>
<tr>
<td>MS-HRM</td>
<td>Qualitative analysis</td>
<td>396-398</td>
</tr>
<tr>
<td>Ms-SnuPE</td>
<td>Quantitative analysis</td>
<td>399</td>
</tr>
<tr>
<td>COBRA</td>
<td>Quantitative analysis</td>
<td>400</td>
</tr>
<tr>
<td>Methylight or Q-MSP</td>
<td>Quantitative analysis</td>
<td>401-405</td>
</tr>
<tr>
<td>MS-DGGE</td>
<td>Complex informational analysis</td>
<td>406</td>
</tr>
<tr>
<td>MS-SSCA</td>
<td>Complex informational analysis</td>
<td>407</td>
</tr>
<tr>
<td>MS-DHPLC</td>
<td>Complex informational analysis</td>
<td>408</td>
</tr>
<tr>
<td>Illumina</td>
<td>Marker discovery</td>
<td>409-411</td>
</tr>
<tr>
<td>MSO</td>
<td>Marker discovery</td>
<td>412</td>
</tr>
</tbody>
</table>

MSP: methylation specific PCR; MS-HRM: Methylation-sensitive high resolution melting; Ms-SnuPE: methylation-sensitive single nucleotide primer extension; COBRA: combined bisulfite restriction analysis; Q-MSP: quantitative methylation specific PCR; MS-DGGE: methylation-specific denaturing gradient gel electrophoresis; MS-SSCA: methylation-specific single-strand conformation analysis; MS-DHPLC: methylation-specific denaturing high-performance liquid chromatography; MSO: methylation-specific microarray.
Prior denaturation of the DNA before bisulfite treatment is a critical step as bisulfite efficiently converts cytosine bases only in single DNA strands. The differential deamination of cytosine by bisulfite treatment involves sequential sulfonation, hydrolytic deamination and alkaline desulfonation. This results in two DNA strands that are no longer complementary to each other because of the C – U conversion of unmethylated cytosines induced in both individual DNA strands by bisulfite. Incomplete C - U conversion is the most important limitation of the bisulfite approach that could lead to false positive results. In addition there is a risk of DNA degradation. Some newly developed techniques, such as a solid-state nanopore-based methylation sensitive assay, do not require bisulfite conversion. This assay also does not require fluorescent labeling and PCR. Such methods could therefore prove very useful in studying the role of DNA methylation in human disease.413

4.3 Methylation specific PCR

Methylation specific PCR (MSP) is widely used for studying the methylation status of CpG dense regions that are frequently found in the promoters of many TSGs. Cytosines in CpG rich regions are usually unmethylated in normal tissues, but become hypermethylated in cancer-associated genes. The differences between methylated and unmethylated alleles after bisulfite modification of cytosine residues make the basis of MSP. The primers for MSP are designed according to these differences to distinguish between the differential methylation statuses. This is why MSP primer designing is a critical step for a successful MSP assay. The MSP primers should be designed with the following considerations:

i. the annealing temperatures of both the primers must be similar ranging between 55 to 65°C;

ii. the size of the PCR product should be between 80 and 250 bp;

iii. each primer should contain at least two CpG dinucleotides;

iv. both the primers should have at least one CpG at its 3’-end;

v. each primer should contain some non-CpG cytosines to avoid false positives results because of amplification of unmethylated but still unmodified DNA.

MSP is a simple and cost efficient technique. As the bisulfite treatment also causes some degradation of the DNA the MSP reaction conditions should be optimized. As the MSP is a very sensitive method and can detect one methylated DNA copy in
thousand unmethylated DNA copies which makes the optimization of the PCR assay such as number of magnification cycles very important for the reliability of the results. A weakness of the system is the subjective nature of the results which, in the future needs to be addressed by developing it into a quantitative method for the analysis of multiple markers such as quantitative digital MSP.

4.4 Multiplex methylation specific PCR

Multiplex methylation specific PCR (MMSP) is a useful technique for studying the methylation status of multiple biomarkers in a single reaction. For MMSP less DNA is needed which makes it more material-economic than simple MSP. Principally, MMSP is based on the same principle as MSP, except that the analysis is run in a single reaction tube and all included genes are amplified proportionally according to optimization by MSP. The increased number of primers needed (two for each gene) makes the primer design more tedious. This method is particularly useful in case of limited amount of sample, e.g. when starting with small biopsies.

4.5 Cell lines or tumor tissues

We employed cell lines from NPC, Burkitt’s lymphoma and NSCLC. Cell lines are excellent tools for mechanistic studies as well as for generating hypothesis about epigenetic regulation of specific genes. However their epigenetic profile may be deformed compared to the tissue/tumor of origin due to explantation and the specific conditions of in vitro growth. There are also concerns regarding the identity and correct tissue origin of cell lines.

To map a more tumor/tissue correct epigenetic profile it is necessary to use ex vivo biopsy material. However, tumor biopsies contain a large proportion of non-tumor infiltrating and stromal cells, which may constitute half or more of the tumor tissue. NPC is, for example, well known to be heavily invaded by CD4+ T-cells. Such cell dilute signals specific for the tumor cells. Tumors are also known to show great intra- and inter-tumoral heterogeneity, which might be evaluated by micro-dissection and the use of large cohorts of patient biopsies.
4.6 Treatment of cells with 5-aza-2’-deoxycytidine

5-aza-C is an analogue of the cytosine nucleotide. When incorporated into DNA, it inhibits DNMTs. Thus, in vitro treatment of cells with 5-aza-C leads to demethylation of the cytosines as de-novo synthesized DNA cannot be methylated. This may result in activation of silent genes, such as hypermethylated TSGs in cancer cells. It is important to optimize time and dosage for 5-aza-C treatment individually for each cell line.

4.7 Bisulfite sequencing of cloned PCR products

Sequencing of bisulfite converted DNA is a means of identifying methylated cytosine residues with a specific sequence/gene. Bisulfite converted DNA can be directly sequenced but unlike the sequencing of clones, which was done in our work, the direct method of bisulfite sequencing produces average methylation values for populations of DNA molecules.

The challenging step for successful bisulfite sequencing is the primer design. MethPrimer is a good online tool for designing primers for bisulfite sequencing (http://www.urogene.org/methprimer/index1.html). The software identifies CGIs the DNA sequence of study. Primers around the CGIs are then suggested by the software (or in other regions specified by the user). The NCBI database “Gene” (http://www.ncbi.nlm.nih.gov/gene) was used to find the molecular location of our target genes. The UCSC genome browser (http://genome.ucsc.edu) was used to get the DNA sequence. RefSeq Genes filter was used to mark the exons.

4.8 Pyrosequencing of bisulfite converted DNA

The pyrosequencing technology is based on sequencing-by-synthesis. It relies on the luminometric detection of inorganic pyrophosphates (PPI) which are released by the incorporation of nucleotides catalyzed by primer-directed DNA polymerase. Unincorporated nucleotides are subsequently degraded by nucleotide-degrading enzyme - Apyrase. Addition of dNTPs is performed sequentially according to the template sequence. DNA stretches up to only one hundred bases can be sequenced with this method, which is a limitation. It provides the advantages of precision, flexibility, parallel processing and can also be easily automated. The technique does not need
labeled primers, labeled nucleotides nor gel electrophoresis. Compared to conventional sequencing based on plasmid clones with inserted DNA fragments, pyrosequencing is fast, simple to use and delivers sequence data in real-time with high accuracy and quantity. This protocol has been reported to detect DNA methylation levels as low as 5%.415

4.9 Immunostaining

Immunostaining is performed for the detection of protein expression in tissue sections with specific antibodies. This can provide information about the location where protein is being expressed and also the intensity of the positive signal. The immunostaining results were analyzed, evaluated and scored by senior pathologists with no knowledge of the clinicopathological outcome of the patients. The most critical aspect of immunostaining is the quality of the antibodies used: their sensitivity and specificity.

4.10 NotI microarray

This method can be used for genome-wide search for methylated or deleted genes. The principle of the method is that NotI restriction enzyme cuts only unmethylated CpG pairs located within the NotI recognition site (5’-GCGGCCGC-3’).416 NotI digested fragments make up only a small fraction (0.1-0.05%) of the human genome and are almost exclusively positioned in CGIs. The NMA we used, covers 188 potential TSGs in chromosome 3, the region we focused on, to compare the normal and tumor cells at the chromosomal level. NotI microarray makes it possible to screen for overall patterns of methylation/deletion patterns in normal and cancer cells.417

4.11 Methylation microarray

Illumina infinium® methylation assay (Illumina Inc. USA) is the most widely used array-based methylation profiling platform. By applying the principle hybridization between two complementary strands of nucleic acids, Illumina Infinium HumanMethylation450 BeadChip can detect the DNA methylation status of 99% of genes and 96% of CGI regions throughout the genome in a single experiment.418 For an individual CpG site, a pair of bead-bound probes is used to detect the presence of T (unmethylated state) or C (methylated state). Beta values range from 0 (completely unmethylated) to 1 (fully methylated) and provide a quantitative readout of relative
DNA methylation for each CpG site, which relates to the percentage of methylation for a given site.

4.12 Expression microarray

Various manufactures produce genome-wide gene expression microarray platforms. Expression arrays from Affymetrix (Affymetrix Inc. USA) and Agilent (Agilent technologies USA) make use of short oligonucleotides to investigate gene expression levels. Illumina ((Illumina Inc. USA) uses BeadArray technology based on 3-micron silica beads. Each bead contains millions of copies of a specific oligonucleotide that act as capture sequences. Developments of the techniques have made it possible to incorporate more probes in a single array allowing higher accuracy in the estimate of gene expression levels. For example, the number of probes per array in Affymetrix GeneChip platform has increased over 100-fold, from the early versions (45,000 probes) to the later (5,500,000 probes).

4.13 Statistical analysis

The correlation between gene expression and clinicopathological parameters were assessed using Chi-Square test. Differences were considered as significant when p < 0.05. Sensitivity was calculated as the percentage of the number of positive cases in the patients divided by total number of cases tested. Specificity was calculated as the percentage of the total number of tested controls minus number of positive cases in controls divided by total number of tested controls.
5 CONCLUDING REMARKS

In this work, we have tried to identify and evaluate the possibility to use aberrant DNA methylation in cancer (NPC and NSCLC) as a marker for detection of the cancer. A number of methods have been used to identify marker genes possible to use in an MMSP assay. These included methylome and expression arrays, Not I microarray, methylation sensitive PCR and literature searches. For some novel candidates in vitro cell assays were used as complement to study impact of control by methylation. The MMSP assay was mostly applied on tumor material, but also on DNA from nasopharyngeal swabs and serum. The main conclusions were:

1. We developed MMSP assays for NPC and NSCLC. This as an assay which allows simultaneous detection of methylated DNA sequences, in our case derived primarily from aberrantly methylated promoter regions. The sensitivity and specificity of MMSP in detecting 49 NPC swabs was 98% and 100% respectively. It is only semi-quantitative, but could probably be developed to be quantitative utilizing multiple color based Q-PCR. Our assay could simultaneously detect EBV and methylation status of multiple viral and cellular genes using DNA from nasopharyngeal swabs.

2. This MMSP assay was modified in an attempt to optimize it for NPC-detection in different high/medium-risk areas. With this modification it worked equally well in three different geographical countries, although some of the materials were small due to shortage of biopsy material. We could detect NPC markers in some of the EBNA1 negative cases. A first effort was made to apply the MMSP assay on a small set of NPC serum samples, but the result was disappointing. NP swabs seemed much more useful. To use blood as a source sensitivity and specificity has to be increased. One way may be to use whole blood or plasma as a source, or even start by isolation of circulating tumor cells. Further development of the approach is definitely needed before it can be tested in population-based screening for early detection of cancer.

3. We discovered two additional genes, ITGA9 and WNT7A, where aberrant promoter methylation partly explain their downregulation. For the MMSP assay purpose we judged ITGA9 to be better due to higher specificity. The possible role of ITGA9 and WNT7A as TSGs in NPC deserves further functional analysis.
4. We developed an MMSP assay for NSCLC. In tissue derived DNA the Sensitivity and Specificity of detecting NSCLC were 87% and 94% respectively. Next step will be to apply the MMSP protocol on DNA from bronchoalveolar lavage fluid, sputum or even serum from lung cancer patients.

There are now many different types of biomarkers with a potential to be used to detect cancer early by screening, to guide choice of and follow effects of treatment and to distinguish subtypes within one group of cancer. In addition to serum proteins and DNA with specific mutations, they include microRNAs and exosomes. Tumor specific methylated DNA should be added as another highly interesting type of biomarker, as aberrant DNA methylation can be one mechanism to silence TSGs and may as such appear very early during tumorigenesis. It has the advantage of being highly specific if the right marker gene is chosen and the molecules are relatively stable. As epigenetic reprogramming is a hallmark of cancer to be added there is a large spectrum of methylated genes/sequences to choose between in each type of cancer. From these it will be important to choose the most specific and informative markers. Individual epigenetic tumor markers have to be distinguished from those common to a tumor type. Novel high through put methods will provide large data sets to select candidate genes/sequences from, but these methods are yet too insensitive to be applied on small amounts of methylated DNA in body fluids. Issues of sensitivity still have to be resolved, and isolation, quality and storage of biological specimens are always important. Isolation of circulating tumor cells from blood combined with methylation profiling is an interesting possibility, as well as biotechnological improvements of assays to detect specific methylated sequences with high sensitivity.

Like with any biomarkers, their usefulness has to be validated in large clinical studies, and it will be a challenge to extract those markers that will make it all the way to such studies.
6 ACKNOWLEDGMENTS

Now as I have finished this long journey I cannot forget how it was when I was starting it. I am in debt of so many wonderful people that have contributed, directly or indirectly, to the completion of this journey. All these people deserve a special place in this thesis. I would like to acknowledge all their contribution and help and I hope I do not miss out anyone's name. But If I do, please know that the deepest appreciation and gratitude will always remain in my heart.

First of all, I would like to express my special gratitude and thanks to my supervisor Prof. Ingemar Ernberg. I feel very lucky to be a part of your group and supervised by you. Not only this but I am greatly thankful for your never-ending support, ideas, motivation and encouragement that will always inspire me in my future life. Thanks a lot Ingemar, you applied both the first and the last touch to my PhD work including thesis write up.

Prof. Li-Fu Hu, you deserve a special appreciation for inviting and introducing me to MTC as my co-supervisor. I am greatly thankful for your continuous support and the patience that you always showed on the mistakes that I made during the course of my studies. I will always remember your politeness and patience.

I would like to thank Prof. Qinghua Zhou his generous offer to work in your research laboratory in China. These visits proved to be really very informative and productive. I would also like to thank all your group members especially Yanjie Qiao, Xiaoming Qiu, Bin Liu for making my stay in China very pleasant.

I would also like to thank Prof. Tomas Ekström for your collaboration and giving me permission to work in your laboratory for pyrosequencing work.

I would like to say thanks to Prof. Rolf Ohlsson for giving me permission to use his laboratory equipments. I would also like to thank his group members: Chengxi, Farzaneh, Li-sophie, Marta, Nori, and for their cooperation:

I also, highly appreciate the educational administrator at MTC Anita Wallentin for her assistance and support in official matters in a highly efficient and professional manner.
I would specially thank Malin Almgren, Ilya Ignatyev, Wang Zuobai, Zi-Ming Du and Klas Haraldson for their assistance and guidance to teach me some molecular techniques.

I would like to acknowledge all my collaborators for having pleasure to publish with. Di Sun, Do Nguyen Van, Guangwu Huang, Heng Wu, Ilya Ignatyev, Khalid Moumad, Prof. Debora Martorelli, Prof. Eugene Zabarovsky (late), Prof. Jaap M. Middeldorp, Prof. Meriem Khayatt, Prof. Qinghua Zhou, Prof. Riccardo Dolcetti, Prof. Moulay Mustapha Ennaji, Sofia M. Haryana, Susanna Hilda Hutajulu, Tatiana Pavlova, Vladimir Kashuba, Yaguang Fan, Yang Li, Zhe Zhang and Zi-Ming Du.

Special thanks to all current and previous members and associates to Ingemar’s group. Special gratitude to Fu Chen (late). You were really always there to help and guide me when I started taking my first steps in the research. But it is a pity that you could not see my graduation today. I cannot express the feelings in my heart for you. Jiezhi Zou, for all your help in the lab and arranging lovely Friday coffees. Elvira Grigorieva for nice discussions on various issues. Qin Li and Xiaoying Zhou best of luck for your theses. Thank you all very much: Anastasia Piskaeva, Anders Wennborg, Andrey Alexeyenko, Anna Aleman, Anna Birgersdotter, Anna-Karin Persson, Aymeric Fouquier d'Herouel, Benedek Bozóky, Elisabeth Lissa Norin, Gosta Winberg, Helena Lonnqvist, Henrik Brändén, Ilgar abdullahyev, Joakim Coster, Konrad Kaufmann, Li-Sophie Zhao Rathje, Lidiia Astakhova, Liudmila Matskova, Prof. Eugene Zabarovsky (late), Prof. Gunnar Klein, Sven Grützmeier, Tatiana Pavlova, Tore Midtvedt, Vladimir Kashuba, Wang Zoubai, Xue Xiao, Yvonne Strandberg and Zi-Ming Du.

I would also like to thank Prof. Geogre Klein, Prof. Eva Klein, Prof. Anneka Ehrnst (late) and all their group members Barbro Ehlin-Henriksson, Liang Wu, Noemi Nagy, Elena Kashuba, Daniel Salamon, Eahsan Rasul, Harsha Madapura, Monika Adori, Suhas Darekar, Muhammad Mushtaq, Rozina Cardha, and Lotta Pramanik for the instructive joint group meetings and the knowledge you ever brought to me.

Thanks to my students Ebba Orava, Olga Axell, Vendela Leifland Berntsson. Working with you was really a wonderful experience.
Special thanks to all the fantastic people in Sweden who made who my life a wonderful experience outside the Lab:

**Muhammad Sharif Hasni**, I would like to thank you in special for all your help and support of all kind from the very first day at the airport till date, long Friday night discussions and all your hospitality.

**Irfan Ahmad, Mohsin Barq** and **Muhammad Mushtaq** for your help, support and lunch-time discussions at MTC. All my friends and colleagues at MTC: **Farzaneh Shahin Varnoosfaderani, Naeem Anwar, Sadia Salam Rana** and **Siti Mariam Zakaria**.


Thank you very much from the bottom of my heart for the family environment that you provided to us in Sweden: **Mr. and Mrs. Abdolwahab Karimzehi, Mr. and Mrs. Akherdad Hosseinbor, Mr. and Mrs. Alamdar, Mr. and Mrs. Hashim, Mr. and Mrs. Irfan, Mr. and Mrs. Khodabakhsh Mirjaveh, Mr. and Mrs. Mohammad Atta Hosseinbor, Mr. and Mrs. Mushtaq, Mr. and Mrs. Osman lala, Mr. and Mrs. Shahzeb Iqbal, Mr. and Mrs. Siddik Azad** and **Mr. and Mrs. Taj Mohammad Breseeg**.

Special acknowledgment to my friend **Yasir Buzdar** who deserves a heartily appreciation for motivating me to pursue PhD studies from abroad.

I would also like to express my gratitude for **Prof. Dr. Masoom Yasinzai, Engr. Ahmed Sohail Bazai, Engr. Farooq Badini, Dr. Tahir Waheed, Dr. Muhammad Alam Mengal, Dr. Farhat Iqbal** and **Abdul Raziq Kakar** for their help and support in all the official matters in the University of Balochistan.

I would not have accomplished whatever I have, had there been not continuous and incredible support and father **Shah Nawaz Bugti**, my wonderful **mother (Ammi)** and **brothers (Sarfaraz, Umair and Shaheer)** and **sisters**, my father-in-law **Mir Abdul**
Haleem Satakzai, my mother-in-law, Aaee, and brother-in-laws Adil Satakzai, Jamal Nasir, Shehzad Satakzai, and Sher Ahmed Satakzai. You all deserve special thanks, gratitude and acknowledgements, that perhaps, I will never be able to express in words.

Last, but by no means least, I am most grateful to my wife for her exceptional display of patience and understanding for all these years. She did all she could do to support and encourage me throughout this journey. I am also grateful to my son Hamza and my daughter for filling the most beautiful colors in my life. I am also thankful to them for their understanding of the importance work that I was engaged in and still making this journey beautiful, pleasant, smooth, and happy.

I would like to thank the Cancerfonden; Cancerföreningen; China-Sweden International Scientific and Technological Cooperative Project; Early Diagnosis and Treatment Project for Lung Cancer of the Central Fiscal; EU-Asialink grant; National Basic Research Program of China; National Eleventh-Five-Year Key Task Projects of China; Natural Science Foundation of China; Nuffic from the Dutch Academy for Sciences; Radiumhemmets Forskningsfonder; SIDA (Swedish International Development Cooperation Agency); SSMF (Swedish Association of Medical Research); STINT (The Swedish Foundation for International Cooperation in Research and Higher Education); Swedish Cancer Society; Transfer payment of China; and University of Balochistan, Pakistan for financing my research projects and PhD studies.
7 REFERENCES


70. Rose NR, Klose RJ. Understanding the relationship between DNA methylation and histone lysine methylation. Biochimica et biophysica acta 2014.


118. Phillips DM. The presence of acetyl groups of histones. The Biochemical journal 1963; 87:258-63.


CCNA1, RARRES1, and HRASLS3 in nasopharyngeal carcinoma. Oral oncology 2008; 44:400-6.


386. Nair SS, Coolen MW, Stirzaker C, Song JZ, Statham AL, Strbenac D, Robinson MD, Clark SJ. Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. Epigenetics: official journal of the DNA Methylation Society 2011; 6:34-44.


