RESTRICTION LANDMARK GENOMIC SCANNING TO IDENTIFY NOVEL METHYLATED AND AMPLIFIED DNA SEQUENCES IN HUMAN LUNG CANCER

DISSEPTION

Presented in Partial Fulfillment of the Requirement for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By Zunyan Dai, M.S.

The Ohio State University

2002

Dissertation Committee:
Christoph Plass, Ph.D., Adviser
W. James Waldman, Ph.D.
Stephen J. Qualman, M.D.
Ming You, M.D., Ph.D.
Gregory A. Otterson, M.D.

Approved by

Adviser
Pathology Graduate Program
ABSTRACT

Lung cancer is the leading cancer related deaths in the United States. The majority of lung cancers are sporadic diseases and attributed to cigarette smoking. Both loss-of-function of tumor suppressor genes and gain-of-function of oncogenes are involved in lung cancer development. Knudson’s two-hit hypothesis postulates that genetic alterations in both alleles are required for the inactivation of tumor suppressor genes. Genetic alterations include small or large deletions and mutations. Over the past years, it has become clear that epigenetic alterations, mainly DNA methylation, are additional mechanisms for gene silencing. Restriction Landmark Genomic Scanning (RLGS) is a two-dimensional gel electrophoresis that assesses the methylation status of thousands of CpG islands. RLGS has been successfully applied to scan cancer genomes for aberrant DNA methylation patterns. So far, the majority of this work was done using NotI as the restriction landmark site. Following the introduction, chapter one, we describe in chapter two the development of RLGS using AscI as the restriction landmark site for genome wide scans of cancer genomes. The availability of AscI as a restriction landmark for RLGS allows for scanning almost twice as many CpG islands in the human genome as compared to RLGS using NotI only. We describe the development of an AscI-EcoRV boundary library that supports the cloning of novel methylated genes. Feasibility of this system is demonstrated in three tumor types: medulloblastomas, lung cancers and head
and neck cancers. We report the cloning of 178 Ascl RLGS fragments via two methods using this library.

Epigenetic changes, including DNA methylation, are a common finding in cancer. In lung cancers methylation of cytosine residues affects tumor initiation and progression in several ways, including the silencing of tumor suppressor genes through promoter methylation and by providing the targets for adduct formation of polycyclic aromatic hydrocarbons present in combustion products of cigarette smoke. Although the importance of aberrant DNA methylation is well established, the extent of DNA methylation in lung cancers has never been determined. In chapter 3, we selected 1184 CpG islands for RLGS analysis and determined their methylation status in 16 primary non-small cell lung cancers. Some tumors did not show methylation while others showed up 5.3% methylation in all CpG islands of the profile. Cloning of 21 methylated loci identified 11 genes and 6 ESTs. We demonstrate that methylation is part of the silencing process of BMP3B in primary tumors and lung cancer cell lines.

BMP3B was previously reported to be a methylation target in lung cancer. DNA methylation was associated with down regulation of transcription of the BMP3B message in both primary human lung cancer and lung cancer cell lines. Although BMP3B is a member of TGF-β superfamily, its role in lung cancer is not well characterized. In chapter 4, in order to understand the possible mechanisms of BMP3B silencing in lung cancer, 96 primary non-small cell lung cancers were used to detect BMP3B promoter methylation and mutation, and 17 paired primary non-small cell lung cancer samples were tested for loss of heterozygosity (LOH). In addition, BMP3B was reexpressed in the A549 lung cancer cell line to evaluate the significance of BMP3B silencing. Our results
showed that approximately 45% of primary lung cancer samples exhibited increased methylation, and 35% demonstrated LOH in microsatellite marker D10S196. Although no significant cell growth rate change was observed after BMP3B reexpression in A549 cell line, the number of colonies formed in anchorage-dependent assays was slightly decreased. In addition, the colony formation ability of A549 cells after BMP3B expression in anchorage-independent assays in soft agar was significantly reduced (P<0.005, t-test). Moreover, A549 cell growth assays in nude mice indicate that BMP3B suppresses tumor formation in vivo. Finally, our microarray expression profiling data, combined with information from literature, suggest that BMP3B and TGF-β might have opposing role, as shown for the regulation of fibronectin expression. In conclusion, this study provides evidence that BMP3B expression is repressed by different mechanisms in lung cancer, and that the silencing of BMP3B might promote lung tumor development.

Amplification of oncogenes is an important mechanism that causes gene overexpression and contributes to tumor development. The identification of amplified regions might have both prognostic and therapeutic significance, such as HER2/neu in breast cancer. It is of great interest to identify novel amplicons and oncogenes in lung cancer. In chapter 5, RLGS was used to screen primary lung cancer and cell lines to identify novel amplified sequences. Enhanced RLGS fragments that might indicate gene amplification were observed in primary tumors and cell lines of both non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). In addition to the previously reported amplicons such as MYCC, MYCL1 and 3q26-27, we have identified 2 novel amplicons in lung cancer. 11q22 had been reported in other types of cancer, but 6q21 has not been previously characterized. All amplifications were confirmed by Southern
hybridization. The amplified region of 11q22 was refined to approximately 1 Mbp region that has been completely sequenced. Genes localized in this region are inhibitors of apoptosis (cIAP1 and cIAP2), matrix metalloproteinases (MMP7, MMP20, MMP27, MMP8, MMP10, MMP1), YAP1 and porimin, as well as approximately 10 novel genes. This study provides a strong basis for the future investigation of amplified genes in 11q22 and the other newly identified regions.

In summary, the current studies have successfully established Ascl RLGS and Ascl-EcoRV boundary plasmid library and library mixing RLGS gels to facility the identification of methylated and amplified genes in cancer development. The overall methylation level of lung cancer was for the first time evaluated and novel methylated genes were characterized. BMP3B as an aberrantly methylated gene, showed down regulation in 5 out 6 human lung cancers. Restoration of gene expression suppresses in vitro and in vivo cell growth of human lung cancer cell line A549. In addition, we have identify novel amplicons, for which candidate oncogenes need to be identified. Overall, this indicates that the characterization of these novel methylated genes with loss-of-function and the amplified genes with gain-of-function during lung cancer development will improve the understanding of lung cancer tumorigenesis. Our study should provide better knowledge of lung cancer mechanism for better diagnosis and treatment of lung cancer, which is one of the greatest threats to health.
Dedicated to my family, teachers

and all friends
ACKNOWLEDGMENTS

I would like to thank my adviser, Dr. Plass, for his excellent guidance during my graduate studies. Dr. Plass encouraged me during my dissertation research project whenever I needed assistance. He maintained the lab in such an efficient way that lead to assured success. He discussed every detail of my project, helped me focus on the proper research directions, and helped me to organize ideas suitable for publication. His endless support is unforgettable, and my research has greatly benefited from his excellent training.

Members of Dr. Plass’ lab are wonderful, friendly, cooperative, and creative. The helpful ideas and support offered by this large group of fellow researchers make success possible. The lab members, including Dominic Smiraglia, Laura Rush, Anthony Popkie, Laura Smith, Aparna Raval, Julie Hall, Yue-Zhong Wu, Romulo Brena, Yu Li, Stephen Lee, Chunhui Liu, Melania Bembea and Michael Fruhwald, provided me great technical support and important discussions about my research project. Only with their experience and guidance was the dissertation possible.

My advisory committee members, consisting of Dr. W. James Waldman, Dr. Stephen J. Qualman, Dr. Ming You, and Dr. Gregory A. Otterson, continually devoted their precious time and experience to direct my research.
The following laboratories at the Ohio State University, Dr. Gregory Otterson’s Lab, especially Wei-Guo Zhu; Dr. Gustavo Leone’s Lab, especially Cynthia D. Timmers; Dr. Tom Rosol’s lab, especially Sarah Tannehill-Gregg, Dr. Ramana V. Davuluri lab, Dr. Carl Morrison, Dr. Denis C. Guttridge, also provided excellent expertise for my project.

Dr. Robert A. Kratzke and Gloria A. Niehans from Minneapolis VA Medical; Dr. Harvey Pass and Anil Wali from Wayne State University, Detroit; Dr. Julian Molina, Mayo Clinic, Rochester, MN, provided us with invaluable primary lung cancer samples for the study.

My family also made great contribution to support me to go this far. My wife, Ji Zhang, lent abundant support during my graduate studies at OSU. My parents and aunt also provided me vital encouragement and support during my extensive training.

Dr. Barth served on the Candidacy Examination Committee and he guided me through my first year of study at the Ohio State University. The Graduate School of the Ohio State University also awarded me the University and Presidential Fellowships to support my first and last year study.

Thanks for Dr. Plass, Dr. Dominic Smiraglia and Laura Smith for their great efforts to improve my dissertation. In addition, I received other suggestions from my friends, family and teachers not mentioned here.

Finally, I would like to say that all of the guidance during my Ph.D. training made the dissertation possible. All of your kindness is greatly appreciated!
VITA

December 17, 1969  …………….Born – Yinan, Shangdong, P.R.China

1993…………………………….Bachelor of Medicine, Shanghai Medical University.

1996…………………………….Master of Science, Shanghai Medical University.

1996-1997………………………Researcher, Liver Cancer Institute, Shanghai Medical University

1997-present……………………Ph.D. Candidate, Department of Pathology, The Ohio State University

PUBLICATIONS

Research Publication


FIELDS OF STUDY

Major Field: Pathology,

Human Cancer Genetics and Molecular genetics.
TABLE OF CONTENTS

ABSTRACT ............................................................................................................... ii
DEDICATION ........................................................................................................... vi
ACKNOWLEDGMENTS ............................................................................................. vii
VITA ........................................................................................................................ xix
LIST OF TABLES .................................................................................................... xvi
LIST OF FIGURES ................................................................................................ xviii
Chapters:

1. Introduction ........................................................................................................... 1
   1.1 Human lung cancer ..................................................................................... 1
   1.2 Molecular genetic changes in human lung cancer ....................................... 3
       1.2.1 Genetic susceptibility to lung cancer ................................................. 3
           1.2.1.1 Inherited susceptibility ............................................................... 3
           1.2.1.2 Genetic polymorphism related to tobacco carcinogens .......... 4
       1.2.2 Tumor suppressor genes .................................................................... 4
           1.2.2.1 Definition of tumor suppressor genes .................................... 5
1.2.2.2 Mechanisms of loss of function of tumor suppressor genes...

1.2.2.3 Tumor suppressor genes in lung cancer:
   - p53, p16/CDKN2A and RB ........................................ 6

1.2.2.4 Examples of newly identified candidate tumor suppressor
genes in lung cancer: FHit, RASSF1A, SEME3B, Kras2...

1.2.3 Oncogenes................................................................. 8
   - Known oncogenes involved in lung cancer ................. 8
   - Mechanisms of amplification ...................................... 9
   - Strategies to identify novel amplicons ....................... 9

1.3 Gene inactivation by epigenetic changes:
   - DNA methylation and histone deacetylation, histone methylation .... 10
   - Introduction of epigenetic changes ............................. 10
   - DNA methylation ...................................................... 11
   - Histone acetylation and histone methylation .................. 13
   - Connection between DNA methylation and histone
     modification for gene transcription regulation ............. 14

1.4 DNA methylation in cancer ........................................... 15

1.5 DNA methylation in lung cancer ................................... 16
   - Methylated genes in lung cancer ......................... 16

1.5.2 DNA methylation as biomarker for early diagnosis in
   human lung cancer ................................................. 17

1.6 Techniques for the study of aberrant methylation ............. 18
   - General introduction .............................................. 18
1.6.2 Methods based on DNA methylation sensitive restriction enzymes

1.6.2.1 Restriction landmark genomic scanning (RLGS)

1.6.2.2 Other techniques based on methylation sensitive restriction enzymes: MCA, MS-AP-PCR, DMH

1.6.3 Methods based on bisulfite treatment: Bisulfite sequencing, MS-PCR, COBRA MS-SNuPE, MS-SSCP, MS-HPLC

1.6.4 Other techniques used in screens for methylated genes

1.7 Hypothesis and the summary of the study

2 An AscI boundary library for the studies of genetic and epigenetic alterations in CpG islands

2.1 Introduction

2.2 Materials and methods

2.3 Results and discussion

2.4 Conclusion

3 Global Methylation Profiling of Lung Cancer Identifies Novel Methylated Genes

3.1 Introduction

3.2 Materials and methods

3.3 Results

3.4 Discussion
LIST OF TABLES

Table 1.1. Genes that have been shown promoter methylation in lung cancer…. 28
Table 1.2. Techniques to detect DNA methylation…………………………….. 29
Table 2.1. Characteristics of NotI and Ascl restriction sites……………………. 48
Table 2.2. Methylation frequencies in various tumor samples determined
by either NotI or Ascl as a restriction enzyme……………………… 49
Table 2.3. Cloned Ascl spots……………………………………………………. 50
Table 3.1. DNA methylation in 16 lung cancer patients………………………… 73
Table 3.2. Distribution of methylated CpG islands in various
histopathological subgroups of non-small cell lung cancer………….. 74
Table 3.3. Cloned RLGS fragments altered in RLGS profiles of
16 primary tumors………………………………………………………… 75
Table 4.1. Summary of LOH of D10S196 and methylation
detected by MS-PCR………………………………………………… 105
Table 4.2. Alteration of gene expression due to expression of BMP3B
measured by cDNA microarrays…………………………………….. 106
Table 5.1. Primers used for probe preparations…………………………………. 130
Table 5.2. Enhanced RLGS fragments in primary lung cancers……………… 131
Table 5.3. Enhanced RLGS fragments in lung cancer cell lines.................. 132
Table 5.4. Properties of cloned RLGS fragments................................. 133
LIST OF FIGURES

Figure 1.1 The addition of a methyl group to cytosine in 5’-CpG-3’

is referred as DNA methylation........................................... 30

Figure 1.2 An artificial DNA sequence and the nucleotide changes

after bisulfite treatment and amplification by PCR...................... 30

Figure 2.1 Outline of the RLGS procedure using Ascl as a

restriction landmark enzyme........................................... 53

Figure 2.2 RLGS profiles using Ascl-EcoRV-HinfI restriction

enzyme combination..................................................... 54

Figure 2.3 RLGS identifies DNA methylation in primary lung cancer........... 55

Figure 3.1 Restriction Landmark Genomic Scanning (RLGS) and

Confirmation of methylation by Southern hybridization........... 76

Figure 3.2 Global methylation patterns in lung cancer are non-random........ 78

Figure 3.3 Abnormal transcription of BMP3B NSCLC samples.................... 79

Figure 4.1 BMP3B promoter methylation detected by bisulfite

sequencing and methylation-specific PCR.......................... 107

Figure 4.2 CTG repeat polymorphism in BMP3B exon 1...................... 109

Figure 4.3 Detection of reexpression of BMP3B in A549 cell line........... 110
Figure 4.4 *In vitro* and *in vivo* cell growth assay of A549 cell line after BMP3B reexpression…………………………………… 111

Figure 4.5 Proposed model of opposing role of TGF-β1 and BMP3B………… 113

Figure 5.1 RLGS sections with enhanced fragments
in primary lung cancer……………………………………………… 134

Figure 5.2 RLGS sections of lung cancer cell lines with enhanced DNA fragments……………………………………………… 135

Figure 5.3 Confirmation of DNA amplification by Southern hybridization…… 136

Figure 5.4 Diagram of all the known genes and ESTs
on chromosome 11q22……………………………………………… 137

Figure 5.5 11q22 amplicon definition by a series of Southern hybridization…. 138
CHAPTER 1

INTRODUCTION

1.1 Human lung cancer

Lung cancer is the leading cause of cancer-related deaths in the United States (1). In 2002, it was estimated that there would be approximately 169,400 new cases and 154,900 deaths attributed to lung cancer (1). The American Cancer Society estimated that lung cancer in the United States was the second most common malignancy in both males and females. In comparison to all cancer types, the incidences of lung cancer were only 14% and 12% of total estimated new cases in males and females respectively. However, only approximately 13% of lung cancer patients survive more than 5 years. Thus, the mortality rate of lung cancer is the highest for both males and females annually. It was estimated that lung cancer might contribute to 31% and 25% of the total cancer-related deaths in the United States in 2002 for males and females respectively. Therefore, lung cancer is a serious threat to human kind.

For clinical diagnosis and treatment, lung cancer is divided into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) (2, 3). Approximately 75% of lung tumors are NSCLC (3). NSCLC tend to metastasize later in development, while
SCLC spread very early. Therefore, NSCLC can be successfully treated by surgery in their early stages; while, SCLC patients are suitable for chemotherapy and radiation therapy. Early diagnosis is critical for improving the survival of lung cancer patients.

Histologically, the World Health Organization (WHO) classified NSCLC into three major subgroups: squamous cell carcinoma, adenocarcinoma, and large cell carcinoma (30%, 40% and 15% of all lung cancer cases in the United States respectively) (2). Some lung cancers are heterogeneous in morphology. For example, both squamous cell carcinoma and adenocarcinoma are present in approximately 10% of cases (2, 4) and are classified as adenosquamous carcinoma. Although the histologic origin of NSCLC and SCLC are not well characterized, Yesner claimed that different subtypes of lung cancer develop from the multipotent endodermal stem cells (5). He also pointed out that the differences between the subtypes were quantitative. That is, the various properties of cells are present in varying degrees in different subtypes of tumors. Yesner used the Y-construct model to explain the dynamic spectrum of lung cancer. Small cell lung cancers were placed at the bottom, and squamous cell carcinomas and adenocarcinomas at the two arms, while large cell carcinomas were placed at the bifurcation. Although large cell carcinomas appear undifferentiated, organelles of both squamous cell carcinomas and adenocarcinomas were frequently found in large cell carcinoma by electron microscopy (5).

Both environmental and genetic factors contribute to lung cancer development. The environmental factors that have been correlated with lung cancer include cigarette smoking and exposure to other carcinogens, such as asbestos and radon. Cigarette smoking has been extensively studied (2). Strong association between lung cancer and
cigarette smoking has been established based on epidemiological studies. Approximately 80 to 90% of lung cancers are attributed to cigarette smoking (2). It is well known that the incidence of squamous cell carcinoma and small cell carcinoma increases with continued exposure to smoke. Recently, Thun et al. reported that the increase of lung adenocarcinoma after 1950s was related to the changes in smoking behavior and cigarette design, such as deep inhalation and the design of filter-tip cigarettes (6).

1.2 Molecular genetic changes in human lung cancer

1.2.1 Genetic susceptibility to lung cancer

1.2.1.1 Inherited susceptibility

Although the majority of lung cancers are related to environmental factors, such as cigarette smoking, previous studies have shown increased risk (~2.4 fold) among families with a history of lung cancer (7). The risk of developing lung cancer is even higher among first degree relatives (~6.1 fold) and offspring (~7.2 fold) of younger (40-59 years old) non-smoking lung cancer patients (8). Using a mouse model, pulmonary adenoma susceptibility (Pas1) was mapped as a major locus affecting inherited predisposition to lung cancer in mice, which is on chromosome 6, close to Kras2 (9). The possibility that Kras2 is the gene responsible for pulmonary adenoma susceptibility in this region was supported by subsequent studies. Wild type Kras2 was demonstrated to have tumor suppressor function in mouse model (10).
1.2.1.2 Genetic polymorphism related to tobacco carcinogens

Although 80-90% lung cancer patients are smokers, it was estimated that approximately 10% of smokers are at risk of developing lung cancer (3). Genetic polymorphisms of enzymes involved in the metabolism and detoxification of smoking related carcinogens may result in different susceptibility for developing lung cancers in smokers. For example, CYP2D6 and CYP1A1, GSTM1 have been extensively studied, although conflicting results have been reported and reviewed (7, 11, 12). For example, an MspI restriction site polymorphism of CYP1A1 has been correlated with lung cancer incidence related to cigarette smoking in a Japanese population (13), but the role of the CYP1A1 genetic polymorphism as a lung cancer risk factor was not confirmed in other population (7).

1.2.2 Tumor suppressor genes

Molecular genetic studies have shown that sporadic lung cancers caused by environmental factors are polygenic diseases, that is, multiple genetic loci contribute to lung cancer development (7). Molecular abnormalities include both growth promoting oncogenes and growth suppressing tumor suppressor genes. The loss-of-function of tumor suppressor genes and gain-of-function in oncogenes contribute to tumorigenesis. The identification of both tumor suppressor genes and oncogenes will be of great significance in understanding the mechanism of carcinogenesis for better diagnosis and treatment of cancer patients.
1.2.2.1 Definition of tumor suppressor genes

The definition of tumor suppressor genes was discussed by Harber and Harlow (14). Genes in this category usually meet the following criteria: tumor suppressor genes acquire a loss of function mutation in both familial and sporadic tumors, and this defect is restored by the wild-type gene that suppresses tumor growth and the transformed phenotype. Haber and Harlow pointed out that the definition has its limitations, for example, genes involved in genomic integrity and DNA repair, such as MLH1, MSH2, do not fit the definition. Therefore, the functional criteria may not apply to the genes with various functions in different cellular pathways. Subsequently, tumor suppressor genes were more broadly defined as genes that undergo loss-of-function mutations during the development of cancer (14).

1.2.2.2 Mechanisms of loss of function of tumor suppressor genes

Genetic alternations, which may cause the loss of the function of tumor suppressor genes, usually fit for Knudson’s “two-hit” model. That is, both alleles must be mutated for gene inactivation. For example, in inherited familial cancers, one allele is inactivated by germline mutation and the second allele becomes inactivated by acquired somatic mutation, such as the RB gene in retinoblastoma. Sporadic cancers acquire somatic mutations of both alleles (15). The genetic alterations may be point mutations, loss of heterozygosity (LOH) or homozygous deletions. Classical studies have shown that regions of LOH harbor tumor suppressor genes (3). Tumor suppressor genes, such as p53, Rb and p16, are all located in regions of LOH (3).

Epigenetic changes, as DNA methylation of the gene promoter region, also
suppress the transcription of tumor suppressor genes. In addition, histone deacetylation, another epigenetic change, modifies the chromatin structure and regulates gene expression. These will be further discussed in chapter 1.3 and 1.4.

1.2.2.3 Tumor suppressor genes in lung cancer: p53, p16/CDKN2A and Rb

p53, p16/CDKN2A and RB are shared targets for inactivation in different types of cancers and are the best studied tumor suppressor genes in lung cancer. p53 mutations were observed in about 50% of NSCLC and 90% of SCLC tumors (3, 16, 17). p16/CDKN2A is inactivated more frequently by homozygous deletions or hypermethylation in NSCLC cell lines which retain wild type RB (18, 19). The RB protein is abnormal in approximately 90% SCLC and 15-30% NSCLC due to mutation, deletion or altered protein expression (7, 20, 21).

1.2.2.4 Examples of newly identified candidate tumor suppressor genes in lung cancer: FHIT, RASSF1A, SEME3B, Kras2

Currently, great efforts are devoted to identify novel candidate tumor suppressor genes. For example, the short arm of chromosome 3, which showed high frequency of loss of heterozygosity (LOH), is thought to harbor several tumor suppressor genes (3, 7, 20). Both the FHIT gene on chromosome 3q14.2, and the RASSF1A and SEME3B genes on 3p21.3, have been described as targets for inactivation in these regions. In addition, wild type Kras2 that has been classified as an oncogene may have tumor suppressor function (10).

**FHIT:** The FHIT (fragile histidine triad) gene, located on 3p14.2, was cloned in
1996 by positional cloning (22). Among lung cancer patients, a much higher frequency (80%, 41/51) of LOH of the FHIT gene was reported in smokers than in nonsmokers (22%, 9/40) (23). Abnormal transcription of the FHIT gene was reported in 40% of NSCLC (24). Transcriptional inactivation of FHIT by methylation of the promoter was demonstrated in esophageal squamous cell carcinoma (25). Recently, it was reported that reexpression of FHIT using adenovirus vector could induce apoptosis and suppress cell growth in nude mice (26).

**RASSF1**

RASSF1 is located on chromosome 3p21.3, a region where LOH has been demonstrated in more than 90% of small-cell lung cancer and in 50-80% of NSCLC. The region has been thoroughly studied in order to identify candidate tumor suppressor genes. Dammann et al. reported that RASSF1, located in the 120-kb minimally deleted region of chromosome 3p21.3, is transcribed in three transcripts, A, B, and C that are derived by alternative splicing (27). The transcripts A and C were detectable in all normal tissues tested. However, only RASSF1A was silenced at the transcriptional level due to promoter methylation. Upon reexpression in human lung cancer cell lines, cell growth was inhibited in both *in vitro* and *in vivo* assays (27). Similar results were reported by Burbee et al. (28).

**SEMA3B:**

Semaphorin 3B is located in the same chromosomal region as RASSF1A. The gene was identified as directly inducible target of p53 (29). Expression of the gene was down regulated in lung cancer. Reexpression of the gene could suppress the growth of HEY cells, an ovarian cancer cell line (30).

**Kras2:**

Although the Ras gene is a well-known oncogene, Zhang et al. reported that heterozygous deletion of Kras2 increases the susceptibility of lung cancer to
chemical induction. In addition, reexpression of wild type Kras2 inhibits tumor formation in transformed NIH/3T3 cells and a mouse lung cell line (10) indicating that wild type Kras2 has tumor suppressor function.

1.2.3 Oncogenes

1.2.3.1. Known oncogenes involved in lung cancer.

Oncogenes are a group of genes which acquire gain-of-function mutations during tumorigenesis. The activation of oncogenes promotes tumor progression. Oncogenes have various cellular functions and include growth factors, growth factor receptors, genes that function in intracellular signal transduction, transcription factors, and positive regulators of the cell cycle. Oncogenes can be activated by point mutations, translocations, DNA amplification and other mechanisms that result in overexpression of proteins.

In NSCLC, the proto-oncogene KRAS, a component in signal transduction pathways, is mutated in approximately 30% of cases. The mutations happen predominantly with a G to T change in codon 12, which render KRAS gene transformation potential (7, 20, 31). Amplification of MYC transcription factors (MYCC, MYCN and MYCL1) is well established in human lung cancer. MYCC is amplified in both SCLC and NSCLC, while amplification of MYCN and MYCL1 are usually limited to SCLC (32). MYCC amplification in small cell lung cancer could be used as prognostic factor (33, 34). Cyclin D1 and the epidermal growth factor
receptor (EGFR) are amplified in 5% and 6% of NSCLC, respectively (35). C-erbB-2 (Her-2/neu) or BCL-2 over-expression are involved in approximately 25% of cases (20).

1.2.3.2 Mechanisms of amplification

Amplification is one of the most important genetic changes that lead to the activation of oncogenes, but the mechanism of amplification is unknown. Different models have been proposed to explain DNA amplification in cancers, such as the breakage-fusion-bridge (BFB) model (36, 37). In brief, the chromosomes break at common fragile sites with induction by hypoxia or drugs. Following replication, the uncapped sister chromatids become fused during metaphase. During anaphase, a bridge is formed between opposite poles. The breakage of the chromosome results in one daughter cell containing three copies of selected genes. As a result, the cells might gain growth advantage. The same process is repeated multiple times, leading to gene amplification (36-38). The amplified DNA sequences present in the chromosome appear as homogenously staining regions (HSRs), or extra, small, separated DNA materials, such as double minutes (DMs).

1.2.3.3 Strategies to identify novel amplicons

Since DNA amplification is one of the most important mechanisms leading to oncogene activation, identification of novel amplicons is of great significance in the search for novel oncogenes. Comparative genomic hybridization (CGH) was introduced as a powerful cytogenetic technique to study genome-wide genetic changes, such as deletions and amplifications in the tumor (39). CGH is based on the hybridization of
differentially labeled fluorescent tumor DNA and normal DNA to normal metaphase spreads. The ratio change of the fluorescence intensity between tumor and normal DNA is quantified to identify deletions or amplifications (39). The resolution of this technique is limited to metaphase chromosomes (40). CGH is suitable to detect genetic changes greater than 5-10 Mbp or 5-10 fold amplification with a size larger than 1 Mbp. In addition, other techniques, such as fluorescent in situ hybridization (FISH), are required to further narrow down the regions identified by CGH. In the last decade, CGH has been widely used to identify novel amplicons in cancer, including non-small cell lung cancer (41).

Recently, the combination of microarray techniques with CGH, called array-CGH, has been performed to identify amplicons in human lung cancer (42). Array CGH directly identifies genetic alterations in cancers without many limitations of CGH.

Restriction landmark genomic scanning is an alternative strategy to screen novel amplification. The methodology and application will be discussed in chapter 1.6.2.1.

1.3 Gene inactivation by epigenetic changes: DNA methylation and histone deacetylation, histone methylation

1.3.1 Introduction of epigenetic changes

Besides genetic inactivation of tumor suppressor genes, epigenetic changes, such as DNA methylation and histone deacetylation, play crucial roles in tumorigenesis (43). Epigenetics study the heritable modifications that regulate gene expression without the
changing the DNA sequence (44). Tremendous progress has been made to understand the significance of epigenetic regulation of gene expression in the last decade. It was estimated that silencing of tumor suppressor genes by aberrant DNA promoter methylation is at least as common as genetic mutations. In addition, approximately 50% of tumor suppressor genes related to familial cancers have been shown to be silenced by methylation in sporadic cancers (43). The status of histone acetylation or histone deacetylation affects chromatin structure and plays a key role in the regulation of gene expression (45). Recent evidence has connected these two epigenetic processes together for the regulation of gene expression (46, 47).

1.3.2 DNA methylation

DNA methylation refers to the addition of a methyl (CH$_3$) group to the cytosine ring in 5’-CpG-3’ (figure 1.1) which is maintained by DNA methyltransferase 1 (DNMT1) after replication (48). DNMT1 has de novo methyltransferase function, but the activity is much higher on hemimethylated DNA, so the main function of DNMT1 is to maintain DNA methylation patterns (49). The other two DNA methyltransferases, DNMT3A and DNMT3B, are de novo methyltransferases and highly expressed in embryonic stem cells (49). DNA methylation is an epigenetic change, since it is heritable and affects chromatin organization and gene expression without changing of the genetic code (50). DNA methylation in mammals is involved in several normal processes including genomic imprinting (51, 52), X chromosome inactivation (53) and aging (54).

Genome-wide evaluation shows that the majority (70-80%) of single CpG dinucleotides are methylated. These methylated regions occur mainly in intronic
DNA, and in the bodies of many genes and repetitive sequences (55-57). In contrast, CpG dinucleotides in CpG islands are usually unmethylated. CpG islands range in size from 0.2 kb to about 4 to 5 kbs and are defined as a stretches of DNA which include a G+C content of greater than 50% with a ratio of observed to expected CpG \[ \frac{number\ of\ CpG}{total\ number\ of\ nucleotides} \] of at least 0.6 (58). It was estimated that there are 29,000 CpG islands in the human genome based on human draft sequences (59). Approximately 50 to 60% of genes are associated with CpG islands. Methylation of CpG islands in gene promoter regions usually is associated with gene silencing (55, 60).

The mechanism for transcriptional repression by DNA methylation is not clear. It is postulated that DNA methylation is involved in different ways to inactivate transcription. First, DNA methylation interferes with the binding of transcription factors, such as (AP-2, NFκB) that contain CpGs in their transcription factor binding sites. Second, transcriptional repressors, such as methyl-CpG-binding domain protein 2 (MBD2) and methyl-CpG-binding protein 2 (MeCP2) bind methylated CpGs and recruit histone deacetylase (HDAC) (61, 62). For example, MBD2 is associated with HDAC1 in the MeCP1 repressor complex (63), while MeCP2 recruits the Sin3-HDAC corepressor complex (64). The recruitment of HDACs results in a hypoacetylation of histone H3 and H4 tails. Finally, the chromatin structure is modified to a condensed form and subsequently inhibits transcription (62). Therefore, methylation of the gene promoter region causes gene inactivation (43, 65). This will be further discussed in chapter 1.3.3 and 1.3.4.
1.3.3 Histone acetylation and histone methylation

In the nucleus, the DNA double helix wraps around histones and forms the nucleosomal structure (66). Histone proteins include H2A, H2B, H3 and H4 and the linker histone H1. The nucleosome fibers in turn are assembled into a higher order structure, called chromatin (67). A portion of the nuclear chromatin called heterochromatin is highly condensed throughout the cell cycle and is transcriptionally inactive. While the remainder of the nuclear chromatin called euchromatin contains transcriptionally active DNA and has a relatively extended conformation is called euchromatin (68). The covalent modifications of histones, such as acetylation, phosphorylation and methylation, alter chromatin structure, thus leading to the inherited transcriptional difference. This information was proposed as “histone code” that extends the information from the genetic code (67).

Histone acetylation and deacetylation are well-studied histone modifications. The deacetylated histones have positive charged lysine residue tails and bind tightly to negatively charged DNA. In this situation, the DNA is not accessible to the transcriptional machinery. However, if the lysine tail is acetylated by histone acetyltransferase (HAT), the bonding between the DNA and the lysine is weak, and the DNA can be easily transcribed. The process is reversible, allowing for the acetyl group to be removed by histone deacetylases to shut off the gene expression (66).

Histone methylation also participates in gene regulation. It has been shown that the methylation of lysine 9 in the tail of histone H3 correlates with transcriptionally repressed chromatin, while methylated lysine 4 on H3 resides in transcriptionally
active chromatin flanked by heterochromatin. Recently, it was found that in *Neurospora crassa* and *Arabidopsis thaliana*, Lys9 histone methyltransferase also control DNA methylation (69, 70). Jones and Baylin pointed out in a recent review that these are extremely important findings, since the methylation of H3 Lys9 might be required to determine the sites of DNA methylation (43).

### 1.3.4 Connection between DNA methylation and histone modification for gene transcription regulation.

In addition to the previous finding that Lys9 histone methyltransferase controls DNA methylation in some species (69, 70), DNA methylation has been related to histone acetylation. First, as mentioned above, methyl-CpG-binding proteins, such as MeCP1 and MeCP2 bind to methylated DNA and recruit histone deacetylase to remodel the chromatin structure (43, 65). Second, tremendous evidences have shown that DNA methyltransferases (DNMT1, DNMT3A and DNMT3B) directly associated with histone deacetylase (HDAC). For example, DNMT1 binds to HDAC1 and HDAC2 to recruit the histone deacetylase activity (71, 72). Similarly, DNMT3a can directly recruit HDAC1 (73). Third, the acetylation of histones was found to be inversely associated with DNA methylation, such as in death-associated protein kinase (DAPK) (74). The acetylation status of histones H3 and H4 related to specific genes can be detected by chromatin immunoprecipitation assay (74). Fourth, the DNA methyltransferase inhibitor, 5-aza-2'-deoxycytidine, and the histone deacetylase inhibitor, trichostatin A (TSA), have synergistic effects to induce the reexpression of silenced genes in cancer cells (75). Therefore, it was proposed that promoter methylation, together with histone
deactylation, maintains condensed heterochromatin, causing the gene to become inaccessible to the transcriptional machinery, resulting in gene silencing (43).

1.4 DNA methylation in cancer

It has been recognized for more than ten years that epigenetic changes in neoplasia include overall genomic hypomethylation, with concurrent hypermethylation in CpG islands which are usually unmethylated in normal tissues (55). The role of methylation in oncogenesis has been explained by one or more mechanisms. First, tumor suppressor genes can be inactivated by hypermethylation of CpG islands in promoter regions. Examples include RB, VHL, p16, p15, E-cadherin and hMLH1 (55). Second, methylated CpG dinucleotides determine hot spots for mutations in the p53 tumor suppressor gene (60, 76). Other proposed mechanisms of DNA methylation include increased chromosomal instability (60) and activation of oncogenes (77) due to hypomethylation.

Jones and Baylin reviewed evidence that demonstrates the silencing of tumor suppressor genes by promoter methylation plays an important role in tumor development (43). First, transcriptional silencing of cancer related genes by aberrant CpG island methylation is observed in every type of cancer (with the exception of seminomas (78)). The list of candidate tumor suppressor genes silenced by promoter methylation continues to grow; see also (65). Second, nearly 50% of the tumor suppressor genes associated with familial cancers that becomes inactivated by mutations, can be silenced by promoter methylation in sporadic cases. Examples of such genes include BRCA1, VHL, MLH1, APC, p16, RB. Third, promoter methylation is regarded as the one “hit” or
both “hits” needed to inactivate suppress tumor suppressor genes in Knudson’s two hit model. The methylation can be the second hit when combined with genetic changes, such as LOH or mutation of the other allele. Fourth, cancers with epigenetic inactivation of important genes, such as MLH1 and MGMT DNA repair genes, predispose the genome to genetic changes. Fifth, tumor suppressor genes that are affected by both epigenetic change and/or genetic changes are usually in a region of LOH (43).

1.5 DNA methylation in lung cancer

1.5.1 Methylated genes in lung cancer

Promoter hypermethylation of one or both alleles is one of the mechanisms leading to aberrant gene function in lung cancer development. More than 30 genes have been reported to show increased promoter methylation in human lung cancer (79). These genes have been shown to be involved in a variety of cellular functions. The genes that have been demonstrated to be silenced by promoter methylation in lung cancer are listed in table 1.1 (65, 79). Since the genes affected by methylation have such important functions in cell cycle control regulation, differentiation, DNA repair, metastasis and cell adhesion, silencing of these genes play critical roles in lung cancer development and progression. Different methylation sensitive strategies have been utilized to detect the methylation, which results in variances in the frequency of methylation in individual genes from different studies (79).

p16/CDKN2A is a well-studied tumor suppressor gene exhibiting promoter methylation in lung cancer (18, 19, 80, 81). p16/CDKN2A abnormalities, frequently
observed in primary NSCLC and cell lines, include homozygous deletions and promoter methylation of both alleles, but are rare in SCLC. In some studies, p16 promoter methylation shows no difference among histological subtypes or tumor stages (81). However, in other studies, p16 has been shown to play a very early role in lung cancer development, and the methylation status increases during tumor progression in both smoking related carcinogen NNK, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butaneone derived from nicotine during smoking, induced lung cancer in rats, and in squamous cell carcinoma of the human lung (82).

In addition, methyl-CpGs determine hot spots of mutation in human lung cancer. For example, guanines in methyl-CpG are believed to be vulnerable to the smoking-related chemical carcinogen, BPDE [(+/-) anti-7β, 8α-dihydroxy-9α, 10α-epoxy-7, 8, 9, 10-tetrahydrobenzo[α]pyrene)]. These sites are the most common point mutations (G to T) of p53 in lung cancer (76).

1.5.2 DNA methylation as biomarker for early diagnosis in human lung cancer.

Tsou et al. discussed the feasibility of using DNA methylation as a possible biomarker for early diagnosis of lung cancer (79). They summarized that promoter methylation of p16/CDKN2A, MGMT, DAPK, GSTP1 and APC could be detected in lung cancer patients in remote places, such as sputum, serum or bronchoalveolar lavage samples. In addition, methylation of p16/CDKN2A and MGMT was detectable in high-risk patients prior to the diagnosis of lung cancer, but the usefulness of these markers as
predictors of lung cancer development from the sputum of high-risk but cancer-free patients is controversial. The great potential of targets of DNA methylation as biomarkers, such as aberrant p16 promoter methylation that claimed as a new biomarker for the early detection of lung cancer (82), needs to be further evaluated for the sensitivity and specificity.

1.6 Techniques for the study of aberrant methylation

1.6.1 General introduction

Traditional strategies to detect DNA methylation rely on methylation sensitive restriction enzymes. These enzymes only digest unmethylated sequences, and leave methylated sequenced uncut. Therefore, the methylated and unmethylated DNA can be distinguished. The disadvantage of methylation sensitive restriction enzymes is that they only detect the methylation in the restriction sites. Bisulfite treatment of genomic DNA was a technique developed in the last decade that makes it possible to detect the methylation status of any CpG dinucleotide (83, 84). The methyl-cytosine is differentiated from cytosine by bisulfite treatment, which converts unmethylated cytosine (C) to uracil (U), while methylated C remains intact (83). After bisulfite treatment, several strategies have been developed for the detection of cytosine methylation. So far all of these techniques are based on PCR amplification following bisulfite treatment. The techniques that are currently used to detect DNA methylation are listed in table 1.2.
1.6.2 Methods based on DNA methylation sensitive restriction enzymes

Southern hybridization can detect methylation of known sequences. Although Southern blot can only detect a limited number of CpGs in the restriction sites of the methylation sensitive enzymes, it remains the gold standard because it avoids problems encountered by PCR amplification of CG rich region. In addition, the degree of partial methylation can be quantified. The genomic DNA is digested by different methylation sensitive restriction enzymes, such as NotI (GC↓GGCCGC), AscI (GG↓CGCGCC), HhaI (GCG↓C) and EagI (C↓GGCCG), and then separated on an agarose gel that is then transferred to a membrane. More than 50 methylation sensitive enzymes are commercially available.

Several strategies have been designed to screen unknown aberrantly methylated sequences in the human genome following digestions with methylation sensitive restriction enzymes. For example, restriction landmark genomic scanning is a two dimensional DNA electrophoresis technique that can detect more than 2000 CpG islands simultaneously.

1.6.2.1 Restriction Landmark Genomic Scanning (RLGS)

RLGS is based on the methylation sensitive NotI and AscI restriction enzymes, which can only digest unmethylated genomic DNA, and is a powerful tool for genome-wide scans for DNA methylation changes. Costello et al. described the protocol in detail (85) and the procedure is outlined in Figure 2.1. RLGS does not rely on prior knowledge of a gene sequence and has been used to clone novel imprinted genes (86) and amplified genes (87-90), as well as for genome wide scans for aberrant DNA methylation
patterns in different human cancers (89, 91, 92).

RLGS is very efficient in cloning genes that are affected by methylation during oncogenesis (91). Since approximately 90% of the NotI restriction enzyme sites (GC\(\downarrow\)GGCCGC) are located in CpG islands, which are linked to genes (91, 93), the methylation changes identified by RLGS are closely related to the coding regions of the genome enabling the screening of the methylation status of gene-related genomic regions.

Because there is no sequence information known regarding the RLGS fragments of interest, the cloning of RLGS fragments is critical. Using a NotI/EcoRV boundary plasmid library constructed in our lab (94), and the RLGS mixing gels from different plates, rows or columns, the individual fragments can be very efficiently correlated with a single plasmid clone in the library (89). These clones are then sequenced and used for database searches to identify corresponding genes. The correlation between DNA methylation and aberrant transcription will be studied in more detail.

Since DNA methylation is one of the most important mechanisms to silence tumor suppressor genes, and the methylated targets are distributed through the human genome, the screening for the novel methylated events may identify candidate tumor suppressor genes (43, 93). This has been demonstrated in mouse models and human hepatocellular carcinomas. Akama et al. found that p16 tumor suppressor gene and \(\alpha4\)-integrin were targets of methylation in the mouse liver tumor induced by SV40 T antigen using RLGS (95). Yoshikawa et al. reported SOCS-1 (suppressor of cytokine signaling 1), as a target of methylation and identified by RLGS in human hepatocellular carcinoma (HCC), was transcriptionally silenced by methylation in HCC. Restoration of the gene inhibits cell growth (96). In addition, global methylation phenotypes have been
evaluated in different kinds of cancers and the pattern of methylation in these tumors is not random (91, 92, 97-99). Moreover, methylated genetic loci identified by RLGS may be used as prognostic factors (97, 100). The study of DNA methylation in cancer has been excellently reviewed (93, 101).

Besides screening of DNA methylation, RLGS is quantitative and can distinguish changes of CpG island copy number without the difficulties of PCR amplification of GC rich regions. Thus, enhanced RLGS fragments from regions of genomic amplification can be readily identified on an RLGS profile. RLGS has been successfully used to survey genomic amplifications in variety of human primary malignancies (87, 88, 102, 103), human cancer cell lines (104-106), as well as mouse tumors (107, 108).

1.6.2.2 Other techniques based on methylation sensitive restriction enzymes: MCA, MS-AP-PCR, DMH

Several novel techniques based on the methylation sensitive restriction enzymes have been developed in the last several years, and include methylated CpG island amplification (MCA) (109), methylation-sensitive arbitrarily primed PCR (MS-AP-PCR) (110, 111) and differential methylation hybridization (DMH) (112, 113). But PCR amplification is necessary for these new techniques, which may have some difficulties to amplify CG rich regions.

Methylated CpG island amplification (MCA) utilizes isoshizomeric restriction enzymes, one methylation sensitive (such as SmaI) and another one methylation insensitive (XmaI). After the first digestion with SmaI, unmethylated DNA is digested and leaves blunt ends. The methylated DNA is then digested with XmaI and produces
sticky end. Therefore, only methylated DNA can be ligated with adaptor and amplified by PCR for further analysis (109).

**Methylation-sensitive arbitrarily primed PCR (MS-AP-PCR)** is based on the digestion of genomic DNA with methylation sensitive enzyme (such as \textit{Hpa}II), and then arbitrarily amplified using random GC rich primers with low annealing temperatures. Similarly, the same DNA is digested with isoshizomer \textit{Msp}I, which digest both methylated and unmethylated DNA, and then PCR amplified by same primers. After separation on high-resolution polyacrylamide gels, a band present in \textit{Hpa}II digested DNA but missing from the \textit{Msp}I digested DNA indicates the DNA fragment containing methylated \textit{Hpa}II site(s) (110, 111).

**Differential methylation hybridization (DMH)** is based on the hybridization of normal and tumor DNA to an array with more than 8000 CpG island fragments that are spotted onto a nylon membranes or glass slides. The normal and tumor DNAs are digested with restriction enzyme and ligated with linker, after methylation sensitive restriction enzyme \textit{BstUI} digestion, the DNAs are amplified. So the sequences with \textit{BstUI} site(s) methylated or unmetylated are differentially amplified. After labeled with two different fluorescent dyes (Cy5 and Cy3), the normal and tumor DNAs are hybridized to the arrays to detect DNA methylation (113).

**1.6.3 Methods based on bisulfite treatment**

**Bisulfite sequencing** is an approach where bisulfite treatment is followed by sequencing the PCR product (figure 1.2). Specific PCR primers are designed that do not include CpGs. Theoretically, both methylated and unmethylated sequences
are amplified simultaneously. The PCR products are directly sequenced using PCR primers, or cloned into a plasmid vector. Individual clones with the correct insert size are sequenced.

**Methylation sensitive PCR** (MS-PCR) is based on designing primer pairs that distinguish methylated and unmethylated sequences following bisulfite conversion (114). Two sets of primers are designed; one pair of primers amplify only methylated sequences which retain cytosines in CpGs after bisulfite treatment, while the other pair amplify only unmethylated sequences in which cytosines in CpGs have been converted to uracils. If the sample contains both methylated and unmethylated molecules, both primers will amplify a product. Modified MS-PCR, such as real-time MS-PCR (ConLight-MSP) has also been designed in ordered to detect unconverted DNA during bisulfite treatment (115).

Other techniques based on bisulfite treatment, such as **combined bisulfite restriction analysis (COBRA)**, **methylation-sensitive single-nucleotide primer extension (Ms-SNuPE)**, **methylation-sensitive single-strand conformational polymorphism (MS-SSCP)** and **MS-HPLC**, are different strategies that have been performed for methylation analysis following bisulfite treatment. Theoretically, primers are designed to amplify both methylated and unmethylated sequences at same efficiency.

Combined bisulfite restriction analysis (**COBRA**) utilizes a restriction enzyme to digest the PCR product following bisulfite treated DNA. The restriction enzymes only cut sequences that have retained the recognition sequence following conversion of bisulfite treated DNA (116).

Methylation-sensitive single-nucleotide primer extension (**Ms-SNuPE**) extends
the PCR product a single nucleotide to detect if the cytosine at a particular CpG site is converted to uracil or remains intact as cytosine. Oligonucleotide(s) probes was designed immediately upstream of the CpG site(s) as primers for extension (117, 118).

Methylation-sensitive single-strand conformational polymorphism (MS-SSCP) is based on the differential band mobility of unmethylated and methylated DNA from same DNA sequence following bisulfite treatment and PCR amplification. The methylated DNA and unmethylated DNA can be distinguished using nondenaturing gel analysis (119). Similarly, the PCR product can be analyzed with HPLC to discriminate the methylated and unmethylated alleles (120).

1.6.4 Other techniques used in screens for methylated genes

For the screening of novel aberrantly methylated sequences, there are other techniques, not based on restriction enzymes digestion, such as using microarray to detect gene reexpression following treatment with inhibitors of DNA methylation and histone deacetylation to identify genes silenced by methylation (121). But the original methylation status of the DNA must be analyzed by using methylation sensitive enzymes or bisulfite treatment. Another technique employs a methylated DNA binding column (122, 123), which “is an affinity matrix that contains a polypeptide derived from the methyl-CpG binding domain (MBD) of the rat chromosomal protein MeCP2, which preferentially binds DNA at symmetrical mCpG sites”. Methylated and unmethylated DNAs are separated based on their different binding affinities on the column.
1.7 Hypothesis and summary of the study

Current knowledge demonstrates that epigenetic changes are closely related to genetic changes, and aberrant DNA methylation of tumor suppressor genes is one of the mechanisms involved in lung cancer development. Further studies of epigenetic changes are crucial to fully understanding the molecular mechanisms of oncogenesis. In addition, amplification of oncogenes plays an important role in lung cancer. Since the major cause of lung cancer is cigarette smoking and multiple genomic loci are involved, we hypothesize that utilizing this genome-wide approach to identify methylation changes, as well as DNA amplifications, can lead to the discovery of aberrant DNA methylation patterns in lung cancer, and the identification of novel genes that are silenced, or activated, in lung cancer. In this study, RLGS was performed on paired lung cancer and normal lung from the same patient, as well as lung cancer cell lines, to screen for genome wide methylation and amplification changes.

The introductory chapter briefly reviews human lung cancer, molecular genetic changes in lung cancer. In addition, the significance of DNA methylation as an epigenetic change that has been connected to histone modification and chromatin remodeling, is discussed as one of the mechanisms that lead to cancer development, such as in lung cancer. The techniques used to detect methylation in known genes and to screen for novel methylated targets are introduced. Chapter 2 describes the establishment of Ascl as the restriction landmark site for restriction landmark genomic scanning (RLGS), allowing for the screening of as many as twice the number of CpG islands in the human genome than when using NotI only. The ability of Ascl RLGS to detect methylated fragments was demonstrated. Moreover, an Ascl-EcoRV boundary plasmid clone library was
created, and the library RLGS mixing gels were prepared. We successfully cloned 178 RLGS fragments. From these cloned fragments, 119 of them are methylated targets in lung cancers, head and neck cancers, medulloblastomas and/or leukemia. **Chapter 3** reports the global methylation analysis of primary lung cancers, and the cloning of novel methylated genes using the NotI-EcoRV library that has previously been established. The overall methylation level showed heterogeneity among patients. Some RLGS fragments are frequent targets of methylation in lung cancer. Among the cloned RLGS fragments, BMP3B showed promoter CpG island methylation and was down regulated in both primary lung cancer and lung cancer cell lines. After inhibition of DNA methylation in NSCLC cell lines by 5-aza-2'-deoxycytidine, the gene was reactivated. **Chapter 4** describes the detailed characterization of BMP3B as a cancer related genes. In a large set of NSCLC, we found that BMP3B showed increased methylation in approximately 45% primary lung cancers and the marker D10S196 closest to BMP3B among the markers that we used showed 35% LOH. We reexpressed BMP3B in lung cancer cell line A549 to test the significance of BMP3B silencing in lung cancer. Although the morphology and cell growth rate was not obviously affected, the colony formation ability was suppressed *in vitro* as seen in anchorage dependent colony formation assays, and anchorage independent colony formation assays. Similarly, our preliminary data also indicated that the *in vivo* tumor growth was also suppressed. Preliminary findings from microarray analysis suggest that BMP3B might suppress genes that become up regulated by TGF-β, including fibronectin variant 1. **In chapter 5,** the amplified RLGS fragments are summarized and four amplicons in primary lung cancer have been cloned. Two of the amplicons, 3q26-27 and MYCL, have been reported previously in lung cancer.
Additional amplicons, including 11q22 and 6q21 have not been well characterized in lung cancer. The 11q22 amplicon, also amplified in other types of cancer, was further characterized by a series Southern hybridizations to define the amplified region.

In conclusion, the study identified both novel methylated and amplified genes in human lung cancer using RLGS. The biological significance of the genetic and epigenetic changes is being further investigated. These findings will lead to a better understanding of the mechanism of lung cancer development. Ultimately, this understanding will lead to an early diagnosis and treatment of human lung cancer, which is discussed in chapter 6, future directions.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retinoic acid receptor β (RARb)</td>
<td>Differentiation</td>
</tr>
<tr>
<td>RAS association domain family member 1 (RASSF1A)</td>
<td>Signal transduction</td>
</tr>
<tr>
<td>Adenomatous polyposis of the colon (APC)</td>
<td>Signal transduction</td>
</tr>
<tr>
<td>Tissue inhibitor of metalloproteins 3 (TIMP3)</td>
<td>Metastasis</td>
</tr>
<tr>
<td>p16/CDKN2A</td>
<td>Cell cycle</td>
</tr>
<tr>
<td>O6-methylguanine-DNA methyltransferase (MGMT)</td>
<td>DNA repair</td>
</tr>
<tr>
<td>Retinoblastoma protein interacting zinc finger (RIZ1)</td>
<td>Histone/protein methyltransferase</td>
</tr>
<tr>
<td>S100 calcium-binding protein A2 (S100A2)</td>
<td>Candidate tumor suppressor gene</td>
</tr>
<tr>
<td>H-cadherin (CDH13)</td>
<td>Cell adhesion</td>
</tr>
<tr>
<td>Tumor suppressor in lung cancer 1 (TSLC1)</td>
<td>Cell adhesion</td>
</tr>
<tr>
<td>Serum deprivation response factor (SRBC)</td>
<td>Candidate tumor suppressor gene</td>
</tr>
<tr>
<td>FHIT1</td>
<td>Others</td>
</tr>
</tbody>
</table>

**Table 1.1 Genes that have been shown promoter methylation in lung cancer**
<table>
<thead>
<tr>
<th>Method</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Based on methylation sensitive restriction enzyme</strong></td>
<td></td>
</tr>
<tr>
<td>Southern hybridization</td>
<td>Detection of methylation in known sequences</td>
</tr>
<tr>
<td>Restriction landmark genomic scanning</td>
<td>Screening for novel methylated genes</td>
</tr>
<tr>
<td>Methylated CpG island amplification (MCA)</td>
<td>Screening for novel methylated genes</td>
</tr>
<tr>
<td>Methylation-sensitive arbitrarily primed PCR (MS-AP-PCR)</td>
<td>Screening for novel methylated genes</td>
</tr>
<tr>
<td>Differential methylation hybridization (DMH)</td>
<td>Screening for novel methylated genes</td>
</tr>
<tr>
<td><strong>Based on bisulfite treatment</strong></td>
<td></td>
</tr>
<tr>
<td>Bisulfite sequencing</td>
<td>Detection of methylation in known sequences</td>
</tr>
<tr>
<td>Methylation sensitive PCR (MS-PCR)</td>
<td>Detection of methylation in known sequences</td>
</tr>
<tr>
<td>Combined bisulfite restriction analysis (COBRA)</td>
<td>Detection of methylation in known sequences</td>
</tr>
<tr>
<td>Methylation-sensitive single-nucleotide primer extension (Ms-SNuPE)</td>
<td>Detection of methylation in known sequences</td>
</tr>
<tr>
<td>Methylation-sensitive single-strand conformational polymorphism (MS-SSCP)</td>
<td>Detection of methylation in known sequences</td>
</tr>
<tr>
<td><strong>Others</strong></td>
<td></td>
</tr>
<tr>
<td>Expression analysis after demethylation</td>
<td>Screening for novel methylated genes</td>
</tr>
<tr>
<td>methylated DNA binding column</td>
<td>Screening for novel methylated genes</td>
</tr>
</tbody>
</table>

**Table 1.2, techniques to detect DNA methylation**
Figure 1.1 The addition of a methyl group to cytosine in 5’-CpG-3’ is referred as DNA methylation.

Figure 1.2 An artificial DNA sequence and the nucleotide changes after bisulfite treatment and amplification by PCR.
CHAPTER 2

AN *Asc*I BOUNDARY LIBRARY FOR THE STUDIES OF GENETIC AND EPIGENETIC ALTERATIONS IN CpG ISLANDS


2.1 Introduction

Multiple genome scanning approaches have been developed in the past years to study genetic and epigenetic alterations in cancer (124). The majority of those techniques target genetic alterations such as deletions, insertions and copy number changes. Restriction Landmark Genomic Scanning (RLGS), a highly reproducible two-dimensional gel electrophoresis, allows scanning of genomes for DNA polymorphisms, DNA amplification, and DNA methylation (86, 91, 125, 126). The use of RLGS to
study human cancers resulted in the identification of several novel genes that were amplified and overexpressed in malignant tissues (87, 88). Furthermore, the use of methylation sensitive restriction enzymes as landmark enzymes makes scanning of genomes for changes in the DNA methylation patterns possible (97, 98, 101, 127). This is of particular interest in cancer genetics, since promoter methylation has been shown to be involved in the silencing of tumor suppressor genes (65, 128, 129). The methylation sensitive restriction landmark enzyme \textit{NotI} has a GC rich recognition sequence, which is preferentially located in CpG islands sequences, found mainly in promoter regions of genes (91). In normal tissue DNAs, these sites are unmethylated (130). However, in tumors, methylation of a \textit{NotI} site results in the absence of an RLGS fragment in the respective profile.

Although RLGS profiles can be generated from any high quality genomic DNA without prior sequence information, subsequent cloning of RLGS fragments is essential for future studies. Several PCR based protocols have been developed allowing the identification of RLGS sequences (131). More efficient however, is a cloning strategy that uses an arrayed human library of \textit{NotI-EcoRV} clones and RLGS mixing gel catalogues (89). This protocol circumvents the need for PCR based amplification, which could be problematic with GC rich sequences. Successful use of this library system resulted in the identification of many methylation targets in several human tumors (93).

The use of the \textit{NotI-EcoRV} boundary library as a cloning tool for RLGS is restricted to RLGS profiles that use the enzyme combination \textit{NotI} and \textit{EcoRV} as the first and second restriction enzymes. In order to increase the potential coverage of CpG islands, we developed reaction conditions for the use of \textit{AseI} as the restriction landmark
enzyme in RLGS. In addition, we prepared an \textit{AscI-EcoRV} library and RLGS mixing gels that allow the efficient recovery of cloned RLGS fragments. We estimate that this novel resource, together with the \textit{NotI-EcoRV} library, will greatly increase the utility of RLGS and, in addition, provide access to up to 15,000 of the estimated 29,000 CpG islands in the human genome (59).

\section*{2.2 Materials and Methods}

\textbf{Tissue samples and cell lines:}

Frozen non-small cell lung tumors paired with normal adjacent tissues were collected through the Cooperative Human Tissue Network (CHTN). Nine paired samples (patient number 2, 3, 5, 7, 10, 11, 13, 14 and 17) and clinical characteristics were previously described (127). Six medulloblastoma samples were described previously (97). Three head and neck cancer tissues were collected at The Ohio State University through the CHTN. All sample collection was performed in accordance with NIH guidelines. Non-small cell line lung cancer lines A549 (from ATCC), H125, H1299 and H2086, head and neck cancer cell line SCC-9, leukemia cell lines HL-60, ML-1 and K-562, medulloblastoma cell lines Daoy, D425 MED, MHH-MED-1 and MHH-PNET-5, used in this study were previously described (98, 99, 106, 127, 132).
Isolation of plasmid and genomic DNAs:

High molecular weight DNA for the RLGS procedure was isolated according to our previously published protocol (89). Plasmid DNA was isolated using QIAprep Spin Miniprep kit (Qiagen Inc.) and the manufacturer’s recommended protocols.

Restriction landmark genomic scanning (RLGS):

RLGS was performed according to published protocols (133) with modifications for the use of AscI as the restriction landmark enzyme. Briefly, to prevent non-specific labeling, the sheared ends of approximately 7 μg of genomic DNA were blocked in a 10 μl reaction by the addition of nucleotide analogues (αS-dGTP, αS-dCTP, ddATP, ddTTP) using 2.5 U DNA polymerase I (Boehringer Mannheim) (37°C, 20 min) followed by enzyme inactivation (65°C, 30 min). The DNA was digested (37°C, 2 h) with 20 U EcoRV (New England Biolabs, Beverly, MA), followed by 20 U AscI (New England Biolabs) in NEB buffer 4 (37°C, 2 h). The resulting restriction sites from AscI were labeled in a fill-in reaction using Sequenase Ver. 2.0 (USB, Cleveland, OH) in the presence of [α-32P] dGTP (6000 Ci/mmol, NEN Life Science Products, Boston, MA) and [α-32P] dCTP (3000 Ci/mmol, NEN) for 30 min and stopped by adding buffer that included dCTP and dGTP. A portion of the reaction was electrophoresed through a 60 cm long, 0.8 % agarose tube gel (first dimension separation). The agarose gel was equilibrated in restriction buffer and the DNA was digested in the gel with 750 U Hinf I (New England Biolabs) at 37°C for 2 h. The agarose gel was placed horizontally across the top of a nondenaturing 5 % polyacrylamide gel, the two gels were connected with
molten agarose, and the DNA was electrophoresed in the second dimension. The gels were dried and exposed to Kodak X-OMAT AR film in the presence of one intensifying screen (Quanta 111, DuPont) for 2 - 10 days.

**Ascl restriction trapper purification:**

A mix of 500 µg of total human genomic DNA from three donors was digested with 1500 U *Ascl* at (37°C for 3 h) and subsequently with 500 U *EcoRV* (37°C overnight), extracted with phenol/chloroform/isoamyl alcohol (PCI), precipitated, and resuspended in H₂O at a concentration of 2 mg/ml. Aliquots of 100 µg of restriction digested DNA were ligated in 150 µl volume to a 0.67% (w/v) DNA Trapper R- *BssHII* (Japan Synthetic Rubber Co.) in the presence of 10% PEG 6000 using 1400 U T4 DNA ligase (New England Biolabs, Beverly) at 18°C overnight. The DNA trapper-ligated DNA was digested twice with 100 U *EcoRV*, centrifuged to remove non-ligated *EcoRV* fragments and then digested with 100 U *Ascl* to release *Ascl-EcoRV* fragments. DNA fragments were PCI purified, precipitated in the presence of glycogen (Boehringer Mannheim) and dissolved in 13 µl TE buffer. A total of 11.8 µg purified DNA was recovered. To determine the quality and purity of the *Ascl-EcoRV* fragments, 1 µg was endlabeled using a fill-in reaction using Sequenase Ver. 2.0 in the presence of [α-32P] dGTP (6000 Ci/mmol, DuPont, Boston) and [α-32P] dCTP (3000 Ci/mmol, DuPont) and subjected to the two dimensional separation in the RLGS system. The resulting profile was compared to the RLGS profile prepared from total genomic DNA.
**Construction of vector KSII\(^{+}\)-AscI:**

To insert an *AscI* site (GGCGCGCC) into vector Bluescript KSII\(^{+}\) (Stratagene, Heidelberg), 50 pmol of each primer *AscI*-1 (CCACCGCGGTGGGCGCGCCT) and *AscI*-2 (CTAGAGGCGCGCCCACCCGCGGGTGGAGCT) (custom made by MWG Biotech, Ebersberg) were annealed and subsequently ligated with 100 ng *SacI*-XbaI cut vector DNA. Appropriate insertion of the annealed primers would not disturb the ORF of the multiple cloning site and, hence, the vector’s capability for blue/white selection on agar medium containing X-GAL. *E. coli* DH10B (Life Technologies, Karlsruhe) were transformed with the ligation mixture and plated onto LB agar containing ampicillin, IPTG and X-GAL. Blue colonies were tested for the presence of KSII\(^{+}\) harbouring an *AscI* site. One of such plasmids, designated KSII\(^{+}\)-AscI, was selected for subsequent library construction.

**Library construction:**

To facilitate reliable double digestions of vector KSII\(^{+}\)-AscI with *AscI* plus *EcoV*, we first shot-gun cloned *AscI-EcoRV* genomic fragments of mouse DNA into the vector. A clone with a 1.6 kb *AscI-EcoRV* insert was then used to prepare the vector for library construction. Two µg of the recombinant plasmid were *AscI-EcoRV* digested and separated on a gel. The vector band was sliced out and run a second time on a gel to improve purity. The band was eluted and dissolved in H\(_2\)O at a concentration of 10 ng/µl. Self ligation of 10 ng and subsequent electroporation of electrocompetent *E. coli* DH10B cells (Life Technologies, Bethesda, MD; transformation efficiency \(~7\times10^9\) transformants/µg pUC19) yielded in 45 clones. This figure indicated the expected non-
recombinants when 10 ng vector DNA were ligated with insert DNA at similar conditions. For library construction, two 10 µl ligation mixtures each containing 10 ng vector DNA, 3 µl human restriction trapper purified DNA and 0.5 U T4 DNA ligase (Roche Diagnostics, Mannheim, Germany) were incubated at 16°C for 16 h. After addition of 2.5M NH₄-acetate (final concentration) and 1 µl glycogen (stock: 20 mg/ml; Roche Diagnostics, Mannheim, Germany) as carrier, the DNA was precipitated and redissolved in a total of 5 µl 0.5X TE. One µl ligated DNA was used per transformation.

**Library picking, replication and preparation of high density hybridization filters:**

Transformed cells were spread onto LB/Agar plates containing ampicillin (100 µg/ml), IPTG and X-gal, and grown at 37°C for 18-20 hours. Clones were manually picked and arrayed into 384 well microtiter plates containing LB/ampicillin (50 µg/ml)/glycerol (7.5%). The arrayed clones were incubated at 37°C for 18 hours, and then frozen at -80°C. Five additional copies of each plate were made using a 384-pin replicating tool (V &P Scientific, San Diego, CA) for inoculation. The replicas were grown for 18 hours at 37°C and then frozen at -80°C. High density hybridization filters were prepared using a "Q-Bot" colony picker/ high-density filter griddes. All clones from plate 1 to 48 were used to spot onto three 22.25 x 22.25 cm nylon membranes using a protocols identical to the one used for BAC clones (134).
RLGS mixing gels with clones from the *AscI*-EcoRV library (A-RV-1):

Plates 1 to 32 from A-RV-1 were chosen for the RLGS mixing gels. Clone pool DNAs for each of the 32 plates, all 16 rows (A-P) and 24 columns (1-24) were prepared as described earlier (89). Individual clones were grown in microtiter plates, overnight cultures were combined and plasmid DNAs for each pool of clones were isolated using spin columns (Qiagen). Genomic DNA from normal lung was labeled by a fill-in reaction using Sequenase Ver. 2.0 (USB) in the presence of [α-32P] dCTP (6000 Ci/mmol, NEN) and [α-32P] dGTP (3000 Ci/mmol, NEN) for 30 min. Pooled clone DNA was digested by EcoRV (Promega) and *AscI* (NEB) sequentially and labeled following same procedure for standard RLGS. Ten pg DNA per clone of labeled pooled clone DNA was mixed with the appropriate amount of labeled genomic DNA and loaded on the first dimension RLGS agarose gel followed by the standard RLGS procedure. The amount of labeled genomic DNA was optimized in order to obtain a four-day exposure of the RLGS gel on X-ray film.

**Sequencing and database analysis:**

All sequence analyses were performed in the Core Facility of the Division of Human Cancer Genetics using an ABI PRISM 377 DNA sequencer. For CG-rich sequences high annealing temperatures were employed using an ABI PRISM BigDye Terminator Cycle Sequencing kit. *AscI*-EcoRV clones were sequenced with M13 forward primer. DNA sequence files were analyzed using DNASTar and Chromas software. For homology searches, sequences were submitted to the publicly available databases.
Bioinformatics:

The standard two-sided Z-test was used to compare the methylation frequencies in NotI and AscI test.

We downloaded the assembled sequences (August 6, 2001 draft assembly of UCSC) of the 24 chromosomes from the UCSC Human Genome Project working draft (http://genome.ucsc.edu). We scanned each of the chromosomes for NotI (GCGGCCGC) and AscI (GGCGCGCC) sites, and retrieved the sequences that contain these sites. Each sequence is of length 1008bp (-500 to +500 of the site). We used a sliding window of length 201bp and counted the percentage of CpG dinucleotides (CpG score) and GC% for each window. The sequence is considered a CpG island if there exists a sliding window with CpG score $\geq 60\%$ and GC$\%\geq 50$. We retrieved all the 30,095 CpG islands mapped in the human genome and counted the number of CpG islands that have NotI and AscI sites. In order to determine whether these sites fall in a gene region or not, we used the public human genome annotations available at UCSC genome server. We counted the number of sites that fall in and around (within the 5kb region of the annotated gene ends) known genes and ESTs.

2.3 Results and Discussion:

AscI and NotI restriction sites predicted in the human genome.

The majority of RLGS gels generated to study DNA methylation profiles in human malignancies have used NotI as the restriction landmark enzyme. Many of these
studies were supported by a NotI-EcoRV library (89). In order to develop an additional landmark enzyme for the purpose of CpG island identification by RLGS, we analyzed the human genome sequence for the frequency and location of restriction sites for rare cutting, methylation sensitive restriction endonucleases. AscI is a restriction enzyme that recognizes the target sequence GG\textit{CGCG}CC and does not cut the methylated recognition sequence. Some characteristics of the loci cut by AscI obtained from the human genomic sequence (August 6, 2001 draft assembly of UCSC) are listed in Table 2.1 and compared to those cut by NotI. Surprisingly, the human genome possesses only half the number of AscI sites (4935) as compared to NotI sites (9628), although both recognition sequences are composed of four guanines and four cytosines each and both contain 2 CG dinucleotides. Nevertheless, NotI and AscI are highly comparable in terms of the types of loci they assess. Of particular note is the fact that 86% and 83% of these sites, respectively, are found in CpG islands, while only 5% and 7% are found in repetitive elements not associated with CpG islands. This strong bias of representation of CpG islands over repetitive elements is a major strength of RLGS using these two enzymes. Furthermore, 86% and 83% of these CpG islands, respectively, are associated with known genes or ESTs. These data indicate that AscI is an excellent choice of landmark enzyme to complement RLGS studies performed using NotI. In addition, since NotI and AscI sites co-localize in only 3.7% of CpG islands, by using AscI as a second landmark enzyme we are able to almost double the number of CpG islands whose methylation status can be analyzed.
RLGS profiles using AscI as the landmark enzyme display up to 2000 distinct CpG islands.

We established the reaction conditions for the use of AscI as a restriction landmark site (see Fig. 2.1 for an outline of the procedure). RLGS profiles with AscI show a lower fragment density than NotI profiles (Fig. 2.2a), as expected from the genome sequence survey that identified fewer AscI restriction sites in the human genome (Table 2.1). An AscI master profile was prepared using total genomic DNA from three donors to maximize coverage of polymorphic spots. The master profile was labeled with a coordinate system of spot numbering (a portion is shown in Fig 2.2d) as was done for the NotI master profile previously described (91). The lower density of RLGS fragments in an AscI profile allows the scoring of more fragments in the higher molecular weight sections. These sections are difficult to score in the NotI profiles due to the high density of spots and are frequently excluded from the analysis. Thus, although the number of fragments on an AscI profile is less than on a NotI profile, a similar number of approximately 2,500 fragments can be analyzed on both.

Since AscI is methylation sensitive, we compared methylation frequencies detected by NotI and AscI in the same samples to determine if both recognition sequences are equal targets for aberrant methylation in human malignancies. Table 2.2 summarizes the data obtained for nine lung cancers, six medulloblastomas and three head and neck cancers. The number of methylated sequences detected with both restriction enzymes is not statistically different (P<=0.05). These data indicate that although these enzymes assess different loci, they are similar in their abilities to detect aberrant methylation in human malignancies.
Establishment and initial characterization of an Ascl-EcoRV library.

The initial step in the construction of the Ascl-EcoRV boundary library was the purification of Ascl-EcoRV fragments from total genomic DNA using the BssHII/Ascl restriction trapper. This procedure results in the enrichment of Ascl-EcoRV fragments and eliminates EcoRV-EcoRV fragments (see Material and Method for details). The quality of the purified Ascl-EcoRV fragments was tested by using an aliquot of these fragments for RLGS separation. A portion of the purified Ascl-EcoRV fragments was labeled and subjected to two-dimensional separation in the RLGS system. The resulting RLGS profile showed the same set of fragments as the original profile without prior purification (Fig. 2.2 a,b), indicating that the purification did not result in loss or gain of certain fragments. The remaining purified material was used for cloning into pBluescript KS-Ascl. The Ascl-EcoRV library (A-RV-1) consists of 19,200 clones picked into fifty 384 well plates. The average insert size was 2.48 kb (n=75) ranging from 0.3 kb to 10 kb. Accordingly, the library has an expected bias towards smaller fragments reflecting the cloning bias of the plasmid vector. Clones from 48 plates were spotted onto filters for hybridization based screening, providing an additional resource for studies of CpG islands. Each filter contains the entire set of clones from the 48 plates spotted in duplicate. The availability of these filters allows for rapid identification of plasmid clones with 5’ end sequences for known genes. In addition these clones provide a unique resource for array-based studies.
AscI-EcoRV library clone sequence characteristics are similar to predicted.

CpG islands are mainly located in the promoter region of genes and are less frequently found in the body or 3’ end of genes. The survey of the human genome for AscI sites described above indicated that the recognition sequence of AscI (GGCGCGCC) has a preferential localization to CpG islands. In order to determine if our library has a similar representation, we sequenced 178 AscI-EcoRV fragments cloned from this library by the two methods described below. A total of 158 sequences (89%) showed CpG island features (see Materials and Methods). We mapped all 178 sequences to the human genome draft sequence (August 6, 2001 draft assembly of UCSC) and found that 137 (77%) mapped to known genes or ESTs. In 84 cases where the CpG island could be mapped within the context of a known gene, 66 (79%) were found in the 5’ end of a gene, 12 (14%) in the body, and 6 (7%) in the 3’ end (Table 2.3). This further supports the assumption that AscI sites are preferentially located in CpG islands near genes and assures that our library is a faithful representation of this.

Establishment of mixing gels as a cloning tool for RLGS fragments.

In order to generate a novel tool that will aid the cloning of RLGS fragments from profiles generated with AscI as the landmark enzyme and EcoRV as the second restriction enzyme, we prepared RLGS mixing gels from plates 1 to 32 of the A-RV-1 library. In addition, the rows and columns from these 32 plates were individually pooled to produce 16 row pool (A-P) and 24 column pool (1-24) mixing gels. The procedure followed the strategy previously used for the generation of the NotI-EcoRV mixing gel catalogue (89).
In RLGS mixing gels, fragments for which a corresponding clone is present in the pool of clones mixed with the genomic DNA will show enhancement. Determination of the plate, row, and column mixing gels in which the RLGS spot of interest is enhanced indicates the unique library address where the corresponding RLGS fragment is cloned. An example of an RLGS mixing gel is shown in Fig. 2.2c with clones from a 384 well plate of the library. The average number of enhanced RLGS fragments per plate is 153 since many of the clone insert sizes fall outside of the window of resolution of a standard RLGS profile. In the 32 plate mixing gels there are 1468 unique RLGS fragments represented.

Use of the *AscI-EcoRV* cloning gel catalogue to identify hypermethylated sequences in various cancers.

*AscI* is preferentially located within CpG islands and is methylation sensitive. Thus *AscI* is useful as a restriction landmark enzyme in RLGS studies to identify methylation changes in two different samples. We used the methylation scanning properties of *AscI* to determine methylation changes in medulloblastoma (MB), lung cancer, and head and neck cancer (HNSCC) primary tumors relative to adjacent normal tissue, as well as cell lines representing all three tumor types and a leukemia cell line. RLGS fragment loss in tumor profiles and cell line profiles is the most prominent observation and is indicative of hypermethylation of those fragments. Less frequently, newly appearing RLGS spots are found on the tumor profiles that may represent hypomethylation. Unfortunately, however, such rare RLGS spots cannot be cloned using this library since they are not present in the RLGS profiles of the DNAs used to create the
library. Figure 2.3 shows an example of an RLGS fragment (RLGS fragment A2E54; where “A” indicates Ascl profile and “2E54” indicates spot #54 in section 2E) that is present in normal adjacent lung tissue DNA, but absent from the lung tumor and two lung cancer cell lines. The corresponding library clone was identified in the plate 4, row I, and column 14 mixing gels. Clone 4I14 was isolated from the library and used in a single clone mixing gel to confirm that the clone represents the intended RLGS spot (Fig 2.3a). Insert DNA from this clone was used as a hybridization probe for Southern blot analysis to confirm methylation of the Ascl site. Lung tumor and normal genomic DNA was digested with both EcoRV and Ascl. Control DNA in lane 1 of Figure 2.3b was digested with EcoRV only. In the Southern analysis, the probe detects either a small (Ascl-EcoRV) fragment or a larger (EcoRV-EcoRV) fragment. The presence of the large fragment indicates that the Ascl site was protected from restriction digestion by DNA methylation. The Southern data for patient 14 confirms the RLGS result. Similarly, 17 additional RLGS fragments have been cloned using this targeted approach and are shown in bold in Table 2.3.

**Large scale identification of RLGS fragment sequences.**

A second non-targeted strategy to identify the sequences of RLGS fragments in Ascl gels was also used. Ascl-EcoRV clones derived from the A-RV-1 were sequenced from the Ascl end. This sequence was used to determine the full-length Ascl-EcoRV and the Ascl-HinfI restriction fragment sizes from the August 2001 freeze of the human genome. DNAs of clones with insert sizes of 0.5-5 kb and an Ascl-HinfI fragment greater than 100 bp were pooled into groups of up to 15 plasmid DNAs and used for RLGS
mixing gels. These mixing gels resulted in the enhancement of exactly the same number of RLGS fragments as the number of plasmid clones in the pool. Since the sequence and thus the predicted mobility of these clones in RLGS gels was known it was possible to unambiguously identify which clone and sequence corresponds with each RLGS fragment. Using this strategy it was possible to identify 160 additional RLGS fragments and their sequences. This data is summarized in Table 2.3, which is sorted by methylation status and chromosomal location. Out of the total number (178) of \textit{Ascl} fragments cloned by the two methods described, we found 70 that were methylated in primary tumors and 119 methylated in cancer cell lines (Table 2.3).

\section*{2.4 Conclusions}

We have developed a valuable resource for isolating and studying CpG rich regulatory human sequences using \textit{Ascl} as a restriction landmark enzyme. \textit{Ascl} is as suitable as \textit{NotI} to determine methylation patterns in human malignancies and nearly doubles the set of loci that can be studied. The \textit{Ascl} recognition sequence occurs less frequently in the genome than the \textit{NotI} sequence but its location is similarly biased towards CpG islands. The cloning gel catalogues now available for \textit{NotI-EcoRV} and \textit{Ascl-EcoRV} allow for the targeted cloning of up to 3,257 RLGS fragments (1789 from \textit{NotI} (89), 1468 from \textit{Ascl}), greater than 90\% of which are expected to represent CpG islands (93). In addition, we have developed a non-targeted, but higher throughput strategy for using these libraries to clone RLGS fragments.
By applying these cloning strategies we have created a novel resource, which when used in conjunction with RLGS analysis of tumor profiles, allows for the identification of large numbers of methylation targets. Even in this limited study of 18 analyzed primary tumor profiles, we have already identified 70 targets of hypermethylation in cancer. Three of the genes that were identified, \textit{HOXA11}, \textit{NELL1} and \textit{ALX3} have previously been identified by others as methylation targets in lung adenocarcinomas (123, 135) or neuroblastomas (136), respectively. As we increase the number of tumor profiles analyzed and go through multiple iterations of the cloning strategies describe in this article, we will significantly increase the number of targets of hypermethylation that we can identify. This is a requisite step in order to begin to understand the mechanisms and consequences of such hypermethylation.

Together there are a little over 15,000 \textit{NotI} and \textit{AscI} restriction sites in the human genome. Considering the total number of 29,000 CpG islands in the human genome (59), these libraries provide access to nearly half the CpG islands. Therefore, these libraries will prove to be excellent tools for the study of aberrant CpG island methylation when used in combination with various methylation-scanning techniques such as RLGS and differential methylation hybridization (93, 137). Although standard RLGS running conditions only resolve a set of approximately 2500 CpG islands with a 1st dimension size of 5kb-500bp, these conditions can be altered to resolve a similar number of fragments with 1st dimension size ranging from 10kb-5kb (138). Thus, by modifying RLGS electrophoresis conditions and by utilizing other technologies that do not require electrophoresis, the full potential of these libraries may be achieved.
<table>
<thead>
<tr>
<th><strong>Characteristics</strong></th>
<th><strong>NotI</strong></th>
<th><strong>AscI</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of restriction sites in the human genome</td>
<td>9628</td>
<td>4935</td>
</tr>
<tr>
<td>Number of restriction sites in CpG islands</td>
<td>8239 (86%)</td>
<td>4071 (83%)</td>
</tr>
<tr>
<td>Number of restriction sites in repetitive elements, not CpG islands</td>
<td>520 (5%)</td>
<td>332 (7%)</td>
</tr>
<tr>
<td>Frequency of restriction sites near 5’ end of a known gene</td>
<td>3357 (34.9%)</td>
<td>1612 (32.7%)</td>
</tr>
<tr>
<td>Frequency of restriction sites near 3’ end of a known gene</td>
<td>1328 (13.8%)</td>
<td>725 (15%)</td>
</tr>
<tr>
<td>Frequency of restriction sites inside a gene</td>
<td>1392 (14.5%)</td>
<td>738 (15%)</td>
</tr>
<tr>
<td>Frequency of restriction sites falling near ESTs</td>
<td>2221 (23.1%)</td>
<td>1001 (20.3%)</td>
</tr>
<tr>
<td>Frequency of CpG islands with both NotI and AscI sites</td>
<td>1100 (3.7%)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.1. Characteristics of NotI and AscI restriction sites**

*August 6, 2001 draft assembly of UCSC*
<table>
<thead>
<tr>
<th>Tumor type</th>
<th>Number of methylated CpG islands</th>
<th>Number of RLGS fragments analyzed</th>
<th>Methylation frequency in NotI gels</th>
<th>Number of methylated CpG islands</th>
<th>Number of RLGS fragments analyzed</th>
<th>Methylation frequency in Ascl gels</th>
<th>Z-test statistic for proportions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung 63</td>
<td>1184</td>
<td>5.3%</td>
<td>78</td>
<td>1614</td>
<td>4.8%</td>
<td>0.5833</td>
<td></td>
</tr>
<tr>
<td>Lung 59</td>
<td>1184</td>
<td>5.0%</td>
<td>62</td>
<td>1614</td>
<td>3.8%</td>
<td>1.4669</td>
<td></td>
</tr>
<tr>
<td>Lung 28</td>
<td>1184</td>
<td>2.4%</td>
<td>49</td>
<td>1690</td>
<td>2.9%</td>
<td>-0.8735</td>
<td></td>
</tr>
<tr>
<td>Lung 8</td>
<td>1184</td>
<td>0.7%</td>
<td>8</td>
<td>1008</td>
<td>0.8%</td>
<td>-0.3234</td>
<td></td>
</tr>
<tr>
<td>Lung 7</td>
<td>1184</td>
<td>0.6%</td>
<td>13</td>
<td>1205</td>
<td>1.0%</td>
<td>-1.3079</td>
<td></td>
</tr>
<tr>
<td>Lung 5</td>
<td>1184</td>
<td>0.3%</td>
<td>18</td>
<td>1734</td>
<td>1.0%</td>
<td>-1.8470</td>
<td></td>
</tr>
<tr>
<td>Lung 3</td>
<td>1184</td>
<td>0.3%</td>
<td>7</td>
<td>1734</td>
<td>0.4%</td>
<td>-0.6822</td>
<td></td>
</tr>
<tr>
<td>Lung 1</td>
<td>1184</td>
<td>0.1%</td>
<td>3</td>
<td>1380</td>
<td>0.2%</td>
<td>-0.8503</td>
<td></td>
</tr>
<tr>
<td>Lung 0</td>
<td>1184</td>
<td>0.0%</td>
<td>5</td>
<td>1614</td>
<td>0.3%</td>
<td>-1.9169</td>
<td></td>
</tr>
<tr>
<td>MB 53</td>
<td>1702</td>
<td>3.1%</td>
<td>20</td>
<td>923</td>
<td>2.1%</td>
<td>1.4091</td>
<td></td>
</tr>
<tr>
<td>MB 32</td>
<td>1768</td>
<td>1.8%</td>
<td>22</td>
<td>1246</td>
<td>1.8%</td>
<td>0.0903</td>
<td></td>
</tr>
<tr>
<td>MB 31</td>
<td>1825</td>
<td>1.7%</td>
<td>21</td>
<td>1093</td>
<td>1.9%</td>
<td>-0.4401</td>
<td></td>
</tr>
<tr>
<td>MB 32</td>
<td>2016</td>
<td>1.6%</td>
<td>19</td>
<td>1172</td>
<td>1.6%</td>
<td>-0.0735</td>
<td></td>
</tr>
<tr>
<td>MB 15</td>
<td>1741</td>
<td>0.9%</td>
<td>19</td>
<td>1327</td>
<td>1.4%</td>
<td>-1.4947</td>
<td></td>
</tr>
<tr>
<td>MB 14</td>
<td>2018</td>
<td>0.7%</td>
<td>9</td>
<td>1421</td>
<td>0.6%</td>
<td>0.2140</td>
<td></td>
</tr>
<tr>
<td>HNSCC 13</td>
<td>1703</td>
<td>0.8%</td>
<td>14</td>
<td>1739</td>
<td>0.8%</td>
<td>-0.1387</td>
<td></td>
</tr>
<tr>
<td>HNSCC 3</td>
<td>1839</td>
<td>0.2%</td>
<td>6</td>
<td>1009</td>
<td>0.6%</td>
<td>-1.9625</td>
<td></td>
</tr>
<tr>
<td>HNSCC 0</td>
<td>2126</td>
<td>0.0%</td>
<td>1</td>
<td>1243</td>
<td>0.1%</td>
<td>-1.3080</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2. Methylation frequencies in various tumor samples determined by either NotI or Ascl as a restriction enzyme

aHNSCC- head and neck squamous cell carcinomas; MB- medulloblastoma

bZ-test statistic value for testing the significant difference in methylation frequencies in NotI and Ascl gels. All of the Z-static values are between - 1.96 to +1.96, which suggest that there is no significant (p<=0.05) difference in methylation frequencies in NotI and Ascl gels.
<table>
<thead>
<tr>
<th>RLGS spot</th>
<th>chrNA (Rdm) position</th>
<th>chr. band</th>
<th>Cpg island</th>
<th>Methylation</th>
<th>Gene homology or context</th>
</tr>
</thead>
</table>

**Table 2.3. Cloned Ascl spots**

<table>
<thead>
<tr>
<th>RLGS spot</th>
<th>chrNA (Rdm) position</th>
<th>chr. band</th>
<th>Cpg island</th>
<th>Methylation</th>
<th>Gene homology or context</th>
</tr>
</thead>
</table>

**Legend:**
- Primary Cell lines: L = LNCaP, HN = HN-200
- Gene or EST homology: DRD4/PTDSS
- Context: 5' end Body
- Body position: 5' end
- Homology: PTDSS2

**Additional Information:**
- Asc island: Primary Cell lines
- Body position: 5' end
- Homology: PTDSS2

**Reference:**
- Table 2.3: Cloned Ascl spots
- chrNA (Rdm) position: chrNA (Rdm) position
- chr. band: chr. band
- Cpg island: Cpg island
- Methylation: Methylation
- Gene homology or context: Gene homology or context
<table>
<thead>
<tr>
<th>Chr</th>
<th>Start</th>
<th>End</th>
<th>Str</th>
<th>5' End</th>
<th>3' End</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>LU PDE4DIP</td>
<td>5' end</td>
</tr>
<tr>
<td>2</td>
<td>21109946</td>
<td>21100008</td>
<td>L</td>
<td>HN LU</td>
<td>EDN3</td>
<td>5' end</td>
</tr>
<tr>
<td>3</td>
<td>31301480</td>
<td>31302856</td>
<td>L</td>
<td>UBE2A</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>4</td>
<td>13408838</td>
<td>13411584</td>
<td>L</td>
<td>HN LU</td>
<td>ADIR</td>
<td>5' end</td>
</tr>
<tr>
<td>5</td>
<td>119332838</td>
<td>119334546</td>
<td>L</td>
<td>HN LU</td>
<td>CPI-5/MEG3035</td>
<td>5' end</td>
</tr>
<tr>
<td>6</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>7</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>8</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>9</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>10</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>11</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>12</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>13</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>14</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>15</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>16</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>17</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>18</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>19</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>20</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>21</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>22</td>
<td>171209431</td>
<td>171213341</td>
<td>L</td>
<td>HN LU</td>
<td>EST</td>
<td>5' end</td>
</tr>
<tr>
<td>X</td>
<td>119332838</td>
<td>119334546</td>
<td>L</td>
<td>HN LU</td>
<td>UBE2A</td>
<td>5' end</td>
</tr>
<tr>
<td>Y</td>
<td>119332838</td>
<td>119334546</td>
<td>L</td>
<td>HN LU</td>
<td>CPI-5/MEG3035</td>
<td>5' end</td>
</tr>
<tr>
<td>3p23.3</td>
<td>88999433</td>
<td>89001018</td>
<td>L</td>
<td>HN LU</td>
<td>IDUA</td>
<td>5' end</td>
</tr>
<tr>
<td>Chr</td>
<td>Start</td>
<td>End</td>
<td>Length</td>
<td>Gene</td>
<td>Methylation</td>
<td>Expression</td>
</tr>
<tr>
<td>----</td>
<td>-------</td>
<td>-----</td>
<td>--------</td>
<td>------</td>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>chr7:35634775-35639133</td>
<td>7p14.3</td>
<td>Y</td>
<td>UMPH1</td>
<td>5' end</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:171991051-171991626</td>
<td>7q36.3</td>
<td>Y</td>
<td>VIPR2</td>
<td>5' end</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:16943828-169442460</td>
<td>7q36.3</td>
<td>Y</td>
<td>EST</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr8:22999489-23000030</td>
<td>8q22</td>
<td>N</td>
<td>EST</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr8:106931787-106938522</td>
<td>8q22</td>
<td>Y</td>
<td>EST</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr8:11654244-11655166</td>
<td>8p23.1</td>
<td>Y</td>
<td>MASL1</td>
<td>5' end</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:11223172-112234856</td>
<td>8q22.2</td>
<td>Y</td>
<td>FLJ13955</td>
<td>5' end</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:15552100-15557689</td>
<td>9p22.3</td>
<td>Y</td>
<td>EST</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:12832254-12835833</td>
<td>9p23.4</td>
<td>Y</td>
<td>FLJ10578</td>
<td>5' end</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:103784748-1037848636</td>
<td>10q26.13</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:131288366-131290808</td>
<td>10q25.3</td>
<td>Y</td>
<td>GPRK5</td>
<td>5' end</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:102645051-102642431</td>
<td>10q24.32</td>
<td>Y</td>
<td>EST</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:1150765407-1150767466</td>
<td>11p12.2</td>
<td>Y</td>
<td>MADD</td>
<td>Body</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:77635388-77638573</td>
<td>12q12.1</td>
<td>Y</td>
<td>MAD2</td>
<td>Body</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:126950393-126951113</td>
<td>12q12.23</td>
<td>Y</td>
<td>DEC2</td>
<td>3' end</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:1348904213-1348905159</td>
<td>13q14.2</td>
<td>Y</td>
<td>PTPRO</td>
<td>5' end</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:1457864802-1457865972</td>
<td>14q23.1</td>
<td>N</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:1030357109-103038095</td>
<td>14q23.2</td>
<td>Y</td>
<td>EST</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:104114601-104116064</td>
<td>14q23.33</td>
<td>Y</td>
<td>EST</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:103329673-103381214</td>
<td>14q23.33</td>
<td>Y</td>
<td>BAG5</td>
<td>5' end</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:1567148586-1567152897</td>
<td>15q23</td>
<td>Y</td>
<td>DDXBP1</td>
<td>5' end</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:156858604-156859143</td>
<td>15q23</td>
<td>N</td>
<td>EST</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:1571744340-157146434</td>
<td>15q23</td>
<td>Y</td>
<td>ARH11</td>
<td>Body</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:154610574-154613722</td>
<td>15q21.1</td>
<td>Y</td>
<td>C15ORF3</td>
<td>5' end</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:154922267-154923623</td>
<td>15q21.2</td>
<td>Y</td>
<td>EST</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:157800452-157801848</td>
<td>15q23.3</td>
<td>Y</td>
<td>RAN</td>
<td>5' end</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:159075188-159079211</td>
<td>15q23.1</td>
<td>Y</td>
<td>WWOX</td>
<td>5' end</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:1787076028-1787077939</td>
<td>17q25.3</td>
<td>Y</td>
<td>MGC10561</td>
<td>5' end</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr7:184722735-184724741</td>
<td>18p11.21</td>
<td>Y</td>
<td>FLJ10352</td>
<td>Body</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr20:18473985-184746482</td>
<td>20p11.23</td>
<td>Y</td>
<td>EST</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr20:1037413619-1037414289</td>
<td>22q21.22</td>
<td>Y</td>
<td>EST</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr22:17007571-17008276</td>
<td>22q11.21</td>
<td>Y</td>
<td>EST</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.3. Cloned Ascl spots

\(a\)BLAT search results based on http://genome.ucsc.edu/cgi-bin/hgBlat, Aug. 2001 Freeze.

\(b\)Y indicates that the Ascl site is within a CpG island, N indicates that it is not.

\(c\)Indicates methylation found in a primary tumor or cell line as indicated. L, lung carcinoma; MB, medulloblastoma; HN, head and neck squamous cell carcinoma; LU, Leukemia

\(d\)Known genes found in the Refseq database or spliced ESTs.

\(e\)5'end indicates that the CpG island includes the region immediately upstream of exon 1 and/or exon 1. 3' end indicates the CpG island is found in 3' most exon. Body indicates the CpG island is found within the genomic structure of the gene excluding the 5' most and 3' most known exons.

\(f\)Genes share same CpG island in their 5' end and are transcribed in opposite directions.
<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>Isolation of high molecular weight genomic DNA from fresh or fresh frozen tissue.</td>
</tr>
<tr>
<td>II.</td>
<td>Blocking of sheared DNA ends with αS-dGTP, αS-dCTP, ddATP, ddTTP and DNA polymerase I to prevent non-specific labeling.</td>
</tr>
<tr>
<td>III.</td>
<td><em>EcoRV</em> restriction digest.</td>
</tr>
<tr>
<td>IV.</td>
<td><em>Asci</em> restriction landmark digest. <em>Asci</em> cuts only unmethylated restriction sites but not methylated.</td>
</tr>
<tr>
<td>V.</td>
<td>Radioactive labeling of <em>Asci</em> restriction half-sites in a fill-in reaction using ([α-32P]) dGTP, ([α-32P]) dCTP and Sequenase Ver. 2.0.</td>
</tr>
<tr>
<td>VI.</td>
<td>First dimension: high resolution 0.8% agarose gel electrophoresis to separate restriction fragments by molecular weight.</td>
</tr>
<tr>
<td>VII.</td>
<td>In-gel digest of restriction fragments using <em>Hinfl</em>.</td>
</tr>
<tr>
<td>VIII.</td>
<td>Second dimension: separation of restriction fragments in a 5% non-denaturing polyacrylamide gel.</td>
</tr>
<tr>
<td>IX.</td>
<td>Gel drying and exposure to X-ray film.</td>
</tr>
</tbody>
</table>

**Figure 2.1 Outline of the RLGS procedure using *Asci* as a restriction landmark enzyme.**
Figure 2.2 RLGS profiles using *Ascl-EcoRV-Hinfl* restriction enzyme combination.

A. RLGS profile of normal lung DNA.

B. RLGS profile using restriction trapper purified *Ascl-EcoRV* fragments derived from peripheral blood lymphocyte DNAs.

C. RLGS mixing gel generated with normal lung DNA as the genomic background and clones from plate 3 pool in the A-RV1 library.

D. Section 4C of the *Ascl* Master RLGS profile, showing the numbers assigned to each *Ascl* fragments.
Figure 2.3 RLGS identifies DNA methylation in primary lung cancer.

A. Sections from RLGS profiles including RLGS fragment A2E54 (arrow). Sections from normal and tumor profiles from patient 14 as well as two lung cancer cell lines (H1299 and H125) are shown. The corresponding Ascl-EcoRV clone was found in plate 4, row I and column 14, and this clone was confirmed by use in a mixing gel.

B. DNA from Ascl clone 4114 corresponding to RLGS spot A2E54 was used for Southern analysis. DNAs from normal lung (NL), lung tumors (T) and adjacent normal tissue (N) from patients 10, 14, 17 and 18, as well as from three lung cancer cell lines H125, H1299 and A549 were digested with Ascl and EcoRV. DNA in the first lane was digested only with EcoRV and shows the size of the EcoRV fragment. In the double digestes, hybridization to the large EcoRV band is indicative of protection of the Ascl site digestion by methylation. The smaller band is indicative of cutting by Ascl.
CHAPTER 3

GLOBAL METHYLATION PROFILING OF LUNG CANCER IDENTIFIES NOVEL METHYLATED GENES


3.1 Introduction

Lung cancer is the leading cause of cancer related death in both males and females world-wide (139). Clinically, lung cancer can be divided into two groups: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) (2, 3). The latter constitutes approximately 75% of all lung cancers (3) and includes squamous cell carcinoma (SCC), adenocarcinoma (AC) and large cell carcinoma (LCC) (approximately 30%, 40% and 15%, respectively in all lung cancer cases in North America) (2).

Molecular abnormalities in lung cancer affect both growth-promoting oncogenes and growth-inhibiting tumor suppressor genes. So far mutations have been
reported in the oncogene K-RAS (7, 20, 31) as well as in tumor suppressor genes p53 (16), CDKN2 (18, 19, 80, 81) and RB (21). In addition to genetic changes, the epigenetic change of DNA methylation, the addition of a methyl group to the cytosine ring in 5’-CpG-3’dinucleotides, may play a significant role during lung cancer development (55, 140-143). DNA methylation is established and maintained by a family of DNA methyltransferases (144) and affects chromatin organization as well as gene expression (145). A well-studied example in lung cancer is the aberrant promoter methylation of the tumor suppressor gene, CDKN2, which correlates with gene silencing (19, 80, 146-148).

Since most of the reports describe methylation in single cancer genes, no measurement of the overall contribution of promoter methylation in lung cancer exists. As an initial step to address this question, Zochbauer-Muller et al. showed that numerous genes, including retinoic acid receptor β-2 (RARβ), tissue inhibitor of metalloproteinnase 3 (TIMP-3), CDKN2, O6-methylguanine-DNA-methyltransferase (MGMT), death-associated protein kinase (DAPK), E-cadherin (ECAD), p14ARF and glutathione S-transferase P1 (GSTP1), were methylated at various degrees in a collection of 107 primary non-small cell lung cancer (148).

Methylation changes in lung cancer appear to be early events leading to malignancy, and thus should be useful in improving early detection of potentially malignant cells (82, 149, 150). For example, CDKN2 promoter methylation is proposed as a biomarker for early detection of lung cancer and monitoring of prevention trials (82, 149, 150). Using sensitive PCR-based methylation analysis, methylation in CDKN2 and/or MGMT promoters was found in sputum of smokers up to three years prior to clinical diagnosis of squamous cell lung carcinoma (82, 149, 150).
There is a need for the identification of novel markers in lung cancer as well as the identification of cancer related genes. Aberrantly methylated target sequences can guide a search for novel genes that may be useful biomarkers, as well as candidate cancer genes. In this study we use, for the first time, Restriction Landmark Genomic Scanning (RLGS) to determine frequencies of DNA methylation and to identify novel methylation targets in NSCLC samples. Transcription patterns for one methylated gene, \textit{BMP3B}, were studied in greater detail in primary tumors as well as in lung cancer cell lines. We show that aberrant methylation of the CpG island of \textit{BMP3B} downregulates gene transcription of this interesting gene product.

### 3.2 Materials and Methods

**Primary human non-small cell lung cancer (NSCLC) samples and cell lines:**

Sixteen frozen paired NSCLC tumors with normal adjacent tissue were collected through the Cooperative Human Tissue Network (CHTN) in accordance with NIH guidelines. Four cancer cell lines, H23, H125, H522 and H1155, were cultured in RPMI-1640 medium (Gibco BRL, Rockville MD) supplemented with 10% fetal bovine serum, 100 units/ml Penicillin and 0.1mg/ml Streptomycin (Gibco BRL).

**Two-dimensional separation by Restriction Landmark Genome Scanning (RLGS):**

RLGS was performed as described previously (89, 151). In summary, high molecular weight DNA was digested with the methylation sensitive restriction enzyme \textit{NotI} (Promega, Madison, WI), end-labeled by $[^{32}\text{P}]-\text{dGTP}$ and $[^{32}\text{P}]-\text{dCTP}$.
(Amersham, Piscataway, NJ), and then digested using the restriction enzyme EcoRV (Promega). NotI-EcoRV DNA fragments were separated in a first dimension through an 0.8% agarose tube gel, followed by an in-gel digestion with a third restriction enzyme, HinfI (New England Biolabs, Beverly MA). Finally, the DNA was separated on a second dimension 5% polyacryamide gel, the gel was dried and exposed to X-ray film for 5-10 days. RLGS profiles of primary tumors and normal adjacent lung tissue were superimposed to visually detect differences in the intensities and/or presence of the radiolabeled fragments.

**RLGS analysis:**

The fragments in the RLGS profiles have been named on our “Master RLGS profile” (see website [http://pandora.med.ohio-state.edu/masterRLGS/](http://pandora.med.ohio-state.edu/masterRLGS/)). The master RLGS profile, derived from normal peripheral blood lymphocyte DNA, is divided into 63 sections. Each RLGS fragment is given a unique identifier (e.g. 3C1) that relates to the position within the RLGS profile. Therefore, data sets from different patients can be compared in order to identify commonly changed fragments.

**Cloning of RLGS fragments:**

A human NotI-EcoRV plasmid library and library mixing gels were created previously to facilitate cloning of RLGS fragments (89, 94). These mixing gels allow the determination of an address for a library clone corresponding to the RLGS fragment by identifying enhancement in the plate, row and column mixing gels (89). Bacterial clones were cultured in LB medium with ampicillin to isolate plasmid DNA using Qiagen®
miniprep kit (Qiagen, Valencia, CA). The plasmid DNA was digested with \textit{Not}I and \textit{EcoRV} (Promega), end-labeled with $[\alpha^{32}\text{P}]-\text{dGTP}$ and $[\alpha^{32}\text{P}]-\text{dCTP}$ (Amersham). 5.2 pg and 10.4 pg labeled DNA per clone were mixed with labeled peripheral blood lymphocytes (PBL) genomic DNA from a normal healthy donor, and subsequently separated in the two dimensional RLGS mixing gel. Enhanced intensities of the RLGS fragment of interest in these mixing gels indicated that the \textit{Not}I/\textit{EcoRV} clone represents the RLGS fragment of interest.

\textbf{Characterization of RLGS fragments:}

The confirmed plasmid clones are sequenced with M13 forward and M13 reverse primers. DNA sequences from M13 forward primer were used to perform standard nucleotide-nucleotide BLAST searches, using non-redundant (nr) and high throughput genomic sequence (htgs) databases at release time March 1, 2001 (http://www.ncbi.nlm.nih.gov/BLAST/). When possible, 2kb genomic sequences from both sides of the \textit{Not}I site were used for subsequent BLAST searches for genes or ESTs. Chromosomal location of cloned DNA fragments were obtained either directly from the information given in Genebank (http://www.ncbi.nlm.nih.gov/), or by searching the OMIM database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM) using gene names, or by searching the BAC resource website (http://www.ncbi.nlm.nih.gov/genome/cyto/hbrc.shtml) that contains cytogenetic data of FISH-mapped and sequence-tagged BAC clones. The properties of CpG islands were determined using a web based program http://www.itba.mi.cnr.it/webgene/, which is provided by the Institute of Advanced Biomedical Technologies (ITBA), Italy.
Southern hybridization:

Southern hybridization was performed as described previously (89). Briefly, control samples, DNA from healthy donors, were digested by EcoRV alone and a NotI/EcoRV combination. The control sample will show the expected fragment size in the EcoRV digestion and a smaller fragment in the double digestion. All tumor samples were digested by NotI/EcoRV. The probes were prepared by restriction enzyme digestion of the clone DNA, purification of the target fragments, and subsequent labeling by random priming using the Prime IT® II kit (Stratagene, La Jolla, CA). Percentage of methylation was quantified by phosphorimager.

5-aza-2’-deoxycytidine treatment of cell lines:

A total of 3×10^6 cells of each NSCLC cell line were seeded into T75 culture flasks and cultured with RPMI-1640 media overnight. 5-aza-2’-deoxycytidine media was freshly made each day from stock solution (10 mmol/L in 100% DMSO) (Sigma, St. Louis, MO) to a final concentration of 1μmol/L. Cells were cultured with 5-aza-2’-deoxycytidine medium for 24 hours and then washed with PBS twice and continued to culture with fresh medium without 5-aza-2’-deoxycytidine for two days (24 hour timepoint). Cells for the 72 hour timepoint were cultured in 5-aza-2’-deoxycytidine medium that was changed daily for three days, then washed with PBS twice and continued to culture in fresh medium without 5-aza-2’-deoxycytidine for one additional day.
Semi-quantitative RT-PCR:

Total RNA from cell lines with or without 5-aza-2’-deoxycytidine treatment, primary tumors and paired normal adjacent lung tissue was isolated using TRIzol (Gibco BRL) and purified by RNeasy Mini Kit (Qiagen, Valencia, CA). Three microgram total RNA was reverse transcribed \textit{in vitro} by random hexamer and oligo dT using SUPERSCRIPT™ First-Strand Synthesis kit (Gibco BRL). cDNA was amplified by PCR. Primers for RT-PCR were designed from the published cDNA sequences. Forward and reverse primers are from different exons to avoid amplification from genomic DNA. Primers sequences were as follows: BMP3B forward: 5’-GGTGGACTTCGCAGACATCG-3’; BMP3B reverse: 5’-GATGGTGGCATGGTTGGATG-3’, product size: 130bp. GPI forward: 5’-GACCCCCAGTTCCAGAAGCTG-3’; GPI reverse: 5’-GCATCAGCTCTCCGTCACC-3’, product size: 178bp. In all reactions the forward primer for each pairs was end labeled by \([\gamma^{-32}\text{P}]\) ATP with T4 kinase (Gibco BRL).

The semi-quantitative radioactive RT-PCR was performed with optimized conditions for both the target gene and an internal control, \textit{glucose}-6-\textit{phosphate isomerase} (\textit{GPI}), in a single reaction tube. Amplification was stopped in the exponential range for both genes. The exponential range was determined by phosphorimager quantification of the PCR product band intensities from different amplification cycles. Each PCR reaction was carried out in 50 µl final volume containing 5 µl of 10× PCR buffer, 1.5mM MgCl₂, 10 pmol of each primer, 200 µM of each dNTP and 2.5 units Platinum Taq DNA polymerase (Gibco). The reactions were initiated with 95°C for 10 minutes to activate DNA polymerase and then followed by amplification. \textit{BMP3B} was
amplified at 96°C for 20 seconds, 63°C for 15 seconds and 72°C for 15 seconds for 8 cycles prior to addition of GPI primers and additional 22 cycles.

**Statistical Methods:**

Tests were performed for heterogeneity in methylation across patients and for preferential methylation of certain CpG island fragments, described in detail in [30]. Briefly, the heterogeneity test is based on a comparison of the mean methylation frequency to its variance in a chisquare statistic. Preferential methylation is assessed using a standard goodness-of-fit test (152) assuming that all spots are lost at equal true frequency. Empirical null distributions for both of these statistics were obtained by performing appropriate 10,000 random permutations of the fragment/patient data. Such an approach accounts for multiple testing (i.e. multiple fragments were examined) and does not rely on asymptotic distribution assumptions.

### 3.3 Results

**Levels of methylation in CpG islands of NSCLC.**

RLGS profiles from 16 matched pairs consisting of lung tumors from NSCLC patients and matched normal lung tissue were prepared using the enzyme combination *NotI-EcoRV-HinfI* (Figure 3.1A). Each tumor profile was compared against the matching normal lung RLGS profile. Previously, we had shown that the loss of an RLGS fragment in the tumor as compared to the matching normal is indicative of DNA methylation in the
NotI site (91, 132). The total number of methylation events out of 1184 RLGS fragments analyzed, as well as clinical data for the patients are shown in Table 3.1. The range of methylation in these samples is from 0 to 5.3%. While most (12 of 16) of the tumor samples showed methylation levels below 1%, four of 16 patients show levels of CpG island methylation above 2%. Of these, patients 5, 11, 14 and 17 show methylation frequencies of 4.9%, 5.3%, 2.4% and 5.0%, respectively. No obvious correlation of overall methylation frequency and clinical data can be seen although tumors from patients 5 and 11 were both classified as stage T2N0, IB. This range of variation is greater than would be expected if all patients had the same underlying methylation rate. A chisquare test (see Statistical Methods) shows significant heterogeneity in methylation levels across the patients (p<.0001). In addition, some CpG islands are preferentially methylated, as indicated by the number of fragments showing relatively high methylation frequency (Figure 3.2, p<.0001). Of the total 1184 analyzed RLGS fragments, 1036 were never methylated. A total of 76 fragments were methylated in only one tumor, 53 fragments were methylated in two tumors, nine fragments (2E61, 2F43, 3D24, 3F82, 3F85, 4D8, 4E53, 4F15 and 4F58) were methylated in 3 tumors, two fragments (2D45 and 3F16) were methylated in four tumors, one fragment (3F28) was methylated in five tumors, one fragment (3G78) was methylated in seven tumors, five fragments (2C35, 3C1, 3E55, 4E1 and one RLGS fragment not present in the RLGS master profile) were methylated in eight tumors and one fragment (1F22) was methylated in nine tumor samples (Figure 3.2).

No correlation between global methylation frequency and any of the clinical parameters including tumor stage, differentiation, histopathological classification, age or
gender could be detected. However, it is interesting to note that a subset of methylation events correlate with the histopathological features of the tumors. For example, 30 CpG islands are methylated only in adenocarcinomas (AC), 28 methylation events are specific for large cell carcinomas (LCC) and 43 are found methylated exclusively in squamous cell carcinomas (SCC). Other methylation events (a total of 17 CpG islands) are shared between all three subtypes. In addition 4 methylation events are found in AC and LCC, 8 are shared between AC and SCC and a total of 17 CpG islands are methylated in both SCC and LCC (Table 3.2).

Cloning of RLGS fragments and properties of cloned fragments.

To further characterize some of the methylated fragments in NSCLC, the NotI/EcoRV plasmid clone library mixing gels were used to locate the corresponding clones (89). Clones with expected insert sizes were analyzed in RLGS mixing gels to confirm that the correct fragment was cloned. Figure 3.1B shows two examples for RLGS fragments 3C1 and 4F15. A total of 21 fragments were cloned and sequenced. Twelve of these clones (2D14, 2D20, 2C35, 2E24, 2E61, 3B36, 3C1, 3E55, 3F16, 3F50, 3F82 and 4E53) have been identified as methylation targets in other types of malignancies (91, 98, 132). BLAST searches identified homologies to 11 genes and 6 EST sequences (Table 3.3). Two sequences (3F16 and 3F82) show high homology to DNA-binding protein A (DBPA) and mouse early B-cell factor 3, respectively and may represent either pseudogenes or novel gene family members. The remaining 2 showed homology to genomic sequences. We found that 20 of 21 NotI sites are located within CpG islands. Six of the 11 CpG islands with homology to genes are located in the 5’ end
of the genes. The CpG islands identified in *insulin promoter factor 1 (IPF1)*, *orthodenticle (drosophila) homolog 1 (OTX1)*, *HOX11*, *T-box brain 1*, *monocarboxylate transporter 3 (MCT3)* are located in the middle or 3’ end of the genes. Chromosomal location of all the fragments were derived from database searches of the human draft sequence of the human genome, Genebank or OMIM database. The detailed information for these 21 fragments, including fragment addresses on master profile, total methylation frequency in primary tumors, CpG island properties and location in the genes, BLAST search results, and chromosomal location of the genes, are listed in Table 3.3.

**RLGS fragment loss is due to DNA methylation.**

Southern hybridization was performed as a more sensitive method to estimate the degree of methylation of the NotI site from RLGS fragments lost in tumors. Southern hybridization has the ability to detect methylation as low as 5-10%, while RLGS allows the detection of 30% methylation (91). Southern blot hybridization was also performed to evaluate the possibility that the RLGS fragment loss was due homozygous deletion. Corresponding NotI/EcoRV plasmid clones for four RLGS fragments were used as probes on Southern blots. These four clones included three fragments with homology to known genes 3C1 (*G-α-olfactory, GNAL*), 3E55 (*insulin promoter factor 1, IPF1*), and 4F15 (*Bone morphogenetic protein 3B, BMP3B*). In addition fragment 2C35 with homology to an EST sequence was used. Representative Southern blots for RLGS fragment 3C1 (*GNAL*) and 4F15 (*BMP3B*) are shown in figures 3.1C and 3.1D, respectively. Tumor samples and paired normal adjacent lung tissue DNAs were digested with NotI/EcoRV. The detection of a fragment in the NotI/EcoRV digests equal in size to
the one detected with EcoRV only was scored as a methylation event in the tumor DNA. The hybridization shows that most tumor samples have a much higher degree of methylation of the NotI site than the normal counterparts. No homozygous deletions were detected since the hybridization signals are present in all tumor samples. Methylation of the NotI site in the 5' end of GNAL was observed in 10 out of 16 patients. Methylation of 4F15 (BMP3B), was detected in five out of six tumors and thus confirmed the methylation events found by RLGS. Tumor specific or increased methylation was observed in 11 out of 16 primary non-small cell lung cancer for fragment 2C35 and 11/16 for 3E55 (IPF1) (Figure 3.1C,D and data not shown).

**Aberrant transcription of BMP3B in primary NSCLC and NSCLC cell lines.**

Radioactive semi-quantitative RT-PCR reactions were performed to determine the expression levels of BMP3B in 6 primary NSCLC samples and their paired normal lung tissue. BMP3B was found hypermethylated in a CpG island that is located in the 5’ end of the gene (Figure 3.3A) and was selected for further analysis. RT-PCR was performed under optimized conditions for both the target gene and the internal control gene glucose-6-phosphate isomerase (GPI). BMP3B was expressed in normal lung, whereas expression of BMP3B in all studied tumor samples was reduced (Figure 3.3B). Interestingly, tumors from patient No. 2 and 6, which did not show methylation of the NotI site exhibited very low levels of BMP3B expression, suggesting heterogeneous methylation of the CpG island or other genetic modes of gene silencing.

In order to investigate the effect of CpG island methylation on the transcription of the associated genes in more detail, we used three NSCLC cell lines that are methylated
in the *Not*I site of *BMP3B* promoter region. H23, H125 and H1155 show more than 50% methylation in the *BMP3B* promoter. The cells were treated with two different timepoints (24 h and 72 h) of 5-aza-2’-deoxycytidine (Figure 3.3C). Similar to the primary tumors, all three cell lines (H23, H125, H1155) did not show any detectable level of *BMP3B* expression at baseline. However, expression was induced after 5-aza-2’-deoxycytidine treatment in all three cell lines (Figure 3.3C).

### 3.4 Discussion

The methylation scanning properties of RLGS have previously been used for the identification of imprinted genes in the mouse genome (86, 125), as well as for the identification of methylated sequences in various human malignancies (91, 92, 132) but not lung cancer. Genome-wide scans for methylated sequences in lung cancer have been performed by two different strategies including the use of a methylated DNA binding column (122) and arbitrarily primed PCR (153). However, these techniques either showed a bias for methylated repetitive sequences or were limited in the number of analyzed sequences. In contrast, Restriction Landmark Genomic Scanning (RLGS), a highly reproducible two-dimensional gel electrophoresis, is a genome-wide scan of DNA methylation changes in CpG islands. Established cloning protocols utilizing an arrayed plasmid clone library facilitate the rapid identification of genomic sequences corresponding to the methylated targets (89, 94). RLGS is based on the digestion of genomic DNA with the methylation sensitive restriction enzyme *Not*I, which can only
digest unmethylated genomic DNA and does not rely on prior knowledge of the gene sequence (89, 91, 92).

We have used RLGS to determine the contribution of CpG island hypermethylation in non-small cell lung cancer (NSCLC) and to identify novel methylation targets. While the importance of DNA methylation in lung tumor development had been demonstrated in several reports (27, 76, 82, 154), the overall extent was previously unknown. For the first time we demonstrate that up to 5.3% of all promoter regions, or 1537 of the estimated total 29,000 CpG islands (59, 155) in the tumor genome could be methylated. The variability in the range of methylation shows that NSCLC represents a heterogeneous group not only with respect to the genetic defects identified but also on the epigenetic level.

Promoter methylation in cancer related genes is well known in lung cancer and is correlated with gene silencing in genes involved in cell cycle (CDKN2) (18, 19, 80, 81), apoptosis (DAP) (156), metastasis H-cadherin (157) and (TIMP-3) (158), differentiation (RARβ) (159), DNA repair (MGMT) (160) and the recently identified candidate tumor suppressor gene (RASSF1A) with homology to the RAS family (27). In this study, we have identified 21 additional genomic loci including eleven genes and six ESTs with aberrant CpG island methylation in NSCLC. None of the cloned genes has been reported to be methylated in lung cancer previously. However, it is interesting to note that CpG island 3F16, is approximately 200 kb away from MGMT (located in sequence contig NT_024100.1). MGMT is a DNA repair gene that was previously found to be methylated in 21% of NSCLC (148). This finding would suggest a more regional effect of methylation, similar to aberrant methylation found in chromosome 17p11.2 in the major
breakpoint cluster region for medulloblastomas (132). Another interesting target sequence is CpG island 4D8 located in chromosome 17q25.1. Chromosome 17q had previously been implicated with frequent (42%) loss of heterozygosity in NSCLC (161), suggesting that both genetic and epigenetic mechanisms could be involved in the silencing of a putative tumor suppressor gene in this region. Other fragments are derived from chromosome segments in 1p (3G78), 5q (4E15), 10q (3F50), 13q (3E55) 18p (3C1) and 22q (2E24) all within regions for which either LOH or homozygous deletions have been reported in lung cancer (162, 163). Whether any of the newly identified genes meet the expected criteria for tumor suppressor genes remains to be determined and will be the focus of future work.

We focused our studies on RLGS fragment 4F15, derived from a CpG island in the promoter region of the \textit{BMP3B} gene and determined the relation of CpG island methylation to transcription. \textit{BMP3B} is located in chromosome 10q11.21-11.23, a region that was shown to be deleted in 20-30% of non-small cell lung cancers and 51% of small cell lung cancers (164) (see online \url{http://www.helsinki.fi/~lgl_www/LOSS/Respiratory.html}) and thus located in a candidate tumor suppressor region. \textit{BMP3B} is a member of the transforming growth factor \(\beta\) (TGF-\(\beta\)) superfamily, originally identified due to their osteoinductive capacity. Members of this family are usually involved in the regulation of cell growth/differentiation during development and were shown to be dysregulated in various human malignancies. Interestingly other members of the BMP family have been shown to induce apoptosis during organ development (165-167). In addition two BMP family members, \textit{BMP4} and \textit{BMP2}, have been shown to induce apoptosis in multiple myeloma cell lines (168) or hematopoetic cells (169), respectively. \textit{BMP2} was also
shown to suppress the transformed phenotype in the human lung carcinoma cell line A549 (170).

*BMP3B* is highly expressed in human adult lung, brain, skeletal muscle, pancreas and testis, an expression pattern that distinguishes it from the closest family member *BMP3* (171). *BMP3B* knockout mice did not show any detectable abnormalities, suggesting a redundant function with that of other members of the TGF-β family (172). However, adult rat lung tissue does not express *BMP3B* suggesting the possibility of different functions of *BMP3B* in rodents and humans (173). We have shown that methylation in *BMP3B* is correlated with transcriptional repression. The repression is reversible by treatment with the demethylating agent 5-aza-2'-deoxycytidine. Thus, our data suggest a causal relationship between methylation of the *BMP3B* promoter and transcriptional repression. We found *BMP3B* down regulated in all NSCLC patient samples and cell lines, even in those without methylation in the *NotI* site, suggesting that methylation patterns in the CpG island are heterogeneous. This assumption was confirmed by COBRA analysis testing the methylation status in four *BstUI* (CGCG) restriction sites in the promoter (Dai *et al.* unpublished). Alternatively, other mechanisms (e.g. LOH or mutations) of gene silencing could be present. Additional work to study the complete genetic and epigenetic mutation spectrum of *BMP3B* in lung tumorigenesis is underway.

The identification of multiple targets for methylation opens the exciting possibility to use these methylation events as biomarkers for the early detection of lung cancer in sputum as demonstrated by others (82). In addition our data also indicate the possibility that certain methylation events may be specific for lung cancer or subtypes
within this group and thus could serve as potential markers for the molecular
classification of lung cancers and different disease stages. Validation of methylation
events as possible markers for early diagnosis, as predictive markers for survival or
markers that classify subtypes will require larger sets of patient samples.
<table>
<thead>
<tr>
<th>Patient No.</th>
<th>Methylation events out of 1184 CpG islands</th>
<th>Percent of methylation</th>
<th>Age</th>
<th>Gender</th>
<th>Tumor stage</th>
<th>Differentiation</th>
<th>Tumor type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>0.4</td>
<td>45</td>
<td>F</td>
<td>T2N1, IIB</td>
<td>Well</td>
<td>AC</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.3</td>
<td>72</td>
<td>M</td>
<td>T1N0, IA</td>
<td>Moderate</td>
<td>SCC</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>0.6</td>
<td>56</td>
<td>F</td>
<td>T2N1, IIB</td>
<td>Poor</td>
<td>LCC with features of AC</td>
</tr>
<tr>
<td>5</td>
<td>58</td>
<td>4.9</td>
<td>75</td>
<td>M</td>
<td>T2N0, IB</td>
<td>Moderate</td>
<td>SCC</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>0.6</td>
<td>62</td>
<td>F</td>
<td>T1N0, IA</td>
<td>Poor</td>
<td>SCC</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0.2</td>
<td>68</td>
<td>M</td>
<td>T3N2, IIIA</td>
<td>Undifferentiated</td>
<td>LCC</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>66</td>
<td>M</td>
<td>T2N0, IB</td>
<td>Well</td>
<td>AC</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>0.4</td>
<td>78</td>
<td>F</td>
<td>T1N0, IA</td>
<td>Poor</td>
<td>SCC</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0.1</td>
<td>69</td>
<td>M</td>
<td>T2N1, IIB</td>
<td>Poor</td>
<td>N/A</td>
</tr>
<tr>
<td>11</td>
<td>63</td>
<td>5.3</td>
<td>67</td>
<td>F</td>
<td>T2N0, IB</td>
<td>Poor</td>
<td>LCC</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>76</td>
<td>M</td>
<td>T1N1, IIA</td>
<td>Poor</td>
<td>SCC</td>
</tr>
<tr>
<td>14</td>
<td>28</td>
<td>2.4</td>
<td>61</td>
<td>M</td>
<td>N/A</td>
<td>Poor</td>
<td>SCC</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>0.3</td>
<td>70</td>
<td>F</td>
<td>T3N0, IIB</td>
<td>N/A</td>
<td>LCC with SCC and AC</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>0.2</td>
<td>79</td>
<td>F</td>
<td>T1N0, IA</td>
<td>N/A</td>
<td>AC (with clear cell features)</td>
</tr>
<tr>
<td>17</td>
<td>59</td>
<td>5.0</td>
<td>81</td>
<td>M</td>
<td>N/A</td>
<td>Well</td>
<td>AC</td>
</tr>
<tr>
<td>18</td>
<td>8</td>
<td>0.7</td>
<td>63</td>
<td>F</td>
<td>N/A</td>
<td>Poor</td>
<td>SCC</td>
</tr>
</tbody>
</table>

**Table 3.1. DNA methylation in 16 lung cancer patients.**

SCC-squamous cell carcinoma, AC-adenocarcinoma, LCC-large cell carcinoma, N/A, no data available.
<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Number of Methylated CpG islands</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC</td>
<td>28</td>
</tr>
<tr>
<td>AC</td>
<td>30</td>
</tr>
<tr>
<td>SCC</td>
<td>43</td>
</tr>
<tr>
<td>LCC+AC</td>
<td>4</td>
</tr>
<tr>
<td>AC+SCC</td>
<td>8</td>
</tr>
<tr>
<td>LCC+SCC</td>
<td>17</td>
</tr>
<tr>
<td>LCC+AC+SCC</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 3.2. Distribution of methylated CpG islands in various histopathological subgroups of non-small cell lung cancer.

LCC- large cell carcinoma; AC adenocarcinoma; SCC-squamous cell carcinoma
<table>
<thead>
<tr>
<th>RLGS Master address</th>
<th>No. of tumors methylated</th>
<th>Methylation found in following subgroups</th>
<th>CpG island</th>
<th>Location of CpG island in gene</th>
<th>Gene or EST</th>
<th>Accession Number</th>
<th>Chromosomal Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>2C35 8</td>
<td>AC/LCC/SCC</td>
<td>Yes</td>
<td>EST</td>
<td>BG142595</td>
<td>AL139281</td>
<td>NT_024073.1</td>
<td>10p12</td>
</tr>
<tr>
<td>3C1 8</td>
<td>AC/LCC/SCC</td>
<td>Yes</td>
<td>5' end</td>
<td>GNAL</td>
<td>U55180</td>
<td>18p11.21-pter</td>
<td></td>
</tr>
<tr>
<td>3E55 8</td>
<td>AC/LCC/SCC</td>
<td>Yes</td>
<td>3' end</td>
<td>IPF1</td>
<td>NM 000209</td>
<td>13q12.1</td>
<td></td>
</tr>
<tr>
<td>3G78 7</td>
<td>AC/LCC/SCC</td>
<td>Yes</td>
<td>5' end</td>
<td>TAL1</td>
<td>AL135960</td>
<td>1p32</td>
<td></td>
</tr>
<tr>
<td>2D20 4</td>
<td>AC/LCC/SCC</td>
<td>Yes</td>
<td>middle</td>
<td>OTX1</td>
<td>AB037501</td>
<td>2p13</td>
<td></td>
</tr>
<tr>
<td>3F16 4</td>
<td>AC/LCC/SCC</td>
<td>Yes</td>
<td></td>
<td>Homologous to EBF</td>
<td>AL354950</td>
<td>NT_024100.1</td>
<td>10q26</td>
</tr>
<tr>
<td>3F50 4</td>
<td>LCC/SCC</td>
<td>Yes</td>
<td>3' end</td>
<td>HOX11</td>
<td>AJ009794</td>
<td>10q24</td>
<td></td>
</tr>
<tr>
<td>2E61 3</td>
<td>SCC</td>
<td>Yes</td>
<td>EST</td>
<td>A2230817</td>
<td>AL354000</td>
<td>17p11.2</td>
<td></td>
</tr>
<tr>
<td>3B55 3</td>
<td>LCC/SCC</td>
<td>Yes</td>
<td>3' end</td>
<td>T-box brain I</td>
<td>XM_002531</td>
<td>2q23-27</td>
<td></td>
</tr>
<tr>
<td>3F82 3</td>
<td>LCC/SCC</td>
<td>Yes</td>
<td></td>
<td>Homologous to DBP4</td>
<td>M24069</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4D8 3</td>
<td>AC/LCC/SCC</td>
<td>No</td>
<td>EST</td>
<td>BF928282</td>
<td>AC012118</td>
<td>NT_010641.1</td>
<td>17q25.1</td>
</tr>
<tr>
<td>4E15 3</td>
<td>LCC/SCC</td>
<td>Yes</td>
<td>EST</td>
<td>BE247619</td>
<td>AC003959</td>
<td>5q</td>
<td></td>
</tr>
<tr>
<td>4E53 3</td>
<td>AC/LCC/SCC</td>
<td>Yes</td>
<td></td>
<td></td>
<td>AL353159</td>
<td>NT_009829.1</td>
<td>13q12.2</td>
</tr>
<tr>
<td>4F15 3</td>
<td>LCC/SCC</td>
<td>Yes</td>
<td>5' end</td>
<td>BMP3B</td>
<td>D49493,   NT_008757.1</td>
<td>10q11.21-11.11</td>
<td></td>
</tr>
<tr>
<td>5E25 3</td>
<td>AC/LCC/SCC</td>
<td>Yes</td>
<td></td>
<td>Formin 2 like</td>
<td>AL359918</td>
<td>NT_004771.1</td>
<td>1q44</td>
</tr>
<tr>
<td>2D14 2</td>
<td>AC/LCC/SCC</td>
<td>Yes</td>
<td>5' end</td>
<td>CD8 β Chain</td>
<td>S87068</td>
<td>2p12</td>
<td></td>
</tr>
<tr>
<td>3B36 2</td>
<td>AC/SCC</td>
<td>Yes</td>
<td>5' end</td>
<td>CYP1b1</td>
<td>XM_002576</td>
<td>2p21-22</td>
<td></td>
</tr>
<tr>
<td>3D44 2</td>
<td>AC/SCC</td>
<td>Yes</td>
<td></td>
<td></td>
<td>AL355304</td>
<td>NT_019429.1</td>
<td>6q23.1-6q24.3</td>
</tr>
<tr>
<td>5C32 2</td>
<td>AC/SCC</td>
<td>Yes</td>
<td>5' end</td>
<td>CD34</td>
<td>M81938</td>
<td>1q2</td>
<td></td>
</tr>
<tr>
<td>2E24 1</td>
<td>AC</td>
<td>Yes</td>
<td>middle</td>
<td>MCT3</td>
<td>AF132611</td>
<td>22q12.3-13.2</td>
<td></td>
</tr>
<tr>
<td>3E34 1</td>
<td>SCC</td>
<td>Yes</td>
<td>EST</td>
<td>A631157</td>
<td>AC000068</td>
<td>22q11.2</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.3. Cloned RLGS fragments altered in RLGS profiles of 16 primary tumors.

AC-adenocarcinoma, LCC-large cell carcinoma, SCC-squamous cell carcinoma,
Figure 3.1 Restriction Landmark Genomic Scanning (RLGS) and confirmation of methylation by Southern hybridization.

A. Representative RLGS profile of normal lung DNA using the enzyme combination \textit{NotI-EcoRV-Hinfl}. Fragment sizes for the 1\textsuperscript{st} dimension (1D) and 2\textsuperscript{nd} dimension (2D) are given in kb.

B. Sections of RLGS profiles from normal lung and lung tumor from patient 17 highlighting RLGS fragment 3C1 (\textit{GNAL}) and 4F15 (\textit{BMP3B}) (arrows). While the lung cancer profile shows decreased intensity of 3C1 and 4F15, the mixing gel shows enhancement, indicating that the correct \textit{NotI-EcoRV} clone was isolated from the library. Only differences in RLGS fragments 3C1 and 4F15 were indicated, other changes between normal and tumor profiles were not marked by arrows.

C. and D. Southern blot analysis of primary lung cancers and matched normal adjacent tissue. Hybridization was used to confirm methylation of RLGS fragment 3C1, \textit{GNAL} (figure 3.1C) and 4F15, \textit{BMP3B} (figure 3.1D) in RLGS profiles of primary lung cancer is due to DNA methylation. The peripheral blood lymphocyte (PBL) DNA in the first lane of each blot was digested with \textit{EcoRV} only. All other tumor (T) and normal adjacent lung (N) DNA samples were double digested by \textit{EcoRV/NotI}. The majority of the DNA is digested by \textit{NotI} and results in the smaller \textit{NotI-EcoRV} (N/RV) fragment. Tumor samples show a much higher degree of methylation. M: indicates the marker lane. H23, H125, H522 and H1155 are lung cancer cell lines.
Figure 3.1
Figure 3.2 Global methylation patterns in lung cancer are non-random.
Frequency distribution of methylation events in CpG islands within 16 non-small cell lung cancers.
Figure 3.3 Abnormal transcription of *BMP3B* NSCLC samples.

**A.** Schematic representation of the genomic structure of *BMP3B*. The location of the *NotI* site relative to exon 1 and the extent of the CpG island are displayed. The star indicates the *NotI* site identified by RLGS.

**B.** Semi-quantitative radioactive RT-PCR was used to determine the relative expression levels of *BMP3B* in six lung cancer patients. Relative expression levels of RNAs obtained from tumor (T) and normal adjacent lung (N) were determined by comparing the intensities to the internal control, GPI. Negative control (-) contains water. GPI: *Glucose-6-phosphate isomerase*.

**C.** Radioactive semi-quantitative RT-PCR was used to determine the relative expression of *BMP3B* levels in RNAs from three cell lines (H23, H125 and H1155) after treatment with 5'-aza-2'-deoxycytidine. The upper panel shows the RT-PCR results and the lower panel shows the relative expression levels of *BMP3B* compared to the internal control GPI. Cell lines were treated with 1μmol/L 5'-aza-2'-deoxycytidine for either 24 hours (24) or 72 hours (72). Untreated control cell lines (C) were harvested in exponential growth phase. Y-axis gives the relative ratio of band intensity for the target gene/GPI quantified by Phosphorimager. Negative control (-) contained water, positive control (+) was normal lung tissue (NL).
Figure 3.3
CHAPTER 4

CHARACTERIZATION OF BMP3B AS A NON-SMALL CELL LUNG CANCER RELATED GENE

4.1 Introduction:

DNA methylation of CpG islands in the promoter regions of genes has been correlated with the silencing of tumor suppressor genes (51, 55, 93). BMP3B was identified as a methylated target in non-small cell lung cancer using restriction landmark genomic scanning (RLGS) (127). This gene was highly expressed in normal human lung (171), but was found to be down regulated in primary non-small cell lung cancers and cell lines. After treating non-small cell cancer cell lines with 5-aza-2’-deoxycytidine, a DNA methyltransferase inhibitor, the gene was reexpressed (127) suggesting that DNA methylation was one of the mechanisms leading to BMP3B gene down regulation in lung cancer. These observations indicated that the biological significance of BMP3B silencing in lung cancer is worthy of further study.

BMP3B belongs to TGF-β superfamily. Both TGF-β and BMP signal through the SMAD signaling transduction pathway (174). TGF-β has a wide variety of biological
activities, including the induction of mesenchymal epithelial transition during embryogenesis, context dependent cell growth inhibition or stimulation, extracellular matrix synthesis, immunoresponse and carcinogenesis (175). The role of TGF-β, especially TGF-β1, has been extensively studied in tumorigenesis (176). Current knowledge indicates that TGF-β has both tumor suppressor and oncogenic activities (174, 176, 177). It is noteworthy that during the late stage of tumor development, TGF-β is often overexpressed in tumors compared to their normal counterparts and promotes tumor progression by different mechanisms (176, 178). For example, TGF-β signaling promotes epithelial mesenchymal transition. During this process the epithelial cells lose their epithelial phenotype and cell-cell adhesion by inhibiting the expression of genes such as E-cadhein, and gain the fibroblastoid phenotype, which facilitates invasion and metastasis. In addition, TGF-β1 secreted from tumor cells induces angiogenesis and suppress host immunoresponse (176).

The function of BMP3B has not been well characterized, however, at the amino acid level, BMP3B showed 82% homology to BMP3, the closest family member of BMP3B (171). Hino et al. reported that the BMP3b mRNA level was increased during rat calvarial osteoblastic differentiation simultaneously with the activity of osteoblast differentiation marker alkaline phosphatase (ALPase). In addition, BMP3b expression could be rapidly and completely suppressed by TGF-β1 and was associated with inhibition of osteoblast differentiation (179). Melnick et al. found that mouse embryonic submandibular glands (SMGs) were induced to apoptosis and decreased proliferation by the NF-κB inhibitor SN50 and that increased BMP3B expression was one of the markers
that could best discriminate the treated SMGs from controls (180). Finally, BMP3B expression was found to be repressed in the rat kidney epithelial cell line RK3E after Ha-ras transformation (181). Taken together, these data suggest that loss of function of BMP3B might promote cell proliferation and/or block differentiation in lung cancer development.

In the current study, a large sample set of primary non-small cell lung cancers were screened for BMP3B promoter methylation, as well as genetic changes such as point mutations and LOH. The gene was also reexpressed in A549 non-small cell cancer cell line to evaluate BMP3B function in carcinogenesis. The growth properties of the cells after BMP3B reexpression were evaluated in both *in vitro* and *in vivo* studies. Finally, possible BMP3B downstream target genes were analyzed using microarray gene profiling.

### 4.2 Materials and methods:

**Cell lines and primary lung cancer samples**

Type II alveolar epithelial human lung cancer cell line A549 was cultured in RMPI-1640, supplemented with 10% fetal bovine serum, 100 U/ml penicillin and 0.1 mg/ml streptomycin (Invitrogen, Carlsbad, CA). Cell lines transfected with pBabe retrovirus were maintained in the same medium as the parental lines but the medium was supplemented with 3 μg/ml puromycin. Phoenix amphotropic packaging cells were cultured in DMEM medium with 15% fetal bovine serum. Twenty paired primary non-small cell lung cancer samples were collected through CHTN at The Ohio State
University. Additional 76 primary lung cancer DNAs and 4 normal lung DNAs were collected at the Minneapolis VA Medical center. All sample collection was performed according to NIH guidelines.

**Bisulfite treatment of genomic DNA**

Bisulfite treatment of genomic DNA was performed with modifications using the previously published protocol (114). Briefly, one microgram genomic DNA was denatured by 3M NaOH for 10 min at 37°C in a volume of 50µl and then incubated at 50°C overnight following the addition of 10µl, 10mM hydroquinone, 520µl, 3M sodium bisulfite. The DNA was then purified using Qiaquick gel extraction kit (Qiagen, Valencia, CA) and eluted into 50µl elution buffer. After adding 5µl fresh 3M NaOH, the reaction was incubated at room temperature for 5 min. Ten microliter 3M NaOAc (pH 5.0) was added to adjust the pH. The treated DNA was again purified using the Qiaquick gel extraction kit and eluted into 30µl elution buffer. One microliter of the product was used for each PCR reaction.

**Bisulfite sequencing**

BMP3B bisulfite sequencing PCR primers were designed to amplify both methylated and unmethylated DNA and do not include any CpG in the primer sequence. The forward primer is 5’-TGTAGGGGTAATTAGTAGGTAGG-3’ and reverse primer is 5’-AACCTCTAAAACACTACAACAAAAA-3’, which is located in the promoter region of BMP3B from –649 to –389 relative to the translation start site (3268 to 3527 in NCBI sequence D49493.1). Bisulfite converted DNA was amplified using the RoboCycler® Gradient 40
PCR machine (Stratagene, La Jolla, CA) at 95°C for 10 minutes to activate Taq DNA polymerase, followed by 35-cycles of amplification. For each cycle, the template was denatured at 96°C for 40 seconds, annealing was performed at 58°C for 1 minute and extension was done at 72°C for 66 seconds. Each PCR reaction was performed in a 50 μl volume containing approximately 33 ng of bisulfite treated genomic DNA, 1× PCR buffer (114), 40 pmol of each primer, 1.25 mM of each dNTP and 2.0 U Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA). The PCR product was gel purified using Qiaquick gel extraction kit and ligated into pCR2.1-TOPO TA cloning vector (Invitrogen, Carlsbad, CA). After transformation into TOP10F’ competent cells, single colonies were picked up for plasmid DNA isolation using QIAprep® Spin Miniprep kit (Qiagen). Ten clones with correct insert sizes were sequenced.

**Methylation-Specific PCR (MS-PCR) analysis of BMP3B promoter.**

MS-PCR analysis was performed on bisulfite treated DNA from 20 lung tumor samples paired with normal adjacent tissue, and an additional 76 lung tumor samples. Both the non-methylation specific primers BMP3B-UF: 5'-TGGTGTTGATATATAGGAGTT-3', BMP3B-UR: 5'-AAATCATCCCTAACCCAACCT-3', and methylation specific primers BMP3B-MF: 5'-CGGCGTGCATATAGGAGTC-3', BMP3B-MR: 5'-AAATCGTCCCTAACCCGACT-3' were designed to amplify a 166 bp region of the BMP3B promoter from –578 to –413 relative to the translation start site (3339 to 3504 in NCBI sequence D49493.1). Each amplification was performed in a 25 μl reaction containing
~30 ng bisulfite treated DNA, 20 pmol each of forward and reverse primer, 1.25 mM dNTPs, 1X Herman's Buffer (114), and 1 unit of Platinum Taq DNA polymerase (Invitrogen). Both methylation and non-methylation specific reactions started at 95°C for 10 min to activate the polymerase. Non-methylation specific reactions were carried out for 38 cycles of 30 sec at 96°C denaturation, 15 sec at 56°C annealing, and 15 sec at 72°C extension. Methylation specific reactions were carried out for 35 cycles of 30 sec at 96°C denaturation, 15 sec at 61°C annealing, and 15 sec at 72°C extension. Both reactions ended with a single extension at 72°C for 7 min. Reactions were carried out in a Perkin Elmer 9700 thermal cycler (Perkin Elmer Applied Biosystems). Electrophoresis of a 10µl aliquot of each reaction was performed on 8% denaturing polyacrylamide gel and stained with ethidium bromide. 100% methylated DNA was prepared from normal lung genomic DNA with M.SssI CpG methylase (New England BioLabs). M.SssI treated DNA was purified using the QIAquick gel extraction kit (QIAGEN). 0.1%, 1%, 10%, and 50% methylated normal lung DNA were prepared by mixing M.SssI methylated normal lung DNA and unmethylated normal lung DNA in 1:1000, 1:100, 1:10, and 1:1 ratios respectively. These samples were then subject to sodium bisulfite treatment as described previously.

**Mutation Screen**

Mutation screening was carried out on 96 lung tumor DNA samples. Four different PCR amplifications products were used to cover the coding regions of all 3 exons of BMP3B. The first 18 bases of the forward and reverse primers were 'tailed' with M13
forward and reverse primer sequences respectively. PCR primer sequences are as follows:

Exon 1: BMP3Bmut1_F 5’-TGTAAAACGACGGCCAGTCCTTCCTCCTCCTCCTGGACTTC-3’; BMP3Bmut1_R 5’-CAGGAAACAGCTATGACCAATGAGGAGAAGGGGTCTCTG-3’

Exon 2:
- BMB3Bmut2.1_F 5’-TGTAAAACGACGGCCAGTCAGCCTGTGGTCTCTCCTCTC-3’
- BMB3Bmut2.1_R 5’-CAGGAAACAGCTATGACCGGGAAGGGGTCGTATCTCTCTG-3’
- BMB3Bmut2.2_F 5’-TGTAAAACGACGGCCAGTGTCCTGTGGTCTCTCCTCTC-3’
- BMB3Bmut2.2_R 5’-CAGGAAACAGCTATGACCGGCGTGGCTTCTTTCCAG-3’

Exon 3:
- BMB3Bmut3_F 5’-TGTAAAACGACGGCCAGTGCTGTGGTCTCTCCTCTC-3’
- BMB3Bmut3_R 5’-CAGGAAACAGCTATGACCGGCGTGGCTTCTCTTCCAC-3’

Each PCR amplification was performed in a 50 μl reaction containing ~30ng genomic DNA, 200nM each forward and reverse primer, 200 μM dNTPs, 1× cloned Pfu DNA polymerase reaction buffer (Stratagene), either 2.4% DMSO (BMP3Bmut1 and BMP3Bmut3) or 5% DMSO (BMP3Bmut2.1 and BMP3Bmut2.2), and 2.5 units of PfuTurbo hot start DNA polymerase (Stratagene). All PCR reactions started at 95°C for 2 min BMP3Bmut1, BMP3Bmut2.1, and BMP3Bmut3 reactions included 35 cycles of 30 sec at 96°C denaturation, 30 sec at 60°C annealing, 50 sec at 72°C extension. BMP3Bmut2.2 reactions included 38 cycles of 30 sec at 96°C denaturation, 30 sec at 60°C annealing, 50 sec at 72°C extension. All reactions ended with a single extension at 72°C for 7 min. A 10 μl aliquot of each PCR amplification products was then subject to treatment with 10 units of Exonuclease I (USB Corp. Cleveland, Ohio USA) and 2 units of
Shrimp alkaline phosphatase (USB Corp. Cleveland, Ohio USA) for 15 min. at 37°C followed by 15 min at 80°C to inactivate the enzymes. The PCR products were sequenced at the Genotyping-Sequencing Unit in the Division of Human Cancer Genetics, The Ohio State University. Differences were verified by sequencing from reverse direction of re-amplified PCR products.

**Genotyping**

Genomic DNA (2 ng/μl) was sent to the Genotyping-Sequencing Unit in the Division of Human Cancer Genetics, OSU, for automated fluorescent microsatellite marker analysis. The chromosome 10 markers were from ABI’s Linkage Mapping Set Version 2, including D10S249, D10S591, D10S189, D10S547, D10S1653, D10S548, D10S197, D10S208, D10S196, D10S1652, D10S537, D10S1686, D10S185, D10S192, D10S597, D10S1693, D10S587, D10S217, D101651 and D10S212. A ratio of the normal/tumor pairs was calculated using formula: (height of normal allele 1/height of normal allele 2)/(height of tumor allele 1/height of tumor allele 2). A threshold of less than 0.67 or greater than 1.5 was determined as LOH.

**Construction of MYC tagged pBabe-BMP3B expression vector.**

IMAGE clone (No. 308743 in vector pT7T3D-Pac) containing the majority of the BMP3B cDNA was obtained from the ATCC. This clone is missing the 5’ end of the cDNA and starts at position + 186 relative to the translation start site of the published coding sequence of BMP3B (NCBI accession number NM_004962.2). The BMP3B
portion of the coding sequence missing at the 5’-end was amplified by RT-PCR from normal lung cDNA using forward primer 5’-AATCCTAAGCTTCCTCCTCGGACTTCG-3’ (AAGCTT is HindIII restriction site) and reverse primer 5’-GAAATACACGGCTTCTGGTCG-3’. The PCR product was digested with HindIII and SalI, and ligated with the BMP3B image clone digested with same enzymes. The HindIII site is located in the multiple cloning sites of the vector and SalI is located in the BMP3B cDNA insert. The cloning of the full-length BMP3B coding region in pT7T3C-Pac was confirmed by sequencing. The full-length cDNA insert was cut out of the vector using HindIII and Pmel and subsequently ligated with pcDNA3.1/hygro (+) vector (Invitrogen) that was digested with HindIII and EcoRV (blunt end).

In order to prepare c-MYC-tagged BMP3B, the c-MYC tag sequence was prepared by PCR amplification of MCS pcDNA3-5×Myc plasmid (gift from Dr. Leone in Human Cancer Genetics, OSU). The forward primer for this reaction was 5’-CCCAAGCTTGGATCCACCATGGA-3’ (GGATCC BamHI restriction site) and reverse PCR primer was 5-TACTAAGCTTTCCAGATCCTCTT-3’ (AAGCTT HindIII restriction site). The product PCR product contains two HindIII restriction sites on both ends of the c-MYC sequence. This HindIII fragment was cloned into the HindIII site of pcDNA3.1/hygro-BMP3B plasmid. The c-MYC tagged BMP3B was digested with BamHI and Pmel (blunt end), and then was ligated into pBabe-puromycin vector that was digested with BamHI and SnaBI (blunt end). Using similar strategy, pBabe-puro-BMP3B without c-MYC tag was also constructed.
**Retrovirus infection of A549 cell line and puromycin selection**

Plasmid DNA of pBabe-vector alone, pBabe-puro-BMP3B, and c-MYC tagged pBabe-puro-BMP3B were transiently transfected into Phoenix Amphotropic cells using calcium phosphate coprecipitation transfection. The retrovirus in the culture medium was harvested after 48 hours post transfection and then added to the A549 cell line at 30-40% confluency. This step was performed twice in a 12 hour interval. After 36 hours post infection, the A549 cells were split and cells containing the retrovirus were selected using 3 ug/ml puromycin medium for 5 days. Non-transfected A549 cells were used as controls for the selection and all cells died within 24-48 hours.

**RT-PCR:**

The cDNA preparation protocol and primers were previously published (127). Total RNA from cell lines was isolated using TRIzol (Invitrogen) and further purified by RNeasy Mini Kit (Qiagen). Three microgram total RNA was used for reverse transcription by both random hexamer and oligo dT using SUPERSCRIPT™ First-Strand Synthesis kit (Invitrogen). Forward and reverse primers are from different exons to avoid amplification from genomic DNA. BMP3B forward: 5’-GGTGGACTTCGCAGACATCG-3’; BMP3B reverse: 5’-GATGGTGCCATGGTGGATG-3’, product size: 130bp. GPI forward: 5’-GACCCCAAGTTCCAGAAGCTG-3’; GPI reverse: 5’-GCATCAGTCCTCCGTCACC-3’, product size: 178bp.
Each PCR was performed in 50 µl final volume containing 1µl cDNA from above, 1× PCR buffer, 1.5mM MgCl₂, 10pmol of each primer, 200µM of each dNTP and 2.5U Platinum Taq DNA polymerase (Invitrogen). The PCR was performed in Perkin Elmer PCR machine 9700 (Perkin Elmer Applied Biosystems, Foster city, CA). The PCR started at 95°C for 10 minutes to activate DNA polymerase. Both PCR reactions were carried out for 30-cycle amplification. For each cycle, the template was denatured at 96°C for 25 seconds, annealed at 63°C for 15 seconds and extended at 72°C for 15 seconds.

**Western Hybridization:**

We followed previously published procedure with slight modification (182, 183). Cells were lysed with lysis buffer [50 mM Tris-HCl, 250 mM NaCl, 5 mM EDTA, 50 mM NaF, 0.1% Igepal CA-630, and protease inhibitor cocktail (Sigma)]. Equal amounts of proteins (150–250 µg) were separated on 9% SDS-PAGE for c-MYC (9E10) monoclonal antibody (Santa Cruz Biotechnology, Inc., Santa Cruz, CA) and α-tubulin (Ab-1) monoclonal antibody (Oncogene Research Products, Boston, MA). Proteins on the gels were transferred onto Hybond™ ECL™ nitrocellulose membranes (Amersham Pharmacia Biotech, Uppsala, Sweden) and blocked in blocking buffer [5% nonfat milk, 200 mM NaCl, 25 mM Tris (pH 7.5), and 0.05% Tween 20]. The blot was incubated with primary antibody at 4°C overnight with shaking. After washing with TBS-T (20 mM Tris, 500 mM NaCl, and 0.1% Tween 20) once, blocking buffer once and then TBS-T four times, 5 min each, membranes were incubated with the appropriate secondary
antibody (antimouse 1:2000 for both c-MYC and α-tubulin antibody) at 4°C for 1–2 hour(s). After same wash for primary antibody, proteins were detected using a chemiluminescent detection system (Amersham Pharmacia Biotech).

**Growth curve assay**

Ten thousands cells of each cell line were seeded into 100-mm dish. Triplicate for each cell lines were cultured in RPMI 1640 medium plus 3 μg/ml puromycin. The cells were harvested and counted everyday using Coulter® ZTM Series Particle Counter (Coulter Corporation, Miami, FL) started at second day for 10 days.

**Anchorage dependent colony formation assay**

One thousands cells were plated into 60-mm dishes. All cell lines were plated in triplicate. The cells were cultured for 14 days, washed twice with PBS, and then fixed with methanol/acetic acid (3:1), stained with 0.1% crystal violate in PBS for 30 min, washed and dried in room temperature.

**Anchorage independent colony formation assay in soft agar**

Four percent noble agar (Sigma) was prepared in H₂O and autoclaved. The 4% agar was melted in the microwave and cooled to 50°C in a waterbath. Five ml 4% agar was mixed with 35 ml medium (RMPI-1640, supplemented with 10% fetal bovine serum, 100 U/ml penicillin, 0.1 mg/ml streptomycin and 3 μg/ml puromycin) pre-warmed to 37°C in a water bath. Four ml of the 0.5% agar in the medium was put into 60-mm dish
to provide the base agar. By same way, 0.3% agar in the culture medium was prepared
and was equalized to 40°C in a water bath, 2\times10^4 cells were suspended into 4 ml 0.3%
agar for each 60-mm plate and were plated on the base agar. All cell lines were plated in
triplicate and experiments were repeated three times. After incubation at 37°C incubator
for 14 to 21 days, the cell colonies were stained with 0.005% crystal violet for 1 hour.
The number of colonies was counted using a dissecting microscope.

**In vivo cell growth assay in nude mice**

Two million of pBabe vector or pBabe-BMP3B infected cells in 0.5 ml culture
medium (RMI-1640, supplemented with 10% fetal bovine serum, 100 U/ml penicillin,
0.1 mg/ml streptomycin and 3 μg/ml puromycin) were subcutaneously injected into
athymic-nu/nu nude mice (Charles River Laboratory, Wilmington, MA). The tumors
were measured weekly. The tumor volume was calculated as width \times length \times height / 2.
After 10 weeks, the mice were sacrificed and the tumor weight was measured.

**Statistic analysis of cell growth**

The t-test was used to assess whether there are statistical difference of the cell
growth between pBabe-puro vector alone control group, pBabe-puro-c-MYC-BMP3B
group or pBabe-puro-BMP3B-group.
Microarray study of pBabe vector and pBabe-BMP3B infected A549

9000 cDNA clones from subtractive RDA libraries (184) have been used as probes. These clones have in common that they are at least enriched for genes expressed at higher levels in tumors than in normal tissue. The libraries will be described in detail elsewhere. PCR products of these cDNAs have been spotted onto poly-L-lysine coated slides using the Promedia microarray printer (Promedia, Larchmont, NY) and treated according to standard protocols available from http://cmgm.stanford.edu/pbrown/protocols/ with minor variations. Briefly, DNA was covalently linked to poly-L-lysine by 60mJ UV, and slides were blocked using succinic anhydride. 20ug total RNA from A549 cells transfected with a retrovirus expressing BMP3B as well as from control cells transfected with an empty virus were reverse transcribed in the presence of either Cy3-dUTP or Cy5-dUTP. After purification of labeled cDNAs by ethanol precipitation, a mixture of both labeled cDNAs was hybridized overnight onto the microarrays. In a dye-swap experiment, the labeling of RNAs was exchanged. After washing, independent images were obtained for Cy3 and Cy5 fluorescence emitted from hybridized microarrays by using the Affymetrix 428 dual channel laser scanner (Affymetrix, Santa Clara, CA). The images were analyzed with GenePix Pro software version 4.0 (Axon Instruments, Foster City, CA) and spots of poor quality, determined by visual inspection, were removed from further analysis.

Net signal intensity (NSI) in each channel (Cy3 and Cy5) was determined by subtracting the local background from signal intensity values and the median of ratios (mR) of NSIs was used to determine regulation of gene expression for each single probe. A correction for differences in overall intensities of both channels was applied via global
normalization. This global normalization was done using the log of the median of all mRs and is based on the assumption that the percentage of mRNA in the total RNA from both samples is identical.

Only features matching the following criteria were used for estimation of gene expression: a) spots with $\geq 66\%$ of feature pixels with intensities more than two standard deviations above the local background pixel intensities at both wavelengths. This parameter helps to eliminate features of low intensities that frequently can produce non-reliable medians of ratios. b) spots with $\geq 0.5\ Rgn\ R^2$, a parameter calculated by the GenePix Pro software that characterizes the coefficient of determination for the current regression value. Using this criterion, we have eliminated features of non-homogenous signals. c) spots with $mR > 1.0$ in one experiment and $mR < 1.0$ in the dye-swap experiment or vice versa. This selection eliminates features showing a severe bias for Cy3- or Cy5-intensity.

**SYBR green RT-PCR.**

cDNAs were prepared as described above. The PCR was performed using iCycler iQ™ Multi-Color Real Time PCR Detection System (Bio-Rad, Hercules, CA). Well factors were collected using 20 µl 1× iCycler iQ™ external well factor solution (Bio-Rad). The PCR reactions were set up using LightCycler-DNA master SYBR Green I (Roche) according to the manufacture’s instruction. Each PCR reaction of 20 µl contained 1× reaction master, 10 pmol of each primer, 2 mM MgCl$_2$, and 1 µl of cDNA. The PCR was carried out for 35 cycles. For each cycle, the DNA was denatured at 95°C
for 20 seconds, annealed at 57°C for 20 seconds and extended at 72°C for 25 seconds. The PCR amplification was monitored each cycle at the extension stage to calculate threshold cycle. Each sample was quadruplicated and average was calculated. Fibronectin and GAPDH RT-PCR forward and reverse primers are located from different exons to avoid genomic DNA contamination. The PCR products were sequenced and confirmed. The fibronectin forward primer is 5’-GATGCTCCCCACTAACCTCCA-3’, reverse primer is 5’-CGCATCTCTTTTCCTGTCCAT-3’, and the product is 501 base pair. The GAPDH forward primer is 5’-TGGAAAGGACTCATGACCACA-3’ and reverse primer is 5’-TTACTCCTTTGGAGGCCATGT-3’, and the product size is 497 base pair.

4.3 Results

**Promoter methylation of BMP3B**

We previously identified BMP3B methylation in a screen for aberrant DNA methylation in NSCLC cancer (127). In order to determine the frequency of aberrant DNA methylation in this gene we studied the methylation pattern by bisulfite sequencing and subsequently designed a methylation specific PCR (MSP) assay. For bisulfite sequencing, samples included primary lung cancer patient No. 3 and 5 (PT3 and PT5) that showed methylation in the BstUI restriction sites by digestion of the PCR product using bisulfite sequencing primers and PT2, P14 and P19 with no aberrant methylation in the BstUI restriction sites (data not shown). PT3 and PT5 showed 62% and 73% methylation of the total CpG dinucleotides, respectively. In contrast, only 9% and 20% were methylated in the normal adjacent controls respectively (Figure 4.1A). PT2, PT14
and P19 showed methylation in less than 10% of the CpG dinucleotides (data not shown). In addition, the lung cancer cell line A549, used for the BMP3B reexpression studies described below, showed 50% methylation (Figure 4.1A).

In order to detect BMP3B promoter methylation in a large set of primary non-small cell lung cancers (NSCLC), methylation specific PCR (MSP) was performed on 20 paired primary non-small cell lung cancer samples and 76 non-small cell lung cancer samples. MSP primers were designed based on the bisulfite sequencing data described above to ensure their ability to discriminate between normal and tumor. Methylation was scored when the ratio of methylated band to unmethylated band within a sample was greater than the ratio in the 1% artificially methylated control. We found that 9 out of 20 (45%) paired samples (table 4.1), showed methylation in the tumor samples. The MSP results in PT2, 3, 5, 14, 19 was consistent with the bisulfite sequencing results described above. Representative examples are shown in figure 4.1B. Methylation was also detected in 4 out of 20 normal adjacent lung tissues, which might due to the contaminating preneoplastic cells. In addition, in a set of 76 NSCLC samples without paired adjacent normal lung tissue, 47% (36/76) showed methylation.

**Genetic alterations in BMP3B**

The BMP3B gene is found on chromosome 10q11.22 and located between markers D10S208 and D10S196. The order is D10S208 (on chromosome 10 short arm), centromere, BMP3B, D10S196 (http://genome.ucsc.edu, June 2002 freeze). BMP3B is 3.38 Mbp proximal to D10S196. LOH data with D10S196, was obtained from 17-paired patient samples. Results were non-informative in 4 cases. Seven patients’ tumors showed
no LOH while 6 tumors had LOH (35%) (Table 4.1).

Sequencing of the BMP3B coding region was used to search for possible mutations. No mutation was detected in the coding region of 83 tumor samples in our study. We found a polymorphism in a region containing 6 CTG repeats (position 502 to 519 in the published cDNA, NCBI accession number NM_004962.2) in exon 1. In frame heterozygous loss of one CTG repeat (figure 4.2) from one allele was identified in 4 out of 83 primary NSCLC samples, but also present in normal adjacent lung tissue. The shorter version of BMP3B results in the omission of a single leucine in the amino acid sequence. The functional consequences of this single amino acid deletion are undetermined.

The results of the LOH and MSP studies on the 20 paired primary lung cancer samples are summarized in table 4.1. The tumors from 12 patients showed either increased methylation or LOH. In particular, 1 of the 12 samples, PT17, showed both methylation and LOH.

Re-expression of BMP3B in A549 cell line using pBabe-puro-BMP3B

In order to determine the biological significance of BMP3B down regulation in human NSCLC lung cancer, full-length BMP3B coding sequence was constructed and ligated into retrovirus expression vector pBabe-puro. We started from IMAGE cDNA No. 308743 in pT7T3D-Pac vector. However, after sequencing we found that the this cDNA clone starts at position +186 of the coding sequence of BMP3B. The missing 5’-end of the BMP3B coding sequence was filled in by adding a PCR product from normal lung cDNA to make full-length coding sequence, which was then verified by sequencing.
The BMP3B full-length coding sequence cDNA was ligated into pcDNA3.1/hygro(+) vector. Second, the c-MYC tag was successfully amplified and placed in the 5’-end of the BMP3B cDNA. The construct was sequenced to confirm that both c-MYC tag and BMP3B were in the right reading frame. Finally, BMP3B full-length coding sequence with or without c-MYC tag was successfully ligated into pBabe retrovirus vector, which is verified by both restriction digestion and sequencing.

After infection of the A549 cell line with pBabe-puro, pBabe-puro-BMP3B or pBabe-puro-MYC-BMP3B retrovirus, the majority of the A549 cells survived after puromycin selection (3 μg/ml) for 5 days, while the non-infected A549 cells died in less than 48 hours. The infected cells were maintained in the culture medium supplemented with 3 μg/ml puromycin for all the assays.

Reexpression of BMP3B was demonstrated by RT-PCR (figure 4.3A). BMP3B was not detectable in A549 cells infected with vector only but was expressed in both pBabe-puro-MYC-BMP3B and pBabe-puro-BMP3B. The possibility of genomic DNA contamination in the mRNA could be excluded since no expression was detectable in the (-) RT control (Figure 4.3A). In addition, the expression at protein level was confirmed using anti-c-MYC monoclonal antibody for Western hybridization (figure 4.3B). The c-MYC tag protein was detected only in the A549 cells transfected with pBabe-puro-MYC-BMP3B.
A549 cell line in vitro and in vivo growth assay of pBabe-BMP3B infected A549 cell line.

Although neither obvious morphologic changes nor statistically significant growth rates were observed between the vector alone, pBabe-puro-MYC-BMP3B or pBabe-puro-BMP3B transfected groups (data not shown), we found that there were approximately 70% of cells in BMP3B expressing cells compared with vector alone at the beginning of the growth curve assay, and this trend remained throughout the assay. The anchorage dependent colony formation assays, repeated five times, showed results consistent with the growth curve assays, demonstrating approximately 60% and 70% of colonies formed in pBabe-puro-MYC-BMP3B and pBabe-puro-BMP3B compared with vector alone. There is a statistically significant difference between vector alone group and pBabe-puro-MYC-BMP3B, or vector alone group and pBabe-puro-BMP3B (P<0.05, t-test) (figure 4.4A). But no statistical significance was observed between pBabe-puro-MYC-BMP3B and pBabe-puro-BMP3B group (figure 4.4A).

The number of colonies formed in anchorage independent assay in soft agar was significantly decreased in the BMP3B transfected A549 cells. Only 10% of colonies formed when BMP3B was expressed from both the pBabe-puro-MYC-BMP3B and pBabe-puro-BMP3B constructs (figure 4.4B). Statistical significance between the vector alone group and pBabe-puro-MYC-BMP3B, or the vector alone group and pBabe-puro-BMP3B (P<0.005, t-test) was observed (figure 4.4B), but not between pBabe-puro-MYC-BMP3B and pBabe-puro-BMP3B group (P>0.1, t-test) (figure 4.4B). Similar results were observed in three independent experiments.

In the in vivo subcutaneous cell growth assay in nude mice, statistical
significance of average tumor volume between vector control group and pBabe-puro-BMP3B group was observed at 42 days of growth and the tumor volume remained significantly different up to day 70 (P<0.05, t-test) (Figure 4.4C). The average of tumor weight of the ten tumors growing subcutaneously was 46% lower in the pBabe-puro-BMP3B group compared with vector alone control group. The t-test indicates there is a statistically significant difference between the vector control group and pBabe-puro-MYC-BMP3B group (P<0.05) (data not shown).

**Expression profiling changes of A549 cells infected with pBabe-puro-BMP3B**

We have measured alteration of transcription caused by the expression of BMP3B in A549 cells, where no endogenous BMP3B could be detected. Using microarrays containing 9000 cDNAs, measuring about 4000 genes, we detected no alteration of transcription of a gene higher than three fold. Two- to three fold regulations is frequently claimed to be the limit of detection in cDNA microarrays (185). Due to this fact, we considered a gene to be a target of BMP3B only, if at least two independent clones showed the same direction of regulation in the dye-swap experiment. Table 4.2 lists the observed genes showing alteration of transcription due to the introduction of BMP3B in A549 cells. For example, three genes, fibronectin transcript variant 1 (FN1), transforming growth factor-β-induced, 68-kd (TGFBI) and stearoyl-CoA desaturase (delta-9-desaturase) (SCD), showed 1.5 to 2 fold decrease of expression that was demonstrated consistently by multiple clones for each gene on the array. SYBR green semi-quantitative PCR was performed to verify FN1 down-regulation.
Compared to vector control, pBabe-puro-BMP3B infected cells showed 1.4 cycles later to reach threshold cycles normalized by GAPDH, which indicated that fibronectin was approximately 3-fold down regulated by BMP3B expression at transcriptional level.

4.4 Discussion:

In the present study, we demonstrate that the CpG island containing BMP3B promoter is hypermethylated in approximately 45% NSCLC primary tumors (n=96) detected by MSP. The MSP assay was designed based on bisulfite sequencing and consistent with the bisulfite sequencing results. Although no mutation was found in the BMP3B coding sequence, it remains possible that BMP3B the region might be affected by other genetic changes, such as mutations in the 5’ and 3’ UTRs or splice sites. In addition, 35% of primary non-small cell lung tumors in this study (n=17) showed LOH in marker D10S196. Previously published data showed that D10S208/D10S196 was a region of frequent allelic loss (41%), in a study of 36 human lung cancer cell lines (186). Therefore, it is clear that the BMP3B locus is affected by both epigenetic and genetic changes, both leading to loss-of-function of the gene product. Furthermore, from a previous study of rat calvarial osteoblast primary culture, BMP3B was completely inhibited at the transcriptional level by TGF-β1, which indicated that transcriptional repression of BMP3B by inhibitory transcription factors would be another mechanism that inhibits BMP3B expression. Overexpression of TGF-β1 in non-small cell lung cancer compared to adjacent normal lung (187) might play a role in BMP3B down regulation in non-small cell lung cancer.
Our study shows that BMP3B could significantly repress anchorage independent cell growth in soft agar. Although no growth rate change was observed in growth curve assay, there were less cells that survived in BMP3B reexpressing A549 cells, which was confirmed in the anchorage dependent colony formation assay and the in vivo cell growth assay in nude mice. This is in agreement with previous findings that BMP3B expression was correlated with rat calvarial osteoblastic differentiation, apoptosis and suppressed growth of mouse embryonic submandibular glands (SMGs) inhibited by NF-κB inhibitor SN50. Our results also indicate that inhibition of colony formation might be mediated through affecting down stream targets that are involved in cell adhesion or extracellular matrix, such as fibronectin (discussed below), instead of genes that directly affect cell growth rate.

Our data also suggest that BMP3B and TGF-β1 might have opposing roles (figure 4.5). Similarly, BMP3, the closest family member of BMP3B, has also been observed having antagonistic roles against TGF-β1 in human bone marrow stoma during cell growth and differentiation (188). Furthermore, three genes suppressed from 1.5 to 2 fold by BMP3B identified in our microarray study have different functions, yet share the common feature that they are all up regulated by TGF-β1. First, fibronectin in A549 lung cancer cell line was significantly induced by TGF-β (189); Based on the gene expression data, Ridley et al. point out that fibronectin might promote migration of cancer cells (190). In addition, fibronectin was up-regulated during Adenovirus E1A 12S mutant HBdl12 induced epithelial mesenchymal transition in primary kidney cells (191). Second, transforming growth factor-β-induced, 68-kd (TGFBI) is a gene that was originally
identified by differential hybridization, as a target that could be induced up to 20 fold increase after 2 days with TGF-β treatment of A549 (192). TGFB1 was also found to be overexpressed in lung cancer and was correlated with later stage of lung cancer (193). Third, stearoyl-CoA desaturase can be induced by TGF-β1, 2 or 3 in cultured human retinal pigment epithelial cells (194). Our data suggested that some TGF-β1 targets could be inversely regulated by BMP3B. Taking together with the complete silencing of BMP3B by TGF-β1 in rat calvarial osteoblasts, we postulated that BMP3B and TGF-β1 might be involved in a same complex network of signaling pathway, such as SMAD pathway that has been shared by TGFs and BMPs, with negative regulation of each other.

In conclusion, our preliminary study and published data suggest antagonistic role of BMP3B and TGF β1 in some degree. BMP3B could be silenced in lung cancer by promoter methylation, genetic change or transcriptional repression. Since TGF-β signaling through SMAD pathway is involved in modulating epithelial mesenchymal transition, including the changes of extracellular matrix and adhesion (175, 195), to promote tumor invasion and metastasis, BMP3B silence in lung cancer might cause the loss of its regulatory role against TGF-β, which finally promotes tumor progression.
<table>
<thead>
<tr>
<th>PT No.</th>
<th>D10S196 LOH</th>
<th>Methylation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Non-informative</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>Normal</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>Normal</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>Normal</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>LOH</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>Non-informative</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>Non-informative</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>Normal</td>
<td>No</td>
</tr>
<tr>
<td>11</td>
<td>LOH</td>
<td>No</td>
</tr>
<tr>
<td>13</td>
<td>Normal</td>
<td>No</td>
</tr>
<tr>
<td>14</td>
<td>LOH</td>
<td>No</td>
</tr>
<tr>
<td>15</td>
<td>Normal</td>
<td>No</td>
</tr>
<tr>
<td>17</td>
<td>LOH</td>
<td>Yes</td>
</tr>
<tr>
<td>18</td>
<td>Normal</td>
<td>No</td>
</tr>
<tr>
<td>19</td>
<td>LOH</td>
<td>No</td>
</tr>
<tr>
<td>20</td>
<td>Non-informative</td>
<td>Yes</td>
</tr>
<tr>
<td>21</td>
<td>LOH</td>
<td>No</td>
</tr>
<tr>
<td>22</td>
<td>Not done</td>
<td>Yes</td>
</tr>
<tr>
<td>23</td>
<td>Not done</td>
<td>Yes</td>
</tr>
<tr>
<td>24</td>
<td>Not done</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 4.1. Summary of LOH of D10S196 and methylation detected by MS-PCR.
<table>
<thead>
<tr>
<th>Gene name</th>
<th>Accession #</th>
<th>Repression by BMP3B*</th>
</tr>
</thead>
<tbody>
<tr>
<td>fibronectin transcript variant 1, (FN1)</td>
<td>NM_002026</td>
<td>2.0</td>
</tr>
<tr>
<td>stearoyl-CoA desaturase, (SCD)</td>
<td>XM_005719</td>
<td>1.7</td>
</tr>
<tr>
<td>transforming growth factor-β-induced, 68-kd (TGFBI)</td>
<td>M77349</td>
<td>1.5</td>
</tr>
<tr>
<td>small cell lung carcinoma cluster 4 antigen, (CD24)</td>
<td>NM_013230</td>
<td>1.4</td>
</tr>
<tr>
<td>clone IMAGE:3625286</td>
<td>BC014103</td>
<td>1.4</td>
</tr>
<tr>
<td>keratin 18 (KRT18)</td>
<td>XM_029624</td>
<td>1.3</td>
</tr>
<tr>
<td>amyloid beta precursor-like protein 2 (APLP2)</td>
<td>NM_001642</td>
<td>1.3</td>
</tr>
<tr>
<td>RNA polymerase II transcription cofactor 4 (PC4)</td>
<td>NM_006713</td>
<td>0.8b</td>
</tr>
</tbody>
</table>

Table 4.2. Alteration of gene expression due to expression of BMP3B measured by cDNA microarrays. Alteration of gene expression due to expression of BMP3B measured by cDNA microarrays; * the degree of repression is calculated as the mean of repression measured by all cones representing the indicated gene on the microarray; b a 0.8 fold repression is equivalent to a 1.2 fold induction.
Figure 4.1. BMP3B promoter methylation detected by bisulfite sequencing and methylation-specific PCR. (A). Bisulfite sequencing of BMP3B promoter region containing 28 CpG dinucleotides for Patient 3 (PT3), PT5 and cell line A549. Open circle indicates no methylation of cytosines in CpG dinucleotides. Filled circle, methylated cytosine in CpG dinucleotides. (B) Representative methylation-specific PCR to detect methylation in primary non-small lung cancer patients. Each sample was amplified using both non-methylation (U) and methylation (M) specific primers. UT: genomic DNA without bisulfite treatment. 50%: artificially methylated 50% methylation control. N: normal adjacent tissue. T: tumor.
Figure 4.2. CTG repeat polymorphism in BMP3B exon 1. Top sequence result shows that both alleles, with a (CTG)$_6$ repeat encoding 6 leucines. Bottom patient sample was heterozygous for the CTG repeat, one allele had a (CTG)$_6$ repeat, while the second allele had (CTG)$_5$ repeat causing a deletion of one leucine.
Figure 4.3. Detection of reexpression of BMP3B in A549 cell line. (A). RT-PCR shows that BMP3B was reexpressed in both pBabe-MYC-BMP3B and pBabe-BMP infected cells. No expression was detectable in cells infected with vector alone. RT: reverse transcriptase. In vitro transcription without reverse transcriptase was used as a control to avoid genomic DNA contamination. (B). Western blotting using anti-c-MYC tag. The c-MYC was expressed in the A549 cell line that was infected with pBabe-MYC-BMP3B, which indicates that BMP3B was expressed as a fusion protein.
Figure 4.4. *In vitro* and *in vivo* cell growth assay of A549 cell line after BMP3B reexpression. (A). Anchorage dependent colony formation. The bar graph shows the average from experiments that had been repeated 5 times. X-axis gives the colony numbers. The error bar indicates the 95% confidence intervals to each mean from triplicate repeats. The right side figure shows the representative colonies formed on the cell culture plates. (B). Anchorage independent colony formation assay in soft agar. The left side bar graph shows the numbers of colonies formed after BMP3B reexpression compared to vector control for three times. The 95% confidence intervals to each mean from triplicate repeats was shown as the error bar. The right side figure shows the colonies formed in soft agar. (C) *In vivo* cell growth assay in nude mice. The volume was calculated as width (mm) × length (mm) × height (mm) / 2. Ten mice were used for pBabe and pBabe-BMP3B groups respectively.
A

Colony Formation Assay

B

Soft Agar Assay

C

Cell Growth In Nude Mice
Figure 4.5. Proposed model of opposing role of TGF-β1 and BMP3B. TGF-β1 has been shown to induce epithelial mesenchymal transition (EMT) during the late stage of tumor development, which would promote tumor invasion and metastasis (196). During the rat calvarial osteoblasts differentiation in primary culture, BMP3B expression was elevated with differentiation, but was completely repressed when TGF-β1 was added to the culture medium. In addition, the three genes that showed 1.5 to 2 folds suppression by BMP3B at mRNA level in our study are inducible targets of TGF-β1. Combining this information, we propose that BMP3B and TGF-β1 might have opposing role. BMP3B might be involved in the reverse pathway of epithelial mesenchymal transition that could be induced by TGF-β1, mesenchymal epithelial transition, a process involved in normal lung differentiation.
CHAPTER 5

NOVEL AMPLICONS IN HUMAN LUNG CANCERS IDENTIFIED BY RESTRICTION LANDMARK GENOMIC SCANNING

5.1 Introduction:

Gene amplification, as a mechanism leading to overexpression of proto-oncogenes, plays a major role in cancer development. Numerous oncogenes have been identified and their characterization has led to a better understanding of tumorigenesis. Activated oncogenes are important prognostic factors. Examples include HER2/neu in breast cancer (197-199), MYCN in neuroblastoma (200, 201), and MYCC in small cell lung cancer (33, 34). In addition, amplified genes might become molecular targets for cancer treatment, as recently demonstrated for HER2/neu in breast cancer (199, 202).

Amplification of the MYC-family oncogenes (MYCC, MYCN and MYCL1) in human lung cancers has been well characterized (32). MYCC is amplified in both small cell lung cancers (SCLC) and non-small cell lung cancers (NSCLC), while MYCN and MYCL1 amplifications are usually limited to SCLC (32). The frequency of amplification of any of the MYC-family members is 18% in SCLC primary tumors and 31% in SCLC
cell lines, respectively. MYCC is amplified in 7% of primary NSCLC tumors and 20% of NSCLC cell lines. Although MYC gene amplification might be a rare event, it was more frequently observed in lung cancer cell lines and primary lung cancer following chemotherapy (34, 203). Other oncogenes are also amplified at low frequency (204). Cyclin D1 and EGFR are amplified in 5% and 6% of NSCLC respectively (35), HER2/neu in 4 out of 344 NSCLC (205), MDM2 in 2 out of 30 NSCLC (206). However, the decoy receptor 3 (DcR3) gene, located on chromosome 20q13 that blocks FAS ligand induced apoptosis was found amplified in 8 out of 16 primary lung tumors (207).

The chromosome 3q26-27 amplicon has recently been identified in non-small cell lung carcinoma by comparative genomic hybridization (CGH) (41). This amplicon is observed mostly in squamous cell carcinomas (SCC) of the lung and rarely found in adenocarcinomas (208, 209). The region was found to be the most frequently amplified in SCC using array comparative genomic hybridization (array-CGH) and appeared to be as large as 30 Mbp (42). Multiple candidate oncogenes including PIK3CA (42) and amplified in SCC (AIS) have been identified (210).

In order to identify novel amplicons in cancer, several techniques have been used in different studies, including comparative genome hybridization (CGH) (41), restriction landmark genomic screening (RLGS) (87), and a newly developed CGH on arrayed BAC (42) or cDNA clones (211). RLGS, as a two-dimensional gel electrophoresis, has the capacity to perform genome-wide scanning for both novel methylated target as well as for DNA amplification (89). Intensities of RLGS fragments correlate precisely with gene copy numbers since RLGS profiles display end labeled restriction ends. The increased copy number leads to an enhanced RLGS fragment as compared to the one found in the
normal tissue DNA. RLGS has been successfully used in variety of human primary malignancies (87, 88, 102, 103), human cancer cell lines (104-106) as well as mouse tumors (107, 108).

Here we report a genome scan for amplified sequences in lung cancer by RLGS. Four amplified regions have been identified and characterized in primary lung cancers. Among them, two are known amplicons, MYCL1 amplicon and an amplicon on 3q26-27. The other two amplicons identified on chromosomal 6q21 and 11q22 are novel in primary lung cancers. Detailed characterization of the 11q22 amplicon allowed us to define the amplified region for the further identification of candidate oncogenes in lung cancer.

5.2 Materials and methods

Primary lung cancer tissue samples and cell lines:

Primary human lung cancer and normal adjacent tissue samples were collected through the Cooperative Human Tissue Network at The Ohio State University, James Cancer Hospital. A total of twenty NSCLCs were studied, sixteen of them had previously been used for DNA methylation profiling (127). These tumors included 8 squamous cell carcinomas, 6 adenocarcinomas, 4 large cell carcinomas and 2 uncharacterized NSCLC. In addition, 5 small cell lung cancers (SCLCs) obtained from Julian Molina, Mayo Clinic, Rochester, MN, were included in the study.

Nine non-small cell lung cancer cell lines, H23, H125, H522, H1155, H1299, H2009, H2086, H2172 and A549, and 8 small cell lung cancer cell lines, H69, H82,
H209, H211 H719, H792, H841 and N417, were obtained from ATCC and cultured in RPMI-1640 medium (Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum, 100 units/ml Penicillin and 0.1mg/ml Streptomycin (Invitrogen).

**Isolation of high molecular genomic DNA**

Frozen tissues were cut into small pieces and suspended in 1 to 2ml lysis buffer (0.15 M EDTA pH 8.0, 10 mM Tris-HCl pH 8.0 and 1% Sarkosyl). The samples were then snap frozen in liquid nitrogen prior to crushing. Five to 15ml pre-warmed lysis buffer containing proteinase K (Roche, Indianapolis, IN) at final concentration 100 μg/ml was added to the samples, mixed and incubated at 55°C for 20 to 30 minutes. After cooling on ice for 10 minutes, an equal amount of phenol, chloroform and isoamyl alcohol (in a ratio of 50:49:1) was added and the sample was rotated gently for one hour. After centrifugation for 20 minutes at 2500 rpm, the top aqueous layer of sample was collected and the viscous sample was again extracted with PCI. DNA samples were dialyzed using dialysis tubing (Gibco) in 4L 10mM Tris pH 8.0 initially for an hour then overnight followed by a final dialysis for one hour at room temperature. The samples were further treated with RNase A (1 μg/μl) for 2 hours. After adding 2.5 volumes of 100% ethanol, the samples were rotated gently for one hour. The precipitated DNA was transferred to a microcentrifuge tube and briefly air dried, redissolved in 1× TE buffer to a final concentration of approximately 1μg/μl and stored at 4°C.
Restriction landmark genomic scanning (RLGS)

RLGS was performed according to published protocols (89, 127). In summary, randomly fragmented ends of high molecular weight DNA were blocked using DNA polymerase I (Boehringer Mannheim), followed by a restriction digestion of the DNA with NotI (Promega, Madison, WI). NotI restriction ends were end-labeled using [α-32P]-dGTP and [α-32P]-dCTP (Perkin Elmer, Foster, CA) in the presence of Sequenase (Amersham, Piscataway, NJ). The DNA was further digested with the restriction enzyme EcoRV (Promega). For first dimension separation, NotI-EcoRV DNA fragments were separated by their molecular weights through an 0.8% agarose gel. After an in-gel digestion using restriction enzyme HinfI (New England Biolabs, Beverly, MA), the second dimension separation was carried out on a 5% polyacrylamide gel. The dried gels were exposed to X-ray film for 5-10 days. The RLGS profiles were superimposed and visualized to compare primary tumors and normal adjacent lung tissues to detect the enhanced RLGS fragments that might represent amplified DNA. The RLGS profiles were exposed to phosphoimager screen for quantification. All the RLGS fragments were named according to the master profile (see online: http://pandora.med.ohio-state.edu/masterRLGS).

RLGS fragment cloning

Clones corresponding to RLGS fragments were identified using the NotI-EcoRV plasmid library and library RLGS mixing gels (89, 94). The plasmid clones corresponding to the RLGS fragments in the library are determined by identifying
enhancements in the plate, row and column mixing gels. Plasmid DNAs from candidate clones were isolated using Qiagen miniprep kit (Qiagen, Valencia, CA). After digestion with $\text{NotI-EcoRV}$, plasmid DNA was end labeled with $[\alpha^{32}\text{P}]-\text{dGTP}$ and $[\alpha^{32}\text{P}]-\text{dCTP}$ (Perkin Elmer). Prior to first dimension loading, radioactively labeled plasmid DNA clones (10pg per clone) were mixed with an appropriate amount of labeled genomic DNA. The resulting enhancements of the expected RLGS fragments confirmed proper cloning.

**Database search.**

Once the plasmid clones were confirmed as the corresponding RLGS fragments, the clones were sequenced in the Core Facility of the Division of Human Cancer Genetics using M13 forward primer. The NCBI database, [http://www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/), and UCSC human genome project working draft database, [http://genome.ucsc.edu/](http://genome.ucsc.edu/), were searched for sequence homologies and chromosomal locations.

**Southern hybridization**

Probes were made from either the restriction enzyme digested fragments of the plasmid clones or using PCR products. Four $\text{NotI-EcoRV}$ clones (IPF1, 5E14, 2F72 and 2F41/42), with sequences listed in table 5.1, were digested with appropriate restrictions enzymes. The IPF1 probe is a 493 bp $\text{XbaI-EcoRV}$ fragment; the 5E14 probe is a $\text{XbaI-EcoRV}$ fragment of 1426 bp; the 2F72 is a 381 bp $\text{PstI-BamHI}$ fragment; and the 2F41/42 probe is a 278 bp $\text{BglII-NotI}$ fragment.
All other probes were made by PCR amplification of genomic DNA. The primers are listed in table 5.1. Each PCR was performed in 50 µl final volume containing 20-50ng genomic DNA, 1x PCR buffer, 1.5mM MgCl₂, 10 pmol of each primer, 200 µM of each dNTP and 2.5 U Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA). The PCR was initiated in RoboCycler® Gradient 40 PCR machine (Stratagene, La Jolla, CA) at 95°C for 10 minutes to activate DNA polymerase, followed by 35-cycle amplification. For each cycle, the template was denatured at 96°C for 40 seconds, annealing was performed at temperatures listed in table 5.1 for 1 minute and extension was at 72°C for 1 minute. One microliter of the PCR product was used as template for a second PCR amplification at the same condition to reduce the genomic DNA contamination. The second round PCR product was separated on an agarose gel, cut out and purified using QIAquick® gel extraction kit (Qiagen, Valencia, CA). All the probes were labeled with [α-32P]dCTP using Prime-It® II Random Primer labeling kit (Stratagene). Southern hybridization followed the published protocol (89). The Southern blots were exposed to phosphorimager screen and quantified.

5.3 Results

A genome-wide search for amplified sequences in human lung cancers by RLGS.

To identify novel regions of DNA amplification in human lung cancers, 20 paired normal tissue and primary NSCLC, 9 NSCLC cell lines, 5 primary SCLC and 8 small cell lung cancer cell lines, were analyzed by RLGS. RLGS profiles of primary lung cancers were compared with a profile derived from normal lung tissue from the same patient. All
enhanced RLGS fragments were named according to the RLGS master profile addresses that had been previously established (212). Two types of enhancements were identified in the tumor profiles. The first type was seen in RLGS fragments that were also present in the normal profiles with diploid copy number intensity. For example, the intensity of 5E14 on the normal lung profile of patient (PT) 2 is the same as the intensity of surrounding fragments and represents a single-copy sequence while the intensity of 5E14 is enhanced in the tumor of PT2 (Figure 5.1A). The second type of enhancement in tumor profiles was seen in RLGS fragments that were not present in the normal profile (Figure 5.1B). The first type represents DNA amplification of a single copy sequence, the second type most likely represents hypomethylation of a normally methylated repeat sequences that contains a NotI site (88, 213). The intensified RLGS fragments of the first type that have been identified in at least one patient are listed in table 5.2. Enhancements were observed in both NSCLCs and SCLCs. Six out of 20 NSCLC tumors (PT1, PT2, PT6, PT11, PT14 and PT24) and 1 of the 5 SCLC primary tumors (PT12) showed amplification. RLGS fragments 1G20, 3D17 and 3D21, 3G68 and 5GXX were enhanced in two primary lung cancers, while all other fragments were found only in one tumor. Three NSCLCs (PT1, PT11 and PT14) showed enhancement of only a single RLGS fragment, while the other four tumors (PT2, PT6 PT12 and PT24) showed multiple enhanced RLGS fragments. For example, RLGS fragments 1G20, 2E46, 2F72, 3E46 3E58, 3D17 and 3D21 were present with low-level but equal intensity in PT6 (figure 5.1C, showing 2F72, 3E46 3E58, 3D17 and 3D21), indicating that they might be derived from a single amplicon. Interestingly, the tumor from PT24 showed enhancement in a subset of these fragments (1G20, 3D17 and 3D21), indicating that this tumor has a
smaller amplicon size. It is noteworthy that both tumor samples with amplification (2 out of 8; 25%) are classified as squamous cell carcinoma (SCC), and that amplification was not seen in other subtypes of lung cancer. The tumor from PT12 also showed several amplified sequences (fragments 2F41/42, 2F50, 3D53, 4D13 and 5C25), however, intensities of those fragments vary, suggesting that they are derived from different amplicons (figure 5.1D, 3D53, not shown).

Each of the lung cancer cell line profiles were compared with two normal lung profiles. Overall, cell lines showed a higher rate of enhancements as compared to primary tumors. Among the 17 cell lines, 8 out of the 9 NSCLC cell lines, and all 8 SCLC cell lines showed amplified sequences (table 5.3). The majority of fragments showed low-level (3- to 5-fold) enhancement on RLGS profiles as measured by phosphorimager quantification (Figure 5.2B, 5.2C). For instance, 2F36 was enhanced approximately 5-fold in cell line H841 (figure 5.2B). Three cell lines, H23, H125 and H82, showed numerous enhanced sequences at low levels (Figure 5.2C for H125), possibly as a result of whole chromosomal gains, or gains of large chromosomal regions. In contrast to low-level amplifications seen in the majority of the enhanced RLGS fragments, fragments 3B35 and 4C3 were amplified in cell lines H23, H82, N417 and H792 with copy numbers greater than 20. Figure 5.2A represents a portion of the RLGS profile with 3B35 and 4C3 enhancements from the H82 cell line. Seven RLGS fragments (2C59, 2F36, 2G55, 3B35, 3G68, 4C3 and 5C1) were enhanced in more than one cell line (table 5.3). Moreover, three RLGS fragments (2F51, 3G68 and 5E14) were found to be amplified in both primary tumors and cancer cell lines. All other amplifications were found exclusively in primary tumors or in cell lines.
Cloning of RLGS fragments

In order to further characterize amplified sequences in lung cancer, it was critical to clone and sequence the enhanced RLGS fragments. Subsequent database searches allowed the mapping of sequences to specific chromosome regions and to identify genes in the vicinity. Some of the sequences (3B35, 3E24, 3C71, 3G68, and 4C3) had been identified in previous studies (88, 106). Fragments 3B35 and 4C3, which are amplified in cell lines H23, H82, N417 (214) and H792, are sequences located in the MYCC gene region on chromosome 8q24 (88). The finding of MYCC amplification in these lung cancer cell lines confirms previous work (214).

In order to identify novel amplified sequences in lung cancer we decided to clone additional fragments, enhanced in primary lung tumors. The NotI-EcoRV plasmid library and corresponding library mixing gels were used to facilitate the cloning (89) using the same strategy described earlier (127). Table 5.4 provides a list of all cloned RLGS fragments and summarizes the information obtained from database searches, including the full-length NotI-EcoRV sequences, chromosomal locations, genes and ESTs related to the NotI-EcoRV sequences. When applicable, sequences in table 5.4 were ordered according to chromosomal location.

Three groups of sequences could be identified from database and literature searches. The first group includes sequences that fell into known regions of DNA amplification in lung cancers. For example, amplified fragments from tumor PT6 (specifically 3D17, 3D21, 1G20, 2F72 and 3E58) are all located on chromosome 3q26-27, a region known to be amplified in SCC of the lung (42), head and neck cancer (215),
cervical and ovarian cancers (216, 217). Fragments 3D17, 3D21 and 1G20 were also amplified in SCC tumor PT24, which indicates that the amplicon in this tumor might be smaller than in PT6. Fragments 2F41/42 and 2F50, amplified in PT12, were mapped to chromosome 1p34 approximately 1 Mbp away from the MYCL1 gene.

The second group of sequences fell into chromosomal regions for which candidate oncogenes had been identified in other types of cancers, but not in lung cancer. For instance, 5E14 was located on chromosomal 11q22, a 1.8 Mbp region that has been reported to be amplified in primary esophageal cancers and cancer cell lines as well as other types of cancer (218). To our knowledge, this chromosomal region has not been well characterized in lung cancer.

The third group of sequences, 4D13 and 5C25 amplified in PT12, are located on chromosome 6q21, a novel region of amplification in lung cancer. To our knowledge, the 6q21-22 region has been identified as chromosomal gains or amplifications in primary breast cancer by CGH study (219) and in the breast cancer cell line BT-20 (220) by fluorescence in situ hybridization. However, no well-defined amplicon or specific oncogene has been reported in this region in any type of cancer.

**Confirmation of DNA amplification by Southern hybridization**

In order to further confirm that the cloned RLGS sequences represent amplified fragments, probes from the corresponding sequences were prepared for Southern hybridizations. Either the RLGS fragments, or adjacent gene sequences, were used as probes. A probe for insulin promoter factor 1 (IPF1) gene located on chromosome 13q12.1 was used as DNA loading control for all the hybridizations. We confirmed DNA
amplification for four chromosomal regions for which RLGS fragments had been identified. Southern hybridization for the amplicon on chromosomal region 3q26-27, identified in PT6 and PT24, shows approximately a 5-fold (3D17) or 10-fold (2F72) increased intensity in the tumor of PT6 as compared to the normal adjacent tissue (figure 5.3A). Southern hybridization using probe 2F41/42 for the second amplicon on chromosome 1p34, showed approximately 7-fold enhanced intensity (figure 5.3B). This region includes MYCL1, which is amplified approximately 12-fold. Amplification of chromosome 11q22 is demonstrated using cIAP1 as probe for Southern hybridization (see Figure 5.3C, and below). We found that the region was amplified approximately 7-fold in PT2, and 5-fold in cell line H2086. The fourth genomic region on 6q21 showed DNA amplification up to 14-fold in PT12 (Figure 5.3D). This region includes the popeye protein 3 (POP3), blood vessel epicardial substance (BVES), and EST LOC135537 and KIAA1320.

**Characterization of the 11q22 amplicon**

Among the four amplicons described above, the 1p34 and 3q26-27 amplicons are known regions of DNA amplification in lung cancer, and include the known oncogene MYCL1 on chromosome 1p34 (221), and multiple candidate oncogenes from chromosome 3q26-27 (210, 222). The amplicon on 6q21 is located in a region that had not yet been refined in any human cancer. The amplicon on 11q22 was also found to be amplified in esophageal cancer within a 1.8 Mbp region (218). Although a big amplified region from 11q13 to 11qter has been reported in the NSCLC cell line RVH6489 (223), 11q22 had not been reported previously as a clearly defined amplicon in lung cancer.
Using sequences from the NCBI database, human chromosome 11 working draft sequence segment NT_009151.8, we found that the RLGS fragment 5E14 is located in the area on 11q22 that has been completely sequenced (Figure 5.4). The region contains multiple members of matrix metalloproteinases (MMPs), inhibitors of apoptosis (cIAP1, cIAP2), YAP1, porimin and some uncharacterized genes (figure 5.4). In order to define the core amplified region, a set of probes (table 5.1) were designed for Southern hybridizations (Figure 5.4 and 5.5). The intensity of the Southern hybridization signals was quantified using phosphorimager and listed in figure 5.5. The amplicon in PT2 could be defined to a region extending from LOC143867 to MMP1. The amplicon in cell line H2086 extends the one of PT2 (Figure 5.5). However, PT7 showed low level amplification (or duplication) of a region localized between markers MMP7 and LOC143868.

In addition, Southern hybridization was performed on another twenty-four non-small cell lung cancers to detect cIAP1 gene amplification, however, no amplification was detected (data not shown), which indicated the amplification of 11q22 in lung cancer might be infrequent, similar to the observation in other types of cancer (218).

5.4 Discussion:

In the present study, we have performed restriction landmark genomic scanning to identify DNA amplification in human lung cancer. The characterization of the enhanced RLGS fragments in primary lung cancer identified amplicons on 1p34, 3q26-27, 11q22 and 6q21 in lung cancer. Amplification of 1p34 had previously been described and this
amplicon contains the MYCL1 oncogene (214). Amplification of 3q26-27 had been reported in about 25-40% of patients with squamous cell carcinoma of the lung (209, 224, 225), and only very rarely in adenocarcinoma (208, 209). Consistent with these published reports, we found amplification of 3q26-27 in two out of eight squamous cell carcinomas. The cloned fragments derived from this amplicon (3D17, 3D21, 1G20, 2F72 and 3E58), are located in an area extending 30 Mbp, including several candidate oncogenes in lung cancers (42, 222, 226). For example, PIK3CA, a gene associated with phosphatidylinositol-3-kinase pathway, is a candidate oncogene in both lung cancer (42) and cervical cancer (217). Amplification of 3q26-27 has been found in a variety of human cancers, including head and neck squamous cell carcinoma (215, 227, 228), cervical cancer (217), primary carcinoma of the fallopian tube (229), osteosarcomas (230), ovarian cancer (216, 231), and nasopharyngeal carcinoma (232). In head and neck squamous cell carcinomas and in ovarian carcinomas, amplification correlates with tumor progression (228, 231).

Amplification of 6q21, found in SCLC, has also been described in breast cancer, but not in lung cancer (219, 220). This amplicon is currently not well defined but includes the known genes prolyl endopeptidase (PREP), blood vessel epicardial substance (BVES) and popeye protein 3 (POP3). PREP is an enzyme involved in the maturation and degradation of peptide hormones and neuropeptides, and displays increased activity in human lung tumors compared to normal lung (233). BVES and POP3 are expressed in heart and skeletal muscle and are essential for coronary vasculogenesis (234-236). However, currently, no candidate oncogene has been characterized.
To our knowledge amplification of 11q22 had only been described in a NSCLC cell line, RVH6489, extending from 11q13 to 11qter (223). However, the amplicon was previously reported in esophageal squamous cell carcinoma cell lines KYSE2270 and KYSE170 and primary tumors (218). The authors described amplification in BACs RP11-21G19 to 23P6. Based on the current human sequences in this region (NT_009151.18) this region is approximately located between LOC160122 to MMP13 (Figure 5.5). Therefore, the region that we identified in PT2 is overlapping with the amplicon in esophageal squamous cell carcinoma. The shared amplified region between our samples and the esophageal cancer cell lines harbors several genes including, inhibitor of apoptosis (cIAP1 and cIAP2), MMPs, promin and additional novel genes.

Based on expression studies as well as functional analysis, cIAP1 has been proposed as candidate oncogene in the 11q22 amplicon in esophageal squamous cell carcinoma (218). Ectopic expression of cIAP1 in mammalian cells inhibited apoptosis induced by serum deprivation (237). It was also shown that expression of cIAP1 resulted in the resistance of apoptosis induced by chemotherapy drugs (218), and correlated with radiation resistance in cervical cancer (238).

In addition, other known genes might also play important roles in carcinogenesis, such as cIAP2, YAP1 and members of the matrix metalloproteinase family. For example, cIAP2 was over-expressed in human malignant pleural mesothelioma and the high expression of cIAP2 correlates with drug resistance in cancer cells (239). YAP1 was also identified as a gene that is preferentially expressed in transformed and metastatic tumor cell lines of murine squamous cell carcinoma compared with primary keratinocytes (240). The MMP family plays a critical role in
tumor progression (241). ESTs, such as LOC160122, that might relate to angiopoietin, are also worthy of further study.

In conclusion, RLGS has been successfully utilized to identify both known and novel amplicons in human lung cancer. Future studies will focus on the identification of candidate oncogenes in the amplified regions, which may contribute to a better understanding of the oncogenic process in lung.
<table>
<thead>
<tr>
<th>Probe Name</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Annealing °C</th>
<th>Probe size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>11q22</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC143868</td>
<td>5'-TTGCTTGCACGTGCCATCATAC-3'</td>
<td>5'-CCTGCCAAGCCAAATCTGAG-3'</td>
<td>58</td>
<td>497 bp</td>
</tr>
<tr>
<td>LOC143867</td>
<td>5'-ATCCCTCTTATATCTCCATGCA-3'</td>
<td>5'-CAGCACTTGGTCTGCTT-3'</td>
<td>58</td>
<td>410 bp</td>
</tr>
<tr>
<td>LOC143866</td>
<td>5'-AGTGCAACAGGCAAGTTTAC-3'</td>
<td>5'-CTTACTCCAGAGCTTACAGAT-3'</td>
<td>58</td>
<td>401 bp</td>
</tr>
<tr>
<td>LOC120317-1</td>
<td>5'-ACAATTCTGTTGATGCTCAG-3'</td>
<td>5'-TTACGTGAGATGTCTAC-3'</td>
<td>58</td>
<td>479 bp</td>
</tr>
<tr>
<td>LOC120317-2</td>
<td>5'-GAACGGTGTTACCTTTGGA-3'</td>
<td>5'-AATCATGGGAACCTGCAAACC-3'</td>
<td>58</td>
<td>484 bp</td>
</tr>
<tr>
<td>LOC160122</td>
<td>5'-TTTCTCAGGCGCTTTTCTT-3'</td>
<td>5'-TGGAGAAGTGTCTCATAATC-3'</td>
<td>58</td>
<td>505 bp</td>
</tr>
<tr>
<td>cIAP2</td>
<td>5'-TTTATCTTAAATTTGTTTTC-3'</td>
<td>5'-AATTTTAAAGTTAATCTC-3'</td>
<td>50</td>
<td>253 bp</td>
</tr>
<tr>
<td>cIAP1</td>
<td>5'-ACATTTGATGTTGATGCTCAG-3'</td>
<td>5'-TTATGAGATGTCTAC-3'</td>
<td>50</td>
<td>366 bp</td>
</tr>
<tr>
<td>MMP7</td>
<td>5'-TCTTCCAGAGGGGTCTTGG-3'</td>
<td>5'-AGGATAGGAAACCTACAGCA-3'</td>
<td>58</td>
<td>603 bp</td>
</tr>
<tr>
<td>MMP20</td>
<td>5'-CTTGTCTGCTCAGTCTCAG-3'</td>
<td>5'-CTTGTCTGCTGCTCAG-3'</td>
<td>58</td>
<td>602 bp</td>
</tr>
<tr>
<td>MMP8</td>
<td>5'-GACAGAAATGGAGCGCTTCT-3'</td>
<td>5'-ACCACATCCTCCAGAAGCAG-3'</td>
<td>58</td>
<td>600 bp</td>
</tr>
<tr>
<td>MMP10</td>
<td>5'-TTATTTCTGGTGTCTCTCA-3'</td>
<td>5'-CTTTCCTCAGAGCTGAATG-3'</td>
<td>58</td>
<td>468 bp</td>
</tr>
<tr>
<td>MMP1</td>
<td>5'-ATGCTGAAACCTGTGAGGTG-3'</td>
<td>5'-AAAGGCAACAGAGCTTCCA-3'</td>
<td>58</td>
<td>501 bp</td>
</tr>
<tr>
<td>LOC120321</td>
<td>5'-AGGCAAAGTGAAAGAGGA-3'</td>
<td>5'-TTTGGCCCTTATGAAGGCAC-3'</td>
<td>58</td>
<td>398 bp</td>
</tr>
<tr>
<td>MMP3</td>
<td>5'-ACCATTGCGTCTCCTCCT-3'</td>
<td>5'-GACAGTGAGCAGGAC-3'</td>
<td>58</td>
<td>506 bp</td>
</tr>
<tr>
<td>Porimin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3q26-27</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D17</td>
<td>5'-GGAGAGAAACACGCACAC-3'</td>
<td>5'-CTGCTTCTCCAGTCTGAGC-3'</td>
<td>58</td>
<td>455 bp</td>
</tr>
<tr>
<td><strong>1p34</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYCL1</td>
<td>5'-GGCATGCAGAAAGGCTTATT-3'</td>
<td>5'-TGCAATCCACCCAGAGTGA-3'</td>
<td>58</td>
<td>542 bp</td>
</tr>
<tr>
<td><strong>6q21</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC135357</td>
<td>5'-CGGCTTCAAGTATGGCATT-3'</td>
<td>5'-GAGGCCCTACCTCGTTCAGC-3'</td>
<td>58</td>
<td>504 bp</td>
</tr>
<tr>
<td>BVES</td>
<td>5'-CCCTTACCTCCCTTCCTTCTT-3'</td>
<td>5'-ACCAACACCTGCAAACAACG-3'</td>
<td>58</td>
<td>502 bp</td>
</tr>
<tr>
<td>KIAA1320</td>
<td>5'-GCTTACCTGCGTCAACAA-3'</td>
<td>5'-GCATGGCAAAATTGGATG-3'</td>
<td>58</td>
<td>441 bp</td>
</tr>
</tbody>
</table>

Table 5.1. Primers used for probe preparations.
<table>
<thead>
<tr>
<th>RLGS Fragment</th>
<th>NSCLC primary tumor</th>
<th>SCLC primary tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AC</td>
<td>SCC</td>
</tr>
<tr>
<td>3G68</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5E14</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>2E1</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>5GXX</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>1G20</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>2E46</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>2F72</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3E46</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3E58</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3D17</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3D21</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>2F51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2F41/42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2F50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4D13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5C25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.2. Enhanced RLGS fragments in primary lung cancers.**

AC: adenocarcinoma; SCC: squamous cell carcinoma; LCC: large cell carcinoma; *Patient 12, poorly differentiated neuroendocrine carcinoma of small cell carcinoma and large cell carcinoma.
## Table 5.3. Enhanced RLGS fragments in lung cancer cell lines

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Cell line</th>
<th>Enhanced RLGS fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>H23</td>
<td>2B29, 2D26, 2F9, <strong>2F51</strong>, 3B35, 3G42, 4B14, 4C3, 4D29</td>
</tr>
<tr>
<td>AS</td>
<td>H125</td>
<td>1E11, 2E8, 2D31, 2D39, 2D43, 2D44, <strong>2C59</strong>, 3B10, 3D65, 3H6, 3H57, 4C30, 4F7, 5F19, 5G65</td>
</tr>
<tr>
<td>AC</td>
<td>H522</td>
<td>None</td>
</tr>
<tr>
<td>N/A</td>
<td>A549</td>
<td>2D60</td>
</tr>
<tr>
<td>LCC</td>
<td>H1155</td>
<td><strong>3G68</strong></td>
</tr>
<tr>
<td>LCC</td>
<td>H1299</td>
<td>2C13, <strong>2C59</strong>, 2F62, 3D9</td>
</tr>
<tr>
<td>AC</td>
<td>H2009</td>
<td>2D23, 3D37</td>
</tr>
<tr>
<td>AC</td>
<td>H2086</td>
<td>2E61, 3E34, 3C71, 3F37, <strong>5C1, 5E14</strong></td>
</tr>
<tr>
<td>N/A</td>
<td>H2172</td>
<td><strong>5C1</strong>, 3C34, 4E10, 4F44</td>
</tr>
<tr>
<td>H69</td>
<td></td>
<td><strong>2G55</strong></td>
</tr>
<tr>
<td>H82</td>
<td></td>
<td>1G35, <strong>3B35</strong>, 3H18, 3H46, <strong>4C3</strong>, 5C40, 5E39, 6C21, 6G13, 6G27</td>
</tr>
<tr>
<td>H209</td>
<td></td>
<td>2D57</td>
</tr>
<tr>
<td>H211</td>
<td></td>
<td><strong>3G68</strong></td>
</tr>
<tr>
<td>N417</td>
<td></td>
<td><strong>2G55</strong>, <strong>3G68</strong>, <strong>3B35</strong>, <strong>4C3</strong></td>
</tr>
<tr>
<td>H719</td>
<td></td>
<td><strong>3G68</strong>, <strong>2F36</strong></td>
</tr>
<tr>
<td>H792</td>
<td></td>
<td><strong>3G68, 3B35, 4C3</strong></td>
</tr>
<tr>
<td>H841</td>
<td></td>
<td><strong>3G68, 2F36</strong></td>
</tr>
</tbody>
</table>

**Table 5.3. Enhanced RLGS fragments in lung cancer cell lines**

AC: adenocarcinoma; LCC: large cell carcinoma; AS: Adenosquamous. Bold fragments indicate shared targets of enhancement.
<table>
<thead>
<tr>
<th>RLGS fragment</th>
<th>Patients &amp; cell lines</th>
<th>NotI-EcoRV clone Sequence*</th>
<th>Chromosomal location*</th>
<th>Related genes, ESTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>5E14</td>
<td>P2, H2086</td>
<td>AP000942.5 (44045-46459)</td>
<td>11q22</td>
<td>LOC143865</td>
</tr>
<tr>
<td>3D17</td>
<td>P6, P24</td>
<td>AC078799.21 (51984-53677)</td>
<td>3q26</td>
<td>LOC200816</td>
</tr>
<tr>
<td>3D21</td>
<td>P6</td>
<td>AC007823.37 (143153-144982)</td>
<td>3q26.32</td>
<td>LOC51193</td>
</tr>
<tr>
<td>1G20</td>
<td></td>
<td>AC117415.6 (92689-96295)</td>
<td>3q26.3-27</td>
<td>SOX-2</td>
</tr>
<tr>
<td>2F72</td>
<td>P6</td>
<td>AC068644.12 (155539-158713)</td>
<td>3q26.33-27.1</td>
<td>MRP5</td>
</tr>
<tr>
<td>3E58</td>
<td></td>
<td>AC068644.12 (158706-161287)</td>
<td>3q26.33-27.1</td>
<td></td>
</tr>
<tr>
<td>2E46</td>
<td>Not cloned</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3E46</td>
<td>Not cloned</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2F41/42</td>
<td>P12</td>
<td>AL139260.19 (20339-23614)</td>
<td>1p34.3</td>
<td>LOC163501</td>
</tr>
<tr>
<td>2F50</td>
<td></td>
<td>AL139158.11 (18113-21073)</td>
<td>1p34.1</td>
<td>POU3F1</td>
</tr>
<tr>
<td>4D13</td>
<td></td>
<td>AL359709.15 (18552-20325)</td>
<td>6q21</td>
<td>BM547890, POP3</td>
</tr>
<tr>
<td>5C25</td>
<td></td>
<td>AL359709.15 (17050-18181)</td>
<td>6q21</td>
<td></td>
</tr>
<tr>
<td>3D53</td>
<td>Not cloned</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.4. Properties of cloned RLGS fragments**

Figure 5.1. RLGS sections with enhanced fragments in primary lung cancer. (a) Amplification of fragment 5E14 (arrow) in PT2. In the RLGS profile from normal adjacent tissue 5E14 has single copy fragment intensity. (b) Newly appearing multi-copy DNA fragment in PT2 tumor profile. This fragment is not present in the normal lung profile. (c) A group of RLGS fragments, 2F72, 3D17, 3D21, 3E46 and 3E58, showed increased intensity to the same level in PT6 tumor profile, indicating that they are derived from the same amplicon. (d) Fragments 2F50, 2F41/42, 4D13 and 5C25, present in both primary tumor and normal lung profile, showed enhancement in PT12 tumor profile.
Figure 5.2. RLGS sections of lung cancer cell lines with enhanced DNA fragments. The RLGS fragments on the RLGS profiles were quantified by phosphoimager. (a) 3B35 and 4C3 show enhancement of more than 20-fold in SCLC cell line H82. (b) Approximately 5-fold increased intensity of 2F36 in cell line SCLC H841. (c) Multiple fragments showed 3- to 5-fold, low-level amplification in NSCLC cell line H125.
Figure 5.3. Confirmation of DNA amplification by Southern hybridization. (a) Chromosome 3q26-27 amplification was demonstrated using probes from RLGS fragments 3D17 and 2F72. All DNAs were digested with EcoRV. The chromosomal map for genes and RLGS fragments was created based on the information from the UCSC database and published literature (Sattler et al., 2000). (b) 1p34 amplification shown with probes 2F41,42 and MYCL1. (c) The 11q22 amplicon verified using the cIAP1 probe. (d) Confirmation of 6q21 amplification using probe KIAA1320, BVES and LOC135537.
Figure 5.4. Diagram of all the known genes and ESTs on chromosome 11q22. The black bar indicates the core amplified region in human primary lung cancer. The region has been completely sequenced (contig NT_009151.8). * Gene or EST location in the sequence contig is not drawn to scale. The estimated amplicon boundaries from literature are shown in red (218).
Figure 5.5. 11q22 amplicon definition by a series of Southern hybridization. All hybridizations were carried out on the same Southern blot with EcoRI digestions using probes described in Table 5.1. The degree of amplification is calculated by densitometry as described in the materials and methods. The shaded area indicates sequences that were enhanced 1.5-fold or higher.
CHAPTER 6

FUTURE DIRECTIONS

6.1 Biological significance of the methylated genes identified by RLGS in human cancer.

Besides BMP3B, other methylated genes identified by RLGS screening, using both \textit{Ascl} and \textit{NotI} RLGS, are of great significance for further studies, due to their important cellular roles (see table 1.1). The genes of interest will be studied in order to correlate methylation and transcriptional silencing. DNA methylation will be confirmed by Southern hybridization and techniques based on bisulfite treatment. The transcriptional level can be detected by RT-PCR or quantitatively Real-Time RT-PCR. In addition, cell lines with promoter methylation that silences candidate genes will be treated with 5-aza-2’-deoxycytidine to measure the reexpression of the genes following demethylation. Genetic alterations, such as LOH and mutations, are also worthwhile to study for the identification of other inactivating mechanisms of candidate tumor suppressor genes. The restoration of the previously silenced genes by promoter methylation in lung cancer cell lines will identify changes in cellular phenotypes through \textit{in vivo} and \textit{in vitro} cell growth growth assays, morphologic changes and metastatic abilities. For instance, the
methylated genes of interest from the 119 RLGS fragments cloned using the *AscI-EcoRV* boundary library are being studied in lung cancer, head and neck cancers and chronic lymphocytic leukemia (CLL) to search for candidate tumor suppressor genes.

### 6.2 Methylation and clinical correlation:

In a previous study using RLGS to screen overall changes in methylation in hepatocellular carcinomas (HCC), the total number of methylated RLGS fragments was shown to be an independent prognostic factor. These data showed that those patients with lower degree of methylation have longer post operative disease-free periods than those patients with higher levels of methylation (100). In a similar study, Fruhwald et al. have reported 30 methylated RLGS fragments, out of which 7 had significant correlation with survival time, independently. Therefore, larger sets of lung cancer samples with well characterized clinical information are needed to study the correlation between clinical outcome and the overall methylation pattern, and/or specific methylation events.

### 6.3 Further characterization the function of BMP3B

#### 6.3.1 Evidence for opposing roles of BMP3B and TGF-β1

As has been discussed in our previous study, BMP3B might have an antagonistic role against TGF-β1. Similarly, BMP3, the closest family member of BMP3B, and TGF-β1 have opposing effects on human bone marrow stromal cell growth and differentiation. In addition, BMP3 decreasing TGF-β1 affinity for its receptors is one of the possible mechanisms inhibiting TGF-β1 function (188).
TGF-β1 has multiple functions in tumor progression. In later stages of tumor development, TGF-β1 promotes tumor progression. One mechanism is through TGF-β1 induction of epithelial-mesenchymal transition (EMT). According to tissue origin, approximately 85% of cancers arise in epithelial tissue. However, the epithelial phenotype is lost to some degree during carcinoma progression. For example, epithelial cells are tightly connected to each other through a variety of cell-cell junctions. Normal epithelial cells are polarized, with distinct apical and basolateral sides. However, carcinoma cells are undifferentiated, have lost their cellular orientation, and loosely connect with each other. During invasion and metastasis, carcinoma cells dislodge from their place of origin and migrate to other sites, acting phenotypically like mesenchymal cells. This process is called epithelial mesenchymal transition (EMT). This process is also important for morphogenesis in multicellular organisms. Thiery reviewed the finding of EMT in tumor progression (196). In summary, if BMP3B can reverse some phenotypes induced by TGF-β1, it may have tumor suppressor function.

Further evidence will be collected to prove the hypothesis that BMP3B and TGF-β1 have opposing roles. First, the three down-regulated genes, fibronectin transcript variant 1 (FN1), transforming growth factor-β-induced, 68-kd (TGFBI) and stearoyl-CoA desaturase (delta-9-desaturase) (SCD), identified from microarray studies will be confirmed at both the transcriptional and translational levels using RT-PCR and Western hybridization respectively. Second, TGF-β1 will be added to the culture medium of A549 cells retrovirally infected with BMP3B to see if TGF-β1 acts as an antagonist of BMP3B. One might argue that the reexpression of BMP3B under the control of the retroviral promoter would be regulated differently than the endogenous gene. To approach this,
TGF-β1 can be added to cells displaying endogenous BMP3B expression and the effect on BMP3B can be monitored. Likewise, purified BMP3B protein can be collected and added in cell culture to study its role against TGF-β1.

### 6.3.2 BMP3B in metastasis

Based on our study that BMP3B cells reexpressing BMP3B have lower colony formation abilities via anchorage dependent and independent colony formation assays, we demonstrate that BMP3B down regulates components of extracellular matrix, such as fibronectin variant 1 (FN1), from microarray analysis and SYBR green semi-quantitative PCR. FN1 is highly expressed in human cancer and correlates with high metastatic ability (190). It has been demonstrated that cancer cell-associated fibronectin induces the release of matrix metalloproteinase-2 (MMP2) bound to collagen from normal fibroblasts to facilitate tumor cell invasion (242). Therefore, the invasive and metastatic ability of A549 cells might be affected following reexpression, which can be detected by *in vitro* or *in vivo* assays.

### 6.4 Identification of oncogenes in the amplicons.

According to our study, the 11q22 amplicon shared in different types of cancers was refined to a completely sequenced region less than 1 Mb. Known genes and predicted genes from the region are listed in figure 5.4. The identification of candidate oncogenes will be the focus of future studies. The overexpression of the genes at the mRNA and protein levels will be tested to identify candidate oncogenes for further study.
In addition, amplification is one of the mechanisms that lead to the overexpression of oncogenes. Same oncogenes may be regulated at the transcriptional or posttranscriptional levels resulting in gene overexpression in other individuals without amplification. Therefore, the detection of amplification-independent overexpression in lung cancer will provide further evidence to prove one gene acts as an oncogene. Furthermore, candidate oncogenes will be verified for their cancer promoting functions. For example, cIAP1 expression inhibits the apoptosis induced by chemotherapeutic drugs in esophageal squamous cell carcinoma cell lines (218) and decreases radiotherapy sensitivity in cervical cancers (238). The MMP family also plays a critical role in tumor progression (241). Thus, the identification of oncogenes in the amplicon will provide therapeutic targets by blocking their cancer promoting function, such as HER2/neu in breast cancer (199, 202).

6.5 Closing remarks.

In the current study, restriction landmark genomic scanning was performed to identify novel methylated and amplified DNA sequences in human lung cancer. RLGS using Ascl as restriction landmark enzyme established in this study doubled the numbers of CpG islands that can be evaluated for methylation or DNA amplification in different types of cancers. Methylated genes identified using both Ascl and NotI RLGS will be further studied for biological significance and clinical correlation. The novel methylated targets have great clinical potential for both diagnosis and therapy in cancer patients as has been reviewed (79, 243). Progress in the last decade makes it feasible to routinely detect DNA methylation of patients’ samples for diagnosis. First, aberrant DNA
methylation was observed in almost of all types of tumors. Second, new techniques, such as methylation detection following bisulfite treatment can detect DNA methylation from minimal amounts of clinical samples. Third, the detection of DNA methylation in patients’ sera and sputum provide non-invasive approaches for early diagnosis to guide therapy. It was proposed that the detection of multiple methylated genes to provide “methylation signature” of lung cancer would increase the specificity and sensitivity of the detection. The methylation signature might be utilized for early diagnosis, classification of histologic groups and guidance of clinical therapy. For instance, aberrant p16/CDKN2A methylation has been reported as a potential biomarker for early diagnosis (82). In regards the therapeutic potential of the novel methylated targets and overall methylation pattern, Brown and Strathdee (243) pointed out that the genome-wide methylation profiles might determine different treatment strategies for patients. In addition, understanding epigenetic silencing will improve the knowledge of tumor development for novel drug developments. Moreover, epigenetic modifiers, such as DNA methyltransferase inhibitor, 5-aza-2’-deoxycytidine, and histone deacetylase inhibitors have been used for clinical trials (243). Lack of specificity of these drugs is a major drawback. The strategy of reactivating specific genes silenced by methylation is one focus for future studies. Furthermore, the inhibition of over-expressed oncogenes may provide targets for the future gene therapy. Since the majority of tumor suppressor genes and oncogenes are shared targets in different types of cancers, I hope my study will benefit all cancer patients for better diagnosis and treatment.
BIBLIOGRAPHY


224. Brass, N., Heckel, D., Sahin, U., Pfleunertschuh, M., Sybrecht, G. W., and Meese, E. Translation initiation factor eIF-4gamma is encoded by an amplified gene and


