

ABERRANT DNA METHYLATION AND CANCER: A GLOBAL ANALYSIS OF
PROMOTER HYPERMETHYLATION IN HUMAN LUNG CANCERS

APPROVED BY SUPERVISORY COMMITTEE

John D. Minna, M.D. (Mentor)

Jerry W. Shay, Ph.D. (Chairman)

David R. Corey, Ph.D.

Michael A. White, Ph.D.

Thomas Kodadek, Ph.D.

DEDICATION

For Marcella, Victor, Heather, and John

ABERRANT DNA METHYLATION AND CANCER: A GLOBAL ANALYSIS OF
PROMOTER HYPERMETHYLATION IN HUMAN LUNG CANCERS

by

DAVID S. SHAMES

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

September 2006

Copyright

by

David S. Shames, 2006

All Rights Reserved

ABERRANT DNA METHYLATION AND CANCER: A GLOBAL ANALYSIS OF
PROMOTER HYPERMETHYLATION IN HUMAN LUNG CANCERS

Publication No.

David S. Shames, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2006

Supervising Professor: John D. Minna, MD.

ABSTRACT

Tumor-acquired alterations in DNA methylation include both genome-wide hypomethylation and locus specific hypermethylation. Global loss of DNA methylation destabilizes chromatin architecture, augments genomic instability, and may reactivate repetitive element expression. Promoter hypermethylation often coincides with loss of heterozygosity at the same loci, and together these events can result in loss of function of the gene in tumor cells. The “rules” governing which genes are methylated during the pathogenesis of individual cancers are unknown; however, it is known that certain genes are

methyated with high frequency in selected tumors, whereas others are methyated across most types of tumors.

The objective of the work described below was to use global profiling platforms (RNA and DNA) to identify epigenetically modulated genes that may be involved in cancer pathogenesis and bring these to the point where they could be developed as targets for diagnostic and treatment strategies.

Using a global expression profiling approach and pharmacological inhibition of the DNA methyltransferases, 132 genes were identified that have 5' CpG islands, are induced from undetectable levels by 5-aza-2'-deoxycytidine (5-aza) in multiple non-small cell lung cancer cell lines, and are expressed in untreated immortalized human bronchial epithelial cells. Methylation analysis of a subset (45/132) of these promoter regions in primary lung cancer (N=20) and adjacent non-malignant tissue showed that 31 genes had acquired methylation in the tumors, but did not show methylation in normal lung or peripheral blood cells. Promoter methylation of eight of these genes were studied in breast cancers (N=37), colon cancers (N=24), and prostate cancers (N=24) along with counterpart non-malignant tissues. We found that seven loci were frequently methyated in both breast and lung cancers, with four showing extensive methylation in all four epithelial tumors.

The data presented below suggest that while tumors differ in their molecular genetic phenotypes and pathogenesis, there may be underlying similarities in the pathways they follow toward malignancy. Some of these similarities may be reflected in the methylation programs tumor cells engage, which in turn, provides an opportunity to exploit for therapeutic applications and diagnosis. The approaches described herein entail a systematic

and reproducible method to identify novel methylation markers in a variety of cancers, and the results of these studies provide a basis for developing a generic set of methylation markers for early detection screening across common epithelial cancers.

TABLE OF CONTENTS

ABSTRACT	v
PRIOR PUBLICATIONS	xii

CHAPTER ONE – DNA METHYLATION IN HUMAN HEALTH, DISEASE, AND CANCER

ABSTRACT	1
INTRODUCTION.....	3
EPIGENETICS AND CHROMATIN STRUCTURE.....	4
DNA METHYLATION AND CHROMATIN STRUCTURE.....	8
DNA METHYLATION AND CANCER.....	14
GENOME-WIDE HYPOMETHYLATION.....	16
PROMOTER HYPERMETHYLATION	22

CHAPTER TWO – METHODS FOR STUDYING DNA METHYLATION

ABSTRACT.....	36
INTRODUCTION.....	37
FINE MAPPING AND QUANTITATIVE ANALYSIS OF DNA METHYLATION...	40
GLOBAL APPROACHES TO STUDYING DNA METHYLATION.....	47
MICROARRAY EXPRESSION PROFILING	48
RESTRICTION LANDMARK GENOMIC SCANNING	51
CHIP-ON-CHIP	54

SUMMARY	57
 CHAPTER THREE - MATERIALS AND METHODS	
MICROARRAY ANALYSIS, BIOINFORMATICS, AND STATISTICS.....	59
WET LAB PROTOCOLS.....	66
 CHAPTER FOUR – DEMETHYLATING DNA	
INTRODUCTION.	76
RNAI-MEDIATED KNOCKDOWN OF DNMT1 EXPRESSION REVERSES PROMOTER HYPERMETHYLATION AND INDUCES EXPRESSION OF MULTIPLE GENES IN HUMAN LUNG AND BREAST CANCER	
ABSTRACT.....	79
INTRODUCTION	81
RESULTS AND DISCUSSION.....	83
 CHAPTER FIVE – COMPARISON OF 5-AZA AND RNAI IN CANCER AND IMMORTALIZED CELLS	
INTRODUCTION	103
DIRECT COMPARISON OF 5-AZA AND RNAI IN H157 CELLS	104
DIRECT COMPARISON OF 5-AZA AND RNAI IN IMMORTALIZED BRONCHIAL EPITHELIAL CELLS	109
SPECIAL METHODS FOR HBECS	111

CONCLUSIONS.....	113
CHAPTER SIX – GENOME-WIDE SCREEN FOR METHYLATED GENES IN NSCLC	
INTRODUCTION	121
A GENOME-WIDE SCREEN FOR HYPERMETHYLATED GENES IN LUNG CANCER IDENTIFIES TUMOR-SPECIFIC METHYLATION MARKERS FOR MULTIPLE MALIGNANCIES	
ABSTRACT.....	123
INTRODUCTION	125
RESULTS	128
DISCUSSION.....	146
CHAPTER SEVEN – UNPUBLISHED DATA	
INTRODUCTION	184
CPG ISLANDS.....	185
ANALYSIS OF METHYLATION CHANGES IN PROGRESSED HBECS	188
INTRODUCTION – SATO ET AL.....	188
METHYLATION IN HBECS	
CHAPTER EIGHT	
CONCLUSIONS AND RECOMMENDATIONS	200

FUTURE STUDIES.....	207
BIBLIOGRAPHY.....	215
VITA.....	230

PRIOR PUBLICATIONS

1. **Shames, D.S.**, Girard, L., Gao, B., Sato, M., Lewis, C.M., Shivapurkar, N., Jiang, A., Shyr, Y., Fong, K., Gerald, W., Pollack J.R., Perou, C.M., Euhus, D.M., Shay, J. W., Gazdar, A.F., Minna, J.D. A Genome-wide Screen for Hypermethylated Genes in Lung Cancer Identifies Tumor-Specific Methylation Markers for Multiple Malignancies. *PLoS Medicine*. *In press*.
2. **Shames, D.S.**, Minna, J.D., Gazdar, A. F. DNA methylation and cancer. *Current Molecular Medicine*. *In press*.
3. **Shames, D. S.**, Minna, J.D., Gazdar, A., F. Methods for Detecting DNA Methylation in Tumors: from Bench to Bedside. *Cancer Letters*. *In press*.
4. Janowski, B.A., Huffman, K. E., Schwartz, J.C., Ram, R., Nordsell, R., **Shames, D.S.**, Minna, J.D., Corey, D.R. Ago1 and Ago2 link mammalian transcriptional silencing with RNAi. *Nature Structural and Molecular Biology*. *In press*.
5. Suzuki, M., Shigematsu, H., **Shames, D. S.**, Sunaga, N., Takahashi, T., Shivapurkar, N., Iizasa, T., Minna, J. D., Fujisawa, T., Gazdar, A. F. Methylation and Gene Silencing of the Ras-related GTPase Gene in Lung and Breast Cancers. *Annals of Surgical Oncology*. *In press*.
6. Sato, M., Vaughan, M. B., Girard, L., Peyton, M., Lee, W., **Shames, D. S.**, Ramirez, R. D., Sunaga, N., Gazdar, A. F., Shay, J. W., and Minna, J. D. Multiple oncogenic changes (K-RAS(V12), p53 knockdown, mutant EGFRs, p16 bypass, telomerase) are not sufficient to confer a full malignant phenotype on human bronchial epithelial cells. *Cancer Res*, 66: 2116-2128, 2006.

7. Suzuki, M., Shigematsu, H., **Shames, D. S.**, Sunaga, N., Takahashi, T., Shivapurkar, N., Iizasa, T., Frenkel, E. P., Minna, J. D., Fujisawa, T., and Gazdar, A. F. DNA methylation-associated inactivation of TGFbeta-related genes DRM/Gremlin, RUNX3, and HPP1 in human cancers. *Br J Cancer*, 93: 1029-1037, 2005.
8. Janowski, B. A., Huffman, K. E., Schwartz, J. C., Ram, R., Hardy, D., **Shames, D. S.**, Minna, J. D., and Corey, D. R. Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs. *Nat Chem Biol*, 1: 216-222, 2005.
9. Suzuki, M., Sunaga, N., **Shames, D. S.**, Toyooka, S., Gazdar, A. F., and Minna, J. D. RNA interference-mediated knockdown of DNA methyltransferase 1 leads to promoter demethylation and gene re-expression in human lung and breast cancer cells. *Cancer Res*, 64: 3137-3143, 2004.
10. Vilenchik, M., Raffo, A. J., Benimetskaya, L., **Shames, D.**, and Stein, C. A. Antisense RNA down-regulation of bcl-xL Expression in prostate cancer cells leads to diminished rates of cellular proliferation and resistance to cytotoxic chemotherapeutic agents. *Cancer Res*, 62: 2175-2183, 2002.

LIST OF FIGURES

FIGURE 1-1	34
FIGURE 1-2	35
FIGURE 3-1	75
FIGURE 4-1	96
FIGURE 4-2A	97
FIGURE 4-2B	98
FIGURE 4-3A	99
FIGURE 4-3B	100
FIGURE 4-3C	101
FIGURE 4-3D	102
FIGURE 5-1	114
FIGURE 5-2	115
FIGURE 5-3	116
FIGURE 5-4	117
FIGURE 5-5	118
FIGURE 5-6	119
FIGURE 5-7	120
FIGURE 6-1	161
FIGURE 6-2	162
FIGURE 6-3	163-64
FIGURE 6-4	165

FIGURE 6-5.....	166
FIGURE 6-6.....	167
FIGURE 6-7.....	168
FIGURE 6-8.....	169
FIGURE 6-9.....	170-71
FIGURE 6-10.....	172-73
FIGURE 6-11.....	174-75
FIGURE 6-12.....	176-77
FIGURE 6-13.....	178
FIGURE 6-14.....	179
FIGURE 6-15.....	180
FIGURE 6-16.....	181
FIGURE 6-17.....	182
FIGURE 6-18.....	183
FIGURE 7-1.....	194
FIGURE 7-2.....	195
FIGURE 7-3.....	196
FIGURE 7-4.....	197
FIGURE 7-5.....	198
FIGURE 7-6.....	199

LIST OF TABLES

TABLE 2-1	58
TABLE 4-1	94
TABLE 4-2	95
TABLE 6-1	152
TABLE 6-2	153
TABLE 6-3	154
TABLE 6-4	155
TABLE 6-5	156
TABLE 6-6	157
TABLE 6-7	158
TABLE 6-8	ONLINE
TABLE 6-9	159
TABLE 6-10	160
TABLE 7-1	191
TABLE 7-2	192
TABLE 7-1	193

LIST OF APPENDICES

LIST OF METHYLATED GENES 213

LIST OF DEFINITIONS

5-aza – 5-aza-2'-deoxycytidine

DNMT (1, 2, 3A, 3B, 3L) – DNA methyltransferase enzymes

CpG – 5'-cytosine phosphodiester guanine – 3'

mC – methyl cytosine

HCC – Hamon Cancer Center

NSCLC – non-small cell lung cancer

HBEC – human bronchial epithelial cells

MSP – methylation specific PCR

TSG – tumor suppressor gene

SAM – significance analysis of microarrays

siRNA – short interfering RNAs

RNAi – RNA interference

DMSO – dimethylsulfoxide

LOH – loss of heterozygosity

CHAPTER ONE

DNA METHYLATION IN HEALTH, HUMAN DISEASE, AND CANCER

ABSTRACT

The spatial arrangement and three-dimensional structure of DNA in the nucleus is controlled through the interdigitation of DNA binding proteins such as histones and their modifiers, the Polycomb-Trithorax group of proteins, and the DNA methyltransferase enzymes. These proteins interact with DNA to form chromatin via covalent modification as well as reversible macromolecular interactions involving heterogeneous protein complexes, RNA, and the nuclear membrane. DNA methylation forms the foundation of chromatin and is crucial to epigenetic gene regulation in mammals. Disease pathogenesis mediated through infectious agents, inflammation, aging, or genetic damage often involves changes in gene expression. In particular, cellular transformation coincides with multiple changes in chromatin architecture, many of which appear to affect genome integrity and gene expression. Infectious agents, such as viruses directly affect genome structure and induce methylation of particular sequences to suppress host immune responses. Hyperproliferative tissues such as those in the gastrointestinal tract and the colon have been shown to gradually acquire aberrant promoter hypermethylation. Below I review recent findings on altered DNA methylation in human disease, with particular focus on cancer and the increasingly large number of genes subject to tumor-specific promoter hypermethylation and the possible role of aberrant methylation in tumor development.

Note: The following chapter is in part made up of a review article written by David S.

Shames under the guidance of Adi F. Gazdar

INTRODUCTION

The primary structure of DNA encodes all the information necessary to establish and maintain the diverse tissues found in higher organisms. In humans, with the notable exception of certain programmed hemopoietic lineages, each cell contains identical DNA that is capable of directing the formation of diverse organs and tissues. There are two interdependent aspects of multicellular life that must be resolved by the organism in one way or another: one is how to establish differentiation programs, and the other is how to maintain a given cellular phenotype. To manage these tasks, complex organisms have evolved different mechanisms to regulate the signal transduction events and gene expression programs that lead to the differentiation and maintenance of specialized tissue types during development.

In humans, cellular differentiation begins when intrinsic and extrinsic signals activate or repress master regulatory transcription factors, which in turn activate or repress downstream factors that impart to the cell its characteristic phenotype (Taylor and Jones 1985; Zingg and Jones 1997; Agarwal and Rao 1998; Lee, Fitzpatrick et al. 2001; Hutchins, Mullen et al. 2002; Lee, Agarwal et al. 2002; Jaenisch and Bird 2003). Often times, the signal transduction events that initiate differentiation cascades are transient in nature, and thus must be recorded in the genomes of cells present at time of the signal. Thus, in order for specialized tissues to develop properly each cell needs to retain a “memory” of where it has been – what genes were expressed, as just as importantly, what genes were not (Urnov and Wolffe 2001).

EPIGENETICS AND CHROMATIN STRUCTURE

In mammals, memories of things past are often mediated by epigenetic mechanisms. Epigenetic phenomena are those that confer heritable phenotypes in cells that cannot be explained by changes in the primary structure of DNA. Epigenetic regulation of gene expression is critical to organismal development and cellular differentiation as well as a key element in the development of cancer (Jaenisch and Bird 2003; Feinberg 2004). The biochemical mechanisms of epigenetic gene regulation are complex, but they seem to be orchestrated through the spatial arrangement of DNA in the nucleus and alterations in chromatin structure. As discussed below, these changes occur through the interactions of the DNA methylation machinery and the Polycomb-Trithorax (Pcg-TrxG) family of proteins.

There are at least two levels of epigenetic gene regulation: one involves DNA methylation and the other occurs through complex protein-protein-DNA interactions involving the Pcg-TrxG family of proteins, histones, and histone modifying enzymes (Turner 2000; Wolffe 2001; Turner 2002; Reiner, Mullen et al. 2003; Ringrose and Paro 2004; Jaenisch, Hochedlinger et al. 2005; Bernstein, Mikkelsen et al. 2006; Boyer, Plath et al. 2006; Lee, Jenner et al. 2006). While relatively new to the field of epigenetics, it is clear that certain RNA species are also involved in heritable changes in gene expression (Lippman, May et al. 2003; Lippman and Martienssen 2004; Bernstein and Allis 2005). RNA may act by directly interfering with transcription in a process called transcriptional gene silencing, indirectly by recruiting silencing complexes, or by physically coating the DNA in association with DNA methylation, as occurs in X-chromosome inactivation and Barr body formation in females (Lee and Jaenisch 1997; Eggan, Akutsu et al. 2000; Janowski, Huffman et al. 2005;

Ting, Schuebel et al. 2005). The major focus of this review involves DNA methylation and its role in cancer development. However, the influence of histones and their modifications, as well as the Pcg-TrxG proteins to the epigenetic control of gene expression in normal and cancer cells should not be overlooked.

While histone modifications are clearly important to gene regulation, there is some question as to whether they are actually epigenetic phenomena because their modification state is not necessarily mitotically heritable (Bird 2002). However, recent evidence suggests that alterations in histone modification may precede and possibly direct more permanent changes in chromatin structure (Baylin and Ohm 2006; Pruitt, Zinn et al. 2006). Classically defined, chromatin is divided into two types based on differential Giemsa staining: heterochromatin and euchromatin (Urnov and Wolffe 2001; Fahrner and Baylin 2003; Huisinga, Brower-Toland et al. 2006). On a molecular and structural level, chromatin can be distinguished by its level of compaction, nuclear localization, and the classes of histone modifications and other protein associated with it (Schubeler, Francastel et al. 2000; Tolhuis, Palstra et al. 2002; de Laat and Grosveld 2003; Fahrner and Baylin 2003; Felsenfeld and Groudine 2003). Euchromatin exists in an open configuration, contains the majority of transcribed protein coding genes, localizes near the center of the nucleus, and is associated with hyper-acetylated histones; conversely, constitutive heterochromatin is highly ordered and compacted, contains transcriptionally inert sequences such as telomeric and pericentromeric repeats, localizes to the nuclear periphery, and is associated with trimethylated-histone 3 lysine 9 (H3K9), trimethylated-histone 3 lysine 27 (H3K27), heterochromatin protein 1 (HP1) and other factors (Ansel, Lee et al. 2003; de Laat and

Grosveld 2003; Fahrner and Baylin 2003; Ragooczy, Telling et al. 2003). A third type of chromatin, called facultative heterochromatin, is differentially compacted in different cell types, presumably reflecting the different gene expression profiles extant in different cells.

The major focus of this chapter involves DNA methylation and its role in cancer development. However, the influence of histones and their modifications, as well as the Pcg-TrxG proteins to the epigenetic control of gene expression in normal and cancer cells is also important. The Pcg-TrxG proteins are most influential in establishing and maintaining cellular phenotypes during early differentiation (Bernstein, Mikkelsen et al. 2006; Lee, Jenner et al. 2006). The PcG proteins form repressor complexes that appear to specifically target the regulatory elements of master regulator transcription factors such as the homeobox (*HOX*), sry-related homeobox (*SOX*), and the muscle segment homeobox (*MSX*) family of genes. Conversely, the TrxG proteins are activators of the same suite of target genes acting primarily through the modification of histones.

Histone proteins together with DNA form nucleosomes, which are the basic subunits of chromatin. Nucleosomes are made up of double helical DNA wound around an octamer of the four core histones – two each of histone 2A, histone 2B, histone 3, and histone 4 (Berger and Felsenfeld 2001; Felsenfeld and Groudine 2003). Each of the core histone particles has a globular domain around which the DNA winds and a positively charged amino (H3 and H4) or carboxy termini (H2A and H2B). These tails contain multiple lysine (K) and arginine (R) residues that protrude from the nucleosome core and act as substrates for a large and growing number of histone modifying enzymes including histone methylases, acetylases, phosphorylases, and ubiquitinases (Turner 2000; Turner 2002; Felsenfeld and Groudine

2003). How the different histone modifications affect gene expression alone and in combination are an active and increasingly complex field of research.

The most important function of chromatin is its influence on gene expression. The earliest evidence to suggest that chromatin affects gene expression comes from genetic studies in *Drosophila*. In the early 1930's H. J. Muller observed that when treated with X-rays, some *Drosophila* embryos developed a mosaic red and white pattern of eye color. Subsequent investigations demonstrated that when the mosaic phenotype was present the *white* gene had undergone a translocation from euchromatin to centromeric heterochromatin where it was subject to position effect variegation. In cells with white eye pigment the *white* gene had undergone heritable silencing at some point early in development due to its proximity to heterochromatin (Wakimoto 1998). In *Drosophila*, position effect variegation is mediated by the PcG proteins. In mammals, heterochromatin formation is also influenced by the PcG proteins. However, the most conspicuous difference between DNA in heterochromatin and euchromatin is the presence of DNA methylation.

DNA METHYLATION AND CHROMATIN STRUCTURE

In mammals, DNA methylation occurs on the 5th carbon of deoxycytidine in the dinucleotides 5'-CpG (Fig 1-1). This dinucleotide is substantially depleted in the human genome because methyl-cytosine (mC) is particularly susceptible to spontaneous deamination (Jones and Baylin 2002). However, ~70% of CpG sites in genome are usually methylated (Bird 2002; Paz, Fraga et al. 2003). mC acts as a foundation for DNA binding proteins such as the methyl-CpG binding domain protein family, which associate with histone deacetylases, histone methylases, and other heterochromatin associated proteins (Bird 2002). It is thought that the major function of DNA methylation in heterochromatin is to ensure that repetitive DNA remains transcriptionally inert; these regions contain an abundance of retroviral elements, long interspersed repeats (LINEs), and short interspersed repeats (SINEs) that could be deleterious to the genome if expressed as well as increase its susceptibility to recombination events (Garrick, Fiering et al. 1998; Walsh, Chaillet et al. 1998; Elena Kolomietz 2002). In cancer and certain rare congenital diseases (immunodeficiency-centromeric instability-facial anomalies (ICF) syndrome), the genome is depleted of repetitive element CpG site methylation (Xu, Bestor et al. 1999; Gaudet, Hodgson et al. 2003; Gaudet, Rideout et al. 2004).

The CpG palindrome is globally depleted in the genome, but there are local enrichments of GC content where the CpG dinucleotides occurs with high frequency. These sequences often occur in the 5' regions of protein coding genes and are called CpG islands (Gardiner-Garden and Frommer 1987; Takai and Jones 2002). CpG islands do not have an operational definition, but are characterized by higher than expected GC content (>50%)

where the CpG dinucleotide is relatively enriched with an observed vs. expected ratio of >0.6 , over a distance of at least 200 base pairs (Gardiner-Garden and Frommer 1987). A more recent analysis of the full sequence information on chromosomes 21 and 22 suggests that a better definition may be a minimum of 500 base pairs, a GC content of $\geq 55\%$, and an observed vs. expected CpG ratio of ≥ 0.65 . The more stringent definition excludes many intergenic CpG rich areas such as those associated with long terminal repeats (LTRs), *Alus*, and other repetitive elements. According to the latter definition approximately 40% of human genes are associated with these elements (Bestor, Gundersen et al. 1992; Takai and Jones 2002).

CpG islands are well conserved in higher vertebrates, and in contrast to most CpG sites, they are normally unmethylated. However, in transformed cells, dense methylation of promoter associated CpG islands is a relatively common event, and age-related methylation of certain CpG islands does occur in certain tissues, particularly in hyperproliferative tissues such as the colon and gastrointestinal (GI) tract (Esteller, Corn et al. 2001; Issa, Ahuja et al. 2001; Jones and Baylin 2002). Thus, there is a mosaic pattern of CpG site methylation in normal cells – highly methylated individual CpG sites along with unmethylated CpG islands – that is reversed in transformed cells – genome-wide hypomethylation with increased CpG island hypermethylation.

The mechanisms that establish genome methylation patterns are complex and occur during a brief period in embryonic development, whereas tissue-specific promoter hypermethylation patterns probably occur during cellular differentiation, but at restricted sites (Li, Beard et al. 1993; Rao and Avni 2000; Ansel, Lee et al. 2003; Jaenisch and Bird

2003). A diverse but conserved family of enzymes mediate the transfer of methyl groups from S-adenosyl-methionine to cytosine through an unusual mechanism whereby a cytosine base, in the context of a CpG dinucleotides, is flipped out of helical DNA, modified, and then replaced into the helix (Jeltsch 2002). In humans, there are at least five of these enzymes (DNA methyltransferase 1, 2, 3A, 3B, and 3L), three of which have *in vitro* methyltransferase activity. A recent study suggests that DNMT2 may act on tRNA, whereas DNMT3L activity may be limited to certain embryonal cell types (Grace Goll and Bestor 2005; Goll, Kirpekar et al. 2006; Jeltsch 2006). For several years it was believed that the different methyltransferases had exclusive activities, where DNA methyltransferase 1 (DNMT1) was the ‘maintenance’ methyltransferase and DNMT3A and 3B were *de novo* methyltransferases. However, recent biochemical and genetic evidence suggest that this distinction is probably more convenient than real.

DNMT1 was designated the maintenance methyltransferase because it localizes with the replication foci during S-phase and the catalytic efficiency of recombinant human DNMT1 is 3-10 fold higher on hemimethylated endogenous substrates (small nuclear riboprotein-associated peptide N (*SNRPN*) - exon-1 and fragile X mental retardation syndrome (*FMR-1*) locus) than on unmethylated substrates, *in vitro* (Bacolla, Pradhan et al. 1999; Pradhan, Bacolla et al. 1999; Rountree, Bachman et al. 2000). However, genetic evidence suggests that these enzymes are at least partially redundant with respect to maintenance activity. The strongest support for this comes from studies based on a series of isogenic cell lines engineered to have different combinations of intact methyltransferase enzymes. Homozygous deletion of any of the DNMT enzymes in the colon cancer cell line

HCT116 had almost no effect on genome-wide and repetitive DNA methylation, or aberrantly methylated promoters. However, when both *DNMT1* and *DNMT3B* were deleted, nearly all genomic methylation was lost, and genes silenced by promoter hypermethylation were reactivated (Rhee, Jair et al. 2000; Rhee, Bachman et al. 2002).

During lymphocyte differentiation, there are changes in the methylation patterns of regulatory elements that control expression of the *Il4* cytokine locus (Guo, Hu-Li et al. 2002; Lee, Fields et al. 2003). In a series of experiments using isolated immature thymocytes, Makar and colleagues co-immunoprecipitated both DNMT1 and DNMT3B at methylated regulatory elements in *Il4* locus, however, upon stimulation and lineage commitment, both proteins were actively excluded from these sites, methylation was lost, gene activation occurred (Makar, Perez-Melgosa et al. 2003). These data suggest that the DNMT enzymes are at least partially redundant with respect to maintenance methyltransferase activity (Rhee, Jair et al. 2000; Lee, Fitzpatrick et al. 2001; Rhee, Bachman et al. 2002; Makar, Perez-Melgosa et al. 2003; Ma, Jacobs et al. 2005).

The evidence for exclusive *de novo* methyltransferase activity is similarly controversial. While there appears to be some sequence specificity for DNMT3A and DNMT3B compared to DNMT1, the Michaelis constants for DNMT1, DNMT3A, and DNMT3B are essentially the same for unmethylated substrates (0.5 - 1.3 +/- 0.1 μ M, 0.9 +/- 0.3 μ M, and 3.5 +/- 1.2 μ M, respectively) when compared *in vitro*, suggesting that all three enzymes have *de novo* methyltransferase activity (Bacolla, Pradhan et al. 1999; Aoki, Suetake et al. 2001). However, the biological relevance of these assays is questionable, since

all three methyltransferase activities complex with large numbers of proteins in the context of chromatin, which substantially affects their activity and efficiency.

The genetic evidence supporting a *de novo* role for any of these enzymes is based on germ-line knockout experiments in mice. Germ-line disruption of *Dnmt1* or *Dnmt3b* results in early embryonic lethality (Li, Bestor et al. 1992; Lei, Oh et al. 1996; Okano, Bell et al. 1999). *Dnmt3a*^{-/-} mice survive to term but are runted and die at around 4 weeks (Okano, Bell et al. 1999). In double homozygous *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-} mice, repetitive elements and proviral DNA remain unmethylated, and the mice die around day 8. However, repetitive elements and proviral DNA in *Dnmt1* null mice was found to be partially methylated by day 8 (Okano, Bell et al. 1999). Interestingly, methylation of intracisternal A particle (IAP) repeats and proviral DNA in single knockout *Dnmt3a* or *Dnmt3b* were almost normal. These data suggest that while Dnmt1 alone cannot establish methylation on repetitive DNA in the embryo without one of the other methyltransferases, it does compensate for loss of only one of the other enzymes. However, it is important to recognize that the biology of embryonic stem cells and germ cells is probably quite different from that of somatic cells or cancer cells.

Studies in tumor cells have complicated matters further. Some reports suggest that different cell types may have different requirements for the different methyltransferase enzymes. siRNA targeting DNMT1 in both the lung cancer cell lines H1299 and the breast cancer cell line HCC1954 reversed promoter hypermethylation at several tumor suppressor genes, whereas similar experiments in HCT116 did not appear to have any effect on promoter hypermethylation (Suzuki, Sunaga et al. 2004; Ting, Jair et al. 2004). A subsequent study by Ting et al. confirmed the findings previously reported by ourselves and

others that in some cases inhibiting DNMT1 seems to be sufficient for reversing promoter hypermethylation (Fournel, Sapieha et al. 1999; Robert, Morin et al. 2003; Ting, Jair et al. 2006). Taken together, the data suggest that the relationship between the various DNMTs, the types of sequences they methylate, as well as the timing of their activities, is complex, and probably context dependent.

DNA METHYLATION AND CANCER

The idea that there was an epigenetic component to neoplasia is several decades old, but its empirical roots are found in several studies performed on teratoma cell lines in the mid-1970's (Mintz and Illmensee 1975; Illmensee and Mintz 1976; Nowell 1976). Illmensee and Mintz combined mouse teratocarcinoma cells with a mouse blastocyst and showed that in the chimeric adult mice the teratocarcinoma cells were substantially involved in producing most of the tissues and organs in the animal. The direct implication of these studies was that teratocarcinoma cells retain many aspects of totipotent embryonic stem cells. More importantly, these studies showed that the malignant phenotype was not entirely intrinsic to or immutable in the tumor cell; crucial aspects of tumor cell behavior must also depend on extrinsic and epigenetic factors as well.

Teratocarcinomas are unusual tumors and may have a large epigenetic component. Jaenisch and co-workers revisited these studies using modern genetic techniques and demonstrated that passage of melanoma cell nuclei through an embryo, which has been shown to remove and then re-establish methylation marks, effectively reverses many of the malignant characteristics of the tumor cells (Hochedlinger, Belloch et al. 2004). Chimeric, but not cloned mice, developed to term. While the mice did have a propensity to develop melanomas and lymphomas, these findings further demonstrate that epigenetic events substantially affect the malignant phenotype.

During cancer development two distinct changes in DNA methylation occur: genome-wide hypomethylation and locus specific gain or loss of cytosine methylation in promoter-associated CpG islands (Fig. 1-2). Genome-wide hypomethylation is associated with the

early stages of cellular transformation and usually targets non-coding regions. In addition, genome-wide hypomethylation influences genomic stability, causes loss of imprinting, and may result in the induction of ectopic onco-fetal gene expression (Jaenisch and Bird 2003; Holm, Jackson-Grusby et al. 2005; Hong, Kang et al. 2005; Tong Ihn Lee 2006). It is likely that genome-wide changes in methylation alter overall chromatin architecture, chromosome segregation in mitosis, and cell ploidy, all of which augment cellular transformation. Promoter region CpG island hypermethylation and in some cases hypomethylation affects the expression of associated genes (Jones and Baylin 2002; Baylin and Ohm 2006; Feinberg, Ohlsson et al. 2006).

GENOME-WIDE HYPOMETHYLATION

Genome-wide hypomethylation occurs early in cellular transformation and usually targets repetitive elements and other non-coding regions (Jaenisch and Bird 2003). Genome-wide hypomethylation may have several implications for the genome of preneoplastic cells, which may be broadly grouped into transcriptional and genetic effects. Transcriptional effects include loss of imprinting, induction of ectopic onco-fetal gene expression, and transcriptional activation of repetitive elements (Yoder, Walsh et al. 1997; Walsh, Chaillet et al. 1998; Holm, Jackson-Grusby et al. 2005). Genetic effects are probably more indirect and involve larger scale processes such as overall chromatin architecture, aneuploidy, and DNA replication (Eden, Gaudet et al. 2003; Gaudet, Hodgson et al. 2003; Jaenisch and Bird 2003).

It has been suggested that the primary function of global hypermethylation is to maintain the integrity of the genome through the heritable repression of repetitive element transcription (Walsh and Bestor 1999). Nearly half of the genome is compromised of repetitive element which range from nearly complete retroviruses that have been trapped in the genome by DNA methylation dependent silencing or mutation of their *env* genes, to *Alus* which are short, inverted repeats approximately 300 bps long, (Englander, Wolffe et al. 1993; Ostertag and Kazazian Jr 2001). In theory, up-regulated expression of retrotranspositionally active elements could increase the probability of insertional mutagenesis, although experimental evidence for this proposition is limited (Jackson-Grusby, Beard et al. 2001). The idea that ectopic expression of transposable and retrotransposable elements might lead to increased insertional mutagenesis probably derives, at least in part, from early molecular genetics techniques where repetitive elements were used to drive

recombination and mutation events in yeast and *Drosophila*. However, most laboratory species do not have endogenous DNA methyltransferase activities, and thus probably have evolved ways to tolerate active transposition (*Drosophila*), or silence repetitive elements through alternative pathways (*C. elegans*; co-suppression). In any case, with relatively rare exception, insertional mutagenesis does not seem to be a common event in human cancer (Griffiths 2001). Although, until recent years, sequence analysis of cancer genes has focused primarily on exonic sequences, thus only those insertion events occurring in, or proximal to, coding regions would be detected. Recent advances in high throughput genomic sequencing will likely reveal whether mutational events such as repetitive sequence retrotransposition into regulatory elements of protein coding genes occurs with any frequency in the cancer genome.

There is evidence that global demethylation results in up-regulation of genes that are silent in normal cells. While the mechanisms behind this phenomenon are unclear, at least two possibilities exist: 1) gene promoters that are methylated during embryogenesis, development, or differentiation become reactivated upon promoter demethylation resulting in aberrant expression; 2) global demethylation results in gene rearrangements that in turn lead to the stochastic integration of silent genes near active promoters. One of the first studies to demonstrate that altered promoter methylation states existed in cancer cells relative to normal cells showed that the promoters of *KRAS* and *HRAS* were relatively hypomethylated compared to companion normal cells (Feinberg and Vogelstein 1983; Feinberg and Vogelstein 1983). In the years since this study most of the focus has been on the reverse process of promoter hypermethylation, which is the focus of the next section. However,

recent studies have begun to revisit this issue and promoter demethylation may be a common event in carcinogenesis, as well. In particular, numerous microarray expression studies using 5-aza-2-deoxycytidine (5-aza) have shown that induction of oncogenes, imprinted genes, and developmentally regulated transcription factors are a direct effect of DNA demethylation, and comparative gene expression profiling studies have shown that tumor cells frequently express clusters of carcino-embryonic antigens (Sato, Maitra et al. 2003; Holm, Jackson-Grusby et al. 2005; Vatolin, Abdullaev et al. 2005; Suzuki, Suzuki et al. 2006).

Genomic hypomethylation has been shown to increase genomic instability and the frequency of spontaneous tumor formation in mice (Chen, Pettersson et al. 1998; Gaudet, Hodgson et al. 2003; Ma, Jacobs et al. 2005). In one of the first studies to address the global differences in mC content between tumor and benign tissues, Ehrlich and colleagues examined 103 tumors ranging from benign fibroadenomas to a variety of secondary neoplasms found at autopsy, and found that mC content was lower in more progressed tumors (secondary malignancy < primary malignancy < benign tumor < normal tissue), whereas there was no statistical difference in cytosine content overall (Gama-Sosa, Slagel et al. 1983).

The first hints of a connection between genome hypomethylation and genomic instability came from studies in a series of colon cancer cells using a selectable (G418/neomycin) retroviral reporter (*β -Gal*) system (Lengauer, Kinzler et al. 1997). After repeated attempts, it was found that while all cell lines were selectable in G418, only half expressed the *β -Gal* reporter after G418 selection at detectable levels. It turned out that the cells without detectable β -Gal expression had methylated the retroviral 5' LTR.

Interestingly, all these cell lines had mutations in the DNA mismatch repair pathway, whereas the cell lines that expressed β -Gal did not. These findings led the authors to speculate that cells without intact mismatch repair pathways were nevertheless able to methylate ectopic DNA, which coincided with high-levels of microsatellite instability. On the other hand, cells with wild-type mismatch repair genes that did not methylate the reporter construct were more susceptible to large-scale chromosomal aberrations and aneuploidy.

Direct evidence for the relationship between genomic instability and genome hypomethylation comes from studies in mice without a functional *Dnmt1* allele (Chen, Pettersson et al. 1998). Mouse embryonic stem (ES) cells null for the *Dnmt1* locus were drug selected using 6-thioguanine, which selects for mutations in hypoxanthine phosphoribosyltransferase gene (*Hprt*). *Dnmt1*^{-/-} ES cells produced 10-fold more 6-thioguanine resistant clones than *Dnmt1* positive cells. Analysis of the *Hprt* locus by PCR and southern blot showed that the major mechanism for loss of *Hprt* function involved large genomic rearrangements rather than point mutations. Further experiments showed that the rearrangements occurred primarily through mitotic recombination events.

The data described above demonstrate a correlation between DNA hypomethylation and genomic instability, but only suggest a link to tumorigenesis. Since *Dnmt1*^{-/-} mice die before gestation, *Dnmt1* hypomorphs were created that harbored 10% of wild-type DNA methyltransferase activity. These mice were viable, but runted, and most developed T-cell lymphomas by 4 months (Gaudet, Hodgson et al. 2003). The authors examined three alternative explanations for the high incidence of tumors in these animals: 1) up-regulation of endogenous retroviral elements and insertional mutagenesis, 2) activation of proto-oncogene

expression via promoter demethylation, 3) increased genomic instability. There was no increase type-C endogenous retrovirus expression, but some increase in IAP expression. There were three insertion events in *c-myc* in the 12 mouse tumors, which is probably significant, but was not pursued by the authors. *C-myc* was over-expressed in the tumors, but not in surrounding tissues from the same mice, suggesting that *c-myc* activation was secondary to tumor formation. The authors used array-based comparative genome hybridization (aCGH) to determine whether tumors from *Dnmt1* hypomorphs exhibited higher incidences of genomic instability overall than tumors derived from mice infected Moloney Murine Leukemia Virus (MMLV). While the number of mice in this study was small and the array technology used was relatively low resolution, there was a clear and statistically significant increase in chromosomal aberrations in the hypomorphic mice (Eden, Gaudet et al. 2003; Lengauer 2003). Considering the above data, it appears that the major influence of genomic hypomethylation on tumorigenesis occurs at the level of gross chromosomal alterations.

DNA hypomethylation may well have other adverse effects on genome. Recent expression profiling experiments using either 35 bp resolution tiling arrays or 5 bp tiling arrays suggest that the transcriptional capacity and complexity of the genome has been vastly underestimated (Cheng, Kapranov et al. 2005; Willingham and Gingeras 2006). Nearly all repetitive elements including endogenous retroviruses, LINEs, SINEs, and *Alus* contain strong promoters. Indeed, *Alus* are essentially clusters of transcription factor binding sites, and the older members of this family are significantly enriched in GC-rich domains of the genome (Polak and Domany 2006). Demethylation of mobile genomic elements may not

only result in transcription of these loci, but also transcription initiation at these loci. However, the technology to test this possibility has only recently become available and thus awaits future examination.

PROMOTER HYPERMETHYLATION AND CANCER

In normal cells, promoter hypermethylation is relatively rare, occurring at imprinted loci such as *H19/Igf2* promoter, as well as at promoters of genes active only during gametogenesis, embryogenesis, and development. There are cases of tissue-specific promoter, or regulatory element hypermethylation in the hematopoietic lineages that become demethylated and activated upon cytokine stimulation, as well as a variety of promoters that are methylated in an age-dependent and tissue-specific fashion in hyperproliferative tissues such as the colon (Yoder, Walsh et al. 1997; Issa, Ahuja et al. 2001; Lee, Fitzpatrick et al. 2001; Makar, Perez-Melgosa et al. 2003).

Imprinting and the related phenomenon of X-chromosome inactivation are mediated through epigenetic mechanisms that are established in early development (Chow, Yen et al. 2005; Pauler and Barlow 2006). However, there are two important differences between these two processes: first, X inactivation occurs randomly in females resulting in a mosaic pattern of X-chromosome derived gene expression, whereas imprinting occurs in a parent-of-origin specific pattern affecting all cells in the same way; secondly, X-chromosome inactivation occurs only in females, whereas imprinting occurs in both sexes. Despite these differences, both imprinting and X-inactivation are dependent on the expression of non-coding RNAs (ncRNA), which act *in cis* and recruit heterochromatin associated factors to the allele from which they are expressed.

Imprinted genes tend to occur in clusters ~1 Mb in length, usually including 5-10 protein-coding genes and a single ncRNA. Depending on the locus, a ncRNA is expressed from either the maternal or the paternal allele and the protein coding genes are expressed

from the other allele. While the ncRNA appears to be the key initial repressor of mRNA *in cis*, in nearly all cases, a conserved imprinting control element (ICE) carries the epigenetic “mark” enabling the alleles to “know” whether they should be transcribing mRNA or ncRNA in the first place (Pauler and Barlow 2006). Interestingly, male alleles (paternally expressed) are regulated through a different mechanism than female alleles (maternally expressed), and the male imprinting mark is established much earlier in gametogenesis and persists longer in the adult than female imprints (Bourc'his and Bestor 2006). Female alleles of paternally expressed genes are usually silenced by mechanisms involving direct methylation of promoter-associated CpG islands, whereas paternal alleles of maternally expressed genes seem to be regulated by methylation of distal enhancers elements (Bourc'his and Bestor 2006).

While the mechanisms involved in imprinting are complex and the details may differ somewhat from cluster to cluster the underlying theme is that at some point in development an expressed ncRNA recruits heterochromatin associated silencing machinery, including the DNMTs, which together impart the permanent epigenetic mark and expression pattern retained in all somatic cells throughout the life of the organism (Reik and Lewis 2005; Lewis and Reik 2006; Pauler and Barlow 2006). X-inactivation is thought to work in a similar way except that silencing affects an entire chromosome. For many years, it was thought that both copies of the X-chromosome were active in early female embryos, and then stochastic expression of *XIST* results in monoallelic silencing of the X-chromosome *in cis*. However, recent evidence suggests that X-inactivation may have two mutually exclusive steps. Early on, paternally derived expression of *XIST* results in silencing of the paternal X-chromosome,

whereupon the epigenetic mark is erased and then reset randomly, affecting all somatic cells in the female embryo and adult (Reik and Lewis 2005). Thus, the main features of gene specific epigenetic silencing in normal cells appear to depend on meiotically heritable “marks” that are interpreted by ncRNAs, which in turn direct the silencing machinery to the promoters and regulatory elements of proximal protein coding genes, *in cis*. It is unclear whether the mechanisms that regulate imprinting and X-inactivation during normal development are involved in establishing the aberrant methylation patterns found in tumor cells, although it would not be surprising if ncRNAs were somehow involved.

There is overwhelming evidence that tumor acquired promoter hypermethylation is a common event in the multi-step pathogenesis of human cancer (Zochbauer-Muller, Fong et al. 2001; Sekido, Fong et al. 2003; Zochbauer-Muller, Lam et al. 2003; Baylin and Ohm 2006; Belinsky, Liechty et al. 2006). Over the past decade, nearly 150 genes have been identified that show tumor-specific methylation in primary tumor samples (*Appendix A*). Gene specific promoter hypermethylation is also an early event in tumorigenesis and occurs in conjunction with transcriptional silencing of the associated gene. In addition, aberrant promoter hypermethylation often coincides with loss of heterozygosity resulting in complete loss of function of the affected locus (Jones and Baylin 2002; Baylin and Ohm 2006). However, the mechanisms that drive acquired promoter hypermethylation in cancer progression have remained elusive (Bestor 2003).

The earliest direct evidence that altered promoter methylation patterns are present in tumor tissues compared to companion normal cells, came from studies by Andrew Feinberg and Bert Vogelstein (Feinberg and Vogelstein 1983). They showed that the genomic

sequences of *HRAS* and *KRAS* were both hypomethylated in colon cancer cells compared to normal colonic epithelium. The link between promoter hypermethylation and loss of gene expression was made first at the calcitonin locus, and then more definitively at the promoter of the tumor suppressor gene *p16* (Baylin, Hoppener et al. 1986; Baylin, Makos et al. 1991; Fukuhara, Hooper et al. 1992; Ottaviano, Issa et al. 1994; Merlo, Herman et al. 1995).

DNA methylation-dependent silencing frequently affects genes that are involved in transcriptional regulation, DNA repair, negative regulation of the cell cycle, as well as growth regulatory signaling pathways (*Appendix A*). More recently, it was shown that multiple members of some gene families are found to be silenced in the same tumor, suggesting that the methylation machinery may have specificity during the pathogenesis certain types of cancer (Akiyama, Watkins et al. 2003; Margetts, Astuti et al. 2005; Suzuki, Toyooka et al. 2005; Shivapurkar, Stastny et al. 2006). Most data suggest that some loci are preferentially methylated in certain cancers, but not others (Baylin, Belinsky et al. 2000; Esteller, Corn et al. 2001). Aberrant promoter hypermethylation has been found in a variety of preneoplastic lesions, which supports the hypothesis that this epigenetic alteration is an early event in carcinogenesis. This observation has resulted in substantial interest from the medical community in that detection of methylation in patient samples may have utility in the early detection of cancer.

Promoter methylation has been detected in premalignant epithelial and hemopoietic cells infected with oncogenic viruses such as Epstein-Barr virus (EBV), SV40 virus, and Hepatitis B (HBV) and C (HCV) viruses. Human Immunodeficiency Virus (HIV) and human papilloma viruses have also been associated with differential methylation, however

the latency of HIV and the obligate presence of HPV in cervical cancers have made these associations more difficult to establish (Mikovits, Young et al. 1998; Fang, Mikovits et al. 2001; Duenas-Gonzalez, Lizano et al. 2005). Studies examining the methylation profiles of virally infected cells suggest that oncogenic viruses may induce specific changes in promoter hypermethylation during cellular transformation that are distinguishable from tumors originating from the same tissue in the absence of virus (Zhong, Tang et al. 2002).

EBV has been associated with a variety of epithelial malignancies as well as >40% of Hodgkin's disease. EBV is a member of the herpes virus family and is ubiquitous in the human population. EBV-associated gastric cancers show elevated levels of tumor suppressor gene promoter hypermethylation as well as CpG island methylator phenotype-associated (CIMP) methylation patterns (Chang, Uozaki et al. 2006). SV40 viral sequences are detectable in malignant mesotheliomas, brain tumors, and several types of lymphoid malignancies with relatively high frequency (Gazdar, Butel et al. 2002; Carbone, Bocchetta et al. 2003). This virus expresses two potent oncoproteins that inhibit the activities of tumor protein 53, retinoblastoma protein, and protein phosphatase 2A. Several studies have shown that the presence of SV40 viral sequences correlates with increased promoter methylation of tumor suppressor genes (Toyooka, Carbone et al. 2002; Shivapurkar, Takahashi et al. 2004; Suzuki, Toyooka et al. 2005). Two strains of HPV (16 and 18) are strongly associated with cervical cancer, which rarely occurs in the absence of these viruses (Shivapurkar, Toyooka et al. 2004; Takahashi, Suzuki et al. 2005; Kitkumthorn, Yanatatsanajit et al. 2006).

HPV, SV40, and EBV all contain oncogenes that probably mediate their tumorigenic effects. However, viruses also induce cellular transformation through persistent infection

(Karin, Lawrence et al. 2006). HBV and HCV do not express oncoproteins, but persistent infection of hepatocytes by these viruses is the major etiologic factor in the development of hepatocellular carcinoma (Thorgeirsson and Grisham 2002; Thorgeirsson, Lee et al. 2006). Long-term infection of the liver by HBV and HCV leads to cirrhosis and chronic inflammation of the infected tissue through localized infiltration of B and T cells. These immune cells secrete high-levels of chemokines and cytokines that induce apoptosis in infected cells, which in turn increases the turnover rate of hepatocytes. It is thought that the high turnover rate of these cells indirectly leads to errors in both DNA replication and possibly DNA methylation (Block, Mehta et al. 2003). A number of studies have shown that several tumor suppressor genes including *p16*, glutathione S-transferase (*GSTP1*), adenomatous polyposis coli (*APC*), and E-cadherin (*ECDH*) are methylated with higher frequency in HBV and HCV-associated hepatocellular carcinomas than in virus-independent malignant liver disease (Yang, Guo et al. 2003; Li, Hui et al. 2004). However, the exact mechanisms that cause HBV and HCV associated hepatocellular carcinoma are complex and direct evidence that either HBV or HCV actually alter DNA methylation has yet to be presented.

As stated previously, persistent viral infection often leads to chronic inflammation. Chronic inflammation is caused by a variety of etiologic agents including bacterial infections such as *H. pylori*, which is the primary cause of gastric cancer (Li, Stoicov et al. 2003; Houghton and Wang 2005). *H. pylori* infection results in the release of reactive oxygen species, which can cause oxidation of DNA resulting in adduct formation, spontaneous deamination of mC, and in rare cases double-stranded DNA breaks (Nardone, Rocco et al.

2004). Besides its initiator role in tumorigenesis, *H. pylori* infection has been associated with a highly significant increase in tumor suppressor gene methylation in benign gastric mucosa in affected patients without cancer compared to patients without cancer or *H. pylori* infection. When promoter hypermethylation patterns were compared between patients with gastric cancer, a clear distinction was found between those with and those without *H. pylori* infection (Maekita, Nakazawa et al. 2006). While there is strong, correlative evidence to suggest that *H. pylori* infection leads to increased tumor suppressor gene hypermethylation, it remains unclear whether this phenotype occurs because of direct or indirect effects of the infection.

Chronic inflammation has also been associated with increased promoter region hypermethylation in the absence of microbial infections. Barrett's esophagus is a premalignant lesion of the esophagus that begins at the junction between the stomach and esophagus. This disease is characterized by morphological changes in the epithelial layers at this junction from stratified squamous epithelium to metaplastic columnar epithelium, which gradually migrates to the distal esophagus (Lambert, Hainaut et al. 2004). The major etiologic factor in this disease is acid reflux, although *H. pylori* infection, smoking, and alcohol consumption also contribute (Crew and Neugut 2004). Several groups have investigated whether methylation markers may be useful in stratifying patients with Barrett's esophagus into at-risk and low risk groups for developing adenocarcinoma of the esophagus, and found that there was increased methylation of multiple markers in Barrett's esophagus, but that *p16*, *HPPI1*, and *RUNX3* methylation correlated with progression (Schulmann,

Sterian et al. 2005). Another study suggested that *TIMP3* and *APC* might also be useful (Geneviève Clément 2006).

Whether the result of environmental, infectious, or idiopathic causes, it has been proposed that cancer develops through the step-wise accrual of genetic mutations that together impart novel characteristics to the affected cell (Fearon and Vogelstein 1990; Hanahan and Weinberg 2000). Recent work has shown that a similar progression of epigenetic events may occur in the pathogenesis of common epithelial malignancies. Increasing promoter hypermethylation exists in age-related cancers such as that of the colon (Chan, Broaddus et al. 2002). Patients with ulcerative colitis, a precursor lesion for age related colon cancer, exhibit frequent methylation of *p16*, *MLH1*, and estrogen receptor in both dysplastic and benign tissue, whereas patients without ulcerative colitis do not (Issa, Ahuja et al. 2001). In other tissues where there is chronic exposure to carcinogens, a continuum of increasing methylation from hyperplasia through invasive carcinoma is evident (Wistuba, Mao et al. 2002; Zochbauer-Muller, Minna et al. 2002; Zochbauer-Muller, Lam et al. 2003; Shivapurkar, Stastny et al. 2006; Wistuba and Gazdar 2006).

The finding that promoter hypermethylation increases with age, carcinogen exposure, and histological progression suggests a mechanism whereby CpG islands progressively acquire methylation over time, which eventually leads to transcriptional silencing. There are at least two ways this gradual accrual of aberrant methylation could take place: in the first scenario, CpG islands, which are usually refractory to methylation, could lose some of their resistance to methylation and increasing CpG site methylation could eventually encroach on the core of the CpG island, which often contains the transcription start site (“encroachment”

model); the major factor here is time and frequent cell cycling (Baylin 2002). The second possibility is that over successive cell divisions the DNMT enzymes randomly methylate one or two CpG sites within a CpG island and this acts as a seed during subsequent cell division cycles (“marauding methylase” model) (Jones and Baylin 2002).

Experimental evidence that directly addresses the distinguishing features of these two possibilities is difficult to come by, and probably both mechanisms exist. However, the key issue to the relationship between aberrant promoter hypermethylation and tumorigenesis is the affect methylation has on transcription. It is well established that inhibition of the DNA methyltransferases, pharmacologically or by genetic approaches, leads to promoter demethylation and gene reactivation (Bender, Pao et al. 1998; Rhee, Bachman et al. 2002; Velicescu, Weisenberger et al. 2002). Most studies imply that DNA methylation is the proximate cause of silencing, however, it has been difficult to prove unequivocally. However, there is some evidence to suggest that methylation may be secondary to histone modification and transcriptional silencing in certain circumstances (Bachman, Park et al. 2003; Pruitt, Zinn et al. 2006).

Using the colon cancer cell line HCT116 engineered to be null for *DNMT1* and *DNMT3B* and allele specific chromatin immunoprecipitation assays, Vogelstein and colleagues demonstrated that methylation of histone 3 lysine 9 preceded promoter methylation as the cause of silencing of wild-type *p16*. HCT116 has one methylated wild-type copy of *p16*, which is not expressed, and one mutant copy, which is unmethylated and expressed at the RNA level. Upon genetic disruption of the two active DNA methyltransferases, *p16* is expressed from both alleles. After 22 passages, wild type *p16*

expression is undetectable, but in the absence of promoter hypermethylation, whereas the mutant allele of *p16* is expressed as before. Interestingly, promoter methylation did eventually return to the wild-type promoter after 86 passages, a finding that was not explored in detail (Bachman, Park et al. 2003).

Whether or not aberrant DNA methylation causes gene silencing in tumor cells initially does not belie the importance of this epigenetic mark to the process of tumorigenesis. Estimates vary, but it has been suggested that a clinically evident lung cancer requires 20 or more mutations in protein coding genes (Fong, Sekido et al. 2003). Accumulated data suggest that aberrant promoter methylation may affect over 100 genes in a single tumor (Sato, Fukushima et al. 2003; Keshet, Schlesinger et al. 2006; Shames 2006). While it is likely that some of these aberrantly methylated promoters are the product of on-going, random events, there is a strong possibility that promoter specific hypermethylation influences and perhaps drives crucial steps in cellular transformation.

It has been suggested that altered DNA methylation is the primary etiologic factor in certain types of colon cancer, where there is an apparent CpG island methylator phenotype (CIMP). CIMP colorectal tumors are defined by the presence of concordant methylation of *p16*, *MLH1*, *MINT31*, *MINT2*, *MINT1*, and some have argued that CIMP tumors may derive from different precursor lesions. Moreover, CIMP tumors appear to have a worse prognosis than tumors with the classical “mutator” phenotype (Suzuki, Itoh et al. 1999; Toyota, Ahuja et al. 1999; Loeb 2001; Issa 2004). Other diseases such as leukemia, pancreatic cancer, and gastric cancers appear to have CIMP characteristics, but future studies will be required to

determine whether these 5 sequences are sufficient to define this phenotype in all cancers (Toyota, Ahuja et al. 1999; Ueki, Toyota et al. 2000; Garcia-Manero, Jeha et al. 2003).

A curious feature of aberrant promoter hypermethylation is that it does not appear to affect all genes with equal probability. The most conspicuous example of this phenomenon is difference between *p16* and *RB*; these two genes interact directly and inactivation of one or the other gene is nearly universal in tumors. Both have large CpG islands in their promoter regions, but only *p16* is methylated with significant frequency, whereas inactivation of *RB* almost always occurs through genetic mechanisms. Moreover, in certain cases such as the tumor suppressor gene *RASSF1A*, promoter methylation appears to be the major mechanism underlying gene inactivation in tumors. This suggests tumor-acquired promoter hypermethylation is non-random, and that there is something about certain loci that makes them particularly susceptible to aberrant methylation, an observation consistent with the CIMP hypothesis (Issa 2004; Baylin and Ohm 2006).

The frequency of tumor-specific promoter methylation suggests that aberrant promoter hypermethylation is an important element in epithelial cell carcinogenesis. The effects of large-scale changes in DNA methylation may be analogous to genomic instability and loss of heterozygosity in the neoplastic process. That the methylation machinery is always present and functional in cancer cells, suggests that tumor cells have usurped its normally protective function and use it to actively repress anti-proliferative and apoptotic signals as well as stabilize gene expression at critical stages during transformation. It may be that tumor cells, which are thought to derive from stem cell populations, randomly methylate genes in response to environmental, genotoxic, hypoxic, or other cues, using extant normal

processes such as those active in the epigenetic gene regulatory cascade of lymphocyte differentiation (Jaenisch and Bird 2003; Reiner 2003; Reiner, Mullen et al. 2003). Alternatively, aberrant DNA methylation may be the result of a more directed process with each type of tumor having a distinctive promoter hypermethylation profile. Either way, the result of these acquired epigenetic changes is a fitter, drug resistant, metastatic neoplasm, with a heritable epigenetic signature capable of further development toward malignancy. Thus, while tumors differ in their molecular genetic phenotypes and gene expression programs, they may engage similar methylation programs during their pathogenesis. Future studies will surely unravel the mechanistic basis of the complex interplay between the epigenetic and genetic components of carcinogenesis.

Figures

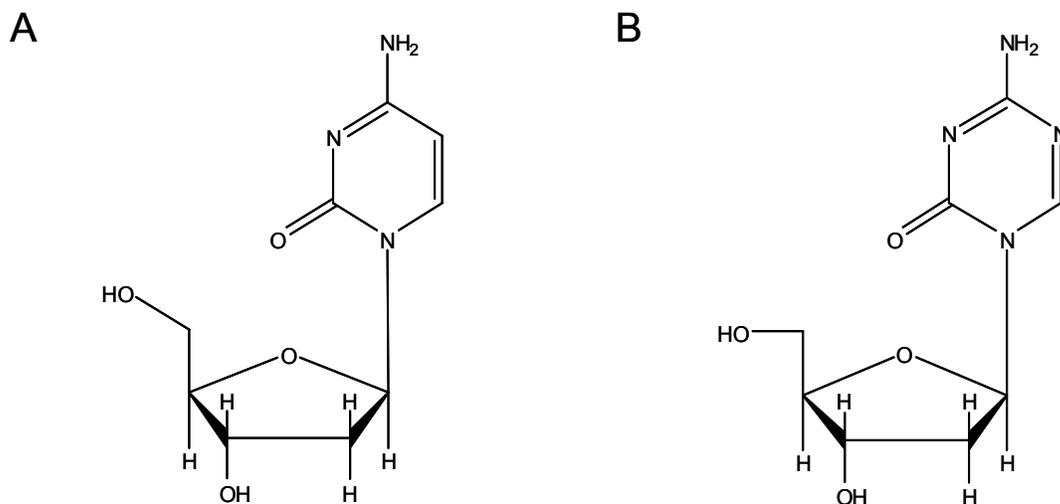


Figure 1-1. Chemical structures of deoxycytidine and 5-aza-2'-deoxycytidine. A) deoxycytidine. B) 5-aza-2'-deoxycytidine (5-aza). 5-aza has a nitrogen substituted for the 5th carbon in the purine ring. This compound is incorporated into DNA during S-phase and acts as a suicide inhibitor of the DNA methyltransferases. 5-aza is highly toxic to proliferating cells and is unstable in aqueous solution. Chemical structures were obtained through the Chemfinder database which may be found at <http://chemfinder.cambridgesoft.com/result.asp>.

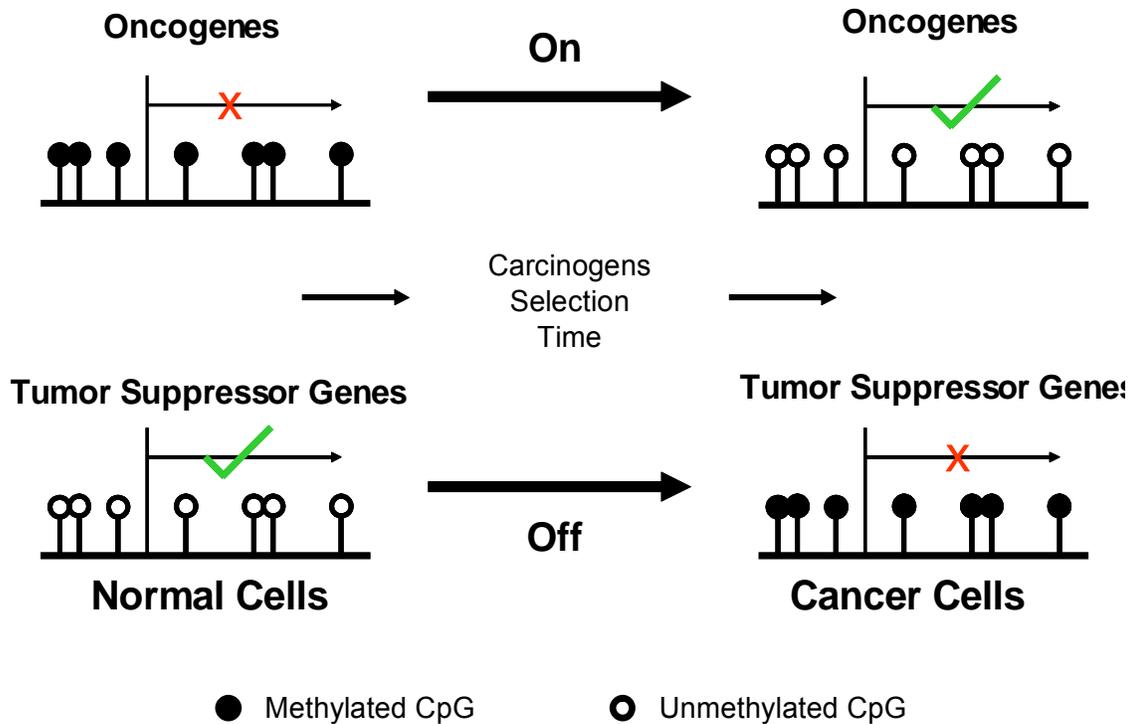


Figure 1-2. Schematic of tumor-acquired promoter hypermethylation and promoter hypomethylation in the development of human cancer. Cellular oncogenes and carcino-embryonic antigens that are normally silent in adult somatic cells are usually associated with methylated promoters. During the step-wise accumulation of mutations in genomic DNA that accompanies the early stages of carcinogenesis, promoter can become demethylated, resulting ectopic expression of these genes. In the reverse process, tumor suppressor genes acquire methylation of time that eventually results in silencing of the associated genes. Lollipops represent CpG sites; full circles are methylated CpGs; open circles are unmethylated CpGs. Red “X” indicates no expression; green check indicates expression.

CHAPTER TWO

METHODS FOR STUDYING DNA METHYLATION

ABSTRACT

Tumor-acquired changes in DNA methylation are the focus of research in an increasing number of basic, translational, and clinical laboratories around the world. In the laboratory, genome-wide technologies such as expression and DNA microarrays have been adapted to analyze patterns of DNA methylation and screen for novel disease markers. Other technologies that are relatively inexpensive and highly sensitive such as methylation specific PCR (MSP), or quantitative, such as quantitative MSP and pyrosequencing are widely used in retrospective studies and have potential in a diagnostic setting. In the near future, it may be possible to screen patients for common cancers using DNA methylation signatures as well as to measure patient responses to treatment, to identify patients at increased risk, or to monitor interventions designed to reduce cancer incidence. In this chapter, I review genome-wide and quantitative, high-resolution methods for methylation analysis that are used in the laboratory and clinic, and discuss their potential for use in a clinical setting.

Note: This chapter is made up of an invited review written by David S. Shames under the guidance of Adi F. Gazdar

INTRODUCTION

Whether or not promoter hypermethylation is the cause or consequence of gene silencing, the key link between the presence of CpG island methylation in the context of a gene promoter is that the associated gene is not expressed, and that pharmacological inhibition (and in most cases RNAi-based or genetic inhibition) of the DNA methylation machinery leads to gene re-activation (Jones and Taylor 1980; Merlo, Herman et al. 1995; Robert, Morin et al. 2003; Suzuki, Sunaga et al. 2004; Ting, Jair et al. 2006). These two observations have led to intense interest in the study of DNA methylation because genes that are silenced specifically in cancer likely affect the evolution of tumors in a negative way: for basic scientists this means that the genes are candidate tumor suppressors; for clinicians, detection of DNA methylation may provide a sensitive and specific, as well as high-throughput and non-invasive way to screen for the presence of cancer in at-risk individuals (Jones and Baylin 2002; Belinsky 2004). As a result, both basic and clinical scientists have developed methods to study, detect, and reverse DNA methylation.

High through-put approaches to discovering novel DNA methylation markers such as restriction landmark genomic scanning, microarray gene expression profiling after 5-aza-2'-deoxycytidine treatment, and ChIP-on-chip approaches have been employed in a variety of experimental contexts and have led to the discovery of several new methylation markers in recent years (Karpf, Peterson et al. 1999; Costello, Fruhwald et al. 2000; Suzuki, Gabrielson et al. 2002; Weber, Davies et al. 2005). A major outcome of experiments using these high through put approaches is the realization that tumor acquired promoter hypermethylation may not be random, and that cancers may have a common underlying promoter methylation

profile (Esteller, Corn et al. 2001; Keshet, Schlesinger et al. 2006; Shames 2006). However, like all high-content screening approaches, these methods are prone to certain biases, which may not be obvious in the initial study design. Therefore, data from these experiments must be interpreted with caution and confirmed independently by other methods.

Higher resolution methods, such sodium bisulfite pyrosequencing and quantitative methylation specific PCR, are gaining wider acceptance in the medical community as a promising method for early detection screening (Belinsky 2004). For several types of cancer, and in particular lung cancer, early detection could substantially increase 5 year survival rates. Several studies have been published that use a combination of markers to accurately predict outcome or the presence of clinically relevant neoplastic disease (Fackler, McVeigh et al. 2004; Shivapurkar, Stastny et al. 2006). Recently, the first prospective study using standard MSP on a cohort of at-risk individuals for lung cancer showed that combined analysis of 6 markers, in particular *p16*, *MGMT*, *RASSF1A*, *PAX5*, *GATA5*, and *DAPK* had both a sensitivity and specificity of 65% (Belinsky, Liechty et al. 2006). These results are promising and future, larger studies perhaps using more markers and quantitative assays will help support the prospect of using DNA methylation as method for early detection screening.

The detection of DNA methylation in clinical samples has other potential uses besides early detection screening. Because PCR-based methylation assays are extremely sensitive, DNA methylation could be used to confirm tumor margins in surgically resected specimens (Goldenberg, Harden et al. 2004). Additionally several recent articles have suggested that DNA methylation profiles could be useful in stratifying risk for breast cancer patients (Pu, Laitala et al. 2003; Fackler, Malone et al. 2006). Other groups including ourselves are

currently investigating whether methylation profiles can be used to predict patient outcome and whether there are relationships between particular methylation profiles and tumor drug sensitivity. These new directions in cancer epigenetics research require the application of global screening approaches and high-resolution methods and are important examples of how clinical and basic researchers will need to work together in the future.

As a result of these advances, we will focus primarily on methods used in basic research to identify global patterns of CpG island methylation, and those with potential use in a clinical setting (Table 2-1). For global methylation pattern recognition, several technologies including DNA and expression microarrays, CpG island microarrays, and ChIP-on-chip will be discussed, whereas for high-resolution and quantitative analysis, quantitative methylation specific PCR and pyrosequencing will be discussed. For a general overview of methods to detect DNA methylation we refer the reader to several excellent reviews (Costello, Smiraglia et al. 2002; Paz, Avila et al. 2002; Fackler, McVeigh et al. 2004; Curtis and Goggins 2005).

FINE MAPPING AND QUANTITATIVE ANALYSIS OF DNA METHYLATION

Because of the stability of DNA in bodily fluids, there is great potential in using tumor-acquired DNA methylation as a marker for early detection screening. Several methods have been described that quantitatively assess methylation at given CpG sites including sodium bisulfite conversion of DNA and direct sequencing, combined bisulfite restriction analysis (COBRA), methylation specific PCR, methylation-sensitive single-nucleotide primer extension (Ms-SNuPE), and more recently MALDI-TOF and have been reviewed elsewhere (Paz, Avila et al. 2002; Ehrich, Nelson et al. 2005). Here, we will focus on quantitative methylation specific PCR (QMSP) and pyrosequencing of bisulfite DNA, and discuss the interpretation of quantitative methylation data from clinical samples.

The study of DNA methylation was greatly facilitated by the discovery that it was possible to convert unmethylated cytosine to uracil by reacting denatured DNA with a saturated solution of sodium bisulfite (Frommer, McDonald et al. 1992). Methylcytosine is protected from conversion by aqueous sodium bisulfite. The resultant DNA can then be analyzed using a variety of methods, including those mentioned below.

The description of methylation specific PCR in the mid-90's, opened up the possibility of using the power and sensitivity of the polymerase chain reaction in cancer epigenetics research (Herman, Graff et al. 1996). In this method, PCR primers are designed to be complementary to completely methylated or completely unmethylated target DNA where methylated and unmethylated primer sets differ only in the CG position of the bisulfite converted primary sequence: methylated primer sets contain CG dinucleotides and unmethylated primers contain TG at these positions. In the laboratory, this method has had a

tremendous impact on the efficiency of DNA methylation analysis as evidenced by the number of citations the original work has received (>1500; SCOPUS). Qualitative MSP was used extensively in the studies described herein. This method was adapted to screen for methylated genes in a 96 well format and has the advantages of being both sensitive and highly specific. However, the limitation of MSP in terms of clinical applications is that it is not quantitative (see below). To address this issue, a quantitative assay based on Taqman technology has been developed (Eads, Danenberg et al. 1999; Eads, Danenberg et al. 2000).

Taqman technology uses fluorescence resonance energy transfer (FRET) to quantify *Taq* polymerase-based 5' - 3' exonuclease activity on DNA-primed, DNA substrates and is proportional to PCR based amplification of target DNA. Sequence specific primers and an intervening probe are designed to cover an amplicon of approximately 100 bps in length. The probe is labeled with a fluorescent reporter dye on the 5' end and a quencher on the 3' end. During sequence amplification in PCR, *Taq* exonuclease activity cleaves the 5' sequence from the probe resulting in fluorescence. Each round of PCR leads to an increase in fluorescence proportional to the amount of target present in the sample. This method was first applied to methylation analysis by using Taqman probes as part of standard MSP where each sample is assayed in two separate reactions: one for methylated DNA and the other for unmethylated DNA (Eads, Danenberg et al. 1999; Eads, Danenberg et al. 2000). More recently, two derivative approaches using this basic method have been described (Fackler, McVeigh et al. 2004; Shivapurkar, Takahashi et al. 2004; Suzuki, Toyooka et al. 2005).

The newer methods are based on the same principle as the original, but they differ in their theoretical assumptions and practical applicability; in the first method, called

quantitative MSP (QMSP), only the methylated reaction is used in a single PCR and the resultant data yield information only on the number of completely methylated alleles of the target DNA, whereas in the alternative approach, called quantitative multiplex MSP (QM-MSP), two rounds of PCR amplification are used to determine the relative percentage of completely methylated and completely unmethylated DNA in the sample.

In QMSP, sodium bisulfite treated DNA is used as a substrate for amplification of the methylated target promoter and a reference sequence in separate reactions. The reactions are set to a predetermined threshold based on standard curves using serially diluted, SssI-treated, sodium bisulfite DNA. Derivative fluorescence data are converted to numbers using a predetermined standard curve, and the ratio of target promoter to a reference sequence (that does not contain CpG sites, for example, *MYOD1*) is calculated. This number is an absolute value in fluorescence units and in theory reflects the normalized, methylated sequence content for a given marker in a given sample. This number can range widely depending on the efficiency of the reaction and prevalence of methyl-CpG in the target sequence. The reference sequence is used to normalize samples for DNA input, and also to determine the quality of bisulfite conversion (Shivapurkar, Stastny et al. 2005).

In QM-MSP, two rounds of PCR are used and both methylated and unmethylated target promoter regions are analyzed. In the first-round reaction, primers targeted to multiple regions of interest are multiplexed in a single reaction. Then two separate quantitative reactions using primers and fluorescent probes specific to completely methylated and completely unmethylated sequences are performed for each marker. The fluorescence data are then converted to percentages based on the amplification of both primer sets and are

additive such that the methylated sequence plus unmethylated sequence must equal 1 ($U + M = 1$). In the first round reaction, primers flanking the target region are used to amplify the region of interest independent of CpG content – i.e. the primers should not include CGs and therefore should amplify target sequences equally (Fackler, McVeigh et al. 2004). This requirement limits the available sequences that can be interrogated using this method.

Careful quality control and calibration steps are necessary to optimize this reaction, as many variables, known and unknown, must be accounted for if the subsequent reaction is to be truly quantitative. The primary purpose of this step is to provide a high quality substrate for the subsequent quantitative reaction. In the next step, aliquots of PCR product from the first reaction are subjected to PCR amplification using fluorescent probes corresponding either to methylated or unmethylated alleles of the target sequence.

Recently, an alternative locus-specific quantitative method was developed by adapting the pyrosequencing method of sequence analysis to bisulfite modified DNA (Colella, Shen et al. 2003). Pyrosequencing was developed as an alternative to dideoxy sequencing and is based on the on the detection of pyrophosphate (PPi), which is liberated from incorporated nucleotides by DNA polymerase during strand elongation. Free PPi molecules are converted to ATP by ATP sulfurylase, which provides the energy to luciferase to oxidize luciferin and generate light. Nucleotides are added sequentially to enable base calling. If the added base is not the proper Watson-Crick partner no luminescence is detected and the base is removed by washing in the solid-phase platform or enzymatically by apyrase in the liquid-phase reaction (Ronaghi 2001).

The major advantage of the pyrosequencing method compared to methylation specific PCR is that the data are actual sequences rather than fluorescence data from PCR based amplification. This means that pyrosequencing can detect partially methylated sequences that are outside of the priming sites, whereas MSP can only detect sequences that are completely complimentary to the primer and probe sequences. However, when the pyrosequencing method is applied to bisulfite treated DNA some special considerations need to be addressed that can affect assay reproducibility. Bisulfite DNA, particularly in dense CpG islands, tends to be highly repetitive, with long homopolymeric tracts of thymines. As with standard direct sequencing of bisulfite DNA, the reliability of pyrosequencing over these regions is limited, and is unlikely quantitative after ~75 bps, depending on the region (Dupont, Tost et al. 2004). As such, the further a particular CpG site is away from the 3' end of the forward primer, the less reliable/quantitative the data will be. Another drawback is that pyrosequencing requires dedicated equipment that is quite expensive.

Whatever method is ultimately used, some general considerations on the biology and technical aspects of DNA methylation analysis need to be addressed prior to committing resources and precious biological samples. These issues relate to the material under study and affect data interpretation. In cell lines, the presence or absence of methylation for the most part is generally binary question because cell lines represent pure DNA samples. This means that in the case of pyrosequencing, there should be a dominant base call at any given position, and a high fluorescence value or methylation percentage by QMSP and QM-MSP, respectively.

The situation is entirely different in clinical samples: the biological material available for use in a diagnostic setting such as a tissue biopsy, ductal lavage, blood, or sputum samples always contains a combination of tumor cells, stroma, leukocytes, lymphocytes, normal cells, and dead cells in any number of proportions. Thus, if the region of interest is differentially methylated in any of these cell types other than the tumor cells, or there are different proportions of tumor cells in different samples, it is not always clear what is actually being quantified. At the very least, this factor means that the selection of controls is especially important. The standard control used for marker development is normal lymphocyte DNA because lymphocytes are present in nearly all tissue and samples. Therefore, if a marker is positive in lymphocytes, it cannot be used to differentiate tumor from normal. For primary samples such as biopsy specimens, while not always practicably obtained, adjacent “normal” tissue is ideal. However, if the distinction between tumor specific methylation and background signal has been independently and empirically determined for a given marker, these control samples are not necessarily required.

There is one final caveat that warrants mention and distinguishes QMSP from QM-MSP. QMSP only measures the presence of completely methylated DNA sequences relative to a reference control, whereas QM-MSP measures both methylated and unmethylated sequences. QM-MSP assumes that the only possible target sequences are either completely methylated or completely unmethylated and the standard curves used to calculate the final percentages are based on this model. However, primary tumor material always contains different fractions of target sequence, some completely methylated, some completely unmethylated, and others harboring only partial methylation. As the Taqman probe/primer

pairs are highly specific and are designed to amplify only cognate target sequences, how the partially methylated fraction interferes with the PCR and affects the final calculations is unclear and probably different for the two above described methods. This means that in some cases the two methods yield different results, but both have been used successfully to discern cancer from normal in clinical samples (Fackler, McVeigh et al. 2004; Shivapurkar, Takahashi et al. 2004).

The methods outlined above are robust and have been used to address a variety of questions related to altered DNA methylation. While there are minor differences in the “quantitative” nature of these methods, as long as the key relationship between methylation and expression is kept in mind, and interpretation is based on independent empirical analysis for each marker, all three are probably valid. Future studies will hopefully include head-to-head comparisons of the three methods.

GLOBAL APPROACHES TO STUDYING DNA METHYLATION

The central focus of my thesis was to develop genome-wide information on gene methylation in lung cancer. As part of this, I also wanted to detect methylated sites that correlated with altered gene expression. Before discussing the methods I ultimately used, I discuss the methods that were available at the time I began this project.

Recently Peter Jones and others in the cancer epigenetics field have called for an “epigenome project” wherein the chromatin profiles of a series of tumors and normal tissue will be examined on a large scale (Jones and Martienssen 2005). Undoubtedly, novel, high-throughput methods will be developed to attack this problem, some of which may be based on current genome-wide approaches to studying methylation. At present, global approaches to methylation analysis are high-throughput with regard to the number of loci that can be analyzed at one time, but they are all relatively expensive and labor intensive. There are three distinct methods to study global (genome-wide) changes in promoter methylation and a variety of alternatives that combine aspects of each: these include microarray expression profiling of cell lines before and after pharmacological inhibition of DNA methylation or inhibiting expression of the DNA methyltransferase enzymes by genetic disruption or RNAi; restriction landmark genomic scanning followed by gel electrophoresis or hybridization to a CpG island microarray; and finally immunoprecipitation of methylated DNA followed by array-based comparative genome hybridization (ChIP on chip) analysis (Suzuki, Gabrielson et al. 2002; Sato, Fukushima et al. 2003; Gius, Cui et al. 2004; Heisler, Torti et al. 2005; Weber, Davies et al. 2005; Smith, Lin et al. 2006). Each of these methods has their advantages and disadvantages which will be discussed below.

MICROARRAY EXPRESSION PROFILING

Transcriptional profiling using high-density microarrays has become a standard method to identify broad biological differences between groups of samples, and has been used in a variety of experimental contexts. Widespread application of microarray technology has led to the development of statistical approaches and algorithms that are now relatively standardized. The relative merits of each approach are reviewed elsewhere (Allison, Cui et al. 2006).

Several groups have used microarray expression profiling to identify novel methylation markers, and each has approached the experiments in slightly different ways (Suzuki, Gabrielson et al. 2002; Sato, Fukushima et al. 2003; Lodygin, Epanchintsev et al. 2005). Most study designs are based on the following scheme: cultured cancer cells are treated with 5-aza and then compared by RNA expression microarray both before and after drug treatment (Karpf, Peterson et al. 1999). 5-aza is a cytidine analogue that integrates into genomic DNA during cell division (Jones and Taylor 1980). The reactive base binds irreversibly to the DNA methyltransferases and over a period of several days effectively depletes the cells of methyltransferase activity, which eventually leads to DNA demethylation (Jones and Taylor 1980). Unfortunately, 5-aza is highly toxic to cells, which can make data interpretation difficult. Recently several other compounds that affect DNA methylation have been identified, however, none are anywhere near as potent as 5-aza with respect to DNA demethylation (Stresemann, Brueckner et al. 2006).

One of the first studies to use the 5-aza expression profiling approach used membrane filter microarrays where radiolabeled cDNA is hybridized to a membrane that has a clone

library spotted onto it. This method is less sensitive than newer platforms, but the authors employed an innovative subtractive hybridization step using pooled mock-treated cancer cell line cDNA as bait and then hybridized the eluted cDNA (thereby excluding expressed genes). In addition the authors combined trichostatin A with 5-aza to increase the induction affected genes. The result of this landmark study was the discovery that multiple members of the secreted frizzled-related protein family – key members of the WNT signaling cascade – and are silenced by DNA methylation in a variety of tumors (Suzuki, Gabrielson et al. 2002).

There are two other types of expression profiling microarray platforms in current use. One uses a single fluorophore such as biotin to label a single RNA sample, and the other uses two separate dyes, which emit at different wavelengths, to label two different samples of RNA – one test and one reference (the same reference is often used against all test samples in an experiment). Both platforms have their advantages and disadvantages, but in the case of DNA methylation studies, the single color array platform is superior because analysis of the two color array platform data requires detection of signal in both Cy3 and Cy5 channels for a spot to be included in subsequent analyses. Since genes that are methylated in untreated cancer cells do not express any RNA for that locus, only the reference dye will emit. Therefore, most studies use the single color method platform such as the Affymetrix Genechip platform.

As mentioned above, 5-aza not only integrates into DNA, but also irreversibly binds protein to DNA. It is both cytotoxic and genotoxic, and also likely affects critical metabolic pathways such as folate metabolism and nucleotide synthetic pathways (Stresemann, Brueckner et al. 2006). This means that microarray experiments comparing gene expression

changes before and after treatment with this drug need to be interpreted carefully; it is absolutely critical to perform biological replicates using the cancer cell lines before and after treatment with 5-aza, and preferably using a least two different doses of drug. It is important to validate the results of the screens at both the level of gene expression (quantitative PCR) and promoter methylation (MSP or bisulfite sequencing).

A further consideration is that it is not entirely clear that the most popular statistical algorithms used to identify differentially expressed genes such as Significance Analysis of Microarrays (SAM) and cluster analysis are suitable for DNA methylation studies that employ transcriptional profiling. Primarily this is because in a standard comparison between control and 5-aza treated cancer cell lines, a “differentially expressed gene” is an unlikely methylation candidate. For methylation studies, interesting genes are those that are *not* expressed in untreated controls but that are induced by 5-aza treatment; genes that are expressed in both controls and treated cells even if they go up after treatment or alternatively those that go down after treatment are interesting, but their relationship to promoter methylation is likely indirect. Therefore, rote application of parametric algorithms in this context is imprudent as they exclude many genes from the “significant gene” list because they are not expressed in most of the samples. One solution to this problem, which I developed in this thesis, is to include controls such as immortalized cell lines that express tumor suppressor genes to properly balance the analysis (Shames 2006).

RESTRICTION LANDMARK GENOMIC SCANNING FOLLOWED BY 2D GEL ELECTROPHORESIS OR CPG ISLAND MICROARRAY

One of the major pitfalls of using 5-aza to screen cells for methylated genes is that 5-aza alters the expression of many genes, and thus, microarray expression profiling experiments do not evaluate methylation *per se*; rather they identify genes that respond to 5-aza treatment. The importance of this difference is highlighted in several studies where nearly half the genes induced by 5-aza do not have CpG islands (Suzuki, Gabrielson et al. 2002; Sato, Fukushima et al. 2003). To look at methylation directly, some groups have employed a method called restriction landmark genomic scanning (RLGS) (Costello, Fruhwald et al. 2000; Costello, Smiraglia et al. 2002; Smith, Lin et al. 2006). In this method genomic DNA from paired samples – ideally from DNA isolated from cancerous and adjacent normal tissue from the same organ in the same patient, however as this method requires a significant amount of high-quality DNA, usually it is performed on unmatched cell lines – is first digested with a six base cutter restriction enzyme that is sensitive to the presence of methyl-cytosine (does not cut methyl-cytosine), such as *NotI*.

After digestion, the DNA is labeled with dCTP³² and dGTP³², and cut again with another six-base cutter. Since each fragment will contain one molecule each of dCTP and dGTP the method is theoretically quantitative, and the sequential restriction digestions yields a product that can be resolved by agarose gel in the first dimension. After a suitable time, another, more frequent cutting restriction enzyme is used to cut the DNA *in situ*, then the DNA is transferred to polyacrylamide gel, turned by 90° and run perpendicular to the previous direction.

Once the gels from the two samples have been resolved in both dimensions, they are compared to identify spots that are present in the normal sample, but not in the cancer. These fragments may represent aberrantly methylated sites, since *NotI* does not cut methyl-CpG. Using this approach identifies only those CpG islands that contain *NotI* sites, which has been estimated at approximately 10% of CpG islands in the human genome –based on an estimate of approximately 45,000 CpG islands which 10-fold lower than the present estimate (Smiraglia and Plass 2002). To increase coverage various approaches have been used including clone libraries, and different combinations of enzymes. Recently bioinformatic algorithms have been designed to create virtual 2D gel maps of RLGS screens enabling easy identification of differentially staining positions on a gel.

Whatever method is used to discriminate novel methylated sites, the fragment still needs to be cloned out of the gel, which can be tedious. Thus while RLGS is more specific with respect to the identification of tumor specific methylation differences than the 5-aza induction approach described in the previous section, there are several disadvantages that make this method more difficult. First, the samples need to be of high quality and relatively abundant, detection is limited to the diversity of the clone library, and distinguishing bands need to be cloned and sequenced. Furthermore, less than half of the *NotI* clones in these libraries represent gene loci, and while tumor-specific differences in methylation of these sites is interesting in general, their biological relevance is still controversial (Costello, Fruhwald et al. 2000). Therefore, for each *NotI* clone that associates with a given gene, the relationship that between methylation and expression needs to be demonstrated independently.

In an adaptation of this method, restriction digested DNA can be analyzed by CpG island microarray (Heisler, Torti et al. 2005). To use this method, differentially digested DNA fragments are labeled with Cy3 or Cy5 and hybridized to microarrays using DNA from the methylation insensitive restriction digestion as reference and the methylation sensitive digestion as the test set. Differential fluorescence indicates the presence or absence of methylated CpG islands.

CHIP ON CHIP

Recently a novel approach, based on chromatin immunoprecipitation analysis, called methylated DNA immunoprecipitation (MeDIP) was developed to tackle the two main problems associated with the methods described above: specificity in the case of microarray expression profiling, and sensitivity in the case of RLGS (Weber, Davies et al. 2005; Keshet, Schlesinger et al. 2006). In this method, sonicated or restriction digested genomic DNA is immunoprecipitated using a monoclonal antibody to 5-methylcytosine. The immunoprecipitated DNA is then labeled and hybridized to a DNA microarray spotted with probes corresponding to particular regions of interest (such as CpG island libraries), promoter arrays, or a whole genome tiling arrays (Bernstein, Mikkelsen et al. 2006; Keshet, Schlesinger et al. 2006; Lee, Jenner et al. 2006).

Whole genome tiling arrays are generally of the two-color variety, but instead of labeling one RNA sample with Cy3 and the other with Cy5, DNA is used instead. The detection method is the same as in the RNA profiling system in that spots enriched for one or the other sample will excite more in the Cy3 or Cy5 channel, respectively. Analysis of CGH microarray data has different analytical hurdles to deal with than RNA microarrays in that positional information must be integrated into the gain or loss of signal. Unlike RNA microarrays, probes are generally “tiled” over a particular locus so each probe overlaps with adjacent probes. This means that data from adjacent probes can be used to average copy number differences over a particular region, which works as an important internal control of data quality. CGH data are relatively ‘noisy’ in that local gains and losses can appear to be large by differential fluorescence intensity on a particular spot, but that are also

discontinuous and thus likely spurious. Therefore, researchers apply “smoothing” algorithms that combine data from several adjacent probes resulting in a kind of moving average along the chromosome for regional gains and losses.

When a whole genome-tiling array is used with meDIP, the reference sample is an aliquot of input DNA and the test sample is the immunoprecipitated portion of the sample. Methylated sequences are detected by comparing the fluorescent signal for each probe corresponding to known genomic sequences for input and test samples. Loci that are enriched in the test sample are potential methylation candidates. In theory, the MeDIP method circumvents the specificity problems associated with 5-aza expression profiling experiments and the sensitivity issues found in RLGS because it selects sequences by a specific antibody first, but can still detect any sequence in the genome. However, it is likely that the DNA shearing or cleaving and immunoprecipitation steps need to be optimized for different cell types similar to conventional chromatin immunoprecipitation experiments making this assay difficult to perform on many samples. Moreover, this step probably introduces bias toward certain types of sequences in the genome where there is a high density of CpG sites, while it is unclear whether CpG site density *per se* correlates with tumor-specific promoter methylation.

For RLGS, CpG island microarrays, and meDIP, the relationship between methylation and expression needs to be established after the initial screen. It is not sufficient to assume that tumor specific methylation always correlates with expression. All of the above-described methods are designed to identify novel methylation targets, and provide the first step in developing clinically relevant markers for early detection screening, but all have

their limitations. Researchers using any of these methods must balance the likelihood of discovering false positives and making the screen too stringent to detect important methylation targets; importantly, these decisions should be fixed firmly in biology, as well as statistics. As with all high throughput methods in molecular biology, confirmation and validation of positive hits from screening procedures are essential.

SUMMARY

After more than 20 years of research, it is now clear that altered patterns of DNA methylation involving both genome wide loss of methylation and locus specific hypermethylation are essential events to cellular transformation (Jones and Taylor 1980; Feinberg and Vogelstein 1983; Baylin and Ohm 2006). Led by innovative technology developments in diverse fields, epigenetics research has entered a new era of discovery and has become central to our understanding cancer biology. The study of these changes has evolved from a candidate gene approach where single alterations in DNA methylation content at promoter regions were thought to result from a random process, to global epigenetic profiles that are beginning to be understood in the context of cancer cell physiology and homeostasis. Undoubtedly, the next series of global profiling studies will include combined analyses of both methylation and expression microarrays. Hopefully these types of studies will shed some light on how aberrant DNA methylation is initiated in the early stages of carcinogenesis, and whether there are common sequence motifs or protein binding elements that coincide with common regions of hyper- or hypomethylation. Development and refinement of these findings hold great promise for our understanding of the basic science of cancer, as well as providing a platform for its early detection and treatment.

Table 2-1. Methods for Detecting DNA Methylation

Method	Through-put	Material	Application	Limitations	Citation
Methylation specific PCR	Relatively high for samples; low for genes	Patient DNA, cell line DNA, bodily fluids (sputum, blood, urine)	Screening for new methylation markers, analyzing relatively large numbers of samples for known markers	Qualitative	Herman et al. (1996), Zochbauer-Muller et al. (2001)
Quantitative MSP (Taqman, sybr-green)	Relatively high for samples, difficult to set up, every gene requires empirical standardization	Patient DNA, cell line DNA, bodily fluids (sputum, blood, urine)	Quantitative analysis of patient DNA, correlations with clinical parameters	Requires substantial optimization, limited target sequences	Shivapurkar et al. (2006), Fackler et al. (2004)
Pyrosequencing	Moderate for samples and genes, high throughput for sequencing	Patient DNA, cell line DNA, bodily fluids (sputum, blood, urine)	Quantitative analysis of patient DNA, correlations with clinical parameters, analysis of more CpG sites than MSP	Expensive, requires dedicated equipment	Collella et al. (2003)
Microarray expression profiling after reversal of DNA methylation	Low for samples, high for genes	Cell lines	Novel marker discovery	Large type I and type II error rates	Suzuki et al. (2001), Shames et al. (2006)
Restriction landmark genomic scanning	Low for samples, high for genes	Cell lines	Novel marker discovery	Limited coverage (restriction libraries), no correlation with expression	Costello et al. (2000)
ChIP-on-chip	Low for samples, high for genes	Cell lines, potentially tumor material	Novel marker discovery	Novel, unknown biases, no correlation with expression	Weber et al. (2005), Keshet et al. (2006)

CHAPTER THREE

MATERIALS AND METHODS

MICROARRAY ANALYSIS AND BIOINFORMATICS

RNA quality and microarray analysis. The quality of total RNA was analyzed by formaldehyde gel and/or by capillary electrophoresis on the Experion System (Bio-Rad). Total RNA was labeled and amplified by our genomics core facility, according to manufacturer's instructions (<http://www.affymetrix.com>). cRNA was re-analyzed after labeling to ensure optimal amplification. cRNA was hybridized to U133 Plus 2.0 (~47000 transcripts) or U133A (~18400 transcripts) (Affymetrix, Santa Clara, CA), and scanned by our microarray core facility (<http://microarray.swmed.edu/>). RNA for other lung cancer samples were obtained as part of collaborations with William Gerald at Memorial Sloan-Kettering Cancer Center (New York dataset) and Chi-Leung Lam and Maria Wong at the University of Hong Kong. All samples were collected with appropriate consent and internal review board approval. Expression analysis of microarray data was performed using several algorithms: Robust Multichip Averaging (RMA) (Bolstad, Irizarry et al. 2003; Irizarry, Bolstad et al. 2003), Microarray Analysis Suite 5.0 (Affymetrix), MATRIX 1.29 (Girard et al., 2006, see below) NIH-DAVID (Dennis, Sherman et al. 2003), Cluster, and TreeView (Eisen, Spellman et al. 1998).

After scanning, arrays were checked for quality using GCOS and then normalized to each using either RMA or MATRIX 1.29. For log ratio calculations using MAS5 normalization

(MATRIX 1.29), the only requirement was that the numerator be present (p -value < 0.065). Data were then logged and renormalized. For RMA normalization, all data were compiled using RMA Express, or RMA through R or BRBArrayTools.

MATRIX (MicroArray TRansformation In eXcel) is a Microsoft Visual Basic program that allows import of multiple CHP files (saved as text file format) from Affymetrix MicroArray Suite 5.0 into an Excel spreadsheet where median normalization, comparison of arrays using log ratios and t-tests, color display and hierarchical clustering can be performed. Specifically, expression signals are first \log_2 -transformed and a color-coded such that higher signals are displayed as darker (blue) colors. Absent (high detection p -value) signals are optionally coded separately on a gray scale. For comparison of samples or classes of samples, \log_2 ratios (i.e. difference of \log_2 -transformed signals) are calculated. If samples are compared, the stronger signals must have a present call (detection p -value < 0.05). If classes of samples are compared (as log ratios of the means), the median of the detection p -values for the class with the highest mean expression value must be less than 0.05. Two-sample t-tests are further calculated to filter out univariate non-significant differential expression. Hierarchical clustering is performed using average linkage with a Pearson correlation metric. All analyses are performed using extensive gene annotation and all probes are BLAST-verified. MATRIX has not been released, as it is still under development. While this program was used extensively in these studies, all analyses were reproduced using publicly available software. Please contact Luc Girard (Luc.Girard@utsouthwestern.edu) for further details.

Statistical methods. For CpG island enrichment analysis, intersect tables between the relevant RefSeq gene lists and CpG island annotations were generated using the Table Browser function at the Genome Browser database (http://genome.ucsc.edu/cgi-bin/hgTables?org=Human&db=hg17&hgsid=73574615&hgta_doMainPage=1). Statistical significance for the resultant data was determined using the χ^2 method where the expected value for 5' CpG islands for RefSeq annotations was ~37% based on the May 2006 genome build.

Statistical analysis for the primary tumor gene expression data was based on the significance analysis of microarray algorithm (SAM) implemented through BRB ArrayTools developed by Dr. Richard Simon and Amy Peng Lam. Statistical significance of the methylation data was determined using the χ^2 method where appropriate.

Correlations between array and QPCR data were determined using the Pearson correlation coefficient. Cluster analysis was performed using Cluster and Treeview either through BRB ArrayTools or directly. Agreement analysis for biological replicate array data was performed as follows: Affymetrix U133 Plus 2.0 .cel files were normalized using RMA implemented through the "Affy" R package (version 1.8.1) from Bioconductor (<http://www.bioconductor.org/packages/bioc/1.7/src/contrib/html>). To evaluate the consistency of the most differentially expressed genes from biological replicate experiments, we considered a gene to be in agreement if in both experiments, the gene was up or down regulated in the same direction compared to control. The agreement analysis consists of the following steps: 1) calculate \log_2 (expression value of the treated cell/expression value of the control cell of RMA normalized data) for each cell line in each experiment; 2) select the top

1000 or 2000 up or down regulated genes from each experiment; 3) extract genes that were common between the replicate experiments (union gene set) – genes that were in the top or bottom 1000 or 2000 genes in both experiments; 4) calculate the proportion of genes in common for each union data set, which yields a point estimate for the proportion of agreement: $(\# \text{ of pairs that move in the same direction})/(\# \text{ of pairs in the union set})$; 5) for each data set obtain 5,000 bootstrap samples drawn with replacement from the original dataset; 6) calculate the median and 95% confidence interval (2.5% and 97.5%) for the agreement proportion (Efron and Tibshirani 2002). The total number of genes and ESTs on the array is 54,675.

Enrichment analysis for gene ontology and chromosomal location was performed using NIH-DAVID (<http://david.abcc.ncifcrf.gov/home.jsp>), using text files containing accession number lists of Affymetrix probe IDs or Genbank accession numbers. Statistical enrichment was determined using a Fisher Exact test where the null hypothesis is that no difference exists between the number of genes falling into a given ontology in the input list than there is in the genome as a whole (Dennis, Sherman et al. 2003).

Normalization: Normalizations were performed using Robust Multichip Averaging (RMA) (Bolstad, Irizarry et al. 2003; Irizarry, Bolstad et al. 2003), Microarray Analysis Suite 5.0 (MAS5) (Affymetrix), MATRIX 1.29 (Girard et al., 2006). RMA normalizations were used prior to cluster analyses as well as some “fold-change” analyses. This is because RMA normalized data tends to be smoother than MAS5 normalized data (Irizarry, Bolstad et al. 2003). The RMA algorithm is available online at <http://www.stat.berkeley.edu/~bolstad/RMAExpress/RMAExpress.html>.

Gene selection: Genes were selected on the basis of their expression patterns in the different cell types before and after 5-aza treatment normally using Microsoft Excel formulas and macros. Different combinations of criteria were used for different analyses and are indicated in the data tables and figures. Examples of freeware packages used in these studies include EASE, BRB ArrayTools developed by Dr. Richard Simon and Amy Peng Lam, Cluster, and TreeView (Eisen, Spellman et al. 1998), which are all available online. The primary array analysis software package (MATRIX 1.29) that was used in these studies was developed by Luc Girard. This program imports multiple .cel files (saved as text files) from MAS 5.0 into an Excel spreadsheet, and performs various analyses: mean and median normalization, grouping, comparison of arrays using log ratios, scatter plots, scaled coloring, hierarchical clustering, etc.

CpG island analysis: A useful website for CpG island analysis and gene associations may be found at <http://www.charite.de/ch/medgen/cpg/>. To determine whether 5-aza induced gene lists were enriched for genes associated with CpG islands, intersections of the RefSeq annotations and CpG island positional databases were performed using the Genome Browser Table website (http://genome.ucsc.edu/cgi-bin/hgTables?org=Human&db=hg17&hgside=73574615&hgta_doMainPage=1). Primary databases comprising different annotation lists derived from microarray analysis were uploaded into Genome Browser and converted into custom tracks. Then each reference database was intersected with the CpG island positional annotation track within the “regulation and expression” group. The basic procedure for this is as follows:

1. From the Table Browser select the refGene table; set the position to genome, and upload the identifier list.
2. Select custom track output; select the "Create 1 BED record per 5' UTR Exons" option, and load the track into the Table Browser.
3. Select the custom track; then intersect it with the cpGIslandExt table.
4. Output the intersection as BED or custom track.
5. Copy and paste BED file into notepad and save.
6. Import data into Excel using "delimited" option; click next; leave tab delimiter and click other and add an underscore.
7. Concatenate "NM" column with identifier adding an underscore in between.

The statistical significance of gene associated CpG islands was performed using a 2x2 Chi-square where the expected association of a 5' CpG island with a RefSeq annotation is ~37% based on the May 2006 genome build (See *Chapter 6* for more details).

Analysis of the distribution of CpG islands in the genome was performed using the UCSC Table Browser (http://genome.ucsc.edu/cgi-bin/hgTables?org=Human&db=hg17&hgsid=73574615&hgta_doMainPage=1) and rendered on the genome using Vega http://vega.sanger.ac.uk/Homo_sapiens/karyoview.

Statistical methods for agreement analysis. Data from Affymetrix U133 Plus 2.0 .cel files were normalized using RMA implemented through the "Affy" R package (version 1.8.1) from Bioconductor (<http://www.bioconductor.org/packages/bioc/1.7/src/contrib/html>). To evaluate the consistency of the most differentially expressed genes from biological replicate experiments, we considered a gene to be in agreement if in both experiments the gene was up

or down regulated in the same direction compared to control. The agreement analysis consists of the following steps: (i) calculate \log_2 (expression value of the treated cell/expression value of the control cell) for each cell line in each experiment; (ii) select top 1000 or 2000 up or down regulated genes; get union set of the two experiments for each cell line between the top 1000 or 2000 up or down gene set; (iii) calculate the proportion of agreement genes for each union data set, which yields a point estimate of agreement proportion; (iv) for each data set obtain 5,000 bootstrap samples (re-sampling labels of the genes), calculate the median and 95% confidence intervals (2.5% and 97.5%) for the agreement proportion. Bootstrap data points are a random sample of size N (same length as the original data) drawn with replacement from the population of n objects (Efron and Tibshirani 2002).

Useful bioinformatics portals -

Skip Garner's website at UTSW - http://innovation.swmed.edu/research/res_inf.html

Several links to array analysis tools and gene databases - <http://www.biotech.ufl.edu/WorkshopsCourses/bioinfoWorkshops/bioinfoTools/bioinfoTools.html>

"R" statistical package website - <http://www.r-project.org/>

Freeware for QPCR primer design - <https://www.genscript.com/ssl-bin/app/primer>

Portal for all kinds of useful websites - <http://www.bioinformatics.vg/biolinks/bioinformatics/PCR%2520and%2520Primer%2520Design.shtml>

LAB PROTOCOLS

Cell lines and 5-aza-2'-deoxycytidine treatment. With the exception of A549, HCT116, SKBR3, ZR-75-1, and MCF7, which were purchased from the American Type Culture Collection (ATCC), all tumor cell lines were established by us and are deposited at the ATCC, or are available upon request (Phelps, Johnson et al. 1996; Gazdar, Kurvari et al. 1998). All cancer cell lines were grown in RPMI-1640 medium (Life Technologies Inc., Rockville, MD) supplemented with 10% fetal bovine serum. Immortalized human bronchial epithelial cells (HBECs) were established by us (Ramirez, Sheridan et al. 2004). In the present study, unless otherwise indicated, HBEC cells ectopically express murine *cdk4* and *hTERT*. HBEC cell lines were grown in KSFM medium supplemented with bovine pituitary extract and recombinant human epidermal growth factor (Gibco, Carlsbad, CA). All cell lines were grown in a humidified atmosphere with 5% CO₂, at 37°C. A 50 mM stock solution of 5-aza-2'-deoxycytidine (5-aza) (Sigma, St. Louis, MO) was prepared in dimethylsulfoxide (DMSO) and kept at -80°C until used. Working dilutions were prepared from aliquots using DMSO prior to each treatment. Cell lines were incubated in culture medium with 100 nM or 1 μM 5-aza for 6 days, with medium changes on days 1, 3, and 5. Cells were harvested and total RNA extracted on day 6 using Trizol (Invitrogen, Carlsbad, CA).

Primary Tumors – Primary, resected, non-small cell lung carcinomas and corresponding normal lung tissue DNA was extracted as previously described (Zochbauer-Muller, Fong et al. 2001). 20 primary lung tumor samples and corresponding nonmalignant lung were randomly selected from a larger panel ($n = 107$) obtained from NSCLC patients who had

been treated with curative resectional surgery in The Prince Charles Hospital (Brisbane, Australia) between June 1990 and March 1993. This cohort of patients has been investigated previously for various genetic abnormalities (Fong, Zimmerman et al. 1994; Fong, Kida et al. 1995; Fong, Schonrock et al. 1995; Fong, Zimmerman et al. 1995; Fong, Zimmerman et al. 1995). There were 76 males and 31 females (age, 28–81 years; mean age at diagnosis, 61 years). Sixty-one patients had stage I disease, 21 patients had stage II disease, 24 patients had stage IIIA disease, and 1 patient had stage IIIB disease. Histological subtypes included 45 adenocarcinomas, 43 squamous cell carcinomas, 11 adenosquamous carcinomas, 4 large cell carcinomas, 3 atypical carcinoids, and 1 typical carcinoid. Ninety-eight patients were smokers (mean pack-years, 31), and the rest of patients were never smokers or nonsmokers. Survival data of 5 or more years were available on most patients. Breast tumor DNA from the University of North Carolina, the University of Chicago, and Thomas Jefferson University was prepared as previously described (Usary, Llaca et al. 2004). All samples were collected with internal review board approval. Breast sample collection from UT Southwestern was approved by the Institutional Review Board at UT Southwestern Medical Center and written informed consent was documented for each subject. Random periareolar fine needle aspiration (FNA) was performed as previously described except that the FNA sample was fixed in Preservcyt (Cytoc Corporation, Marlborough, MA) (Lewis, Cler et al. 2005). DNA was extracted using the Puregene kit (Gentra Systems, Inc., Minneapolis, MN). All samples in this set were obtained from patients diagnosed with \leq stage IIb breast cancer. Benign and malignant prostate and colon DNAs were obtained through the University of Texas Southwestern Tissue Resource (UTSTR) overseen by the University of Texas

Southwestern Medical Center Institutional Review Board. Tissues were retrieved from the operating room and samples were snap frozen in liquid nitrogen within 30 minutes off of blood supply. The samples were stored at -80 C until the DNA was isolated using the Qiagen DNA Isolation Kit (cat # 51306). The final DNA product was stored in TE buffer at -80 C until retrieved for sodium bisulfite modification. All DNAs in this group of samples were obtained from patients with stage 2 or 3 malignancies.

Sodium Bisulfite Treatment, Sequencing, and Methylation Specific PCR. Sodium bisulfite treatment for the UTSW breast FNAs were performed as previously described, using yeast tRNA as a carrier (Clark, Harrison et al. 1994). Sodium bisulfite modification of genomic DNA for the remaining samples and methylation specific PCR were performed as reported by Herman *et al.* with some modification to increase through-put (Herman, Graff et al. 1996). We modified the protocol to work in 96-well format as follows: 2 µg of genomic DNA was subjected to sodium bisulfite treatment as before except that samples were incubated in deep-well (1 ml) 96-well plates using a silicon seal (Nunc, Rochester, NY), and reagent concentrations were modified to allow the use of a repeat pipettor (Eppendorf, Hamburg, Germany). An equal volume of membrane binding solution (Promega, Madison, WI) or 4M guanidine isothiocyanate (Sigma) was added to the bisulfite reaction after 16 hrs at 50°C. The mixture from each well was transferred into the same well on a binding plate held in a 96-well vacuum manifold and the mixture was evacuated. Bound DNA was washed 3 times with 80% isopropanol, then desulfonated *in situ* with 100 µl of 0.2N NaOH for 10 minutes at room temperature. 100 µl of membrane binding solution or 4M guanidine isothiocyanate was added, then evacuated. The desulfonated, bisulfite DNA was washed 2

more times in 80% isopropanol, and kept under vacuum for 4 minutes after the last wash to dry the membrane. DNA was eluted into a collection plate with 100 μ l of warm ($\sim 65^{\circ}\text{C}$) nuclease free water and further diluted to 250 μ l before analysis.

Methylation specific PCR primers were designed in part by using MethPrimer, however substantial modification was necessary in most cases (Li and Dahiya 2002). 45 of the 132 gene 5-aza induction panel were selected for methylation analysis because this number enabled accommodation to a 96-well plate format including 2 control sequences (*TKTL1* and *GAPDH*; total 94 primer sets), and two blank wells for negative controls. We ensured that the MSP primers targeted a region within 250 bps of the annotated transcription start site, where possible (UCSC Genome Browser and RefSeq), and that they contained ≥ 3 CpG sites per primer (most contained ≥ 4), had a 3'-proximal CpG site, and had predicted annealing temperature of $\geq 55^{\circ}\text{C}$. PCR conditions and primer sequences may be found in (Supplementary Methods I). Primers were purchased from ([Integrated DNA Technologies](#), Coralville, IA) in 96-well format and diluted to 1 μM . 2 μ l of mixed primers were added to the corresponding well on a pre-aliquoted 96-well PCR plates (Invitrogen). PCR plate set up was based on T_m (Fig. 3-1). 2 μ l of diluted bisulfite DNA was added to each well. PCR products were resolved by electrophoresis using 3% agarose in TBE and ethidium bromide. Gels were visualized using a Kodak (Rochester, NY) CCD camera and images were collated using Adobe Photoshop CS2 (San Jose, CA). Several controls gels were performed using different combinations of bisulfite DNA, agarose, and running buffers to ensure that the resolving power of the gel was sufficient to identify the appropriately sized bands from primer dimers, which did appear when no amplicon was present. We were unable to

differentiate bands from background for amplicons <90 bps using our final conditions, which precluded use of *GAPDH* as a control. An optically visible band of the appropriate size was called positive for each primer pair.

Sodium bisulfite sequencing was performed using TA cloning (Invitrogen) as described previously (Janowski, Huffman et al. 2005). Sequencing data was compiled and rendered into lollipop diagrams using BiQ Analyzer software (Bock, Reither et al. 2005).

Real-time RT-PCR. Expression of *LOX*, *NRCAM*, *BNC1*, *CCNA1*, *MAF*, *ALDH1A3*, *CTSZ*, *IRX4*, *MSX1*, *KLF11*, *SERPINB5*, *TKTL1*, *GAPDH*, *r18s*, and *CDKN2A* was analyzed by quantitative real-time RT-PCR. Primers and probes were purchased from Applied Biosystems assay-on-demand, with the exception of *p16*, which was an assay-by-design (Hs00923893_m1) (ABI, Foster City, CA). All samples were run on the Chromo 4 Real Time Detector (MJ Research, Bio-Rad, Hercules, CA) twice, each time in duplicate. We averaged expression of *GAPDH* and *r18s* as internal reference genes to normalize input cDNA. Quantitative real-time RT-PCR was performed in a reaction volume of 20 μ l including 1 μ l of cDNA. We used the comparative Ct method to compute relative expression values.

Comparative Genome Hybridization Array (aCGH) – Cell line DNA was isolated using a phenol/chloroform extraction and ethanol precipitation. Each cell line was fingerprinted prior to analysis. aCGH were performed as previously reported (Pollack, Perou et al. 1999; Ramirez, Sheridan et al. 2004).

Preparation and transfection of siRNAs. siRNAs targeting *DNMT1* were designed and prepared as described previously (Elbashir, Harborth et al. 2001). The two siRNA sequences

against *DNMT1* were 5'-CGGUGCUCAUGCUUACAACCTT-3' (sense) and 5'-GUUGUAAGCAUGAGCACCGTT-3' (anti-sense), and 5'-CGAGUUGCUAGACCGCUUCTT-3' (sense) and 5'-GAAGCGGUCUAGCAACUCGTT-3' (anti-sense). The siRNA sequences against the human T cell leukemia virus gene (*Tax*) and *Lamin A/C* were as previously reported (Elbashir, Harborth et al. 2001; Verma, Surabhi et al. 2003). The siRNA target sequences were tested in a BLAST search of GenBank (NCBI database) to ensure that only the corresponding gene is the target. RNA oligonucleotides were obtained from the core facility in University of Texas Southwestern Medical Center (see http://cbi.swmed.edu/pages/oligonet_index.htm for details). The sense and antisense oligonucleotides were annealed to make siRNA (Tuschl 2001) and stored at -20°C before use. One day prior to transfection, cells were seeded such that they were 30-50% confluent the next day. Cells were transfected with 100 nM of siRNA using Oligofectamine transfection reagent (Invitrogen, CA) in Opti-MEM I reduced serum medium (Invitrogen) at 37°C in a 5% CO₂ atmosphere for 6 h. The medium was removed and replaced with fresh RPMI supplemented with 5% fetal bovine serum. Control cells were treated with Oligofectamine alone or with *Tax* and *Lamin A/C* siRNA. Transfection was repeated at 2, 4, and 6 days for a total of 4 treatments. Cells were grown and harvested at 3, 5, 7, 9, 14, and 23 days after the initial transfection for further analysis.

Western blot analysis. Cells were grown and harvested at 80-90% confluency, and cellular proteins were extracted with lysis buffer (40 mM Hepes-NaOH, pH 7.4, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% SDS, 150 mM NaCl) containing Complete Mini, a cocktail of protease inhibitors (Roche, IN). Total protein was electrophoresed on

SDS/polyacrylamine gel and transferred to nitrocellulose membranes (Schleicher & Schuell, NH). After blocking with 5% nonfat dry milk and 0.1% Tween 20 in Tris-buffered saline, membranes were incubated with the mouse monoclonal anti-DNMT1 (IMGENEX, CA), the rabbit polyclonal anti-DNMT3B (a kind gift from Dr. A. Robert MacLeod), the rabbit monoclonal anti-p16^{ink4A} (Santa Cruz, CA), or the mouse monoclonal anti-Actin (Sigma, MO) antibodies. The membranes then were developed with peroxidase-labeled antibodies (Amersham Pharmacia, NJ) by Super Signal chemiluminescence substrate (Pierce, IL). Actin protein levels were used as a control for equal protein loading.

Cell fractionation: Frozen cell pellets were resuspended in 4 volumes of hypotonic buffer (10mM HEPES-KOH, 1.5mM MgCl₂, 10mM KCl, 0.5mM DTT (dithiothreitol), 0.2mM PMSF (phenylmethylsulfonyl fluoride), 50 mM NaF, 1 mM Na Orthovanadate, 0.5mM B-glycerophosphate, H₂O, 1 tablet of complete mini protease inhibitor (- EDTA)), and incubated on ice for 10 minutes, vortexing on low speed every 2 mins. Cell suspensions were then pelleted by centrifugation at 14000 rpm for 30s on a bench top centrifuge. Supernatant was removed and snap frozen for the cytoplasmic fraction. Pelleted nuclei were lysed in 4-6 volumes of SDS lysis buffer (50mM NaCl, 1mM EDTA, 2.5mM Tris pH 7.4, 0.1% SDS, 1% NP40, dH₂O) and incubated on ice for 15-20 minutes. Lysates were then centrifuged for 30 minutes at 4°C at 12000g.

Soft agar-growth assay. Cells were transfected with siRNAs for a total of 4 treatments, and 7 days after the initial transfection, cells were replated for soft agar-growth assay. Briefly, 300 viable cells were suspended and plated in 0.33% agar in RPMI-1640 medium (Life Technologies Inc.) supplemented with 20% fetal bovine serum and layered over a 0.50%

agar base medium in 12 well plates. After 2 weeks, the number of colonies >100 cells were counted in triplicate plates.

Microtiter growth assay: Cell growth was assayed by MTT or MTS (CellTiter 96® AQueous One Solution Cell Proliferation Assay, Promega) assays according the manufacturer's instructions. Cells were plated 24 hrs prior to addition of drug. Absorbance measurements were determined at 570nm for MTT and 490 nm for MTS 96 hrs after addition of drug. Assays were performed in duplicate 96 well plates until a minimum of three plates produced a standard deviation smaller than the mean.

Quantitative methylation-specific PCR (MSP) Assay. Genomic DNA was obtained from cell lines by digestion with proteinase K (Life Technologies), followed by phenol/chloroform (1:1) extraction. One µg of genomic DNA was denatured with 2N NaOH and modified with Sodium bisulfite, as previously described (Herman, Graff et al. 1996). The modified DNA was purified using the Wizard DNA purification kit (Promega, Madison, WI), treated with 3N NaOH, and precipitated with ethanol and resuspended in water. Sodium bisulfite-treated genomic DNA was amplified by fluorescence-based real-time MSP (Perkin-Elmer Corp., Foster City, CA) as described previously (Toyooka, Toyooka et al. 2002). For the internal reference gene, *MYOD1*, the primers and probe were designed to avoid CpG nucleotides. The methylation ratio is defined as the ratio of the fluorescence emission intensity values for the target PCR products to those of the *MYOD1* PCR products, multiplied by 100. The ratio is then divided by the ratio of the non-treated sample and multiplied by 100 to yield a percentage. The sequences of the primers and probes are shown in Table 1. Quantitative real-time MSP assays were performed in a reaction volume of 25µl by using components

supplied in a TaqMan PCR Core Reagent Kit (Perkin-Elmer Corp.). Each assay was performed in triplicate. The final reaction mixtures contained the forward and reverse primers at 300 nM each; the probe at 100 nM; 200 μ M each of dATP, dGTP, dCTP and dTTP; 5.5 mM MgCl₂; 1x TaqMan Buffer A; 1 unit of HotStarTaq DNA polymerase (QIAGEN Inc., Valencia, CA); and 2 μ l bisulfite-converted genomic DNA. PCR was performed under the following conditions: 95°C for 12 minutes, followed by 50 cycles of 95°C for 15 seconds and 60°C for 1 minute. We performed quantitative real-time MSP with the Gene Amp 5700 Sequence Detection System (PerkinElmer Corp.). DNA from lymphocytes of a healthy volunteer treated with Sss1 methyltransferase (New England BioLabs, Beverly, MA) was used as a positive control. The same untreated, unmethylated DNA was used as a negative control for methylated alleles. Water blanks were included with each assay.

Figures

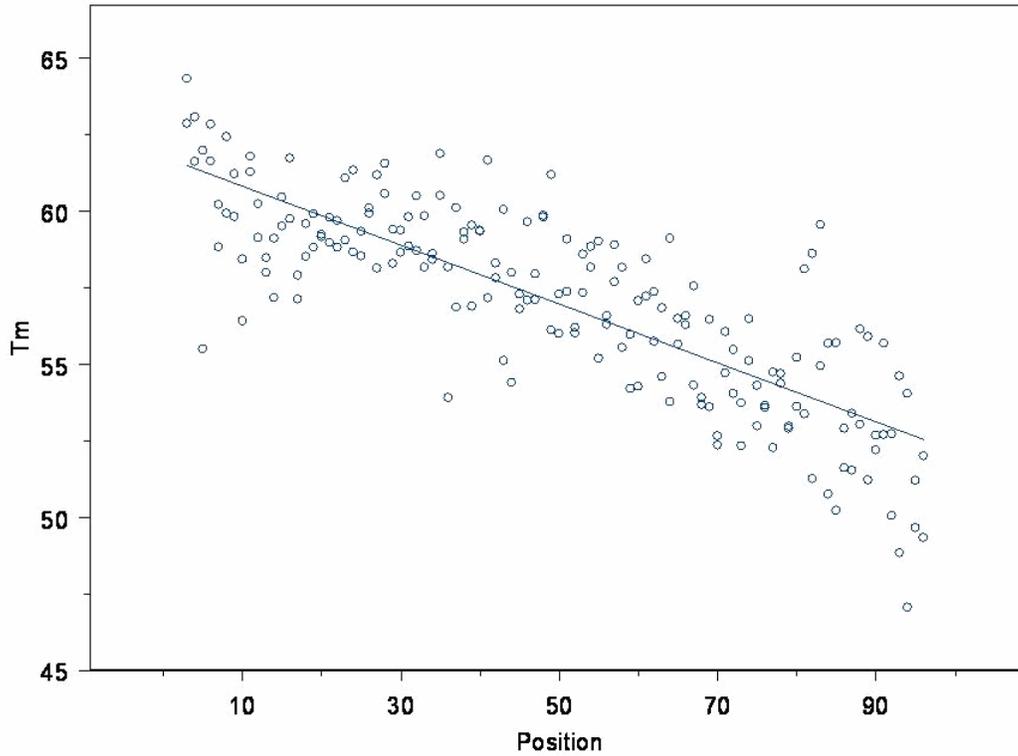


Figure 3-1. Regression of primer melting temperature by position on PCR plate. The strategy was to align primers with position on the PCR plate such that different annealing temperatures could be used if required. Pre-aliquoted PCR plates were used for these reactions which had been perforated every three columns. This enabled the use of two sets of conditions which turned out to be all that was needed for this reaction.

CHAPTER FOUR

INTRODUCTION

The primary goal of the work described in this thesis, was to identify all of the genes subject to aberrant promoter hypermethylation in lung cancer. The studies described in this and the following chapter were pilot experiments designed to determine the appropriate approach to use in subsequent microarray studies, which are described in *Chapter 6*. This chapter is derived from a published manuscript, and describes experiments using RNAi approaches to demethylate DNA in lung and breast cancer cells. Both promoter methylation and gene expression are analyzed using quantitative methods. *Chapter 5* details experiments that were not published, but are a logical extension of the manuscript reproduced below. These studies explore the phenotypic effects of RNAi based approaches to demethylating DNA in cancer and immortalized bronchial epithelial cells compared to 5-aza. In addition, a direct, head-to-head comparison of RNAi and 5-aza in the NSCLC cell H157 by microarray is discussed.

There are three basic methods to demethylate genomic DNA: genetic knockout of the DNA methyltransferase enzymes, RNAi against the DNA methyltransferases, or pharmacological inhibition of the DNA methyltransferases. All were considered for these studies, and both pharmacological and RNAi based approaches were evaluated experimentally. The relative advantages and disadvantages will be discussed briefly below.

Genetic knockout of the DNA methyltransferases

In this method, homologous recombination is used to sequentially remove the DNA methyltransferase loci from host genomic DNA. A targeting vector is used where DNA homologous to the intended endogenous locus flanks a resistance marker such as hygromycin. This strategy enables exogenous Cre-mediated excision of the drug marker after targeting of the first allele and thus the use of the same targeting construct to delete the second allele (Rhee, Jair et al. 2000; Rhee, Bachman et al. 2002). This method requires substantial manipulation of cells and is therefore not suitable for testing multiple different cell lines, as there are three functional methyltransferase enzymes, which means that 6 sequential recombination events are required.

RNA interference mediated silencing of the DNA methyltransferases

Soon after its rediscovery in 1998, RNA interference (RNAi) was shown to be useful for studying gene function in mammalian cells (Fire, Xu et al. 1998; Elbashir, Harborth et al. 2001; Elbashir, Lendeckel et al. 2001). Antisense oligonucleotides targeting the DNA methyltransferases had been shown to suppress DNA methyltransferase protein expression suggesting RNAi approaches might work as well (Fournel, Sapieha et al. 1999).

Pharmacological inhibition of the DNA methyltransferases

Several compounds are available that inhibit the DNA methyltransferase enzymes. The most widely used is 5-aza-2'-deoxycytosine (5-aza), which is a cytidine analogue that has a nitrogen substitution at the 5th carbon of the cytosine ring. This molecule is incorporated into DNA during genome replication and inhibits DNA methylation through

covalent linkage of the DNA methyltransferases to the nascent strand (Jones and Taylor 1980; Jones 1985). DNA demethylation occurs through a passive process of methyltransferase depletion (Michalowsky and Jones 1987).

**RNAI-MEDIATED KNOCKDOWN OF DNMT1 EXPRESSION
REVERSES PROMOTER HYPERMETHYLATION AND INDUCES
EXPRESSION OF MULTIPLE GENES IN HUMAN LUNG AND
BREAST CANCER**

ABSTRACT

DNA methyltransferase 1 (DNMT1) is required to maintain DNA methylation patterns in mammalian cells, and is thought to be the predominant maintenance methyltransferase gene. Recent studies indicate that inhibiting DNMT1 protein expression may be a useful approach for understanding the role of DNA methylation in tumorigenesis. To this end, we used RNA interference (RNAi) to specifically down-regulate DNMT1 protein expression in NCI-H1299 lung cancer and HCC1954 breast cancer cells. RNAi-mediated knockdown of DNMT1 protein expression resulted in >80% reduction of promoter methylation in RASSF1A, p16ink4A, and CDH1 in NCI-H1299, and RASSF1A, p16ink4A, and HPP1 in HCC1954, and re-expression of p16ink4A, CDH1, RASSF1A, and SEMA3B in NCI-H1299, and p16ink4A, RASSF1A, and HPP1 in HCC1954. By contrast, promoter methylation and lack of gene expression was maintained when these cell lines were treated with control small interfering RNAs (siRNAs). The siRNA treatment was stopped and 17 days later, all the sequences showed methylation and gene expression was again dramatically down-regulated, indicating the tumor cells still were programmed for these epigenetic changes. We saw no effects on soft agar colony formation of H1299 cells 14 days after DNMT1 knockdown indicating that either these genes are not functioning as tumor

suppressors under these conditions or that more prolonged knockdown or other factors are also required to inhibit the malignant phenotype. These results provide direct evidence that loss of DNMT1 expression abrogates tumor-associated promoter methylation and the resultant silencing of multiple genes implicated in the pathogenesis of human lung and breast cancer.

Note: The first part of this chapter is substantially the product of a published manuscript (Suzuki, Sunaga et al. 2004), where most of the primary data was generated by Makato Suzuki. The siRNA sequences were designed by Noriaki Sunaga. The original idea for the manuscript came from discussions between Noriaki Sunaga and David Shames. The manuscript was written by David Shames. Noriaki Sunaga and David Shames contributed equally to the data submitted for the revised version of the manuscript. This manuscript is included because it is the result of pilot experiments performed as part of the comparison between pharmacologic and RNAi based approaches to demethylating DNA.

INTRODUCTION

Tumor acquired, aberrantly methylated CpG dinucleotides in the promoter regions of tumor suppressor genes (TSGs) is a hallmark and major means of TSG inactivation in human cancer (Baylin, Herman et al. 1998; Jones and Baylin 2002). Substantial evidence indicates that promoter methylation is associated with loss of TSG expression in lung and breast cancers (Toyooka, Toyooka et al. 2001; Zochbauer-Muller, Minna et al. 2002). The repressed state conveyed by the presence of DNA methylation in TSG promoters can be reversed by administration of the nucleotide analogue 5-aza-2'-deoxycytidine (5-aza) (Bender, Zingg et al. 1998). However, this drug is cytotoxic even at low concentrations, which may lead to expression changes not directly related to DNA methylation (Bender, Pao et al. 1998; Suzuki, Gabrielson et al. 2002). To address this concern, genetic approaches have been used to analyze DNA methylation in cancer.

At present three active DNA methyltransferases (*DNMT1*, *DNMT3A*, and *DNMT3B*) and one candidate gene, *DNMT2*, have been identified in mammals (Jeltsch 2002). *DNMT1*, the first DNA methyltransferase to be cloned, is responsible for maintaining DNA methylation patterns during DNA replication (Szyf 2001). Recently, Fournel and MacLeod showed that ablation of DNMT1 expression with anti-sense oligonucleotides resulted in loss of promoter methylation, re-expression of *p16^{ink4A}*, and inhibition of cell proliferation in the bladder cancer cell line, T24 (Fournel, Sapiha et al. 1999). In contrast, Rhee *et al.* demonstrated that targeted deletion of *DNMT1* by homologous recombination in the colon cancer cell line HCT116 was not sufficient to cause promoter demethylation and gene re-expression. In these experiments, *DNMT1* deletion resulted in only a small decrease

(~20%) in overall genomic methylation, and imprinted genes were not re-expressed (Rhee, Jair et al. 2000). Rhee *et al.* further showed that deletion of both *DNMT1* and *DNMT3B* reduced overall genomic methylation by >95% as well as promoter methylation of specific genes, and caused the re-expression of multiple genes (*p16^{ink4A}* and *TIMP-3*), resulting in substantial growth suppression of HCT116 cells (Rhee, Bachman et al. 2002). Paradoxically, a more recent publication by the MacLeod group showed that DNMT1 depletion using either anti-sense or siRNA techniques led to demethylation of *p16^{ink4A}* and *MLH1* promoters and re-expression of *p16^{ink4A}* in the same HCT116 cells (Robert, Morin et al. 2003). Therefore, it is still unclear how the different *DNMT* genes act alone or in concert, to maintain or establish DNA methylation patterns in individual types of human cancers.

To address this issue, we used RNA interference (RNAi) technology to knock down DNMT1 protein expression in the non-small cell lung cancer (NSCLC) cell line, NCI-H1299, and the breast cancer cell line, HCC1954. Using quantitative assays for DNA methylation and mRNA expression, we found that DNMT1 knockdown led to a dramatic loss of methylation (>80%) compared to non-treated controls at the promoters of *RASSF1A*, *p16^{ink4A}*, *CDH1*, and *HPPI*, and re-expression of *RASSF1A*, *p16^{ink4A}*, *CDH1*, *HPPI*, and *SEMA3B* in lung and breast cancer cells. These findings provide quantitative evidence of the role of DNMT1 activity in both lung and breast cancer cells.

RESULTS AND DISCUSSION

RNAi-mediated knockdown of DNMT1 protein expression in NCI-H1299 and HCC1954

We used RNAi technology to examine the effect of DNMT1 expression on the stability of tumor-associated promoter methylation in lung and breast cancer cells (Caplen, Parrish et al. 2001; Elbashir, Harborth et al. 2001; Sharp 2001). Two siRNAs targeting different sequences of *DNMT1* mRNA were used to verify that our results were a consequence of specific inhibition of DNMT1 expression. In addition, siRNA targeting the human T-cell leukemia virus *Tax* oncogene was used as a negative control, since this viral protein is not expressed in epithelial cells (Verma, Surabhi et al. 2003). Another negative control involved targeting of the expressed *Lamin A/C* gene, since Lamin A/C protein is nonessential in cultured mammalian cells (Harborth, Elbashir et al. 2001). siRNAs both against *DNMT1* (*DNMT1-1* and *DNMT1-2*), *Tax*, and *Lamin A/C* were transfected into NCI-H1299 and HCC1954 cells every 2 days for a week. Cells were harvested at 3, 5, 7, 9, 14, and 23 days after the initial transfection, and Western Blot analysis was conducted to monitor endogenous DNMT1 protein expression (Fig. 4-1). Both siRNAs targeted to *DNMT1* mRNA led to substantial down-regulation of DNMT1 expression 3 days after the initial transfection in the NCI-H1299 cell line, and HCC1954. These effects continued until at least day 9 (Fig. 4-1). We routinely observe targeted gene silencing in ~90% of the NCI-H1299 cells transfected with siRNA as detected by immunofluorescent staining of individual cells (data not shown). DNMT1 down-regulation was specific as evidenced by a consistent level of Actin protein (Fig. 4-1) and DNMT3B protein (Fig. 4-3D) in the context of siRNA targeted

to *DNMT1*, while DNMT1 protein expression was not affected by siRNA targeted to *Lamin A/C* (data not shown).

RNAi-mediated knockdown of DNMT1 expression led to demethylation of tumor suppressor gene promoters in lung and breast cancer cell lines

To assess the effect of RNAi-mediated down-regulation of DNMT1 expression on aberrant methylation in the promoter regions of genes thought to be involved in the pathogenesis of lung and breast cancer in NCI-H1299 and HCC1954 lines, we used a real-time MSP assay to quantitate the degree of methylation before and after DNMT1 knockdown (Toyooka, Toyooka et al. 2002; Toyooka, Carbone et al. 2002). Direct quantitation of the extent of methylation in a particular region of a promoter yields important information about the specificity of DNMT1 activity in terms of the methylation of specific CpGs in the regulatory sequence of particular genes. Several groups have reported that DNMT1 does not have a preference for certain CpG sites or promoters: it appears to act as a general methyltransferase (Szyf 2001; Jeltsch 2002). The kinetics of demethylation for the genes we assessed were not significantly different; therefore, our results are consistent with this hypothesis.

To compare the methylation levels of each gene before and after treatment with siRNA, we converted the mean ratio of promoter methylation to a percentage. RNAi-mediated down-regulation of DNMT1 protein expression resulted in a significant decrease in methylation levels at the *RASSF1A*, *p16^{ink4A}*, and *CDH1* promoters in NCI-H1299 (Fig. 4-2A), and similar effects were observed for *RASSF1A*, *p16^{ink4A}*, and *HPP1* in HCC1954 (Fig. 4-2B, $p < 0.001$; all genes examined, repeated measures ANOVA).

Importantly, the kinetics of demethylation correlate with the loss of DNMT1 expression (Fig 4-1, 4-2A, 4-2B). The level of methylation for all genes was reduced on day

3, more so on day 5, and reached a nadir on day 7, whereupon maximal demethylation seems to have been reached (Fig. 4-2A, 4-2B). As the data indicate, there does not appear to be a significant decrease in methylation levels between day 7 and 9. The methylation level of all the genes tested was reduced by up to 80% when compared with promoter methylation levels in untreated cells. Reduction of promoter methylation was greatest in the *HPP1* promoter in HCC1954 (Fig. 4-2B), and *p16^{ink4A}* in NCI-H1299 (Fig. 4-2A), yet in neither case was methylation completely lost.

It is known that siRNA can be used to specifically knock down target genes, but RNAi never completely eliminates the targeted gene products (Shi 2003). Thus, the presence of basal amounts of promoter methylation we observed, even with extended siRNA treatment, may result from residual DNMT1 protein, or other DNMTs. Other methyltransferases such as DNMT3B or methyl-DNA binding proteins may affect methylation levels in the promoters of TSGs. Rhee *et al.* demonstrated that genetic disruption of *DNMT1* by homologous recombination did not lead to promoter demethylation and re-expression of *p16^{ink4A}* in the colon cancer cell line HCT116, whereas *p16^{ink4A}* was demethylated and re-expressed in HCT116 cells, in which both *DNMT1* and *DNMT3B* were disrupted (Rhee, Bachman *et al.* 2002). Therefore, knockdown of both DNMT1 and DNMT3B or other factors may be required to achieve complete demethylation of genes involved in cancer pathogenesis.

5-aza treatment results in global demethylation of genomic DNA in many cancer cell lines. Upon removal of 5-aza and continued culture, re-methylation occurs slowly and in a sequence specific manner (Velicescu, Weisenberger *et al.* 2002). The propensity of

particular regions of DNA to become re-methylated may result from selective pressure, such as TSG function, or some cryptic sequence information within loci that are preferentially re-methylated. Recent research using 5-aza indicates that *de novo* methylation of CpG sites in the *p16^{ink4A}* promoter is not stochastic. Thus, the kinetics of selective CpG island re-methylation in the promoters of genes may reflect differences in the contribution individual CpG sites have to gene repression. However, due to the non-specificity, and cytotoxicity of 5-aza, it is unclear which DNMT is responsible for the apparent nascent methylation, or whether re-methylation is really the result of the expansion of a resistant subclone within the treated population of cells (Velicescu, Weisenberger et al. 2002).

To address these issues, and to determine how persistent loss of promoter methylation was in the context of the specific down-regulation of DNMT1 protein, we maintained the treated cell lines in the absence of any further siRNA treatment. We then reexamined the methylation level of all genes at day 14 and day 23 after initial treatment. The kinetics of re-methylation varied between genes in both cell lines, however re-methylation (returning to 40-80% of starting levels) and loss of gene expression was observed in all cases by day 23 (Fig. 4-2A, 4-2B). These results indicate that the appearance of *de novo* methylated CpG sites within multiple gene promoters occurs in tandem with the re-expression of DNMT1 protein. This finding clarifies the results from the 5-aza experiments described above, because it suggests that DNMT1 as opposed to DNMT3A, DNMT3B has important *in vivo*, *de novo* DNA methyltransferase activity. A previous report has demonstrated that DNMT1 has *de novo* methylases activity, but only *in vitro* (Yoder, Soman et al. 1997). The variation in re-methylation kinetics between the two cell lines may

result from differences in their doubling times (NCI-H1299 have a doubling time of 25 h, whereas HCC1954 double every 31 h), since *de novo* methylation has been shown to be dependent on cell division (Velicescu, Weisenberger et al. 2002).

Demethylation induced by RNAi-mediated DNMT1 knockdown restored the expression of several tumor suppressor genes in lung and breast cancer cell lines

To establish whether loss of promoter methylation mediated by *DNMT1* siRNA resulted in the quantitative re-expression of genes, we analyzed the expression status of *RASSF1A* and *SEMA3B* genes in NCI-H1299, and *RASSF1A* and *HPP1* genes in HCC1954 line by real-time RT-PCR (Fig. 4-3C, 4-3B). RNAi-mediated DNMT1 knockdown induced the expression of all genes examined ($p < 0.001$, repeated measures ANOVA). The expression levels of all genes in *DNMT1* siRNA-treated cells were 2-8 fold higher than that of untreated cells. We examined the expression status of *SEMA3B* because it is silenced by tumor-associated promoter methylation in NCI-H1299, and is located on 3p21, a known tumor suppressor locus as reported by ourselves and others (Tomizawa, Sekido et al. 2001; Kuroki, Trapasso et al. 2003). The expression level of *p16^{ink4A}* and *CDHI* genes in NCI-H1299, and *p16^{ink4A}* in HCC1954 were examined by 37-cycle end-point RT-PCR. NCI-H1299 cells treated with *DNMT1* siRNA expressed *p16^{ink4A}* mRNA from day 5 to day 23 and expressed *CDHI* from day 3 to day 23 (Fig. 4-3C). HCC1954 cells treated with *DNMT1* siRNA expressed *p16^{ink4A}* from day 3 to day 23 (Fig. 4-3C).

Since the *p16^{ink4A}* gene locus has a complicated structure, it was not possible to design an isoform specific TaqMan probe. Thus, we sought to verify gene induction by Western blot. Both of two different siRNAs targeted to *DNMT1* restored *p16^{ink4A}* protein expression (Fig. 4-3D). Thus, there is a clear inverse relationship between the presence of methyl-CpGs in the promoter of *p16^{ink4A}*, and the expression of *p16^{ink4A}* mRNA and protein. We further compared the effect of the *DNMT1* siRNA (*DNMT1-2*) on the restoration of gene

expression with that of 5-aza treatment in these cell lines. siRNA inhibitors of DNMT1 protein expression are at least as effective at restoring mRNA expression as 5-aza treatment (Table 4-2).

We found that specific inhibition of DNMT1 expression by RNAi is a useful technique to examine the relationship between DNMT1 activity and aberrant promoter methylation in cancer cells. RNAi-mediated knockdown of DNMT1 expression persisted for over 9 days, and was sufficient for achieving the loss of promoter methylation at *RASSF1A*, *CDH1*, *p16^{ink4A}*, and *HPP1*, and re-expression of *p16^{ink4A}*, *CDH1*, *RASSF1A*, *SEMA3B*, and *HPP1* mRNA, which also persisted for 9–14 days, in lung and breast cancer cells. Our findings support and extend the conclusion of MacLeod *et al.* (who used HCT116 colon cancer cells) that DNMT1 siRNA-mediated knockdown alone is sufficient to achieve inhibition of gene methylation with associated gene re-expression (Robert, Morin *et al.* 2003). Our findings, like MacLeod's, differ from that of Rhee *et al.* who found that in HCT116 colon cancer cells, using recombinant knockout techniques, both *DNMT1* and *DNMT3B* had to be removed to achieve loss of methylation and gene re-expression (Rhee, Bachman *et al.* 2002).

We assessed the effect of the DNMT1 knockdown on *in vitro* growth of NCI-H1299 cells by soft agar growth assay. Surprisingly, there was no significant difference in colony number between treatment of *DNMT1* siRNAs (*DNMT1-1* and *DNMT1-2*) and that of *Tax* siRNA (means \pm SD of colony number by treatments with *DNMT1-1*, *DNMT1-2* and *Tax* siRNAs were 115 ± 13 , 100 ± 14 and 108 ± 9 , respectively). Thus, we could not demonstrate an obvious phenotypic effect of the loss of DNMT1 expression on the *in vitro* tumor growth of

NCI-H1299 cells. While this was unexpected, there are several possible explanations for this result. The first is that the genes we monitored (e.g. *RASSF1A* or *SEMA3B*) really do not function as “tumor suppressor genes.” While study of *SEMA3B* as a TSG is early, there are multiple methylation and functional studies of the role of *p16^{ink4A}* and *RASSF1A* strongly implicating them as TSGs in lung and other cancers (Dammann, Li et al. 2000; Ballestrero, Coviello et al. 2001; Burbee, Forgacs et al. 2001; Tomizawa, Sekido et al. 2001; Kuroki, Trapasso et al. 2003). While the tumor cells were plated after 7 days and 4 RNAi treatments the colonies were not scored until 14 days later, it is possible that transient re-expression of the tumor suppressor genes by DNMT1 siRNA was not sufficient to inhibit colony formation due to the short term inhibition of DNMT1 expression. In fact, a recent study showed that prolonged knockdown of DNMT1 by a tetracycline-inducible vector-based siRNA induced growth arrest while growth resumed 1-2 days after the siRNA knockdown was relaxed (Matsukura, Jones et al. 2003). It is also possible that the tumor cells have developed other ways to bypass these growth regulatory molecules. For example, the p53 null status of H1299 cells (they are homozygously deleted for *p53*) prevents transient re-expression of the proteins from inducing apoptosis. In fact, a previous study showed that adenovirus-mediated exogenous p16 expression alone did not induce apoptosis in H1299 cells, but only exhibited apoptosis after the addition of exogenous p53 expression (Kataoka, Shimomura et al. 2000).

Finally, it is possible that DNMT1 knockdown led to the expression of proteins (e.g. those involved in the differentiated state) which either made the cells resistant to tumor suppressor function or caused growth arrest, preventing subsequent induction of apoptosis by other re-expressed proteins. All of these mechanisms will require future study. However,

the lack of a dramatic effect on growth of H1299 cells by DNMT1 knockdown indicates to us that the use of agents that block methylation may have to be combined with other approaches before being clinically active. Further studies of single cells and clones after knockdown will be needed to verify that individual cells can undergo DNMT1 knockdown, loss of promoter methylation, and re-expression of genes followed by later promoter re-methylation and gene silencing. In this regard, further investigations using RNAi vectors that can stably suppress the expression of other DNMTs and/or methyl-DNA binding proteins will elucidate how DNA methylation contributes to cancer pathogenesis, and enable us to systematically analyze the DNA methylation machinery as a target for therapeutic intervention of cancer.

Note: The following data are part of a published manuscript (Suzuki, Sunaga et al. 2004), where most of the primary data was generated by Makato Suzuki. The siRNA sequences were designed by Noriaki Sunaga. The original idea for the manuscript came from discussions between Noriaki Sunaga and David Shames. The manuscript was written by David Shames. Noriaki Sunaga and David Shames contributed equally to the data submitted for the revised version of the manuscript.

Tables

Table 4-1. Primer and TaqMan probe sequences for MSP and mRNA expression studies

Gene	Type	Forward primer	Reverse primer
<i>RASSF1A</i>	M	5'-GTGGTTTCGTTTCGGTTCGC-3'	5'-CGATACCCCGCGCA-3'
	M-probe	6FAM-5'-CCGACATAACCCGATTAACCCGTAAGTTCG-3'-TAMRA	
	RT	5'-GCTCGTCTGCCTGGACTGTT-3'	5'-TGGGCATTGTACTCCTTGATCTT-3'
	RT-probe	6FAM-5'-TGTGGAGTGGGAGACACCTGACCTTTCT-3'-TAMRA	
<i>p16^{ink4A}</i>	M	5'-CGCAACCGCCGAACG-3'	5'-TTTTTTCGTTAGTATCGGAGGAAGA-3'
	M-probe	6FAM-5'-CGGATCCGCCCCACCCT-3'-TAMRA	
	RT	5'-TTCGGCTGACTGGCTGGCCA-3'	5'-AGCTCCTCAGCCAGGTCCAC-3'
<i>CDH1</i>	M	5'-AATTTTAGGTTAGAGGGTTATCGCGT-3'	5'-TCCCAAACGAACTAACGAC-3'
	M-probe	6FAM-5'-CGCCACCCGACCTCGCAT-3'-TAMRA	
	RT	5'-TTTCTTGGTCTACGCCTGGGACTC-3'	5'-CACCTTCAGCCATCCTGTTTCTC-3'
<i>HPP1</i>	M	5'-GTTATCGTCGTCGTTTTTGTGTC-3'	5'-GACTTCCGAAAACACAAAATCG-3'
	M-probe	6FAM-5'-CCGAACAACGAACTACTAAATCCCAGC-3'-TAMRA	
	RT	5'-TGCTTCCCTACCTCCTTAAGTGA-3'	5'-CTGTCATCATAACCAGCAATTCC-3'
	RT-probe	6FAM-5'-TGCCAACGCCACCGGC-3'-TAMRA	
<i>SEMA3B</i>	RT	5'-CTGGCTCAATGAGCCCAAGT-3'	5'-CTACCGCCGTCTCACGAAAG-3'
	RT-probe	6FAM-5'-AGGTATTTGGATCCCGGAGAGCGAGAATA-3'-TAMRA	
<i>MYOD1</i>	M	5'-CCAACTCCAATCCCCTCTCTAT-3'	5'-TGATTAATTTAGATTGGGTTTAGAGAAGGA-3'
	M-probe	6FAM-5'-TCCCTTCTATTCTAAATCCAACCTAAATACCTCC-3'-TAMRA	
<i>GAPDH</i>	RT	5'-GACCACAGTCCATGCCATCACT-3'	5'-GCTTCACCACCTTCTTGATGTCA-3'
<i>TBP</i>	RT	5'-TGCTGCGGTAATCATGAGGAT-3'	5'-TGGAACCAACTTCTGTACAAC-3'
	RT-probe	6FAM-5'-AGAGAGCCACGAACCACGGCACTG-3'-TAMRA	

Table 4-2. Gene expression ratios in NCI-H1299 and HCC1954 in response to treatment

Gene	NCI-H1299			HCC1954		
	Treatment					
	Nontreated ^a	DNMT1siRNA ^b	5-Aza-CdR	Nontreated	DNMT1siRNA	5-Aza-CdR
<i>RASSF1A</i>	1 (3.3)	7.0 (23)	8.2 (27)	1 (4.2)	3.1 (13)	6.4 (27)
<i>HPP1</i>	—	—	—	1 (26)	1.8 (46)	1.7 (44)
<i>SEMA3B</i>	1 (1.1)	6.5 (7.1)	5.6 (6.2)	—	—	—

^a Using quantitative RT-PCR (see "Materials and Methods") the actual levels of gene expression were determined relative to that for TBP mRNA (TSG:TBP) x 100. A ratio for gene:TBP expression was calculated and is given in parentheses for each gene and nontreated sample. The other ratios are then given relative to an untreated ratio of 1.

^b siRNA, small interfering RNA; 5-Aza-CdR, 5-Aza-2'-deoxycytidine; RT-PCR, reverse transcription-PCR; TBP, TATA box binding protein; TSG, tumor suppressor gene.

Figures

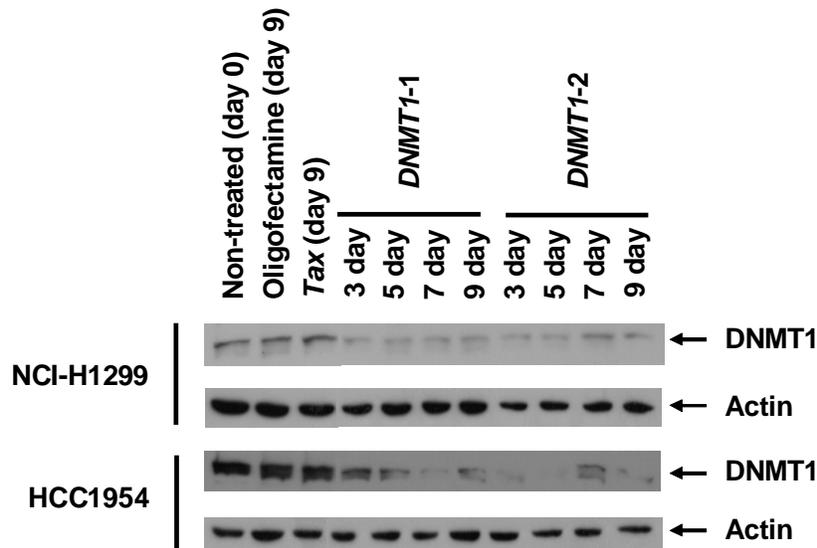


Fig. 4-1. RNAi-mediated knockdown of DNMT1 protein expression in NCI-H1299 NSCLC and HCC1954 breast cancer cell lines. NCI-H1299 and HCC1954 cells were untreated, or treated with Oligofectamine alone, *Tax* siRNA, or two different sequences of siRNA targeted to *DNMT1* (*DNMT1-1*, *DNMT1-2*) four times (on days 0, 2, 4, and 6). Western blots were performed on lysates from untreated cells at day 0, and oligofectamine treated cells and *Tax* siRNA treated cells at day 9, and *DNMT1* siRNA- treated cells at 3, 5, 7, and 9 day. Twenty μg of total protein were loaded per lane.

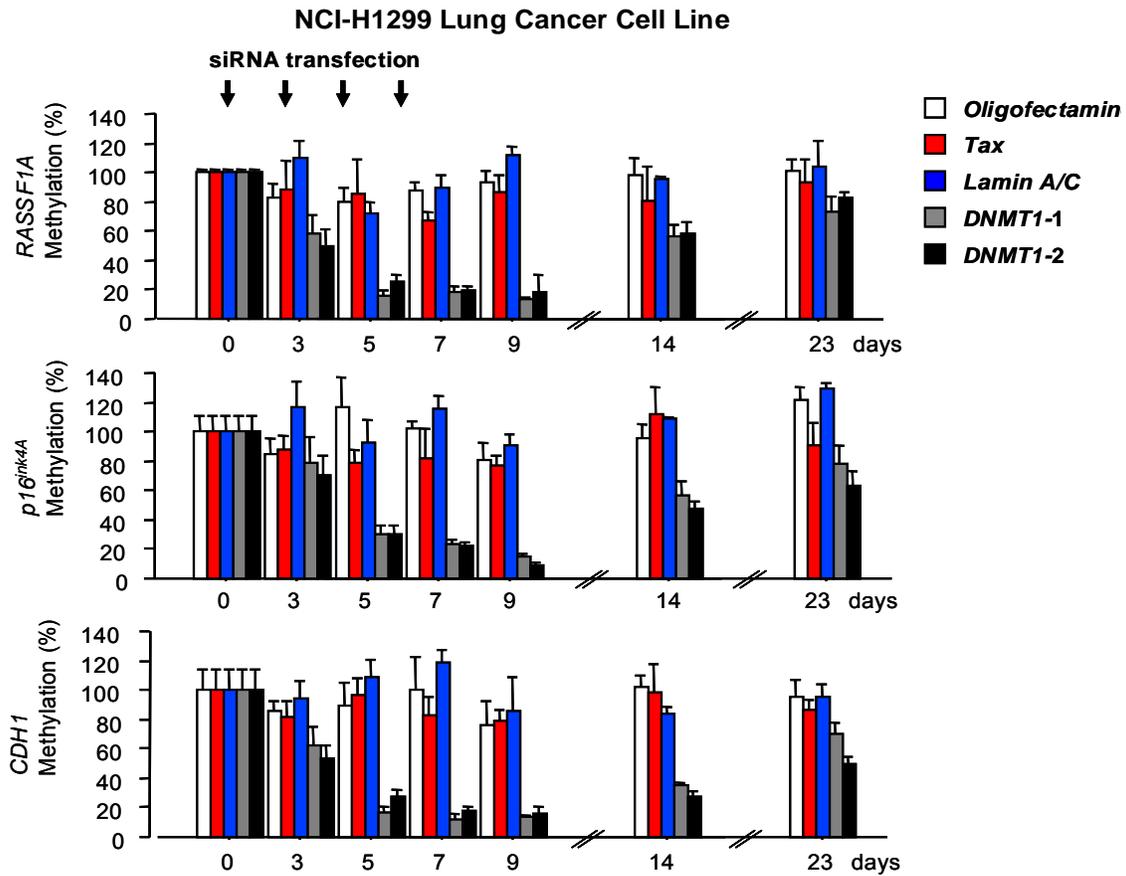


Fig. 4-2A. Time course and kinetics of promoter hypermethylation changes of *RASSF1A*, *p16^{ink4A}*, and *CDH1* genes in NCI-H1299, treated with siRNAs targeted to *DNMT1*, as determined real-time methylation-specific PCR assay. Cells were treated with Oligofectamine alone (blank), *Tax* (red), *Lamin A/C* (blue), *DNMT1-1* (gray), or *DNMT1-2* (black) siRNA four times (on days 0, 2, 4, and 6). Cells were harvested at 0, 3, 5, 7, 9, 14, and 23 days and DNA was extracted and treated with sodium bisulfite. Real-time MSP was performed as described (Materials and Methods). Each ratio was normalized to *MYOD1* and converted to a percentage based on the same ratio in untreated cells. Each point represents averages from three independent experiments \pm S.E.

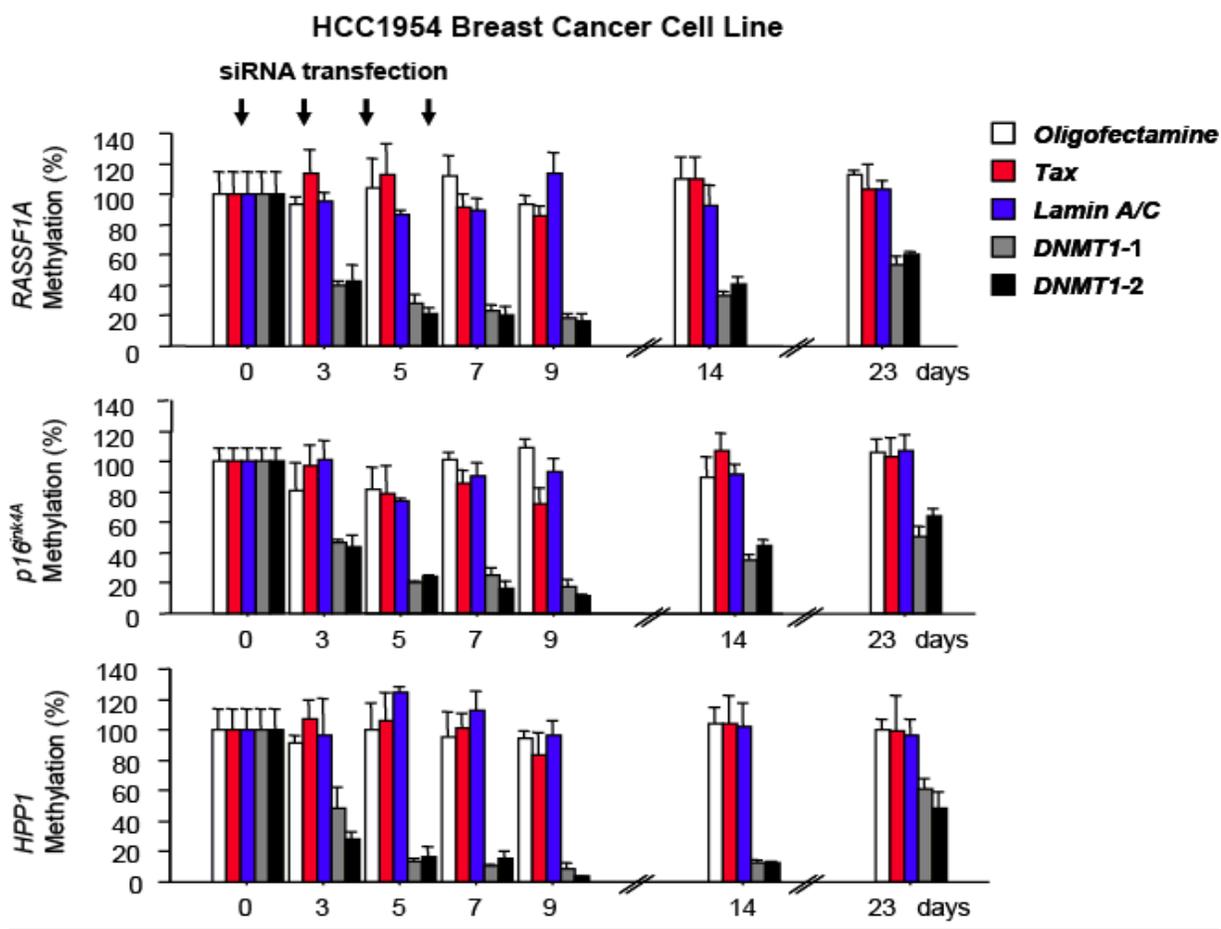


Fig. 4-2B. Time course and kinetics of promoter hypermethylation changes of *RASSF1A*, *p16^{ink4A}*, and *HPP1* genes in HCC1954 treated with siRNAs targeted to *DNMT1*, as determined by real-time Methylation-Specific PCR assay. Cells were treated with Oligofectamine alone (blank), *Tax* (red), *Lamin A/C* (blue), *DNMT1-1* (gray), or *DNMT1-2* (black) siRNA four times (on days 0, 2, 4, and 6). Cells were harvested at 0, 3, 5, 7, 9, 14, and 23 days and DNA was extracted and treated with sodium bisulfite. Real-time MSP was performed as described (Materials and Methods). Each ratio was normalized to *MYOD1* and converted to a percentage based on the same ratio in untreated cells. Each point represents averages from three independent experiments + S.E.

NCI-H1299 Lung Cancer Cell Line

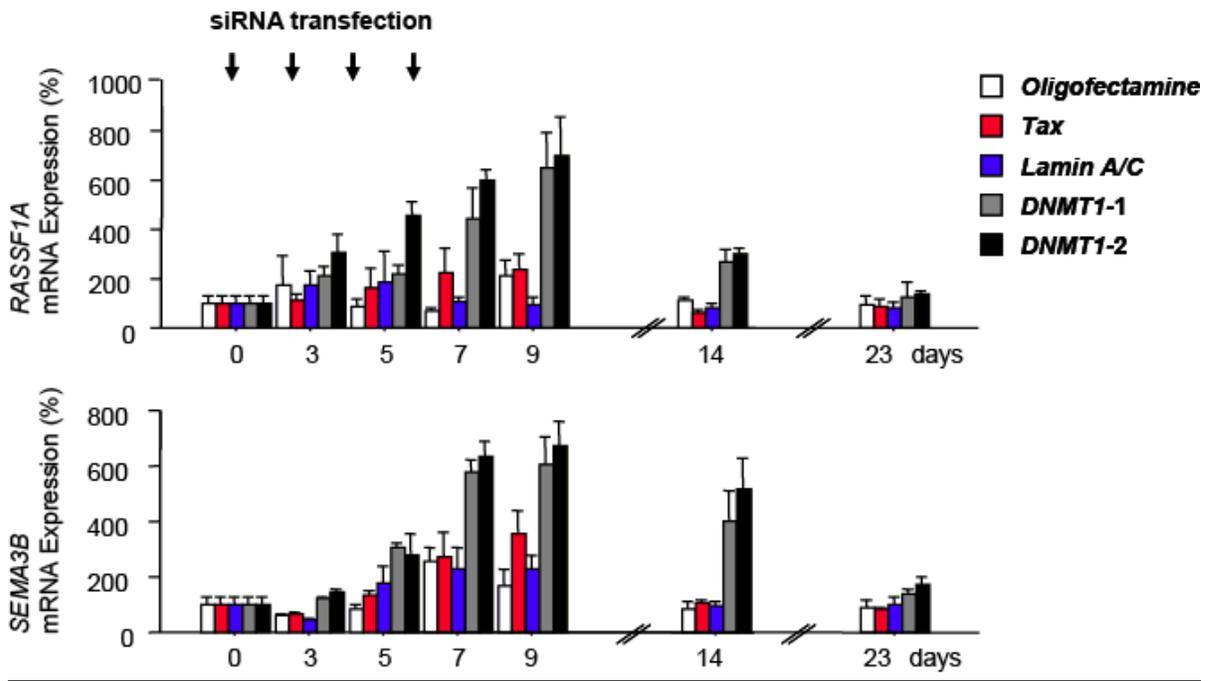


Fig. 4-3A. Time course of mRNA expression level of *RASSF1A* and *SEMA3B* genes by real-time RT-PCR in H1299. Cells were harvested, RNA was extracted, and cDNA synthesized. Real-time RT-PCR was performed as described (Materials and Methods). Each ratio was normalized to *TBP* and converted to percentage based on the same ratio in non-treated cells. Each point represents averages from three independent experiments + S.E.

HCC1954 Breast Cancer Cell Line

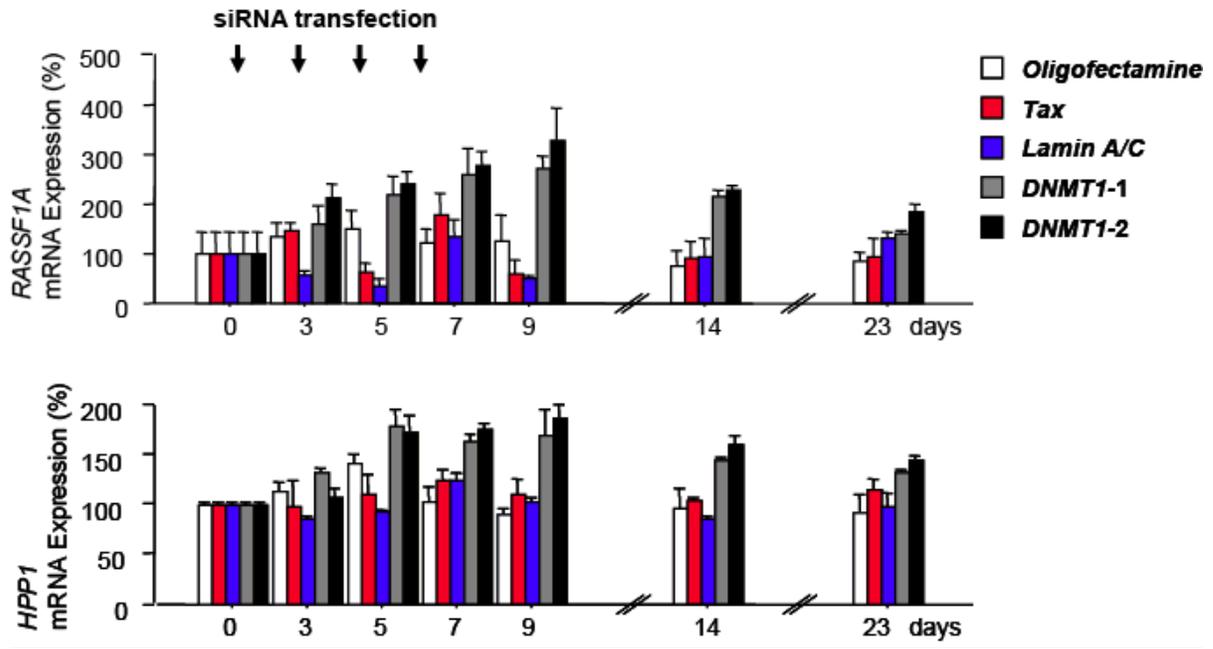


Fig. 4-3B. Time course and kinetics of mRNA induction levels for *RASSF1A* and *HPP1* in HCC1954 by real-time RT-PCR. Cells were harvested, RNA was extracted, and cDNA synthesized. Real-time RT-PCR was performed as described (Materials and Methods). Each ratio was normalized to *TBP* and converted to percentage based on the same ratio in non-treated cells. Each point represents averages from three independent experiments + S.E.

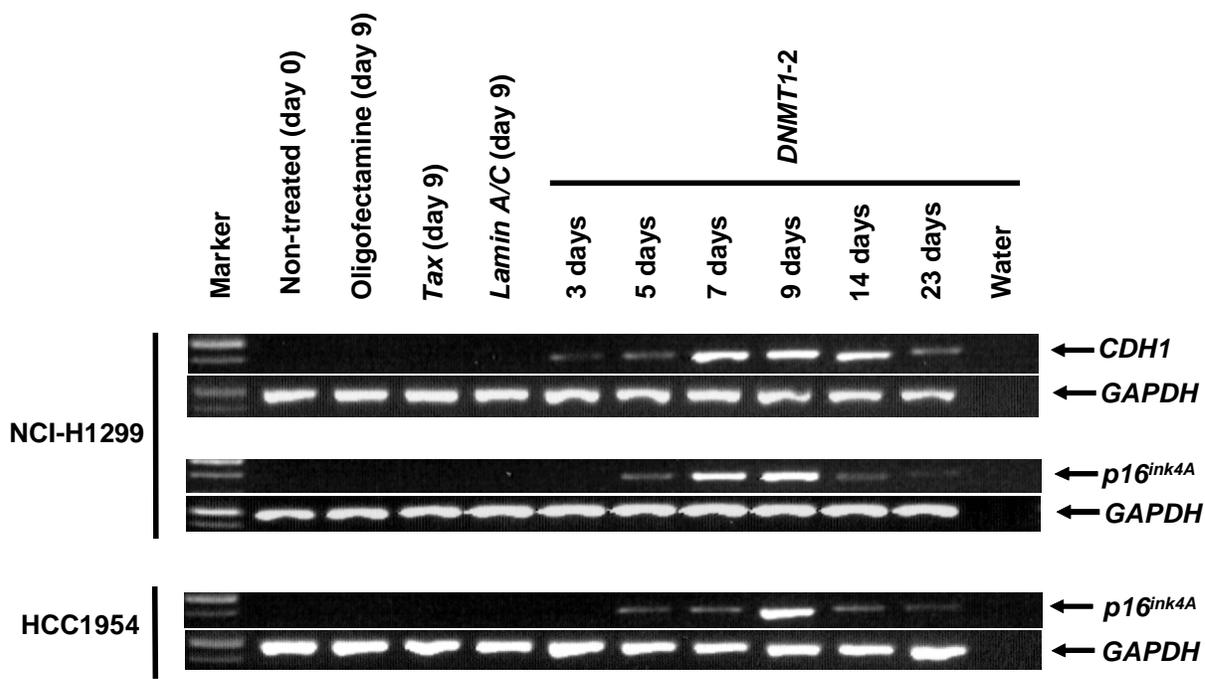


Fig. 4-3C. Time course and kinetics of mRNA induction levels for *CDH1* and *p16* in HCC1954 and H1299 RT-PCR. Cells were harvested, RNA was extracted, and cDNA synthesized. The PCR products were separated on 2% agarose gel. *GAPDH* was run as a control for RNA integrity.

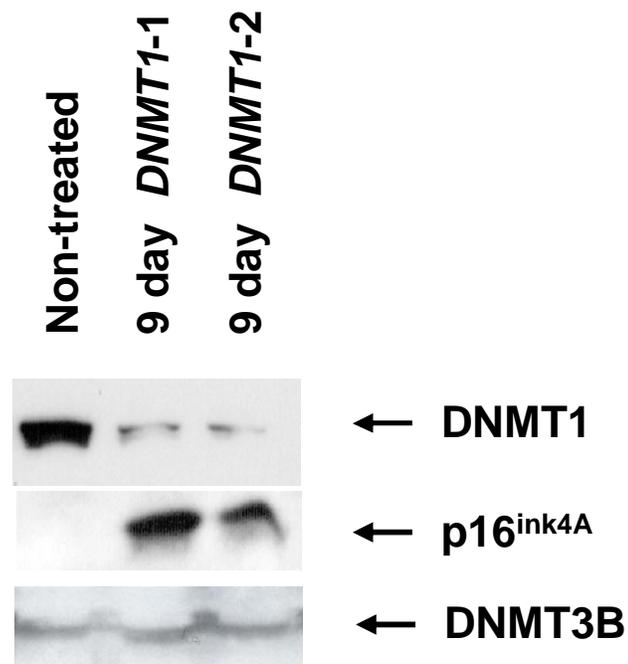


Fig. 4-3D. Western blot showing induction of p16^{ink4A} protein expression in H1299 cells after DNMT1 siRNA. Western blot for DNMT3B shows no change in this methyltransferase in the presence of DNMT1 siRNAs.

CHAPTER FIVE

COMPARISON OF RNAI AND 5-AZA

INTRODUCTION

The studies described above showed that it would be possible, at least in principle, to use RNAi approaches to demethylate DNA and screen for genes subject to promoter hypermethylation in lung cancer. However, as noted in the discussion section above, there were two findings in particular that were unexpected. The first was that DNMT1 knockdown alone was sufficient for gene re-expression, and the second was that while multiple TSGs were induced in the different cell lines, there was little evident toxicity. Moreover, the above study did not address the question of whether RNAi approaches would be sufficient to reverse DNA methylation to the extent necessary for detection by gene expression microarray. Finally, the experimental strategy to identify tumor-specific promoter hypermethylation entailed the use of immortalized human bronchial epithelial cells, which had not been tested for their responses to DNA demethylation, siRNA oligos, or 5-aza treatment. To resolve these issues and to further evaluate whether RNAi approaches were comparable to 5-aza, several experiments were performed including growth assays by colony formation in liquid culture, mass culture growth assays (MTT/MTS), quantitative RT-PCR, western blotting, and microarray gene expression profiling.

DIRECT COMPARISON OF RNAI AND 5-AZA IN H157 CELLS

Based on the study described above, it was clear that siRNA targeting DNMT1 was sufficient to induce tumor suppressor genes silenced by promoter hypermethylation in both H1299 and HCC1954. However, three issues remained to be resolved before we could make a decision on whether to use RNAi or 5-aza. Previous studies had shown that microarray technology was relatively insensitive compared to PCR based detection techniques, and it was not yet apparent whether RNAi approaches resulted in gene induction levels that were higher than that of 5-aza treatment. In addition, previous studies had also shown that 5-aza treatment is quite non-specific with respect to altering gene expression. However, it was unknown what the effects of RNAi based approaches might be on general changes in gene expression. We were particularly concerned about the possibility of inducing the interferon response. Although several studies have claimed otherwise, in our hands siRNAs sometimes induce cytotoxicity independent of target sequence (see *Tax* siRNA treatments in Fig. 5-2, below), which appears to vary between cell lines (Elbashir, Harborth et al. 2001; Elbashir, Lendeckel et al. 2001; Harborth, Elbashir et al. 2001; Bridge, Pebernard et al. 2003; Moss and Taylor 2003; Sledz, Holko et al. 2003; Dunn, Sheehan et al. 2005). The third issue was whether the HBEC cell lines would tolerate sequential siRNA treatments, or whether they would survive genome demethylation.

To evaluate whether RNAi approaches to reducing DNA methylation resulted in gene induction levels comparable to 5-aza in cancer cell lines, gene induction levels were evaluated for the tumor suppressor genes *MASPIN* and *RASSF1A* in H157 cells. In these experiments, cells were harvested after the second and third transfections. For comparison,

H157 cells were treated with 5-aza and TSA. These data show that transfection of H157 cells with siRNA targeted to DNMT1 alone after either 2 or 3 transfections induces both *RASSF1A* and *MASPIN*, however, to lower levels than with 1 μ M 5-aza. There was an increase in gene induction levels between transfections 2 and 3 for both *RASSF1A* and *MASPIN* transcripts, suggesting a specific effect of DNA demethylation. Trichostatin A (TSA), a potent histone deacetylase inhibitor that synergizes with 5-aza, induced *MASPIN* expression, but not *RASSF1A* expression; however, when TSA was combined with siRNA targeted to DNMT1, both *RASSF1A* and *MASPIN* were induced. These data suggest that siRNA targeted to DNMT1 leads to the induction of TSGs silenced by promoter hypermethylation in H157, and that gene induction levels could be increased by addition of the histone deacetylase inhibitor, TSA.

To determine whether 5-aza and siRNA targeted to DNMT1 and DNMT3B had similar effects on gene expression in a different cell line, and to determine whether siRNA treatment resulted in cytotoxicity in either cell line, we compared gene induction levels and cytotoxicity in H157 with H1299 cells (Fig. 5-2). These data confirm that 5-aza and DNMT1 inhibition led to gene induction in both H1299 and H157. Moreover, loss of DNMT1 activity is less toxic to NSCLC cells than either 5-aza treatment or DNMT3B knockdown, even though DNMT3B knockdown did not appear to affect *RASSF1A* expression.

Based on the experiments described above and the findings of Rhee et al. we decided to examine how 5-aza compared with siRNA targeted to DNMT1 and DNMT3B in terms of global gene expression changes using the Affy Plus 2.0 chips. To confirm that transfection of siRNA targeted to DNMT1 and DNMT3B led to reduction of target proteins, western

blots were performed with the appropriate antibodies (Fig. 5-3). To ensure that the interferon response was not activated during transfection, we assayed for an increase phosphor-PKR levels. Activated phosphor-PKR was not detected in these cells (data not shown).

Microarray analysis of H157 after 5-aza treatment or siRNA targeted to either DNMT1 or DNMT3B exhibited a varied gene expression phenotype. The overall gene expression profiles for these samples varied greatly as indicated by the scatter plots below. Control samples were well correlated with R^2 values ranging from 0.83 – 0.89 (Fig. 5-4 A-C). As expected, substantial alterations in gene expression were observed for the 1 μ M 5-aza treatment (Fig. 5-4D). Interestingly, siRNA targeted to DNMT3B also affected gene expression substantially, but in most cases, the gene expression levels appeared to be reduced compared to Oligofectamine and *Tax* controls. 5-aza treatment was more similar to DNMT1 knockdown than DNMT3B knockdown, while overall the two siRNA treatments were more similar to each other than either was to 5-aza.

To further explore the similarities and differences in gene expression found in these microarray experiments, sample clustering was performed using unsupervised average linkage hierarchical cluster analysis of all seven samples (Fig. 5-5). This analysis shows that while the control samples (DMSO, Oligo, and *Tax*) group together, both 5-aza treated cells and the DNMT1 and DNMT3B transfectants form their own trees apart from each other and the control samples. The simplest interpretation of these data is that siRNA treatments targeting the DNMT enzymes in H157 cells results in a different gene expression profile than 5-aza treatment.

Taken together the data highlighted above suggest that siRNA targeted to DNMT1 and DNMT3B result in a markedly different gene expression phenotypes in H157 cells when compared to 5-aza treatment. The differences observed in these experiments may be explained in part by the different mechanisms of action of RNAi vs. small molecule inhibitors. RNAi approaches target mRNA and thus reduce the levels of proteins, whereas 5-aza acts as a suicide substrate for the DNMTs. The DNMT enzymes form large macromolecular complexes in the nucleus and may be involved in many of aspects of chromatin structure regulation independent of DNA methylation. Thus, depleting these proteins in cells could destabilize other regulatory complexes indirectly, leading to differential gene expression compared to 5-aza. Alternatively, the mechanism of action of 5-aza involves its incorporation into the nascent strand of DNA during S-phase. Since DNMT1 localizes to the replication fork during S-phase, it is possible that covalent linkage of DNMT1 to 5-aza could block DNA synthesis. Another possibility is that covalent linkage of DNMT1 to DNA leads to large-scale double-strand breaks independent of cell division.

While the above experiments do not directly address the question of specificity with respect to differential gene induction between RNAi vs. 5-aza treatment, it is noteworthy that DNMT1 knockdown induced well described TSGs, but did not result in significant cytotoxicity. The reasons for this finding are probably complex, but one explanation may be that once cells undergo transfection they become epigenetically heterogeneous. This means that some cells may express p16, others may express RASSF1A, and still others express both; expression of either of these genes would more likely lead to growth arrest as opposed to cell death. On the other hand, treatment with a small molecule such as 5-aza, which is a

DNA analogue that is incorporated into DNA during replication, probably has more acute toxicity to cells that is independent of its inhibitory activity, which could lead to cell death as opposed to growth arrest. Another possibility is that 5-aza may inhibit all of the methyltransferases, whereas siRNA presumably targets one at a time. The difference here is that cancer cells may be able to tolerate the limited demethylation that results from RNAi, whereas they cannot tolerate the extant on DNA demethylation that results from 5-aza treatment (see discussion *Chapter 4*).

The purpose of the above experiments was to determine whether RNAi approaches might be better than 5-aza treatment to use as part of a demethylation screen. The data above suggests that RNAi approaches may be more specific, but 5-aza induces genes to higher levels than RNAi. Moreover, the serial transfections necessary for the RNAi approach are technically challenging and may introduce other, unpredictable effects in different cell lines.

COMPARISON OF 5-AZA AND RNAI IN IMMORTALIZED HUMAN BRONCHIAL EPITHELIAL CELLS

As part of the demethylation screen, we planned to use a series of immortalized human bronchial epithelial cell lines (HBEC) created by our lab. These cell lines were derived from patients undergoing different forms of lung surgery. The clinical history for the patients from whom these cell lines were developed appears in Table 6-1. These cell lines were created using retroviral expression vectors containing *cdk4* and *hTERT* and were maintained according to the protocols described in the special methods section below.

The HBECs are immortal, clonable, can be genetically manipulated, but do not form colonies in soft agar nor do they form tumors in nude mice (Ramirez, Sheridan et al. 2004). In three-dimensional culture they can undergo differentiation into fully ciliated cells (Vaughan, Ramirez et al. 2006). They have very few genetic alterations and they are a novel and important normal tissue control for 5-aza or siRNA targeted to DNMT1 gene induction experiments. However, at the time, it was unknown whether these cells would tolerate DNA demethylation or RNAi treatment.

To determine whether the HBEC cell lines would tolerate siRNA oligos, HBEC2, 3, and 4 were treated with 100nM, 50nM, 10nM, and 1nM oligos targeting lentiviral gene *Tax*, green fluorescent protein (GFP), Lamin A/C, DNMT1, and DNMT3B complexed in either oligofectamine or lipofectamine reagents. All doses were well-tolerated 72 hrs after a single transfection (data not shown). Pilot experiments suggested that to achieve promoter demethylation in H1299, H157, and HCC1954, three to four sequential transfections were required (Figs. 4-1 and 5-3). This approach requires initially plating the cells in 24-well

dishes and then transfecting every other day and re-plating the cells after the second transfection in 6 well or 60mm plates, depending on the cell's doubling time. It was found that after two transfections, at 100 nM, all oligos induced significant toxicity in the HBEC cell lines. Only the 50nM dose was used for later experiments.

To determine whether the HBEC2 and 3 would tolerate 5-aza treatment, cells were treated with 100 nM and 1 μ M 5-aza every other day for 6 days and then counted using a trypan exclusion assay (Fig. 5-6). These data suggest that HBEC2 and HBEC3 differ in their tolerance to siRNA treatment, and neither cell line tolerated siRNAs targeted to DNMT3B (Fig. 5-6 and 5-7). This finding suggests that loss of DNMT3B expression may necessary for HBEC survival. 1 μ M 5-aza treatment resulted in a different phenotype to both DNMT1 and DNMT3B knockdown in HBEC3.

There were significant morphological changes in HBEC3 cells in response to 5-aza. Normally the HBECs exhibit highly regular morphology and contain single, elliptical, moderately-sized nuclei (Fig. 5-7 A&D) (Ramirez, Sheridan et al. 2004; Sato, Vaughan et al. 2006). Upon 5-aza treatment, the cells became highly irregular, with some cells becoming extremely large, vacuolated, and multinucleated. Some had distended, pleomorphic nuclei with little cytoplasm (Fig. 5-7F). Invariably however, there was significant growth inhibition as opposed to cell death as determined by trypan exclusion assay in both HBEC2 and HBEC3 (Fig. 5-6). After treatment with 5-aza (both 100 nM and 1mM) we re-plated the HBEC cells at several densities to determine whether the cells would recover from the treatment. After two weeks, none of the treatment groups attached to the substrate or formed colonies in liquid culture (data not shown).

SPECIAL METHODS FOR HBECS

Infection Protocol (as per Shelley Sheridan)

- 1) When the cells are about 50% confluent, they are ready to be infected.
- 2) Remove 1 vial (stored in -80) of *cdk4* or E6/E7 infected media for each flask/dish and thaw at 37C. (1 vial will infect 2x T25 flasks, 1x T75 flask, or 1x 10cm dish).
- 3) Place 4 ml of KSFM and 4ug/mL hexadimethoine bromide in a 15ml tube.
- 4) Add the thawed vial of *cdk4* or E6/E7 infected media to the previous mixture to a total volume of 6 ml.
- 5) Immediately add mixture to your flask of cells and incubate for 16 hours. (Do not incubate for more than 18 hours.)
6. Place fresh KSFM to the flask and allow to recover for 24 hours.
Repeat steps 2 - 6.
- 7) When the *cdk4* or E6/E7 infected cells look healthy its time to select the infected cells: add 30 µg/ml G418 to your growth media and grow for 10 days.
- 8) Replace with fresh drug every 2-3 days.
- 9) When the selection is finished, feed with fresh media to allow recovery.
- 10) When the flask is confluent subculture as usual. Make sure you plate an extra flask of the *cdk4* infected cells to add *hTERT*.
- 12) When the *cdk4* infected flask is about 50% confluent infect with *hTERT*.
- 13) Go through steps 2 - 7 using the *hTERT* infected media. When the *hTERT* infected cells look healthy its time to select the infected cells;

- 14) Add 250 ng/ml Puromycin to your growth media and grow for 3 days. When the selection is finished feed with fresh media and allow to recover.

When the flask is confluent subculture as usual

Continue to subculture to PD 100 to confirm immortalization is complete.

CONCLUSIONS

The goal of the above-described pilot studies was to determine the appropriate parameters to properly compare gene induction after demethylation in tumor and normal cells. To examine this question we study the phenotypic and gene expression changes in NSCLC cell lines and HBECs to treatment with 5-aza and siRNA targeted to DNMT1 and DNMT3B in HBECs. These experiments allowed us to establish conditions for the main part of this thesis (see below). In addition, we learned that normal HBECs undergo changes in expression of genes that are probably methylated as part of a differentiation program specific to lung epithelium. We found that DNMT1 reversed promoter hypermethylation at important TSGs in both H1299 and H157, but to lower levels than 1 μ M 5-aza treatment. In addition, we found that DNMT3B knockdown is toxic to both NSCLC and HBEC cell lines.

As a result of these findings, and the technical difficulties involved with performing three sequential transfections in the HBECs, we chose to pursue the 5-aza approach for comparing gene expression changes between the HBECs and NSCLC cell lines.

Figures

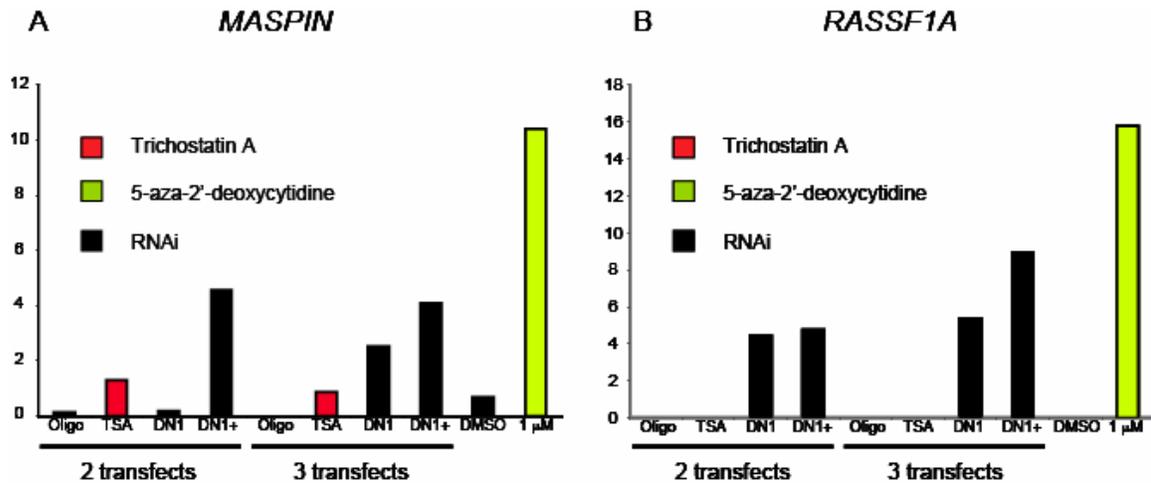


Fig. 5-1. Comparison of gene induction levels for *MASPIN* and *RASSF1A* in H157 by QPCR. H157 squamous cell carcinoma cell lines were treated with Oligofectamine alone, the general histone deacetylase inhibitor Trichostatin A (TSA), siRNA targeted to *DNMT1* (DN1), siRNA to DNMT1 plus TSA (DN1+), DMSO, or 5-aza (1 μ M). Cells were harvested and RNA extracted after the second and third transfections for the siRNA and TSA treatments. For TSA alone treatments, cells were treated three times, every other day for 6 days; siRNA + TSA, cells were treated with TSA of the off day between transfections, either two or three times depending on the number of transfections. 5-aza treatment was according to standard treatment protocol. Black bars indicate samples from transfections; maroon bars indicate samples from TSA treatment; and green bars indicate 5-aza treatment. Values on the y-axis indicate log₂ values according to the $2^{-\Delta\Delta C_t}$ method. QPCR was performed using Taqman fluorescent probes purchased from Applied Biosystems. The *MASPIN* primers and probe sequences were from an assay-on-demand, whereas primers and probe for *RASSF1A* from (Suzuki, Sunaga et al. 2004).

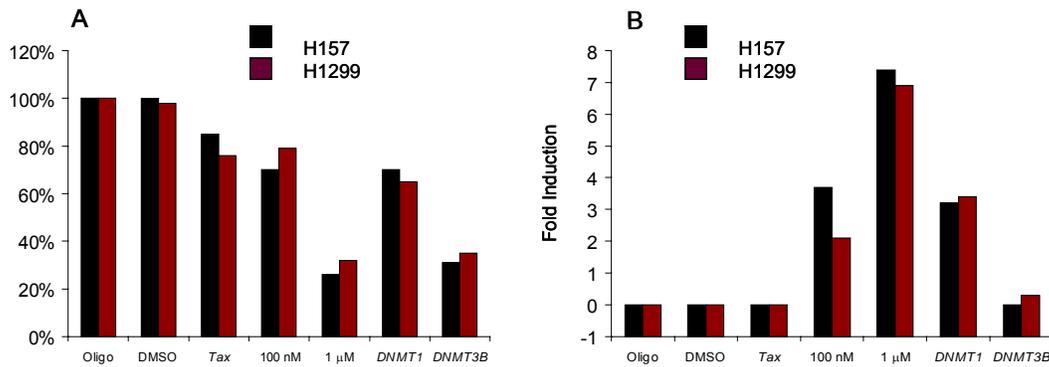


Fig. 5-2. Comparison of cytotoxicity and gene induction levels in 5-aza and siRNA (DNMT1 or DNMT3B) treated H157 cells. Cells were treated as in Fig. 4-4. Cytotoxicity was determined using the trypan exclusion assay on the final day of treatment. Gene induction levels were determined by QPCR. **A)** Percentage of viable cells after indicated treatments. **B)** Gene induction levels based on QPCR for RASSF1A in H157 and H1299. Data were calculated using $2^{\Delta\Delta Ct}$ method. Black bars indicate H157 cells; red bars indicate H1299 cells.

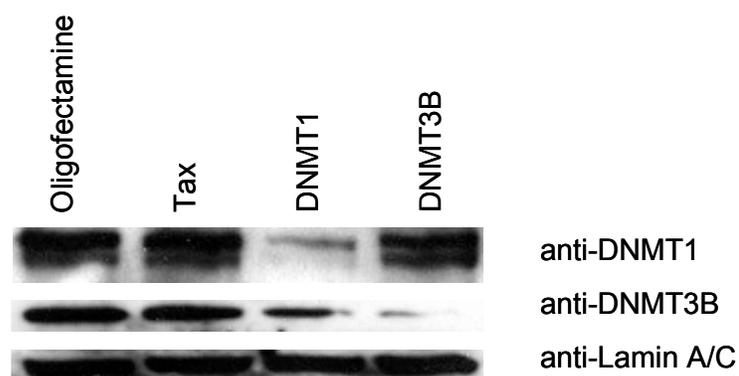


Fig. 5-3. Western blot for DNMT1 and DNMT3B after three transfections with 100 nM siRNA in H157 cells. Crude nuclear extracts were prepared as described in materials and methods. 30 μ g of nuclear lysate was loaded per lane, resolved on 4% acrylamide gel, and transferred to nitrocellulose membranes. Anti-DNMT1 was obtained from Imgenex; DNMT3B rabbit anti-serum was obtained as a kind gift from Robert Macleod; anti-Lamin A/C was obtained from Santa Cruz Biotechnology and was used as a loading control.

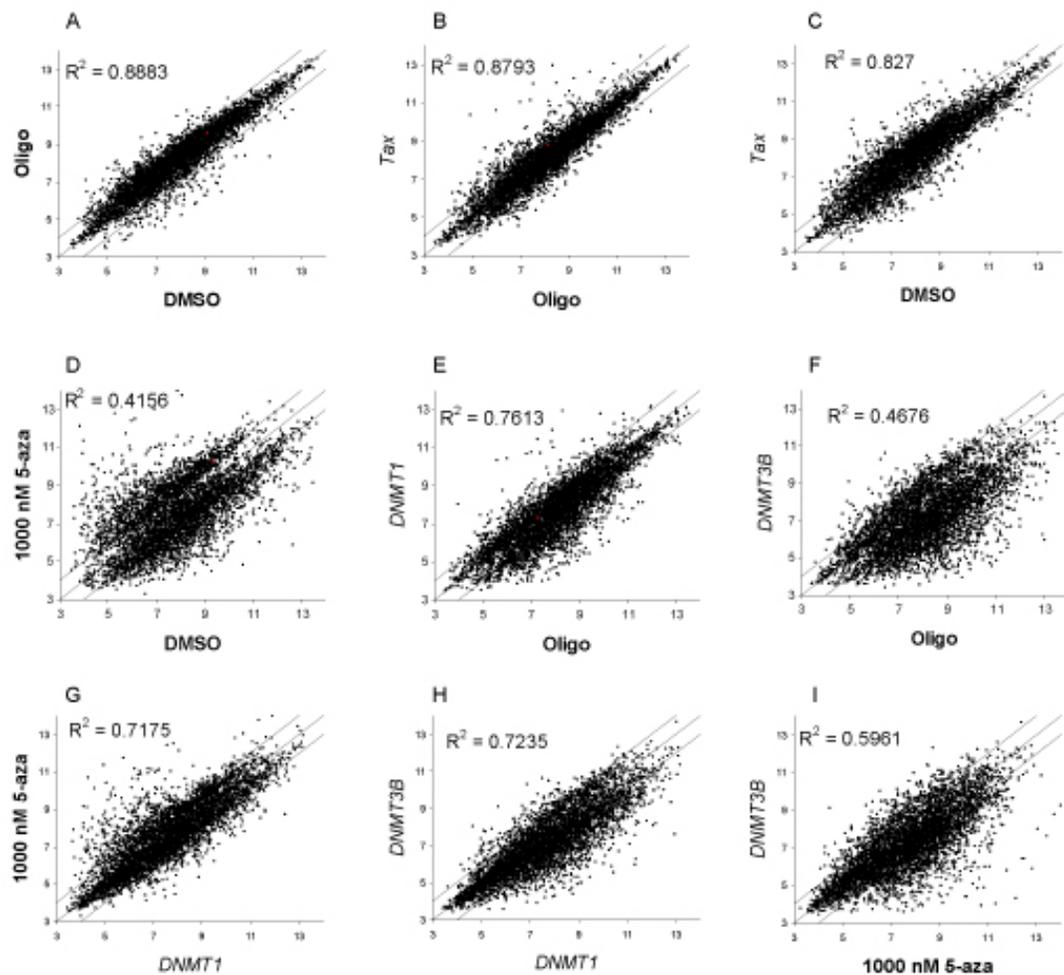


Fig. 5-4 A-I. Scatter plots for gene expression differences in H157 cells by microarray analysis. Data were normalized using RMA (methods) and filtered to exclude genes that were altered <1.5 fold in <20% of samples, which left 5187 genes. Integers on each axis indicate log₂ of the absolute intensity for each probe. Each point represents a single probe, and lines bisecting the data points indicate 2-fold changes either up (above) or down (below). R^2 values were calculated using R and Excel.

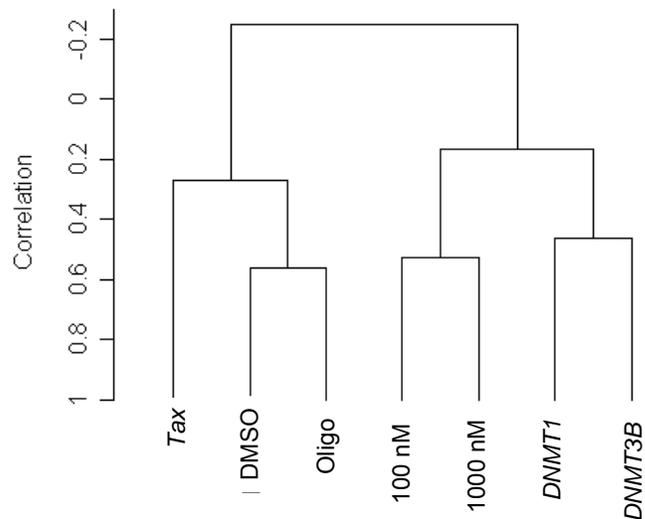


Fig. 5-5. Hierarchical cluster analysis of Affymetrix U133 Plus 2.0 mRNA expression profiles in H157 cells after siRNA or 5-aza treatment. Data were normalized and filtered as in Fig. 5-4, and clustered using the Pearson correlation coefficient and average linkage clustering implemented through BRB array tools. These data indicate that while control treated (DMSO, *Tax*, Oligofectamine) cell lines on the whole have similar gene expression profiles, 5-aza treatment (100 nM, 1 μ M) is different from siRNA targeted to either of the DNMT proteins.

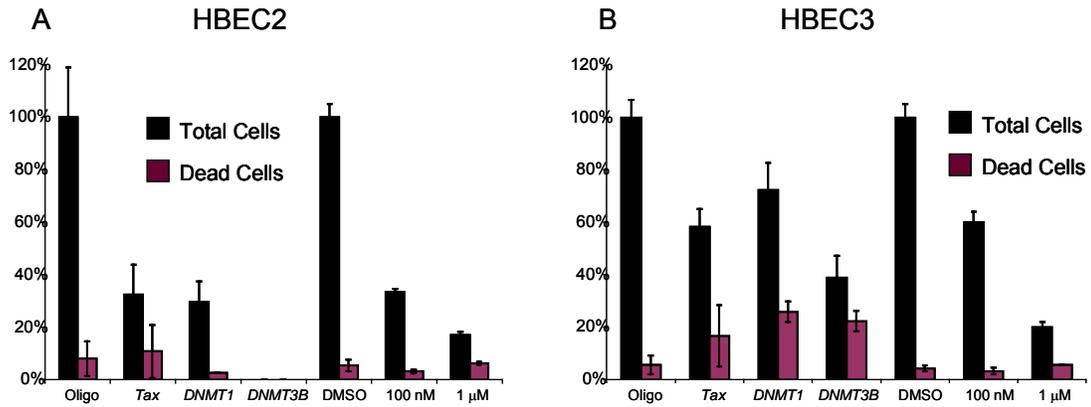


Fig. 5-6. Trypan exclusion assay for HBEC2 and HBEC3 after treatment with either siRNA or 5-aza-2'-deoxycytidine. Cells were treated with 50 nM siRNA targeted to *Tax*, *DNMT1*, or *DNMT3B* every other day for 8 days. Cells were re-plated between the second and third transfections into to 6-well plates and counted on day 10. For 5-aza treatments, cells were treated every other day with the indicated doses and harvested and counted on day 6. Black bars indicate total cells; red indicates dead cells. In general, HBEC3 tolerated both 5-aza and siRNA better than HBEC2.

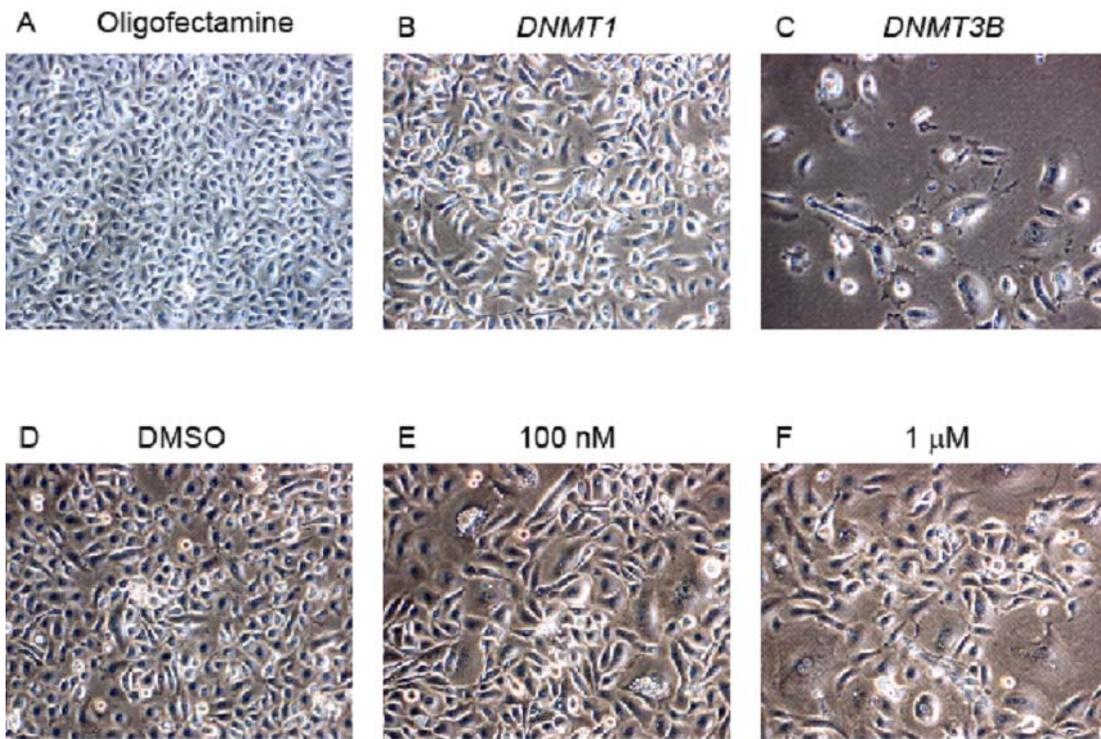


Fig. 5-7 A-F. Phase contrast images of HBEC3 indicating morphological changes associated with various treatments. Images were collected on day 10, prior to harvesting after three sequential siRNA transfections for panels A-C, or three sequential treatments of 5-aza for panels D-F. Magnification is 200x.

CHAPTER SIX

INTRODUCTION

In the previous chapter, it was determined that for our purposes, the appropriate method to use for demethylating DNA in the HBECs and cancer cell lines was 5-aza. This decision was based on two major criteria: first, 5-aza treatment is simpler to perform than serial transfections, and second, 5-aza induced genes to a higher level than siRNA treatment. The chapter below describes experiments that form the major part of this thesis as well as the publication included herein. Besides the material that was included in the publication, which focused primarily on the clinical aspects of the promoter hypermethylation profiles that were identified, I have added a section that addresses the cell line selection process and as well as some bioinformatic analysis of the gene expression profiles of the NSCLC cell lines.

Part of the data included in the publication are derived from general Minna lab resources such as the lung and breast cancer cell lines, the microarray expression data, and array CGH (aCGH) data (Phelps, Johnson et al. 1996). Primary tumor materials were acquired from a variety of sources: breast tumor DNAs were obtained as part of a collaboration with Chuck Perou from the University of North Carolina at Chapel Hill and Olufunmilayo Olopade from the University of Chicago, as well as Cheryl Lewis and David Euhus at in the Hamon Center at UT Southwestern. The benign lung and lung tumor DNA was obtained by the Minna lab from Kwun Fong at the University of Queensland in Brisbane. Prostate and colon cancer samples were obtained from the Cancer Center tissue repository at UTSW. Some of the primary tumor microarray expression data was obtained

through collaborations with William Gerald at the Memorial Sloan-Kettering Cancer Center in New York, and David Lam and Maria Wong at Hong Kong University. All of the microarray expression data was generated in the Minna lab and Luc Girard has been responsible for establishing this major resource for the Minna lab. He created several of the software programs, which handle these data, and supervised the collection and processing of samples that were used to produce this resource. Cheryl Lewis assisted in performing some of the methylation specific PCRs for the breast tumors.

A GENOME-WIDE SCREEN FOR HYPERMETHYLATED GENES IN LUNG CANCER IDENTIFIES TUMOR-SPECIFIC METHYLATION MARKERS FOR MULTIPLE MALIGNANCIES

ABSTRACT

Background: Promoter hypermethylation coupled with loss of heterozygosity at the same locus results in loss of gene function in many tumor cells. The “rules” governing which genes are methylated during the pathogenesis of individual cancers, how specific methylation profiles are initially established, or what determines tumor-type specific methylation are unknown. However, DNA methylation markers that are highly specific and sensitive for common tumors would be useful for the early detection of cancer and those required for the malignant phenotype identify pathways important as therapeutic targets.

Methods and Results: In an effort to identify new cancer-specific methylation markers, we employed a high-throughput global expression profiling approach in lung cancer cells. We identified 132 genes that have 5' CpG islands, are induced from undetectable levels by 5-aza-2'-deoxycytidine (5-aza) in multiple non-small cell lung cancer cell lines, and are expressed in immortalized human bronchial epithelial cells. As expected, these genes were also expressed in normal lung, but often not in companion primary lung cancers. Methylation analysis of a subset (45/132) of these promoter regions in primary lung cancer (N=20) and adjacent non-malignant tissue showed that 31 genes had acquired methylation in the tumors, but did not show methylation in normal lung or lymphocytes. We studied the 8 most

frequently and specifically methylated genes from our lung cancer data set in breast cancer (N=37), colon cancer (N=24), and prostate cancer (N=24) along with counterpart non-malignant tissues. We found that 7 loci were frequently methylated in both breast and lung cancers, with 4 showing extensive methylation in all 4 epithelial tumors.

Conclusions: By using a systematic biological screen, we identified multiple genes that are methylated with high penetrance in primary lung, breast, colon, and prostate cancers. The cross-tumor methylation pattern we observed for these novel markers suggests that we have identified a partial promoter hypermethylation signature for these common malignancies. These data suggest that while tumors in different tissues vary substantially with respect to gene expression, there may be commonalities in their promoter methylation profiles that represent generic targets for early detection screening or therapeutic strategies.

INTRODUCTION

Tumor-acquired alterations in DNA methylation include both genome-wide hypomethylation and locus specific hypermethylation. Genomic hypomethylation occurs early in cellular transformation and affects both genome stability and imprinted gene expression (Feinberg and Tycko 2004; Holm, Jackson-Grusby et al. 2005; Simpson, Caballero et al. 2005). Promoter hypermethylation often coincides with loss of heterozygosity at the same locus, which often results in loss of function of the gene in tumor cells. These changes often occur at tumor suppressor gene loci and are hypothesized to participate in cancer development (Jones and Baylin 2002).

Even though genome methylation patterns are deranged in cancer cells, the DNA methyltransferases themselves are rarely if ever mutated or aberrantly expressed (Bestor 2003). The “rules” governing which genes are methylated during the pathogenesis of individual cancers, as well as the timing of their methylation and silencing (e.g. during preneoplasia or in metastatic progression) are unknown, and it is not yet clear how specific methylation profiles are initially established in tumor cells (Baylin and Ohm 2006; Feinberg, Ohlsson et al. 2006). However, it is known that aberrant promoter hypermethylation is common to most tumors, and in many cases, appears to have tumor-type specificity (Costello, Fruhwald et al. 2000). A few genes, such as the cyclin-dependent kinase inhibitor (*p16*) and the tumor suppressor gene ras association domain family protein 1A (*RASSF1A*) are methylated across many tumor types, but it is unknown whether there are more genes that collectively represent a common promoter hypermethylation profile for multiple epithelial cell malignancies (Merlo, Herman et al. 1995; Burbee, Forgacs et al. 2001; Dammann, Yang

et al. 2001; Esteller, Corn et al. 2001). Since CpG island methylation is readily detectable in tissues and fluids, the identification of a promoter hypermethylation gene set that is common to multiple malignancies – with high frequency and specificity for tumors compared to normal tissues – would have important implications for patient screening, diagnosis, and therapeutic intervention (Esteller, Corn et al. 2001; Belinsky 2004).

In an effort to identify genes subject to frequent promoter hypermethylation in human cancers, we used a genome-wide microarray-based approach. In previous studies of this type, cancer cells were treated with the DNA methylation inhibitor, 5-aza-2'-deoxycytidine (5-aza), or vehicle, and compared by gene expression profiling (Suzuki, Gabrielson et al. 2002; Sato, Fukushima et al. 2003). In the present study, an array platform covering 47,000 transcripts over the whole genome (Affymetrix U133 Plus 2.0) was used. In addition, we increased the efficiency and specificity of the initial screen by performing the 5-aza induction experiments in 7 non-small cell lung cancer (NSCLC) cell lines and 3 different immortalized bronchial epithelial cell lines (HBECs) (Ramirez, Sheridan et al. 2004; Sato, Vaughan et al. 2006). Comparison of the gene expression profiles between NSCLC and HBEC before and after 5-aza treatment allowed us to identify genes expressed in normal bronchial epithelial cells but not in cancer cells that were also selectively induced in multiple lung cancers. Next, we applied a series of bioinformatic filters to exclude genes that are poorly annotated or that do not contain 5' CpG islands to arrive at a set of 132 genes that are candidate methylation markers for NSCLC. Finally, we performed similar experiments in colon and breast cancer cell lines, and found that many of the genes induced specifically in NSCLC cell lines were also induced in these cancer models.

As a result of our findings in the 5-aza gene induction screen, we first examined the promoter methylation status of 45/132 genes in a panel of lung cancer cell lines as well as 20 primary lung tumors with counterpart non-malignant tissue and found that 43 were methylated in many tumor cell lines and primary tumors, but not in normal tissues. We selected 8 loci that were frequently and specifically methylated in lung tumor tissue and examined promoter methylation for these loci in a broader panel of breast, colon, and prostate tumors and counterpart benign tissues. These studies revealed an extensive, cross-tissue methylation pattern for 4/8 of these genes. Our data suggest that while tumors in different tissues may vary substantially with respect to gene expression, there may be commonalities in their promoter hypermethylation profiles, perhaps reflecting common events in their pathogenesis, as well as representing generic targets for early detection or therapeutic strategies.

RESULTS

***Selection of cell lines for microarray analysis**

It is well established that tobacco smoke is the major etiological factor in the development of lung cancer; however, NSCLC remains a heterogeneous disease. There are 26 cell types in the lung, and at least 3 distinct types of NSCLC including, squamous cell carcinoma, adenocarcinoma, and large cell carcinoma. Squamous cell carcinomas (SCC), adenocarcinomas (Ad), and large cell carcinomas (LCC) are all strongly associated with smoking, but SCC occurs primarily in the major bronchi, whereas Ad and LCC usually occur deeper in the lung periphery. It is believed that for SCC, Ad, and LCC progenitor cells acquire multiple genetic and epigenetic changes that coincide with histological progression from normal epithelium through dysplasia to invasive cancer (Fig. 6-1). The molecular pathology and histology of the three different types of NSCLC are different and substantial effort has been directed toward identifying the particular molecular lesions that are associated each type of NSCLC (Sekido, Fong et al. 2003; Wistuba and Gazdar 2006).

As indicated in the figure below, certain lesions have been associated with particular steps in the oncogenic process. Both hypo- and hypermethylation occur early along with microsatellite alterations (particularly on 3p) and the first hints of morphological changes in the cells. Recently, the epidermal growth factor receptor (*EGFR*) was shown to be a frequent target of mutations affecting the kinase domain. This type of mutation is common in female non-smokers of Asian decent (Shigematsu, Lin et al. 2005; Tang, Shigematsu et al. 2005; Shigematsu and Gazdar 2006).

In certain cases, it has been possible to identify gene expression signatures that correspond to distinctive histopathological types of some cancers, notably ductal carcinoma of the breast. There is some evidence that NSCLC can be distinguished along these lines as well (Garber, Troyanskaya et al. 2001). We used hierarchical cluster analysis to explore this finding in our panel of NSCLC cell lines. Our first question was whether the cell lines grouped according to the histopathology of the parental tumor. As indicated in Fig. 6-2, the cell lines do not group according to their histopathology. However, unsupervised cluster analysis of the expression data from the 31 cancer cell lines using 2377 genes that were differentially expressed across the panel by SAM analysis indicates the presence of at least five and possibly six distinct groups of cell lines. Bearing in mind the heterogeneity of NSCLC, we selected one cell line from each part of the dendrogram in our initial experiments (blue arrows). The purpose of this selection process was to identify NSCLC cell lines for 5-aza treatment based on the diverse biology of this disease.*

Standardizing 5-aza-2'-deoxycytidine treatment for HBECs and cancer cell lines

To analyze the gene expression changes associated with loss of promoter methylation in lung cancer cells compared to HBECs, we treated seven non-small cell lung cancer cell lines (NCI-H460, H1299, H157, H2347, H1819, H1993, and A549) and three HBEC lines (HBEC2, 3, and 4) with low (100 nM) and high (1 μ M) doses of 5-aza (Table 6-1). To determine whether low and high dose 5-aza induced genes silenced by promoter methylation in NSCLC cell lines, we performed quantitative reverse transcriptase-PCR (QPCR) for *p16*. We also ran standard RT-PCR for *p16* in several cell lines to ensure that the QPCR primer set did not amplify the alternate splice-form *p14*, which is expressed in some of these cell lines (Sato, Horio et al. 2002). We observed induction of *p16* mRNA for both low and high dose 5-aza in tumor lines that harbor *p16* promoter methylation (Fig. 6-3A & B). Since *p16* could not be used as a positive control for NSCLC lines with homozygously deleted or unmethylated *p16*, we used the universally methylated gene transketolase-like 1 (*TKTL1*) as a positive control for loss of DNA methylation and gene induction. *TKTL1* was induced by 5-aza in all cell lines examined (Fig. 6-3A & C).

Microarray analysis of gene expression changes after 5-aza-2'-deoxycytidine treatment in lung cancer cell lines

We performed microarray expression profiling on the seven NSCLC and three HBEC cell lines before and after treatment with 100 nM and 1 μ M doses of 5-aza, and compared the resultant gene expression profiles. The microarray data have been deposited at the GEO database (<http://www.ncbi.nlm.nih.gov/projects/geo/>). We confirmed our array data in three ways: 1) each cell line was treated with 100 nM and 1 μ M doses of 5-aza in a single experiment to confirm array reproducibility and the ability of both doses to induce gene expression (Table 6-2); 2) biological replication was performed on the three HBEC cell lines 18 months apart on the U133 Plus 2.0 GeneChip, and for 4/7 NSCLC cell lines on the U133A GeneChip, and subsequently on the U133 Plus 2.0 platform (Table 6-3, below); 3) QPCR was performed on at least 15 genes in each cell line and at each dose of drug (Table 6-4, data not shown).

Since 5-aza alters gene expression independent of methylation changes, we reasoned that dose dependent changes in gene expression are probably more specific to aberrant promoter hypermethylation. We found a highly significant relationship between both the genes induced in the two treatments, and those induced in a dose-dependent manner (Table 6-2). We determined whether genes were reproducibly inducible by 5-aza over long term culture by comparing replicates on different types of Affymetrix arrays (U133A GeneChip and U133 Plus 2.0) for 4/7 NSCLC cell lines (A549, H2347, H1299, and H157), as well as data collected on the same type of chip for biological replicates performed 18 months apart on the three HBEC lines. Agreement between HBEC experiments performed 18 months

apart were highly significant (Table 6-3). Gene expression patterns across platforms also correlated well (Pearson correlation coefficients for overlapping gene sets on the two platforms in independent experiments ranged from 0.90 for H157 treated with 1 μ M 5-aza to 0.98 for H157 treated with DMSO).

Our analysis of the gene expression profiles of lung cancer cells before and after treatment with 5-aza identified 866/47000 transcripts that were up-regulated ≥ 4 fold in at least two lung cancer cell lines (Fig. 6-4). Individually, the cell lines exhibited substantial variations in expression phenotype: H1819 had the fewest (268), whereas H460 had the most (1100) (Fig. 6-5). The diversity in gene expression we observed may derive from several factors including etiology and histopathology (Table 6-1).

To further validate the induction patterns observed by microarray, we performed QPCR on 15 genes across all cell lines (Table 6-4). We found that, with the exception of Cathpesin Z (*CTSZ*), QPCR analysis correlated well with microarray expression changes. Disagreement between the array and QPCR data for *CTSZ* likely derives from the sensitivity of the Pearson correlation algorithm to small deviations above and below a mean-centered value.

Isolation of tumor-specific promoter methylation candidates

To identify genes that are methylated specifically in cancer cells, we performed similar induction experiments in three immortalized bronchial epithelial cell lines. All three HBEC cell lines exhibited changes in gene expression after 100 nM and 1 μ M 5-aza treatment (Fig. 6-6). In contrast to the cancer cell lines, the HBECs appeared to be relatively similar in their responses to 5-aza treatment, and bioinformatic analysis of the genes induced ≥ 4 fold in the HBECs suggests that many may be expressed specifically during development or only in certain tissues (Fig. 6-7; Table 6-5).

Beginning with the 866 transcripts that were induced ≥ 4 -fold in at least two NSCLC lines, we excluded 133 that were induced ≥ 4 -fold in HBECs, and we required that a given gene was expressed at a robust median level (MAS5 normalization procedures were used because this method gives an indication of whether a given probe signal is present or absent) in the HBECs with an Affymetrix p-value of ≤ 0.065 . Of the remainder, 460 were excluded on the basis of low (undetectable) expression in the untreated HBEC lines. We further filtered this list of genes by excluding 66 genes without defined 5' ends or that were otherwise poorly annotated, and 11 that were duplicate probes. This left 196 genes that were induced in the NSCLCs and that met the various filtering criteria.

5-aza can affect the expression of genes independent of their methylation status (Suzuki, Gabrielson et al. 2002). Before restricting the gene set to those with CpG islands, we asked whether our approach had identified a set that was enriched for genes associated with 5' CpG islands. The null hypothesis was that our selection criteria would make no difference on the frequency of selecting a gene with a CpG island. The expected rate for a

RefSeq annotated gene to contain a 5' CpG island (>500 bps in length) within 2 kb of its transcription start site is ~35% (Robinson, Bohme et al. 2004). Based on the March 2006 build, ~37% of the RefSeq 5'-UTR annotations contain 5' CpG islands within 500 5' bases. The 866 transcripts we identified on the basis of their induction pattern in NSCLC alone contained 435 RefSeq annotations, while 134 of the 196 transcripts that remained after filtering out genes as described above had RefSeq annotations (Fig. 6-4). Both of these groups had significant increases in CpG frequency (Table 6-5).

On the basis of these data, we examined each of the 196 genes and excluded those that did not have CpG islands defined as >300 bps, a GC content of $\geq 55\%$, and an observed vs. expected CpG ratio of ≥ 0.65 . The remaining 132 transcripts correspond to genes (listed in Fig. 6-8) that are candidates for tumor specific methylation in NSCLC on the basis of their expression pattern in HBECs (i.e. were expressed) and lung cancer cell lines (i.e. were not expressed in several or more lines), their response to 5-aza in lung cancer cells (induced ≥ 4 fold), and the presence of a 5' CpG island (Fig. 6-8).

Lung cancer vs. normal lung expression patterns of the 5-aza induction gene set

While there were other interesting gene sets in our data, particularly those genes that were induced by 5-aza in the HBEC lines, but were expressed in the NSCLC panel (i.e. genes that were candidates to have undergone tumor-specific promoter hypomethylation and thus potentially function as oncogenes) in this study we focused on genes that were likely to have undergone tumor specific promoter hypermethylation leading to inactivation of their expression. We first determined whether our 5-aza induction gene set reflected the gene expression phenotype of a broader set of NSCLC cell lines and HBECs. Using Affymetrix microarray mRNA expression data for NSCLC (N=31; combined U133A and B chips) and HBEC (N=7; U133 Plus 2.0) cell lines, we found that all HBEC lines express relatively high levels of these genes, but the lung cancers, while of diverse histologies, express much less (Fig. 6-9A). These facts suggest that loss of expression of the genes in the 5-aza induction gene set is a common event in NSCLC.

To determine whether the expression patterns we identified *in vitro* accurately represent those identified by microarray profiling in primary lung cancers, we explored whether the 132 genes in the induction set could distinguish uncultured normal lung from primary lung cancer in two separate microarray data sets. These data are derived from different lung tumor sources (see methods), collected over a period of several years and comprise expression phenotypes for primary NSCLC (N=45) and counterpart normal lung (N=29), and were randomly selected from a larger panel of array samples. After extracting the relevant probes and filtering the data, we found that the majority of genes were on average expressed at higher levels in the normal samples, and that the 5-aza induction gene

set clearly distinguished normal lung from lung cancer in these data (Fig. 6-10). 94/117 unique genes in this group were differentially expressed between tumor and benign tissue based on the SAM algorithm (90th percentile confidence, false discovery rate among the 94 significant genes was 0.11 and the delta value (false discovery rate) used to identify significant genes was 0.54 (Table 6-7).

Tumor-acquired promoter methylation often coincides with allele loss. To determine whether any of the 132 candidate genes were also subject to copy number losses, we analyzed CGH data for the same panel of NSCLC cell lines that were used for the microarray studies (N=31). Of the 132 genes, approximately half (58/132) had corresponding probes with high quality data on the Stanford array CGH platform. Of these, 62% (36/58) exhibited a net (median) allele loss across the panel of 31 NSCLC lines (Fig. 6-12, data not shown). Thus, beginning with 5-aza induction data in lung cancer, we have identified 132 genes with 5' CpG islands that are differentially expressed in primary lung cancer compared to normal lung tissues, many of which are also subject to frequent copy number losses in corresponding NSCLC lines.

Methylation analysis of 45/132 5-aza induction candidates in lung cancer and HBEC cell lines, and normal lymphocytes

To determine whether the genes identified in our screen are methylated in lung cancer cell lines, we designed MSP primers sets (methylated and unmethylated specific) for 45 of 132 candidate genes and two control gene primer sets and tested these on the seven lung cancer lines used for the 5-aza induction studies (primers, setup, and protocols maybe found in methods and supplementary material). As determined by MSP, between 19 and 25 genes out the 45 loci were methylated in any given tumor cell line, whereas at most 7 were methylated in the HBECs (Fig. 6-11). Interestingly, we found that several loci were positive for both methylated and unmethylated alleles. This is consistent with previous studies (Sato, Fukushima et al. 2003). As a further control for tumor specific methylation, and to determine whether these markers might be useful in a clinical setting, we tested whether any of the genes were methylated in normal lymphocyte DNA. This control is important because peripheral blood lymphocytes are almost always present in biopsy specimens as well as tumors, and the presence of methylation in these cells would preclude use of a given marker for patient screening purposes. Although we found different promoter hypermethylation profiles between different sources of lymphocytes (data not shown), in this study, a gene promoter was counted as methylated if there was a methylated product in any source of lymphocytes. On this basis, we found that 11 genes were methylated in at least one lymphocyte source. We grouped the genes according to their methylation patterns as follows: genes with tumor-specific methylation (group I; 31 genes); genes with some

methylation in HBEC, but not in normal lymphocyte DNA (group II; 5 genes); genes with methylation in lymphocyte DNA (group III; 11 genes).

Methylation analysis of the 45/132 5-aza induction gene set in primary lung cancers and normal lung

It has been suggested that tumor cell lines acquire methylation in culture and as a result may not accurately reflect the methylation patterns of tumors *in vivo* (Bestor 2003; Paz, Fraga et al. 2003). To address this issue, and to determine whether any of the markers we found were methylated in primary tumor samples, we tested all 45 markers in twenty matched pairs of primary NSCLC and counterpart normal lung tissue (Fig. 6-12). The frequency of methylation in a given tumor ranged from 33/45 to 17/45 genes. When all genes were included, methylation was more frequent in the matched tumor samples than in the normal control samples, which was statistically significant ($P = 4.72 \times 10^{-6}$, paired t-test). Basonucleolin (*BNC1*) and lysyl oxidase (*LOX*) were methylated in nearly all of the primary tumors examined, but were not methylated in normal lymphocytes, and infrequently in normal lung. By comparison, *p16* and *RASSF1A* were methylated in this same NSCLC panel at 30% and 40% rates, respectively (Zochbauer-Muller, Fong et al. 2001). The appearance of low-level methylation in some normal counterpart tissue may result from field effects and/or tumor cell contamination. Some markers were methylated at high frequency in tumors (>30%; compared to *p16* and *RASSF1A*, 30% and 40%, respectively) and never in matched normal tissue such as *CTS2* and placental growth factor (*PGF*).

In general, the methylation frequency of group I genes was similar to that of the cell lines we used in this study; where there was frequent methylation in the cell lines, there was frequent methylation in the primary tumors (Fig. 6-11 & Fig 6-12). Group II and III genes also followed the patterns identified in the cell lines; where methylation was found in the

HBEC, we found frequent methylation in both primary tumors and matched normal lung. When methylation was detected in normal lymphocyte DNA and/or HBEC DNA, we also found methylation in both primary tumor and normal lung DNA samples (which could have lymphocyte contamination). While all of these genes could be involved in lung cancer pathogenesis through promoter methylation and concomitant loss of expression, we focused on the 31 group I genes as being the best candidates for diagnostic markers to avoid genes found to be methylated in normal lung or lymphocytes.

Expression and methylation analysis of select genes in breast cancer cells

While there was some overlap between genes induced by 5-aza among the NSCLC lines, the predominant pattern we found reflects significant expression differences within the same tissue type (Fig. 6-9). The diversity we observed in NSCLC led us to explore whether other epithelial cancers differed dramatically in their response to 5-aza. When we compared other cell types (colon cancer (HCT116), breast cancer (MCF7), SCLC (H526)) after 5-aza induction by significance analysis of microarray (SAM) and cluster analysis, we found that while each cell line clustered with itself independent of treatment, SCLC and breast cancer cells but not the colon cancer cell line HCT116, clustered apart from NSCLC (Fig. 6-13A). However, after supervised hierarchical cluster analysis using our final 5-aza induction gene set tissue-of-origin distinctions were no longer apparent (Fig. 6-13B).

To further explore the finding that 5-aza induction patterns in cancer cell lines may be independent of tissue of origin differences, we compared our data set to those of Sato et al., who used Affymetrix's U133A chip to examine gene induction patterns after 5-aza treatment in four pancreatic cancer cell lines (Sato, Fukushima et al. 2003). The authors reported that 475 genes were up-regulated >5 fold in at least 1 cell line. Of these 475 genes, 203 were also up-regulated in at least one of our NSCLC cell lines, with 127 up-regulated in two or more (Table 6-8). Bioinformatic analysis of the overlapping gene set between Sato et al. and our data indicates some highly significant similarities in the position of the genes induced by 5-aza in these two organ sites (Table 6-9). Multiple genes in two chromosomal regions, Xp11.2-11.4 and 6p21.3, were induced in both types of cell lines, and based on the gene

density in these genomic regions, each enrichment was highly significant (3.01×10^{-09} and 1.01×10^{-07} , respectively, Fisher's Exact Test).

The overlap between the gene induction patterns we found in NSCLC, SCLC, breast, and colon cancer cells in our 5-aza induction microarray experiments, and that previously reported in pancreatic cancer cells led us to explore whether any of the genes we found are also methylated in breast cancers. First we examined 15 of the genes found to be frequently induced by 5-aza and methylated in NSCLC in 6 breast cancer cell lines (HCC3153, HCC1143, HCC1937, SKBR3, ZR-75-1, and MCF7) and found nearly all to be induced by 5-aza in these cells (Table 6-4; Fig. 6-14). We then analyzed the expression pattern of the 5-aza induction gene set across a panel breast cancer cell lines and found that for the 5-aza induction panel (by average linkage cluster analysis) most of the lung cancer cells and approximately half of the breast cancers fall into a major cluster, distinctly apart from the remaining breast cancer cells and the immortalized HBECs, which form their own tight cluster with a minimum Pearson correlation coefficient of >0.7 (6-15A & B). These data suggest that tumor rather than tissue-specific gene expression patterns are the predominant factor driving the clustering algorithm for the 5-aza induction gene set.

We selected 8/15 markers, found to be induced by 5-aza in both lung and breast cancer cells, for analysis in primary breast tumor material. 23 of the primary breast tumors used in this study form part of a large data set used in several earlier studies where fundamental histological and phenotypic differences were defined between subtypes of ductal breast carcinomas (Perou, Sorlie et al. 2000). The DNA from these samples was derived from bulk tumor specimens upon surgical resection from the primary tumor site,

metastatic sites, or at autopsy. All of these tumor specimens, with one exception were stage IIb or later. We found that among the 8 genes tested in 23 breast carcinomas, 7 were frequently methylated (60-90%) (Fig. 6-16). These breast cancer samples did not have counterpart normal tissue.

To address whether methylation for these 8 genes was detectable in benign breast tissue a further 14 tumor samples that have matched benign tissue were examined (see methods); these samples are primarily early stage tumors (\leq stage IIb) collected upon surgical resection of the primary tumor. The counterpart benign tissue was collected by fine needle aspiration in the ipsilateral breast except where indicated and have not been described previously. Again methylation was frequent, although overall there was more frequent methylation in the more advanced tumor stage UNC group. Only *SOX15* exhibited frequent methylation in benign breast material (Fig. 6-16 & Fig. 6-18; Table 6-10).

Methylation specific PCR, while robust, is extremely sensitive and can detect methylated sequences in the presence of large amounts of unmethylated DNA. We used sodium bisulfite DNA sequencing to confirm that the MSP primer sets used in these studies amplified the appropriate target sequences and that these sites were *bona fide* hypermethylated CpG islands. We designed primers that flank the MSP priming sites for the 8 genes examined and then cloned and sequenced PCR products from bisulfite treated HBEC and/or lymphocyte DNA and tumor cell DNA. Between 8-20 subclones from each selection plate for each cell type and gene were analyzed. With the exception of *NRCAM*, all sequences were heavily methylated in the tumor cells but not in lymphocytes or HBEC DNA

(Fig. 6-17). Based on these data, and its infrequent methylation in breast cancer, we excluded *NRCAM* from subsequent analyses.

Examination of the methylated gene set in matched pairs of colon and prostate cancers

Tumor-specific promoter hypermethylation is often also tissue-specific. To explore whether the 7 genes (*BNC1*, *LOX*, *ALDH1A3*, *MSX1*, *CCNA1*, *CTSZ*, *SOX15*) we identified in the previous section were methylated in other tissues besides breast and lung, we examined an independent set of primary colon and prostate cancers and their matched normal tissues. For comparative purposes, we included methylation data for *p16* and *RASSF1A* for all tumor types examined (Fig. 6-17, Table 6-10). Data for *RASSF1A* and *p16* are derived from published work as annotated in the table legend (Esteller, Fraga et al. 2001; Zochbauer-Muller, Fong et al. 2001; Maruyama, Toyooka et al. 2002; Holst, Nuovo et al. 2003; Lewis, Cler et al. 2005; Takahashi, Shigematsu et al. 2006).

BNC1, *MSX1*, and *CCNA1* were frequently methylated in all 4 tumor types. However, *CCNA1* exhibited significant methylation in benign prostate and colon tissues. This suggests that *CCNA1* may undergo tissue-specific methylation during cellular differentiation in certain tissues, but not others. *BNC1* and *MSX1* showed high sensitivity and specificity for tumors when compared to benign counterpart tissues (estimated values and 95% confidence intervals: 0.81 (0.75-0.86); 0.67 (0.60-0.75), respectively). For *BNC1* and *MSX1*, both prostate and colon benign tissues did have some methylation, however the pattern was different from *CCNA1*. *ALDH1A3* was specifically methylated in all tumor types, albeit less frequently than *BNC1* or *MSX1*, showing the highest sensitivity in breast and prostate and highest specificity in lung. *LOX* and *CTSZ* methylation was restricted to lung and breast tumors, and in both cases was highly specific. *SOX15* was methylated in most benign tissues and has been omitted from the histogram for purposes of clarity.

DISCUSSION

We used global gene expression profiling (47,000 transcripts) of 7 lung cancer cell lines before and after treatment with 5-aza to identify genes that were significantly up-regulated by this treatment. We performed similar experiments in three newly available immortalized human bronchial epithelial cells to identify genes whose expression was selectively lost in lung cancer, expressed in normal lung epithelium, but subject to 5-aza induction. The use of these cells as part of a global methylation induction screen has not been described previously. We applied a series of biological filters to extract a list of methylation candidates, and statistical analyses of the major steps in this process suggested that successive lists were enriched for genes with 5' CpG islands. Only those genes that were induced in more than one lung cancer and had well-defined CpG islands in their putative promoter regions were selected. This led us to identify 132 candidate genes of which 45 have been studied in detail in 20 primary lung cancers and companion normal lung tissue.

The large majority of the 132 genes we have identified have not been described to undergo tumor specific promoter hypermethylation. Expression of these genes distinguishes primary lung cancers from normal lung in the same patient. While there are probably many genes that are methylated – perhaps at random – during carcinogenesis, we found that the vast majority of the 45 genes studied here undergo tumor specific methylation in multiple primary lung cancers. Eight of these 45 genes were studied in a panel of 105 primary tumors from NSCLC, breast, colon and prostate cancers and 82 histologically normal companion tissues, which showed that many undergo methylation in these common epithelial cancers.

Thus, this approach has identified many new genes subject to frequent tumor acquired methylation in lung, breast, colon, and prostate cancers. Frequent methylation of specific genes in multiple independent cancers strongly suggests, although does not prove, that these genes are functionally relevant to cancer pathogenesis.

One goal of this study was to identify new genes involved in tumor specific methylation for follow up functional analysis. To this end, our screen uncovered some well-established methylation markers that have tumor suppressor activity, including *TIMP3*, *CDH1*, and *SFRP1*, but missed others such as *p16* and *RASSF1A*. The reason for this highlights one of the limitations of current microarray technology in that commercial arrays cannot always discriminate between alternative splice forms of genes; both *p16* and *RASSF1* have constitutively expressed alternative isoforms that can hybridize to probes specific for these loci. Since both genes have expressed isoforms (*p14* and *RASSF1C*) that differ only in their 5' regions, none of the probes specific to these genes detected differences in expression. This limitation means that we have probably missed isoforms of genes that are subject to tumor-specific methylation, but that are part of an active transcription locus.

Most of the genes identified in this study are novel methylation candidates in NSCLC, however some have been described in other tissues. *LOX* was frequently methylated in our panel of cell lines and NSCLC tumors, and was recently shown to be methylated in gastric cancers (Kaneda, Wakazono et al. 2004). *CCNA1* was shown to be methylated in head and neck cancers and was inversely correlated with *p53* mutation (Tokumaru, Yamashita et al. 2004). Interestingly, in our study, *CCNA1* was methylated in A549, which has wild-type *p53*, but not in NSCLC with mutant *p53*. Loss of dual specificity

phosphatase I (*DUSP1*) expression as determined by immunohistochemistry inversely correlates with increasing malignancy of prostate cancers, and methylation of its promoter appears to be an early event in this disease (Rauhala, Porkka et al. 2005). In another recent report, tissue factor pathway inhibitor 2 (*TFPI2*) methylation was used as part of a six gene panel to screen for cancer in pancreatic juice specimens (Matsubayashi, Canto et al. 2006). Promoter methylation of the transcription factor *TWIST1*, has been described in several reports, and is very frequent in neuroblastoma, cervical and breast cancers, although, curiously, high expression of *TWIST1* seems to be necessary for breast cancer metastasis (Alaminos, Davalos et al. 2004; Mehrotra, Vali et al. 2004; Yang, Mani et al. 2004; Feng, Balasubramanian et al. 2005). The proapoptotic *BCL2* family member, *BIK*, was identified in a global screen for promoter methylation in melanoma using restriction landmark genomic scanning (Pompeia, Hodge et al. 2004).

Our data suggests that some genes, such as *CCNA1*, clearly undergo both tissue and tumor-specific methylation. Tissue-specific promoter hypermethylation arises in response to both extrinsic and intrinsic signals during cellular differentiation, and may account for the distinctive pattern we observed for this particular cyclin (Jaenisch and Bird 2003). The biological basis of frequent tumor-specific hypermethylation in multiple tissues coincident with tissue-specific methylation in another tissue is unknown. However, two well characterized tumor suppressors, *p16* and *RASSF1A*, exhibit similar tumor and tissue-specific promoter methylation profiles; *p16* methylation is frequently observed in benign breast tissue, even in young women, and *RASSF1A* is observed in benign liver and colonic epithelium (Holst, Nuovo et al. 2003; Lehmann, Berg-Ribbe et al. 2005). Thus, the presence

of promoter methylation in selected normal tissues does not exclude a gene from being an important tumor suppressor. Nevertheless, the information on such methylation is important for clinical applications.

Another pattern of promoter hypermethylation evident in our data, exemplified by *LOX* and *CTSZ*, is characterized by frequent but exclusive methylation in certain tumor types. According to available data provided through various online databases (Genecard (Weizmann Institute), and Source (Stanford University)), both *LOX* and *CTSZ* are widely expressed. Both genes also have a several homologues that may be partially redundant, or they may have tissue-specific functions important to tumorigenesis in breast and lung, but not in prostate or colonic epithelium. Several other genes exhibit a similar, restricted methylation profile, such as *BRCA1* in breast and ovarian tumors and *GSTP1* in liver and prostate cancers (Esteller, Silva et al. 2000; Jeronimo, Henrique et al. 2004). Genes that are methylated with high frequency and specificity only in certain tumors might also be of value in the development of a promoter hypermethylation profile to screen for several different cancers at once.

Perhaps the most important profile identified in this study is that of tumor-acquired methylation involving the four most common epithelial tumors. When all matched tumors were combined, *BNC1* and *MSX1* were both highly sensitive and specific for tumor detection. As yet, relatively few loci have been identified that exhibit frequent (>50%), tumor-specific methylation across several types of malignancies. There are several genes that exhibit frequent methylation in NSCLC and other tumor types, such as the tumor suppressor gene adenomatosis polyposis coli (*APC*) or retinoic acid receptor beta (*RARβ*),

but these genes are often methylated in counterpart benign tissue, especially in tumors where field effects are often seen, such as NSCLC (Zochbauer-Muller, Fong et al. 2001; Shivapurkar, Stastny et al. 2006). The identification of more loci like *BNC1* and *MSX1* will be a key element in developing a promoter hypermethylation profile for the early detection of human cancer.

There are relatively few tumor-specific lesions that occur with significant frequency in all types of tumors with the important exceptions of p53 mutation, genomic instability, and constitutive reactivation of telomerase (Hollstein, Sidransky et al. 1991; Mitsudomi, Steinberg et al. 1992; Kim, Piatyszek et al. 1994). The wealth of data available in the scientific literature suggests that aberrant DNA methylation may be another key contributor to cellular transformation. The frequency and diverse patterning of tumor-specific promoter methylation in our panel of lung, colon, prostate, and breast carcinomas coupled with the findings recently reported by others, indicates that tumor-acquired promoter hypermethylation patterns are non-random (Baylin and Ohm 2006; Keshet, Schlesinger et al. 2006). It is possible that there are ongoing random methylation events in cancer cells; however, that some genes are so frequently methylated in multiple tumor types but not in their companion benign tissues, suggests to us that there is something about their function or primary sequence that makes them particularly susceptible to promoter hypermethylation and silencing in the context of cellular transformation.

By contrasting the genome-wide changes in gene expression of normal and lung cancer cells, we were able to gain insight into the complexity of the methylation program required for cells to become fully malignant. Even though we began with a highly

structured, organ-specific screen, by applying successive biological and statistical filters we identified several genes with exceptionally high methylation frequencies and tumor specificity in primary lung and breast tumors. Several of these genes also show significant methylation in colon and prostate tumors but not in counterpart benign tissues. We conclude that while tumors differ in their molecular phenotypes and pathogenesis, the pathways they follow toward malignancy may be similar and may be reflected in the methylation programs they engage, which in turn may provide an opportunity to exploit in early diagnosis or therapeutic strategies. Subsequent studies will be needed to determine whether these novel methylated loci could be useful in early detection screening, and whether loss of expression of their associated genes contributes to tumor initiation and pathogenesis.

Tables

Table 6-1 . Clinicopathological Features of Cell Lines Used in Microarray Studies

Sample Name	Cell Type	Dx	Age	Gender	Race	Smoker/PkYrs
HBEC2	Bronch. Epith	NSCLC	68	M	Caucasian	Y
HBEC3	Bronch. Epith	No Cancer	65	M	Caucasian	Y
HBEC4	Bronch. Epith	Lung Cancer	71	F	Caucasian	Y
A549	Lung Cancer	Adeno	58	M	Caucasian	Y
H460 [‡]	Lung Cancer	Large Cell	~45	M	Caucasian	Y/~40
H1299	Lung Cancer	Large Cell	43	M	Caucasian	Y/50
H1819	Lung Cancer	Adeno	58	F	Caucasian	Y/80
H157	Lung Cancer	Squa	59	M	Caucasian	Y/?
H1993	Lung Cancer	Adeno	46	F	Caucasian	Y/30
H2347	Lung Cancer	Adeno	54	F	Caucasian	N/0
H526	Lung Cancer	SCLC	55	M	Caucasian	?
MCF7 [†]	Breast Cancer	Adeno	69	F	Caucasian	N/A
HCT116 [†]	Colon Cancer	Colorectal Carcinoma	Adult	M	?	N/A

Data are from Phelps et al.

‡John D. Minna, unpublished observations

†American Type Culture Collection Website

Table 6-2 Reproducibility and dose-dependence of gene induction by 5-aza-2'-deoxycytidine

Cell Line	# of Genes Induced >4x by 5-aza			p-Value	
	100 nM	1 μ M	Up in both [†] 1 μ M > 100 nM [‡]		
HBEC2	602	591	99.7%	357	6.07×10^{-6}
HBEC3	148	289	96.6%	106	2.24×10^{-7}
HBEC4	88	219	100%	72	1.19×10^{-9}
H2347	127	387	99.2%	77	0.021
H1299	74	402	100%	64	8.96×10^{-11}
A549	60	188	96.7%	56	9.08×10^{-13}
H1993	11	80	100%	11	9.77×10^{-4}
H157	114	416	98.2%	100	1.72×10^{-15}
H460	27	501	100%	27	1.49×10^{-8}
H1819	25	67	100%	24	1.55×10^{-6}

[†] A given gene was "up in both" if it was induced >4x in the 100 nM experiment and went up any amount in the 1 μ M experiment.

[‡] A given gene was counted as dose-dependent when induction relative to control was >4x in the 100 nM array **and** 1 μ M induction was > 100 nM for the same gene.

Table 6-3. Agreement and 95% confidence intervals for biological replicates performed 18 months apart

Cell Line	Genes Compared	Overlap*	Point Estimate	95% Confidence Interval
HBEC2	Top 1000	1620	0.746	(0.724 - 0.767)
	Top 2000	3291	0.711	(0.695 - 0.727)
HBEC3	Top 1000	1682	0.762	(0.741 - 0.782)
	Top 2000	3431	0.711	(0.695 - 0.726)
HBEC4	Top 1000	1606	0.810	(0.790 - 0.829)
	Top 2000	3278	0.732	(0.717 - 0.747)

*Agreement analysis was performed as described in methods. Comparisons were made between the top 1000 and 2000 genes for each replicate at 1 μ M. The overlap is the number of genes that are coincident between the two datasets: 2000 for top 1000, and 4000 for top 2000.

Table 6-4. Correlation between microarray and QPCR data

Gene	Induction Frequency†		Range of Expression Change‡		Pearson
	Array	QPCR	Array	QPCR	
ADRB2	2/7	5/7	-0.81 - 2.71	0 - 3.66	0.829
ALDH1A3	2/7	5/7	-0.56 - 3.88	0.00 - 3.31	0.333
BNC1	6/7	7/7	1.40 - 8.40	1.48 - 12.06	0.892
CCNA1	5/7	5/7	-1.06 - 6.76	-1.03 - 6.6	0.965
CDH1	2/7	5/7	1.63 - 5.17	-1.91 - 7.10	0.603
CTSZ	2/7	4/7	-1.56 - 2.38	0.44 - 1.56	-0.181
IRX4	3/7	6/7	0.31 - 5.78	2.14 - 17.09	0.978
LOX	3/7	5/7	-2.05 - 2.45	-2.16 - 44.42	0.932
MAF	2/7	5/7	-1.61 - 2.26	-0.26 - 14.31	0.906
NRCAM	3/7	5/7	-0.39 - 2.44	0.74 - 6.37	0.807
PHLDA1	3/7	3/7	-0.43 - 2.28	0.12 - 3.34	0.758

†Induction frequency is the number of cell lines out of 7 where a given gene was induced >4-fold.

‡Expression change is indicated in Log₂

Table 6-5. EASE Analysis of Genes Induced >4 fold in HBEC lines

SYSTEM	TERM	COUNT	PERCENT	Fisher's Exact Test
Biological Process	RESPONSE TO WOUNDING	9	11.11	7.007E-05
Molecular Function	PATTERN BINDING	5	6.17	3.536E-04
Molecular Function	GLYCOSAMINOGLYCAN BINDING	4	4.94	3.342E-03
Molecular Function	POLYSACCHARIDE BINDING	4	4.94	3.740E-03
Biological Process	INFLAMMATORY RESPONSE	5	6.17	6.572E-03
Biological Process	DNA PACKAGING	5	6.17	8.555E-03
Biological Process	PROTEIN COMPLEX ASSEMBLY	5	6.17	9.106E-03
Biological Process	RESPONSE TO STRESS	10	12.35	9.868E-03
Biological Process	CHROMATIN ASSEMBLY OR DISASSEMBLY	4	4.94	1.014E-02
Biological Process	RESPONSE TO EXTERNAL STIMULUS	12	14.81	1.176E-02
Cytoband	6P22-P21.3	2	2.47	1.359E-02
Biological Process	RESPONSE TO EXTERNAL BIOTIC STIMULUS	7	8.64	1.378E-02
Biological Process	REGULATION OF CELL DIFFERENTIATION	3	3.70	1.468E-02
Biological Process	CHROMOSOME ORGANIZATION AND BIOGENESIS	5	6.17	1.592E-02
Biological Process	CHROMOSOME ORGANIZATION AND BIOGENESIS (SENSU EUKARYOTA)	5	6.17	1.592E-02
Biological Process	DEVELOPMENT	14	17.28	1.729E-02
Molecular Function	BINDING	41	50.62	1.949E-02
Cytoband	6P21.3	3	3.70	1.954E-02
Molecular Function	HEPARIN BINDING	3	3.70	2.261E-02
Molecular Function	ION BINDING	15	18.52	2.470E-02
Molecular Function	METAL ION BINDING	15	18.52	2.470E-02
Molecular Function	CATION BINDING	14	17.28	2.532E-02
Biological Process	CELL ADHESION	7	8.64	2.805E-02
Cytoband	XQ28	3	3.70	3.396E-02
Biological Process	NEGATIVE REGULATION OF MONOCYTE DIFFERENTIATION	2	2.47	3.637E-02
Biological Process	CELL COMMUNICATION	22	27.16	3.859E-02
Biological Process	NEGATIVE REGULATION OF CELL DIFFERENTIATION	2	2.47	3.993E-02
Biological Process	ESTABLISHMENT AND/OR MAINTENANCE OF CHROMATIN ARCHITECTURE	4	4.94	4.210E-02
Biological Process	ORGANELLE ORGANIZATION AND BIOGENESIS	7	8.64	4.276E-02
Biological Process	MONOCYTE DIFFERENTIATION	2	2.47	4.349E-02
Biological Process	NEGATIVE REGULATION OF MYELOID BLOOD CELL DIFFERENTIATION	2	2.47	4.349E-02
Molecular Function	RECEPTOR BINDING	6	7.41	4.497E-02
Biological Process	NUCLEOSOME ASSEMBLY	4	4.94	4.831E-02
KEGG Pathway	HSA04512:ECM-RECEPTOR INTERACTION	3	3.70	4.890E-02
KEGG Pathway	HSA04512	3	3.70	4.890E-02
Molecular Function	CARBOHYDRATE BINDING	4	4.94	4.901E-02
Biological Process	REGULATION OF MONOCYTE DIFFERENTIATION	2	2.47	5.055E-02

Table 6-6. Analysis of CpG island enrichment for genes induced by 5-aza in microarray experiments

Gene List	RefSeq Annotation	5' CpG Island	No CpG Island	% with CpG Island	χ^2	χ^2 Monte Carlo (10^5)
All RefSeq	17820	6704	11116	37.6%	N/A	N/A
Up in NSCLC	435	240	195	55.2%	4.1×10^{-14}	1×10^{-5}
5-aza Induction	134	98	36	73.1%	2.2×10^{-16}	1×10^{-5}

Intersect tables were downloaded from the UCSC genome browser for each annotation list using the May 2006 genome build. Chi-square statistics were determined using the expected value of 37.6% with one degree of freedom. The Monte Carlo method was used to simulate samples of various sizes with 100,000 represented here.

Table 6-7. Significant gene list from SAM analysis of 5-aza gene set in primary NSCLC and normal lung.

Mean Intensity Class 1: Cancer	Mean Intensity Class 2: Normal	Fold difference of geom means	Gene Symbol
5819.9	1245.7	4.672	IGFBP3
7497.1	2349.7	3.191	TACSTD1
1457.9	301.3	4.839	CD24
5224.2	2159.6	2.419	SPINT2
4461.2	1734.5	2.572	KRT19
3362.3	1271.6	2.644	EHF
4988.3	2309.6	2.16	MAL2
1282.7	615.5	2.084	PLAU
2186.7	805.5	2.715	DSP
784.5	427.8	1.834	LGALS8
291.3	95.2	3.06	IMP-3
470.7	201.2	2.339	HIP1R
984.9	495.6	1.987	SERPINE2
187.4	55.6	3.371	LY6K
521.8	307.2	1.699	NOL3
912.9	522	1.749	DUSP23
271	137.5	1.971	LGALS8
103.7	50.8	2.041	TWIST1
1404.5	747.5	1.879	HIST2H2AA
2680.2	1487.1	1.802	CSPG2
3046.6	1556.2	1.958	KRT7
915.2	506.5	1.807	EPPK1
61.5	31.4	1.959	PTHLH
49.5	25.9	1.911	PTHLH
49.2	27.1	1.815	PGF
173.1	88.6	1.954	TFPI2
4040.5	2755.8	1.466	RNASET2
30	16.7	1.796	SERPINB5
1651.6	1208.8	1.366	DDX18
780.2	597.5	1.306	PPIG
637.7	463.3	1.376	BAZ1A
213.3	157.7	1.353	C16orf5
391.8	302	1.297	RAGE
520.5	376	1.384	AF1Q
393.4	307.2	1.281	SLC35D1
292.2	224.6	1.301	ZNF268
123.2	97	1.27	ZNF256
5177.8	4166.1	1.243	GPX1
77	59.9	1.285	BNC1
469.6	368.2	1.275	ALDH1A3
18648.6	15750.4	1.184	S100A6
391.5	335.7	1.166	GNA15
41.8	34.2	1.222	HES2
2502.2	2199.7	1.138	MYO5C
154.8	134.7	1.149	JUB
52.5	42	1.25	NEFL
297.1	259	1.147	SERPINI1
1446.9	11241.5	0.129	FHL1
1020.8	6061.7	0.168	KLF4
4011.4	15061.9	0.266	EPAS1
2301.1	10498.8	0.219	RGC32
4311.4	5903	0.073	FOSB
347.9	1128.1	0.308	AKAP12
390.9	1125.9	0.347	ADRB2
756.2	1780.8	0.425	LHFP
2239.6	5357.1	0.418	CTGF
810.3	1899.5	0.427	ANXA3
801.3	1436.8	0.558	MCAM
718.8	1796.1	0.4	TIMP3
957.7	1717.2	0.558	HSPC159
9261.3	19727.6	0.469	DUSP1
337.4	733.8	0.46	BMP2
490.3	754.3	0.65	KLF11
837	1396.5	0.599	CDKN1C
137.9	236.5	0.583	MSX1
357.1	537.3	0.665	PRKODBP
1848.5	2952.2	0.626	FCER1G
227.1	366	0.62	HOXA3
219	339	0.646	IL17D
487.4	748.4	0.651	CD109
880.1	1706.2	0.516	PTGS2
190.9	317.4	0.601	FAD158
857.4	1281	0.669	JAG1
239.8	359.9	0.666	DMD
2978.4	5164.8	0.577	GPCR5A
1260	1820.8	0.692	DOC1
305.6	443	0.69	NOTCH1
451.1	688	0.656	SLCO3A1
513.1	740.4	0.693	ARRDC4
1012.9	1501.7	0.675	F3
524.2	892.5	0.587	AREG
354.3	512.2	0.692	MAD
518.1	699.9	0.74	MAF
171.1	224.9	0.761	SNCA
493.8	677.9	0.728	ARMCK1
876.1	1083.7	0.808	TJP2
605.3	850.6	0.712	CXCL1
521.8	709.3	0.736	CTSZ
164.5	199.8	0.823	WNT5A
654.8	896.6	0.73	THBS1
225.9	279.3	0.809	ULBP2
348.6	413.5	0.843	ZNF559
981.6	1161.2	0.845	TFM2
709.4	842.8	0.842	PYGARD

Table 6-8 may be obtained online as supplementary table I in Shames et al.

Table 6-9. Analysis of overlapping genes between Sato et al. and Shames et al.

SYSTEM	TERM	COUNT	PERCENT	Fisher's Exact Test
Cytoband	XP11.4-P11.2	5	2.53	3.01E-09
Cytoband	6P21.3	9	4.55	1.01E-07
Biological Process	RESPONSE TO WOUNDING	17	8.59	1.72E-06
Molecular Function	ENZYME INHIBITOR ACTIVITY	13	6.57	2.92E-06
Biological Process	INFLAMMATORY RESPONSE	12	6.06	7.46E-06
Biological Process	RESPONSE TO EXTERNAL BIOTIC STIMULUS	17	8.59	1.23E-04
Biological Process	DEVELOPMENT	37	18.69	1.26E-04
Molecular Function	RECEPTOR BINDING	17	8.59	2.14E-04
Biological Process	RESPONSE TO STRESS	23	11.62	4.79E-04
Molecular Function	ENDOPEPTIDASE INHIBITOR ACTIVITY	8	4.04	5.11E-04
Molecular Function	PROTEASE INHIBITOR ACTIVITY	8	4.04	5.31E-04
Molecular Function	ENZYME REGULATOR ACTIVITY	16	8.08	5.93E-04
Cytoband	XP11.23	5	2.53	7.07E-04
Biological Process	RESPONSE TO EXTERNAL STIMULUS	28	14.14	8.68E-04
Cytoband	19P13.3-P13.2	3	1.52	9.69E-04
Biological Process	REGULATION OF CELL CYCLE	13	6.57	1.21E-03
Molecular Function	PROTEIN KINASE INHIBITOR ACTIVITY	4	2.02	1.26E-03
Molecular Function	KINASE INHIBITOR ACTIVITY	4	2.02	1.98E-03
KEGG Pathway	HSA04610:COMPLEMENT AND COAGULATION CASCADES	6	3.03	2.18E-03
KEGG Pathway	HSA04610	6	3.03	2.18E-03
Genetic association database	INFECTION	7	3.54	2.64E-03
Biological Process	MORPHOGENESIS	24	12.12	2.80E-03
Biological Process	MAP KINASE PHOSPHATASE ACTIVITY	3	1.52	3.45E-03
Biological Process	CELL-CELL SIGNALING	14	7.07	4.99E-03
Molecular Function	ISOPRENOID BINDING	3	1.52	5.00E-03
Molecular Function	RETINOID BINDING	3	1.52	5.00E-03
Biological Process	ORGAN DEVELOPMENT	20	10.10	5.25E-03
Biological Process	ORGANOGENESIS	20	10.10	5.37E-03
Cytoband	XQ28	5	2.53	6.74E-03
Molecular Function	GROWTH FACTOR ACTIVITY	7	3.54	7.14E-03
Biological Process	CELL PROLIFERATION	22	11.11	9.26E-03

Table 6. Frequency of promoter hypermethylation for 8 genes as determined by MSP for indicated tumors

D _x	LOX		MSX1		BNC1		CTSZ		ALDH1A3		CCNA1		NRCAM		SOX15	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
Breast Tumor*	16/23	70%	21/23	91%	19/23	83%	14/23	61%	17/23	74%	12/23	52%	4/23	17%	23/23	100%
Lung Tumor	19/20	95%	11/20	55%	18/20	90%	10/20	50%	9/20	45%	14/20	70%	18/20	90%	17/20	85%
Lung Benign	4/20	20%	3/20	15%	3/20	15%	0/20	0%	3/20	15%	7/20	35%	8/20	40%	15/20	75%
Breast Tumor ¹	5/14	36%	11/14	79%	9/14	64%	6/14	43%	4/14	29%	6/14	43%	ND	ND	11/14	79%
Breast Benign ¹	0/14 [†]	0%	5/14 [†]	35%	0/14 [†]	0%	0/14	0%	0/14	0%	1/14	7%	ND	ND	8/14	57%
Prostate Tumor	0/24	0%	20/24	83%	18/24	75%	0/24	0%	5/24	21%	19/24	79%	3/24	13%	24/24	100%
Prostate Benign	0/24	0%	10/24	42%	9/24	38%	0/24	0%	7/24	29%	6/24	25%	1/24	4%	21/24	88%
Colon Tumor	0/24	0%	21/24	88%	22/24	92%	0/24	0%	11/24	46%	24/24	100%	7/24	29%	24/24	100%
Colon Benign	0/24	0%	13/24	54%	10/24	42%	0/24	0%	7/17	29%	23/24	96%	4/24	17%	20/24	83%

Tissue procurement procedures and clinical information for samples may be found in the methods section. In brief, all prostate and colon tumors were Stage II or later, lung tumors ranged from Stage I to IIb. For breast tumors, see below. Benign tissue was obtained from the same patient in all cases except for the UNC samples, see below and methods.

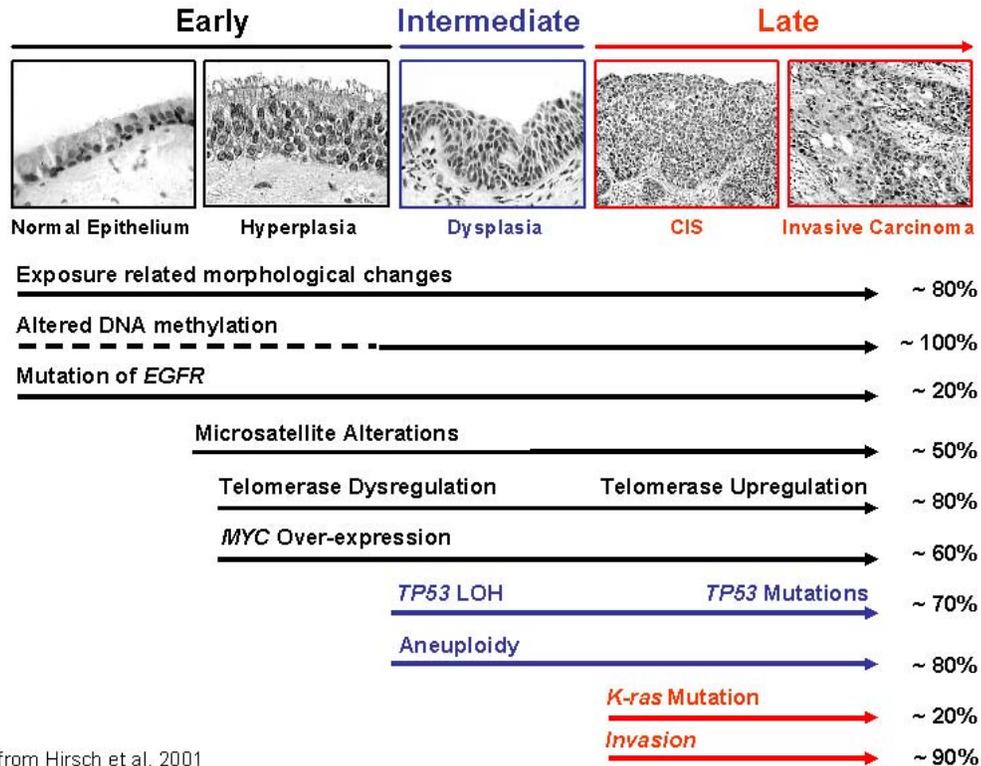
Cells with bold face type show a statistically significant difference in methylation frequency between tumor and normal samples according to a χ^2 statistic ($p < 0.05$).

[†]Benign breast was obtained from the ipsilateral breast except for one sample for LOX and BNC1, and two samples for MSX1, which were obtained from the contralateral breast in the same patient.

*Breast tumor samples were obtained through a collaboration with Chuck Perou at UNC. Samples in this group were all Stage IIb or later, with the exception of a single Stage I tumor.

¹Breast tumor samples were obtained through a collaboration with David Euhus at UTSW. All samples in this group are Stage IIb or earlier.

Figures



Adapted from Hirsch et al. 2001

Figure 6-1. Histological changes that occur in association with the common molecular alterations found in NSCLC. Adapted from (Hirsch, Franklin et al. 2001). The progression from benign bronchial epithelium through hyperplasia, dysplasia, carcinoma *in situ*, a finally to invasive carcinoma is accompanied by characteristic lesions in the lung. Almost all cases of squamous cell carcinoma of the lung are associated with smoking and altered DNA methylation. Other early events include microsatellite instability, telomerase dysregulation, and later on, p53 mutation.

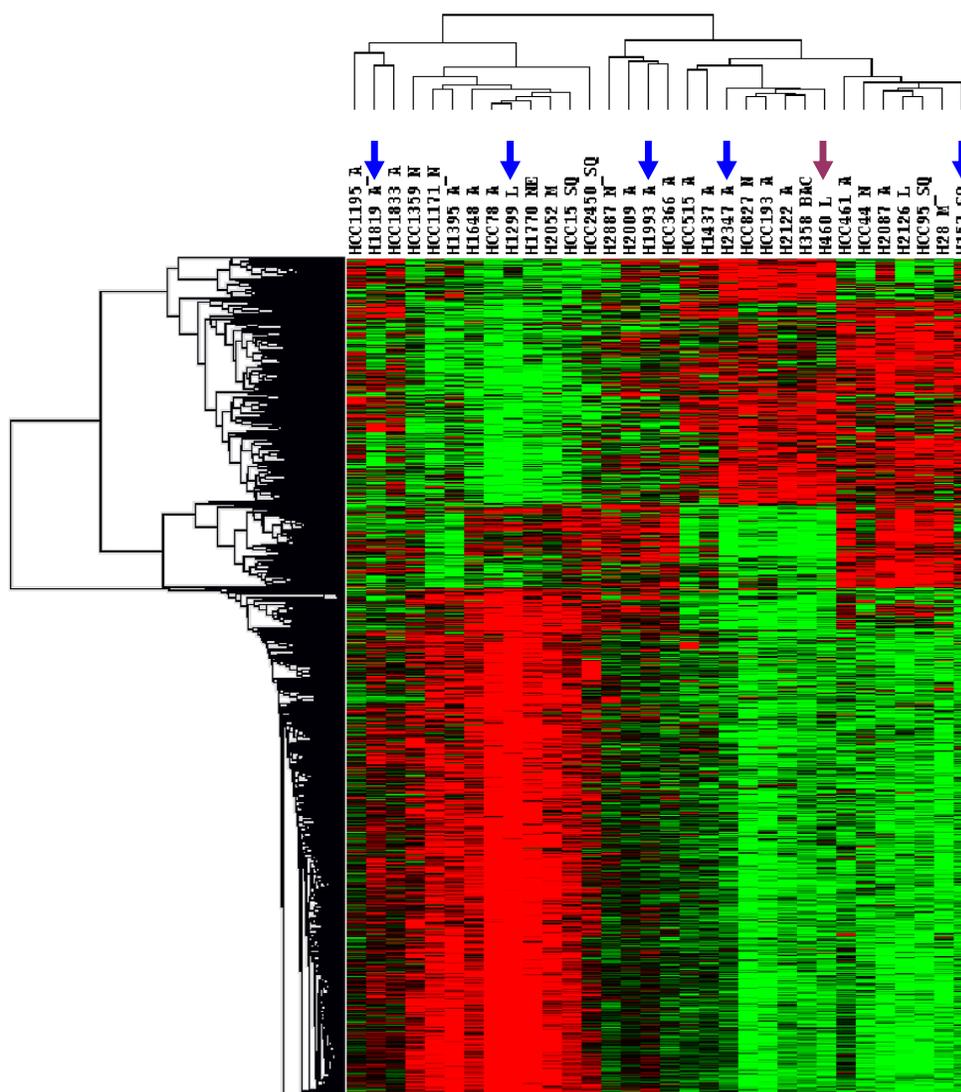
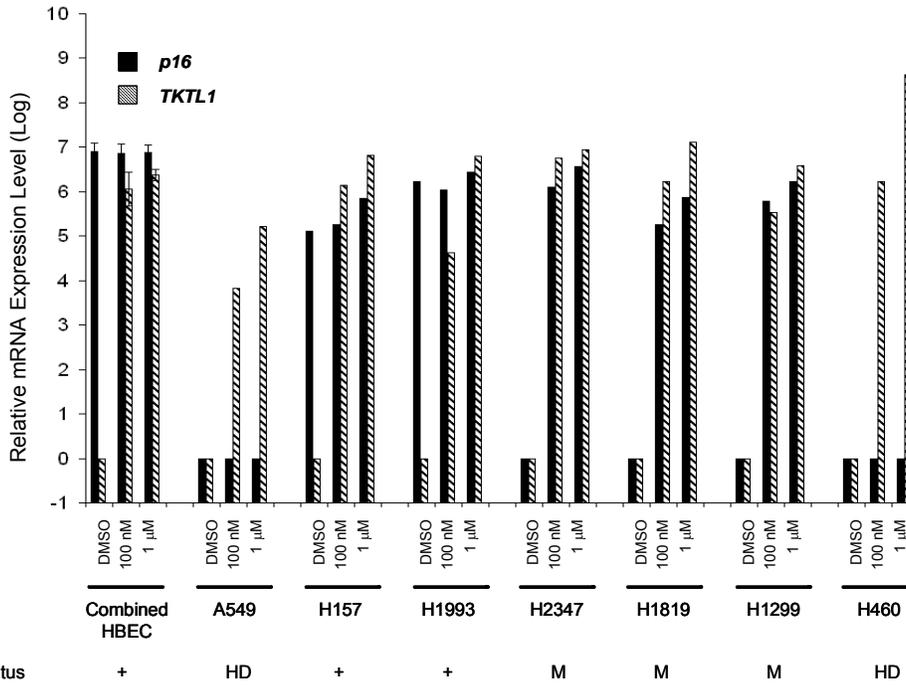
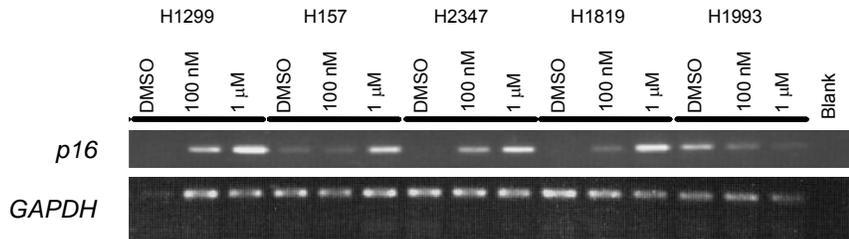


Figure 6-2. Unsupervised hierarchical cluster analysis of microarray data for NSCLC (N=31) cell line panel. Microarray data were normalized using RMA and then filtered using a minimum 2 fold difference between 50% of the samples for a given gene for each gene across samples. Data were then analyzed using the SAM algorithm with a delta value of 0.1 and 1000 permutations. Unsupervised hierarchical clustering using an average linkage algorithm was used. Arrays were centered about the median; genes were not. Red indicates 2 fold above the median value; green indicates 2 fold below the median value for a given gene. Blue arrows indicate cell lines selected as representatives of a part of the dendrogram. Purple arrow indicates a cell line that was selected as part of a second round of microarray studies. A549 was not included as part of the NSCLC cell line panel. **A** = adenocarcinoma; **SQ** = squamous cell carcinoma; **BAC** = bronchioalveolar carcinoma; **L** = large cell carcinoma carcinoma; **NE** = neuroendocrine; **NSCLC** = non-small cell lung carcinoma

A)



B)



C)

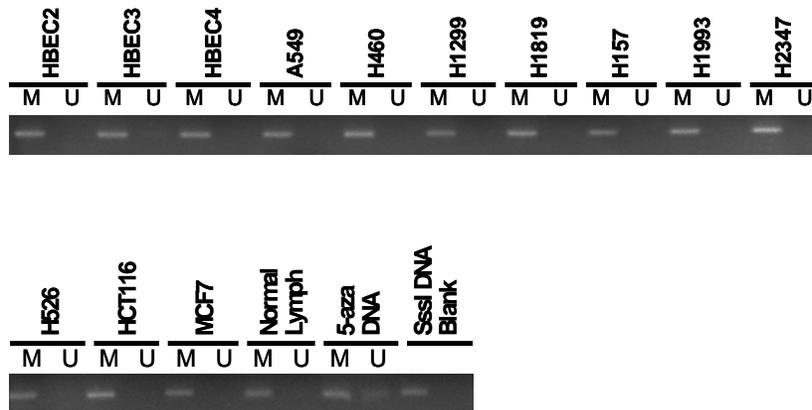


Figure 6-3. 5-aza-2'-deoxycytidine treatment induces genes silenced by promoter methylation in HBEC and NSCLC cancer cell lines. **A)** Quantitative RT-PCR for *p16* and *TKTL1* in HBEC and NSCLC. Solid bars are *p16* and cross-hatch bars are *TKTL1*. Data are normalized relative mRNA expression levels according to the $2^{\Delta\Delta Ct}$ method. HBEC2, 3, 4 all had similar profiles and were combined. *p16* status is indicated below each cell line; (+) expressed, (M) methylated, (HD) homozygous deletion. **B)** RT-PCR for *p16* in the indicated cell lines. *GAPDH* is a loading control. **C)** Methylation specific PCR for *TKTL1* in the indicated samples indicates complete methylation in all samples examined; (M) methylated, (U) unmethylated. *SssI in vitro* methylated DNA in positive control for methylated primer, 5-aza treated DNA is a positive control for the unmethylated primer sets (methods for PCR conditions and primer sequences).

<u>Gene Selection Criteria</u>	<u># of Genes</u>	
Genes Induced $\geq 4x$ by 5-aza in NSCLC	866	} 132 Tumor-specific methylation candidates
Subtract		
Genes Induced $\geq 4x$ by 5-aza in HBECs	(133)	
Subtract		
Genes not expressed in HBEC	(460)	
Subtract		
Poorly annotated genes (66) or duplicate probes (11)	(77)	
Subtract		
Genes without CpG islands	(64)	

Figure 6-4. Strategy used to identify methylation candidates by gene expression microarray. NSCLC and HBEC cell lines were treated with 5-aza and compared to controls (DMSO). We subtracted transcripts induced ≥ 4 -fold in HBEC (133) from the total number induced ≥ 4 -fold in 2/7 NSCLC lines (866/47000) since methylation of these genes is unlikely to be tumor specific. For practical purposes, we removed genes that were not expressed in HBEC (460), were duplicate probes (11), had poor annotation (66). Finally, we excluded genes without identifiable 5' CpG islands (64). The number of genes subtracted from the total induced ≥ 4 -fold in 2/7 NSCLC cell lines (866) is indicated next to each description in parentheses. We used the percentage of transcripts associated with 5' CpG islands as a measure of enrichment for the major steps in the filtering process. 37% of all RefSeq transcripts contain 5' CpG islands; 55% of the 866 5-aza induced transcripts had 5' CpG islands; 73% of the final 196 genes had CpG islands. Statistical analysis of these data appears in Table 6-5.

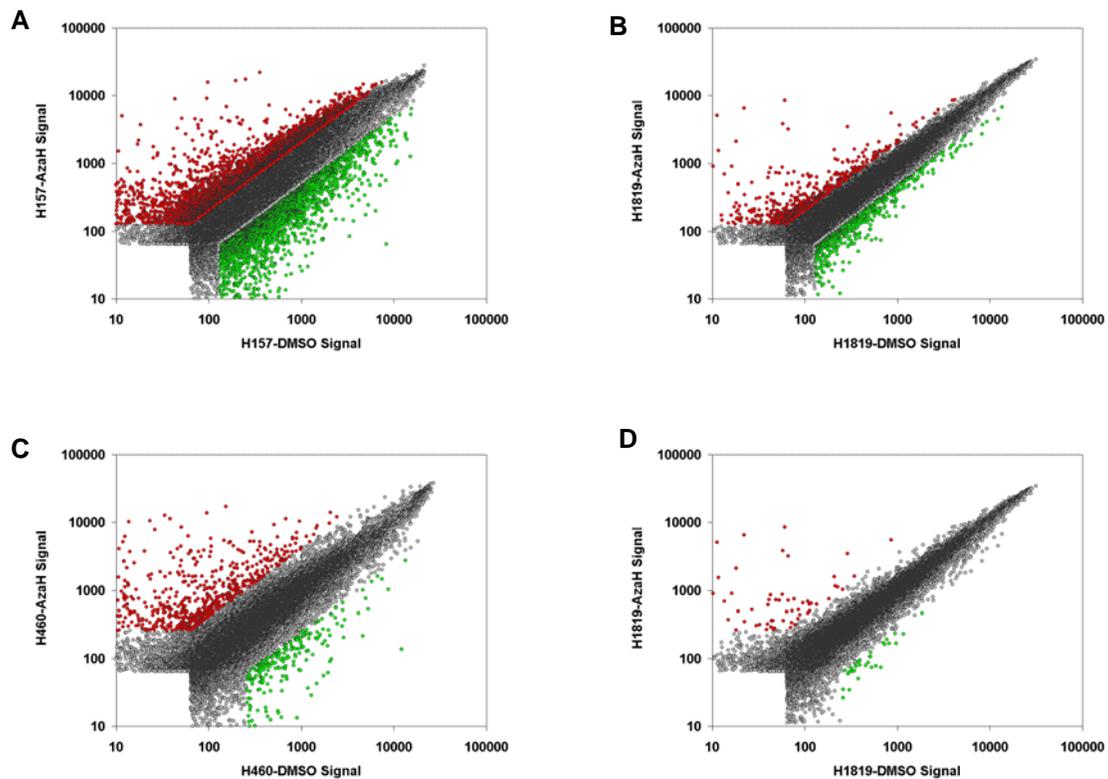


Figure 6-5. Scatter plots of microarray data for indicated cell lines before and after 5-aza treatment. Each point represents signal form a single probe. In panels A and B, a red point indicates 4-fold up-regulation; green indicates 4-fold down-regulation. The cell lines ranged broadly in terms of the number of genes affected by 5-aza treatment: in H157, 4407 genes were induced, whereas in H1819, 972 genes were induced. When we increased the threshold of a “significant” induction to ≥ 4 -fold, 866 the cell lines still exhibited substantial variations in expression phenotype: H1819 still had the fewest (268), whereas now, H460 had the most (1100).

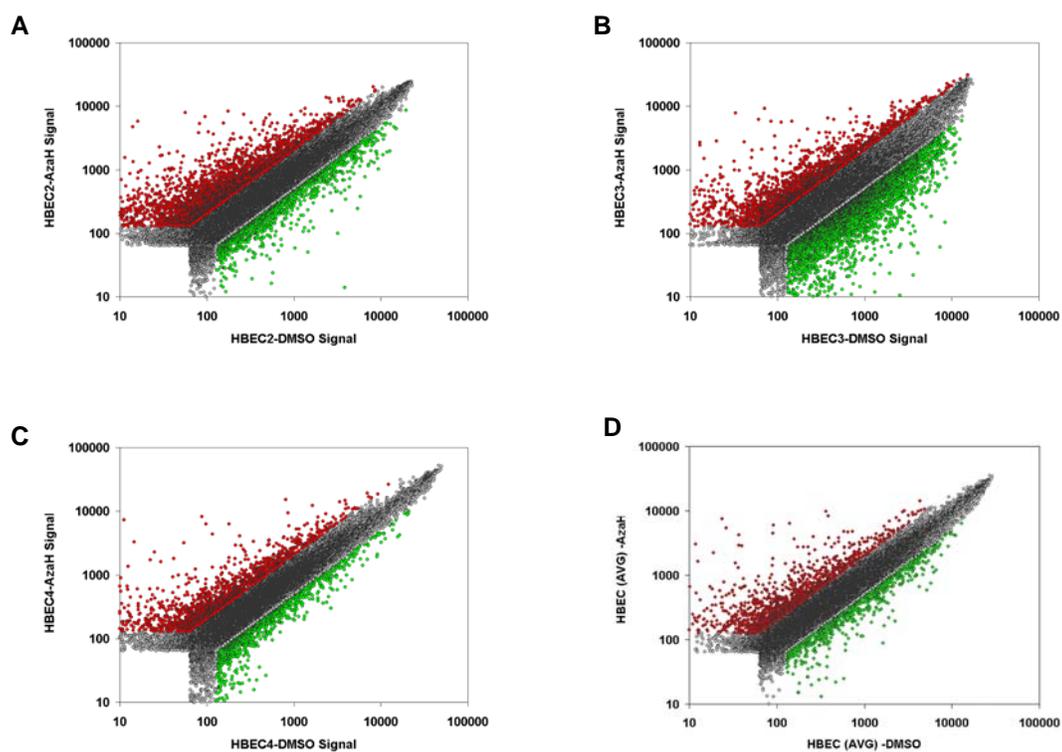


Figure 6-6. Scatter plots of microarray data for the HEBC cell lines before and after 5-aza treatment. Each point represents signal form a single probe. In panels A, B, and C compare gene expression changes in HBEC2, 3, and 4, respectively; panel D is an average of all 3 HEBC cell lines. Red points indicate 4-fold up-regulation; green indicates 4-fold down-regulation. All three HEBC cell lines exhibited substantial, dose-dependent changes in gene expression after 5-aza treatment. In contrast to the cancer cell lines, the HEBCs appeared to be relatively similar in their responses to 5-aza treatment. Bioinformatic analysis of these data appears in Table 6-5.

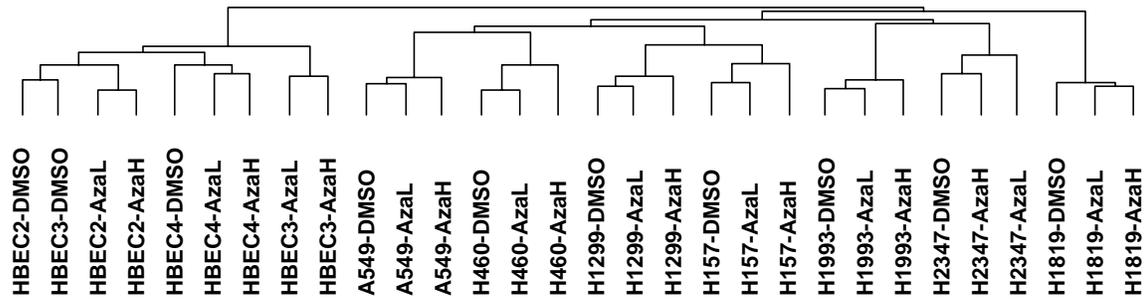


Figure 6-7. Hierarchical cluster analysis of the gene expression profiles of NSCLC and HBEC cell lines before and after 5-aza treatment. Microarray data were normalized using RMA and filtered using a minimum standard deviation of 20% across the genes. Clustering was performed using the average linkage method. While 5-aza induced substantial changes in gene expression in all cell lines with some overlap the NSCLC cell lines retained their expression phenotype differences. Interestingly, two untreated HBEC cell lines grouped together prior to treatment, but were quite different upon 5-aza treatment, suggesting that the gene expression profiles of these two cell lines is quite similar, but their underlying methylation may be quite different.

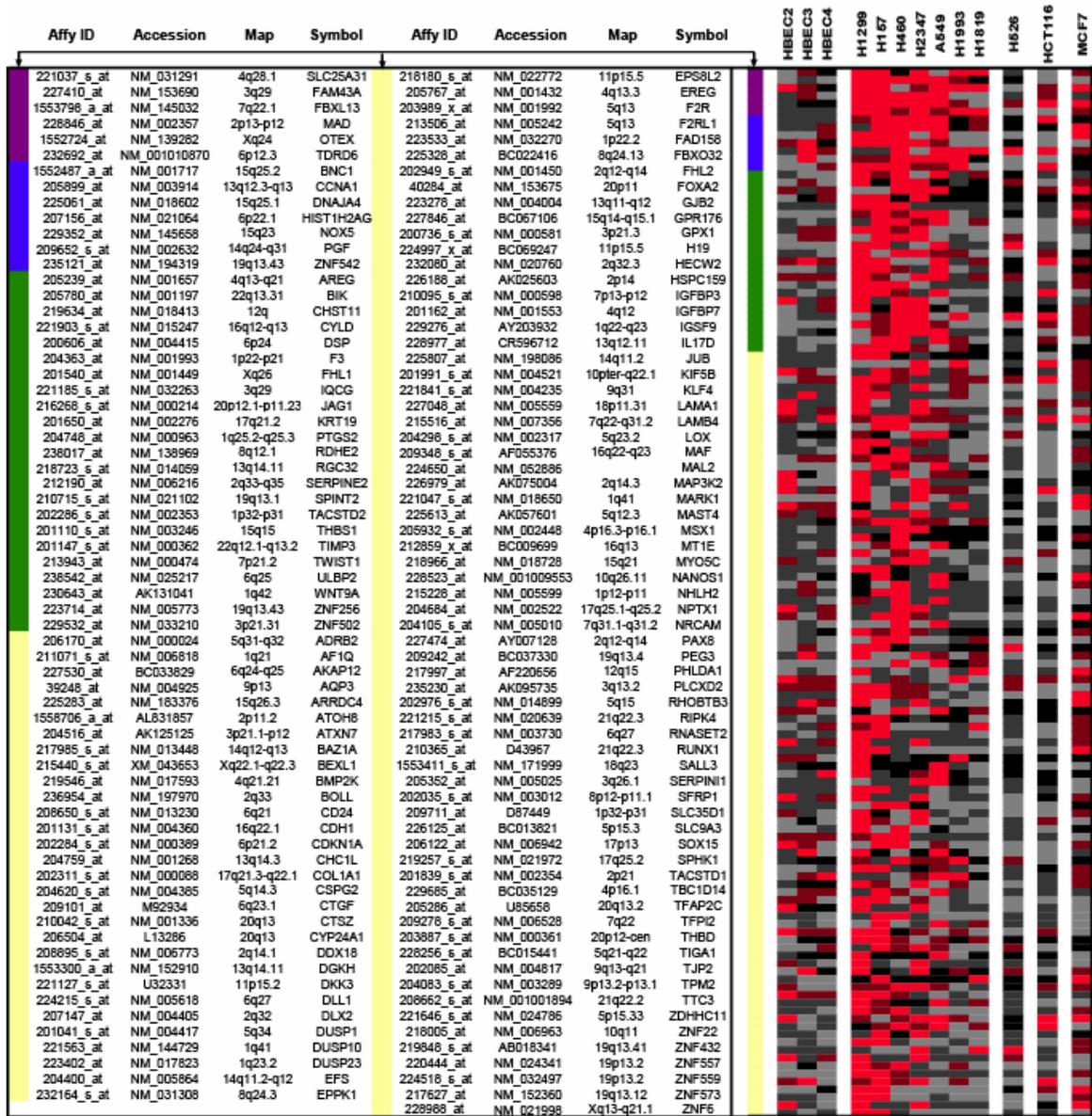


Figure 6-8. 5-aza-2'-deoxycytidine induced gene set in NSCLC, SCLC, breast, and colon cancer cell lines. Heat map for gene induction across NSCLC and other cancer cell lines as indicated. Data are \log_2 changes between mock-treated and 1 μM 5-aza treatment in each cell line. Bright red indicates ≥ 4 -fold up-regulation; intermediate red ≥ 2 -fold induction; grey indicates < 2 -fold induction; black indicates no data. The data are ordered from top to bottom according to the frequency of 4-fold induction across the NSCLC cancer cell lines. The vertical, colored bars parallel to the heat map represent the frequency of 4-fold induction in the NSCLC 5-aza induction experiments. Annotations are represented in order from top to bottom with the colored bars from the heat map indicating fold induction; purple indicates 5/7, blue indicates 4/7, green indicates 3/7, yellow indicates 2/7

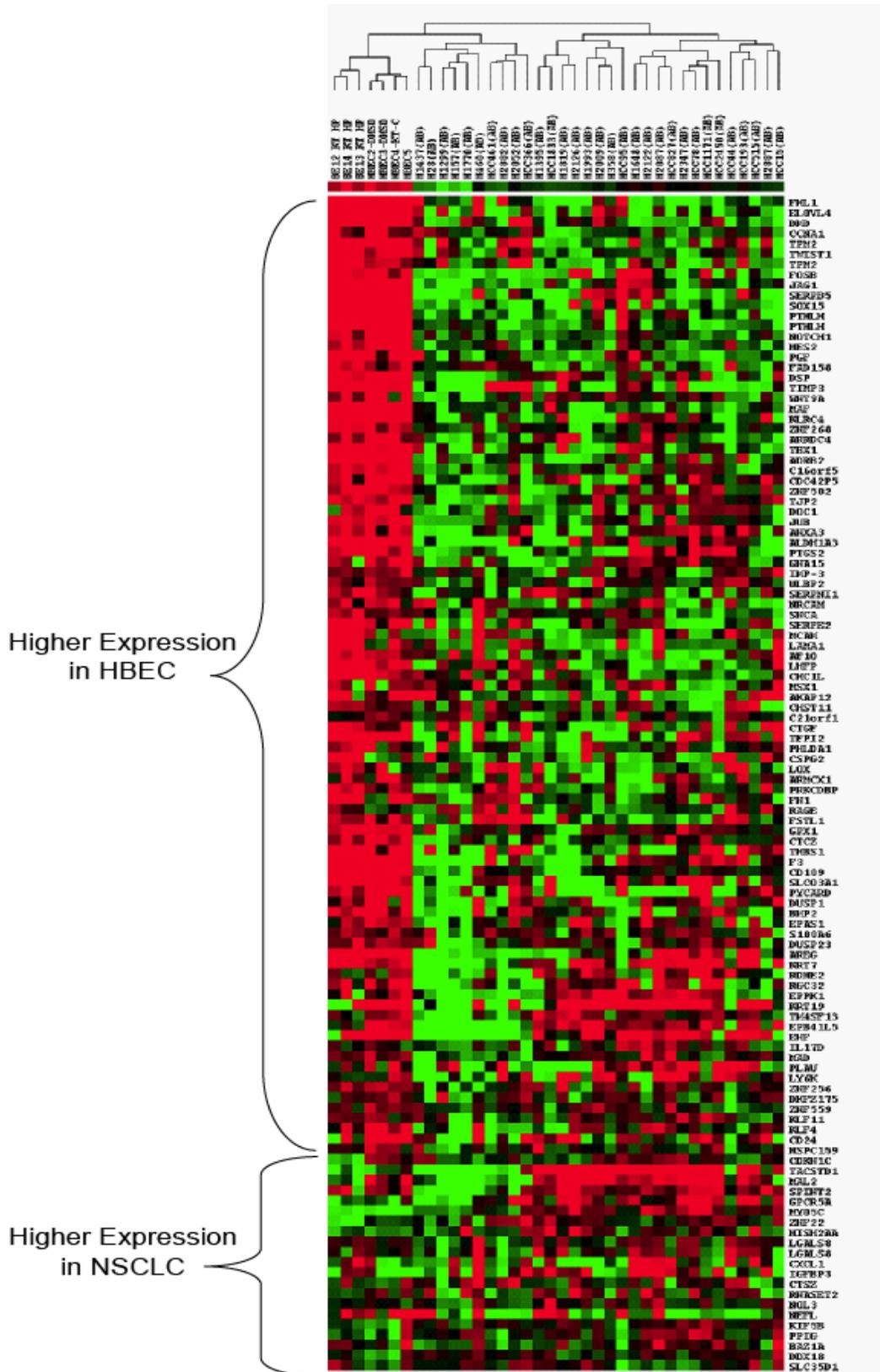


Figure 6-9. Average-linkage cluster analysis of 5-aza induced methylation candidates in independent microarray data sets (cell lines). The NSCLC microarray panel data were established on the Affymetrix U133 A and B chips. These arrays contain approximately 42,000 probes, all of which appear on the U133 Plus 2.0 chip. Data for the 132 gene 5-aza induction gene set, which was established from using the U133 Plus 2.0 chip, were extracted from the U133 A and B chips (120 genes) for the panel of 31 NSCLC cell lines (U133A and B) and 7 HBEC cell lines (U133 Plus 2.0). Data were then renormalized using the median array method, and transformed to \log_2 signal intensities, mean-centered and then clustered using the average linkage method. Red indicates above the mean, green indicates below the mean. 5-aza induction gene set separates cancer from HBEC cell lines and are highly expressed in HBEC. Some genes on the bottom of the figure (Higher in NSCLC) passed filtering because of the cell lines that were used in the study. In addition centering the data has the effect of forcing red and green colors on to the heat map when the differences may not be that great. This is an unfortunate artifact of the algorithm used in this analysis.

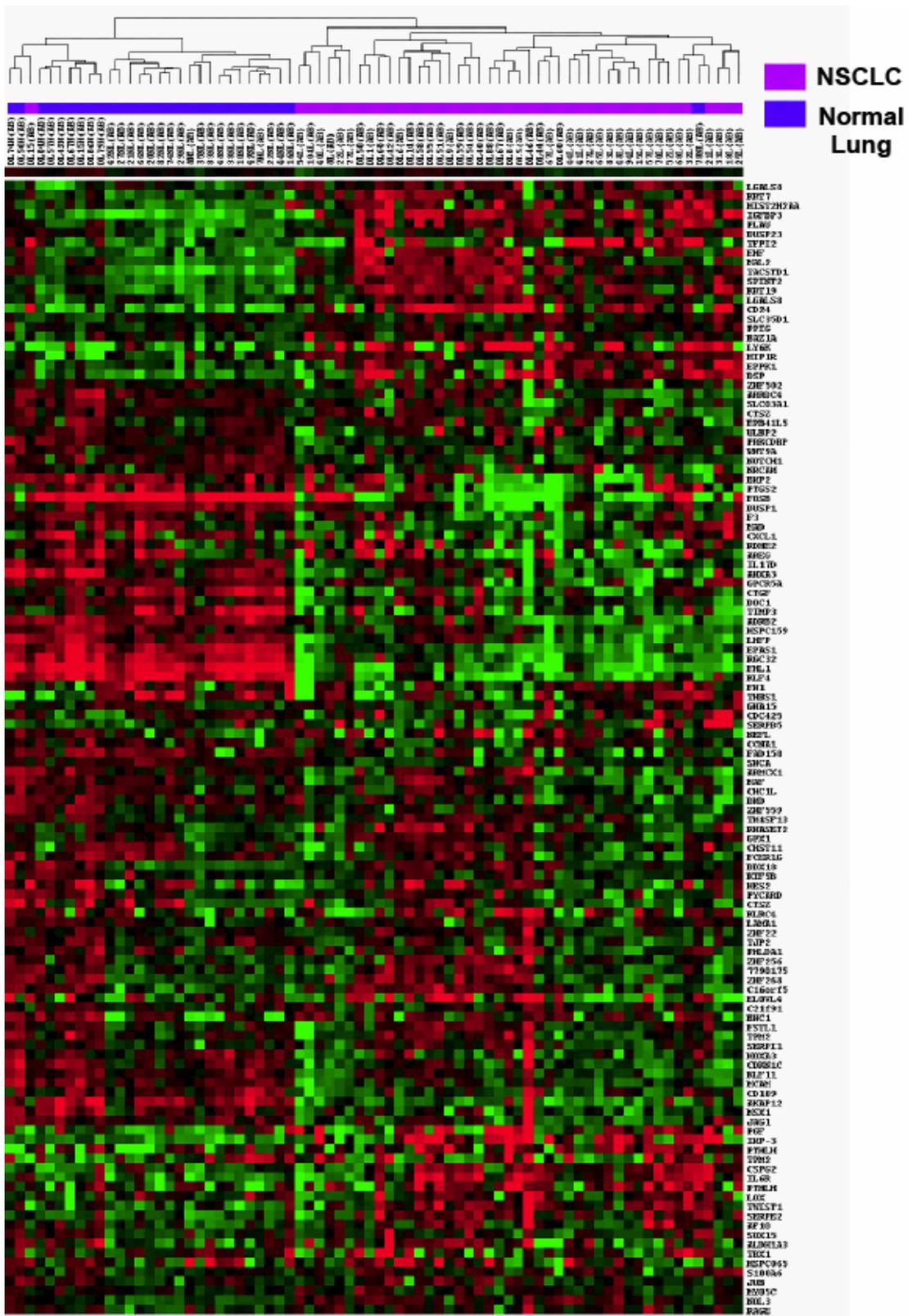


Figure 6-10. Average-linkage cluster analysis of 5-aza induced methylation candidates in normal lung and primary NSCLC. The NSCLC microarray panel data were established on the Affymetrix U133 A and B chips. Microarray expression data in a panel of 46 primary NSCLC and 29 counterpart normal lung tissues. Data are median-centered and colored as in Fig. 6-9. Blue bar indicates normal lung, purple bar indicates tumor tissue. 5-aza induced gene set clearly distinguishes cancer from normal. Most genes are expressed at higher levels in normal tissues, although not all.

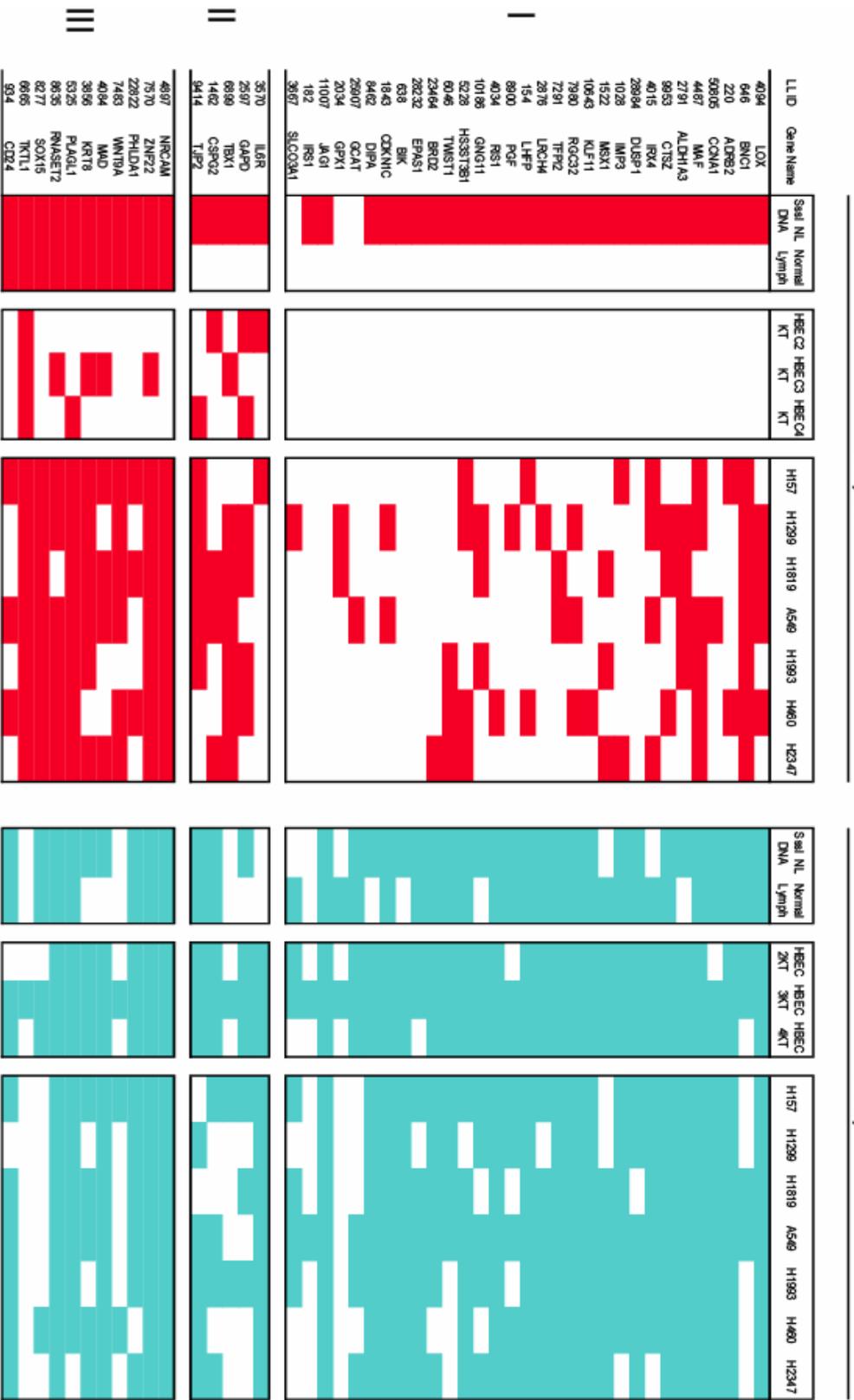


Figure 6-11. Summary of methylation-specific PCR in cell lines. From the left, *in vitro* methylated DNA mixed with lymphocyte DNA (SssI), normal lymphocyte DNA, HBEC cell lines, NSCLC lines, as indicated. Red fill indicates positive methylated product; aqua indicates positive unmethylated product. Data are grouped as follows: group I, no methylation in either HBECs or lymphocytes; group II, methylation in HBEC, but not lymphocytes; group III, methylation in lymphocytes. Data are ordered based on the frequency of methylation in primary lung tumors (Fig. 6-12)

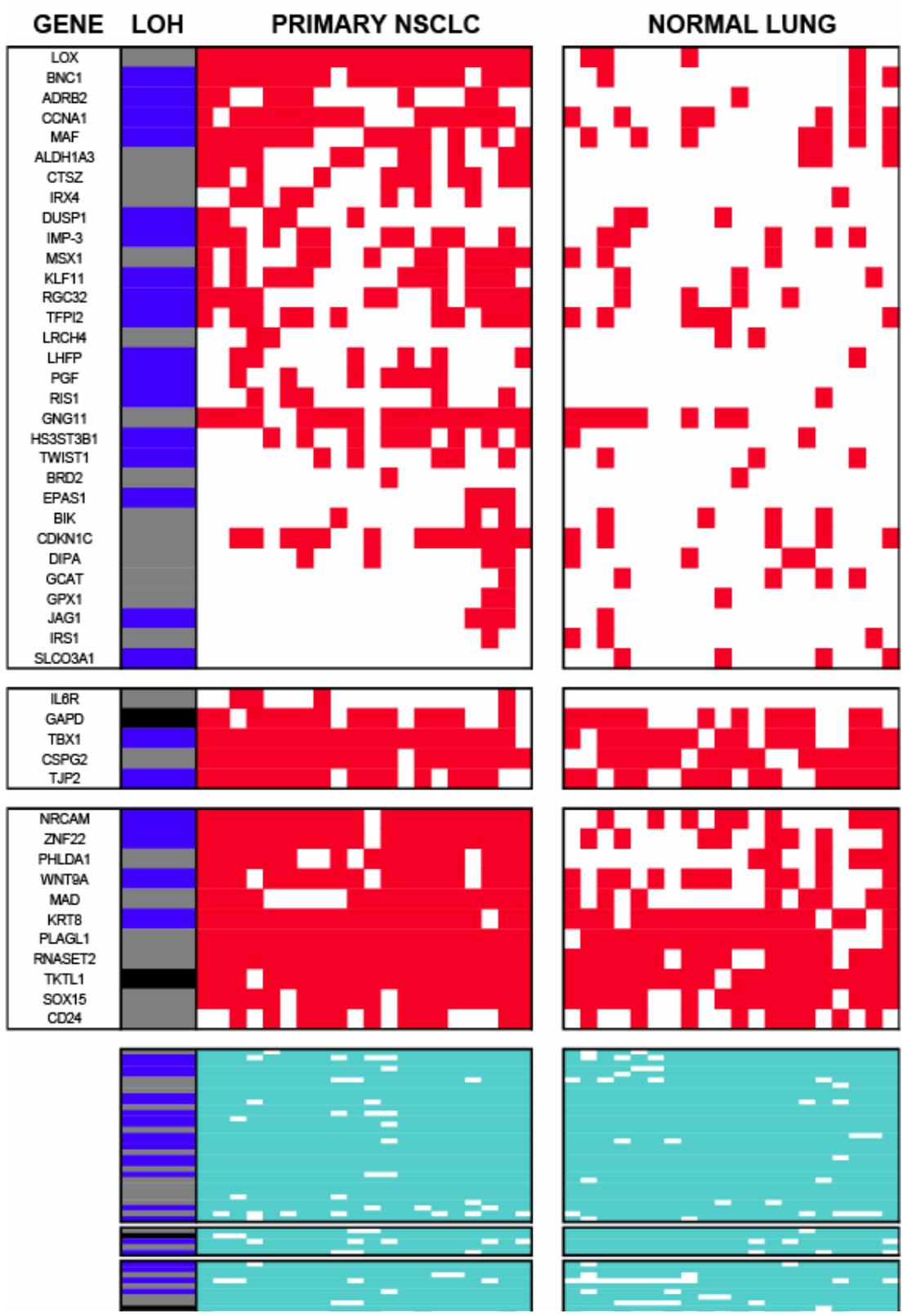
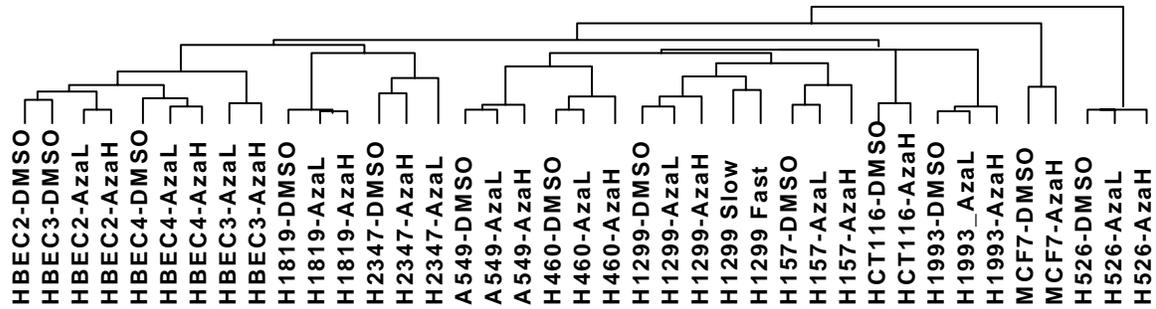


Figure 6-12. Summary of methylation-specific PCR in primary NSCLC and counterpart normal tissue. Data are organized as in Fig. 6-10. Each row represents one gene promoter, each column represents one sample. The samples are in the same order from left to right in both the methylated and unmethylated product panel. Summary of methylation specific PCR in matched primary NSCLC and adjacent non-malignant tissue. Data are colored and grouped as in (A), and ordered from top to bottom according to the frequency of methylation in primary NSCLC. Left column indicates net copy number changes across a panel of 31 NSCLC cell lines by aCGH.

A)



B)

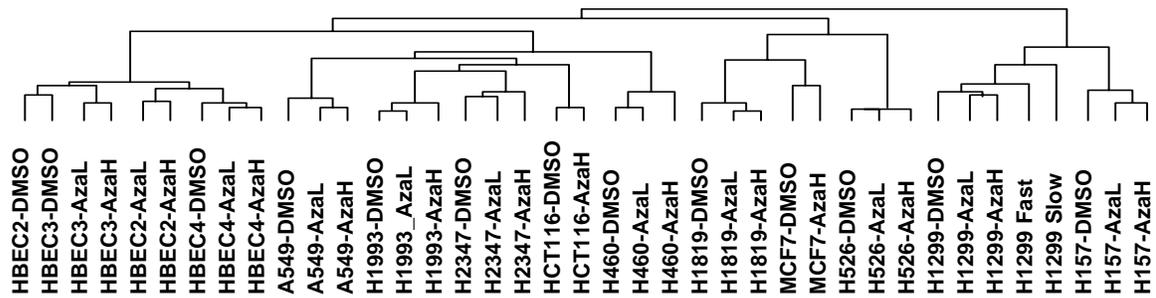


Figure 6-13. Hierarchical cluster analysis of the gene expression profiles of cell lines treated with 5-aza. A) Average linkage clustering of cell lines before and after treatment with 5-aza after SAM analysis. In this analysis, the breast and small cell lung cancer cell lines cluster apart the main grouping of cell lines, suggesting that tissue-specific methylation patterns predominate the expression phenotype of these cells. For these clusters, data was normalized using RMA and filtered to remove genes that were not differentially expressed across all cancer cell lines. MATRIX 1.29 was used to cluster the data. B) Average linkage cluster analysis of cell lines before and after treatment with 5-aza shows that the 5-aza induction gene set removes tissue-specific patterns of gene expression as indicated by the integrated grouping pattern. Data were extracted using MATRIX 1.29 (partial gene list) function.

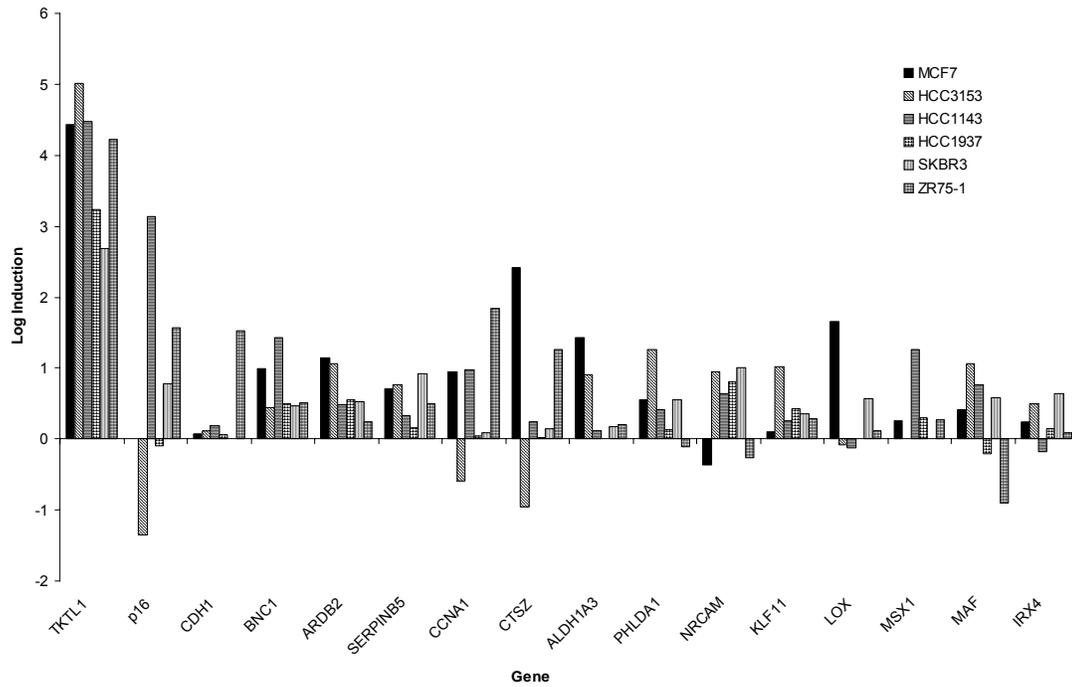
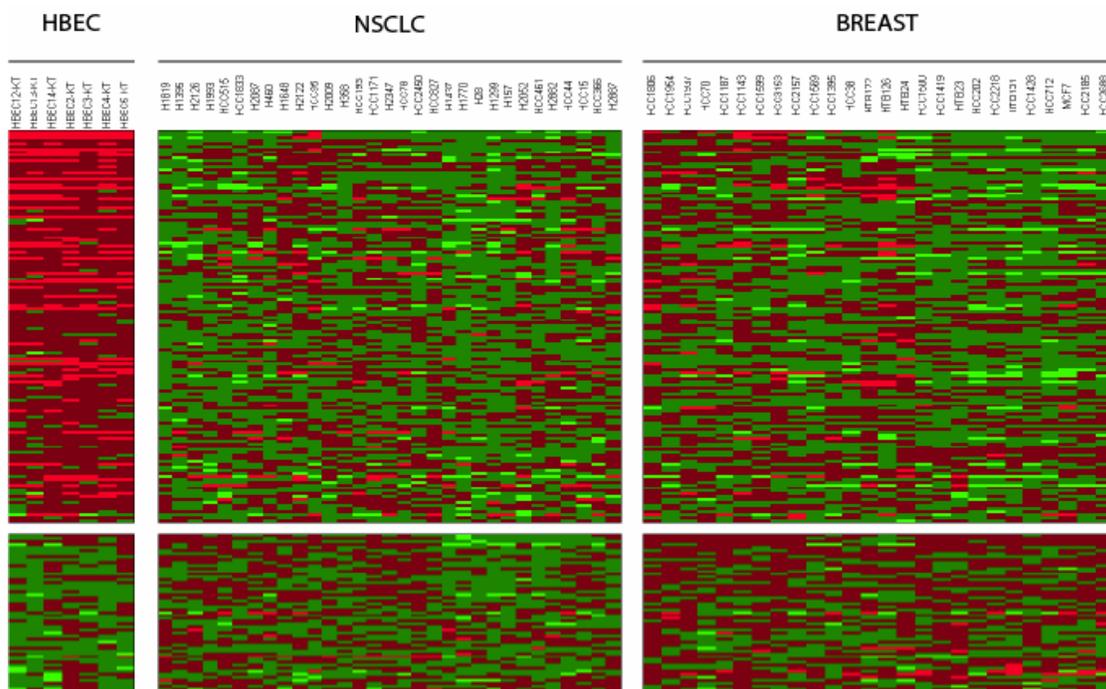


Figure 6-14. Quantitative RT-PCR for genes induced and methylated in NSCLC in six breast cancer cell lines before and after 5-aza treatment. Breast cancer cells were treated with 5-aza according to standard protocols. RNA was extracted using Trizol. cDNA was prepared using Superscript II and random primers. Probes and primers for QPCR were obtained from ABI. Induction is in \log_{10} . Genes induced by 5-aza treatment in NSCLC are also frequently induced in breast cancer cells.

A)



B)

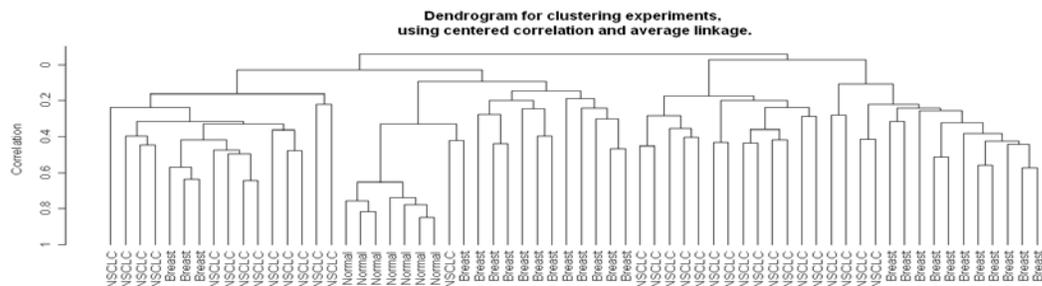


Figure 6-15. Comparison of the gene expression patterns for the 5-aza induced gene set in HBEC, NSCLC, and breast cancer cell lines. **A)** 120 genes from the original 132 were compared across all cell lines using RMA normalized data. While there is intra-tumor type variability, the overall pattern of gene expression between lung and breast cancers were not significantly different (p -value >0.24 , two-tailed t -test, unequal variances), but both were different from HBEC (6×10^{-147} for NSCLC vs. HBEC; 4×10^{-136} for Breast vs. HBEC). **B)** Cluster analysis shows that breast cancer and NSCLC cell lines overlap in their clustering patterns when analyzed using all 196 genes in the 5-aza induction panel, whereas both remain distinct from HBEC.

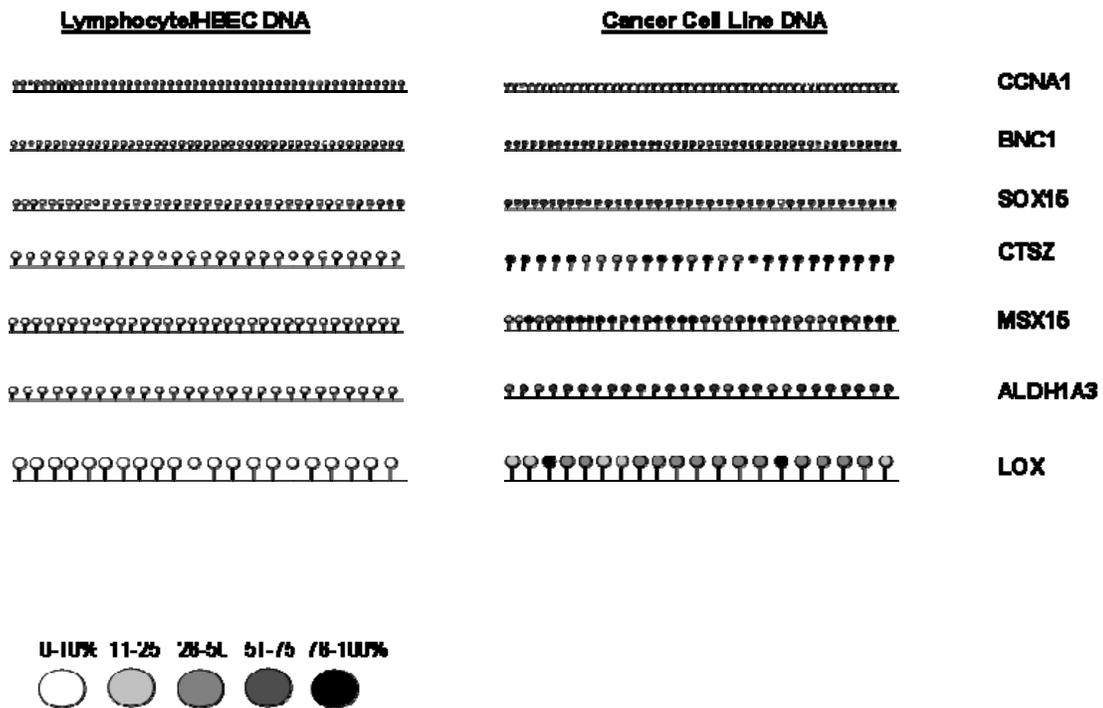


Figure 6-17. Summary of sodium bisulfite sequencing for seven genes in lymphocytes, HBEC, and NSCLC cells as indicated. Between 8 and 20 clones were sequenced for each locus in each cell type. Sequencing primers were designed to flank the MSP priming sites and do not include any CpG sites with the exception of *BNC1* which we were not able to amplify outside of the MSP priming sites for cells that harbored methylation. There was no amplification of the methylated primer set in HBEC or lymphocytes, and no amplification of the unmethylated primer set in the cell lines examined. On some occasions the methylated primer set for *BNC1* amplified a 289 bp amplicon from an unrelated locus on chromosome 1. The sequence corresponds to a CpG island in an intronless gene (*GPR25*) that was heavily methylated in tumors. The unmethylated primer set did not amplify this sequence. Each lollipop represents a composite of clones for that CpG site. Open lollipops indicate 0-10% methylation, light grey indicates 11-25%, middle grey indicates 26-50%, dark grey indicates 51-75%, black indicates 76-100% methylation. Primers and PCR conditions are available upon request. Raw data for sequencing is available online (PLoS Medicine).

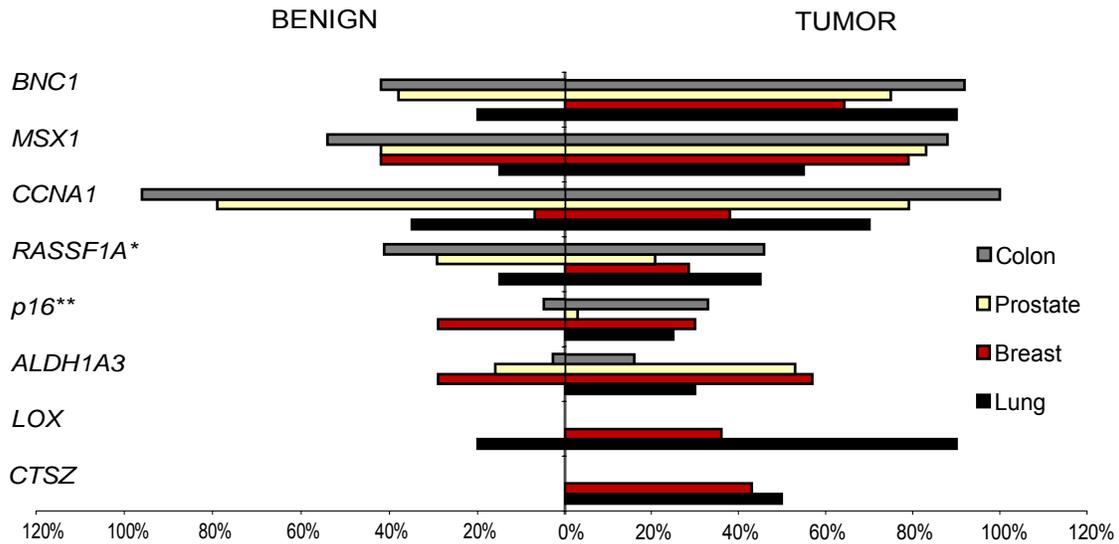


Figure 6-18. Summary of MSP data for indicated genes in breast (N=14; red bars), lung (N=20; black bars), prostate (N=24; pale yellow bars), colon (N=24; grey bars) tumors and benign tissue. Only samples with matching benign and tumor tissue are represented in the histogram. Gels were run and scored as in Fig. 6-17. *SOX15* was omitted from this figure for purposes of clarity. Data for RASSF1A was obtained from (Zochbauer-Muller, Fong et al. 2001; Maruyama, Toyooka et al. 2002; Lewis, Cler et al. 2005; Takahashi, Shigematsu et al. 2006); data for p16 was obtained from (Esteller, Fraga et al. 2001; Zochbauer-Muller, Fong et al. 2001; Holst, Nuovo et al. 2003; Takahashi, Shigematsu et al. 2006).

CHAPTER SEVEN

INTRODUCTION

In the following chapter, unpublished data are presented. The first section details the CpG island annotation method that was developed for analyzing the microarray experiments, as well as the results of some of these bioinformatic studies. The second section presents findings from a methylation analysis of the progressed HBEC system developed by Mitsuo Sato in the Minna lab. At this writing, these data were still preliminary.

CPG ISLANDS

As discussed in previous sections, CpG islands are short repetitive sequences that are highly conserved, contain relatively high levels of GC, and are enriched for the CpG dinucleotide. The importance of these sequences derives from their putative role in regulating transcription. As part of the analysis performed on the microarray data collected from the 5-aza induction experiments, I developed a strategy to examine whether there was any relationship between the overall gene density in the genome, the presence of Refseq genes that are associated with CpG islands, and whether there was detectable gene induction in multiple NSCLC lines in the 5-aza screen at a given position.

To examine this question we downloaded the CpG island annotation from the UCSC database and aligned that to the Affymetrix Gene chip (details of this approach are presented in materials and methods). A schematic of the strategy for this process is presented in Fig 7-1. The algorithm to align the 5' sites of transcripts was written by Luc Girard in the Minna lab, but I edited and optimized it. The general strategy was to take the most 5' positional information from a probe annotation track in Affymetrix database and align that with the CpG island positional information using a Microsoft Access database (Fig 7-2). While this strategy is relatively straightforward in concept, we had to resolve several difficult technical issues.

The primary complication was that the assembly scheme used by Affymetrix in building their U133 Genechip was based on an incomplete genome sequence database (at the time); many probes are based on partial sequence information such as ESTs and incomplete cDNAs – for details see Affymetrix probe set design white papers

(<http://www.affymetrix.com/support/technical/whitepapers.affx>). Each EST/cDNA/partial sequence, etc., has its own accession and Affymetrix used many that were partially redundant or had the incorrect annotation (Benson, Karsch-Mizrachi et al. 2005)(Luc Girard, unpublished information). Moreover, redundant alignments that comprise full RefSeq annotations are usually biased against the 5' end of gene. This led to many instances where the same gene had multiple alignments, some of which did not agree (Fig. 7-3).

Despite this limitation in the underlying data and the alignment method, an overview of CpG island positioning in the genome was undertaken. The number of DNA sequences scored as CpG islands in the genome depends critically on the parameters used to define these sequences. The standard definition is based on the work of Garden and Frommer, and defines CpG islands as short stretches of DNA with higher than expected GC content (>50%) where the CpG dinucleotide content is relatively enriched with an observed vs. expected ratio of >0.6, over a distance of at least 200 base pairs (Gardiner-Garden and Frommer 1987). More recently other groups have used modern bioinformatic techniques and the full sequence of the human genome have developed a new definition where a sequence has a minimum of 500 base pairs, a GC content of $\geq 55\%$, and an observed vs. expected CpG ratio of ≥ 0.65 (Takai and Jones 2002). The major limitation of the older definition is that it includes many intergenic CpG rich areas such as those associated with long terminal repeats (LTRs), *Alus*, and other repetitive elements. According to the newer definition approximately 40% of human genes are associated with these elements (Bestor, Gundersen et al. 1992; Takai and Jones 2002). Table 7-1 shows the relationship between CpG island length and frequency in the genome.

The purpose of the above analysis was to determine whether there was a relationship between the location and density of genes within a given locus and the density of gene induction our 5-aza profiles. To analyze whether there was a relationship between the frequency of gene induction and gene density, the positional information for the RefSeq annotation database and the CpG island database was intersected using the UCSC genome table browser. Then we overlaid this positional information with the 5-aza induction gene set that intersected with RefSeq database. This union set included approximately 480 genes (see previous chapter). The data for this analysis appears in the Fig. 7-5. These data show that for the most part, the location of genes that were induced by 5-aza essentially followed the gene density in the genome (genes with CpG islands). There were some exceptions however including chromosomes 4, 12, and 16 where there were regions of localized gene density that were not represented in the 5-aza gene induction set. In addition, there were areas of enriched 5-aza gene induction (e.g. chromosome 9) that did not seem particularly enriched for CpG island associated genes.

Taken together the gene density vs. gene induction density plots indicate that the 5-aza gene induction set was distributed across the genome, and identifies localized regions of tumor suppressor loci for future study. One particular region of interest lies in 9p21, which was highly enriched for 5-aza dependent gene induction and also is a LOH hot spot region contains the tumor suppressor locus *p16*. These data form a part of a planned future study (see next chapter).

**ANALYSIS OF PROMOTER HYPERMETHYLATION IN ONCOGENICALLY
PROGRESSED IMMORTALIZED HUMAN BRONCHIAL EPITHELIAL CELLS**

INTRODUCTION – SUMMARY OF SATO ET AL

The following section describes data from an ongoing project that is part of a collaboration between the Adi Gazdar, Narayan Shivapakur, Mitsuo Sato, and myself. The basis for this project derives from (Sato, Vaughan et al. 2006), where Mitsuo Sato generated a series of isogenic HBEC cell lines containing a variety of oncogenic changes (described below). These HBECs have differential tumorigenic properties whereupon some grow in soft agar and some form tumors in mice. We hypothesized that alterations in cellular phenotype, particularly toward cellular transformation may coincide with alterations in promoter hypermethylation. The results of these experiments are presented below.

The following abstract from a paper I co-authored is transposed directly from (Sato, Vaughan et al. 2006).

ABSTRACT

We evaluated the contribution of three genetic alterations (p53 knockdown, K-RASV12, and mutant EGFR) to lung tumorigenesis using human bronchial epithelial cells (HBEC) immortalized with telomerase and Cdk4-mediated p16 bypass. RNA interference p53 knockdown or oncogenic K-RASV12 resulted in enhanced anchorage-independent growth and increased saturation density

of HBECs. The combination of p53 knockdown and K-RASV12 further enhanced the tumorigenic phenotype with increased growth in soft agar and an invasive phenotype in three-dimensional organotypic cultures but failed to cause HBECs to form tumors in nude mice. Growth of HBECs was highly dependent on epidermal growth factor (EGF) and completely inhibited by EGF receptor (EGFR) tyrosine kinase inhibitors, which induced G1 arrest. Introduction of EGFR mutations E746-A750 del and L858R progressed HBECs toward malignancy as measured by soft agar growth, including EGF-independent growth, but failed to induce tumor formation. Mutant EGFRs were associated with higher levels of phospho-Akt, phospho-signal transducers and activators of transcription 3 (but not phospho-extracellular signal-regulated kinase (ERK) 1/2), and increased expression of DUSP6/MKP-3 phosphatase (an inhibitor of phospho-ERK1/2). These results indicate that (a) the HBEC model system is a powerful new approach to assess the contribution of individual and combinations of genetic alterations to lung cancer pathogenesis; (b) a combination of four genetic alterations, including human telomerase reverse transcriptase overexpression, bypass of p16/RB and p53 pathways, and mutant K-RASV12 or mutant EGFR, is still not sufficient for HBECs to completely transform to cancer; and (c) EGFR tyrosine kinase inhibitors inhibit the growth of preneoplastic HBEC cells, suggesting their potential for chemoprevention.

METHYLATION ANALYSIS OF HBECS

The data for these studies are presented in Table 7-2 and Fig 7-6. As mentioned above, the goal of these studies is to identify methylated loci that correlate with the acquisition of specific oncogenic changes (loss of p53 and gain of oncogenic p53). Narayan Shivapurkar performed quantitative methylation analysis using Taqman assays (see methods). I performed standard MSP using primers for 45/132 genes in the 5-aza induction gene set (*Chapter 6*).

In these studies, we found that methylation has occurred in the progressed HBECS, but it is inconsistent. In particular, several subclones of the HBEC3R_L53 were examined and compared to the parental lines. There were inconsistencies between the different subclones in terms of which genes were methylated and the combination of all subclones did not reconstitute the parental methylation profile. The C1 subclone was shown to be reproducibly tumorigenic in nude mice, but these cells showed different methylation profiles before going into mice compared to afterwards. There are biological explanations for these findings however more experiments are needed to interpret these data with confidence. As stated above, these are preliminary findings and will be pursued after this writing.

Tables

Table 7-1. Percentage of transcripts associated with CpG Islands on the Affy 2.0 chip

Min. length of CpG Island	Window size					No. of CpG Islands
	+/- 500	+ 500	+/- 250	+250	0	
200	45.9%	37.6%	37.8%	32.8%	26.5%	568326
300	35.3%	30.1%	30.3%	27.3%	23.8%	287917
400	29.3%	25.9%	26.0%	24.1%	21.9%	149444
500	26.7%	24.1%	24.0%	22.6%	20.9%	96774

Table 7-1. Effect of different parameters on the number of CpG islands in the genome. Window size means the number of bases flanking the annotated “start” site for a given probe set in the Affymetrix alignment scheme. While many of these were wrong, they should affect each parameter change equally and so can be ignored.

Official name	Sample #	Vector 1	Vector 2	Vector 3	Form Tumors in Nude Mice?	Soft Agar Colony Formation?
HBEC3-KT	1				(-)	(+)
HBEC3-KTR _L (2)	2	pSRZ	plenti-KRAS		ND	(+)
HBEC3-KTR _L 53	3	p53pSRZ	plenti-KRAS		(-)	(+)
HBEC3-KTR _L 53-s1	4	p53pSRZ	plenti-KRAS		(+)	(+)
HBEC3-KT53	5	p53pSRZ			(-)	(+)
HBEC3-KTR _B	6		pBabe-KRAS		(-)	(+)
HBEC3-KTR _B 53	7	p53pSRZ	pBabe-KRAS		(-)	(+)
HBEC3-KTG	8	pSRZ	plenti-wt-EGFR		ND	(+)
HBEC3-KTGd	9	pSRZ	plenti-del-EGFR		ND	(+)
HBEC3-KTGm	10	pSRZ	plenti-L858R-EGFR		ND	(+)
HBEC3-KT53G	11	p53pSRZ	plenti-wt-EGFR		ND	(+)
HBEC3-KT53Gd	12	p53pSRZ	plenti-del-EGFR		ND	(+)
HBEC3-KT53Gm	13	p53pSRZ	plenti-L858R-EGFR		(-)	(+)
HBEC3-KTR _L 53M22	14	p53pSRZ	plenti-KRAS	C-MYC-C22-pMSCV	(-)	(+)
HBEC3-KTR _L 53B	15	p53pSRZ	plenti-KRAS	Bcl2-pMSCV	ND	ND
HBEC3-KTR _L 53t	16	p53pSRZ	plenti-KRAS	st-pMSCV	(+)	ND
HBEC3-KTB	17	Bcl2-pMSCV			ND	(+)
HBEC3-KTP	18	PTEN-pSRB			ND	(+)

These data were obtained from Mitsuo Sato. He performed all of the transformations and tumorigenicity assays. I was not directly involved in any of the experiments that support the above data. They are included in this thesis for informational purposes only.

Table 7-2. Tumorigenic properties of isogenic HBEC3 and nomenclature. Several clones were tested for tumorigenicity in nude mice and most did not form tumors; see (Sato, Vaughan et al. 2006) for details. In addition, several clones were derived from a soft agar screen. Importantly, two clones from this screen were tumorigenic, and one was reproducibly so. Clone 1 will be the focus of future studies. **Nomenclature:** K = *cdk4*; T = *hTERT*; R = *KRAS* (V12); R_L = *KRAS* (V12) (high titer) lentiviral vector; R_B = *KRAS* (V12) in a retroviral vector (low titer); 53 = *p53* knockdown with shRNA; G = *EGFR*; G_d = *EGFR* deletion mutant (activated); G_m = *EGFR* missense mutant; M22 = *c-MYC*; B = *BCL2*; P = *PTEN* shRNA; t = Small T antigen from SV40.

HBEC	Minna Name	Modifications	Media	3-OST-2	Cytlb	GRP38	RASSF1A	p16	E-CAD	DcR1	DcR2	SOCS3	APC	MYOD1
BE3	HBEC2-KTR-pSRZ_pBABE	cdk4-hTERT-v12Kras-control viruses for BE6	KSFM	0	0	0	0	0	0	0	0	0	0	39.95
BE4	HBEC2-KT53-pSRZ_pBABE	cdk4-hTERT-p53shRNA-control viruses for BE6	KSFM	0	0	0	0	0	0	0	0	0	0	40.27
BE5	HBEC2-pSRZ_pBABE	cdk4-hTERT-control viruses for HBEC2 series	KSFM	0	0	0	0	0	0	0	0	0	0	39.66
BE6	HBEC2-P53, KRAS	cdk4-hTERT-p53shRNA-v12Kras	KSFM	0	0	0	0	0	0	0	0	0	0	41.01
BE7	HBEC3-KTG	cdk4-hTERT-wt EGFR	KSFM	0*	0*	0	0	0	0	0	0	0	0	38.91
BE8	HBEC3-KTGd	cdk4-hTERT-EGFR del. Mutant	KSFM	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
BE9	HBEC3-KTGm	cdk4-hTERT-missense EGFR	KSFM	0*	0*	0	0	0	0	0	0	0	0	38.93
BE10	HBEC3-KTp53-LACZ	cdk4-hTERT-p53shRNA-control virus for BE11-13	KSFM	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
BE11	HBEC3-KT53G	cdk4-hTERT-p53shRNA-wt EGFR	KSFM	0	0	0	0	0	0	0	0	0	0	38.62
BE12	HBEC3-KT53Gd	cdk4-hTERT-p53shRNA-del. EGFR	KSFM	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
BE13	HBEC3-KT53Gm	cdk4-hTERT-p53shRNA-missense EGFR	KSFM	0	0	0	0	0	0	0	0	0	0	40.30
BE14	HBEC3-KTRb53	cdk4-hTERT-v12Kras (low titer - retrovirus)-p53shRNA	KSFM	0.22	0.65	1.05	0	0	0	0	0	0	0	31.13
BE15	HBEC3-KTRI53	cdk4-hTERT-v12Kras (high-titer lentivirus)-p53shRNA	KSFM	0*	1.48	0.67	0	0	0	0	0	0	0	36.27
BE16	HBEC3-KTRI(2)	cdk4-hTERT-lentiviral v12Kras	KSFM	0	0	0	0	0	0	0	0	0	0	38.85
BE17	HBEC3-KTRI53B	cdk4-hTERT-lentiviral v12Kras-p53shRNA-Bcl2shRNA	KSFM	0	0	0	0	0	0	0	0	0	0	41.26
BE18	HBEC3-KTRI53M22	cdk4-hTERT-lentiviral v12Kras-p53shRNA-c-MYC	KSFM	0	0	0	0	0	0	0	0	0	0	41.00
BE19	HBEC3-KTRI53t	cdk4-hTERT-lentiviral v12Kras-p53shRNA-SV40 small t	KSFM	0	0	0	0	0	0	0	0	0	0	39.63
BE20	KTRI53 C.2	cdk4-hTERT-v12Kras (high-titer lentivirus)-p53shRNA	KSFM	0	0	ND	0	ND	ND	ND	ND	ND	ND	ND
BE21	KTRI53 C.3	cdk4-hTERT-v12Kras (high-titer lentivirus)-p53shRNA	KSFM	0	4.9	ND	0	ND	ND	ND	ND	ND	ND	ND
BE22	KTRI53M22 PARENTAL IN R10	cdk4-hTERT-lentiviral v12Kras-p53shRNA-c-MYC	R10	0	0	ND	0	ND	ND	ND	ND	ND	ND	ND
BE23	KTRI53 C.6	cdk4-hTERT-v12Kras (high-titer lentivirus)-p53shRNA	KSFM	0	0.27	ND	0	ND	ND	ND	ND	ND	ND	ND
BE24	KTRI53 C.5	cdk4-hTERT-v12Kras (high-titer lentivirus)-p53shRNA	KSFM	0	+/-	ND	0	ND	ND	ND	ND	ND	ND	ND
BE25	KTRI53 C.4	cdk4-hTERT-v12Kras (high-titer lentivirus)-p53shRNA	KSFM	0	0	ND	6.76	ND	ND	ND	ND	ND	ND	ND
BE26	KTRI53 PARENTAL IN R10	cdk4-hTERT-v12Kras (high-titer lentivirus)-p53shRNA	R10	0	0	ND	0	ND	ND	ND	ND	ND	ND	ND
BE27	KTRI53 C.5 explant cell line (mouse #453)	cdk4-hTERT-v12Kras (high-titer lentivirus)-p53shRNA	R10	0	0	ND	0	ND	ND	ND	ND	ND	ND	ND
BE28	KTRI53 C.1 IN R10	cdk4-hTERT-v12Kras (high-titer lentivirus)-p53shRNA	R10	0	4.17	ND	0	ND	ND	ND	ND	ND	ND	ND
BE29	KTRI53 C.1	cdk4-hTERT-v12Kras (high-titer lentivirus)-p53shRNA	KSFM	0.72	+/-	ND	0	ND	ND	ND	ND	ND	ND	ND
BE30	KTRI53 C.5 IN R10	cdk4-hTERT-v12Kras (high-titer lentivirus)-p53shRNA	R10	0	0.68	ND	0	ND	ND	ND	ND	ND	ND	ND
BE31	KTRI53M22 EXPLANT CELL LINE FROM MOUSE 524	cdk4-hTERT-lentiviral v12Kras-p53shRNA-c-MYC	R10	1.41	11.22	ND	0	ND	ND	ND	ND	ND	ND	ND

Mitsuo Sato was responsible for creating the cell lines involved in this study. The author isolated DNA and performed the bisulfite treatments. Narayan Shivapurkar performed the quantitative analysis presented here.

Table 7-3. Quantitative methylation data for isogenic HBEC2 and HBEC3 series using Taqman assays. Mostly the HBECs are unmethylated, but in certain cases, there is detectable methylation at some loci. The left most column indicates the sample names that correspond to the genetic backgrounds indicated in columns 2 and 3. These sample names also correspond to Fig. 7-6, below. See previous page for nomenclature.

Figures

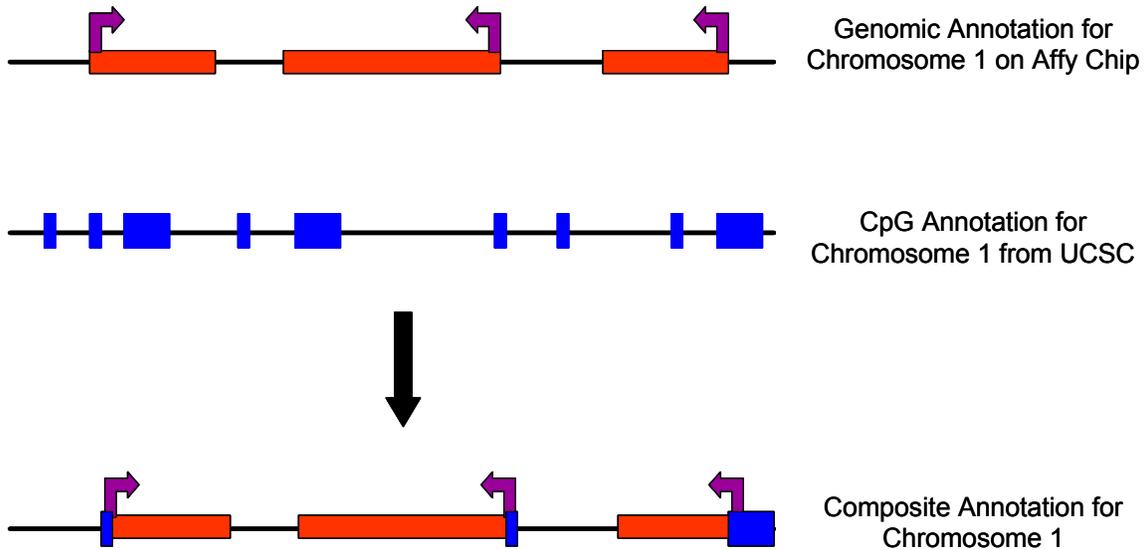


Figure 7-1. Schematic for CpG island/Affymetrix Genechip alignment. Data was obtained through the UCSC genome browser and input into a Microsoft Access database. Chromosomal position information was used to align the databases using an excel macro. Purple arrows indicate putative (RefSeq annotated) transcription start sites; red indicates openreading frames (RefSeq annotated); and blue indicates CpG islands as determined by Gardiner-Garden (Gardiner-Garden and Frommer 1987).

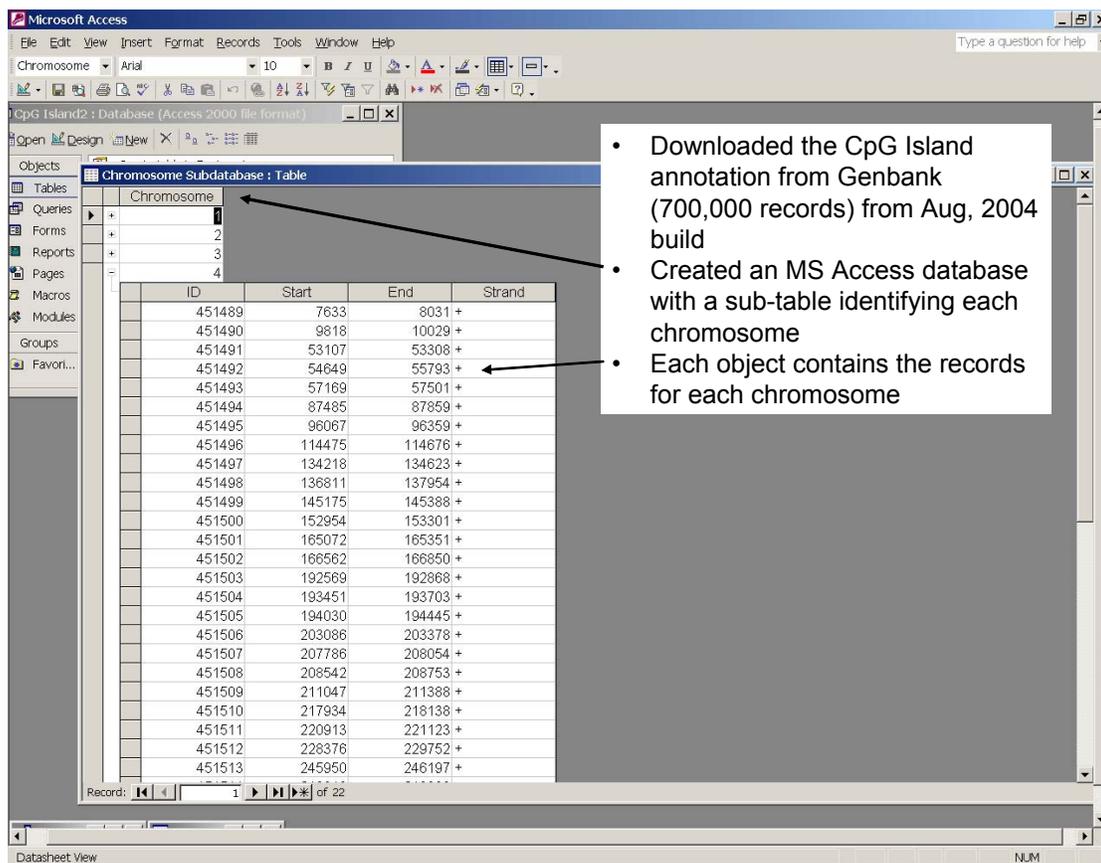


Figure 7-2. Microsoft Access database containing the CpG island annotation from UCSC. The structure of the database was based on position such that each sub-table contained positional (start and end) information for each CpG island was grouped according to chromosome. Each CpG island has its own accession number which can be seen above in the column titled "ID". This database was used to create 8</ref-number><ref-type name='Journal Article'>17</ref-type><contributors><authors><author>Lewis, C. M.</author><author>Cler, L. R.</author></contributors> to look at the effect of changing the parameters defining a CpG island and their association with protein coding genes (see below).

	K	L	M	P	Q	R	U	V	W	X	Y	Z
	Start	End	UCSC Cytoband	GO: Molecular Function	GO: Biological Process	GO: Cellular Component	Symbol		CpG Yes or No			
1												
2	113691170	113752718	2q13	ATP binding	CTP biosynthesis	nucleus	PAX8		Yes			
3	241975325	242013429	2q37.3	GTP binding	cell cycle	contractile ring	NEDD5		No			
4	55371474	55374396	2p16.1	structural constituent of ribosome	protein biosynthesis	intracellular	RPS27A		No			
5	96272478	96296437	2q11.2				STARD7		No			
6	68180987	68201781	2p14	DNA binding		nucleus	CTD		No			
7	96362348	96369525	2p11.2	ATP binding	RNA splicing	spliceosome complex	U5-200KD		No			
8	219862123	219867143	2q35				C2orf24		No			
9	203871184	203881548	2q33.2						No			
10	232145123	232154680	2q37.1	DNA binding		nucleolus	NCL		Yes			
11	241888133	241975225	2q37.3	RNA binding	cholesterol metabolism	nucleus	HDLBP		No			
12	54793855	54796834	2p16.2	actin binding		cytoskeleton	SPTBN1		No			
13	54665157	54807513	2p16.2	actin binding		cytoskeleton	SPTBN1		Yes			
14	20154040	20173073	2p24.1		transport	integral to membrane	LAPTM4A		No			
15	9674703	9721704	2p25.1	protein domain specific	exocytosis		YWHAQ		Yes			
16	206849974	206853157	2q33.3	translation elongation factor	protein biosynthesis	eukaryotic translation	EEF1B2		No			
17	65400048	65407882	2p14				ACTR2		No			
18	65366628	65408946	2p14				ACTR2		Yes			
19	85677961	85684059	2p11.2	ATP binding	one-carbon compound metabolism		MAT2A		No			
20	85677945	85684059	2p11.2	ATP binding	one-carbon compound metabolism		MAT2A		No			
21	232398763	232403220	2q37.1						Yes			
22	232398740	232403754	2q37.1		development	nucleus	PTMA		Yes			
23	201508264	201512806	2q33.1						Yes			
24	201502152	201514065	2q33.1	translation initiation	regulation of translational initiation		BZW1		Yes			
25	242009475	242011956	2q37.3	GTP binding	cell cycle	contractile ring	NEDD5		Yes			
26	10531105	10539051	2p25.1	lyase activity	polyamine biosynthesis	cellular component	ODC1		No			
27	198179296	198184983	2q33.1						No			
28	198176815	198190146	2q33.1	ATP binding	mitochondrial matrix	mitochondrial matrix	HSPD1		No			
29	73373105	73391787	2p13.2	ATP binding	protein folding	cytoplasm	CCT7		Yes			
30	62006913	62027442	2p15	ATP binding	protein folding	cytoplasm	CCT4		No			
31	46524244	46525486	2p21	RNA polymerase II transcription	angiogenesis	nucleus	EPAS1		Yes			
32	46436213	46525486	2p21	RNA polymerase II transcription	angiogenesis	nucleus	EPAS1		No			
33	104659281	104704442	2q37.3	hematopoietic factor	Leukocyte lineage	cytoplasm	STAT1		Yes			

Figure 7-3. Results of an alignment between the CpG island annotation and Affymetrix annotation for chromosome 2. The data above shows an example of an alignment between the CpG annotations for chromosome 2 and the Affymetrix annotation for chromosome 2. The red boxes indicate disagreements between the resultant output from the program. The red underlining shows that the “start” coordinates for these annotations. The two annotations correspond to the same gene but the source sequences are different.

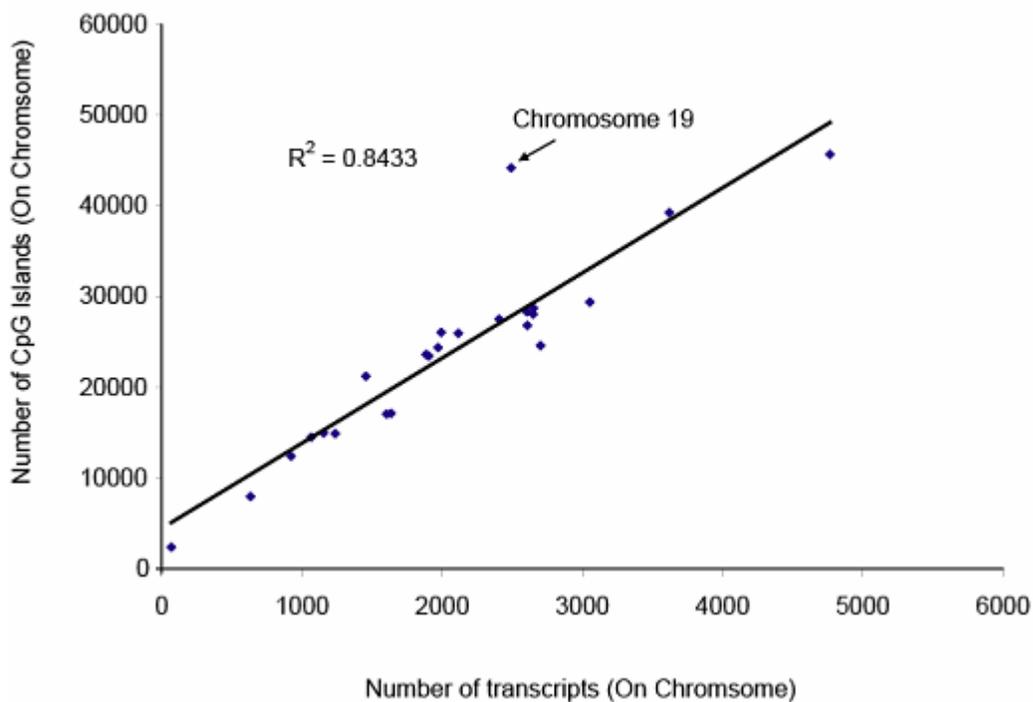


Figure 7-4. Scatter plot showing the number of transcripts per chromosome versus the number of CpG islands per chromosome. CpG island annotations were downloaded from the UCSC genome database and aligned to the Affymetrix probe alignments using a Visual Basic macro. There is no functional definition for a CpG island and the total number in the genome depends heavily on the criteria used: see table 7-1. For the scatter above, the standard Gardiner-Garden definition (length >200 bps; GC content >55%; CpG >0.6) was used and the number of transcripts was based on the Affy Gene chip annotation. Chromosome 19 has an unusually large number of CpG islands relative to the rest of the genome. This did not result in a larger than expected number of genes from chromosome 19 in the ultimate 5-aza induced gene set.

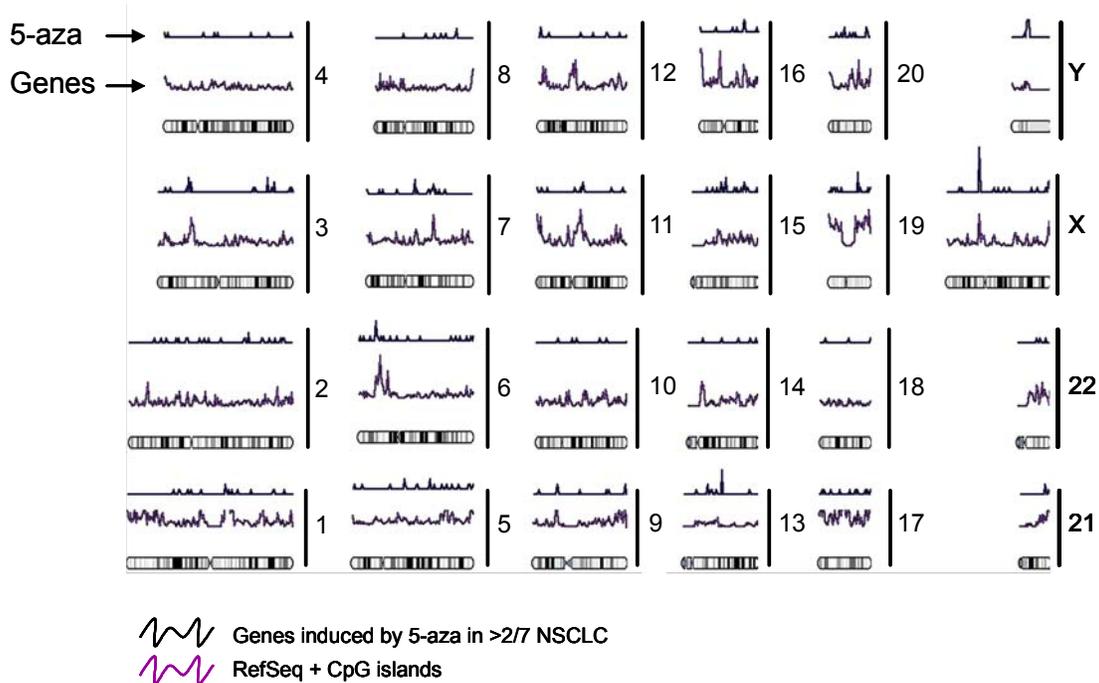


Figure 7-5. Karyogram of genes induced by 5-aza in 2/7 NSCLC compared to gene density. Data for each chromosome is shown against a representation of the expected geimsa cytobanding patterns. The middle, purple line plot represents the positional and frequency data for CpG island associated genes based on the current RefSeq database. Frequency (density) is indicated in the vertical direction where all all heights are relative within an annotation series – i.e. densities are comparable between chromosomes for “RefSeq + CpG” island or “Genes induced by 5-aza”. The black line plot represents the gene frequency data for genes that were induced >4 fold in more than 2/7 NSCLC but already expressed in HBEC (see *Chapter 6*). For the most part, the locations of induced genes followed the gene density in the genome, although not always. For example, chromosomes 8, 12, and 16 have regions of high density that where not represented on the induced gene list, whereas chromosome 9 has a region of frequent induction but relatively lower gene density.

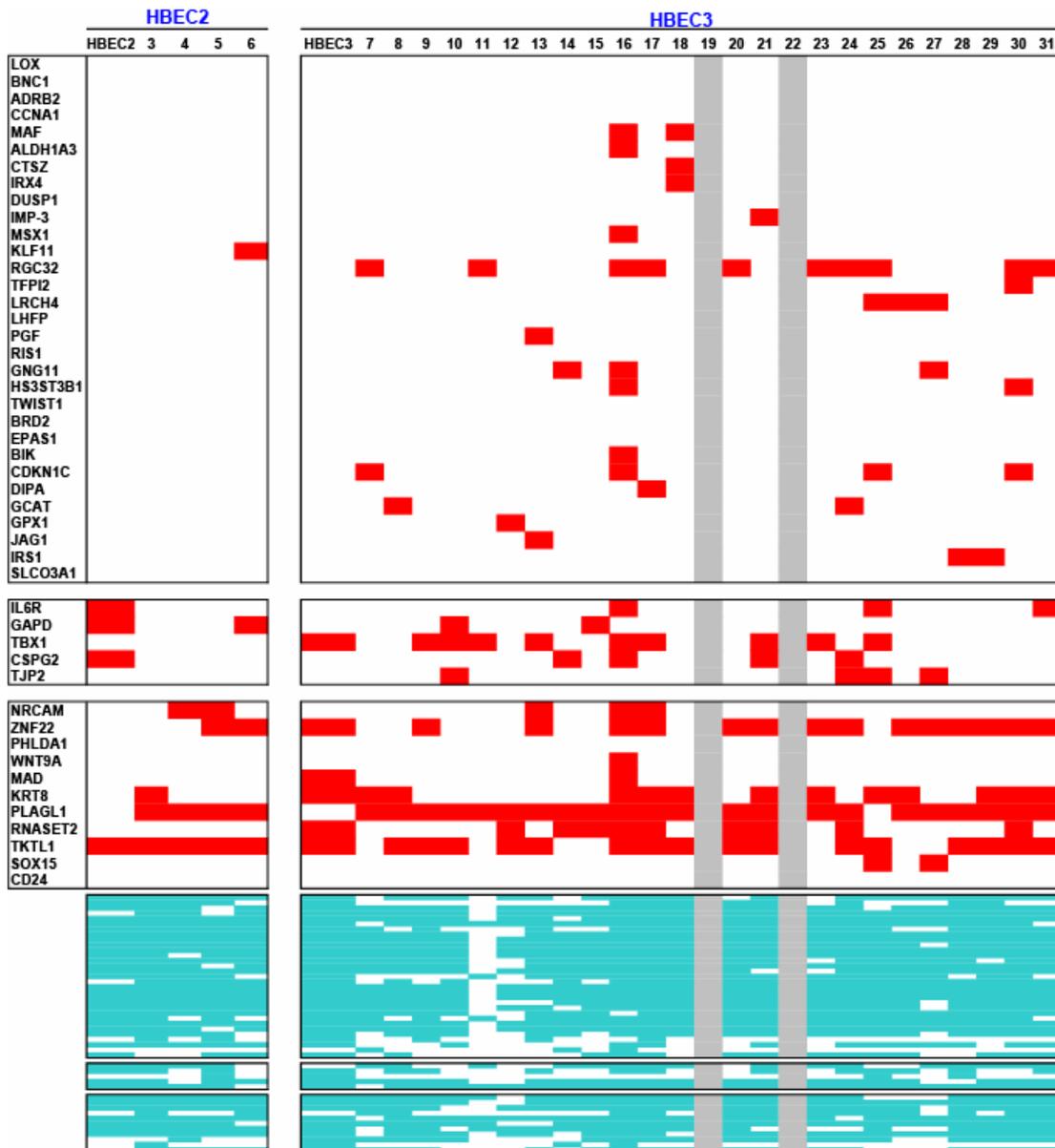


Figure 7-6. Summary of MSP for 5-aza induction gene set in HBEC2 and HBEC3 isogenic cell lines. Red fill indicates positive methylated product; aqua indicates positive unmethylated product. Data are grouped as follows: group I, no methylation in either parental HBECs or lymphocytes; group II, methylation in HBEC, but not lymphocytes; group III, methylation in lymphocytes. Data are ordered based on the frequency of methylation in primary lung tumors. Grey columns were bad DNA. See Table 7-2 for explanation of different samples and their genetic backgrounds.

CHAPTER EIGHT

CONCLUSIONS AND RECOMMENDATIONS

The objective of my thesis work was to use global profiling platforms (RNA and DNA) to identify epigenetically modulated genes that may be involved in cancer pathogenesis and bring these to the point where they could be developed as targets for diagnostic and treatment strategies. The first part of this work encompassed pilot studies that were designed to identify the appropriate approaches to demethylate genomic DNA. These studies led to the conclusion that while RNAi approaches could be used to demethylate DNA in cancer cells, RNAi was not particularly well suited to these experiments because of the potential for artifact associated with serial transfection. Thus, I decided that 5-aza treatment would be the best and most efficient approach to use in subsequent studies. In the second part of this thesis, microarray technology was used to identify novel, cancer-specific methylation markers that are present at high frequency in multiple common epithelial cancers.

The microarray studies form the major part of this work and were designed to avoid the pitfalls of previous studies. We used several approaches not in the literature to do this including different doses of 5-aza, a novel series of immortalized human bronchial epithelial cells, and stringent bioinformatic methods. In the first approach, different doses (100 and 1000 nM) of 5-aza covering the cytotoxic and subcytotoxic levels of the drug were used and only those genes that were induced > 4 fold as well as those that were induced incrementally

between the higher and lower doses were selected. The principle idea here was that “dose-dependent” changes in gene expression would likely be more specific than a response that is variable at the two doses. We obtained more information from this approach than would have been possible using a straightforward replicate-type experiment. In particular, we found that most genes did respond in a dose-dependent manner.

A second, new approach I developed was to perform the global gene expression profiling (47,000 transcripts) before and after treatment with 5-aza on 7 lung cancer cell lines selected by expression profiling and genetic studies to represent distinct subclasses of NSCLCs. Thus, I wanted to sample diverse types of NSCLC. I also performed similar experiments in three newly available immortalized human bronchial epithelial cells (HBECs) to identify genes whose expression was selectively lost in lung cancer, expressed in normal lung epithelium, but subject to 5-aza induction. These HBECs are immortal, clonable, can be genetically manipulated, but do not form colonies in soft agar nor do they form tumors in nude mice (Ramirez, Sheridan et al. 2004). In three-dimensional culture they can undergo differentiation into fully ciliated cells (Vaughan, Ramirez et al. 2006). They have very few genetic alterations and they are a novel and important normal tissue control for 5-aza or siRNA targeted to DNMT1 gene induction experiments. The use of non-malignant epithelial cells such as these as part of a global methylation induction screen had not been described previously in other cancer types.

A third aspect of these studies was that we made extensive use of bioinformatic processes to reduce the size of the 5-aza gene induction set. We applied a series of biological filters to extract a list of methylation candidates that we believed to have a strong likelihood

of being methylated specifically in tumors. Statistical analyses of the major steps in this process suggested that successive lists were enriched for genes with 5' CpG islands. We selected only those genes that were induced in more than one lung cancer and had well-defined CpG islands in their putative promoter regions were selected. This led us to identify 132 candidate genes of which 45 have been studied in detail in 20 primary lung cancers and companion normal lung tissue.

The large majority of the 132 genes identified in these studies have not been described previously to undergo tumor specific promoter hypermethylation. I found that expression of these genes distinguishes primary lung cancers from normal lung tissue in the same patient. While there are probably many genes that are methylated – perhaps at random – during carcinogenesis, we found that the vast majority of the 45 (out of 132) genes studied here undergo tumor specific methylation in multiple primary lung cancers. Eight of these 45 genes were studied in a panel of 105 primary tumors from NSCLC, breast, colon and prostate cancers and 82 histologically companion normal tissues, which showed that many undergo methylation in these common epithelial cancers. Thus, this approach has identified many new genes subject to frequent tumor acquired methylation in primary lung, breast, colon, and prostate cancers. Frequent methylation of specific genes in multiple independent cancers strongly suggests, although does not prove, that these genes are functionally relevant to cancer pathogenesis.

Based on the frequency and specificity of methylation in the promoters of some of these genes (*LOX*, *BNC1*, *CTSZ*, *PGF*) in both lung and breast primary tumors, and the absence of methylation in normal lymphocytes, these markers are candidates for prognostic

or diagnostic evaluation studies. Thus, the promoter hypermethylation profiles of these genes maybe generic markers of malignancy, and thus useful for clinical screening assays.

One goal of this study was to identify novel markers of tumor specific methylation for follow up functional analysis. To this end, our screen uncovered some well-established methylation markers that have tumor suppressor activity, including *TIMP3*, *CDH1*, and *SFRP1*, but missed others such as *p16* and *RASSF1A*. The reason for this highlights the limitations of current microarray technology in that commercial arrays cannot discriminate between alternative splice forms of genes; both *p16* and *RASSF1* have constitutively expressed alternative isoforms that can hybridize to probes specific for these loci. Since both genes have expressed isoforms (*p14* and *RASSF1C*) that differ only in their 5' regions, none of the probes specific to these genes detected differences in expression. This limitation means that we have probably missed isoforms of genes that are subject to tumor-specific methylation, but that are part of an active transcription locus.

For the most part, the genes we identified are novel methylation candidates. However, some of the genes we identified as cancer-specific methylation markers in NSCLCs have been identified previously as methylation candidates in other tissues. *LOX* was frequently methylated in our panel of cell lines and tumors, and was recently shown to be methylated in gastric cancers (Kaneda, Wakazono et al. 2004). *CCNA1* was shown to be methylated in head and neck cancers and was inversely correlated with *p53* mutation (Tokumaru, Yamashita et al. 2004). Interestingly, in our study, *CCNA1* was methylated only in NSCLC A549 cells, which has wild-type *p53*. Loss of dual specificity phosphatase I (*DUSP1*) expression as determined by immunohistochemistry inversely correlates with

increasing malignancy of prostate cancers, and methylation of its promoter appears to be an early event in this disease (Rauhala, Porkka et al. 2005). In another recent report, tissue factor pathway inhibitor 2 (*TFPI2*) methylation was used as part of a six gene panel to screen for cancer in pancreatic juice specimens (Matsubayashi, Canto et al. 2006). Promoter methylation of the transcription factor *TWIST1*, has been described in several reports, and is very frequent in neuroblastoma, cervical and breast cancers, although, curiously, high expression of *TWIST1* seems to be necessary for breast cancer metastasis (Alaminos, Davalos et al. 2004; Mehrotra, Vali et al. 2004; Yang, Mani et al. 2004; Feng, Balasubramanian et al. 2005). The proapoptotic *BCL2* family member, *BIK*, was identified in a global screen for promoter methylation in melanoma using restriction landmark genomic scanning (Pompeia, Hodge et al. 2004). It is intriguing that several of the genes we identified as cancer-specific methylation markers in primary NSCLC and breast cancer are also clearly methylated in diverse epithelial tumors.

Another goal of this study was to begin to understand how pervasive aberrant promoter hypermethylation is within a given cancer. Our data suggests that tumor-acquired promoter methylation is widespread, affecting at least 20 promoters in any given tumor, and probably many more, given that we examined methylation for only part of the 5-aza induction gene set. Interestingly, many of the genes we found, while conforming to the standard assumptions of candidate methylated genes – no expression in cancer and induction by 5-aza, but expressed in normal cells – appear to have both methylated and unmethylated alleles in tumor cell lines, even though the cell lines represent pure sources of tumor DNA. Whether this combination of methylated and unmethylated alleles represents

haploinsufficiency, and/or indicates other mechanisms for loss of gene expression/function remains to be determined. Other studies have reported similar results but did not comment on them (Sato, Fukushima et al. 2003).

It is possible that some loci in the 5-aza induction gene set are imprinted. We did find multiple examples of imprinted genes in our original gene set (*H19*, *Xist*, and *CDKN1C*). It would have been difficult to exclude these genes completely from our analysis and may not have been prudent to do so, as several recent reports have implicated biallelic inactivation or activation of imprinted loci to be crucial to the oncogenic process (Feinberg 2004; Holm, Jackson-Grusby et al. 2005; Hong, Kang et al. 2005). However, for most genes on our list, we find this explanation unlikely. In the simplest case, where the heritable trait is direct promoter methylation, imprinted loci should have a similar pattern of methylation in primary tumors and their counterpart normal tissue, which was not evident for most genes in our panel (with the exception of genes that were methylated everywhere, e.g. *TKTL1* and *RNASET2*). Second, imprinted genes would be evident as hemimethylated in the HBEC cell lines and normal lung tissue, whereas most were not. However, due to diverse mechanisms of genomic imprinting, we cannot exclude the possibility that at least some of the loci we have discovered are parentally imprinted (Feinberg and Tycko 2004; Holm, Jackson-Grusby et al. 2005).

Another possibility is that we have uncovered multiple loci that are specifically targeted for methylation and upon allele loss and consequent, partial loss of expression, contribute to tumor formation through haploinsufficiency. Somatic loss of heterozygosity through locus deletion is the most common genetic alteration in human cancer (Girard,

Zochbauer-Muller et al. 2000; Loeb 2001). But, the extent to which allele loss leads to haploinsufficiency or furthers malignancy has not been studied in as much detail as some other aspects of cancer biology. Where mouse models are available, many classical tumor suppressor genes such as *p53* and *PTEN*, exhibit a haploinsufficient phenotype (Santarosa and Ashworth 2004). Of particular importance to lung cancer, mice heterozygous for *RASSF1A* are as sensitive to chemically induced carcinogenesis as null mice (Tommasi, Dammann et al. 2005). Therefore, the effects of large-scale single allele loss through promoter hypermethylation may be analogous to genomic instability and loss of heterozygosity in the neoplastic process.

By contrasting the genome-wide changes in gene expression of normal and lung cancer cells, we were able to gain novel insight into the complexity of the methylation program required for cells to become fully malignant. Even though we began with a highly structured, organ-specific screen, by applying successive biological and statistical filters we identified several novel loci with exceptionally high penetrance in primary lung and breast tumors, suggesting that we may have captured the variation in promoter hypermethylation signatures found in at least two, and probably more, common epithelial cancers. We conclude that while tumors differ in their molecular phenotypes and pathogenesis, the pathways they follow toward malignancy may be similar and may be reflected in the methylation programs they engage, which in turn may provide an opportunity to exploit in early diagnosis or therapeutic strategies. Subsequent studies will be needed to determine whether these novel methylated loci could be useful in early detection screening, or whether loss of expression of their associated genes contributes to tumor initiation and pathogenesis.

FUTURE STUDIES

There are at least five lines of investigation that lead directly from the studies described here:

- 1) Study the remaining 87 genes in the 5-aza induction gene list for methylation in NSCLC and other types of cancer and determine whether the genes that are methylated in lung cancer are the similar or different to those that are methylated in other tissues. Additionally, compare the methylation profiles of new genes to other previously described genes and determine whether there are common patterns between different combinations of genes.
- 2) Construct a methylation platform for early cancer detection and diagnostic screening using the data provided here and that recently published in other microarray screens
- 3) Bioinformatic investigation of transcription factor binding sites and comparative sequence analysis of the proximate DNA sequences associated with the 5-aza induction gene set
- 4) Functional studies of the genes subject to frequent promoter hypermethylation and loss of expression as well as LOH
- 5) Study genes that are expressed in NSCLC and are off in HBEC, but come on in HBEC after 5-aza treatment. These genes are candidates for hypomethylation associated over expression in tumors and could act as oncogenes.

Study the remaining genes in the 5-aza induction gene list for methylation in NSCLC and other types of cancer and construct a methylation platform for early detection and diagnostic screening using the data provided here and that recently published in other microarray screens

Several aspects of DNA methylation have made it the focus of research in an increasingly large number of clinical and translational cancer research laboratories. First, DNA methylation is a tumor specific change in molecular composition of DNA that is readily detectable in patient samples including blood, sputa, urine, stool, and biopsy specimens. Second, alterations in DNA methylation – both hypermethylation and hypomethylation – are early events in carcinogenesis. Third, there are distinctions in the promoter hypermethylation profiles between different types of tumors from the same organ site as well as between tumors from different tissues. Fourth, DNA methylation is pharmacologically reversible – Decitabine (5-aza-2'-deoxycytidine) has been approved for use in acute myeloid leukemia (AML) and the myelodysplastic syndromes (MDS), and is currently in early stage trials for solid tumors (Lung Cancer Trial NCI-02-C-0205).

If detected early, many epithelial cancers are curable through surgical resection and adjuvant radiation or chemotherapy (Kelloff, Lippman et al. 2006). Unfortunately, with the exception of breast cancer and prostate cancer, early detection screening modalities are invasive and prohibitively expensive (Belinsky 2004). However, the altered DNA methylation patterns found in tumors, as well as in many preneoplastic syndromes such as Barrett's esophagus and ulcerative colitis (above), present a unique opportunity to detect cancer early on. This is because DNA is stable and can be amplified using PCR. As a result,

several labs including our own are currently examining the feasibility of using quantitative PCR based methods to detect the presence of methylation in biopsy material (Fackler, McVeigh et al. 2004; Shivapurkar, Stastny et al. 2005; Belinsky, Liechty et al. 2006; Fackler, Malone et al. 2006; Shivapurkar, Stastny et al. 2006).

The detection of DNA methylation in clinical samples has other potential uses besides early detection screening. Some groups have explored using DNA methylation as a marker to confirm tumor margins in surgically resected specimens (Goldenberg, Harden et al. 2004; Guo, House et al. 2004). However, some have questioned the utility of this approach as a single determinant. Several recent studies have explored whether specific combinations of methylation markers are useful in risk assessment for lung cancer, breast cancer, and pancreatic cancer. These studies have applied both quantitative and qualitative approaches to DNA isolated from sputa, ductal lavage, and pancreatic juice (Belinsky, Liechty et al. 2006; Fackler, Malone et al. 2006; Matsubayashi, Canto et al. 2006). Other groups are investigating similar applications using fine needle aspirates as source material. Some of these studies have reported promising results and future, larger studies perhaps using more markers, will be useful in diagnostic settings. Other applications for DNA methylation profiling in a clinical setting include prognosis, defining drug sensitivity profiles, as well as monitoring therapeutic interventions. Indeed, a tumor sensitive to a particular chemotherapy may have a different methylation profile compared to a tumor that is resistant to that particular drug.

The list of 132 candidates represents our best attempt to reduce the list to a manageable number while allowing for the limitations of available technology platforms.

The 45 genes that were studied in NSCLC were selected at random out of the 132. Determining whether the remaining 87 genes (excluding those known to be methylated) are methylated would be important because the results of these studies would have direct translational impact through the identification of pathways important as therapeutic targets as well as provide a basis for development of the promoter hypermethylation screening platform. Moreover, once the methylation data for these genes is known, it would be of interest to go back to the primary tumor expression arrays and determine whether these data could have been used to further reduce the starting gene list, or increase the possibility of finding biologically relevant genes.

Bioinformatic investigation of transcription factor binding sites and comparative sequence analysis of the proximate DNA sequences associated with the 5-aza induction gene set

One of the major questions surrounding DNA methylation and cancer involves how tumor cells establish their characteristic methylation patterns in the first place. Clearly, there is some type of sequence specificity associated with the activities of the different methyltransferases in that only certain CpG island associated genes are affected. The evident frequency and diverse patterns of methylation we found in the above-described studies, along with the data of others, suggest that most lung cancers have active epigenetic pathways. It could be that tumors co-opt the methylation machinery at critical stages in a given tumor's evolution and actively methylate single alleles of multiple loci in response to environmental cues such as genotoxic or hypoxic stress. This situation is analogous to normal processes

such as the cascade of epigenetic changes observed in differentiating lymphocytes during an immune response, except that in cancer cells, this process is likely dysregulated and random (Jaenisch and Bird 2003; Reiner 2003; Reiner, Mullen et al. 2003). Evolving cancer cells probably require an operational epigenetic program to react to changes in the tumor microenvironment and adapt to the continual insults raised by the body.

As discussed in *Chapter 1*, the biochemical data do not really indicate how the specificity of which genes become methylated during tumorigenesis comes about. One possibility is that chromatin binding proteins such as transcription factors or histone modifying proteins control sequence specific methylation in particular contexts. Following this line of reasoning, preliminary investigations into the sequence context of the 5-aza induced gene set suggest that there was a significant enrichment of genes associated with the transcription factor *c-REL*. This transcription factor is a proto-oncogene that may have a dominant role in lymphoma development, but may also be involved in other types of tumors. Other approaches identified several other transcription factors that may be enriched in the 5-aza induction gene set.

Enrichment for *c-REL* binding sites was initially identified using the TRASFAC database, which can be found through BIOBASE at <http://www.gene-regulation.com/>. To perform this analysis, I created gene lists of interest using their accession numbers, downloaded their upstream sequences (2KB or 10KB), and then used the TRANSFAC algorithm to identify transcription factor binding sites in these sequences.

To further investigate these data we have performed a whole genome transcription factor alignment in collaboration with David States at the University of Michigan, and

Alexander Pertsemlidis in the McDermott center. These data will be complimented with ChIP-on-ChIP analytical methods using antibodies for the various transcription factors using two novel HBEC lines with matched pair NSCLC tumor lines. The results of these studies would be important for understanding the influence of DNA methylation of cancer development, and potentially provide novel targets for therapeutics development.

The last two decades of research into the mechanisms of cancer epigenetics have begun bare fruit not just in our understanding of cancer pathogenesis, but also in terms of therapeutic strategies. As mentioned above, Decitabine has been approved for use in leukemia, and may be useful in treating solid tumors as well. However, there has been an explosion in the identification and screening of compounds that specifically target other epigenetic components including histone deacetylases and histones methylases. These drugs may be of use both by themselves and in combination with other agents. As a result of the recent findings in basic research, some of which have been described here, and the increasing interest from the medical community in using DNA methylation as a tool to detect cancer, the influence of epigenetics research on our understanding of cancer pathogenesis and progression has become significant. A greater understanding of the role of epigenetics in the development of cancer will no doubt have a positive impact on the mortality and morbidity of these diseases in the future.

APPENDIX A – COMPENDIUM OF METHYLATED GENES

Appendix A. Genes found to be methylated in tumors

Gene	Name	Accession Number	Cytoband	Methylated	References
14-3-3 Sigma	Stratfin	BC023552	1p36.11	Breast, pancreas, NSCLC, gyn. GI	Ferguson et al.
ABHD9	Amylase domain containing 9	AK026061	19p13.12	GI	Yamashita, S et al.
ABL1	V-abl Abelson murine leukemia viral oncogene homolog 1	NM_007313	9q34.1	50-100% CML, Some ALL	Issa et al.
ADRB2	Adrenergic, beta-2-, receptor, surface	M15169	5q31-q32	NSCLC, GI	Shames et al.
ALDH1A3	Aldehyde dehydrogenase 1 family, member A3	AF198444	15q26.3	NSCLC, GI, Breast, Prostate, Colon	Shames et al., Yamashita, et al.
APC	Adenomatous polyposis coli	NM_000038	5q21-q22	Colon, gastric and esophageal cancer	Hiltunen et al.
AR	Androgen receptor	NM_000044	Xq11.2-q12	Prostate, Colon ACFs	Jarrard et al.
BDNF	Brain-derived neurotrophic factor	NM_170735	11p13	GI	Yamashita, S et al.
BIK	BCL2-interacting killer (apoptosis-inducing)	AB051441	22q13.31	NSCLC	Shames et al.
BMP7	Bone morphogenetic protein 7 (osteogenic protein 1)	AK094764	20q13	GI	Yamashita, S et al.
BNC1	Basonuclin 1	L03427	15q25.2	NSCLC, Breast, Colon, Prostate	Shames et al.
BRCA1	Breast cancer 1, early onset	NM_007295	17q21	10-20% Breast cancer, some ovarian	Dobrovic et al.
CALCA	Calcitonin/calcitonin-related polypeptide, alpha	X02330	11p15.2-p15.1	25-75% Colon, NSCLC, hematologic neoplasms	Baylin et al.
CASP8	Caspase 8, apoptosis-related cysteine peptidase	BC017031	2q33-q34	Pediatric tumors, NSCLC, HCC, leukemia	Telitz et al.
CAV1	Caveolin 1	NM_001753	7q31.1	Breast, NSCLC, Prostate, Sarcoma	Cui et al.
CCNA1	Cyclin A1	CR616879	12q12.3-q13	NSCLC, Breast, Colon (benign), Prostate (benign), HCC, Head and Neck	Shames et al.
CD24	CD24 molecule	AK057112	6q21	NSCLC	Shames et al.
CD44	CD44 molecule (Indian blood group)	NM_000610	11p13	Prostate cancer, stroma	Lou W
CDH1	Cadherin 1, type 1, E-cadherin (epithelial)	NM_004360	16q22.1	NSCLC, Breast, GI	Graff et al.
CDH13	Cadherin 13, H-cadherin (heart)	BX53273	16q24.2-q24.3	NSCLC, Breast, colon, HCC	Sato et al.
CDH2	Cadherin 2, type 1, N-cadherin (neuronal)	NM_001792	18q11.2	GI	Yamashita, S et al.
CFTR	Cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7)	NM_000492	7q31.2	Cell Lines, HCC	Ding et al.
CLDN3	Claudin 3	NM_001306.2	7q11.23	GI	Yamashita, S et al.
COX2	Cyclooxygenase 2	NM_000963.1	10q25.2-q25.3	Colon, Pancreas, Prostate, NSCLC	Toyota et al.
CSPG2	Chondroitin sulfate proteoglycan 2 (version)	NM_004385	5q14.3	Methylated in NSCLC, Colon	Issa et al.
CTSL	Cathepsin L	AK055599	9q21-q22	GI	Yamashita, S et al.
CTSZ	Cathepsin Z	NM_001336	20q13	NSCLC, Breast	Shames et al.
CX26	Gap junction protein, beta 2, 26kDa (connexin 26)	NM_004004	13q11-q12	NSCLC, Breast	Tan et al.
DAFK	Death associated protein kinase	NM_004938.1	9q34.1	HCC, Head and Neck, Colon, NSCLC, Breast	Esteller et al.
DBCCR1	Deleted in bladder cancer 1	CR627401	9q32-q33	Astrocytoma, Bladder, HCC	Habuchi et al.
DCR1	Tumor necrosis factor receptor superfamily, member 10c, decoy without an intracellular domain	NM_003841	8p22-p21	NSCLC, Breast, Leukemia/Lymphoma	van Noesel MM
DOR2	Tumor necrosis factor receptor superfamily, member 10d, decoy with truncated death domain	NM_003840	8p21	NSCLC, Breast, Leukemia/Lymphoma	van Noesel MM
DIPA	Coiled-coil domain containing 85B	BMS58600	11q12.1	NSCLC	Shames et al.
DUSP1	Dual specificity phosphatase 1	AK127679	5q34	NSCLC	Shames et al.
EDNRB	Endothelial Receptor B	NM_000115	13q22	Prostate, head and neck, nasopharyngeal, bladder, NSCLC	Pao M et al.
EPAS1	Endothelial PAS domain protein 1	NM_001430	2p21-p16	NSCLC	Shames et al.
EPO	Erythropoietin	X02157	7q22	Neuroblastoma, astrocytoma	Yu et al.
ER	Estrogen Receptor	NM_001437	14q23.2	Colon, CML-BC, NSCLC	Ottaviano et al.
F2R	Coagulation factor II (thrombin) receptor	NM_001992	5q13	GI	Yamashita, S et al.
FADS1	Fatty acid desaturase 1	NM_013402	11q12.2-q13.1	GI	Yamashita, S et al.
FHIT	Fragile histidine triad gene	BX641016	3p14.2	10-20% Esophageal SCC	Tanaka et al.
FSD1	Fibronectin type III and SPRY domain containing 1	AY032617	19p13.3	GI	Yamashita, S et al.
FST	Follistatin	NM_006350	5q11.2	GI	Yamashita, S et al.
GATA-4	GATA binding protein 4	NM_002052	9p23.1-p22	NSCLC, colon, GI, ovarian	Akiyama et al.
GATA-6	GATA binding protein 6	BC047790	20q13.33	Colon, GI, NSCLC	Akiyama et al.
GCAT	Glycine C-acetyltransferase (2-amino-3-ketobutyrate coenzyme A ligase)	AK123190	22q13.1	NSCLC	Shames et al.
GNB11	Guanine nucleotide binding protein (G protein), gamma 11	BF971151	7q21	NSCLC	Shames et al.
GPC3	Glypican 3	CR627361	Xq26.1	Mesothelioma, Ovarian, Embryonal, Breast	Huber et al.
GPX1	Glutathione peroxidase 1	BM478682	3p21.3	NSCLC	Shames et al.
GREM1	Gremlin 1, cysteine knot superfamily, homolog (Xenopus laevis)	NM_013372	15q13-q15	NSCLC, Breast, GI	Suzuki, M et al.
GST-pi	Glutathione S-transferase pi	BM926728	11q13	Prostate, Liver, Colon, Breast, Kidney	Esteller et al.
HIC1	Hypermethylated in cancer 1	NM_006497	17p13.3	Breast, Brain, Colon, Prostate, Breast, NSCLC, Kidney, Liquid tumors	Issa et al.
HMLH1	MutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)	BX648844	3p21.3	Colon, Endometrial, GI, NSCLC, breast, GBM, liquid tumors	Kane et al.
HOXA5	Homeobox A5	NM_153631	7p15-p14	Breast	Shirasi et al.
HG3ST3B1	Heparan sulfate (glucosamine) 3-O-sulfotransferase 3B1	BC063301	17p12-p11.2	NSCLC, Breast	Shames et al.
IGFBP3	Insulin-like growth factor binding protein 3	NM_001013398	7p13-p12	NSCLC, Breast, GI	Yamashita, S et al.
IGFBP7	Insulin-like growth factor binding protein 7	BX648756	4q12	NSCLC, GI	Yamashita et al.
IL6R	Interleukin 6 receptor	NM_000565	1q21	NSCLC	Shames et al.
IMP-3	Insulin-like growth factor 2 mRNA binding protein 3	U97188	7p11	NSCLC	Shames et al.
IRF7	Interferon regulatory factor 7	BC021078	11p15.5	Astrocytoma, NSCLC, HCC	Yu et al.
IRS1	Insulin receptor substrate 1	NM_005544	2q36	NSCLC	Shames et al.
IRX4	Iroquois homeobox protein 4	NM_016358	5p15.3	NSCLC	Shames et al.
JAG1	Jagged 1 (Alagille syndrome)	AF003837	20p12.1-p11.23	NSCLC	Shames et al.
KRT8	Keratin 8	NM_002273	12q13	NSCLC	Shames et al.
LHFP	Lipoma HMGIC fusion partner	CR749848	13q12	NSCLC	Shames et al.
LKB1	Serine/threonine kinase 11	AB209553	19p13.3	Colon, Testicular, Breast	Esteller et al.
LOX	Lysyl oxidase	NM_002317	5q23.2	NSCLC, Breast, GI	Kameda et al.
LRCH4	Leucine-rich repeats and calponin homology (CH) domain containing 4	NM_002319	7q22	NSCLC	Shames et al.
MAD	MAX dimerization protein 1	BC098396	2p13-p12	NSCLC	Shames et al.
MAF	V-maf musculoaponeurotic fibrosarcoma oncogene homolog (avian)	NM_001031804	16q22-q23	NSCLC	Shames et al.

MARK1	MAP/microtubule affinity-regulating kinase 1	NM_018650	1q41	GI	Yamashita, S et al.
MGMT	O-6-methylguanine-DNA methyltransferase	AK126049	10q26	Brain, colon, NSCLC, breast, NHL etc.	Esteller et al.
MINT1	Amyloid beta (A4) precursor protein-binding, family A, member 1 (X11)	AK127028	9q13-q21.1	AML, Colon	Toyota M et al.
MINT2	Amyloid beta (A4) precursor protein-binding, family A, member 2 (X11-like)	AF047348	15q11-q12	AML, Colon	Toyota M et al.
MLF1	Myeloid leukemia factor 1	BX641078	3q25.1	GI	Yamashita, S et al.
MSX1	Msh homeobox homolog 1 (Drosophila)	NM_002448	4p16.3-p16.1	GI, NSCLC, Breast, Colon, Prostate	Yamashita, S et al.; Shames, D et al.
MT1a	Metallothionein 1A (functional)	NM_005946	16q13	HCC, astrocytoma	Yu et al.
MTSS1	Metastasis suppressor 1	NM_014751	8p22	GI	Yamashita, S et al.
MUC2	Mucin 2, oligomeric mucous/gel-forming	NM_002457	11p15.5	Colon, pancreas, GI	Gratchev et al.
MYO1D	Myogenic differentiation 1	NM_002478	11p15.4	Colon, Breast, Bladder, NSCLC, Liquid tumors.	Cheng et al.
N33	Tumor suppressor candidate 3	BX641112	8p22	Colon, prostate, brain.	Li et al.
NEP	Membrane metallo-endopeptidase (neutral endopeptidase, enkephalinase, CALLA, CD10)	NM_007289	3q25.1-q25.2	Prostate cancer	Usmani et al.
NRCAM	Neuronal cell adhesion molecule	NM_001037132	7q31.1-q31.2	NSCLC, some Breast	Shames et al.
p14	Cyclin-dependent kinase 4 inhibitor A - alternate reading frame	NM_058195	9p21	Colon	Esteller et al.
p15	Cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)	NM_078487	9p21	Colon/NSCLC/Breast, AML	Herman et al.
p16	Cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)	NM_000077	9p21	Nearly all	Merlo et al.
P27KIP1	Cyclin-dependent kinase inhibitor 1B (p27, Kip1)	NM_004054	12p13.1-p12	Breast, melanoma, CNS	Niwa et al.
p57 KIP2	Cyclin-dependent kinase inhibitor 1C (p57, Kip2)	BC057842	11p15.5	Gastric cancer cell lines, NSCLC	Corn et al.
p73	Tumor protein p73	NM_005427	1p36.3	Lymphoma, leukemia, NSCLC, prostate	Kawano et al.
PAX6	Paired box gene 6 (aniridia, keratitis)	AB209177	11p13	GI	Yamashita, S et al.
PGF	Placental growth factor, vascular endothelial growth factor-related protein	AK023843	14q24-q31	NSCLC	Shames et al.
PR	Progesterone Receptor	NM_000526	11q22-q23	Breast cancer	Lapidas et al.
PHLDA1	Pleckstrin homology-like domain, family A, member 1	NM_007350	12q15	NSCLC	Shames et al.
PLAGL1	Pleiomorphic adenoma gene-like 1	CR749329	6q24-q25	GI, NSCLC	Yamashita, S et al.
RARβ	Retinoic acid receptor, beta	BX640880	3p24	Colon, Breast, NSCLC Cancer	Sirohita SM
RASSF1A	Ras association (Raf/GDS/AP-5) domain family 1	NM_170715	3p21.3	All	Damman et al.
RB1	Retinoblastoma 1 (including osteosarcoma)	L41870	13q14.2	10-20% Retinoblastomas	Sakai T
RBP4	Retinol binding protein 4, plasma	BQ054708	10q23-q24	GI	Yamashita, S et al.
RGCC32	Response gene to complement 32	AK095079	13q14.11	NSCLC	Shames et al.
RIS1	Ras-induced senescence 1	CR615050	3p21.3	NSCLC	Shames et al.
RNASEH2B	Ribonuclease H2	AK001769	6q27	NSCLC	Shames et al.
RORA	RAR-related orphan receptor A	NM_0209916	15q21-q22	GI	Yamashita, S et al.
SCRN1	Secernin 1	AB2014766	7p14.3-p14.1	GI	Yamashita, S et al.
SFRP1	Secreted frizzled-related protein 1	BC036503	8p12-p11.1	Colon, NSCLC, Breast, esophageal, Head and Neck	Suzuki H, et al.
SFRP2	Secreted frizzled-related protein 2	NM_003013	4q31.3	Colon, NSCLC, Breast, esophageal, Head and Neck	Suzuki H, et al.
SFRP4	Secreted frizzled-related protein 4	AF026592	7p14.1	Colon, NSCLC, Breast, esophageal, Head and Neck	Suzuki H, et al.
SFRP5	Secreted frizzled-related protein 5	AF117758	10q24.1	Colon, NSCLC, Breast, esophageal, Head and Neck	Suzuki H, et al.
SLCO3A1	Solute carrier organic anion transporter family, member 3A1	AF205074	15q26	NSCLC	Shames et al.
SOC1	Suppressor of cytokine signaling 1	AK127621	16p13.13	NSCLC, Breast	Yoshikawa H
SOC3	Suppressor of cytokine signaling 3	NM_003655	17q25.3	NSCLC, Breast	He, B et al.
SOX15	SRY (sex determining region Y)-box 15	AB006867	17p13	NSCLC, Breast, Prostate, Colon, benign tissues	Shames et al.
TBX1	T-box 1	AF373867	22q11.21	NSCLC	Shames et al.
TBX3	T-box 3 (ulnar mammary syndrome)	NM_016569	12q24.1	GI	Yamashita, S et al.
TESTIN	Testis derived transcript (3 LIM domains)	NM_015641	7q31.2	Hematopoietic malignancies, ovarian cancer	Tobias ES
TFAP2C	Transcription factor AP-2 gamma (activating enhancer binding protein 2 gamma)	NM_003222	20q13.2	GI	Yamashita, S et al.
TFPI2	Tissue factor pathway inhibitor 2	AK129833	7q22	NSCLC	Shames et al.
THBS1	Thrombospondin 1	NM_003246	15q15	5-10% Colon Cancer 30-40% GBM 20-30% AML 0% Endometrial/Breast	Li, Q
TIMP3	TIMP metalloproteinase inhibitor 3 (Sorsby fundus dystrophy, pseudoinflammatory)	AB051444	22q12.1-q13.2	Human brain (10-50%) and Kidney (20%) cancers, NSCLC, breast, colon	Sachman KE
TJP2	Tight junction protein 2 (zona occludens 2)	AB209630	9q13-q21	NSCLC	Shames et al.
TMEFF2	Transmembrane protein with EGF-like and two follistatin-like domains 2	DQ133599	2q32.3	Colon, NSCLC, Breast, esophageal	Young J
TWIST1	Twist homolog 1 (acrocephalosyndactyly 3; Saethre-Chotzen syndrome) (Drosophila)	NM_000474	7p21.2	NSCLC	Shames et al.
uPA	Plasminogen activator, urokinase	NM_002658	10q24	Breast, NSCLC, prostate	Fakhshan, P
VHL	Von Hippel-Lindau tumor suppressor	NM_000551	3p26-p25	10-20% Renal Cell cancers 0% Common solid and liquid tumors	Herman et al.
WIF1	WNT inhibitory factor 1	AY358344	12q14.3	GI	Yamashita, S et al.
ZNF177	Zinc finger protein 177	NM_003451	19p13.2	GI	Yamashita, S et al.
ZNF22	Zinc finger protein 22 (KOX 15)	NM_006963	10q11	NSCLC	Shames et al.
ZNF559	Zinc finger protein 559	AK092348	19p13.2	GI	Yamashita, S et al.

Data are derived from several sources including Pubmed, Jean-Pierre Issa's website (<http://www.mdanderson.org/departments/methylation/index.cfm?pr=D02B3250-57D7-4F61-88358636A8073A08>), MethDB, Source (<http://genome-www5.stanford.edu/bin/sourceSearch>), Genecard (<http://genome-www.stanford.edu/genecards/index.shtml>), UCSC Genome Browser (<http://genome.ucsc.edu/cgi-bin/hgGateway>).

This list includes genes that are methylated in primary tumors (at least 10 have been tested)

BIBLIOGRAPHY

- Agarwal, S. and A. Rao (1998). "Modulation of chromatin structure regulates cytokine gene expression during T cell differentiation." *Immunity* **9**(6): 765-75.
- Akiyama, Y., N. Watkins, et al. (2003). "GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer." *Mol Cell Biol* **23**(23): 8429-39.
- Alaminos, M., V. Davalos, et al. (2004). "Clustering of gene hypermethylation associated with clinical risk groups in neuroblastoma." *J Natl Cancer Inst* **96**(16): 1208-19.
- Allison, D. B., X. Cui, et al. (2006). "Microarray data analysis: from disarray to consolidation and consensus." *Nat Rev Genet* **7**(1): 55-65.
- Ansel, K. M., D. U. Lee, et al. (2003). "An epigenetic view of helper T cell differentiation." *Nat Immunol* **4**(7): 616-23.
- Aoki, A., I. Suetake, et al. (2001). "Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases." *Nucleic Acids Res* **29**(17): 3506-12.
- Bachman, K. E., B. H. Park, et al. (2003). "Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene." *Cancer Cell* **3**(1): 89-95.
- Bacolla, A., S. Pradhan, et al. (1999). "Recombinant human DNA (cytosine-5) methyltransferase. II. Steady-state kinetics reveal allosteric activation by methylated dna." *J Biol Chem* **274**(46): 33011-9.
- Ballestrero, A., D. A. Coviello, et al. (2001). "Reverse-transcriptase polymerase chain reaction of the maspin gene in the detection of bone marrow breast carcinoma cell contamination." *Cancer* **92**(8): 2030-5.
- Baylin, S. B. (2002). "Mechanisms underlying epigenetically mediated gene silencing in cancer." *Semin Cancer Biol* **12**(5): 331-7.
- Baylin, S. B., S. A. Belinsky, et al. (2000). "Aberrant methylation of gene promoters in cancer--- concepts, misconcepts, and promise." *J Natl Cancer Inst* **92**(18): 1460-1.
- Baylin, S. B., J. G. Herman, et al. (1998). "Alterations in DNA methylation: a fundamental aspect of neoplasia." *Adv Cancer Res* **72**: 141-96.
- Baylin, S. B., J. W. Hoppener, et al. (1986). "DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas." *Cancer Res* **46**(6): 2917-22.
- Baylin, S. B., M. Makos, et al. (1991). "Abnormal patterns of DNA methylation in human neoplasia: potential consequences for tumor progression." *Cancer Cells* **3**(10): 383-90.
- Baylin, S. B. and J. E. Ohm (2006). "Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction?" *Nat Rev Cancer* **6**(2): 107-16.
- Baylin, S. B. and J. E. Ohm (2006). "Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction?" *Nat Rev Cancer* **6**(2): 107-116.
- Belinsky, S. A. (2004). "Gene-promoter hypermethylation as a biomarker in lung cancer." *Nat Rev Cancer* **4**(9): 707-17.
- Belinsky, S. A., K. C. Liechty, et al. (2006). "Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort." *Cancer Res* **66**(6): 3338-44.
- Bender, C. M., M. M. Pao, et al. (1998). "Inhibition of DNA methylation by 5-aza-2'-deoxycytidine suppresses the growth of human tumor cell lines." *Cancer Res* **58**(1): 95-101.

- Bender, C. M., J. M. Zingg, et al. (1998). "DNA methylation as a target for drug design." Pharm Res **15**(2): 175-87.
- Benson, D. A., I. Karsch-Mizrachi, et al. (2005). "GenBank." Nucl. Acids Res. **33**(suppl_1): D34-38.
- Berger, S. L. and G. Felsenfeld (2001). "Chromatin goes global." Mol Cell **8**(2): 263-8.
- Bernstein, B. E., T. S. Mikkelsen, et al. (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-26.
- Bernstein, E. and C. D. Allis (2005). "RNA meets chromatin." Genes Dev **19**(14): 1635-55.
- Bestor, T. H. (2003). "Unanswered questions about the role of promoter methylation in carcinogenesis." Ann N Y Acad Sci **983**: 22-7.
- Bestor, T. H., G. Gundersen, et al. (1992). "CpG islands in mammalian gene promoters are inherently resistant to de novo methylation." Genet Anal Tech Appl **9**(2): 48-53.
- Bird, A. (2002). "DNA methylation patterns and epigenetic memory." Genes Dev **16**(1): 6-21.
- Block, T. M., A. S. Mehta, et al. (2003). "Molecular viral oncology of hepatocellular carcinoma." Oncogene **22**(33): 5093-107.
- Bock, C., S. Reither, et al. (2005). "BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing." Bioinformatics **21**(21): 4067-4068.
- Bolstad, B. M., R. A. Irizarry, et al. (2003). "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias." Bioinformatics **19**(2): 185-93.
- Bourc'his, D. and T. H. Bestor (2006). "Origins of extreme sexual dimorphism in genomic imprinting." Cytogenet Genome Res **113**(1-4): 36-40.
- Boyer, L. A., K. Plath, et al. (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells." Nature **441**(7091): 349-53.
- Bridge, A. J., S. Pebernard, et al. (2003). "Induction of an interferon response by RNAi vectors in mammalian cells." Nat Genet **34**(3): 263-4.
- Burbee, D. G., E. Forgacs, et al. (2001). "Epigenetic inactivation of RASSF1A in lung and breast cancers and malignant phenotype suppression." J Natl Cancer Inst **93**(9): 691-9.
- Caplen, N. J., S. Parrish, et al. (2001). "Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems." Proc Natl Acad Sci U S A **98**(17): 9742-7.
- Carbone, M., M. Bocchetta, et al. (2003). "SV40 and human brain tumors." Int J Cancer **106**(1): 140-2; author reply 143-5.
- Chan, A. O., R. R. Broaddus, et al. (2002). "CpG island methylation in aberrant crypt foci of the colorectum." Am J Pathol **160**(5): 1823-30.
- Chang, M.-S., H. Uozaki, et al. (2006). "CpG Island Methylation Status in Gastric Carcinoma with and without Infection of Epstein-Barr Virus." Clin Cancer Res **12**(10): 2995-3002.
- Chen, R. Z., U. Pettersson, et al. (1998). "DNA hypomethylation leads to elevated mutation rates." Nature **395**(6697): 89-93.
- Cheng, J., P. Kapranov, et al. (2005). "Transcriptional Maps of 10 Human Chromosomes at 5-Nucleotide Resolution." Science **308**(5725): 1149-1154.
- Chow, J. C., Z. Yen, et al. (2005). "SILENCING OF THE MAMMALIAN X CHROMOSOME." Annual Review of Genomics and Human Genetics **6**(1): 69-92.
- Clark, S. J., J. Harrison, et al. (1994). "High sensitivity mapping of methylated cytosines." Nucleic Acids Res **22**(15): 2990-7.

- Colella, S., L. Shen, et al. (2003). "Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites." *Biotechniques* **35**(1): 146-50.
- Costello, J. F., M. C. Fruhwald, et al. (2000). "Aberrant CpG-island methylation has non-random and tumour-type-specific patterns." *Nat Genet* **24**(2): 132-8.
- Costello, J. F., D. J. Smiraglia, et al. (2002). "Restriction landmark genome scanning." *Methods* **27**(2): 144-9.
- Crew, K. D. and A. I. Neugut (2004). "Epidemiology of upper gastrointestinal malignancies." *Semin Oncol* **31**(4): 450-64.
- Curtis, C. D. and M. Goggins (2005). "DNA methylation analysis in human cancer." *Methods Mol Med* **103**: 123-36.
- Dammann, R., C. Li, et al. (2000). "Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3." *Nat Genet* **25**(3): 315-9.
- Dammann, R., G. Yang, et al. (2001). "Hypermethylation of the cpG island of Ras association domain family 1A (RASSF1A), a putative tumor suppressor gene from the 3p21.3 locus, occurs in a large percentage of human breast cancers." *Cancer Res* **61**(7): 3105-9.
- de Laat, W. and F. Grosveld (2003). "Spatial organization of gene expression: the active chromatin hub." *Chromosome Res* **11**(5): 447-59.
- Dennis, G., Jr., B. T. Sherman, et al. (2003). "DAVID: Database for Annotation, Visualization, and Integrated Discovery." *Genome Biol* **4**(5): P3.
- Duenas-Gonzalez, A., M. Lizano, et al. (2005). "Epigenetics of cervical cancer. An overview and therapeutic perspectives." *Mol Cancer* **4**: 38.
- Dunn, G. P., K. C. Sheehan, et al. (2005). "IFN unresponsiveness in LNCaP cells due to the lack of JAK1 gene expression." *Cancer Res* **65**(8): 3447-53.
- Dupont, J. M., J. Tost, et al. (2004). "De novo quantitative bisulfite sequencing using the pyrosequencing technology." *Anal Biochem* **333**(1): 119-27.
- Eads, C. A., K. D. Danenberg, et al. (2000). "MethyLight: a high-throughput assay to measure DNA methylation." *Nucleic Acids Res* **28**(8): E32.
- Eads, C. A., K. D. Danenberg, et al. (1999). "CpG island hypermethylation in human colorectal tumors is not associated with DNA methyltransferase overexpression." *Cancer Res* **59**(10): 2302-6.
- Eden, A., F. Gaudet, et al. (2003). "Response to Comment on "Chromosomal Instability and Tumors Promoted by DNA Hypomethylation" and "Induction of Tumors in Mice by Genomic Hypomethylation"." *Science* **302**(5648): 1153c-.
- Eden, A., F. Gaudet, et al. (2003). "Chromosomal instability and tumors promoted by DNA hypomethylation." *Science* **300**(5618): 455.
- Efron, B. and R. Tibshirani (2002). "Empirical bayes methods and false discovery rates for microarrays." *Genet Epidemiol* **23**(1): 70-86.
- Eggan, K., H. Akutsu, et al. (2000). "X-Chromosome inactivation in cloned mouse embryos." *Science* **290**(5496): 1578-81.
- Ehrich, M., M. R. Nelson, et al. (2005). "Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry." *Proc Natl Acad Sci U S A* **102**(44): 15785-90.
- Eisen, M. B., P. T. Spellman, et al. (1998). "Cluster analysis and display of genome-wide expression patterns." *Proc Natl Acad Sci U S A* **95**(25): 14863-8.
- Elbashir, S. M., J. Harborth, et al. (2001). "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells." *Nature* **411**(6836): 494-8.

- Elbashir, S. M., W. Lendeckel, et al. (2001). "RNA interference is mediated by 21- and 22-nucleotide RNAs." Genes Dev **15**(2): 188-200.
- Elena Kolomietz, M. S. M. A. P. J. A. S. (2002). "The role of *Alu* repeat clusters as mediators of recurrent chromosomal aberrations in tumors." Genes, Chromosomes and Cancer **35**(2): 97-112.
- Englander, E. W., A. P. Wolffe, et al. (1993). "Nucleosome interactions with a human *Alu* element. Transcriptional repression and effects of template methylation." J Biol Chem **268**(26): 19565-73.
- Esteller, M., P. G. Corn, et al. (2001). "A gene hypermethylation profile of human cancer." Cancer Res **61**(8): 3225-9.
- Esteller, M., M. F. Fraga, et al. (2001). "DNA methylation patterns in hereditary human cancers mimic sporadic tumorigenesis." Hum Mol Genet **10**(26): 3001-7.
- Esteller, M., J. M. Silva, et al. (2000). "Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors." J Natl Cancer Inst **92**(7): 564-9.
- Fackler, M. J., K. Malone, et al. (2006). "Quantitative multiplex methylation-specific PCR analysis doubles detection of tumor cells in breast ductal fluid." Clin Cancer Res **12**(11 Pt 1): 3306-10.
- Fackler, M. J., M. McVeigh, et al. (2004). "Quantitative multiplex methylation-specific PCR assay for the detection of promoter hypermethylation in multiple genes in breast cancer." Cancer Res **64**(13): 4442-52.
- Fahrner, J. A. and S. B. Baylin (2003). "Heterochromatin: stable and unstable invasions at home and abroad." Genes Dev **17**(15): 1805-12.
- Fang, J.-Y., J. A. Mikovits, et al. (2001). "Infection of Lymphoid Cells by Integration-Defective Human Immunodeficiency Virus Type 1 Increases De Novo Methylation." J. Virol. **75**(20): 9753-9761.
- Fearon, E. R. and B. Vogelstein (1990). "A genetic model for colorectal tumorigenesis." Cell **61**(5): 759-67.
- Feinberg, A. P. (2004). "The epigenetics of cancer etiology." Semin Cancer Biol **14**(6): 427-32.
- Feinberg, A. P., R. Ohlsson, et al. (2006). "The epigenetic progenitor origin of human cancer." Nat Rev Genet **7**(1): 21-33.
- Feinberg, A. P. and B. Tycko (2004). "The history of cancer epigenetics." Nat Rev Cancer **4**(2): 143-53.
- Feinberg, A. P. and B. Vogelstein (1983). "Hypomethylation distinguishes genes of some human cancers from their normal counterparts." Nature **301**(5895): 89-92.
- Feinberg, A. P. and B. Vogelstein (1983). "Hypomethylation of ras oncogenes in primary human cancers." Biochem Biophys Res Commun **111**(1): 47-54.
- Felsenfeld, G. and M. Groudine (2003). "Controlling the double helix." Nature **421**(6921): 448-53.
- Feng, Q., A. Balasubramanian, et al. (2005). "Detection of hypermethylated genes in women with and without cervical neoplasia." J Natl Cancer Inst **97**(4): 273-82.
- Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*." Nature **391**(6669): 806-11.
- Fong, K. M., Y. Kida, et al. (1995). "Loss of heterozygosity frequently affects chromosome 17q in non-small cell lung cancer." Cancer Res **55**(19): 4268-72.
- Fong, K. M., J. Schonrock, et al. (1995). "Human papillomavirus not found in squamous and large cell lung carcinomas by polymerase chain reaction." Cancer **75**(9): 2400-1.

- Fong, K. M., Y. Sekido, et al. (2003). "Lung cancer. 9: Molecular biology of lung cancer: clinical implications." Thorax **58**(10): 892-900.
- Fong, K. M., P. V. Zimmerman, et al. (1994). "Correlation of loss of heterozygosity at 11p with tumour progression and survival in non-small cell lung cancer." Genes Chromosomes Cancer **10**(3): 183-9.
- Fong, K. M., P. V. Zimmerman, et al. (1995). "Microsatellite instability and other molecular abnormalities in non-small cell lung cancer." Cancer Res **55**(1): 28-30.
- Fong, K. M., P. V. Zimmerman, et al. (1995). "Tumor progression and loss of heterozygosity at 5q and 18q in non-small cell lung cancer." Cancer Res **55**(2): 220-3.
- Fournel, M., P. Sapieha, et al. (1999). "Down-regulation of human DNA-(cytosine-5) methyltransferase induces cell cycle regulators p16(ink4A) and p21(WAF/Cip1) by distinct mechanisms." J Biol Chem **274**(34): 24250-6.
- Frommer, M., L. E. McDonald, et al. (1992). "A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands." Proc Natl Acad Sci U S A **89**(5): 1827-31.
- Fukuhara, T., W. C. Hooper, et al. (1992). "Use of the polymerase chain reaction to detect hypermethylation in the calcitonin gene. A new, sensitive approach to monitor tumor cells in acute myelogenous leukemia." Leuk Res **16**(10): 1031-40.
- Gama-Sosa, M. A., V. A. Slagel, et al. (1983). "The 5-methylcytosine content of DNA from human tumors." Nucl. Acids Res. **11**(19): 6883-6894.
- Garber, M. E., O. G. Troyanskaya, et al. (2001). "Diversity of gene expression in adenocarcinoma of the lung." Proc Natl Acad Sci U S A **98**(24): 13784-9.
- Garcia-Manero, G., S. Jeha, et al. (2003). "Aberrant DNA methylation in pediatric patients with acute lymphocytic leukemia." Cancer **97**(3): 695-702.
- Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." J Mol Biol **196**(2): 261-82.
- Garrick, D., S. Fiering, et al. (1998). "Repeat-induced gene silencing in mammals." Nat Genet **18**(1): 56-9.
- Gaudet, F., J. G. Hodgson, et al. (2003). "Induction of tumors in mice by genomic hypomethylation." Science **300**(5618): 489-92.
- Gaudet, F., W. M. Rideout, 3rd, et al. (2004). "Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing." Mol Cell Biol **24**(4): 1640-8.
- Gazdar, A. F., J. S. Butel, et al. (2002). "SV40 and human tumours: myth, association or causality?" Nat Rev Cancer **2**(12): 957-64.
- Gazdar, A. F., V. Kurvari, et al. (1998). "Characterization of paired tumor and non-tumor cell lines established from patients with breast cancer." Int J Cancer **78**(6): 766-74.
- Geneviève Clément, R. B. N. P. F. T. B. J. B. (2006). "Methylation of *APC*, *TIMP3*, and *TERT*: a new predictive marker to distinguish Barrett's oesophagus patients at risk for malignant transformation." The Journal of Pathology **208**(1): 100-107.
- Girard, L., S. Zochbauer-Muller, et al. (2000). "Genome-wide allelotyping of lung cancer identifies new regions of allelic loss, differences between small cell lung cancer and non-small cell lung cancer, and loci clustering." Cancer Res **60**(17): 4894-906.
- Gius, D., H. Cui, et al. (2004). "Distinct effects on gene expression of chemical and genetic manipulation of the cancer epigenome revealed by a multimodality approach." Cancer Cell **6**(4): 361-71.

- Goldenberg, D., S. Harden, et al. (2004). "Intraoperative molecular margin analysis in head and neck cancer." Arch Otolaryngol Head Neck Surg **130**(1): 39-44.
- Goll, M. G., F. Kirpekar, et al. (2006). "Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2." Science **311**(5759): 395-8.
- Grace Goll, M. and T. H. Bestor (2005). "Eukaryotic cytosine methyltransferases." Annu Rev Biochem **74**: 481-514.
- Griffiths, D. J. (2001). "Endogenous retroviruses in the human genome sequence." Genome Biol **2**(6): REVIEWS1017.
- Guo, L., J. Hu-Li, et al. (2002). "In TH2 cells the Il4 gene has a series of accessibility states associated with distinctive probabilities of IL-4 production." Proc Natl Acad Sci U S A **99**(16): 10623-8.
- Guo, M., M. G. House, et al. (2004). "Promoter hypermethylation of resected bronchial margins: a field defect of changes?" Clin Cancer Res **10**(15): 5131-6.
- Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.
- Harborth, J., S. M. Elbashir, et al. (2001). "Identification of essential genes in cultured mammalian cells using small interfering RNAs." J Cell Sci **114**(Pt 24): 4557-65.
- Heisler, L. E., D. Torti, et al. (2005). "CpG Island microarray probe sequences derived from a physical library are representative of CpG Islands annotated on the human genome." Nucleic Acids Res **33**(9): 2952-61.
- Herman, J. G., J. R. Graff, et al. (1996). "Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands." Proc Natl Acad Sci U S A **93**(18): 9821-6.
- Hirsch, F. R., W. A. Franklin, et al. (2001). "Early detection of lung cancer: clinical perspectives of recent advances in biology and radiology." Clin Cancer Res **7**(1): 5-22.
- Hochedlinger, K., R. Blalock, et al. (2004). "Reprogramming of a melanoma genome by nuclear transplantation." Genes Dev **18**(15): 1875-85.
- Hollstein, M., D. Sidransky, et al. (1991). "p53 mutations in human cancers." Science **253**(5015): 49-53.
- Holm, T. M., L. Jackson-Grusby, et al. (2005). "Global loss of imprinting leads to widespread tumorigenesis in adult mice." Cancer Cell **8**(4): 275-85.
- Holst, C. R., G. J. Nuovo, et al. (2003). "Methylation of p16(INK4a) promoters occurs in vivo in histologically normal human mammary epithelia." Cancer Res **63**(7): 1596-601.
- Hong, J. A., Y. Kang, et al. (2005). "Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells." Cancer Res **65**(17): 7763-74.
- Houghton, J. and T. C. Wang (2005). "Helicobacter pylori and gastric cancer: a new paradigm for inflammation-associated epithelial cancers." Gastroenterology **128**(6): 1567-78.
- Huisinga, K., B. Brower-Toland, et al. (2006). "The contradictory definitions of heterochromatin: transcription and silencing." Chromosoma **115**(2): 110-122.
- Hutchins, A. S., A. C. Mullen, et al. (2002). "Gene silencing quantitatively controls the function of a developmental trans-activator." Mol Cell **10**(1): 81-91.
- Illmensee, K. and B. Mintz (1976). "Totipotency and normal differentiation of single teratocarcinoma cells cloned by injection into blastocysts." Proc Natl Acad Sci U S A **73**(2): 549-53.
- Irizarry, R. A., B. M. Bolstad, et al. (2003). "Summaries of Affymetrix GeneChip probe level data." Nucleic Acids Res **31**(4): e15.
- Issa, J. P. (2004). "CpG island methylator phenotype in cancer." Nat Rev Cancer **4**(12): 988-93.

- Issa, J. P., N. Ahuja, et al. (2001). "Accelerated age-related CpG island methylation in ulcerative colitis." *Cancer Res* **61**(9): 3573-7.
- Jackson-Grusby, L., C. Beard, et al. (2001). "Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation." *Nat Genet* **27**(1): 31-9.
- Jaenisch, R. and A. Bird (2003). "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals." *Nat Genet* **33** **Suppl**: 245-54.
- Jaenisch, R., K. Hochedlinger, et al. (2005). "Nuclear cloning, epigenetic reprogramming and cellular differentiation." *Novartis Found Symp* **265**: 107-18; discussion 118-28.
- Janowski, B. A., K. E. Huffman, et al. (2005). "Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs." *Nat Chem Biol* **1**(4): 216-22.
- Jeltsch, A. (2002). "Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases." *Chembiochem* **3**(4): 274-93.
- Jeltsch, A. (2006). "Molecular enzymology of mammalian DNA methyltransferases." *Curr Top Microbiol Immunol* **301**: 203-25.
- Jeronimo, C., R. Henrique, et al. (2004). "A Quantitative Promoter Methylation Profile of Prostate Cancer." *Clin Cancer Res* **10**(24): 8472-8478.
- Jones, P. A. (1985). "Effects of 5-azacytidine and its 2'-deoxyderivative on cell differentiation and DNA methylation." *Pharmacol Ther* **28**(1): 17-27.
- Jones, P. A. and S. B. Baylin (2002). "The fundamental role of epigenetic events in cancer." *Nat Rev Genet* **3**(6): 415-28.
- Jones, P. A. and R. Martienssen (2005). "A blueprint for a Human Epigenome Project: the AACR Human Epigenome Workshop." *Cancer Res* **65**(24): 11241-6.
- Jones, P. A. and S. M. Taylor (1980). "Cellular differentiation, cytidine analogs and DNA methylation." *Cell* **20**(1): 85-93.
- Kaneda, A., K. Wakazono, et al. (2004). "Lysyl oxidase is a tumor suppressor gene inactivated by methylation and loss of heterozygosity in human gastric cancers." *Cancer Res* **64**(18): 6410-5.
- Karin, M., T. Lawrence, et al. (2006). "Innate immunity gone awry: linking microbial infections to chronic inflammation and cancer." *Cell* **124**(4): 823-35.
- Karpf, A. R., P. W. Peterson, et al. (1999). "Inhibition of DNA methyltransferase stimulates the expression of signal transducer and activator of transcription 1, 2, and 3 genes in colon tumor cells." *Proc Natl Acad Sci U S A* **96**(24): 14007-12.
- Kataoka, H., T. Shimomura, et al. (2000). "Hepatocyte growth factor activator inhibitor type 1 is a specific cell surface binding protein of hepatocyte growth factor activator (HGFA) and regulates HGFA activity in the pericellular microenvironment." *J Biol Chem* **275**(51): 40453-62.
- Kelloff, G. J., S. M. Lippman, et al. (2006). "Progress in Chemoprevention Drug Development: The Promise of Molecular Biomarkers for Prevention of Intraepithelial Neoplasia and Cancer--A Plan to Move Forward." *Clin Cancer Res* **12**(12): 3661-3697.
- Keshet, I., Y. Schlesinger, et al. (2006). "Evidence for an instructive mechanism of de novo methylation in cancer cells." *Nat Genet* **38**(2): 149-53.
- Kim, N. W., M. A. Piatyszek, et al. (1994). "Specific association of human telomerase activity with immortal cells and cancer." *Science* **266**(5193): 2011-5.
- Kitkumthorn, N., P. Yanatatsanajit, et al. (2006). "Cyclin A1 promoter hypermethylation in human papillomavirus-associated cervical cancer." *BMC Cancer* **6**: 55.

- Kuroki, T., F. Trapasso, et al. (2003). "Allelic Loss on Chromosome 3p21.3 and Promoter Hypermethylation of Semaphorin 3B in Non-Small Cell Lung Cancer." *Cancer Res* **63**(12): 3352-5.
- Lambert, R., P. Hainaut, et al. (2004). "Premalignant lesions of the esophagogastric mucosa." *Semin Oncol* **31**(4): 498-512.
- Lee, D. U., S. Agarwal, et al. (2002). "Th2 lineage commitment and efficient IL-4 production involves extended demethylation of the IL-4 gene." *Immunity* **16**(5): 649-60.
- Lee, G. R., P. E. Fields, et al. (2003). "Regulation of the Th2 cytokine locus by a locus control region." *Immunity* **19**(1): 145-53.
- Lee, J. T. and R. Jaenisch (1997). "The (epi)genetic control of mammalian X-chromosome inactivation." *Curr Opin Genet Dev* **7**(2): 274-80.
- Lee, P. P., D. R. Fitzpatrick, et al. (2001). "A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival." *Immunity* **15**(5): 763-74.
- Lee, T. I., R. G. Jenner, et al. (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." *Cell* **125**(2): 301-13.
- Lehmann, U., I. Berg-Ribbe, et al. (2005). "Distinct Methylation Patterns of Benign and Malignant Liver Tumors Revealed by Quantitative Methylation Profiling." *Clin Cancer Res* **11**(10): 3654-3660.
- Lei, H., S. P. Oh, et al. (1996). "De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells." *Development* **122**(10): 3195-205.
- Lengauer, C. (2003). "CANCER: An Unstable Liaison." *Science* **300**(5618): 442-443.
- Lengauer, C., K. W. Kinzler, et al. (1997). "DNA methylation and genetic instability in colorectal cancer cells." *PNAS* **94**(6): 2545-2550.
- Lewis, A. and W. Reik (2006). "How imprinting centres work." *Cytogenet Genome Res* **113**(1-4): 81-9.
- Lewis, C. M., L. R. Cler, et al. (2005). "Promoter hypermethylation in benign breast epithelium in relation to predicted breast cancer risk." *Clin Cancer Res* **11**(1): 166-72.
- Li, E., C. Beard, et al. (1993). "DNA methylation, genomic imprinting, and mammalian development." *Cold Spring Harb Symp Quant Biol* **58**: 297-305.
- Li, E., T. H. Bestor, et al. (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." *Cell* **69**(6): 915-26.
- Li, H., C. Stoicov, et al. (2003). "Helicobacter and gastric cancer disease mechanisms: host response and disease susceptibility." *Curr Gastroenterol Rep* **5**(6): 459-67.
- Li, L. C. and R. Dahiya (2002). "MethPrimer: designing primers for methylation PCRs." *Bioinformatics* **18**(11): 1427-31.
- Li, X., A. M. Hui, et al. (2004). "p16INK4A hypermethylation is associated with hepatitis virus infection, age, and gender in hepatocellular carcinoma." *Clin Cancer Res* **10**(22): 7484-9.
- Lippman, Z. and R. Martienssen (2004). "The role of RNA interference in heterochromatic silencing." *Nature* **431**(7006): 364-70.
- Lippman, Z., B. May, et al. (2003). "Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification." *PLoS Biol* **1**(3): E67.
- Lodygin, D., A. Epanchintsev, et al. (2005). "Functional epigenomics identifies genes frequently silenced in prostate cancer." *Cancer Res* **65**(10): 4218-27.
- Loeb, L. A. (2001). "A mutator phenotype in cancer." *Cancer Res* **61**(8): 3230-9.
- Ma, Y., S. B. Jacobs, et al. (2005). "DNA CpG hypomethylation induces heterochromatin reorganization involving the histone variant macroH2A." *J Cell Sci* **118**(Pt 8): 1607-16.

- Maekita, T., K. Nakazawa, et al. (2006). "High Levels of Aberrant DNA Methylation in Helicobacter pylori-Infected Gastric Mucosae and its Possible Association with Gastric Cancer Risk." Clin Cancer Res **12**(3): 989-995.
- Makar, K. W., M. Perez-Melgosa, et al. (2003). "Active recruitment of DNA methyltransferases regulates interleukin 4 in thymocytes and T cells." Nat Immunol **4**(12): 1183-90.
- Margetts, C. D., D. Astuti, et al. (2005). "Epigenetic analysis of HIC1, CASP8, FLIP, TSP1, DCR1, DCR2, DR4, DR5, KvDMR1, H19 and preferential 11p15.5 maternal-allele loss in von Hippel-Lindau and sporadic pheochromocytomas." Endocr Relat Cancer **12**(1): 161-72.
- Maruyama, R., S. Toyooka, et al. (2002). "Aberrant promoter methylation profile of prostate cancers and its relationship to clinicopathological features." Clin Cancer Res **8**(2): 514-9.
- Matsubayashi, H., M. Canto, et al. (2006). "DNA methylation alterations in the pancreatic juice of patients with suspected pancreatic disease." Cancer Res **66**(2): 1208-17.
- Matsukura, S., P. A. Jones, et al. (2003). "Establishment of conditional vectors for hairpin siRNA knockdowns." Nucleic Acids Res **31**(15): e77.
- Mehrotra, J., M. Vali, et al. (2004). "Very high frequency of hypermethylated genes in breast cancer metastasis to the bone, brain, and lung." Clin Cancer Res **10**(9): 3104-9.
- Merlo, A., J. G. Herman, et al. (1995). "5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers." Nat Med **1**(7): 686-92.
- Michalowsky, L. A. and P. A. Jones (1987). "Differential nuclear protein binding to 5-azacytosine-containing DNA as a potential mechanism for 5-aza-2'-deoxycytidine resistance." Mol Cell Biol **7**(9): 3076-83.
- Mikovits, J. A., H. A. Young, et al. (1998). "Infection with Human Immunodeficiency Virus Type 1 Upregulates DNA Methyltransferase, Resulting in De Novo Methylation of the Gamma Interferon (IFN-gamma) Promoter and Subsequent Downregulation of IFN-gamma Production." Mol. Cell. Biol. **18**(9): 5166-5177.
- Mintz, B. and K. Illmensee (1975). "Normal genetically mosaic mice produced from malignant teratocarcinoma cells." Proc Natl Acad Sci U S A **72**(9): 3585-9.
- Mitsudomi, T., S. M. Steinberg, et al. (1992). "p53 gene mutations in non-small-cell lung cancer cell lines and their correlation with the presence of ras mutations and clinical features." Oncogene **7**(1): 171-80.
- Moss, E. G. and J. M. Taylor (2003). "Small-interfering RNAs in the radar of the interferon system." Nat Cell Biol **5**(9): 771-2.
- Nardone, G., A. Rocco, et al. (2004). "Review article: helicobacter pylori and molecular events in precancerous gastric lesions." Aliment Pharmacol Ther **20**(3): 261-70.
- Nowell, P. C. (1976). "The clonal evolution of tumor cell populations." Science **194**(4260): 23-8.
- Okano, M., D. W. Bell, et al. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-57.
- Ostertag, E. M. and H. H. Kazazian Jr (2001). "BIOLOGY OF MAMMALIAN L1 RETROTRANSPOSONS." Annual Review of Genetics **35**(1): 501-538.
- Ottaviano, Y. L., J. P. Issa, et al. (1994). "Methylation of the estrogen receptor gene CpG island marks loss of estrogen receptor expression in human breast cancer cells." Cancer Res **54**(10): 2552-5.
- Pauler, F. M. and D. P. Barlow (2006). "Imprinting mechanisms--it only takes two." Genes Dev **20**(10): 1203-6.

- Paz, M. F., S. Avila, et al. (2002). "Germ-line variants in methyl-group metabolism genes and susceptibility to DNA methylation in normal tissues and human primary tumors." Cancer Res **62**(15): 4519-24.
- Paz, M. F., M. F. Fraga, et al. (2003). "A systematic profile of DNA methylation in human cancer cell lines." Cancer Res **63**(5): 1114-21.
- Perou, C. M., T. Sorlie, et al. (2000). "Molecular portraits of human breast tumours." Nature **406**(6797): 747-52.
- Phelps, R. M., B. E. Johnson, et al. (1996). "NCI-Navy Medical Oncology Branch cell line data base." J Cell Biochem Suppl **24**: 32-91.
- Polak, P. and E. Domany (2006). "Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes." BMC Genomics **7**(1): 133.
- Pollack, J. R., C. M. Perou, et al. (1999). "Genome-wide analysis of DNA copy-number changes using cDNA microarrays." Nat Genet **23**(1): 41-6.
- Pompeia, C., D. R. Hodge, et al. (2004). "Microarray analysis of epigenetic silencing of gene expression in the KAS-6/1 multiple myeloma cell line." Cancer Res **64**(10): 3465-73.
- Pradhan, S., A. Bacolla, et al. (1999). "Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation." J Biol Chem **274**(46): 33002-10.
- Pruitt, K., R. L. Zinn, et al. (2006). "Inhibition of SIRT1 reactivates silenced cancer genes without loss of promoter DNA hypermethylation." PLoS Genet **2**(3): e40.
- Pu, R. T., L. E. Laitala, et al. (2003). "Methylation profiling of benign and malignant breast lesions and its application to cytopathology." Mod Pathol **16**(11): 1095-101.
- Ragoczy, T., A. Telling, et al. (2003). "A genetic analysis of chromosome territory looping: diverse roles for distal regulatory elements." Chromosome Res **11**(5): 513-25.
- Ramirez, R. D., S. Sheridan, et al. (2004). "Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins." Cancer Res **64**(24): 9027-34.
- Rao, A. and O. Avni (2000). "Molecular aspects of T-cell differentiation." Br Med Bull **56**(4): 969-84.
- Rauhala, H. E., K. P. Porkka, et al. (2005). "Dual-specificity phosphatase 1 and serum/glucocorticoid-regulated kinase are downregulated in prostate cancer." Int J Cancer **117**(5): 738-45.
- Reik, W. and A. Lewis (2005). "Co-evolution of X-chromosome inactivation and imprinting in mammals." Nat Rev Genet **6**(5): 403-10.
- Reiner, S. L. (2003). "Immunity and the animation of the genome." Immunity **19**(6): 775-80.
- Reiner, S. L., A. C. Mullen, et al. (2003). "Helper T cell differentiation and the problem of cellular inheritance." Immunol Res **27**(2-3): 463-8.
- Rhee, I., K. E. Bachman, et al. (2002). "DNMT1 and DNMT3b cooperate to silence genes in human cancer cells." Nature **416**(6880): 552-6.
- Rhee, I., K. W. Jair, et al. (2000). "CpG methylation is maintained in human cancer cells lacking DNMT1." Nature **404**(6781): 1003-7.
- Ringrose, L. and R. Paro (2004). "Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins." Annu Rev Genet **38**: 413-43.
- Robert, M. F., S. Morin, et al. (2003). "DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells." Nat Genet **33**(1): 61-5.

- Robinson, P. N., U. Bohme, et al. (2004). "Gene-Ontology analysis reveals association of tissue-specific 5' CpG-island genes with development and embryogenesis." *Hum Mol Genet* **13**(17): 1969-78.
- Ronaghi, M. (2001). "Pyrosequencing sheds light on DNA sequencing." *Genome Res* **11**(1): 3-11.
- Rountree, M. R., K. E. Bachman, et al. (2000). "DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci." *Nat Genet* **25**(3): 269-77.
- Santarosa, M. and A. Ashworth (2004). "Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way." *Biochim Biophys Acta* **1654**(2): 105-22.
- Sato, M., Y. Horio, et al. (2002). "The expression of DNA methyltransferases and methyl-CpG-binding proteins is not associated with the methylation status of p14(ARF), p16(INK4a) and RASSF1A in human lung cancer cell lines." *Oncogene* **21**(31): 4822-9.
- Sato, M., M. B. Vaughan, et al. (2006). "Multiple Oncogenic Changes (K-RASV12, p53 Knockdown, Mutant EGFRs, p16 Bypass, Telomerase) Are Not Sufficient to Confer a Full Malignant Phenotype on Human Bronchial Epithelial Cells." *Cancer Res* **66**(4): 2116-2128.
- Sato, N., N. Fukushima, et al. (2003). "Discovery of novel targets for aberrant methylation in pancreatic carcinoma using high-throughput microarrays." *Cancer Res* **63**(13): 3735-42.
- Sato, N., A. Maitra, et al. (2003). "Frequent hypomethylation of multiple genes overexpressed in pancreatic ductal adenocarcinoma." *Cancer Res* **63**(14): 4158-66.
- Schubeler, D., C. Francastel, et al. (2000). "Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus." *Genes Dev* **14**(8): 940-50.
- Schulmann, K., A. Sterian, et al. (2005). "Inactivation of p16, RUNX3, and HPP1 occurs early in Barrett's-associated neoplastic progression and predicts progression risk." *Cancer Res* **65**(25): 4138-4148.
- Sekido, Y., K. M. Fong, et al. (2003). "Molecular genetics of lung cancer." *Annu Rev Med* **54**: 73-87.
- Shames, D., Girard, L, Gao, B, Sato, M, Lewis, CM, Shivapurkar, N, Jiang, A, Perou, CM, Kim, YH, Pollack, JR, Fong, KM, Lam, C, Wong, M, Shyr, Y, Nanda, R, Olopade, OI, Gerald, W, Euhus, DM, Shay, JW, Gazdar, AF, Minna, JD (2006). "A Genome-wide Screen for Hypermethylated Genes in Lung Cancer Identifies Tumor-Specific Methylation Markers for Multiple Malignancies." *PLoS Med* **3**.
- Sharp, P. A. (2001). "RNA interference--2001." *Genes Dev* **15**(5): 485-90.
- Shi, Y. (2003). "Mammalian RNAi for the masses." *Trends Genet* **19**(1): 9-12.
- Shigematsu, H. and A. F. Gazdar (2006). "Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers." *Int J Cancer* **118**(2): 257-62.
- Shigematsu, H., L. Lin, et al. (2005). "Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers." *J Natl Cancer Inst* **97**(5): 339-46.
- Shivapurkar, N., V. Stastny, et al. (2006). "Application of a methylation gene panel by quantitative PCR for lung cancers." *Cancer Lett*.
- Shivapurkar, N., V. Stastny, et al. (2005). "Novel real-time PCR assay using a universal molecular marker for diagnosis of hematologic cancers." *Int J Cancer* **116**(4): 656-60.
- Shivapurkar, N., T. Takahashi, et al. (2004). "Presence of simian virus 40 DNA sequences in human lymphoid and hematopoietic malignancies and their relationship to aberrant promoter methylation of multiple genes." *Cancer Res* **64**(11): 3757-60.

- Shivapurkar, N., S. Toyooka, et al. (2004). "Aberrant methylation of trail decoy receptor genes is frequent in multiple tumor types." *Int J Cancer* **109**(5): 786-92.
- Simpson, A. J., O. L. Caballero, et al. (2005). "Cancer/testis antigens, gametogenesis and cancer." *Nat Rev Cancer* **5**(8): 615-25.
- Sledz, C. A., M. Holko, et al. (2003). "Activation of the interferon system by short-interfering RNAs." *Nat Cell Biol* **5**(9): 834-9.
- Smiraglia, D. J. and C. Plass (2002). "The study of aberrant methylation in cancer via restriction landmark genomic scanning." *Oncogene* **21**(35): 5414-26.
- Smith, L. T., M. Lin, et al. (2006). "Epigenetic regulation of the tumor suppressor gene TCF21 on 6q23-q24 in lung and head and neck cancer." *Proc Natl Acad Sci U S A* **103**(4): 982-7.
- Stresemann, C., B. Brueckner, et al. (2006). "Functional diversity of DNA methyltransferase inhibitors in human cancer cell lines." *Cancer Res* **66**(5): 2794-800.
- Suzuki, H., E. Gabrielson, et al. (2002). "A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer." *Nat Genet* **31**(2): 141-9.
- Suzuki, H., F. Itoh, et al. (1999). "Distinct methylation pattern and microsatellite instability in sporadic gastric cancer." *Int J Cancer* **83**(3): 309-13.
- Suzuki, K., I. Suzuki, et al. (2006). "Global DNA demethylation in gastrointestinal cancer is age dependent and precedes genomic damage." *Cancer Cell* **9**(3): 199-207.
- Suzuki, M., N. Sunaga, et al. (2004). "RNA interference-mediated knockdown of DNA methyltransferase 1 leads to promoter demethylation and gene re-expression in human lung and breast cancer cells." *Cancer Res* **64**(9): 3137-43.
- Suzuki, M., S. Toyooka, et al. (2005). "Aberrant methylation profile of human malignant mesotheliomas and its relationship to SV40 infection." *Oncogene* **24**(7): 1302-8.
- Szyf, M. (2001). "Towards a pharmacology of DNA methylation." *Trends Pharmacol Sci* **22**(7): 350-4.
- Takahashi, T., H. Shigematsu, et al. (2006). "Aberrant promoter methylation of multiple genes during multistep pathogenesis of colorectal cancers." *Int J Cancer* **118**(4): 924-31.
- Takahashi, T., M. Suzuki, et al. (2005). "Aberrant methylation of Reprimo in human malignancies." *Int J Cancer* **115**(4): 503-10.
- Takai, D. and P. A. Jones (2002). "Comprehensive analysis of CpG islands in human chromosomes 21 and 22." *Proc Natl Acad Sci U S A* **99**(6): 3740-5.
- Tang, X., H. Shigematsu, et al. (2005). "EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients." *Cancer Res* **65**(17): 7568-72.
- Taylor, S. M. and P. A. Jones (1985). "Cellular differentiation." *Int J Obes* **9 Suppl 1**: 15-21.
- Thorgeirsson, S. S. and J. W. Grisham (2002). "Molecular pathogenesis of human hepatocellular carcinoma." *Nat Genet* **31**(4): 339-46.
- Thorgeirsson, S. S., J. S. Lee, et al. (2006). "Functional genomics of hepatocellular carcinoma." *Hepatology* **43**(2 Suppl 1): S145-50.
- Ting, A. H., K. W. Jair, et al. (2006). "Differential requirement for DNA methyltransferase 1 in maintaining human cancer cell gene promoter hypermethylation." *Cancer Res* **66**(2): 729-35.

- Ting, A. H., K. W. Jair, et al. (2004). "CpG island hypermethylation is maintained in human colorectal cancer cells after RNAi-mediated depletion of DNMT1." Nat Genet **36**(6): 582-4.
- Ting, A. H., K. E. Schuebel, et al. (2005). "Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation." Nat Genet **37**(8): 906-10.
- Tokumaru, Y., K. Yamashita, et al. (2004). "Inverse correlation between cyclin A1 hypermethylation and p53 mutation in head and neck cancer identified by reversal of epigenetic silencing." Cancer Res **64**(17): 5982-7.
- Tolhuis, B., R. J. Palstra, et al. (2002). "Looping and interaction between hypersensitive sites in the active beta-globin locus." Mol Cell **10**(6): 1453-65.
- Tomizawa, Y., Y. Sekido, et al. (2001). "Inhibition of lung cancer cell growth and induction of apoptosis after reexpression of 3p21.3 candidate tumor suppressor gene SEMA3B." Proc Natl Acad Sci U S A **98**(24): 13954-9.
- Tommasi, S., R. Dammann, et al. (2005). "Tumor susceptibility of Rassf1a knockout mice." Cancer Res **65**(1): 92-8.
- Tong Ihn Lee, R. G. J., Laurie A. Boyer, Matthew G. Guenther, Stuart S. Levine, Roshan M. Kumar, Brett Chevalier, Sarah E. Johnstone, Megan F. Cole, Kyo-ichi Isono, Haruhiko Koseki, Takuya Fuchikami, Kuniya Abe, Heather L. Murray, Jacob P. Zucker, Bingbing Yuan, George W. Bell, Elizabeth Herbolsheimer, Nancy M. Hannett, Kaiming Sun, Duncan T. Odom, Arie P. Otte, Thomas L. Volkert, David P. Bartel, Douglas A. Melton, David K. Gifford, Rudolf Jaenisch, Richard A. Young (2006). "Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells." Cell **125**: 301-313.
- Toyooka, K. O., S. Toyooka, et al. (2002). "Establishment and validation of real-time polymerase chain reaction method for CDH1 promoter methylation." Am J Pathol **161**(2): 629-34.
- Toyooka, S., M. Carbone, et al. (2002). "Progressive aberrant methylation of the RASSF1A gene in simian virus 40 infected human mesothelial cells." Oncogene **21**(27): 4340-4.
- Toyooka, S., K. O. Toyooka, et al. (2001). "DNA methylation profiles of lung tumors." Mol Cancer Ther **1**(1): 61-7.
- Toyota, M., N. Ahuja, et al. (1999). "CpG island methylator phenotype in colorectal cancer." Proc Natl Acad Sci U S A **96**(15): 8681-6.
- Toyota, M., N. Ahuja, et al. (1999). "Aberrant methylation in gastric cancer associated with the CpG island methylator phenotype." Cancer Res **59**(21): 5438-42.
- Turner, B. M. (2000). "Histone acetylation and an epigenetic code." Bioessays **22**(9): 836-45.
- Turner, B. M. (2002). "Cellular memory and the histone code." Cell **111**(3): 285-91.
- Tuschl, T. (2001). "RNA interference and small interfering RNAs." ChemBiochem **2**(4): 239-45.
- Ueki, T., M. Toyota, et al. (2000). "Hypermethylation of multiple genes in pancreatic adenocarcinoma." Cancer Res **60**(7): 1835-9.
- Urnov, F. D. and A. P. Wolffe (2001). "Above and within the genome: epigenetics past and present." J Mammary Gland Biol Neoplasia **6**(2): 153-67.
- Usary, J., V. Llaca, et al. (2004). "Mutation of GATA3 in human breast tumors." Oncogene **23**(46): 7669-78.

- Vatolin, S., Z. Abdullaev, et al. (2005). "Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes." Cancer Res **65**(17): 7751-62.
- Vaughan, M. B., R. D. Ramirez, et al. (2006). "A three-dimensional model of differentiation of immortalized human bronchial epithelial cells." Differentiation **74**(4): 141-148.
- Velicescu, M., D. J. Weisenberger, et al. (2002). "Cell division is required for de novo methylation of CpG islands in bladder cancer cells." Cancer Res **62**(8): 2378-84.
- Verma, U. N., R. M. Surabhi, et al. (2003). "Small Interfering RNAs Directed against beta-Catenin Inhibit the in Vitro and in Vivo Growth of Colon Cancer Cells." Clin Cancer Res **9**(4): 1291-300.
- Wakimoto, B. T. (1998). "Beyond the nucleosome: epigenetic aspects of position-effect variegation in *Drosophila*." Cell **93**(3): 321-4.
- Walsh, C. P. and T. H. Bestor (1999). "Cytosine methylation and mammalian development." Genes Dev **13**(1): 26-34.
- Walsh, C. P., J. R. Chaillet, et al. (1998). "Transcription of IAP endogenous retroviruses is constrained by cytosine methylation." Nat Genet **20**(2): 116-7.
- Weber, M., J. J. Davies, et al. (2005). "Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells." Nat Genet **37**(8): 853-62.
- Willingham, A. T. and T. R. Gingeras (2006). "TUF Love for "Junk" DNA." Cell **125**(7): 1215-1220.
- Wistuba, II, L. Mao, et al. (2002). "Smoking molecular damage in bronchial epithelium." Oncogene **21**(48): 7298-306.
- Wistuba, I. I. and A. F. Gazdar (2006). "LUNG CANCER PRENEOPLASIA." Annual Review of Pathology: Mechanisms of Disease **1**(1): 331-348.
- Wolffe, A. P. (2001). "Transcriptional regulation in the context of chromatin structure." Essays Biochem **37**: 45-57.
- Xu, G. L., T. H. Bestor, et al. (1999). "Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene." Nature **402**(6758): 187-91.
- Yang, B., M. Guo, et al. (2003). "Aberrant promoter methylation profiles of tumor suppressor genes in hepatocellular carcinoma." Am J Pathol **163**(3): 1101-7.
- Yang, J., S. A. Mani, et al. (2004). "Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis." Cell **117**(7): 927-39.
- Yoder, J. A., N. S. Soman, et al. (1997). "DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe." J Mol Biol **270**(3): 385-95.
- Yoder, J. A., C. P. Walsh, et al. (1997). "Cytosine methylation and the ecology of intragenomic parasites." Trends Genet **13**(8): 335-40.
- Zhong, S., M. W. Tang, et al. (2002). "Silencing of GSTP1 Gene by CpG Island DNA Hypermethylation in HBV-associated Hepatocellular Carcinomas." Clin Cancer Res **8**(4): 1087-1092.
- Zingg, J. M. and P. A. Jones (1997). "Genetic and epigenetic aspects of DNA methylation on genome expression, evolution, mutation and carcinogenesis." Carcinogenesis **18**(5): 869-82.
- Zochbauer-Muller, S., K. M. Fong, et al. (2001). "Aberrant promoter methylation of multiple genes in non-small cell lung cancers." Cancer Res **61**(1): 249-55.

Zochbauer-Muller, S., S. Lam, et al. (2003). "Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers." Int J Cancer **107**(4): 612-6.

Zochbauer-Muller, S., J. D. Minna, et al. (2002). "Aberrant DNA methylation in lung cancer: biological and clinical implications." Oncologist **7**(5): 451-7.

VITA

David Stuart Shames was born in Johannesburg, South Africa on January 15, 1975, the son of Marcella Elaine and Victor Roger Shames. After completing his degree at Aiglon College in Chesièries-Villars, Switzerland in 1993, he entered the University of Pennsylvania where he majored in environmental studies with a concentration in geology. He left Penn in 1997, to join Macsteel International where he worked in the trading department while completing his senior thesis. He received a Bachelor of Arts degree from the University of Pennsylvania in 1998. After leaving Macsteel in 1999, David matriculated into Columbia University's post-baccalaureate pre-medical program and volunteered in the laboratory of Cy Stein in the Department of Pharmacology at Columbia College of Physicians and Surgeons. In June 2001, David entered the Graduate School of Biomedical Sciences at the University of Texas Southwestern Medical Center at Dallas. He performed his thesis work under the tutelage of Dr. John D. Minna. David received his Ph.D. in September 2006.

Permanent Address: 7421 Frankford Rd

Apt. 2032

Dallas, TX 75252