INTERPLAY BETWEEN DNA METHYLATION AND CHROMATIN STRUCTURE DURING THE EPIGENETIC RESILENCING OF TUMOR SUPPRESSOR GENES

By

CAROLINA E. PARDO BUITRAGO

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2011

To Juanita Pardo
My favorite person in the world

ACKNOWLEDGMENTS

I am extremely grateful to my mentor Dr. Michael Kladde, Mike, for trusting me with MAPit. Mike believed in my abilities and guided me in the process of becoming a scientist with great enthusiasm and patience. I admire his intense passion for science and research. Mike has supported me through every step during the last five years. During very difficult personal times he has always been understanding and generous. Thank you Mike, I will always be grateful to you.

I am very grateful to the members of my committee Dr. Kevin Brown, Dr. Jörg Bungert and Dr. Jim Resnick, for all their comments, suggestions and patience.

There are no words to express how grateful I am to the members of the Kladde lab: Nancy Nabilsi, Russell Darst, Amber Delmas and Mayank Talwar. You have become my extended family and I love you all very much. You are the most amazing group of scientists and friends that anyone can wish for. I am especially grateful to Russell for being a wonderful, constant and extremely patient friend. I am deeply grateful to Nancy “Pants” for all your help and unconditional support and friendship, you are super awesome. Sharing this time with you has made me a happy person.

My deepest gratitude goes to my family for their unflagging support and love. Your pride in me is humbling and inspires me to be a better person.

Finally, I am very grateful to all my friends and teachers at Texas A&M and to all my friends, teachers and staff at the Biochemistry and Molecular Biology Department, at Shands Cancer Center and at the Interdisciplinary Program in Biomedical Sciences at the University of Florida.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CDH1	Classical cadherin 1 (E-cadherin)
CGI	CG island
DNMT	DNA methyltransferase
DTT	Dithiothreitol
EPM2AIP1	EPM2A laforin-interacting protein
ES cells	Embryonic stem cells
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
M.CviPI	First DNA methyltransferase from *Chlorella virus* P
LINE1	Long interspersed nuclear element 1
MAPit	Methyltransferase accessibility protocol for individual templates
MBD	Methyl CG-binding domain
MBP	Methyl CG-binding protein
MSD	Myelodysplastic syndrome
MLH1	MutL homolog 1
M.SssI	First DNA methyltransferase from *Spiroplasma* sp. strain MQ1
ncRNA	Non-coding RNA
NDR	Nucleosome-depleted region
NFR	Nucleosome-free region
PCNA	Proliferating cell nuclear antigen
PRC1/2	Polycomb repressive complex 1 or 2
PMSF	Phenylmethylsulfonyl fluoride
SAM	*S*-adenosyl-*L*-methinonine
SIM2	Single minded 2 gene

TIMP3	Tissue inhibitor of metalloproteinases 3
TSA	Trichostatin A
TSG	Tumor suppressor gene
TSS	Transcription start site
UHRF1	Ubiquitin-like plant homeodomain and RING finger domain-containing protein 1

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

INTERPLAY BETWEEN DNA METHYLATION AND CHROMATIN STRUCTURE DURING THE EPIGENETIC RESILENCING OF TUMOR SUPPRESSOR GENES

By

Carolina E. Pardo Buitrago

August 2011

Chair: Michael P. Kladde
Major: Medical Sciences – Biochemistry and Molecular Biology

Traditionally, cancer has been viewed as a genetic disease, characterized by aberrant gene expression patterns. However, since the discovery of epigenetic phenomena and their heritable role in controlling gene expression, interest in how DNA methylation and chromatin organization modulate gene transcription has exploded and highlighted the relevance of epigenetic mechanisms in the initiation and progression of human cancer. The most studied epigenetic event contributing to carcinogenesis is the transcriptional silencing of tumor suppressor genes. So far, epigenetic marks (i.e. DNA hypermethylation and histone modifications) have strongly been associated with repressive or permissive transcriptional states. DNA hypermethylation and the presence of repressive histone marks have been strongly associated with aberrantly-silenced TSG promoters in cancer cells. On the other hand, there is little information regarding the interplay between DNA methylation and nucleosome positioning, which regulates transcription by impeding access of trans-acting factors to their regulatory sequences. This is due in part to a lack of techniques that can simultaneously query both epigenetic features. Therefore, in Chapter 2, we developed MAPit, a method for simultaneous mapping of endogenous mammalian DNA methylation and chromatin accessibility at the

single-molecule level. In Chapter 3, in collaboration with researchers at the University of Leeds, we developed a computational resource for rapid analysis and visual representation of MAPit datasets. Chapter 4 uses MAPit in part to address a medically relevant question, namely what changes in gene expression and chromatin structure occur when 5-aza-2'-deoxycytidine, a chemotherapeutic agent that demethylates DNA, is added and then removed from cultured cancer cells. Differences in chromatin structure and DNA methylation were first mapped at specific loci in colorectal cancer cell lines with either actively transcribed or silenced TSGs. Then, we used MAPit to investigate how treatment with 5-aza-dC and its subsequent withdrawal altered DNA methylation and chromatin accessibility as TSGs were resilenced. Our data support a model in which changes in chromatin accessibility rather than endogenous DNA methylation drive the epigenetic resilencing of TSGs. Future studies would be aimed at identifying and interfering with molecular events that promote epigenetic resilencing with the goal of increasing the long-range efficacy of epigenetic chemotherapy.

CHAPTER 1
INTRODUCTION

Epigenetic Events and Cancer

Cancer is a stepwise process characterized mainly by aberrant gene expression patterns. The genetic basis of tumor development and progression in cancer has been firmly established; accumulation and selection of mutations leading to oncogenic activation, tumor suppressor gene (TSG) inactivation and chromosomal aberrations (Baylin and Ohm, 2006; Feinberg et al., 2006; Hanahan and Weinberg, 2000; Stratton et al., 2009; Vogelstein and Kinzler, 2004). Epigenetic aberrations, on the other hand, have only recently become recognized as major mediators of carcinogenesis (Baylin and Ohm, 2006; Sawan et al., 2008). Epigenetic abnormalities have been reported in almost every cancer type and stage studied to date.

Our understanding of how epigenetic events such as DNA methylation, histone modifications and chromatin remodeling affect gene expression has greatly increased during the last ten years. This knowledge has revealed that epigenetic regulation is a primary determinant in cancer development and progression (Baylin and Ohm, 2006; Feinberg et al., 2006; Jones and Baylin, 2002, 2007; Sawan et al., 2008; Tarakhovsky, 2010; Ting et al., 2006). Certain epigenetic events, e.g. gene specific hypermethylation patterns, seem to be cancer type specific, while others like global changes in histone modifications and global DNA hypomethylation are shared between cancer types. Interestingly, and in accordance with lessons learned from genetic aberrations in cancer, DNA hypermethylation-dependent TSG silencing affects the same pathways that are deregulated by genetic mutation in cancer (Sawan et al., 2008; Vaissiere et al., 2008; Vogelstein and Kinzler, 2004). As opposed to genetic aberrations, alterations in

the epigenetic layers are characterized by gradual appearance and potentially more facile reversibility, offering a promising arena for improvements in cancer diagnosis, therapy and prevention (Jones and Baylin, 2007; Sincic and Herceg, 2011). Figure 1-1 depicts epigenetic networks that are deregulated in human cancer.

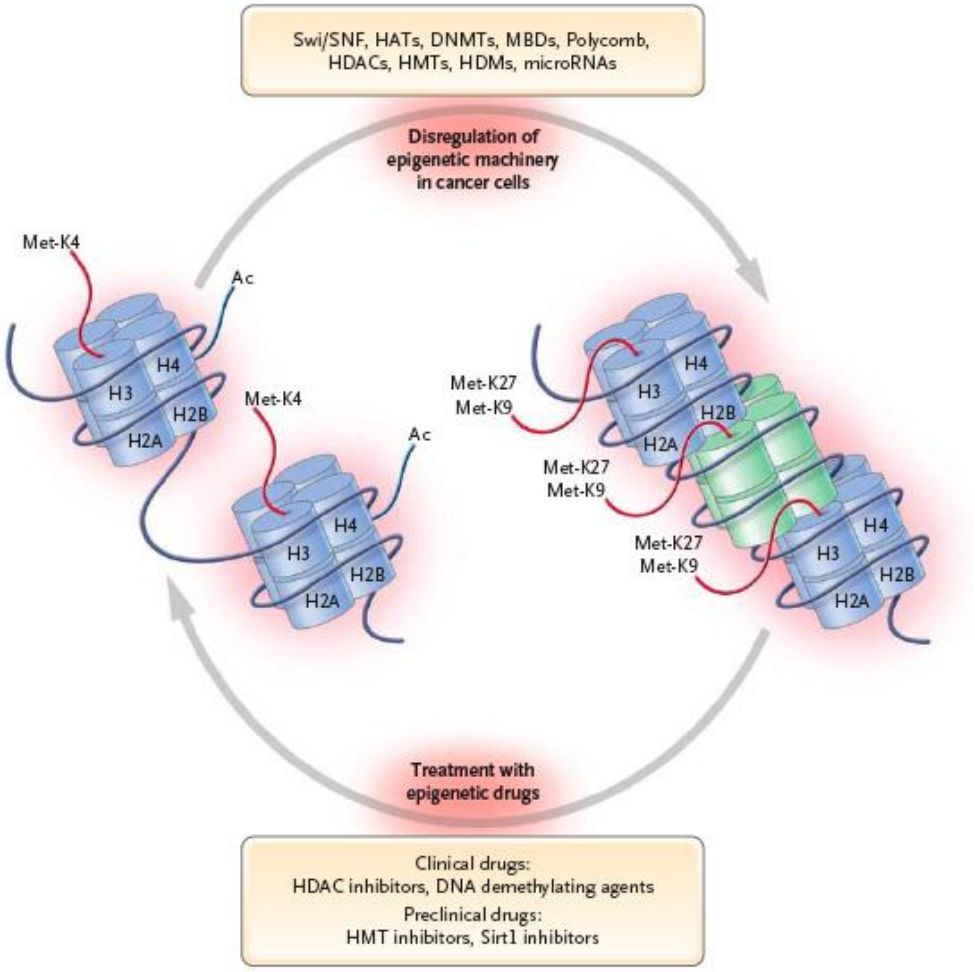

Figure 1-1. Disruption of epigenetic networks in human cancer. Epigenetic events contribute to the establishment of transcriptional permissive (left) and repressive (right) chromatin environments. The box at the top lists examples of members of the epigenetic machinery that, when dysregulated, promote formation of abnormal repressive chromatin environments (e.g. at TSGs). Unlike genetic mutations in cancer, epigenetic events are potentially more reversible. The box at the bottom lists examples of agents that can be used to reestablish an open chromatin environment that is permissive for transcription (Adapted from: Esteller, 2008).

Epigenetic Marks and Tumor Suppressor Gene Silencing in Cancer

Epigenetic gene silencing is a widely-studied epigenetic phenomenon. Epigenetic changes are essential for normal vertebrate functions, such as embryonic development, tissue differentiation and cell identity (Jones and Baylin, 2007; Sincic and Herceg, 2011; Vaissiere et al., 2008). DNA methylation associated transcriptional silencing is required for X-chromosome inactivation, imprinting, and genome protection from parasitic elements. Hence, dysregulation of the epigenetic landscape can lead to disease states (Jones and Baylin, 2007; Portela and Esteller, 2010). A well-characterized example of aberrant epigenetic regulation causing disease is a common cancer phenomenon leading to the transcriptional inactivation of TSGs. TSGs are master regulators of cell functions such as proliferation and survival, cell cycle progression, chromatin remodeling, DNA repair and apoptosis (McCabe et al., 2009; Sawan et al., 2008).

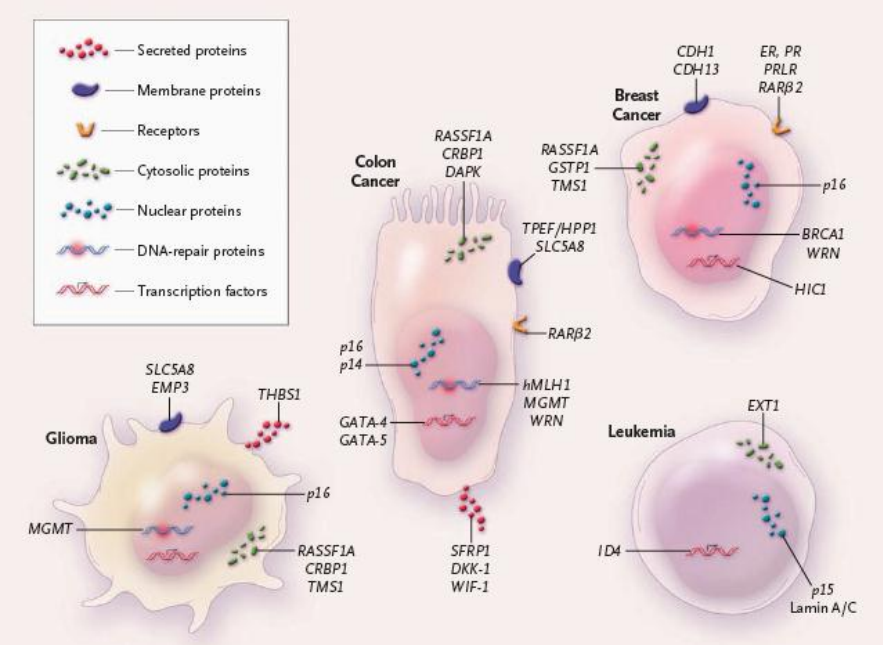

Figure 1-2. Human TSGs affected by promoter CGI hypermethylation in human cancer. This figure shows a non-comprehensive list of TSGs commonly affected by hypermethylation-dependent silencing in different cancer types (Adapted from: Esteller, 2008).

Recently, DNA repair genes have engendered special interest. Due to their function as “guardians of the genome”, epigenetic silencing of genes encoding DNA repair functions exemplifies a means by which epigenetic changes can precede and be causative of genetic change (Sawan et al., 2008; Ushijima and Asada, 2010). Recent studies have shown that a large number of TSGs is silenced in a DNA hypermethylation-dependent manner in other malignancies besides cancer (Portela and Esteller, 2010; Rodriguez-Paredes and Esteller, 2011).

As with genetic mutations, the epigenetic inactivation of TSGs confers on tumor cells a proliferative advantage, increased genetic and epigenetic instability, and may predispose them to metastasis (Herman and Baylin, 2003; Robertson, 2005). TSG silencing has been reported in virtually every cancer studied to date. In cancers where progression is defined well, e.g. cancers of the colon, aberrant TSG hypermethylation is found in the earliest precursor lesions, placing it as an early step, possibly causative, in cancer progression (Baylin and Ohm, 2006; Harris and McCormick, 2010; Jones and Baylin, 2007; Rodriguez-Paredes and Esteller, 2011; Sharma et al., 2010; Taby and Issa, 2010; Vaissiere et al., 2008). TSG silencing has been studied from the perspective of different epigenetic processes like DNA methylation, histone modifications and chromatin remodeling, because each of these events can affect chromatin structure and thus the transcriptional state of genes.

DNA Methylation and Transcriptional Regulation

In mammals, the best-characterized epigenetic phenomenon is cytosine methylation in DNA. This covalent modification of DNA occurs at the C-5 position of cytosine (m^5C) residues, almost exclusively in the context of CpG (hereafter, CG) dinucleotides. Mammalian genomes are globally depleted of CG dinucleotides. This can

be explained by the frequent occurrence of spontaneous deamination of m^5C to thymine and the subsequent accumulation of these mutations in the genome during evolution (Bird, 1986). As a result only 21% of the expected amount of CG dinucleotides is present in the human genome (Lander et al., 2001). However, interspersed in this low CG background, there are conspicuous CG-rich regions, known as CG islands (CGI), where the C+G frequency is closer to that expected when accounting for genomic GC content. Depending of the algorithm used to define CGIs, approximately 60-70% of human gene promoters contain CGIs. CGIs range from 500 bp to several kilobases in length, and are usually localized between –2 kb and +1 kb of the transcription start site (TSS) of genes (Gardinergarden and Frommer, 1987; Takai and Jones, 2002; Wang and Leung, 2004).

Interestingly, most, if not all, housekeeping genes have been shown to possess at least one promoter CGI, while the promoters of only 40% of genes with a tissue-restricted expression profile have CGIs (Saxonov et al., 2006; Weber et al., 2007; Zhu et al., 2008). While most CG dinucleotides distributed throughout the genome are methylated, promoter associated CGIs are maintained methylation free and are characterized by transcriptional-permissive chromatin states (Figure 1-3). In this fashion, expression of a CGI-associated gene can easily occur when the appropriate transcription factors are present (Weber et al., 2007; Zhu et al., 2008). Transcriptional permissiveness is associated with lack of methylation at promoter CGIs, while hypermethylation at promoter CGIs has been strongly associated with stable transcriptional silencing (Bird, 2002; Bird and Wolffe, 1999; Esteller, 2002, 2007b; Herman and Baylin, 2003; Sincic and Herceg, 2011).

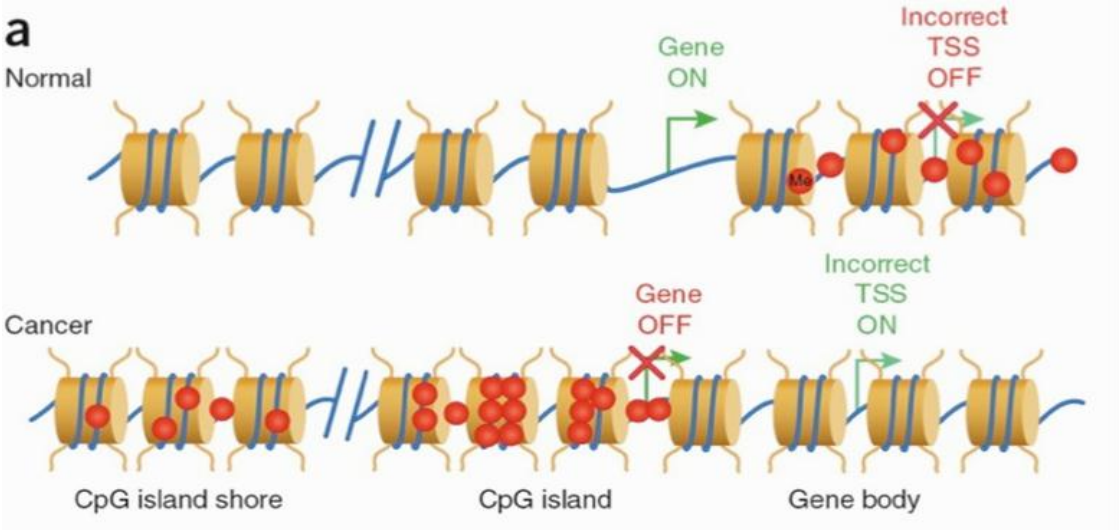

Figure 1-3. Characteristic DNA methylation and chromatin structures at human CGIs in normal and cancer cells. In normal cells, CGIs are maintained DNA methylation free and in an open chromatin structure, associated with active transcription. In cancer cells, CGIs show DNA hypermethylation and closed chromatin structure, associated with transcriptional repression. CGI shores are maintained DNA methylation free in normal cells and accumulate DNA hypermethylation in cancer states (Adapted from: Rodriguez-Paredes and Esteller, 2011).

As recently described by Irizarry et al., CGI “shores” are also involved in epigenetic transcriptional regulation. CGI shores refer to low CG density areas located as far as ~2 kb from CGIs, which are subjected to tissue-specific differential DNA methylation (Irizarry et al., 2009). CGI shores may influence alternative TSS usage (Rauch et al., 2009). In cancer cells, a strong association between TSG silencing and CGI shore methylation has been described, this association is independent of the DNA methylation status of the proximal CGI (Irizarry et al., 2009).

It is important to note that not all CGIs co-localize with promoter regions occasionally, they are found inside gene bodies or at the 3' ends of genes. The transcriptional relevance of intragenic CGIs is not clear but several findings suggest that regardless of location, m^5C, plays an important role in transcriptional regulation. So far,

it has been shown that rare transcripts expressed during specific developmental stages originate from intragenic CGI TSSs. Hence, some of these intragenic CGIs may be linked to true TSSs for uncharacterized genes with highly-regulated expression patterns (Gardinergarden and Frommer, 1994; Kleinjan et al., 2004; Macleod et al., 1998; Rauch et al., 2009). Similarly, some intragenic CGIs can serve as TSSs for regulatory non-coding RNAs (ncRNAs) or antisense transcripts that can negatively regulate the sense transcript, for example *HOTAIR* (Rinn et al., 2007), *AIR* (Sleutels et al., 2002) and *XIST* (Panning and Jaenisch, 1996). When not associated with known 5' promoters, CIGs are prone to methylation (Illingworth and Bird, 2009; Nguyen et al., 2001). In the cases mentioned before, methylation of intragenic CGI can serve as a means to maintain transcriptional repression of highly regulated transcripts or of ncRNAs, which when untimely expressed, can affect the transcription of the primary gene. Recent reports suggest that gene body methylation, first described in *Arabidopsis* (Zilberman et al., 2007) and subsequently in mammals (Ball et al., 2009; Hellman and Chess, 2007), is a common event in ubiquitously expressed genes. Gene body methylation has been proposed to have a protective function against spurious transcription and also as a way to promote transcriptional efficiency by elongation (Illingworth and Bird, 2009).

Aberrant patterns of DNA methylation can promote disease. In terms of DNA methylation, cancer genomes are characterized by two main aberrant events: gene specific hypermethylation and global hypomethylation (20-60% overall reduction in m^5C) (Goelz et al., 1985; McCabe et al., 2009; Sincic and Herceg, 2011). Figure 1-4 shows a comparison between the DNA methylation distribution in normal cells and their abnormal cancerous counterparts. Global hypomethylation may very well have been the

first epigenetic alteration to be described in cancer cells (Diala and Hoffman, 1982; Feinberg and Vogelstein, 1983; Goelz et al., 1985). Despite this, and the occurrence of global hypomethylation in every cancer described to date (Feinberg et al., 2006; Jones and Baylin, 2007), its role in cancer initiation and progression remains poorly understood and controversial (Jones and Baylin, 2007; Sincic and Herceg, 2011). Global loss of m^5C in cancer cells is thought to compromise gene repression in areas that are normally transcriptionally inactive or silenced. In normal cells, long repetitive sequences like retrotransposons (Bird, 2002) and centromeres (Bird, 2002; Suzuki and Bird, 2008) are densely methylated and repressed. Hence, the main contribution of hypomethylation to cancer etiology is thought to be through reactivation of retrotransposons (Iskow et al., 2010) and endoparasitic elements (Howard et al., 2008). Also, increased incidence of aberrant chromosomal rearrangements at repetitive sequences (Eden et al., 2003; Gaudet et al., 2003) is thought to promote reactivation of endogenous retroviral sequences and repetitive sequences (Figure 1-1D). These changes have deleterious consequences on genome integrity contributing to genetic instability in the cancer genome. Gene specific hypomethylation can also contribute to cancer progression through the reactivation proto-oncogenes (Watt et al., 2000) and imprinted genes (Ogawa et al., 1993).

Consistent with the positive association between gene body DNA methylation and high levels of transcription, loss of gene body methylation can also contribute to aberrant transcription and gene expression (Figure 1-4C), but more research is needed to assess the true contribution of alterations on gene body methylation to cancer development (Hellman and Chess, 2007).

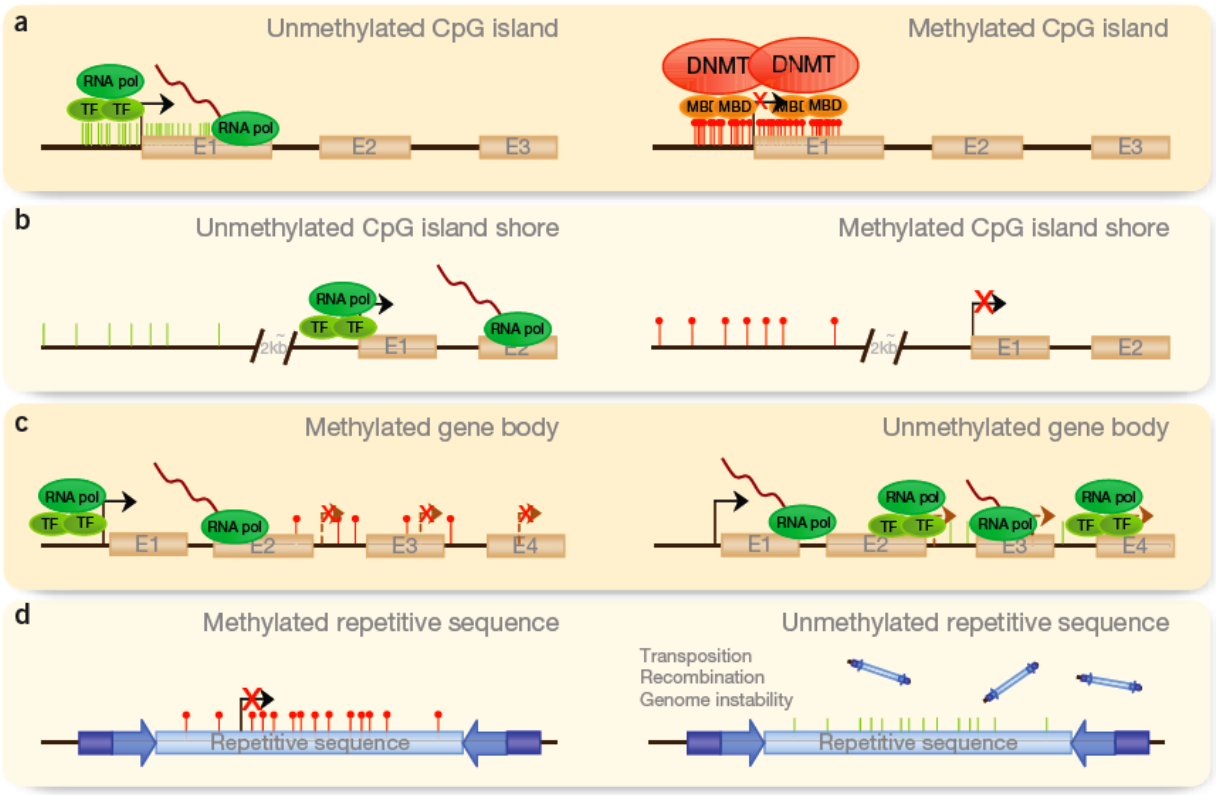

Figure 1-4. DNA methylation patterns in normal and cancer cells. DNA methylation can occur in different regions of the genome. The alteration of these patterns leads to disease. The normal scenario is depicted in the left column and alterations of this pattern are shown on the right. A) CGIs at promoters of genes are normally unmethylated, allowing transcription. Aberrant hypermethylation leads to transcriptional inactivation. B) The same pattern is observed when studying CGI shores. C) However, when methylation occurs at the gene body, it facilitates transcription, preventing spurious transcription initiation. In disease, the gene body tends to demethylate, allowing transcription to be initiated at several incorrect sites. D) Finally, repetitive sequences appear to be hypermethylated, preventing chromosomal instability, translocations and gene disruption through the reactivation of endoparasitic sequences. This pattern is also altered in the diseased state. (Source: Portela and Esteller, 2010)

As mentioned before and discussed in detail in later chapters, aberrant DNA hypermethylation of TSGs promoters and their concomitant transcriptional silencing is a well-documented outcome of aberrant DNA methylation in cancer. As opposed to global DNA hypomethylation, in malignant processes, DNA hypermethylation is a gene specific event affecting the promoters of TSGs (Bird, 2002; Herman and Baylin, 2003). While

CGI and CGI shores are maintained in a methylation-free state in normal cells, during malignant transformation, aberrant DNA methylation can accumulate in both CGIs and/or CGI shores, resulting in transcriptional silencing of the associated gene (Figure 1-4 A-B). Of note, CGI shore DNA methylation can affect TSS selection that can contribute to cancer (Irizarry et al., 2009; Portela and Esteller, 2010; Sincic and Herceg, 2011).

Mechanisms driving the aberrant accumulation of DNA methylation at sites like CGIs, which have been defined by their characteristic transcriptionally permissive environments, remain ill defined (Illingworth and Bird, 2009; McCabe et al., 2009; Portela and Esteller, 2010). Nevertheless, the association of promoter CGI hypermethylation with TSG silencing in cancer has been firmly established. To point, silenced TSGs have proven useful as biomarkers for diagnostic and prognosis as well as to inform treatment of several cancer types (Lima et al., 2010).

Mechanisms of DNA Methylation

DNA methylation is carried out by the DNA methyltransferase family of enzymes that catalyze the transfer of a methyl group from the universal methyl donor *S*-adenosyl-*L*-methionine (SAM) to the 5 position of cytosine residues in DNA (Goll and Bestor, 2005). At least three DNA methyltransferases (DNMTs) are responsible for mammalian DNA methylation patterns (Bestor, 2000; Goll and Bestor, 2005). DNA methyltransferase 1 (DNMT1) was the first mammalian DNMT described (Bestor et al., 1988). Since then, due to some of its properties DNMT1 has been referred to as the maintenance enzyme. DNMT1 shows a 30 to 40-fold preference for hemimethylated DNA (Goyal et al., 2006; Hermann et al., 2004b). In addition, the *DNMT1* gene is

transcribed mostly during S-phase of the cell cycle facilitating the copying of DNA methylation patterns onto the newly-synthesized DNA strand (Robertson et al., 2000).

DNMT1 has been shown to possess *de novo* DNA methylation activity (Goyal et al., 2006; Jeltsch, 2006; Pradhan et al., 1999) and to interact with DNA polymerase proliferating cell nuclear antigen (PCNA), which localizes to replication forks during S phase (Chuang et al., 1997). DNMT1 has also been shown to interact with the ubiquitin-like plant homeodomain and RING finger domain-containing protein 1 (UHRF1). UHRF1 has been shown to bind hemimethylated DNA, possibly recruiting *DNMT1* to sites of DNA hemimethylation, even outside of S phase (Arita et al., 2008; Sharif et al., 2007). Knockout experiments have shown that DNMT1 is responsible for the bulk of genomic methylation and is essential for embryonic development (Li et al., 1992). Disruption of *DNMT1* in cancer cells results in cell death, suggesting that cancer cell survival is dependent on DNMT1 activity (Chen et al., 2007). Nevertheless, complete knock out of DNMT1 in differentiated cells also results in death mediated by p53 (Robertson et al., 2000).

DNMT3a and DNMT3b are referred to as *de novo* enzymes and are thought to be responsible for establishing DNA methylation patterns, especially during early embryogenesis and germ cell development (Chen et al., 2003; Okano et al., 1999). The *de novo* DNMTs are highly expressed in embryonic stem (ES) cells and are downregulated in differentiated cells (Esteller, 2007b). Another member of this protein family, DNMT2, shows weak DNA methylation activity *in vitro*. Targeted disruption of the *Dnmt2* gene in mouse ES cells did not show an altered DNA methylation phenotype, suggesting that it might not be involved in establishing DNA methylation patterns (Goll

and Bestor, 2005). Recent work has shown that it is involved in tRNA methylation (Goll et al., 2006). DNA methyltransferase 3 like (DNMT3L) has been described as an enzymatically-inactive regulatory factor that stimulates DNMT3a and DNMT3b activity. DNMT3L has been shown to colocalize and interact directly with them in the nucleus DNMT3L is required for maternal imprinting and has an expression pattern similar to DNMT3a and DNMT3b (Holz-Schietinger and Reich, 2010; Zhao-Xia Chen, 2005). Members of the DNMT3 subfamily can catalyze non-CG methylation and at CA and CT dinucleotide sites (Ramsahoye et al., 2000). As much as 25% of the DNA methylation content of embryonic cells is found outside of CG sites. Non-CG methylation is preferentially localized to the transcribed strand of active genes (Lister et al., 2009). Disruption of maternally-inherited alleles of *Dnmt3a* and *Dnmt3b* is lethal due to failure to establish maternal imprinting. Disruption of paternally-inherited alleles results in meiotic defects during spermatogenesis and reactivation of retrotransposons (Chen et al., 2003; Ooi et al., 2007). Targeted deletions of the genes encoding Dnmt1, Dnmt3a and Dnmt3b in mouse models have demonstrated that these enzymes are essential for normal development (Li et al., 1992; Okano et al., 1999).

Mechanisms of establishment and maintenance of DNA methylation patterns in mammalian cells were initially explained by the existence of a maintenance enzyme and *de novo* enzymes; however, several observations suggest that the mechanisms are more complex than previously thought. The catalytically active DNMTs have both *de novo* and maintenance activities and show redundancy in certain functions (Hermann et al., 2004a; Jones and Liang, 2009; Sincic and Herceg, 2011). It has been shown that DNMT3A and DNMT3B are required for the maintenance of DNA methylation in specific

sequences. Furthermore, DNMT1 cannot maintain DNA methylation patterns in ES cells that lack DNMT3A and DNMT3B, as methylation is lost after subsequent replications (Liang et al., 2002; Okano et al., 1999). On the other hand, in some cases DNA methylation patterns can be maintained in the absence of DNMT1 (Rhee et al., 2000).

DNA hypermethylation has been strongly correlated with transcriptional gene silencing, both in normal regulation of gene expression programs and in pathological situations like TSG silencing in cancer. There are two proposed mechanisms by which DNA methylation can affect transcription: by impeding transcription factor binding to its cognate sequences (Choy et al., 2010; Klose and Bird, 2006; Knoepfler et al., 1999; Palacios et al., 2010), or through the interaction with methyl-CG-binding proteins (MBPs) (Jones et al., 1998; Nan et al., 1993; Nan et al., 1998; Nan et al., 1997; Zhang et al., 1999). Figure 1-5C shows a DNA methylation-independent gene-silencing event mediated by polycomb repressive complexes (PRCs) that will be discussed later. Recent studies have reported co-localization of DNMTs at promoters of silenced genes with hypermethylated DNA, and several interactions with repressor complexes (Brenner et al., 2005; Burgers et al., 2002; Di Croce et al., 2002; Fuks, 2005; Fuks et al., 2001; Fuks et al., 2003). However, the precise epigenetic mechanisms by which DNA methylation affects transcriptional states remain elusive.

MBPs provided the first evidence for cooperation between epigenetic mechanisms. In 1998. Jones et al., Nan et al. and others showed that MeCP2, an MBP, can recruit histone-modifying enzymes known as histone deacetylases (HDACs) that exert a repressive effect on transcription. A year later, Wade et al. (1999) and Zang et al. (1999) showed that chromatin remodeling complexes also interact with DNA

methylation marks and HDACs (Jones et al., 1998; Nan et al., 1998; Zhang et al., 1999). These early studies seeded the idea that transcriptional gene silencing involving DNA hypermethylation takes place in the context of histone modifications and nucleosome positioning (Figure 1-5B). The combined action of epigenetic events is thought to eventually lead to the formation of stable repressive chromatin environments.

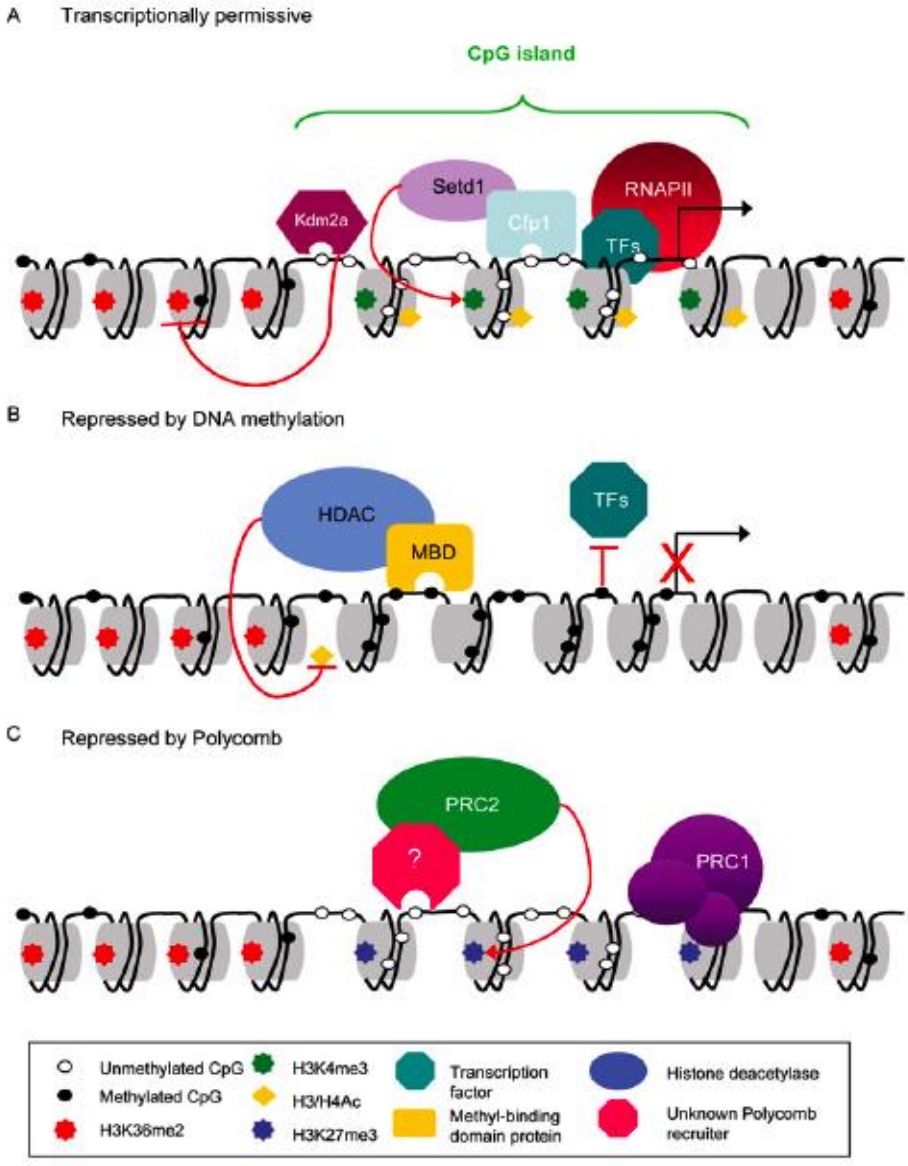

Figure 1-5. Diverse chromatin states at CGIs. A) CGIs usually exist in an unmethylated transcriptionally permissive state. They are marked by acetylation of histones H3 and H4 (H3/H4Ac) and histone H3 lysine 4 trimethylation (H3K4me3), which is directed by Cfp1, and show Kdm2a-dependent depletion of histone

H3 lysine 36 dimethylation (H3K36me2). Nucleosome deficiency and constitutive binding of the preinitiation complex and RNA polymerase II (RNAPII) may also contribute to this transcriptionally permissive state. B) DNA methylation is associated with stable long-term silencing of CGI-containing promoters. This can be mediated by MBD proteins, which recruit corepressor complexes associated with HDAC activity, or may be due to direct inhibition of transcription factor binding by DNA methylation. C) CGIs can also be silenced by PcG proteins and may be key elements involved in polycomb recruitment. An unknown CGI-binding factor could be responsible for recruiting PRC2 to CGIs that then trimethylates H3K27. This H3K27me3 is recognized by PRC1 complexes that act to impede transcriptional elongation, thereby silencing genes. Note that the transcriptionally permissive and polycomb-repressed states can coexist at bivalent CGIs, predominantly in totipotent embryonic cells. (Source: Deaton and Bird, 2011).

Nucleosome Positioning

Nucleosomes are the basic repetitive units of chromatin. The nucleosome core particle, composed of ~147 bp DNA wrapped around an octamer formed by two copies of each of the four core histones (H2A, H2B, H3 and H4), is one of the most thermodynamically stable protein-DNA complexes in eukaryotic cells (Kouzarides, 2007; Li et al., 2007; Liotta and Petricoin, 2000). The histone amino terminal tails project out from the nucleosome core particle, providing a polypeptide substrate for regulatory covalent histone modifications that impart functionality to the nucleosome (Kouzarides, 2007). The way in which chromatin is packaged has proven crucial for DNA-dependent processes like DNA repair and transcriptional regulation. Chromatin organization (nucleosome positioning, occupancy, histone variants and histone modifications) regulate gene expression and are essential in setting transcriptional levels throughout the genome. Additionally, at the level of the organism, cell-to-cell variation is specified by differential activation and repression of genes that defines cell type (Bai and Morozov, 2010; Zhang and Pugh, 2011).

Despite intense research on how nucleosome positioning is determined, our understanding of the factors involved is remains limited. The relationship between DNA sequence and nucleosome positioning has been extensively studied. Unlike DNA binding proteins that bind DNA in a highly sequence-specific manner, nucleosomes do not contain many precise contacts between protein functional groups and specific atoms of the DNA bases (Luger et al., 1997b). This sequence flexibility is likely crucial for nucleosome function and can explain why nucleosomes do not typically adopt a single static position over a DNA sequence. This concept can be illustrated by visualizing a population of cells or chromatin assembled *in vitro*. A nucleosome is designated as well positioned or "phased" if it is spaced from adjacent nucleosomes and occupies a finite range of basepairs within the population, e.g. preferred position varying by ~10 bp. By contrast, a nucleosome is described as not strongly positioned or "fuzzy", if it occupies a random continuous distribution throughout the array (Jiang and Pugh, 2009b).

Figure 1-6. Nucleosome positioning. In a population, nucleosomes can be: well positioned or "phased", occupying a defined range of base pairs at a given locus and constantly maintaining the spacing to the other nucleosomes in the array (left panel). Note that the nucleosome is still moving occupying slightly different translational positions; or randomly or quasi-randomly positioned or "fuzzy" state at a given locus (right panel) (Adapted: Jiang and Pugh, 2009b).

Nucleosome preference for specific DNA-sequences can be determined by the ability of the DNA to bend and alter its helical twist to facilitate wrapping around the histone octamer (Kaplan et al., 2009; Widom, 2001). It has been proposed that there

are two types of DNA sequences when it comes to nucleosome positioning. Sequences that favor nucleosome formation and sequences that are excluded from nucleosomes. It is well known that sequences with high GC content or with AA or TT dinucleotides in periodic 10 bp intervals favor nucleosome formation. In contrast, sequences rich in poly deoxyadenosine tracks (dA•dT in runs of 10 to 20 bp or more) disfavor nucleosome assembly (Kaplan et al., 2009; Lee et al., 2007; Miele et al., 2008; Satchwell et al., 1986; Segal and Widom, 2009; Valouev et al., 2008; Widom, 2001; Yuan et al., 2005). Most of these sequence definitions are based on the periodicity and properties of dinucleotides however longer DNA motifs have also been shown to possess characteristics favorable to nucleosome formation (Valouev et al., 2008).

Supporting the idea that chromatin organization is at least partially encoded by DNA sequence, recent studies have shown that the distribution of nucleosome occupancy determined *in vivo* and *in vitro* are highly similar and that the *in vitro* sequence based models of nucleosome occupancy are highly predictive of *in vivo* nucleosome occupancy (Field et al., 2008; Kaplan et al., 2010; Kaplan et al., 2009; Segal et al., 2006). However, other lines of evidence suggest that DNA sequence is not a major determinant for *in vivo* nucleosome positioning. Zhang et al. (2009) observed substantial differences between *in vivo* and *in vitro* nucleosome positioning maps at the 5' end of coding regions. *In vitro* examination of purified histones assembled onto yeast genomic DNA showed both less pronounced nucleosome-depleted regions (NDRs) and loss of positioned nucleosomes in the nucleosomal array downstream of the NDR. This study proposed that a barrier, most likely established at the core promoter region by a transcriptional-dependent event, determines positioning within the nucleosomal array

(Zhang et al., 2009). This model of “statistical positioning” (Fedor et al., 1988) considers the close packing of nucleosomes into arrays, suggesting that the position of one nucleosome will restrict positioning of others (Jiang and Pugh, 2009b). *In vivo*, both sequence dependent factors and statistical positioning might cooperate to determine global nucleosome positioning.

DNA methylation may also play a role in nucleosome positioning but so far the data have been controversial. It is important to consider that effects of DNA methylation can be direct by altering nucleosome formation ability or indirect by regulating factor binding, which, in turn, may trigger nucleosome repositioning (Bai and Morozov, 2010). Recent studies show that the addition of a methyl group can influence the flexibility of the DNA sequence (Diekmann, 1987; Hagerman, 1990; Hodges-Garcia and Hagerman, 1992) negatively affecting its ability to be incorporated into nucleosomes (Nathan and Crothers, 2002). Additionally, nucleosome affinity towards methylated DNA has been shown to decrease in a sequence- and methylation level-dependent manner (Davey et al., 1997; Davey et al., 2004). It is important to realize, however, that nucleosome positioning can also affect DNA methylation patterns making attribution of causality difficult (Hinshelwood et al., 2009).

Dynamic nucleosome positioning controls transcription by facilitating or impeding access of transcription factors to their corresponding regulatory DNA sequences in gene promoters (Cairns, 2009; Radman-Livaja and Rando, 2010). Figure 1-7 shows how nucleosome positioning can affect transcription factor binding at the promoter, and control gene expression. Accessing DNA sites that are inside in the nucleosome requires ATP-dependent chromatin remodeling (Jiang and Pugh, 2009b).

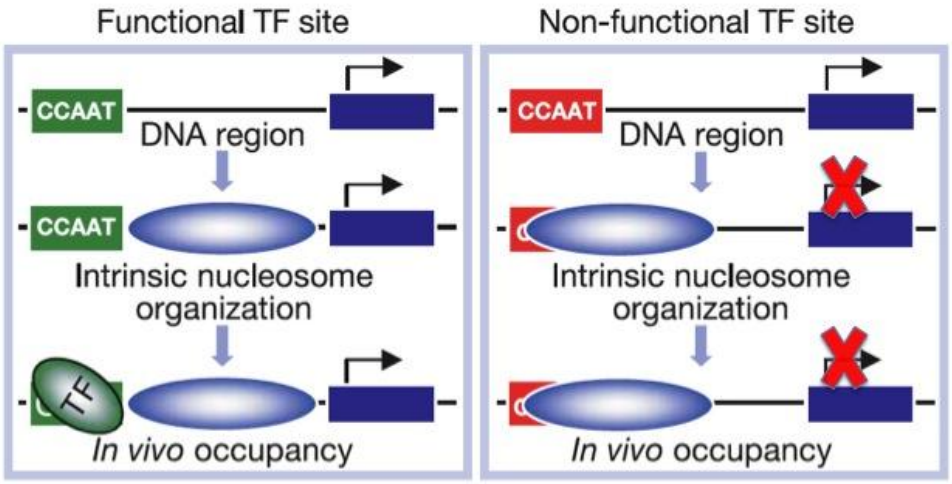

Figure 1-7. Nucleosome positioning regulates transcription. DNA motifs (green and red boxed sequences) are accessible to their cognate DNA binding factors on protein-free DNA. Specific nucleosome positioning controls accessibility of a transcription factor (TF) to its corresponding binding motif in DNA. A) A functional TF binding site is not occluded by nucleosomes, allowing for TF binding and activation of transcription. B) The non-functional TF binding site is occluded by the presence of the nucleosome, impeding TF binding and repressing transcription (Adapted: Segal et al., 2006).

Comparative studies between eukaryotic genomes have shown that promoters and regulatory sequences tend to be nucleosome depleted, whereas transcribed regions are occupied by well-positioned arrays of nucleosomes (Barski et al., 2007; Bernstein et al., 2004; Jiang and Pugh, 2009a; Lieb and Clarke, 2005; Mavrich et al., 2008; Mito et al., 2005; Ozsolak et al., 2007; Schones et al., 2008; Yuan et al., 2005). Recently, genome wide studies combined with the development of global nucleosome positioning mapping techniques have provided several nucleosome positioning maps for most of the model organisms and human. Although differences in conditions, experimental approaches and data analysis, account for some variation and discrepancies between studies, nevertheless striking global conclusions can be extracted from these efforts (Bai and Morozov, 2010). For the purposes of this dissertation, general findings relevant to human promoter nucleosome positioning are

discussed. Not surprisingly, nucleosome positioning is clearly not random. One of the first and most striking observations form genome-wide nucleosome positioning studies is the contrast between nucleosome occupancy in regulatory *versus* transcribed regions. While regulatory regions (promoters and enhancers) tend to be nucleosome depleted, transcribed regions are highly occupied (Barski et al., 2007; Bernstein et al., 2004; Jiang and Pugh, 2009a; Lieb and Clarke, 2005; Mavrich et al., 2008; Mito et al., 2005; Ozsolak et al., 2007; Schones et al., 2008; Yuan et al., 2005). A stereotypical accepted view of nucleosome positioning at promoters from different organisms is shown in Figure 1-8.

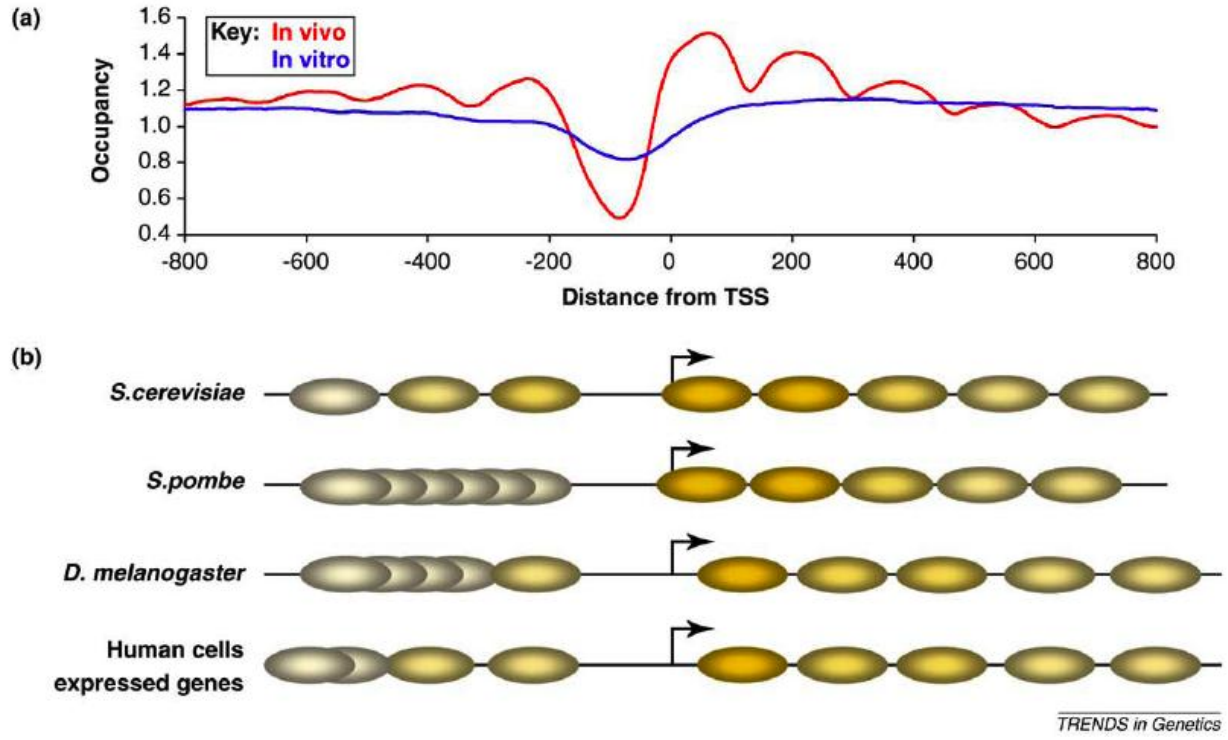

Figure 1-8. Stereotypical view of nucleosome positioning at gene promoters.
A) Averaged *in vivo* (red line) (Kaplan et al., 2009) and *in vitro* (blue line) (Zhang et al., 2009) nucleosome occupancy of *Saccharomyces cerevisiae* genes aligned by their TSSs. B) Schematic representation of typical *in vivo* nucleosome positions in *S. cerevisiae* and in several other species (based on similar measurements). Arrow, TSS; ovals, nucleosomes. The more yellow the nucleosome color, the better it is positioned relative to the TSS. Gray

ovals overlapping with each other represent nucleosomes that are not preferentially positioned .Note that the +1 nucleosome is further downstream in *Drosophila* and human cells compared with yeast. The nucleosomal repeat length (average distance between neighboring nucleosomes) is also different between species species. (Adapted from: Bai and Morozov, 2010).

This stereotypical chromatin conformation is to some extent conserved in humans, especially at CGI containing promoters. A characteristic NDR is observed around and upstream of the TSS. NDR size can vary but usually exclude a single nucleosome (~150 bp in yeast, ~200 bp in mammals); this region is thought to play a crucial role in transcription by allowing assembly of the transcription machinery. Sequences upstream and downstream of the NDR are usually occupied by well-positioned nucleosomes. The first nucleosome downstream of the NDR is typically very strongly positioned and subsequent nucleosomes become less positioned as a function of distance from the TSS. The first nucleosome upstream of the NDR is also usually strongly positioned, and nucleosomes farther upstream tend to be less positioned (Boyle et al., 2008; Ozsolak et al., 2007; Schones et al., 2008; Tillo et al., 2010). Importantly, even though the presence of a NDR is compatible with and permissive for transcriptional activation, it is not sufficient for transcription to occur. Many genes that are transcriptionally repressed shared this stable and common open promoter chromatin structure (Albert et al., 2007; Mavrich et al., 2008; Shivaswamy et al., 2008). Likewise, not all promoters in the human genome share this chromatin conformation. Some tissue-specific repressed genes have completely covered promoters, inducible and highly regulated genes have positioned nucleosome immediately upstream of the TSS (Cairns, 2009; Schones et al., 2008). Finally, a recent study have reported the existence of a highly unstable nucleosome located in the previously reported NDR (Jin et al., 2009).

Nucleosome occupancy has been associated with defined transcriptional states in yeast and flies. In yeast, transcriptional responses to heat shock or shifts in carbon sources that altered transcriptional programs resulted in increased nucleosome occupancy at repressed promoters and decreased nucleosome occupancy at promoters that became active (Bernstein et al., 2004; Lee et al., 2004). In contrast, little is known about the specific role of nucleosome occupancy at mammalian promoters in transcriptional regulation. Even though nucleosome depletion and nucleosome re-positioning or sliding seem to be conserved as regulatory transcriptional mechanisms in eukaryotic genomes, these phenomena further examination in mammalian cells (Heintzman et al., 2007). We and others have shown that nucleosome occupancy at the *MLH1* promoter NDR is associated with transcriptional silencing (Lin et al., 2007; Pardo et al., 2010).

Histone Modifications and Chromatin Structure

Histone modifications present a higher level of complexity than DNA methylation, mainly because several different residues in each histone tail (lysine, arginine, threonine and serine) can be targets of various post-translational modifications (Jenuwein and Allis, 2001; Kouzarides, 1999, 2000, 2007; Stancheva, 2005). These modifications include acetylation, methylation, phosphorylation, ADP-ribosylation, ubiquitination, and sumoylation (Figure 1-9). Adding to this complexity, several histone-modifying enzymes regulate the addition and removal of different modifications onto specific residues. At the same time, distinct protein domains found in chromatin remodeling complexes and in histone-modifying enzymes themselves, are capable of “reading and interpreting” these modifications, translating them into specific chromatin states (Klose and Yi, 2007; Kouzarides, 2007; Li et al., 2007; Zhang and Reinberg, 2001). Hence, histone post-

translational modifications affect DNA-histone interactions, histone-histone interactions and the interactions between histones and their regulatory factors. Histone modifications can also affect nucleosome positioning through interactions with ATP-dependent chromatin remodeling complexes (Bai and Morozov, 2010).

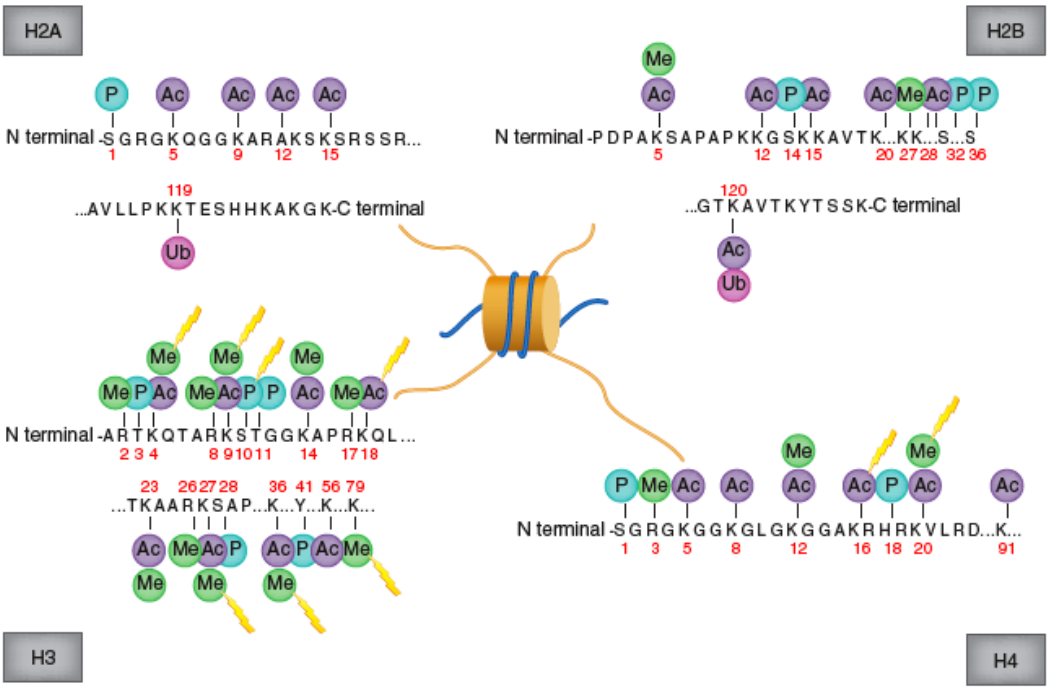

Figure 1-9. Histone modification patterns in normal and cancer cells. Mainly along their protruding N-terminal tails, but also within their C-terminal regions, histones can undergo diverse post-translational modifications. In the correct combination and translated by the appropriate effectors, these modifications contribute to establishing the global and local condensed or decondensed chromatin states that dictate gene expression levels. This figure depicts the main modifications of the four core histones in normal cells (type and position in the amino acid sequence). Furthermore, because disruption of their normal patterns is related to cancer, histone modifications typically associated with the disease have also been highlighted. Ac, acetylation; Me, methylation; P, phosphorylation; Ub, ubiquitination (Source: Rodriguez-Paredes and Esteller, 2011).

Acetylation of histones H3 and H4 is correlated with transcriptional activity both globally and at the level of specific genes. Histone acetylation patterns are maintained by the opposing activities of histone acetyltransferases (HATs) and histone

deacetylases (HDACs) (Vogelauer et al., 2000). Histone acetylation is enriched at promoters and enhancers and is maintained at low levels in gene bodies to prevent spurious transcription (Kouzarides, 2000, 2007). Also a mark that correlates positively with transcription , histone H3 lysine 4 tri- and dimethylation (H3K4me2/3) are enriched at promoters of active genes (Bernstein et al., 2004; Kim et al., 2005; Kouzarides, 2002; Kouzarides, 2007; Li et al., 2007)., while enhancers tend to be marked by H3K4me1 (Ng et al., 2003; Santos-Rosa et al., 2002). Histone 3 lysine 36 trimethylation (H3K36me3) is found in actively-transcribed gene bodies and is considered a positive mark for transcriptional elongation (Carrozza et al., 2005; Keogh et al., 2005).

Conversely, gene silencing seem to be enforced by histone deacetylation and histone H3 lysine 9 trimethylation (H3K9me3) and/or histone H3 lysine 27 trimethylation (H3K27me3) (Komashko et al., 2008). Comparative studies have shown that most genes are silenced by either H3K9me3 or H3K27me3, suggesting that these marks may not be used redundantly in normal cells. H3K27me3 for example is preferentially associated with silenced developmental genes (Barski et al., 2007; Komashko et al., 2008). Gene-by-gene studies have found that silenced gene promoters, especially those with hypermethylated CGIs, display particular histone modifications characteristic of a repressive chromatin environment. These modifications include reduced histone H3K4me3, histone H3 acetylation (specifically at H3K9 and H3K14), as well as increased trimethylation of both H3K9 and H3K27 (Cameron et al., 1999; Fahrner et al., 2002; Lehnertz et al., 2003). Figure 1-10 shows the mechanisms by which H3K9me3 and H3K27me3 are thought to influence chromatin accessibility and transcriptional silencing. Note that PRC mediated silencing can be DNA methylation independent,

while the proposed mechanism for H3K9me3 mediated silencing involves DNA methylation (Figure 1-10).

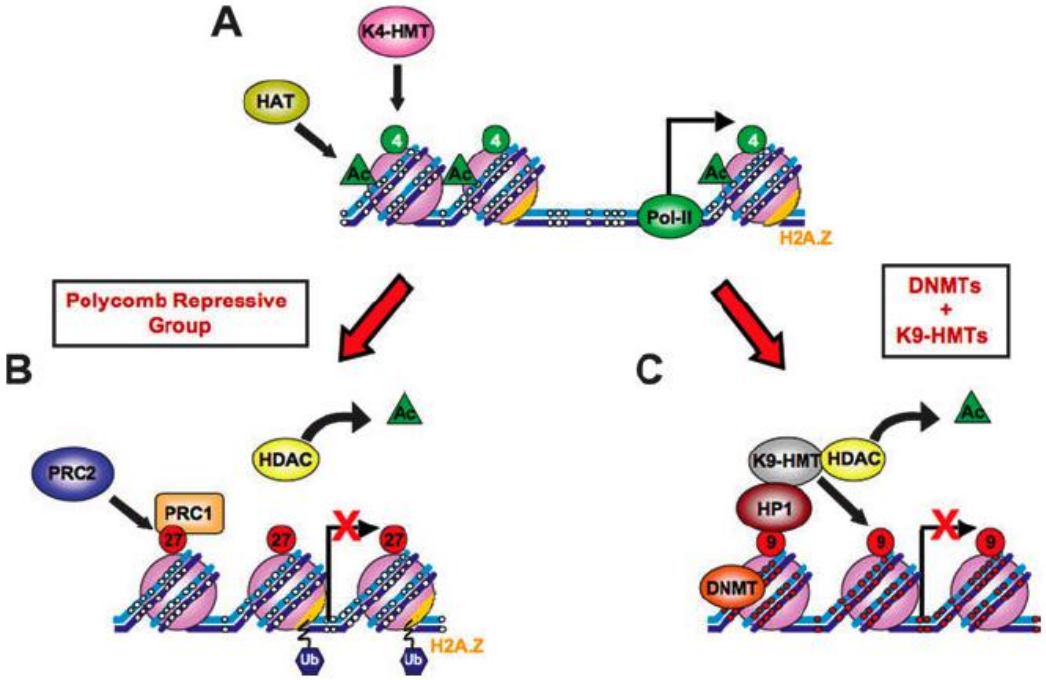

Figure 1-10. Epigenetic silencing mechanisms in mammals. A) An active gene shows an open chromatin structure consisting of an unmethylated promoter region (small white circles on DNA strands), a NDR upstream of the TSS (black bent arrow), an enrichment of active histone marks such as acetylation (green triangle, Ac) and H3K4 methylation (green circles, 4) and high levels of H2A.Z (orange) incorporated in nucleosomes flanking the TSS. The open chromatin structure is permissible for binding of transcription factors and RNA polymerase II (Pol-II), which mediates active transcription on such promoters. Alternatively, repression of such active genes can be achieved in normal cells by two main mechanisms. B) Gene repression by the action of PRC1 and PRC2 that mediate repressive H3K27 methylation (red circles, 27) is accompanied by the removal of acetylation by HDACs, loss of H3K4 methylation, chromatin compaction, nucleosome occupancy at the former NDR and ubiquitylation of H2A.Z. C) Long-term silencing through DNA methylation is performed by DNA methyltransferases. DNA methylation (small red circles on DNA strands) is often accompanied by the repressive H3K9 methylation (red circles, 9) on promoters, which leads to chromatin compaction by recruitment of HP1. Promoters with dense DNA methylation are silenced and show depletion of H2A.Z as well as both H3K4 methylation and histone acetylation. Ac, acetylation; EZH2, enhancer of zeste homolog 2; HP1, heterochromatin protein 1; K4-HMT, histone H3 lysine 4 histone methyltransferase; K9-HMT, histone H3 lysine 9 histone methyltransferase;

Pol-II, RNA polymerase II; PRC1 and PRC2, polycomb repressive complexes 1 and 2; Ub, ubiquitination. (Source: Sharma et al., 2010)

At the genome-wide level, recent studies of global patterns of histone H4 in normal, tumor and cancer cell lines showed that transformed cells exhibited global reduction of monoacetylated H4K16 and trimethylated H4K20. Global loss of H3K4me3 has also been reported in cancer cells. These alterations can be involved in disruption of heterochromatic regions associated with repetitive sequences and parasitic elements, thereby generating genomic instability and predisposing cells to cancer development (Fraga et al., 2005; Fraga and Esteller, 2005).

Several lines of evidence support cross talk between various layers of epigenetic regulation. Co-localization of DNA hypermethylation and repressive histone marks at promoters of silent genes links these two events to TSG silencing. The association of DNA hypermethylation with the formation of nuclease-resistant chromatin, characteristic of regions with dense arrays nucleosomes, provides more evidence for a connection between DNA methylation and nucleosome occupancy (Fraga and Esteller, 2005; Ting et al., 2006). The challenge that we now face is to unveil the more complex picture of interactions between each of these epigenetic features. Exploring these connections will help us understand how epigenetic events work together to establish and maintain transcriptional programming like persistent silencing of TSGs in cancer.

Interplay between Epigenetic Events

The importance of the coordinated action of the different layers of the epigenetic mechanisms to the complex control of gene expression governing diverse biological cell processes is now a well-accepted concept (Fuks, 2005; Harris and McCormick, 2010; McCabe et al., 2009). Different layers of epigenetic regulation can interact either in a

signaling-like manner or can work together to reinforce each other (Esteller, 2007a; Fuks, 2005; Schreiber and Bernstein, 2002; Ting et al., 2006). It is widely accepted that TSG silencing is the result of the coordinated interplay between DNA methylation as well as chromatin modifications and remodeling. However, little is known about the order in which these events occur during establishment of the transcriptional silencing at TSG promoters in cancer cells; either as a *de novo* process or as a resilencing event following pharmacologic derepression and resilencing of TSGs.

Transcriptional epigenetic repression can be achieved through several paths. Biochemical evidence linking the different layers of epigenetic regulation provides some possible models: (a) DNA methylation can modulate histone modifications; (b) histone modifications can modulate DNA methylation; or (c) chromatin remodeling may be an initial step leading to exposure of DNA to the methylation machinery in the cell or impeding the access of the transcriptional machinery to regulatory sequences (Ducasse and Brown, 2006; Fahrner et al., 2002; Hatziapostolou and Iliopoulos, 2011; Li, 2002; Lund and van Lohuizen, 2004). Figure 1-11 illustrates the possible mechanisms involved in these different silencing models. Controversy has characterized the possible hierarchy between DNA and histone modifications. Evidence supporting a dominant role of DNA methylation is based on its recruitment of MBDs, which interpret and mediate the repressive activities of DNA hypermethylation via HDAC and DNMT activity at hypermethylated promoters (Burgers et al., 2002; Fuks et al., 2001; Schreiber and Bernstein, 2002). At some promoters, reversal of DNA hypermethylation with demethylating agents must occur before HDAC activity is inhibited and gene expression is restored (Cameron et al., 1999; Suzuki et al., 2002). Supporting this model further,

demethylation causes the release of MBDs from promoters and presumably HDAC dissociation as well (Bakker et al., 2002; El-Osta et al., 2002; Nguyen et al., 2001).

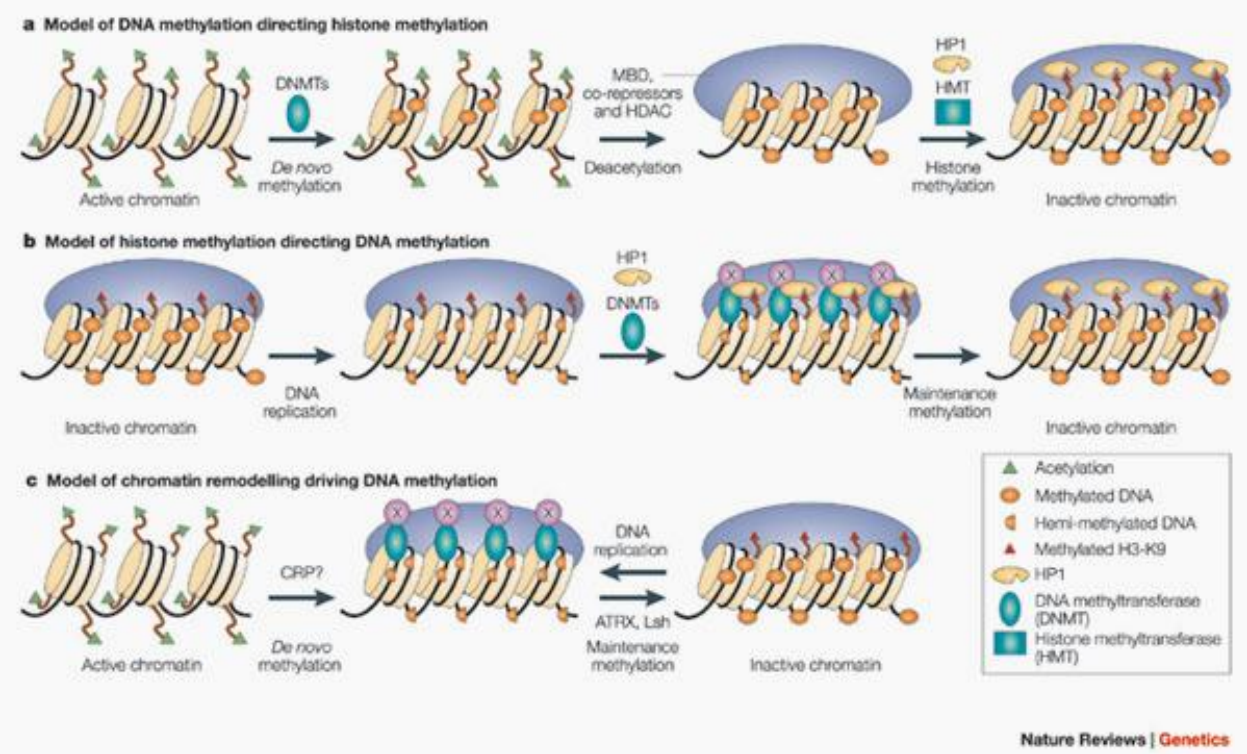

Figure 1-11. Interplay between DNA methylation, histone modification and chromatin remodeling leading to gene silencing. In mammalian cells, both DNA methylation and histone modification are involved in chromatin silencing. DNA methylation and histone modification are believed to be interdependent processes. Three possible models for how they might influence each other are shown. A) A model of DNA methylation directing histone methylation. DNA methylation patterns are established through *de novo* methylation by the DNA methyltransferases DNMT3A and DNMT3B, which are maintained by DNMT1. Methyl-CG-binding proteins (MBD) and histone deacetylase (HDAC) complexes, such as the MECP2–Sin3a–HDAC complex, are believed to then be recruited to the methylated region to induce histone deacetylation and silencing. Histone methyltransferases (HMTs), such as Suv39h or G9a, are then recruited which methylate the lysine 9 residue on histone H3 (H3-K9) and stabilize the inactive state of the chromatin. B) A model of histone methylation directing DNA methylation. Methyl H3-K9 acts as a signal for inactive chromatin by recruiting HP1 to methylated histones, which might, in turn, recruit DNMTs directly or indirectly (through an unknown factor, X) to the silent chromatin to maintain DNA methylation and stabilize the inactive chromatin. C) A model of chromatin remodeling driving DNA methylation. The ATP-dependent chromatin-remodeling and DNA-helicase activities of proteins, such as ATRX and Lsh, might facilitate DNA methylation and

histone modification by exposing nucleosomal DNA to increase its accessibility to DNMTs, HDACs and HMTs. Loss of ATRX and Lsh in *Arabidopsis* impairs both DNA methylation and histone methylation. The mouse chromatin-remodeling protein (CRP) that is involved in *de novo* methylation has yet to be identified (Source: Li, 2002)

Supporting the dominance of histone marks in gene silencing, in *Arabidopsis* and *Neurospora*, histone methyltransferases (HMTs, e.g. Suv39h homologs) that place the H3K9me3 repressive mark have been shown to be required for DNA methylation (Johnson et al., 2002; Soppe et al., 2002; Tamaru et al., 2003). In mammalian X-chromosome inactivation, HMTs responsible for H3K9me3 and H3K27me3 repressive repressive marks recruit DNMTs to the silencing sites (Cao and Zhang, 2004; Plath et al., 2003). Similarly, at pericentromeric satellite repeats in mouse ES cells, deletion of both Suv39h homologs impairs DNA methylation, whereas wild-type levels of H3K9me3 persist in cells double-null for *Dnmt1* and *Dnmt3a* (Lehnertz et al., 2003). Studies on the reactivation of the TSG *P16INK4A* following treatment with the demethylating agent 5-aza-dC showed that re-expression is associated with increased acetylation and loss of H3K9me3. However, after 22 further passages without the drug, the gene was re-silenced and the H3K9me3-silencing mark was reacquired, while DNA methylation was not (Bachman et al., 2003; Dumont et al., 2009).

As mentioned before, little data regarding nucleosome positioning in cancer are available. Hence, information about the detailed mechanisms by which nucleosome occupancy and DNA methylation interact to contribute to TSG remains elusive. Further investigation is clearly needed to resolve current controversies concerning the temporal sequence of molecular events accompanying epigenetic gene silencing and also to shed light on how epigenetic events contribute to different stages of cancer progression.

DNA Demethylation and Resilencing by 5-aza-dC

DNA demethylating agents like 5-aza-dC (Decitabine) and 5-aza-cytidine (5-aza-C; Vidaza) are at the front line of epigenetic altering drugs used for therapeutic treatment of cancer. When compared to other cytotoxic chemotherapeutic agents, DNA demethylating agents have shown higher response rates and increased survival in clinical patients (Oki et al., 2007; Shen et al., 2010). These two compounds are widely used in the treatment of myelodysplastic syndromes (MSD), where response rates are about 30% to 60%. Increased patient survival as compared to other chemotherapeutic agents and supportive care are usually observed (Issa and Kantarjian, 2009). Figure 1-7 shows the pleiotropic therapeutic events of DNA demethylation and gene reactivation in cancer induced by 5-aza-dC.

Widespread therapeutic use of 5-aza-dC and 5-aza-C faces several hurdles. First, treatment of solid tumors has proven challenging, in part because efficient drug delivery to solid tumors is more difficult. Second, while demethylation has been shown to occur *in vivo* in patients, different degrees of demethylation have been observed, depending on the assayed DNA region. Third, it remains controversial as to whether levels of DNA demethylation can predict clinical response (Issa and Kantarjian, 2009; Yang et al., 2010). Fourth, gene remethylation and resilencing are observed, in most cases, to the levels observed pretreatment within few weeks of drug administration (Issa and Kantarjian, 2009; McCabe et al., 2009). At present, the mechanisms by which gene expression is both derepressed by demethylating agents and returns to the silenced state are ill defined. Overcoming these hurdles will no doubt benefit from a more complete understanding of mechanisms of drug resistance and sensitivity (Issa and Kantarjian, 2009; Oki et al., 2008; Stewart et al., 2009).

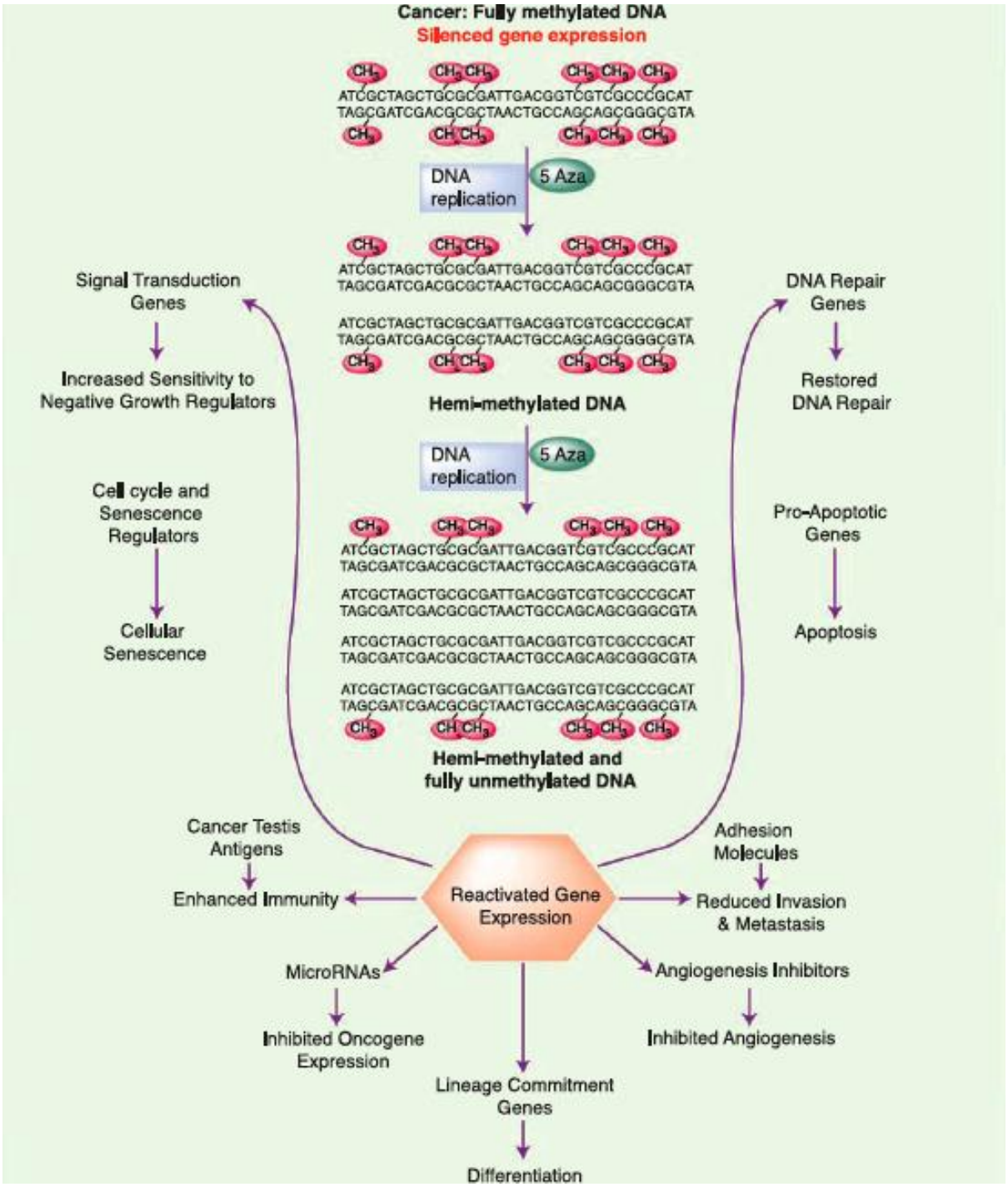

Figure 1-12. Pleiotropic therapeutic effects of DNA methylation inhibition and gene reactivation in cancer. DNA methylation is maintained post-replicatively by the action of DNMTs. The current model for 5-aza-C and 5-aza-dC (5 Aza) action is degradation of cellular DNMTs, Subsequent DNA replication results in passive demethylation that eventually results in reactivated gene expression. Activated gene expression, in turn, has effects on multiple different pathways, each of which could contribute to a clinical response (Source: Issa and Kantarjian, 2009)

Given the interest in expanding the clinical use of epigenetic therapies and controversy regarding the interplay between epigenetic mechanisms; we sought to develop an assay that could simultaneously interrogate multiple epigenetic features. If the following chapters, we first describe the development of MAPit, a novel technique for simultaneous mapping of endogenous DNA methylation and nucleosome positioning at the level of single molecules. Next, we describe development of MethylViewer, a computational analytical tool for analysis of complex MAPit datasets and any bisulfite genomic sequencing project. Finally, we describe the application of MAPit in the characterization of tumor-cell specific chromatin structures and the identification of epigenetic features that contribute to TSG resilencing following 5-aza-dC treatment and withdrawal in cultured cancer cells.

CHAPTER 2
SIMULTANEOUS SINGLE-MOLECULE MAPPING OF PROTEIN-DNA INTERACTIONS AND DNA METHYLATION BY MAPit[1]

Introductory Remarks

Crucial to a complete understanding of any biological function of DNA is the footprinting or mapping protein-DNA interactions at high resolution. Formerly, footprinting methods have relied on assaying accessibility of sites in DNA to probing reagents that result in DNA cleavage, such as nucleases (e.g., DNase I and micrococcal nuclease) or chemicals (e.g. dimethylsulfate). Locations of protein binding to DNA are inferred by comparing sites of protection against damage in the absence and presence of the putative DNA-binding protein. Such conditions can be set up either *in vitro* or *in vivo*, e.g. wild-type cells versus the same cells with expression knock-down or bearing a null mutation in the gene of the factor of interest. While conventional footprinting methods have proven enormously informative, they are subject to several theoretical and practical limitations, as discussed in the Commentary. In particular, as only a single DNA break can be mapped per DNA molecule, conditions must approach limiting or single-hit kinetic levels of cleavage. By mapping cleavages at a given site over all molecules in a sample, the inherent complexity of protein-DNA interactions in biological systems is obscured by population averaging (Pondugula and Kladde, 2008).

These problems are overcome by probing protein-DNA interactions with DNA methyltransferases (DNMTs) that modify C followed by bisulfite sequencing(Clark et al., 1994; Darst et al., 2010; Frommer et al., 1992), termed MAPit (Jessen et al., 2006;

[1] Reprinted with permission of John Wiley & Sons, Inc. As published in: Carolina E. Pardo, Russell P. Darst, Nancy H. Nabilsi, Amber L. Delmas and Michael P. Kladde. 2011. Simultaneous Single-Molecule Mapping of Protein-DNA Interactions and DNA Methylation by MAPit. Current Protocols in Molecular Biology. Chapter 12.

Pardo et al., 2009). A key advantage of MAPit over other techniques is that it reports the methylation status (i.e. accessibility versus protection) of every C residue along one strand of individually-cloned and sequenced DNA molecules. This provides a single-molecule, non-averaged view of protein-DNA interactions that permits correlation between different footprints in a region within a sample population.

This modification is undetectable in commonly studied model invertebrates, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Caenorhabditis elegans*. In contrast, significant levels of m^5C are present in vertebrates, predominantly at CG sites (CG hereafter; (Goll and Bestor, 2005; Gruenbaum et al., 1981)), where it plays important roles in regulation of gene expression (Deaton and Bird, 2011). Non-CG methylation at CHH and CHG (H is A, C or T) also occurs in land plants (Cokus et al., 2008; Henderson and Jacobsen, 2007; Zilberman et al., 2007), a key cellular activity for silencing transposable elements (Goll and Bestor, 2005). Methylation at CHH and CHG has also been detected in human embryonic stem cells, some cancers, and Purkinje neurons, although its functions are not well understood (Grandjean et al., 2007; Hawkins et al., 2010; Kriaucionis and Heintz, 2009; Latham et al., 2008; Laurent et al., 2010b; Lister et al., 2009; Ramsahoye et al., 2000). A DNMT with different sequence specificity is needed to fully leverage MAPit. To this end, we cloned and characterized M.CviPI, an enzyme that methylates GC sites at C-5 (Xu et al., 1998a). Bisulfite sequencing of mammalian chromatin probed with M.CviPI therefore allows the methylation status of both CG and GC sites to be determined along a single DNA strand (Kilgore et al., 2007; Pardo et al., 2009). This chapter describes the final optimized protocol for MAPit using M.CviPI for simultaneous

mapping of both endogenous cytosine methylation and protein-DNA interactions in cultured mammalian cells.

Development and Optimization of the Protocol for Probing Mammalian Nuclear Chromatin with DNMTs

The three basic steps of MAPit are: (1) delivery of a suitable C-modifying DNMT to probe accessibility of DNA or chromatin; (2) bisulfite sequencing, including bisulfite conversion of isolated and denatured DNA, PCR amplification of deaminated DNA, and sequencing cloned individual molecules from the PCR amplicon; and (3) assignment of the methylation status to each potential DNMT target sequence.

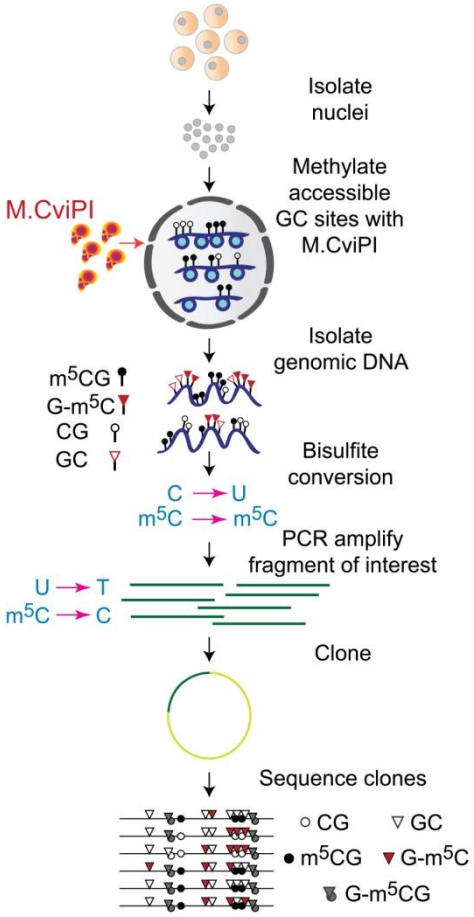

Figure 2-1. MAPit overview for mapping m5CG and chromatin accessibility in mammalian nuclei. Nuclei are isolated form cultured cells grown under desired experimental conditions. Isolated nuclei are then probed with the DNA methyltransferase M.CviPI, which methylates cytosines at accessible GC sites, not bound by nucleosomes or proteins. After isolation cytosines in the

DNA are either non-methylated or, modified at CG sites by endogenous DNMTs (black circles) or at GC sites by the M.CviPI probe (red inverted triangles) . Probed DNA is then subjected to bisulfite conversion. Unmethylated Cs are converted to Us while methylated Cs remain resistant to bisulfite conversion. During the PCR step Us will be read as Ts and methylated Cs will be read as Cs, allowing the assessment of the methylation status of every C residue in the DNA molecule. To obtain a single molecule readout of both the endogenous methylation CG and the chromatin accessibility, PCR products are cloned and sequenced. Sequences are processed using a sequence alignment program. Sequence alignment data is analyzed using the program MehtylViewer which generates an accessibility map of each molecule. In the map each horizontal line represents a single DNA molecule, circles (CG) and triangles (GC) represent the location of potential methylation sites in each molecule. Filled figures represent unmodified residues and open figures represent unmodified residues.

MAPit Protocol Description

MAPit as describe here can be used for the simultaneous probing of DNA methylation and chromatin accessibility in any type of cultured cells. Cell lines and growing conditions will vary according to the question being addressed and researcher discretion. Cells should be cultured using standard tissue culture techniques under desired experimental conditions until at least 1.5 X 10^6 cells per experimental sample (e.g. DNMT dose) are obtained. To obtain the sample data in this chapter, MCF10A immortalized mammary cells were cultured in Dubelcco's modified eagle medium (DMEM) supplemented with 10% (v/v) fetal bovine serum and 1% (v/v) penicillin and streptomycin, following the provider recommendations.

Reagents and solutions needed for the MAPit protocol described in here are listed in Table 2-1.

Table 2-1. Reagents and solutions needed for MAPit experiments.

Reagent and Solutions	Storage conditions
Trysin-EDTA solution	Store for up to 6 months at -20°C
Phosphate buffered saline (PBS)	Store indefinitely at room temperature
1 M dithiothreitol (DTT)	Store in single-use aliquots at -20°C
100 mM phenylmethylsulfonyl fluoride (PMSF)	Dissolve in absolute ethanol. Store for up to 6 months at -20°C

Table 2-1. Continued

Reagent and Solutions	Storage conditions
0.4% trypan blue solution	Store indefinitely at room temperature
32 mM S-adenosyl-L-methionine (SAM)	Store in single-use aliquots at -80°C
80 U/µl M.CviPI	Aliquot and store in non-frost-free-freezer
20 mg/ml proteinase K	Store in non-frost-free-freezer for up to 4 months
5 U/µl HotStar Taq (Qiagen)	Store in non-frost-free-freezer
10.0 M ammonium acetate, pH 8.0	Store indefinitely at room temperature

Buffers and solutions that need to be prepared for the MAPit protocol described in here are listed in Table 2-2.

Table 2-2. Buffers needed for MAPit experiments.

Buffer/Solution	Composition	Notes
Cell Resuspension Buffer	20 mM HEPES, pH 7.5 70 mM NaCl 0.25 mM EDTA, pH 8.0 0.5 mM EGTA, pH 8.0 0.5% (v/v) glycerol 10 mM DTT 0.25 mM PMSF	Make fresh before use. DTT and PMSF should be added fresh immediately before use and stored at -20°C
Cell Lysis Buffer	Cell resuspension buffer + 0.19% (v/v) Nonidet P-40	Make fresh before use DTT and PMSF should be added fresh immediately before use and stored at -20°C
Nuclei Methylation Buffer	Cell resuspension buffer + 290 µM SAM	Make fresh before use DTT, PMSF and SAM should be added fresh immediately before use and stored at -20°C
Methylation Stop Buffer	100 mM NaCl 10 mM EDTA, pH 8.0 1% (w/v) SDS	Store for up to a month at room temperature
M.CviPI Storage Buffer	15 mM Tris-HCl, pH 7.4 200 mM NaCl 1 mM DTT 0.1 mM EDTA 200 µg/ml acetylated BSA 50% (v/v) glycerol	Store at -20°C

Table 2-2. Continued

Buffer/Solution	Composition	Notes
Enzyme Dilution Buffer	1-part M.CviPI storage buffer 7-parts nuclei methylation buffer	Make fresh before use
Phenol chloroform solution	Mix 25:24:1 Equilibrated phenol, pH 8.0: Chlorophorm: Isoamyl alcohol	Store for up to a year at room temperature
0.1X TE buffer	1 mM Tris-HCl, pH 8.0 0.1 mM Na_2EDTA, pH 8.0	Autoclave, store indefinitely at room temperature
70 % (v/v) ethanol	Mix 37 ml absolute ethanol :13ml 0.1X TE	Caution: Flammable

Note: Reagents should be prepared in sterile disposable labware. Use only distilled H_2O in all steps and solutions. Nuclei isolation and methylation buffers should be freshly prepared on the day of the experiment. DTT, PMSF and SAM should be added to solutions immediately before use to avoid oxidation or hydrolysis. M.CviPI activity is very dependent on fresh addition of DTT.

A refrigerated centrifuge and microcentrifuge or one in a cold room is recommended for isolation of nuclei. An automated cell counter or a hemacytometer and a light microscope are used for tissue culture protocols and 37°C and a 50°C are needed for temperature controlled incubations.

Cell harvesting

To harvest cells add an appropriate volume of 37°C trypsin-EDTA solution to remove cells from tissue culture plates or flasks (e.i. 10 ml of trypsin-EDTA for a 10 cm culture dish). Incubate cells at room temperature until they detach from the growth surface. The time needed for cell detachment varies from one cell line to another (~2-10 min), and can be determined by visualization with a light microscope. Alternatively, cells can be harvested by adding ice-cold PBS directly to plates and scraping into 50 ml conical tubes on ice. Add cell growth medium plus serum pre-warmed to 37°C (three times the volume of trypsinization solution used in step 1) to terminate trypsinization. Trypsin activity is inhibited by the serum protease inhibitor alpha-1-antitrypsin. After

trypsinization, centrifuge cells for 5 min at 1,000 × g at 4°C to pellet cells. Carefully aspirate supernatant and add 5 ml ice-cold PBS. Resuspend cell pellet gently by pipetting up and down and wash cells. PBS wash should be repeated once. After the second wash, resuspend cells with ice cold PBS to an approximate concentration of 10^6 cells/ml and keep cells on ice. This is equivalent to resuspending HeLa cells from a 90% confluent 10 cm plate into 3 ml of PBS. To determine cell number mix an aliquot of 20 µl cells with 20 µl 0.4% (w/v) trypan blue solution. Pipet cells up and down several times to disperse and make a homogeneous cell suspension and count the number of live cells that exclude trypan blue either manually with a hemocytometer or using an automated cell counting device.

Once the cell number is known, dispense 1.1×10^6 cells per experimental sample into pre-labeled 1.7 ml microcentrifuge tubes on ice. Each DNMT probing reaction requires 10^6 cells. Starting with 1.1×10^6 cells per reaction (one reaction is one DNMT dose) allows for some loss during preparation of nuclei. We recommend setting up an untreated sample (0 U DNMT) and two concentrations of M.CviPI, therefore requiring 3.3×10^6 cells per experimental condition. In our experience, 30 and 100 U M.CviPI are good starting doses for either the M.CviPI-MBP or M.CviPI-GST reagents. Using two different concentrations of enzyme, while not essential, allows one to assess different degrees of chromatin accessibility and the extent of saturation of methylation by exogenously added M.CviPI at each GC site. The untreated sample (0 U DNMT) serves as a background control to monitor non-conversion of C in GC sites by bisulfite and/or sequencing errors. The untreated sample also shows the level of endogenous CG methylation in the sample before probing, which should be taken into account when

inferring whether GCG sites were likely methylated by endogenous DNMTs or exogenously-added by DNMT probe.

After aliquoting the required number of cells for the experiment, a wash with cell resuspension buffer is needed, for this microcentrifuge cells for 5 min at 1,000 × g at 4°C, to pellet cells. Aspirate the supernatant and add 200 µl ice-cold cell resuspension buffer per 1.1 × 10^6 cells (i.e. add 600 µl, if 0, 30 and 100 U M.CviPI are used). Resuspend pellet by tapping tube gently. Isolating all nuclei for each experimental condition together in a single tube and aliquoting to separate tubes in step 17 ensures that the only variable will be the DNMT concentration.

Mammalian nuclei isolation

To isolate nuclei centrifuge cells for 5 min at 1,000 × g at 4°C. Aspirate supernatant and resuspend cell pellet in 38.5 µl of ice cold cell lysis buffer per 1.1 ×10^6 cells (i.e. add 115.5 µl, if 0, 30 and 100 U M.CviPI are used). Incubate for 10 min at 4°C to lyse cells. Inclusion of the non-ionic detergent Nonidet P-40 in cell lysis buffer allows for cell membrane lysis while maintaining nuclear integrity. Nonidet P-40 concentration and lysis time may need to be optimized for different cell types in order to obtain complete cell lysis without disrupting integrity of the nuclear envelope. To preserve nuclear structural integrity and native protein-DNA interactions, all steps for nuclei preparation should be done at 4°C. Nuclei should be handled carefully as they are prone to lysis. Avoid pipetting of nuclei instead, resuspend by gentle tapping of the tube with a finger.

While cells are undergoing lysis prepare ice-cold methylation buffer by mixing 61 µl ice-cold cell resuspension buffer and 0.55 µl freshly-thawed 32 mM SAM. These volumes are per each sample containing about 10^6 nuclei. Make enough extra solution

to account for pipetting error. The methylation buffer contains 290 µM SAM, which will be diluted to a final concentration of 160 µM in the methylation reactions in step 20. A minute before the lysis incubation is over, check the structural integrity of the nuclei. To do this, stain 2 µl aliquot of nuclei solution by adding 2 µl of 0.4% (w/v) trypan blue solution in a separate tube. Mix by gently tapping the tube and examine nuclei immediately by light microscopy. Nuclei should stain blue as well as appear round and granular with no attached cytoplasmic debris. If intact cells are observed incubate the nuclei a little longer and check them again. Nuclei may swell slightly during isolation and subsequent manipulations.

After lysis is completed add 61 µl ice cold methylation buffer per 10^6 nuclei (i.e. add 183 µl, if 0, 30 and 100 U M.CviPI are used) to dilute Nonidet P-40 concentration. Mix by gently tapping the tube. Dilution of Nonidet P-40 detergent to 0.07% (v/v) in this step helps maintain nuclear integrity. Dispense 90 µl of nuclei resuspension containing 10^6 nuclei into 1.7 ml microcentrifuge tubes pre-labeled with each unit concentration of M.CviPI being used.

Nuclear chromatin structure probing by methylation with exogenous M.CviPI

For the 30 and 100 U M.CviPI samples, appropriate volumes of 3 U/µl and 10 U/µl M.CviPI solution are needed, respectively. Immediately before use, on ice, prepare M.CviPI dilutions as follows: Dilute 80 U/µl commercial stock of M.CviPI by eight-fold with ice-cold methylation buffer to make 10 U/1 µl dilution. Make a 3.33-fold serial dilution of the 10 U/µl dilution with enzyme dilution buffer to make the 3 U/µl dilution. Setting up a dilution series ensures that all samples are subjected to identical conditions in parallel (i.e., salt and 160 µM SAM), with the DNMT concentration being the only variable. Always use freshly made enzyme dilution buffer to make M.CviPI dilutions.

Before probing, pre-warm the nuclei dispensed in each tube by incubation for 2 min at 37°C. At the same time, pre-warm to 50°C a sufficient volume of 2X methylation stop buffer (100 µl per methylation reaction plus some extra to allow for pipetting error). 20. Staggering M.CviPI addition to each tube by 30 sec, add 10 µl of the corresponding M.CviPI dilution to each pre-warmed sample. (e.i. add 10 ul of 3U/ul M.CviPI dilution to the 30 U sample and 10 µl of the 10 U/ µl M.CviPI dilution to the 100 u sample). Pipet up and down gently to mix and methylate for 15 min at 37°C. Staggered addition of enzyme and respective staggered termination of methylation by addition of methylation stop buffer, ensure that the incubation time with the chromatin-probing enzyme is held constant.

Parameters used during the chromatin probing reaction can be changed according to the requirements of the experiment. We recommend performing a pilot experiment under the conditions described here. Time and enzyme concentration can be adjusted accordingly (Commentary at the end of this chapter).

After nuclei methylation, terminate each reaction by adding 100 µl of 2× methylation stop buffer pre-warmed to 50°C, following the staggering scheme described before. Vortex each sample immediately. Complete removal of protein from the DNA is necessary to achieve complete denaturation and hence bisulfite conversion (Warnecke et al., 2002). Add 1 µl of 20 mg/ml proteinase K to a final concentration of 100 µg/ml. Mix by inverting tubes and incubate overnight at 50°C. In our experience, incubation with proteinase K for at least 16 hr is required.

Mammalian genomic DNA isolation

After overnight incubation with proteinase K, extract proteins from the genomic DNA solution by adding 200 µl (an equal volume) of phenol chloroform solution. Vortex

vigorously for 30 sec to obtain a homogeneous suspension. Separate the aqueous and organic phases by centrifugation for 5 min at 14,000 × g at room temperature in a microcentrifuge. Transfer the aqueous (top) phase to a new 1.7 ml microcentrifuge tube carefully avoiding transfer of denatured protein and SDS (white material located at the phase interface). Add 1/4 volume of 10.0 M ammonium acetate (i.e. 2.5 M final), and vortex briefly to mix. Add 2.5 volumes of absolute ethanol, mix thoroughly by gentle inversion. At this point, samples can be stored overnight to indefinitely at −20°C. Overnight incubation at −20°C increases recovery of low concentrations of nucleic acid. Pellet the nucleic acid by centrifugation for 5 min at 14,000 × g at room temperature in a microcentrifuge. Draw off supernatant carefully so as not to dislodge the nucleic acid pellet and add 0.4 ml 70% (v/v) ethanol/0.1X TE to wash nucleic acid pellet, vortex gently. Centrifuge for 5 min at 14,000 × g at room temperature in a microcentrifuge to pellet nucleic acid. Carefully draw off supernatant without disturbing the pellet and air-dry pellet for ~10 min. Finally, resuspend genomic DNA in 50 µl 0.1× TE. Genomic DNA usually requires overnight incubation at 4°C to solubilize completely. Removal of RNA prior to bisulfite sequencing is not necessary. Samples can be stored at 4°C for many months or indefinitely at −20°C.

Bisulfite sequencing of mammalian DNA

After DNA isolation, approximately 5-15 µg DNA are recovered from each reaction containing 10^6 nuclei. Bisulfite sequencing, including bisulfite conversion of purified DNA, PCR amplification of sequences of interest, cloning individual molecules from the PCR product, and sequencing cloned molecules, should ideally be performed as described in Darst et al 2010. Once clones of individual molecules have been sequenced the data are analyzed by MethylViewer (Pardo et al., 2010). This computer

program can concurrently score the methylation status of up to four user-defined sites either directly from *.ab1 sequencing files or from a FASTA file of sequences aligned with another program. For MAPit analysis of mammalian chromatin with M.CviPI, MethylViewer is used to concurrently score methylation at CG and GC sites along each sequenced molecule, and export publication-quality images. Other features, such as verification of bisulfite-conversion efficiency at non-CG and non-GC sequences can also be obtained. Occasional sequences with conversion efficiencies of <97% are typically omitted from further analyses, but this is up to the discretion of the investigator.

MAPit protocol time considerations

This protocol typically requires 4-5 days to complete, plus the time required to obtain DNA sequencing data. Probing isolated nuclei with exogenous DNMT can be performed in one day, including an overnight proteinase K digest, but may take more time depending on specific experimental goals and design. DNA purification takes 4 hr, plus an overnight elution step. Together, bisulfite conversion of purified DNA samples, PCR amplification, ligation and transformation take as many as 20 hr, which may be broken into separate days. After growing colonies overnight, analysis of cloning efficiency and preparation of 96-well sequencing plates takes under 6 hr. Plates are grown overnight, and preparation for transfer to a sequencing facility takes 1 hr the next day. Sequencing time depends on the sequencing facility.

Verification of GC DNA Methylation by M.CviPI

When using a new enzyme preparation, we recommend determining enzyme activity before investing time in sequencing and analysis of MAPit data. It is possible to methylate purified plasmids and test with various restriction enzymes. However, higher enzymatic activity is needed to methylate chromatin. It is convenient to assay activity by

methylation of nuclear DNA, using the actual experimental samples. To confirm that the DNMT used was active, one of two methods may be used to screen for GC methylation, either quantitative methylation-sensitive restriction enzyme digestion (qMSRE) or methylation-specific PCR (MSP).

Quantitative methylation-sensitive restriction enzyme digestion (qMSRE)

For qMSRE, 20-250 ng of purified genomic DNA (not bisulfite treated) is subject to digestion with the methylation sensitive enzyme R.HaeIII. This enzyme can digest unmethylated GGCC sites but not GG-m^5CC sites. A parallel "mock" reaction containing all reaction components except R.HaeIII (replaced with glycerol) is included for each sample. DNA from the R.HaeIII-digested or mock reaction is then amplified by real-time PCR with primers to a known open region containing a HaeIII site, such as the human GAPDH promoter (primers TACTAGCGGTTTTACGGGCG and TCGAACAGGAGGAGCAGAGAGCGA). Results are normalized to each sample's mock digestion control and quantified using the $\Delta\Delta C_T$ method to determine the levels of protection from R.HaeIII digestion achieved by each dose of DNMT.

Methylation-specific PCR (MSP) for human long interspersed nuclear element 1 (LINE 1)

For MSP, 20 ng of bisulfite-treated DNA is amplified with two sets of primers that target human long interspersed nucleotide elements (LINE-1). One primer pair amplifies GC unmethylated or "U" LINE (primer sequences AGGTATTGTTTTATTTGGGAAGTGT and CCTTACAATTTAATCTCAAACTACTATA) and the second pair amplifies GC-methylated or "M" LINE (primers CATTGCTTTATTTGGGAAGCGC and CTTGCAATTTAATCTCAAACTGCTATG) DNA. The product of each PCR reaction is

visualized on an agarose gel: the M product will be more abundant than the U product if the DNMT was active.

Commentary

Background Information

Protein-DNA interactions play crucial roles in mediating all biological functions of DNA in every organism. Eukaryotes package their DNA into chromatin comprising a protein content of roughly half non-histone regulatory factors and half core histones. The fundamental repeating unit of eukaryotic chromosomes is the nucleosome core particle, composed of a histone octamer (central histone tetramer (H3−H4)2 and H2A−H2B dimers) wrapped by a left-handed superhelix consisting of 1.65 turns or 147 bp of DNA (Luger et al., 1997a). Individual nucleosomes are repeated at a distance characteristic for each eukaryotic species, i.e. separated by a modal length of histone-free linker DNA in bulk chromatin. In contrast, at the single-molecule level, there can be considerable variation in linker length within a given region of chromatin. Nucleosomes are among the most stable protein-DNA interactions in eukaryotic chromosomes and act in concert with DNA-binding factors and other chromatin-associated factors to exert tight control of gene expression and other DNA functions (Kouzarides, 2007; Li et al., 2007).

In many eukaryotes, endogenous DNMTs post-replicatively modify the DNA component of chromatin at the 5 position of the cytosine base ring. C-5 methylation (m^5C) in vertebrates appears to occur exclusively at CG sites in most cell types, and plays essential roles in diverse aspects of vertebrate genome function (Bestor, 2000; Bird, 2002). These include repression or silencing of transcription, embryonic development, genomic imprinting of either the paternal of maternal alleles of some

genes, inactivation of one of two X chromosomes in normal females of Eutherian mammals, and suppression of the mobility of parasitic genetic elements, e.g. retrotransposons (Bestor and Bourc'his, 2004; Goll and Bestor, 2005; McCabe et al., 2009; Robertson, 2005; Robertson and Wolffe, 2000). Aberrant DNA methylation is frequently associated with human aging and diseases, such as cancer (Bird, 2002; Feinberg et al., 2006; Jaenisch and Bird, 2003; Jones and Baylin, 2007; Robertson, 2001, 2005). In other cases, m^5CG has been shown to activate transcription when it blocks binding of proteins to DNA that exert transcriptional repression (Lai et al., 2010; Nabilsi et al., 2009; Wu et al., 2010). At lower levels, m^5C is also present at non-CG sites, CHG and CHH (H is either A, C or T), in undifferentiated human embryonic stem cells (Grandjean et al., 2007; Hawkins et al., 2010; Kouidou et al., 2005; Latham et al., 2008; Laurent et al., 2010a; Lister et al., 2009).

Chromatin structure is highly dynamic; nucleosomes are constantly being mobilized to different positions and/or are disassembled via the action of ATP-dependent chromatin remodelers, histone chaperones, or both (Clapier and Cairns, 2009; Langst and Becker, 2004; Saha et al., 2006). Nucleosome depletion at transcription start sites, for example, is often diagnostic of transcription initiation (Boeger et al., 2004; Jiang and Pugh, 2009b; Korber et al., 2004; Mito et al., 2005). Although a hallmark of epigenetic m^5CG is its heritability from one cell to another, DNA methylation is also dynamic. First, methylation is not precisely maintained and thus modification of specific CG sites can fluctuate considerably. Second, cellular differentiation has recently been shown to involve oxidation of m^5C to hydroxymethyl C (hm^5C), which is subsequently removed by an as yet unknown mechanism (Chadee et al., 1995; Ito et

al., 2010; Tahiliani et al., 2009). In sum, dynamic changes in DNA methylation as well as occupancy by nucleosomes and non-histone regulatory factors lead to considerable epigenetic heterogeneity in chromatin.

Detection of the diverse epigenetic signatures present at a given region of interest by conventional footprinting methods poses several challenges (Pondugula and Kladde, 2008). Most of these stem from the nature of mapping DNA breaks introduced by nucleases or by genomic footprinting with chemical agents (e.g. dimethylsulfate). Limited digestion or chemical treatment of the footprinted sample is employed to achieve so-called single-hit kinetic levels of DNA cleavage, which are supposed to approximate a random Poisson distribution of cut sites. In practice, however, adherence to random Poisson statistics is hampered by biological complexity and non-randomness, especially when footprinted samples are of cellular origin. Second, even when single-hit digestion is achieved, only one cut site proximal to a radiolabeled DNA end, hybridizing primer, etc. can be mapped per single DNA molecule. Therefore, the position of nucleosomes or DNA-bound factors relative to one another on the same molecule, which requires mapping >1 cleavage site, cannot be determined. Third, a population of cut DNA molecules must be analyzed to identify a footprint. Such population-ensemble methods average away differences between molecules and thus obscure molecular heterogeneity.

These problems are overcome by MAPit; single-molecule detection of protein-DNA interactions by exogenously-supplied C-methylating DNMTs (Jessen et al., 2006; Kilgore et al., 2007; Pardo et al., 2009; Pardo et al., 2010). MAPit builds on a large body of earlier studies by us and others demonstrating the usefulness of DNMTs as

chromatin structural probes. To our knowledge, the earliest hints that chromatin structure might affect susceptibility to a DNMT were the preferential depletion of endogenous methyl-N6-adenine upon incubation of *Tetrahymena* nuclei with micrococcal nuclease (Pratt and Hattman, 1981, 1983). This suggested accessibility of linker DNA to the DNMT and its exclusion from nucleosome core DNA. Fehér et al. (1983) were the first to suggest that chromatin impeded access of specific sites in a yeast minichromosome to a C-5 DNMT expressed *in vivo*. This observation was repeated almost a decade later, when *E. coli* M.Dam was used to differentiate between "open" and "closed" chromatin in budding yeast (Gottschling, 1992; Singh and Klar, 1992). We subsequently demonstrated that positioned nucleosomes and factors bound site-specifically to DNA impeded accessibility of M.Dam (Kladde and Simpson, 1994).

With the advent of bisulfite sequencing for detection of m5C (Clark et al., 1994; Frommer et al., 1992), C 5 DNMTs became the logical choice for use as probes of protein-DNA interactions. M.SssI was used to probe chromatin structure first, because of its CG dinucleotide resolution and commercial availability (Kladde et al., 1996). In bisulfite sequencing, denatured DNA is subject to hydrolytic deamination of C to U with bisulfite ion, whereas m^5C is relatively non-reactive under optimal conditions (Hayatsu, 1976; Hayatsu et al., 2008). Initially, we used C-5 DNMTs as *in vivo* probes of chromatin structure in yeast and for *in vitro* footprinting of yeast and mammalian factors (Dong et al., 1999; Duan et al., 1999; Jessen et al., 2004; Kladde et al., 1996; Samudio et al., 2001; Vyhlidal et al., 2000; Xu et al., 1998b). In these studies, PCR products obtained from DNMT-probed and bisulfite-converted samples were sequenced directly, generating a population-averaged view of chromatin accessibility. Subsequent work by

us and others (Fatemi et al., 2005; Jessen et al., 2006) published within two months of each other, took the further steps to clone and sequence individual DNA molecules from PCR amplicons. This yielded the methylation status and hence accessibility state of the cytosine in each and every potential DNMT target site along single DNA strands; a powerful single-molecule view of chromatin accessibility. A later manuscript (Gal-Yam et al., 2006), introduced the name methylase-based single promoter analysis (M-SPA).

An important consideration for the utility of DNMT-base footprinting in vertebrate systems is to employ a DNMT with a sequence specificity that differs from the CG methylated by endogenous enzymes. To this end, we cloned the gene encoding the GC DNMT M.CviPI (Xu et al., 1998a). The first M.CviPI footprinting of single mammalian promoters proved its utility for chromatin-structure analysis (Kilgore et al., 2007; Pardo et al., 2010). A second advantage is that, unlike M.SssI, M.CviPI footprinting resolution is not limited by density of CG dinucleotides. Thus, MAPit need not be limited to studies of CG islands, and may well be extended to studies of chromatin at regulatory elements besides promoters.

As shown by the example in Figure 2.2, MAPit footprinting with M.CviPI has many advantages over other footprinting techniques. First, it is not at all subject to the constraints of single-hit kinetics, meaning that methylation of many CG and GC sites can be detected per sequenced molecule. This makes MAPit the only method capable of correlating footprints, i.e. sequential or cooperative binding events, along individual DNA molecules (Gal-Yam et al., 2006; Jessen et al., 2006). In contrast, as only the first DNA cut can be mapped in nuclease-based footprinting, there is much potential for multiple cuts to "mask" signal at locations farther removed from the mapping primer,

hybridization probe, etc. Second, MAPit data sets include molecules with no accessibility. Nuclease footprinting cannot score such molecules, as all signal generated from uncut molecules coalesces in the "parent" or "run-off" band. Lastly, the single-molecule view of footprints completely sidesteps population averaging and is thus able to detect distinct subclasses of molecules.

Critical Parameters and Troubleshooting

Exogenous DNMT, concentration, and treatment time

These are probably the most important variables to control when performing MAPit. Perhaps as expected, the wild-type M.CviPI polypeptide appears to be the most efficient probe for use in MAPit. This may be because DNMT fusion proteins have decreased affinity for DNA, catalytic activity, or both (Xu and Bestor, 1997). However, insolubility of wild-type M.CviPI led us to construct two commercially-available versions, M.CviPI fused to either maltose binding protein (MBP) or glutathione S-transferase (GST). While a high level of modification is desired for single-molecule footprinting, excessive DNMT activity (concentration and/or time of methylation) has the potential to physically compete for DNA-binding sites of proteins being footprinted. For example, we have observed that very high DNMT concentrations can invade the edges of nucleosomes *in vitro*. By no means is this a problem unique to probing with DNMTs, as all enzymes that act on DNA bind their substrate with measurable affinity. One advantage of DNMTs over nucleases is that multiple sites can be methylated per enzyme binding event. This is likely because DNMTs, like many proteins that associate with DNA, can slide or scan along DNA (Holz-Schietinger and Reich, 2010; Matsuo et al., 1994; Renbaum and Razin, 1992; Vilkaitis et al., 2005). That stated, we have not

observed nor are we aware of any situations in which DNMTs have displaced either site-specific DNA-binding factors from DNA or histone octamers from nucleosomes.

It is recommended that pilot experiments be conducted to optimize footprinting results. In principle, enzyme concentration, time, or both can be varied. We have opted, however, to vary enzyme concentration in pilot studies in keeping with most footprinting protocols. Longer times of incubation may also lead to potential loss of DNMT activity, hydrolysis of SAM cofactor, and dissociation of factors of interest from DNA. It is important to realize that methylation by exogenous DNA probes is irreversible during the methylation probing step. Therefore, factors that subsequently bind to methylated sites cannot be footprinted. It is equally important to use a consistent number of nuclei (i.e. mass of chromatin) in each experiment. The conditions indicated in this basic protocol (number of nuclei, DNMT dosages, time, solutions, and temperature) have been standardized to provide an adequate level of modification in our hands.

Buffer composition

Buffers adopted in this protocol have been previously established as maintaining the structure of native chromatin (Richard-Foy and Hager, 1987). Buffers can be changed to suit specific needs, but care needs to be exercised to avoid reagents that affect DNMT activity. High salt concentrations, for example, inhibit DNMT activity, which are also undesirable as they disrupt protein-DNA interactions. It is critical when using M.CviPI to add DTT to a final concentration of 10 mM immediately prior to conducting the chromatin methylation reaction. SAM, the universal cofactor and methyl donor for methyltransferases (Hermann et al., 2004a), hydrolyzes with repeated freeze-thaw cycles. It is therefore important to store SAM at −80°C as single-use aliquots and add freshly immediately prior to methylating chromatin.

DNA isolation

Thorough degradation of DNA-bound proteins with proteinase K is required in order to obtain DNA of high purity. Incomplete proteinase K treatment can interfere with the efficiency of bisulfite conversion (Warnecke et al., 2002). To avoid denaturation, proteinase K should not be vortexed. Digest for at least 16 hr at 50°C. Removal of RNA has also been reported to be necessary for efficient bisulfite conversion; however, in our experience, using the bisulfite treatment protocol described in (Darst et al., 2010), this does not appear to pose a problem. Perhaps this is because the described "home brew" method uses a solution saturated with sodium metabisulfite and thus contains a higher concentration of reactive bisulfite ion than most other protocols. In addition, RNA is completely hydrolyzed under the alkaline and high temperature conditions used to denature DNA prior to deamination.

PCR amplification of sequences from bisulfite-converted DNA

Performing PCR with deaminated DNA as template presents several challenges. The main hurdle is that, although the genome remains the same size, it is reduced in complexity by bisulfite conversion, i.e. it has reduced GC content. Considerations for PCR with deaminated templates are discussed extensively in (Darst et al., 2010). It is worth mentioning that DNA strands are no longer complementary after bisulfite conversion, so strand-specific amplification is determined by primer design (below). Ideally, one would design primer pairs for amplification of both strands of the locus of interest. Artifacts of DNA sequence can impair amplification or cloning of sequences corresponding to certain chromatin conformations at some loci, causing amplicon bias. Because each strand will produce a different sequence, they would not likely share amplicon biases. Comparison of data from both strands will therefore identify most

biases, which must be known for quantitative interpretation of MAPit data. Alternatively, bisulfite sequencing of a 50:50 mixture of placental DNA (primarily unmethylated):methylated DNA (genomic DNA methylated *in vitro* with M.SssI and/or M.CviPI) can be used to provide a direct test for amplification and cloning biases.
As it is single-stranded, deaminated DNA is prone to forming secondary structures that lead to spurious amplification. Performing hot-start PCR will avoid this amplification problem. In our hands HotStar Taq (Qiagen) has given good results with mammalian DNA templates. DNA polymerases can vary in tolerance to uracil containing templates, such as deaminated DNA. Long extension times of 2-4 min per kb can improve amplification yield, as can increasing the number of PCR cycles. Finally, for loci that are difficult to amplify, we employ PCR enhancers such as trimethylammonium chloride (TMAC; titrate concentration around 0.75 mM) or the Coral buffer supplied with HotStar Taq.

Primer design

Considerations for primer design are discussed in detail in (Darst et al., 2010). A main concern when working with native mammalian DNA or that which has been probed with the CG DNMT M.SssI is the presence of m5CG. In such samples, PCR primers for amplification of bisulfite converted samples are designed to avoid CG sites, which may be potentially methylated. When using MAPit with the GC probe M.CviPI to footprint protein-DNA interactions, avoid CG and GC sites with primer binding sites as much as possible. When this is not feasible, degenerate bases should be incorporated into primers to avoid PCR bias towards molecules in which the primer binding sites are either methylated or unmethylated. Conventional guidelines for primer design, PCR

conditions, and cycling parameters for PCR with bisulfite-converted DNA template are described in the PCR amplification step for bisulfite sequencing (Darst et al., 2010).

Results

MAPit analysis of a mammalian tumor suppressor gene promoter is shown in Figure 2.2 as an example of obtained results. The *SIM2* (single-minded 2) gene encodes a transcription factor that is highly expressed in breast tissue, where it has recently been reported to have tumor suppressor function (Kwak et al., 2007; Metz et al., 2006). We performed MAPit with wild-type M.CviPI on the immortalized human mammary epithelial cell line MCF10A. The zero M.CviPI control shows the level and distribution of endogenous m^5CG, as would be seen in any bisulfite sequencing experiment. For both m^5CG and G-m^5C, some background level is expected to result from incomplete deamination, base misincorporation during PCR, and sequencing error. The background can be estimated as equal to the percent unconverted cytosine outside methylation sites. Where endogenous m^5CG is not above background, as seen at SIM2, GCG methylation, which is otherwise ambiguous, can be ascribed to M.CviPI. This increases the resolution of MAPit.

The sequences reveal a nucleosome-free region of about 147 bp, located upstream of the TSS, and flanked by two protected areas that may accommodate at least one nucleosome on each side. Nucleosomes bound to DNA will generate protection footprints of ~150 bp. In a population of molecules, nucleosomal footprints can be shifted by several base pairs to either side due to different translational positions. Footprints comprising smaller sizes can be interpreted as DNA-bound factors, especially when located at known factor binding sites (Hoose and Kladde, 2006; Jessen et al., 2004; Kladde and Simpson, 1996; Xu et al., 1998b). Higher-order chromatin

structures and areas where nucleosomes are closely packed may generate larger footprints (Dechassa et al., 2010).

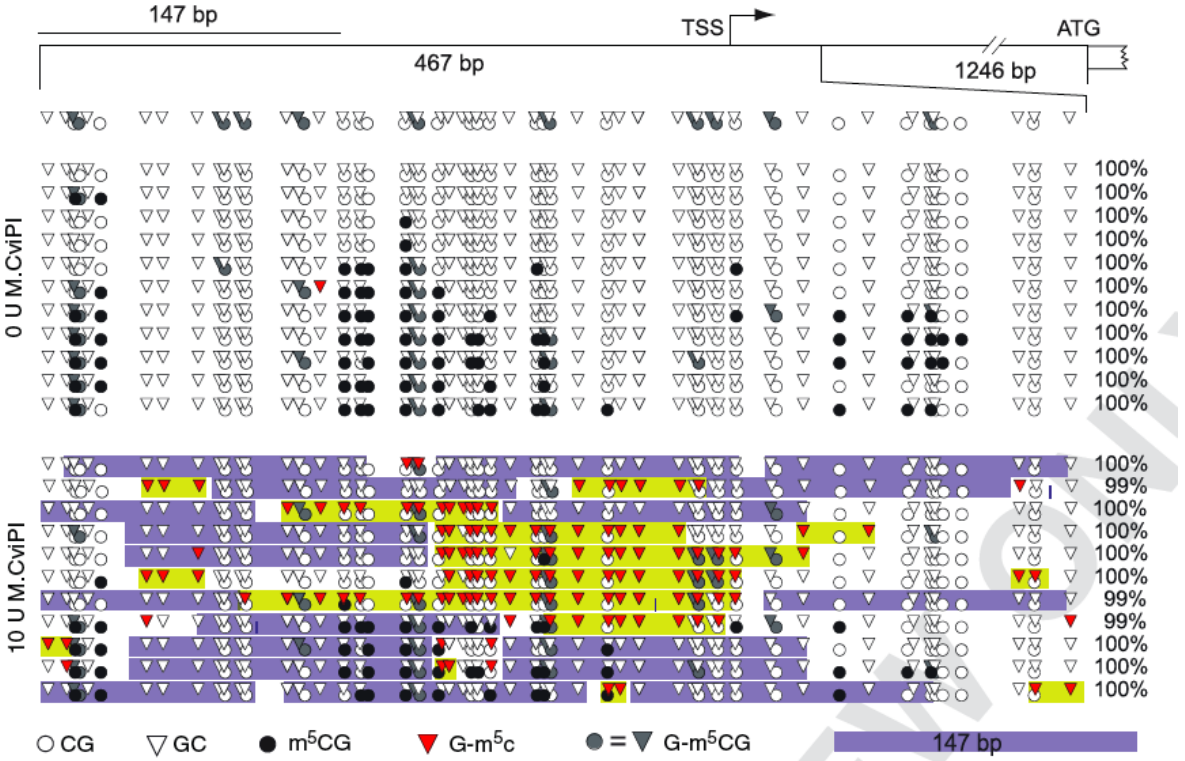

Figure 2-2. MAPit analysis of the TSS region of human SIM2 in MCF10A cells. Nuclei (106) were probed with 10 U of wild-type M.CviPI for 30 min at 37◦C. SIM2 is expressed in MCF-10A cells. Each horizontal line represents 524 bp of chromatin from a single cell. Circles represent CG sites and triangles represent GC sites. Black filled circles and red filled triangles, represent m5CG and G-m5C, respectively. GCG sites are represented by both gray triangles and circles. GCG site methylation cannot rigorously be discriminated as being placed by endogenous or exogenous DNMT, but this can often be inferred from context (Anticipated Results for discussion). Blue highlighted areas represent 147 bp of contiguous M.CviPI DNA footprint. Note that about half of the alleles have relatively high levels of endogenous methylation (black filled circles). Based on molecules from cells not treated with M.CviPI, it can be inferred that gray GCG sites in these densely methylated MCF-10A alleles were likely methylated by endogenous DNMTs. The other half of the molecules is almost free of endogenous methylation but shows an accessible, nucleosome-length region high in M.CviPI methylation (red triangles) highlighted in red. No other technique can determine this bipartite pattern of chromosome structure. The high accessibility to M.CviPI is probably due to histone depletion near the TSS. In contrast, this putative histone-free region is flanked by protected spans of median length ~150 bp. Numbers at the right of each molecule depiction indicate the percentage of C conversion to T in non-

CG and non-GC sequences. Nucleotides that failed to convert or reverted to a C during PCR amplification are indicated by vertical blue tick marks.

CHAPTER 3
MethylViewer: A COMPUTATIONAL ANALYSIS AND EDITING FOR BISULFITE SEQUENCING AND METHYLTRANSFERASE ACCESSIBILITY PROTOCOL FOR INDIVIDUAL TEMPLATES (MAPit) PROJECTS[2]

Introductory Remarks

Methylation of cytosine bases is a commonly occurring modification of deoxycytidine monophosphate (dCMP) pre-replicatively and of deoxycytidine (dC) in DNA post-replicatively (Bird, 2002; Feng et al., 1978; Goll and Bestor, 2005; Grossman, 1981). Post-replicative methylation of dC in prokaryotic genomes, either of carbon 5 (m^5C) or of the exocyclic N4 atom ($m^{N4}C$), is mediated by site-specific DNA methyltransferases (DNMTs). These modifications protect against degradation by cognate restriction endonucleases (Wilson and Murray, 1991).

Many organisms regulate diverse genomic processes through C-5 methylation of CG (CG) sites (i.e. m^5CG) by endogenous DNMTs. These processes include inactivation of: gene expression at the level of transcription initiation or elongation, one of two X chromosomes in normal female mammals, either the maternal or paternal copy of a gene (genomic imprinting) and mobility of parasitic genetic elements, e.g. retrotransposons (Bestor and Bourc'his, 2004; Bird, 2002; Goll and Bestor, 2005). Aberrant patterns of DNA methylation are frequently associated with human disorders, aging, carcinogenesis and developmental defects (Bird, 2002; Feinberg et al., 2006; Jaenisch and Bird, 2003; Jones and Baylin, 2007; Robertson, 2001, 2005). In addition

[2] Reprinted with permission of: Oxford University Press. As published in: Carolina E. Pardo, Ian M. Carr, Christopher J. Hoffman, Russell P. Darst, Alexander F. Markham, David T. Bonthron and Michael P. Kladde. 2010. MethylViewer: Computational analysis and editing for bisulfite sequencing and methyltransferase accessibility protocol for individual templates (MAPit) projects. Nucleic Acids Research. 39.

to CG methylation, land plants, such as Arabidopsis, employ non-CG methylation of transposable elements (Cokus et al., 2008; Feng et al., 2010; Goll and Bestor, 2005; Henderson and Jacobsen, 2007; Zemach et al., 2010; Zilberman et al., 2007). Short transposable elements are particularly enriched for m^5CHH, where H is a degenerate base equal to A, C or T. *Ascomycetes*, such as the filamentous fungi *Ascobolus immersus* and *Neurospora crassa*, 5-methylate dC within repeats in a relatively sequence-independent manner (Selker, 1990; Zemach et al., 2010). Abundant non-CG methylation (m^5CHG and m^5CHH) has also been reported in undifferentiated human embryonic stem cells (Grandjean et al., 2007; Hawkins et al., 2010; Kouidou et al., 2005; Latham et al., 2008; Laurent et al., 2010a; Lister et al., 2009). Non-CG methylation is also found in the genomes of lytic *Chlorella* viruses that infect and degrade the genome of fresh water Chlorella-like green algae (Chan et al., 2004a; Nelson et al., 1998; Nelson et al., 1993; Xu et al., 1998a). In this case, progeny viral genomes are protected post-replicatively against degradation by site-specific introduction of m^5C by DNMTs, e.g. M.CviPI (G-m^5C) and M.CviPII (m^5CCD, m^5Cm^5CAA and m^5Cm^5CCG), that they encode (Chan et al., 2004a; Nelson et al., 1998; Nelson et al., 1993; Xu et al., 1998a).

Bisulfite genomic sequencing (BGS) is a widely used technique for assaying cytosine methylation status in DNA (Clark et al., 1994; Frommer et al., 1992). Bisulfite ion can quantitatively convert unmethylated C in denatured DNA to U, whereas m^5C, $m^{N4}C$ and hm^5C resist chemical deamination and thus are retained (Hayatsu and Shiragami, 1979; Huang et al., 2010; Vilkaitis and Klimasauskas, 1999). BGS can be used to display these three and perhaps other types of cytosine modification as they are

discovered (Klimasauskas et al., 2002). Individual molecules from a PCR product population can be cloned and sequenced, mapping the methylation status of every cytosine along a single DNA strand at nucleotide resolution (Clark et al., 1994; Frommer et al., 1992).

DNMTs are also effective probes, either as purified enzymes or as transgenes expressed *in vivo*, for mapping protein–DNA interactions (Gottschling, 1992; Jessen et al., 2004; Jessen et al., 2006; Kilgore et al., 2007; Kladde et al., 1996; Pardo et al., 2009; Singh and Klar, 1992). DNMTs methylate target sites that are unoccupied by proteins and hence are easily accessed (Gottschling, 1992; Jessen et al., 2004; Kladde and Simpson, 1994; Kladde et al., 1996; Singh and Klar, 1992). Conversely, occupancy of DNMT target sites by histones or non-histone proteins protects against methylation and produces a 'footprint' at the site of factor interaction (Kladde and Simpson, 1994; Kladde et al., 1996; Samudio et al., 2001; Vyhlidal et al., 2000; Xu et al., 1998b). Availability of DNMTs that recognize and methylate cytosines in specific dinucleotide sites, e.g. M.CviPI (GC) and M.SssI (CG) (Nur et al., 1985; Renbaum et al., 1990; Xu et al., 1998a), allows moderate- to high-resolution footprinting, depending on local site density. Additional enzymes with novel short recognition specificities would increase probing resolution and hence obtainable mapping information.

Combining DNMT probing with BGS provides a powerful non-averaged, single-molecule view of chromatin structure, termed methyltransferase accessibility protocol for individual templates (MAPit). Because a continuous strand of DNA is cloned, i.e. regions are not fragmented by nuclease digestion, multiple methylation events and hence footprints can be detected along a single DNA strand (Fatemi et al., 2005; Gal-

Yam et al., 2006; Jessen et al., 2006; Kilgore et al., 2007; Pardo et al., 2009; Pondugula and Kladde, 2008). Manual assignment of the methylation status of every C of many sequenced molecules is time consuming, labor intensive and subject to human error. Manual analysis of raw BGS results is therefore impractical for large-scale projects.

Attempts to alleviate these problems have led to development of several software programs that deal with individual steps in the BGS process: (i) primer design (Li and Dahiya, 2002; Ordway et al., 2005); (ii) alignment of sequenced bisulfite-converted molecules (Bock et al., 2005; Rohde et al., 2008; Xu et al., 2007); and (iii) generation of graphical or text-based outputs (Bock et al., 2005; Grunau et al., 2000; Hetzl et al., 2007; Ordway et al., 2005; Singal and Grimes, 2001; Xu et al., 2007). Some of these programs require computer literacy beyond that of the typical biomedical researcher. In addition, only one program, CGviewer (Carr et al., 2007), which we developed previously, offers a stand-alone solution to the BGS pipeline. However, CGviewer and each of the above programs is limited to methylation analysis of CG, or additional sites in plants, CHG, CHH and CNG, where N is any nucleotide and H is either A, C or T. Furthermore, currently available programs are unable to simultaneously analyze and diagram methylation by multiple enzymes, which may vary from one experiment to another, as is often required for MAPit footprinting studies. We have developed a BGS analysis program, called MethylViewer, which circumvents these and other limitations. MethylViewer was written to facilitate analysis of MAPit experiments, and is a substantially improved and more versatile version of CGviewer (Carr et al., 2007). A key feature of MethylViewer is that it can simultaneously analyze and distinguish between cytosine methylation in bisulfite-converted sequences at as many as four different, user-

defined sequence motifs, including C by itself. The program also has a utility for designing primers for MAPit, directly aligns most common sequencing file formats or processes pre-aligned FASTA files, and generates publication-quality images. Here, the usefulness of MethylViewer is demonstrated by analyzing *.ab1 sequencing files from MAPit analysis of the *hHMLH1* tumor suppressor gene promoter from colorectal cancer cell lines. Our use as chromatin probe of M.CviPI, which we previously cloned and found methylates GC (Xu et al., 1998a), permitted simultaneous detection of endogenous CG methylation and chromatin accessibility within single mammalian sequences. As such, our data provide the first single-molecule view of chromatin accessibility at an endogenously hypermethylated CG island. Also, at a well-characterized yeast promoter, we showcase for the first time use of M.CviPII, a CCD DNMT, as an *in vivo* chromatin probe. We conclude that MethylViewer is a powerful computational resource for accurate and rapid BGS analysis of complex DNA methylation data sets, including those with methylation at degenerate or multiple sites of any sequence of interest.

Materials and Methods

Software Development and Requirements.

Microsoft Visual Studio 2005 was used for programming using the Visual Basic language. The MethylViewer program has been tested only on Microsoft Windows XP, Vista and Windows 7, and requires installation of .NET framework 2.0. The stand-alone graphical user interface program and accompanying documentation are freely available for download at http://dna.leeds.ac.uk/methylviewer/.

Cell Lines, Yeast Plasmids and Strains, Growth Media and Cell Culturing

HCT116 and RKO colorectal cancer cell lines were obtained from the American Type Tissue Culture Collection. Cells were cultured in minimal Eagle's medium (MEM) supplemented with 10% (v/v) fetal bovine serum and 1% (w/v) penicillin and streptomycin, following the provider's recommendations.

The full-length coding region for M.CviPII (Chan et al., 2004b) was subcloned under control of estrogen induction into a single-copy integration vector as described earlier in Jessen et al. 2004 (Jessen et al., 2004). The resulting LYS2-marked plasmid, pCF1439, was digested with R.AscI and R.SalI and integrated as a single copy at the HO locus in budding yeast strain, SCY3854 (MATa leu2Δ0 lys2Δ0 ura3Δ0 pho3Δ::R $PHO5^{HhaI}$ promoter) (Jessen et al., 2006). The $PHO5^{HhaI}$ promoter contains several single nucleotides substitutions into the wild-type PHO5 promoter that introduced several HhaI sites and was described previously (Jessen et al., 2006). R in this strain is a single copy of the recombinase site for *Zygosaccharomyces rouxii* that is a remnant after removal of the marker used to delete the PHO3 coding region (Roca et al., 1992). The constructed strains CFY4011 and CFY4012 were grown to mid-log phase in rich YPD medium [1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose] and grown 10 h more in the presence of 100nM 17b-estradiol to induce M.CviPII expression.

Nuclei Isolation and MAPit Analysis

Nuclei from HCT116 and RKO cells were isolated from ~ 4–7 X 10^6 cells at 4ºC under buffer conditions that preserve the integrity of nuclei and chromatin structure (Kilgore et al., 2007). After harvesting, cells were washed twice with phosphate-buffered saline (PBS) and resuspended in 1X cell resuspension buffer [CRB; 20mM HEPES, pH

7.5, 70mM NaCl, 0.25mM EDTA, 0.5mM EGTA, 0.5% glycerol (v/v), 10mMDTT, 0.25mM phenylmethylsulfonyl fluoride]. After pelleting by centrifugation at 1000 g, cells were resuspended in cell lysis buffer (CLB; 1X CRB plus 0.19% NP-40) for 10 min on ice. Nuclei were then washed twice with CRB and a 2 μl aliquot was stained with 4% (w/v) trypan blue and visualized by light microscopy to confirm their integrity. One million nuclei were resuspended in methylation buffer (MB; 1X CRB plus 160 mM S-adenosyl-L-methinonine). After pre-warming nuclei to 37°C for 5 min, 100U of M.CviPI (New England Biolabs) were added for 15 min at 37°C. Methylation reactions were stopped by adding an equal volume of 2X methylation stop buffer [MSB; 100mM NaCl, 10mM EDTA, pH 8.0, 1% SDS (w/v)], and then incubated overnight with 100 mg/ ml proteinase K at 50 C. DNA was isolated by extraction with phenol:chloroform:isoamyl alcohol (25:24:1) and concentrated by ethanol precipitation.

m^5C Analysis

Bisulfite conversion of total genomic DNA from yeast and cultured mammalian cells was carried out as described in Darst et al. 2010 (Darst et al., 2010). Briefly, 1–2 ug DNA were denatured in alkali for 15 min at room temperature, followed by 5 min at 98°C. Saturated sodium metabisulfite solution, pH5.0 was added directly to each denatured sample, vortexed and incubated in the dark for 4–6 h at 50°C. Bisulfite-converted DNA was desalted, desulfonated and purified with the EZ bisulfite DNA clean-up kit (Zymo Research). Oligonucleotide primers were designed as described in Pardo et al. 2009 (Pardo et al., 2009) to avoid as much as possible not only potential endogenous methylation sites (CG) but also M.CviPI (GC) sites. A 755-bp region encompassing the two transcription start sites (TSS) of *hHMLH1* was analyzed for DNA

methylation. Hot-start PCR was performed with HotStar Taq Plus (Qiagen). To minimize stochasticity due to PCR, at least three separate PCR reactions were performed for each amplicon and pooled prior to cloning individual products. Oligonucleotides CPO1842 (TaaATaTaAACAaATaATTTCTaaAATaAATa, with G to a transitions) and CPO1843 (GGAGGGAYGAAGAGAttt, with C to t transitions and one degenerate pYrimidine, i.e. C and T) were used for PCR amplification of the upper strand of *hHMLH1* sequences from bisulfite-converted DNA. For the yeast experiment, genomic DNA was prepared by the phenol/chloroform extraction method (Adams et al., 1997) and deaminated as in (Shiraishi et al., 1999). DNA methylation was confirmed by McrBC digest (Stewart et al., 2000). PHO5 promoter sequences (500 bp) were amplified from bisulfite-converted DNA with primers WJO766 (ATATATCTCGAGAATATaTCAACaTATTTaaAAaTCATCTTATa; 50 XhoI site underlined; lower case indicating G to a transitions) and WJO769 (ATATATAAGCTTCAAtATTGGTAATtTtGAATTTGtTTGtT; 50 HindIII site underlined; lower case indicating C to t transitions). After purification of PCR products with the QIAEX II gel extraction kit (Qiagen), individual DNA molecules were cloned by either TOPO TA cloning (Invitrogen; *hMLH1* sequences) or directionally into pBluescript (Stratagene) digested with R.XhoI and R.HindIII (PHO5 sequences). White Escherichia coli colonies that screened positive for potential recombinant plasmids were inoculated into 100 ml LB medium with 100 mg/ml kanamycin (TOPO TA cloning) or 100 mg/ml ampicillin (directional cloning), and incubated overnight at 37 C. Plasmid DNA was amplified directly from *E. coli* cells using TempliPhi rolling circle amplification (GE Healthcare). Cloned inserts in the resulting single-stranded DNA were subjected to

BigDye sequencing and analyzed on an automated 3730 sequencer (Applied Biosystems) at the Interdisciplinary Center for Biotechnology Research at the University of Florida. The PHO5 and *hMLH1* *.ab1 files can be downloaded using: http://dna.leeds.ac.uk/methylviewer/Example_Files_ h*MLH1*.zip and http://dna.leeds.ac.uk/methylviewer/ Example_Files_PHO5.zip, and are also available at NAR Online. Note that the wild-type reference sequences should be moved to location on the hard drive outside of the folder containing the *.ab1 files prior to analysis. Analysis of bisulfite-converted sequences The alignment algorithm in MethylViewer uses an array of overlapping DNA fragments created from the wild-type reference sequence that can be changed from the 10-bp default length to word sizes of 6–15 bp via Analysis>Alignment option>Word size . . . . These words are then used to find regions of 100% homology between the reference sequence and the query sequence. Each region is extended until the ends of the alignment have < 80% identity. Local alignments are then concatemerized to create the longest possible global alignment with the minimum number of gaps. Since the sequence of bisulfite-treated DNA may be significantly different from the original sequence, alignments are created between both the original sequence and the theoretical bisulfite treated reference sequence to ensure that an alignment is created. Similarly, the alignment is also performed against the forward and reverse complement of the native and bisulfite-treated reference sequence to identify the orientation of the cloned DNA. The reference sequence is scanned for the presence of any methylation sites and this information is superimposed on the alignment to identify the methylation status of each site. This information is then collated

for each of the experimental sequences and used to form the interactive grid displayed by MethylViewer as well as other exported images.

Results

MAPit for Simultaneous Detection of Endogenous m^5CG and Chromatin Accessibility in Individual Mammalian DNA Strands

Transcriptional inactivation of the human tumor suppressor gene *hMLH1* by hypermethylation of the CG island associated with its promoter is a frequent contributor to colon cancer progression (Bronner et al., 1994; Esteller et al., 1998; Veigl et al., 1998). The *hMLH1* promoter is active in HCT116 cells where its CG island is hypomethylated (Esteller, 2007a; Esteller et al., 2001; Feinberg et al., 2006; Herman and Baylin, 2003; Veigl et al., 1998). By contrast, the *hMLH1* CG island is hypermethylated and the gene is transcriptionally silent in RKO cells. In a recent study, Lin et al. (Lin et al., 2007) reported that chromatin near the two transcription start sites (TSS), TSSa and TSSb, of active alleles of *hMLH1* in HCT116 cells was accessible to nucleases and M.SssI at individual promoters within isolated nuclei. Regions upstream of both TSS in this cell line were also shown to be depleted for nucleosomes as assayed by chromatin immunoprecipitation of the globular region of histone H3 in isolated mononucleosomes. However, single-molecule methylation footprinting of the hypermethylated *hMLH1* promoter in RKO cells using M.SssI was ambiguous because the probe has the same specificity for CG as vertebrate DNMTs. We have overcome this drawback by using as chromatin probe M.CviPI, a DNMT from Chlorella virus NYs-1 that specifically methylates C in GC sites (Xu et al., 1998a). Thus, sequencing cloned, bisulfite-converted molecules after probing mammalian chromatin with M.CviPI should allow clear discrimination between non-overlapping Gm^5C and endogenous m^5CG

within accessible regions of mammalian chromatin. We performed MAPit by probing nuclei isolated from HCT116 and RKO cells that respectively express and silence the *hMLH1* promoter with 0 or 100 U of M.CviPI. The original *.ab1 sequence files, FASTA alignments of four sequences from each sample, MethylViewer, and other files to familiarize users with the program's features are available for downloading on the companion website (http://dna.leeds.ac.uk/methylviewer/download.php). This site also contains detailed instructions for use of the program.

Analysis of Mammalian MAPit Data with Methylviewer

Users are first prompted to Select a function from the menu above when MethylViewer is executed (Figure 3-1A). The program supports two different means for analysis of bisulfite-converted sequences (Figure 3-1B). First, raw sequencing data in either *.txt, *.ab1, or *.scf file format can be aligned without pre-processing by Analysis>Interactive view . . . . A *.txt reference that is the wild-type sequence is required, avoiding the need to generate bisulfite-converted reference sequences *in silico*. Because MethylViewer can interpret methylation at non-palindromic sites, and DNA strands are no longer complementary after bisulfite conversion, the wild-type reference sequence must be that of the DNA strand that was sequenced. After navigating to and selecting the reference sequence file and the folder containing data files, the program generates the alignment via a local extension algorithm that is similar to BLAST (Altschul et al., 1990).

Analysis>Import FASTA alignment provides a second analysis option (Figure 3-1B). After clicking on this option, users navigate to a FASTA file in the format shown in Figure 3-2 that contains sequences pre-aligned by another program. The first entry in this *.txt file must be the wild-type reference sequence of the DNA strand being

analyzed to which the bisulfite-converted sequences are aligned. Choosing either option in the Analysis menu launches a floating window for designating the sequences of sites at which methylation is to be scored (Figure 3-3A). Preset selection buttons for CG only, GC only and CG and GC are provided. Presets for CG only or GC only employ as defaults the usual convention of depicting unmethylated CG or GC as white-filled circles and methylated sites (m^5CG or G-m^5C) as black-filled circles, organized in horizontal rows.

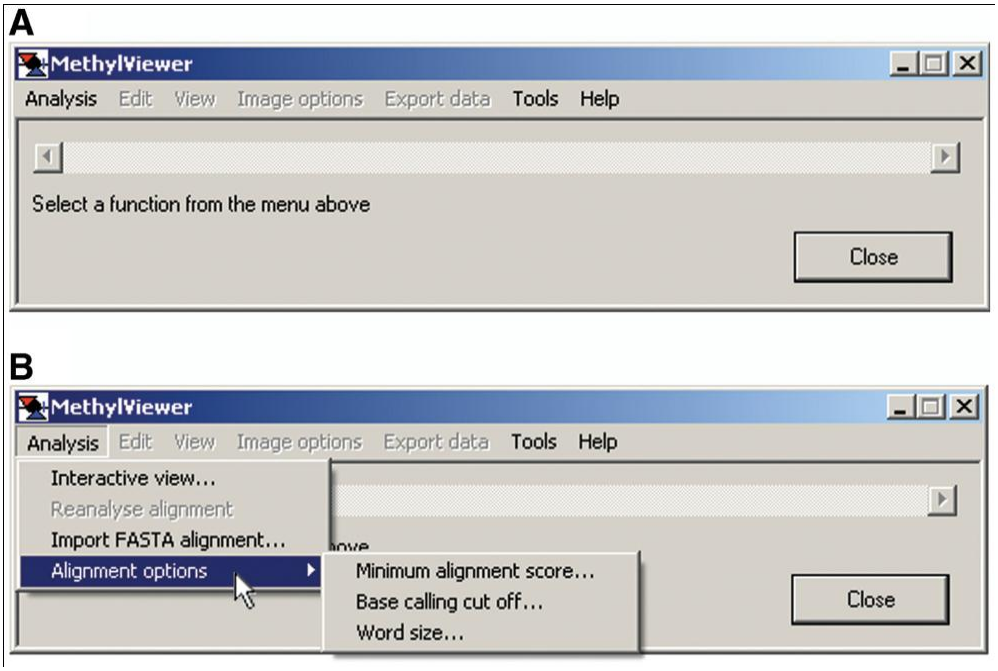

Figure 3-1. MethylViewer initial window. The initial window after loading MethylViewer.exe A) and Analysis>Alignment options B). Minimum alignment score . . . can be set with increased stringency from 15 to 50 upon activating the pop-up window. Likewise, users can set the Base calling cut off . . . for the signal threshold (15 to 300) for calling bases in sequencing traces and Word size . . . of the base pair length of the array of overlapping DNA fragments (6 to 15) created from the wild-type reference sequence for use in the BLAST-like local extension algorithm. Clicking the mouse on Interactive view . . . or Import FASTA alignment . . . activates the Methylation sites window shown in Figure 3-3A .

```
>hMLH1_Wild-Type_Reference
GGAGGGACGAAGAGACCCAGCAACCCACAGAGTTGAGAAATTTGACTGGC
ATTCAAGCTGTCCAATCAATAGCTGCCGCTGAAGGGTGGGGCTGGATGGC
GTAAGCTACAGCTGAAGGAAGAACGTGAGCACGAGGCACTGAGGTG
>HCT116_hMLH1_A01 <ID:1071327>
GGAGGGATGAAGAGATTTAGTAATTTATAGAGTTGAGAAATTTGACTGGT
ATTTAAGTTGTTTAGTTAATAGTTGTTGTTGAAGGGTGGGGTTGGATGGT
GTAAGTTATAGTTGAAGGACGAATGTGAGTATGAGGTATTGAGGTG
>HCT116_hMLH1_D01 <ID:1071363>
GGAGGGATGAAGAGATTTAGTAATTTATAGAGTTGAGAAATTTGATTGGT
ATTCAAGTTGTTTAATTAATAGTTGTTGTTGAAGGGTGGGGTTGGATGGT
GTAAGTTATAGTTGAAGGAAGAATGTGAGTATGAGGTATTGAGGTG
```

Figure 3-2. FASTA file format. The wild-type sequence and proper strand of the region of interest must appear as the first entry. Names of individual aligned sequences can include any number of alpha-numeric characters; however, the program will reduce the font size of labels in exported images to fit within the allotted space. Users must be certain to erase all non-alpha-numeric characters from the file, except>and <.

The default setting, CG and GC, facilitates analysis of MAPit data for vertebrate samples that require analysis of methylation status at endogenous CG sites as well as at GC sites probed by M.CviPI exogenously added to nuclei. The default symbols and colors can be visualized and changed by clicking the Custom preset button (Figure 3-3A), which launches the Custom methylation sites floating window (Figure 3-3B). As usual, CG and m^5CG are depicted as white- and black-filled circles, respectively. White- and red-filled inverted triangles are the depictions for GC and G-m^5C sites, respectively. Selecting GC: 2 in the toggle window under Include the site displays the default settings of methylation of residue 2 and red color (not shown). The options buttons for circle and inverted triangle symbols in the Select symbol bar at the bottom of the window are gray and unavailable as they have already been designated for scoring CG and GC sites. Overlapping methylation, e.g. at GC and CG sites (G-m^5CG), is indicated as gray-filled cells in each grid. This is because they constitute ‘gray areas’ that in principle could have been methylated by either endogenous DNMTs or M.CviPI chromatin probe. This default convention cannot be changed and thus does not appear in the Custom methylation sites window.

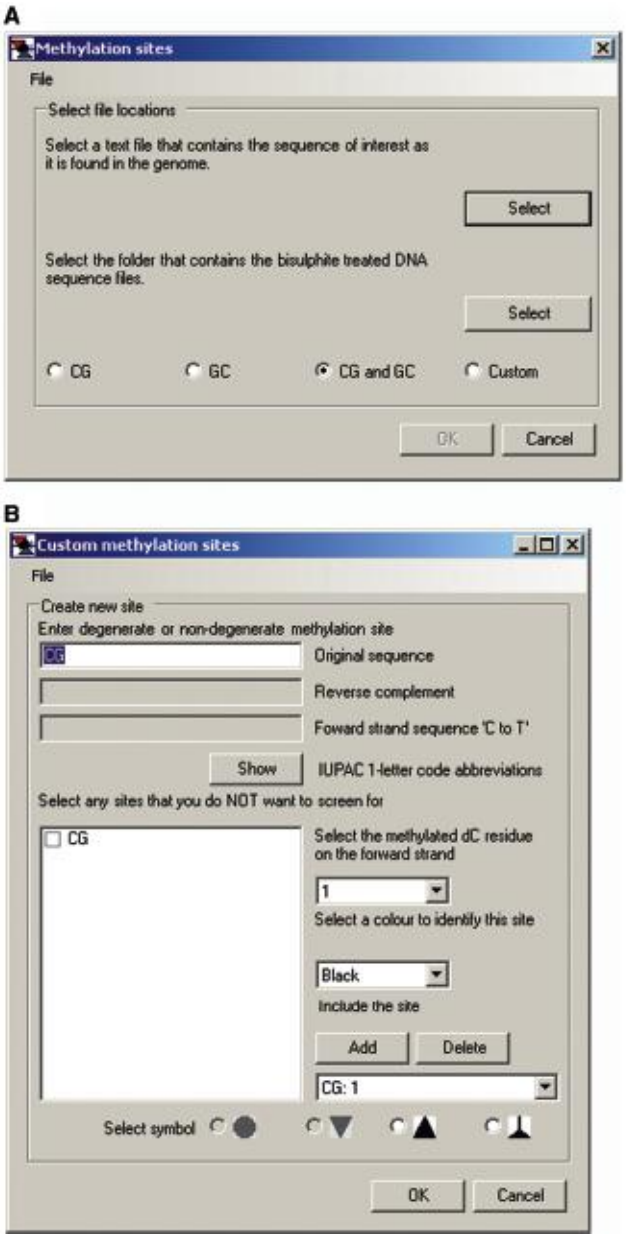

Figure 3-3. Default settings for MAPit methylation footprinting analysis. A) Methylation sites window activated by clicking on Analysis>Interactive view . . . or Import FASTA alignment . . . . If the default setting of CG and GC (keys in Figures 3-4A and 3-10D) is acceptable, the user then selects the indicated files as prompted in the window. B) Custom methylation sites window activated by clicking the Custom button in A. Shown are the default settings for CG methylation analysis: methylated dC residues at position 1 indicated in black. Note that the Include the site field indicates CG: 1. Toggling to GC: 2 shows the second setting, methylation of dC residues at position 2 in GC sites to be indicated in red. The circle and triangle symbols are gray because they are set as defaults for indicating methylation status of CG and GC sites, respectively, in exported images (e.g. Figure 3-10).

Viewing Mammalian Mapit Data by MethylViewer

User-defined symbols and colors are drawn in Export data>Save image drawn to scale as discussed further below. Initially, however, MethylViewer processes and shows the methylation status for every C in sites aligned to the reference sequence as an interactive grid composed of color-coded cells as chosen by the user. Figure 3-4 shows representative BGS grids generated from *.ab1 files of four bisulfite-converted *hMLH1* promoter molecules that were amplified, cloned and sequenced from HCT116 and RKO nuclei incubated with either 0 or 100 U M.CviPI. The default CG and GC setting was used for analysis of the 0U M.CviPI samples so that the number of scored sites and hence grid sizes would be identical to aid comparisons of different samples. MethylViewer can be executed multiple times to generate as many grids as needed.

Rows 2–5 of each grid correspond to residues in DNMT sites in each of the four bisulfite-converted sequences from the indicated cancer cell line ±M.CviPI. The key in Figure 3-4A is for grids generated with the default MAPit settings shown in Figure 3-3B. A key with colors assigned to each methylated site can also be displayed in row 1 of grids by selecting View>Identify methylation sites (Figure 3-4C). The size of grid cells can be increased via View>Cell width of interface>Large to better visualize assigned colors, especially if more DNMT sites are analyzed. Placing the cursor over any cell in a grid, except for header blue cells in column 1, also displays corresponding site sequence(s) in a small pop-up window (Figure 3-4B, D and E). If sequencing data files have been analyzed, left-clicking on these header cells (left- or right-clicking for analysis of FASTA files) indicates the site number and position relative to base pair 1 in the wild-type reference sequence (Figure 3-4D). Cells representing residues in DNMT sites that sequenced as T are colored white. Cells for non-overlapping methylation that

sequenced as C are colored according to the program's default settings (Figure 3-4A), or as otherwise designated in Custom methylation sites. With a view toward analysis of large data sets, Analysis>Reanalyze alignment allows analysis of methylation at additional sites while the window is open (Figure 3-1B).

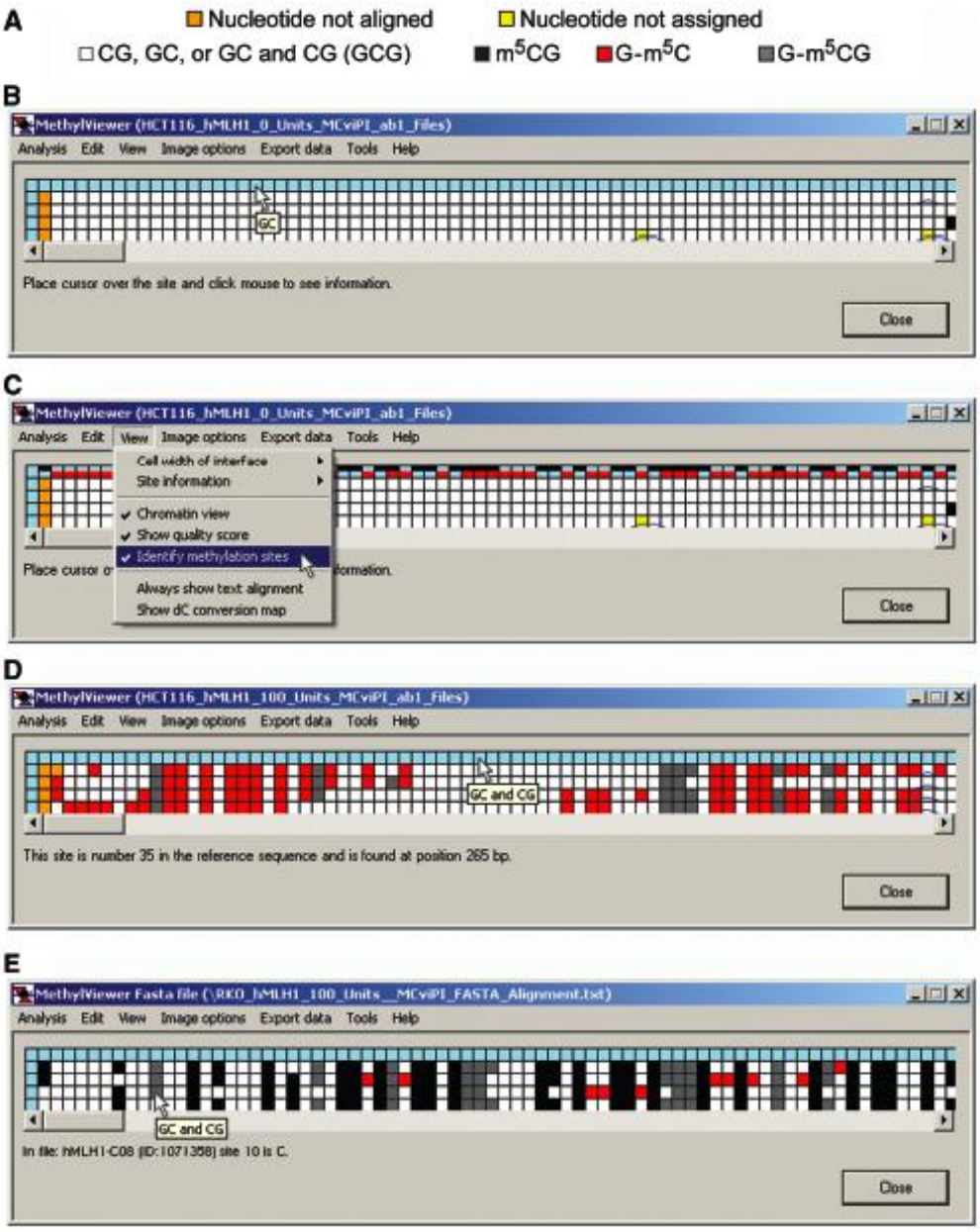

Figure 3-4. MethylViewer interactive data grids. A) Data key. B–E) Interactive grids showing raw, unedited methylation data processed from *.ab1 files via Analysis>Interactive view . . .> CG and GC for HCT116 nuclei treated with 0U M.CviPI (B), HCT116 nuclei+100U M.CviPI (C), RKO nuclei+0U M.CviPI (D)

and RKO nuclei+100U M.CviPI (E). The shown grids were generated with the default Base calling cutoff of 20 and maximum alignment Word size of 15. Each column, except for column 1, contains data for a specific potential methylation site in the reference sequence. Each row, except for the header row 1, represents the unedited sequencing calls (C or T) for each of four cloned and sequenced molecules. Placing the cursor over any cell in the grid, except for those in the header column 1, displays the corresponding sequence of the scored site in a small pop-up window as shown in B, D and E. An alternate way to view sites is to select View>Identify methylation sites, which displays the colors of each methylated dC site as shown in C. In addition, clicking on a light blue or color-coded header cell in row 1, except the first cell, displays specific information about a site as shown in D. Lastly, left-clicking on any cell in the grid displays the file number and sequence of that scored site (C or T) as seen in E. Pop-up windows and sub-menus only obscure white cells with unmethylated residues.

Data Editing

MethylViewer has been written to facilitate accurate analysis of BGS data; a central feature therefore is user-friendly editing of sequences within the interactive grid interface. These capabilities include viewing the quality of both sequencing data and local alignment of sites and flanking sequences, and re-assigning methylation status, if warranted. Cells representing sites in the grid at which local sequencing data could not be aligned to the reference sequence are colored orange (Figure 3-5A), and cannot be edited. In contrast, residues within sites of molecules that were aligned to the reference sequence but unable to be assigned as C (unmethylated) or T (methylated) appear as yellow cells (Figures 3-4B and C and 3-5A and B). This can be due to poor quality sequencing data, the presence of single nucleotide polymorphisms, or mutations that occurred during PCR amplification or cloning. Right-clicking on header blue cells in column 1 for any molecule offers the option in a pop-up window to view the entire ABI trace (not shown) of the sequencing data and examine its quality (Figure 3-5A).

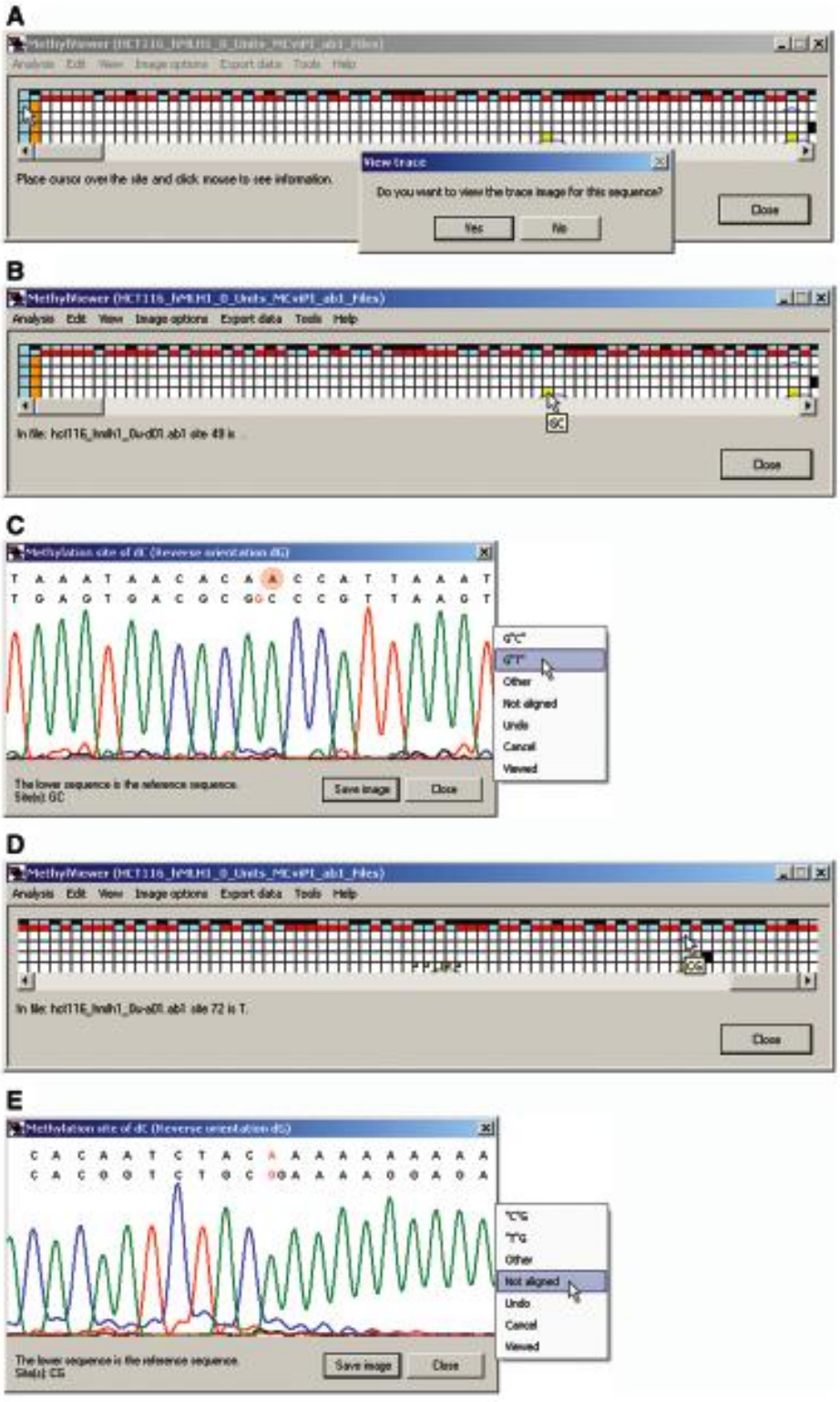

Figure 3-5. Data editing. A) Right-clicking on the light blue header cell offers the option to view the entire sequencing trace of a molecule. B) Right-clicking on any non-header cell in the grids displays C) the local alignment between the reference and query sequences as well as the sequencing trace

encompassing the query site. The methylatable residue in both the reference and query sequences is indicated in red type in properly aligned sequences. In this example, the A residue in the upper query sequence (manually highlighted by red circle; reverse strand sequenced) is in black type, because it should have been aligned to the G in the reference sequence. That is, the gap in the alignment (indicated in A and B as a blue tracer line within the corresponding cell) should be placed at one of the three immediately downstream C residues. Therefore, clicking anywhere in the white field of this local alignment window displays a pop-up window to correct assignment of the nucleotide sequence; specifically, G‘T’, because the corresponding nucleotide in the query molecule sequenced as ‘A’ (when aligned correctly). The base-calling cutoff of 20 is indicated by the red horizontal line at the base of the peaks. D) The corrected G‘T’ appears as a large white square within the original yellow cell for that nucleotide. Cells for which sequencing data was inspected and verified, and hence had unchanged sequencing calls, are marked by small green squares. The window scroll bar has been moved all the way to the right in this panel. E) Right-clicking on the CG cell marked by the cursor in D showed deletion of an A in the query molecule. As this A or any other in the downstream run of adenines could have been deleted, the nucleotide assignment of the residue was changed to Not aligned. If FASTA files are used as input, right clicking on any non-header cell in the grid will display the entire text alignment between the reference and query sequence, with the inspected nucleotide marked by an asterisk.

Left-clicking on any non-header (non-blue) cell in the grid indicates the sequence of a residue, the name of the sequenced file and the site number relative to site 1, the first one aligned (Figure 3-3B). Nucleotide numbers relative to site 1 associated with yellow grid cells are left blank because their sequence could not be unequivocally determined by MethylViewer. As all residues in DNMT sites should be C or T, residues that sequenced as G or A are also colored yellow (not shown).

In grids of *.ab1 sequencing files, right-clicking on any non-blue cell allows inspection of both the ABI data trace and alignment between the reference and query sequence around that DNMT site (Figure 3-5C). This allows users to verify sequencing calls made by MethylViewer as well as to determine if the quality of sequencing data associated with yellow cells merits assigning a specific nucleotide to it. In each sequencing trace, the residue indicated in red corresponds to that clicked on and thus

being queried for methylation status (Figure 3-5C). In this example and throughout, the G rather than C in GC site 49 was queried because the reverse strand was sequenced. It is clear in the fourth analyzed clone that one of the three cytosines inclusive of the C in GC site 49 was deleted (indicated by a blue trace over cells 49 and 50 in row 5 corresponding to sequenced clone 4 of Figure 3-5B). Therefore, correct alignment between the red G in the reference and A in the query sequence supports assignment of site 49 as T or unmethylated. Clicking anywhere within the white area of the sequencing trace launches another pop-up window with several editing options (Figure 3-5C). Selection of G‘T’ re-assigns the cell to its proper call of T. Such edited squares are identified by a square in the upper left corner of the original cell in the color scheme originally chosen for the grid (Figure 3-5D). This square is intermediate in size between the tracking green square and the data cell. If no change to an interrogated cell is warranted, simply closing the ABI trace or selecting Viewed from the pop-up window marks the cell with a small tracking green square in its upper left corner (Figure 3-5D).

MethylViewer also marked CG site 72 in two molecules with blue traces, because they had deletions of one (row 2) or two (row 5) G residues immediately downstream of the query G (Figure 3-5D and E). The nucleotide in CG site 72 in row 2 was aligned to the reference and called as T (A in the sequenced reverse strand). However, it is formally possible that this residue was methylated and deleted instead of the downstream G, and thus we elected to change its assignment to Not aligned (Figure 3-5E). In grids created from pre-aligned FASTA files, a text alignment of the entire sequence of the reference and query molecule appears on right-clicking non-blue cells.

The specific residue represented in the right-clicked cell is demarcated by an asterisk that can be located by scrolling through the alignment.

Edit>Save edited data to file . . . stores all changed nucleotides and green tracking squares in *.edi files. Re-opening these files in MethylViewer via Edit>Open edit data file . . . allows return to data editing at one's convenience.

Viewing Bisulfite Conversion Efficiency and Location of Non-Converted Cytidines

In BGS, most residues that sequence as C do so because they are methylated and unconverted by bisulfite ion. Less frequently, C residues can also arise from failure to deaminate during the bisulfite conversion procedure. It is also possible that reversion of U or T to C occurs during PCR amplification of bisulfite-converted cytosines or during cloning. View>Show dC conversion map displays all C residues to scale, i.e. appropriately separated according to base pair coordinates. C nucleotides not converted by bisulfite deamination (or mutations during PCR or cloning), excluding those within methylated DNMT target sites, are depicted as vertical blue tick marks (Figure 3-6A). Placing the cursor over any residue in the dC conversion map indicates its base pair coordinate relative to base pair 1 of the reference sequence in a pop-up window. Residues scored in DNMT sites that sequenced as T are depicted as gray vertical ticks, whereas those that sequenced as C are shown as black ticks. The level of m^5CG in the four HCT116 molecules is at background as only 1 of 168 scored CG sites sequenced as C, whereas there were two non-converted C nucleotides outside of DNMT sites (compare Figure 3-6A to Figure 3-5A and B). Clicking on any residue in this view displays both the ABI data trace and alignment between the reference and query sequence encompassing the DNMT site (Figure 3-6B). As an alternative, View>Always show text alignment can be selected to produce a full-length text alignment of the

reference and query sequences in a scrollable window, with an asterisk marking the clicked on residue (not shown). This is the only option for inspecting alignments of sites in grids derived from FASTA files. As above, the G indicated in red is queried, because it is complementary to the C in the reverse strand that was sequenced.

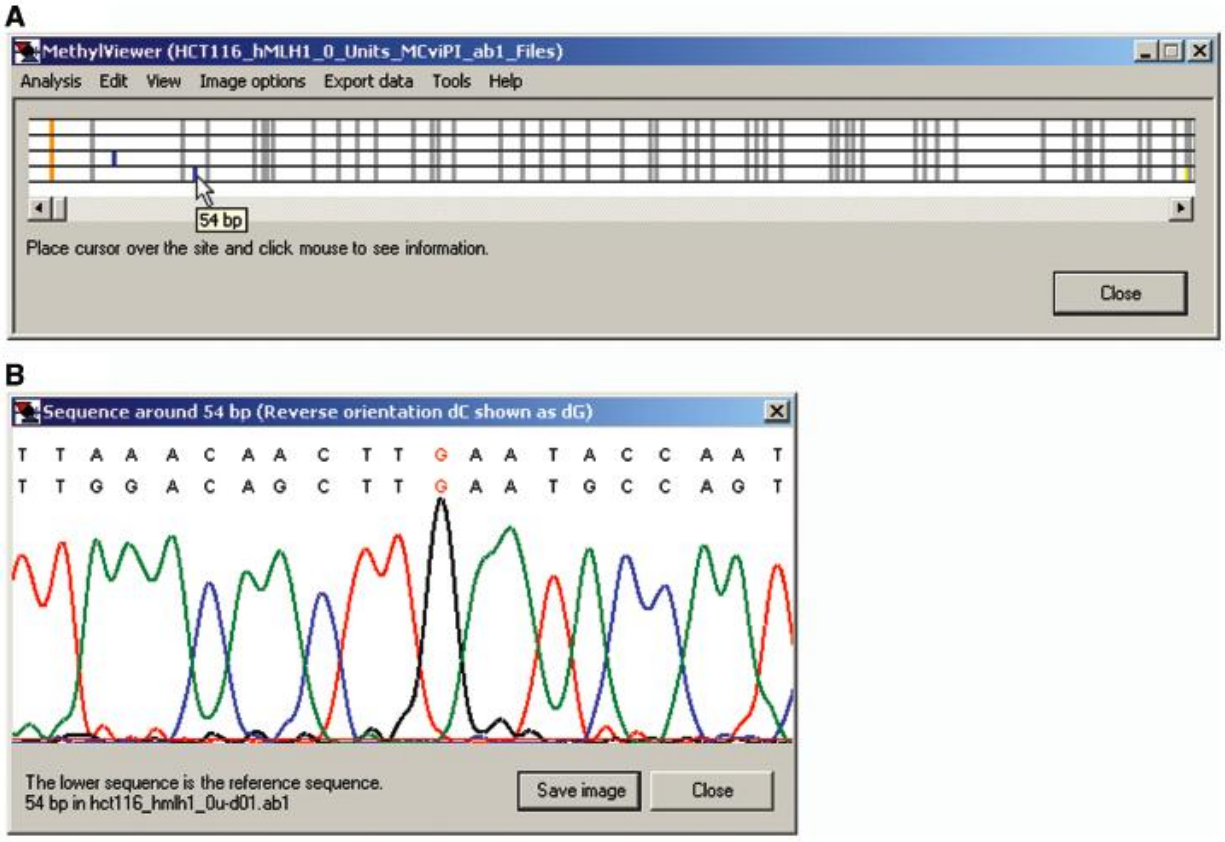

Figure 3-6. Bisulfite conversion status of dC sites not in queried methylation sites. A) Visualization by View>Show dC conversion map. Placing the cursor over any nucleotide position in the map displays its exact position in base pairs relative to base pair 1 in the reference in a pop-up window. Residues that sequenced as C (G on reverse strand) and lie outside of queried methylation sites are marked by a vertical blue line. Black and gray vertical lines indicate residues within target methylation sites that are methylated and unmethylated, respectively. B) Clicking on any nucleotide as shown in A displays the local sequencing trace and alignment to which it is linked.

Single-Molecule View of Endogenous m^5CG and Chromatin Accessibility by Mapit

Cursory visual inspection of the four edited sample data sets reveals that the *hMLH1* fragment contains a m^5CG level at the background of bisulfite-conversion in

HCT116 cells and dense m^5CG in RKO cells (Figure 3-7, compare black cells in A to C and B to D). In contrast to endogenous DNMTs, M.CviPI clearly accessed and methylated more sites at the *hMLH1* promoter in HCT116 than in RKO colorectal cancer cells (Figure 3-7, compare red cells in B to D). These results are consistent with previous findings of epigenetic silencing by hypermethylation of *hMLH1* in the latter cell line (Esteller, 2007a; Esteller et al., 2001; Feinberg et al., 2006; Herman and Baylin, 2003; Veigl et al., 1998). A summary breakdown between unmethylated and methylated residues (unedited) for all molecules in a grid is obtained by clicking on the blue cell at the head of row 1 (Figure 3-7A and C). This function sums the total number of sites scored and the absolute number of unmethylated and methylated sites, including overlapping sites. In the sample data, each cloned and sequenced 755-bp molecule of the *hHMLH1* promoter encompasses 84 DNMT sites, including 30 non-overlapping CG, 42 non-overlapping GC and 12 overlapping GCG sites. Thus, each sample grid with four analyzed molecules contains a total of 336 cells or sites.

Further details about individual molecules are obtained by clicking on blue cells that head each row in the grid, excluding row 1 (Figure 3-7B). This displays the name of the specific sequence file that was scored to obtain that row of data, the total number of sites for which methylation status was assigned, and breakdown of the total number and percentage of methylated non-overlapping sites for each queried DNMT site motif. The total number of residues expected to be converted from dC to dT (not present in selected methylation sites) as well as the number of unconverted dC residues is also tallied. This function is useful as the summary blue header in row 1 sums the total number of unmethylated (or methylated) residues in all queried DNMT sites, even of

different sequence. The overall breakdown in methylation of each DNMT site can be summed from each of the individual values. However, this is cumbersome when many molecules have been analyzed.

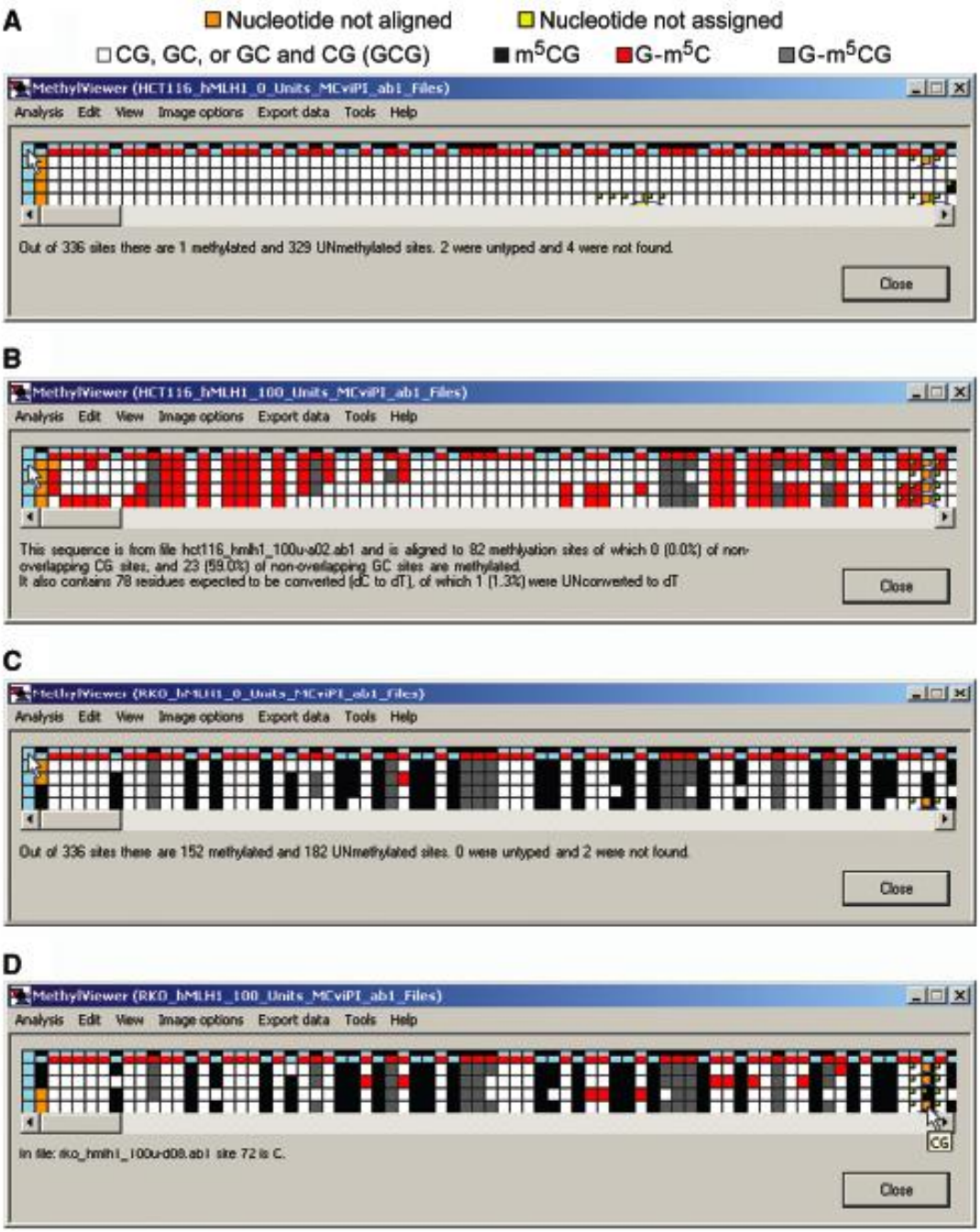

Figure 3-7. Edited interactive grids for mammalian MAPit data. Nuclei from cell line HCT116+0U M.CviPI A), HCT116+100U M.CviPI B), RKO+0U M.CviPI C) and RKO+100U M.CviPI D). The same *.ab1 files used in Figures 3-4–6 were used. Summary information for unedited data for all molecules in a grid is

displayed by clicking the light blue cell at the upper left as shown in A and C. Note that the numbers need to be adjusted to accommodate cells with manually-changed sequencing calls. Left-clicking on the light blue header cell for each row, except row 1, shows information associated with that specific molecule as shown in B.

Therefore, in practice, it is simpler to obtain data summaries by querying for methylation of one DNMT sequence motif at a time. Selecting CG in the Methylation sites window for samples not treated with M.CviPI reveals only 1 of 162 accurately scored CG sites sequenced as C at the *hMLH1* promoter in HCT116 cells (Figure 3-7A). This excludes four sites not aligned with the reference plus the two sites that we judged could not be aligned unequivocally. In contrast to this background level of m^5CG (<1%), 150 of 165 of typed CG sites in RKO cells were methylated (91%) (Figure 3-7C).

To separately determine the frequency of non-overlapping methylation at CG and GC sites in samples treated with M.CviPI, custom NCG and GCN sites, respectively, are scored. This is done by choosing Analysis>Interactive view . . .>Custom or Analysis>Import FASTA alignment>Custom, which activates the floating window for changing default settings (Figure 3-8). Default sites (CG and GC) are first removed by clicking Delete twice. Next, NCG (or GCN) is input under Enter degenerate or non-degenerate methylation site sequence. This automatically displays the four possible NCG (or GCN) sequences under Select any sites that you do NOT want to screen for, and the box for GCG is checked. A toggle window is used to Select the methylated dC residue on the forward strand that is being queried for methylation in each DNMT site. Users also Select a color to identify this site from a toggle window and Select symbol at the bottom of the panel to depict each type of site in an exported image (below). Clicking Add inputs the site designation and then OK is clicked to load the custom settings and close the window.

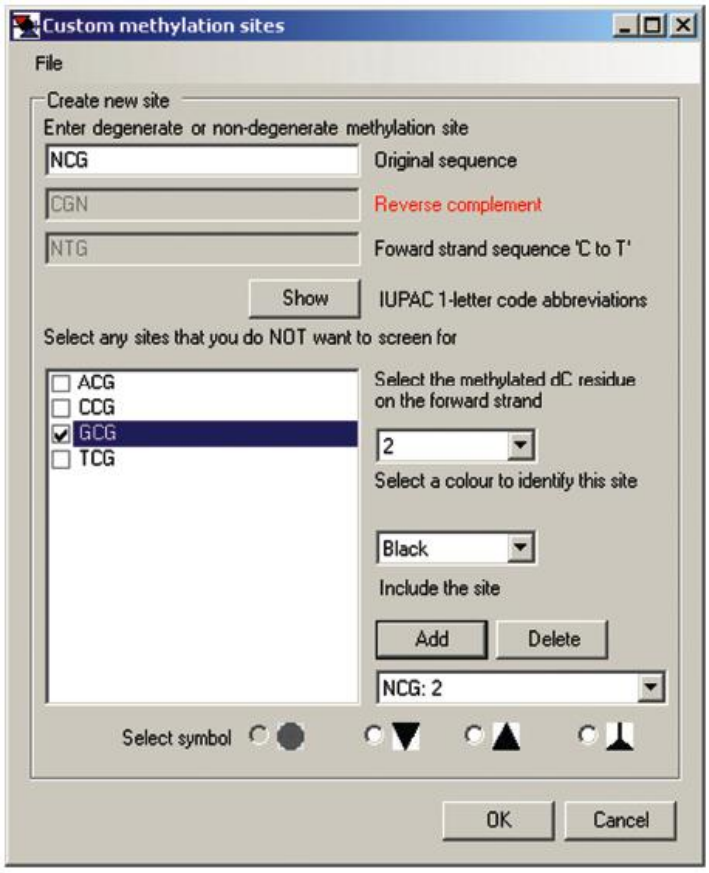

Figure 3-8. MethylViewer maps any input methylation specificity. The shown window was activated by Analysis>Interactive view . . .>Methylation sites>Custom to determine frequencies of non-overlapping methylation in the MAPit data of Figure 3-7. The default CG: 1 and GC: 2 sites were deleted; the settings in each field were input as shown, and site NCG: 2 was entered by clicking Add. Checking the box for GCG eliminates the overlapping site from the analysis. As many as four sites of any sequence containing one or more deoxycytidines, including non-palindromic sites, can be designated.

Using these settings, MethylViewer generated a grid reporting methylation of 110 of 115 (96%) scored non-overlapping CG sites at the *hHMLH1* promoter in RKO cells (grid not shown). This high level of m^5CG is similar to the level in the minus M.CviPI sample from RKO cells (96-91%). Moreover, the high level of m5C corresponded to methylation of only 13 of 168 (7.7%) of scored non-overlapping GC sites (GCN with GCG omitted) by M.CviPI, indicating that chromatin is highly inaccessible in epigenetically-silenced copies of *hHMLH1*. In contrast, in HCT116 cells where *hHMLH1*

was essentially unmethylated at endogenous CG sites and actively transcribed, the summary cell in the grid for GCN with GCG omitted indicate a high degree of accessibility to M.CviPI, with 92 of 167 (55%) scored non-overlapping GC sites being methylated.

Data Saving Options and Publication-Quality Images

After making and editing an alignment, users can choose from several options to customize information included in exported files via the Image options menu (Figure 3-9A). These options include: edited data; image resolution; and bisulfite dC conversion frequencies [(expected number of C residues to convert observed after conversion) / expected 100], which are indicated at the right of each diagrammed molecule in exported images. In addition, selecting Include labels in image files appends the name of the original sequence data file (or sequence name following the first>symbol in FASTA files; Figure 3-2) at the left of each diagrammed molecule and labels DNMT site numbers. Including labels activates the further option Include base position with labels, which places these labels below the DNMT site numbers, if selected. A final option is to Select range of site shown in images (Figure 3-9B), which allows users to export a subset of the data within a grid to an image. For example, sites not aligned between the reference and query sequences, i.e. orange cells, can be omitted from the exported image. The exported image can retain its original sites numbers or the user can elect to Renumber sites.

Once the desired Image options have been selected, MethylViewer can save data in several formats suitable for publication or further manipulation, using the Export data menu (Figure 3-9C). Grid images, a text file and scaled images with the user-selected symbols and colors can be exported as bitmap (*.bmp), portable network graphics

(*.png) or scalable vector graphics (*.svg) files. Export data>Save as text file provides a detailed, tab-delimited output that summarizes the sequence assigned to each scored DNMT site to facilitate further data analysis. If sequencing files were used to generate the grid, the alignment can also be exported as a text FASTA file, with or without the reference sequence.

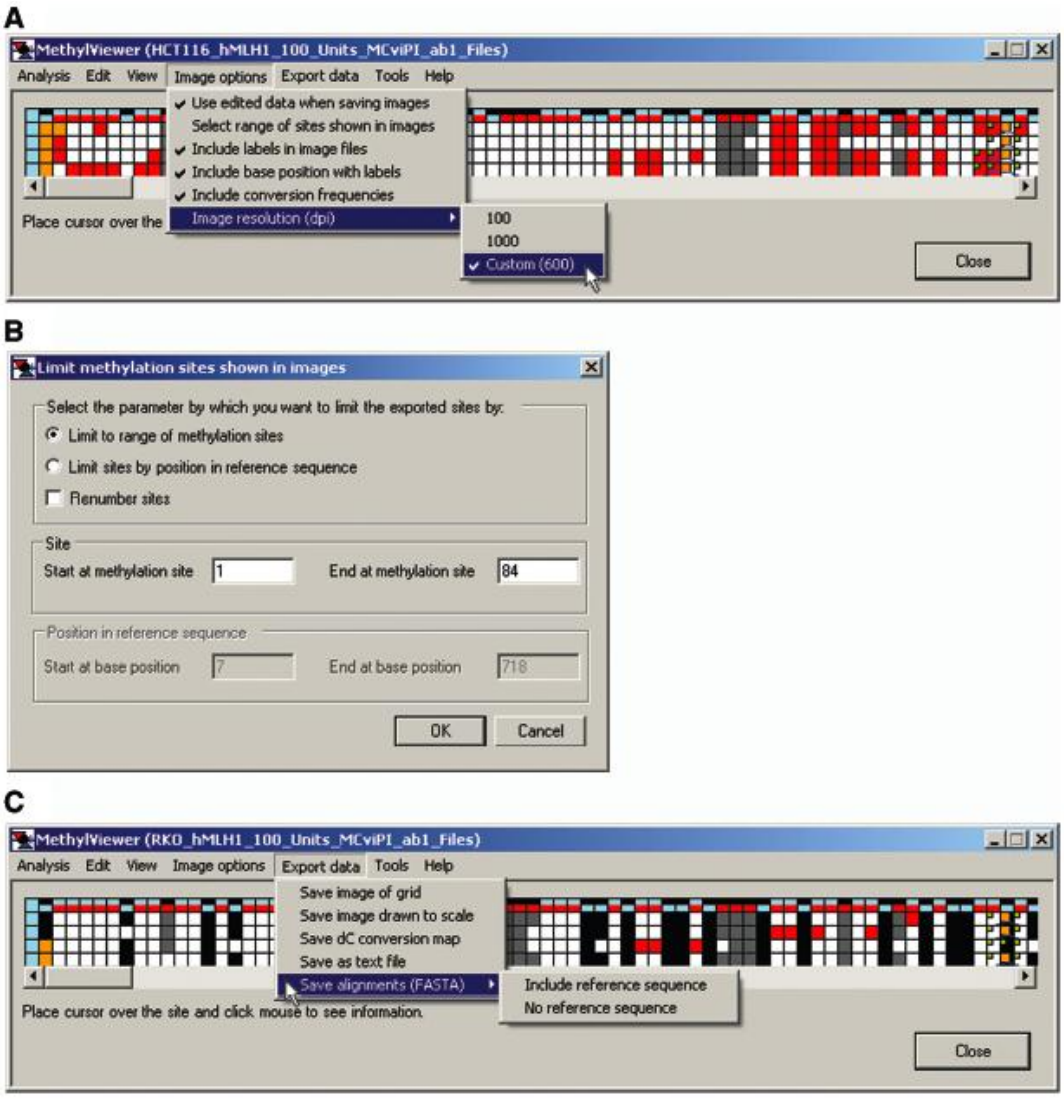

Figure 3-9. Image options and data export. A) Various features for generating and labeling exported images are selected from the shown drop-down menu. In this image, the default image resolution has been reset from 100 to 600 dots per inch (dpi). B) Window launched by Image options>Select range of sites shown in images. Shown are the default settings, which can be changed, for example, to remove sites not aligned at the beginning or end of aligned sequences. C) Export data options are in the shown drop-down menu.

In MAPit analysis, sites in DNA are methylated and hence accessible to DNMT probe if they are unoccupied by non-histone proteins or are not incorporated into a nucleosome (Jessen et al., 2004; Jessen et al., 2006; Kladde et al., 1996). Spans of accessibility to M.CviPI were manually added according to a 2:2 definition in which two consecutively methylated GC sites denote accessible regions (Fatemi et al., 2005) (Figure 3-10). By this definition, continuity of these 'open' regions is only broken by ≥ 2 unmethylated GC sites. We imposed an additional caveat whereby spans of accessibility were broken if two consecutive G-m^{5}C sites were separated by ≥ 30 bp. Based on experience, this is a reasonable footprint size for a non-histone protein that could be bound, but is thus undetectable due to low GC site frequency. G-m^{5}CG is ignored in denoting these accessible regions because it cannot be determined if cytosines in these sequences were methylated by endogenous DNMTs or M.CviPI. Nevertheless, given the near absence of m^{5}CG in HCT116 cells (Figure 3-10A), it is highly likely that G-m^{5}CG in Figure 3-10B was catalyzed by exogenously-added M.CviPI in addition to stand alone G-m^{5}C (not in GCG sites). This same inference does not apply to RKO cell chromatin probed with M.CviPI in Figure 3-10D. It can be inferred, however, that most G-m^{5}CG is attributable to endogenous DNMTs due to high-density, stand alone m^{5}CG in the control sample not probed with M.CviPI (Figure 3-10C). In HCT116 cells, two spans of consecutive sites were accessible to M.CviPI in each molecule (Figure 3-10B). Seven of these eight regions were ~150 bp in length and thus may correspond to nucleosome-free regions mapped by accessibility to M.SssI in transcriptionally-active LD419 cells (Lin et al., 2007). Each pair of M.CviPI-accessible regions flanks a ~150 bp protected region, which is inferred to correspond to a single

nucleosome of different translational position. Regions upstream and downstream of hyperaccessible regions associated with TSSa and TSSb of the *hMLH1* promoter in HCT116 cells, respectively, are likely to be protected by nucleosomes; however, their exact positions cannot be inferred in the absence of a defined linker region. Nonetheless, the majority of promoter sequences in these cells was inaccessible to M.CviPI and thus likely occupied by nucleosomes whose precise positions cannot be assigned.

Alignment and User-Defined Analysis of Cytosine Methylation of Any Specific Site in Bisulfite-Converted Sequences

MethylViewer supports analysis of cytosine modification of any user-defined sequence. Analysis of C methylation in sequences containing degenerate bases is also supported by inputting IUPAC one-letter definitions, which are provided in a pop-up window upon selecting Custom>Custom methylation sites>Show. As many as four custom user-defined sites can be entered by selecting Add. The program is versatile enough to analyze C methylation in any sequence context by entering ‘C’ in Create new site. Four different symbols and six different colors are available for diagramming methylation of each bisulfite-converted sequence. As in the default CG and GC setting, if more than one site is selected; overlapping methylation is depicted as gray symbols. Custom site definitions can be saved by Custom methylation sites>File>Save as and reloaded by Methylation sites>File>Open.

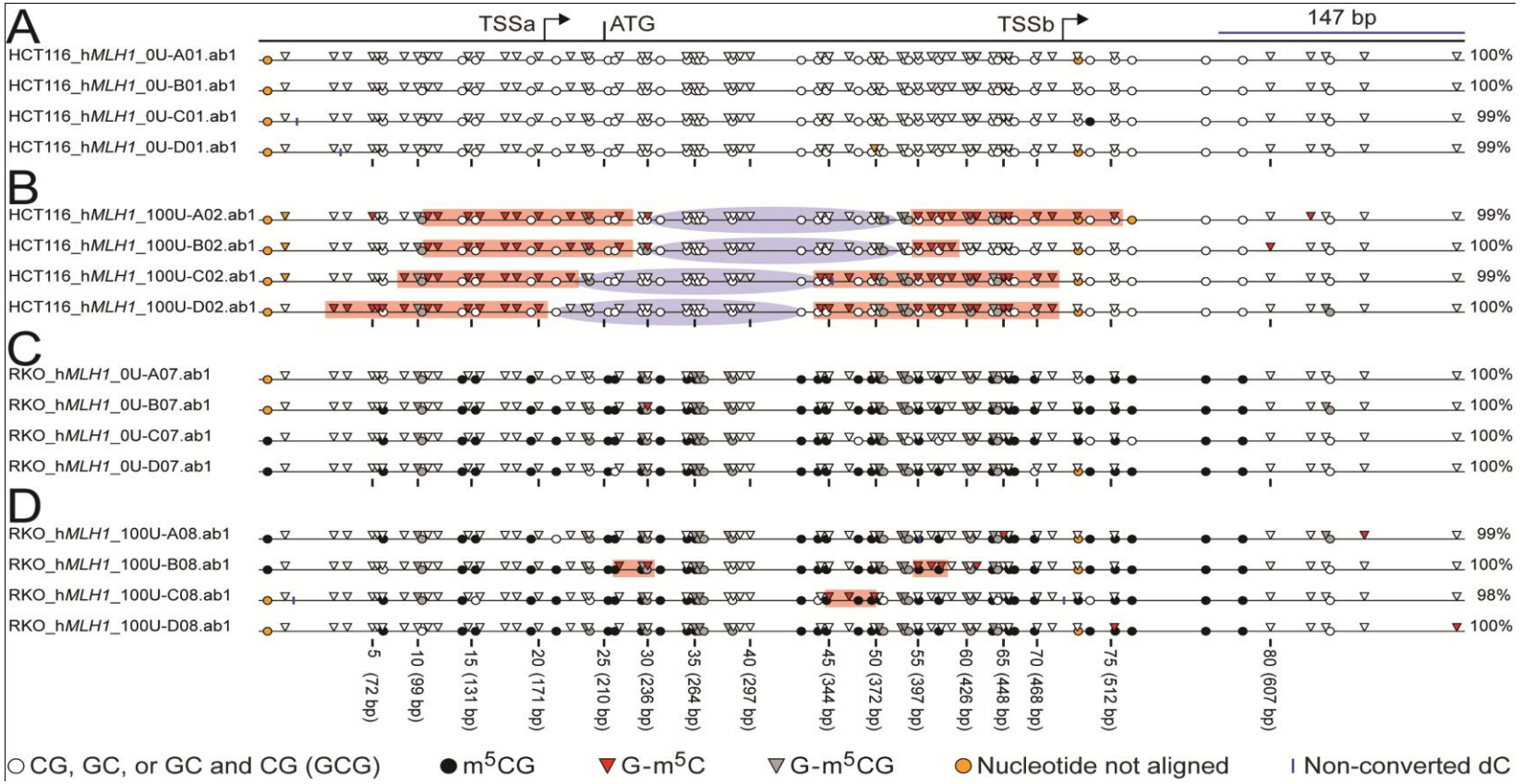

Figure 3-10. Publication-quality, scaled images of *hMLH1* MAPit data. Edited data from each of the grids in Figure 3-7 were exported with Image options as shown in Figure 3-9, via Export data>Save image drawn to scale, and saved using the *.svg option. Nuclei from cell line HCT116+0U M.CviPI A), HCT116+100 U M.CviPI B), RKO+0 U M.CviPI C), and RKO+100 U M.CviPI D). The image is identical to that produced by MethylViewer, except that the amplicon map at the top and legend at the bottom were added in Adobe Illustrator and labels at left were edited slightly (increased font size as well as capitalization and italicization as appropriate and to taste). Also inserted were red rectangles and blue ovals, depicting spans of GC sites methylated by exogenously added M.CviPI chromatin probe (according to 2:2 convention,(text)) and inferred positions of nucleosomes, respectively. Bent arrows, TSSa and TSSb; translational initiation codon, ATG. The blue bar is to scale and indicates the 147 bp length of DNA within a nucleosome core particle.

To demonstrate MethylViewer's capability to analyze C modification at degenerate, non-palindromic sites, we integrated a single estrogen-inducible copy of the gene encoding M.CviPII into the budding yeast genome. This second DNMT from *Chlorella* virus NYs-1 was recently cloned and shown to methylate the first C in CCD sites, where D equals A, G, or T (Chan et al., 2004b). The enzyme was reported to also methylate the first two cytosines in CCAA and CCCG sites. Genomic DNA was isolated from yeast cells incubated with 100nM 17 β-estradiol inducer or ethanol vehicle in rich medium. Following bisulfite conversion, a 500-bp region of the PHO5 promoter, a locus with a well-characterized chromatin structure consisting of five upstream positioned nucleosomes (N–1 to N–5), was PCR amplified. Three and 17 independent clones from uninduced and estradiol-induced cells, respectively, were sequenced and analyzed by MethylViewer with custom sites of CCD, CCAA and CCCG designated (Figure 3-11). As expected, the final scaled *.svg image of site accessibility to M.CviPII in the PHO5 promoter amplicon shows no cytosines and hence no methylated sites among the three clones from cells where the M.CviPII transgene was uninduced (Figure 3-11A and B, top panels). In contrast, 91 of 285 (32%) scored CCD sites were methylated in induced cells (Figures 3-11A and B, bottom panels). Among 85 aligned CCAA sites, 25 were methylated, of which all were modified at the residue overlapping with the CCD specificity; i.e. the first, but not the second, C residue. No CCCG sites are present in the analyzed PHO5 amplicon. Summing the methylation percentage at each site over the 17 molecules methylated by M.CviPII shows that sequences in linkers between and at the edges of nucleosomes exhibited the highest frequencies of methylation (Figure 3-11C), as we have previously shown (Gottschling, 1992; Singh and Klar, 1992), with two

exceptions. The first exception was a high level of m^5CCD at site 16, which we have observed on occasion, and is possibly due to maximum curvature in nucleosomal DNA that occurs near this location (Luger et al., 1997b). The second exception was relatively high accessibility to M.CviPII of various spans of sites (2:2 convention broken by separation of m5CCD sites by ~30 bp), which are occupied by N–2 when transcription of the PHO5 is repressed. Several accessible patches of varying length were also observed at the edge of N–3 and its adjacent linker DNA in a subset of molecules. We conclude that accessibility of chromatin can be probed by M.CviPII and rapidly analyzed and visualized by MethylViewer.

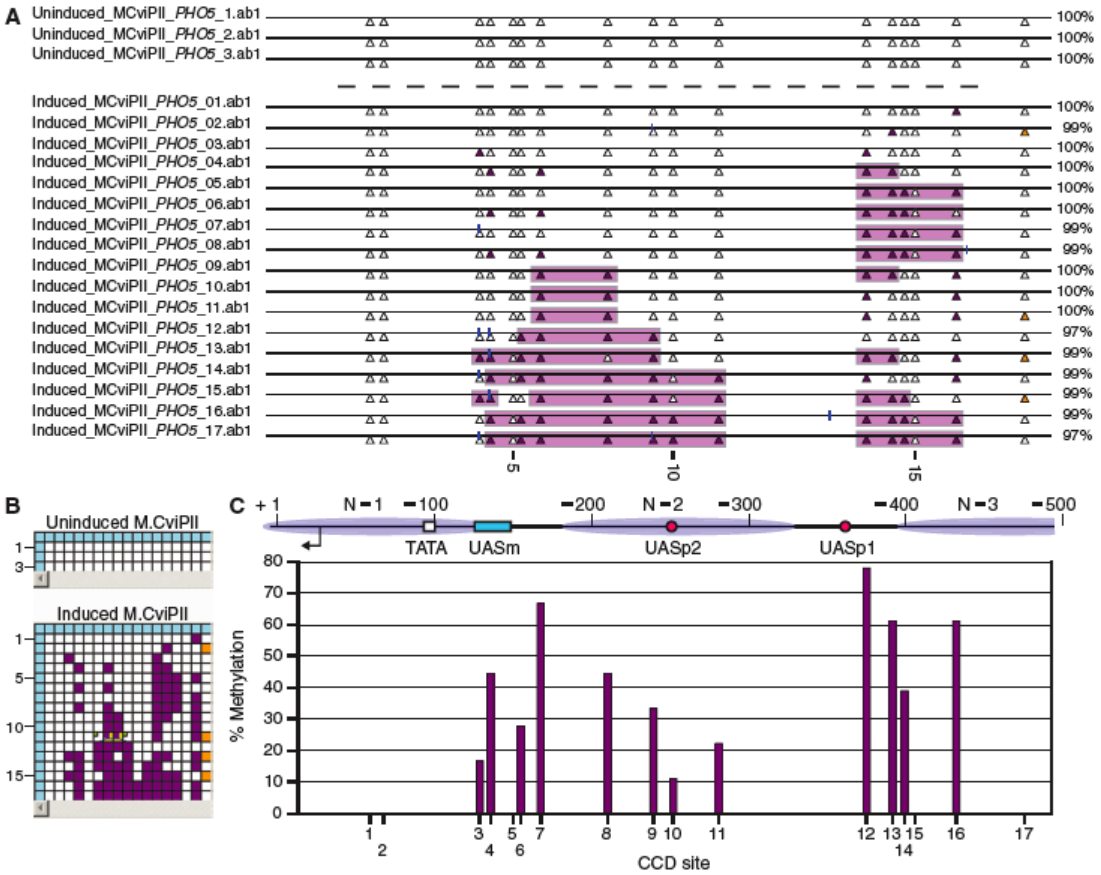

Figure 3-11. MethylViewer analyses MAPit data using degenerate DNMT probe. A) Publication-quality, scaled *.svg image of bisulfite-converted PHO5 sequences obtained from cells with uninduced (top panel) or estrogen-induced (bottom panel) expression of chromatin probe M.CviPII. Colors are

as in other figures, except that CCD and m5CCD sites are indicated by filled white and filled purple triangles, respectively. B) Edited grids used to export images in A. The settings for Base calling cutoff and Word size, were 200 and 15, respectively. C) Overall methylation frequencies at each CCD site are indicated below the map showing placement of: positioned nucleosomes (ellipses labeled N-1, N-2 and N-3; (Almer and Horz, 1986); upstream activating sequences at which Pho4 binds [red-filled circles labeled UASp1 and UASp2 as mapped by Vogel and Hinnen (Vogel et al., 1989)]; the compound Mcm1-Fkh site [cyan-filled rectangle labeled UASm as described in Pondugula et al. (Pondugula et al., 2009)]; and TATA box (white-filled square); major TSS (bent arrow). According to the convention used in budding yeast, base pair coordinates are indicated relative to the first nucleotide of the ATG translational initiation start codon.

Discussion

MethylViewer is a versatile, user-friendly and intuitive graphical user interface program for processing BGS data. It can either directly process and align raw sequencing files or accept text files with pre-aligned sequences in FASTA format. An interactive grid of cells is returned that simplifies data editing and visually marks cells that have been edited and inspected for quality of sequencing calls, which are used to infer site methylation status. Viewed and edited cells in the methylation grid can be saved to files that can be opened at a future time and/or exported to images. With the exception of minor further editing to taste, the program also automatically generates publication-quality images of either the grid or standard ‘lollipop’ images that maintain widely accepted conventions for representing DNA methylation patterns. Users can also choose to include additional information in images, including the name of the sequence file to assist with data tracking, base pair coordinates of each queried methylation site, and dC bisulfite conversion efficiencies of cytosines not in queried methylation sites. The tab-delimited data in exported text files also make possible further external mathematical manipulations, such as correlation matrices and other statistical analyses.

MethylViewer provides a substantial advance over currently available BGS analysis programs, because it permits rapid and accurate interrogation of DNA methylation status of as many as four user-defined sites at a time in any biological system. The capability to specify analysis of multiple methylation sites is ideal for MAPit methylation footprinting that may employ more than one DNMT probe and/or detection of endogenous m5C (Figures 3-7 and 3-10). A crucial feature of MethylViewer for MAPit absent from other programs is the generation of scaled images with symbols spaced according to their relative position in the sequence. This allows more accurate placement of footprints. The graphical user interface of MethylViewer also constitutes a stand-alone, integrated solution for MAPit projects in that Tools>Bisulfite primer design . . . aids users in designing primers that meet the further challenge of avoiding probed GC sites in addition to endogenous CG sites. Analysis and primer design are also therefore not limited to sites commonly methylated in vertebrates (CG) and plants (CGN, CHG and CHH or CNG) as are other programs (Bock et al., 2005; Carr et al., 2007; Grunau et al., 2000; Hetzl et al., 2007; Ordway et al., 2005; Rohde et al., 2008; Singal and Grimes, 2001; Xu et al., 2007).

Furthermore, as MethylViewer is the only available program that can perform custom analysis of any and all cytosines in a sequence, computational analysis of BGS data is not limited to sites modified by DNMTs with known specificity. Edit>Create consensus . . . can be used determine the recognition site and methylated C for newly discovered or uncharacterized DNMTs, or the consensus site of a footprinted region in MAPit studies. In addition, the capability to query methylation status of C within any nearest-neighbor sequence context makes it possible for MethylViewer to support a

myriad of additional BGS applications. For example, the program can be used to analyze BGS datasets from undifferentiated human embryonic stem cells that are reported to containabundant non-CG methylation (m^5CHG and m^5CHH) (Grandjean et al., 2007; Hawkins et al., 2010; Kouidou et al., 2005; Latham et al., 2008; Laurent et al., 2010a; Lister et al., 2009). MethylViewer can also analyze BGS data from organisms, such as *Neurospora crassa* and *Ascobolus immersus*, which, respectively, pre-meiotically methylate DNA repeats at C-5 without strict site specificity (Selker, 1990). Lastly, MethylViewer can be used to analyze methylation by BGS that involves replicative incorporation of modified nucleotides, e.g. bacteriophage T4 or potential PCR applications. We have shown that M.CviPI is as effective as M.SssI for probing chromatin structure of active loci at single-molecule resolution [Figures 3-7B and 3-10B; (Fatemi et al., 2005; Gal-Yam et al., 2006; Lin et al., 2007)]. In particular, similar accessibility of mostly nucleosome-length regions (~147 bp) to both DNMT probes was found near *hMLH1* TSSa and TSSb in the colorectal cell lines HCT116 and LD419. However, unlike M.SssI, probing accessibility of GC sites with M.CviPI (Xu et al., 1998a) enables simultaneous determination of both chromatin accessibility and endogenous m5CG in mammalian cells (Figures 3-7 and 3-10). Availability of two other DNMTs with short recognition sites in addition to M.SssI, M.CviPI (GC) and M.CviPII (CCD, CCAA and CCCG), significantly increases the resolution for mapping chromatin structure. As seen in Figure 3-11, 17-fold coverage of the PHO5 promoter region probed with M.CviPII was able to detect overall preferential accessibility in linker regions between both N–1 and N–2 (sites 4, 6 and 7) and N–2 and N–3 (sites 12–14). Similar patterns of overall accessibility of the PHO5 promoter and other loci to M.CviPII, M.CviPI and

M.SssI have been observed. This demonstrates that single-molecule footprinting is reproducible with different DNMTs in diverse systems (cell types, loci and in vitro-assembled chromatin), and in different laboratories (Hoose and Kladde, 2006; Jessen et al., 2004; Kilgore et al., 2007; Kladde et al., 1999; Pardo et al., 2009). The single-molecule view afforded by MAPit allows one to obtain an accurate picture of how methylation is partitioned amongst the molecules. A good illustration of this occurred in the region occupied by PHO5 N–2; sequenced promoter molecules clearly clustered into one subpopulation inaccessible to M.CviPII and another with stretches of accessible sequence. Different lengths of accessible sequence may represent partial or complete disassembly of N–2 nucleosomes or differential degrees of sliding of the N–2 octamer on individual molecules (Dechassa et al., 2010). Sites at the edges of positioned nucleosomes may also be accessed by DNMTs due to 'breathing' or site exposure at the entry-exit sites (Li and Widom, 2004; Polach and Widom, 1995, 1996). The two distinct subpopulations may also be indicative promoters that have remodeled or are in the process of remodeling PHO5 promoter chromatin. The experiment in Figure 3-11 was performed on a population of yeast cells grown asynchronously in rich medium, which contains cells that have activated PHO5 transcription in M phase and have repressed transcription in G1 to early S phase (Pondugula et al., 2009). The MAPit results are thus consistent with subpopulations of yeast cells containing transcriptionally-active or -inactive copies of PHO5. Verification of this hypothesis and distinguishing between different possibilities of chromatin remodeling require more extensive MAPit analysis as well as additional studies. In closing, MethylViewer should greatly facilitate efforts of genome-wide analysis of m^5C via BGS. We will publicly host

our server to the increasing number of researchers studying the role of DNA methylation in epigenetic regulation who we expect will find MethylViewer a valuable freely available resource.

CHAPTER 4
INTERPLAY BETWEEN DNA METHYLATION AND CHROMATIN STRUCTURE DURING EPIGENETIC RESILENCING OF TUMOR SUPPRESSOR GENES IN COLON CANCER CELLS

Introductory Remarks

Epigenetic regulation is crucial for the maintenance of gene expression patterns in normal cells. Abnormal global and gene specific epigenetic changes in DNA methylation and histone modification patterns are common occurrences in cancer. Gene-specific TSG silencing is one of the best-documented epigenetic aberrations affecting all cancer types described to date (Esteller and Herman, 2002; Feinberg, 2008; Jones and Baylin, 2007; Lima et al., 2010). Active TSG promoters are found in transcriptionally-permissive chromatin conformations characterized by DNA hypomethylation as well as the presence of active histone marks, including histone H3 and H4 acetylation and histone H3 lysine 4 di- and trimethylation (H3K4me2 and H3K4me3, respectively). Upon epigenetic silencing, TSG promoters shift to a repressive epigenetic organization characterized by increased DNA methylation, loss of active histone marks and accumulation of repressive marks, such as trimethylation of lysines 9 and 27 of histone H3 (H3K9me3 and H3K27me3, respectively). In concert with other events, DNA methylation and histone modifications are thought to collaborate to establish transcriptionally-repressive chromatin environments (Esteller, 2007a; Rodriguez-Paredes and Esteller, 2011). Despite much recent attention, how the interplay between DNA methylation and histone modifications (and their characteristic profiles) contributes to establishment and maintenance of permissive *versus* repressive transcriptional states remains ill defined. Even more obscure is our understanding of the relationship between

nucleosome positioning and chromatin organization, and how they contribute to epigenetic gene silencing of TSGs.

Epigenetic silencing of TSGs is now accepted as a primary event contributing to cancer development and progression, occurring with a frequency on par with the genetic component of the disease. Epigenetic silencing of TSGs has evoked substantial interest in the areas of cancer diagnosis, prognosis and treatment. In contrast to genetic mutations, epigenetic phenomena have the distinguishing potential for reversibility without needing gene therapy, making them promising targets for cancer treatment (Feinberg, 2008; Harris and McCormick, 2010; Rodriguez-Paredes and Esteller, 2011).

The DNA demethylating agents 5-aza-2'-deoxycytidine (5-aza-dC; Decitabine) and 5-aza-cytidine (5-aza-C; Vidaza) are at the front line of epigenetic altering drugs for therapeutic treatment of cancer. These two compounds have become standard of care for a hematological disorder of myeloid cells called myelodysplastic syndrome (MDS). In comparison to other prevalently-used cytotoxic chemotherapeutic agents, patients treated with DNA demethylating agents have shown higher responsiveness to therapy and increased survival rates (Oki et al., 2007; Shen et al., 2010). DNA demethylating agents have also shown promise as a treatment modality in leukemia and clinical trials have begun on solid tumors, where its implementation has proven more challenging (Issa and Kantarjian, 2009; Oki et al., 2008; Stewart et al., 2009). Upon 5-aza-dC treatment, TSG demethylation and derepression are observed in cells from both patients and cultured cancer lines. Frequently, after cessation of 5-aza-dC treatment , derepressed or reactivated TSGs return to a transcriptionally-silent state and reacquire

DNA methylation. This clearly hampers the usefulness of this class of anticancer drug (Issa and Kantarjian, 2009; Oki et al., 2007).

Enhancing the clinical success of 5-aza-dC is likely to rely on increasing our knowledge base for how the compound effects gene reactivation. Furthermore, relatively little is known about the molecular events that accompany epigenetic gene resilencing after 5-aza-dC withdrawal, and if they are general or gene-specific. The best described mechanism of 5-aza-dC and 5-aza action is based on their inhibitory activity of DNMTs that modify C-5, mainly in the context of CpG (hereafter, CG) dinucleotides (Jones and Baylin, 2002). Both nucleoside analogs can be substituted for cytosine in DNA during replication, whereas only 5-aza-C can also be incorporated into RNA. DNA methyltransferase DNMT1 becomes covalently linked to or “sequestered” by DNA during electrophylic attack of the cytosine ring when either analog is post-replicatively incorporated into CG (Wu and Santi, 1987)., These effects are thus cytotoxic at high doses of 5-aza-C. At lower doses, DNMT1 sequestration allows replication to occur in the absence of DNA methylation. Consequently, initial methylation patterns are erased and not passed on to progeny cells. The resulting 5-aza-dC-mediated demethylation of hypermethylated TSG promoters is thought to restore critical tumor suppressor function. The demethylated status of CG dinucleotides is maintained through subsequent divisions in the absence of *de novo* methylation (Goll and Bestor, 2005; Hermann et al., 2004a; Hermann et al., 2004b; Ushijima et al., 2003; Zhu et al., 2001). In addition, recent data suggest that 5-aza-dC also directly targets DNMT1 for degradation by the 26S proteasome (Datta et al., 2009; Ghoshal et al., 2005). Therefore, the precise mechanism of action of 5-aza-dC remains controversial (Issa and Kantarjian, 2009).

The physiological effects of 5-aza-C incorporation into RNA remain unexplored. For this reason, we have limited remaining discussion and our experimental investigations to the DNA-specific nucleoside analog,5-aza-dC.

DNMTs are also known to act in a coordinated manner with other layers of epigenetic regulation through their association with chromatin modifying enzymes, such as those possessing histone methyltransferase (HMT) and histone deacetylation (HDAC) activity (Harris and McCormick, 2010; Lima et al., 2010; Sharma et al., 2010; Taby and Issa, 2010). Histone modifications are altered at TSG promoters after treatment with 5-aza-dC indicating that the interplay between distinct epigenetic layers is more complicated than previously appreciated (Cameron et al., 1999; Chiurazzi et al., 1999). Understanding the mechanisms by which 5-aza-dC achieves transcriptional reactivation through DNA demethylation and chromatin reorganization, and the extent to which the derepressed state is maintained (or lost), are crucial to the improvement of cancer treatment by epigenetic altering drugs.

Herein, we have studied how DNA methylation and nucleosome positioning are altered during reactivation of densely methylated promoters upon 5-aza-dC treatment, and the extent to which gene expression is resilencing after drug withdrawal. We found that three different TSGs relevant to colon cancer (*CDH1*, *MLH1* and *TIMP3*) exhibited varying folds of derepression and DNA demethylation in the presence of a low dose of 5-aza-dC. Despite these differences, all three genes became resilenced within 7-11 d of 5-aza-dC removal, whereas overall reductions in DNA methylation at each of them persisted for at least 21 d. Addition of the deacetylase inhibitor Trichostatin A (TSA) did not lead to derepression, suggesting that DNA demethylation was necessary for TSG

reactivation. Single-molecule MAPit methylation footprinting showed that a low dose of 5-aza-dC and subsequent removal caused regiospecific changes in CG methylation. Furthermore, nucleosome occupancy changed dramatically during both promoter derepression as well as subsequent resilencing and corresponded well with levels of TSG transcript. These results support a model in which DNA methylation must be removed for TSG reactivation but is not necessary for subsequent TSG resilencing, which instead correlates with changes in chromatin accessibility.

Materials and Methods

Cell Culture and 5-aza-dC Treatments

RKO and HCT116 colorectal cancer cells were obtained from the American Type Tissue Culture Collection. Cells were cultured in minimal Eagle's medium (MEM, Cellgro) supplemented with 10% (v/v) fetal bovine serum and penicillin and streptomycin at 1% (w/v) each, as recommended by the suppliers. Experiments were carried out with cells that had undergone less than 20 passages. For the high dose treatment with 5-aza-dC, 6×10^6 RKO cells were plated in 10 cm culture dishes. After 24 h, the cells were treated with 10 μM (Sigma) for 3 d with daily replacement of old growth medium with new containing fresh nucleoside analog. For the low dose treatment, 5×10^5 RKO cells were plated in 10 cm dishes, and after 24 h they were treated with 500 nM 5-aza-dC for 3 d with daily exchange of old medium for new containing fresh drug. Twenty-four hours after addition of the last drug dose, cells were washed with phosphate-buffered saline (PBS) and fresh media without 5-aza-dC was added, designated Day 0 of the resilencing time course (cf. Figure 4-5A). Cells were given fresh drug-free medium every 3 d for the duration of the resilencing time course.

Cells were harvested with trypsin-EDTA solution and separate aliquots used for isolation of total RNA, DNA and nuclei on the indicated days in the absence of 5-aza-dC.

To test if RKO cells surviving the first 5-aza-dC treatment regimen remained sensitive to 5-aza-dC, on day 22 of the resilencing time course, cells were collected and reseeded at 5×10^5 cells per 10 cm dish. After 24 h they were treated with 500 nM 5-aza-dC for 3 d with daily exchange of old medium for new containing fresh drug. Twenty-four hours after addition of the last drug dose, cells were washed with phosphate-buffered saline (PBS) and fresh media without 5-aza-dC was added. Cells were collected 24 h later and separate aliquots were used for isolation of total RNA and DNA.

For treatment with the of the histone deacetylase (HDAC) inhibitor Trichostatin A (TSA) RKO cells were seeded at 5×10^6 cells per 10 cm dish. After 24 h they were treated with 100 nM TSA for 24 h. Treated cells were collected and separate aliquots were used for isolation of total RNA and DNA

Derivation of Clonal RKO Cell Lines

RKO cells were incubated with 500 nM 5-aza-dC (Sigma) as described above. On day 18 of the resilencing time course, isolated single colonies were removed from the culture dishes and expanded. An aliquot of cells was frozen on day 35 after drug withdrawal. The remaining cells were harvested and separate aliquots processed to purify RNA for expression analysis and DNA for methylation analysis by pyrosequencing.

RNA Isolation and Quantitative-Reverse Transcription PCR (qRT-PCR)

Total RNA was isolated from harvested cells using TRIzol® reagent following the manufacturer's recommendations. RNA was resuspended in diethyl pyrocarbonate-treated RNase-free water and incubated with recombinant RNase-free DNase I (Roche). Specific cDNA copies were obtained by reverse transcription with Superscript II (Invitrogen) using gene-specific reverse primers (Table 4-1). Each RT reaction was carried out in triplicate with 60 ng of total RNA template in a 96-well plate in a 5 μl reaction volume. Thermocycling conditions were 50ºC for 30 min followed by 72ºC for 5 min. Gene-specific transcript abundance was determined by quantitative real-time PCR using SYBR Green (Applied Biosystems) in a StepOne™ Real-Time System (Applied Biosystems). Thermocycling conditions included an initial reverse transcriptase activation step at 95ºC for 10 min followed by 40 cycles of 95ºC for 15 sec and 60ºC for 1 min. A melting curve cycle was performed to confirm the presence of a single PCR product. Melting curve cycling conditions were 95ºC for 15 sec, 60ºC for 1 min and 95ºC for 15 sec. No template and no reverse transcriptase controls were carried out in parallel to assess primer specificity and possible contamination of RNA with DNA. Gene-specific primer pairs were optimized for each gene (Table 4-1). All assays were normalized to β-ACTIN gene transcript. Relative expression levels were determined using the $\Delta\Delta C_T$ method relative to the control sample or by standard curve quantification.

Table 4-1. qRT-PCR primer pair sequences

Gene	Primer sequence
EPM2AIP1	CPO1809 (+) 5' ATG CCG GAA CCG GGC TTG TG
	CPO1811 (−) 5'- GCC GCA AGG GGA GAG GAG GA
MLH1a	CPO1809 (+) 5'- CAG CTA ATG CTA TCA AAG AGA TGA TTG[1]
	CPO1811 (−) 5'- GTT GTA AGA GTA ACA TGA GCC ACA TG[1]
MLH1b	CPO1810 (+) 5'- GAG ACC TTT TAA GGG TTG TTT GG[1]
	CPO1811 (−) 5'- GTT GTA AGA GTA ACA TGA GCC ACA TG[1]
TIMP3	CPO1872 (+) 5'- ACG CTG GTC TAC ACC ATC AAG C[2]
	CPO1873 (−) 5'- CCG AAA TTG GAG AGC ATG TCG[2]
CDH1	CDH1 (+) 5'- CAG TGA ACA ACG ATG GCA TT[3]
	CDH1 (−) 5'- CTG GGC AGT GTA GGA TGT GA[3]
GAPDH	CPO1697 (+) 5'- CTG CAC CAC CAA CTG CTT AG[4]
	CPO1698 (−) 5'- AGG TCC ACC ACT GAC ACG TT[4]
β−ACTIN	NNO2098 (+) 5'- CCC TGG CAC CCA GCA C
	NNO2099 (−) 5'- GCC GAT CCA CAC GGA GTA C
18S rRNA	NNO2096 (+) 5'- CGG CTT AAT TTG ACT CAA CAC
	NNO2097 (–) 5'- ATC AAT CTG TCA ATC CTG TCC

(+) Forward primer, (−) reverse primer
Taken from: [1](Lin et al., 2007), [2](Datta et al., 2009), [3](Lin et al., 2010), [4](Ai et al., 2008).

DNA Methylation Analysis by Pyrosequencing

DNA was isolated as described in Chapter 2. Briefly, cells were harvested at various time points by scraping. After proteinase K treatment, total genomic DNA was isolated by extraction with phenol:chloroform:isoamyl alcohol (25:24:1) and concentrated by ethanol precipitation. DNA was resuspended in 1 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, (0.1 × TE). Sodium bisulfite conversion was carried out as described in Chapter 2. Bisulfite-converted DNA (1 μl) was amplified using HotStar Taq® DNA polymerase (Qiagen) and gene specific primers in a 20 μl reaction. Thermocycling conditions included a 95°C incubation for 5 min followed by 49 cycles of 94°C for 45 sec, primer specific annealing temperature for 45 sec and 72°C for 2 min, followed by a final extension at 72°C for 10 min. PCR amplification of a single species was confirmed by Tris-acetate-EDTA agarose gel electrophoresis in the presence of 0.5 μg/ml ethidium bromide. Amplified DNA (~5 μg) was purified with streptavidin-

coated sepharose beads and subjected to pyrosequencing using a PyroMark ID instrument per the manufacturer's instructions. Pyrosequencing assays were tested for amplification bias on a set of five standards containing mixtures (100:0, 75:25, 50:50, 25:75 and 0:100) of total DNA from RKO cells methylated in vitro to completion by M.SssI and placental DNA, which contained undetectable levels of methylation at the studied loci. Primers used for pyrosequencing are listed in Table 4-2.

Table 4-2. Pyrosequencing primer pair sequences

Gene	Primer sequences
EPM2AIP2	CPO2464 (+) 5'- GTA GAA GTT TTA TTA GGG T
	CPO2465 (–) 5'- TAA AAC CCT ATA CCT AAT CT
MLH1	CPO2143 (+) 5'- GGA GTA AGt Tat AGt TGA AGG AAG AA
	CPO2145 (–) 5'- Biotin/CCA ATT CTC AAT CAT CTC TTT AAT AAC
TIMP3	TIMP3F (+) 5'- GGG TtA GAG AtA ttt AGT GGt tt
	TIMP3R (–) 5'- TTA CCT CAT CAA CCC TCC
	TIMP3Seq: 5' GGt ttA GGT GGG
CDH1	CDH1F (+) 5'- GGA AtT GtA AAG tAt tTG TGA Gt
	CDH1R (–) 5'- Biotin/CTC CAA AAA CCC ATA ACT AAC C
	CDH1Seq: 5'-GTt AGT TtA Gat Ttt AGt t

(+) Forward primer, (–) reverse primer, Seq = sequencing primer

MAPit Chromatin Accessibility Analysis

MAPit is described in detail in Chapter 2 and in Pardo et al. (2011). Briefly, nuclei were isolated from 3.1×10^6 cells at 4°C under buffer conditions that preserve the integrity of nuclei and chromatin structure (Pardo, 2011). Harvested cells were washed twice with PBS and resuspended in Cell Resuspension Buffer (CRB: 20 mM HEPES, pH 7.5, 70 mM NaCl, 0.25 mM EDTA, 0.5 mM EGTA, 0.5% (v/v) glycerol, 10 mM DTT, 0.25 mM PMSF). After pelleting by centrifugation at 1,000 × *g*, cells were lysed in 108 μl of Cell Lysis Buffer (CLB: 1 CRB plus 0.19% Nonidet P-40) for 10 min on ice. Nuclei were examined by light microscopy to confirm integrity. Finally, 170 μl of M.CviPI Methylation Buffer (M.CviPI MB: CRB plus 160 μM *S*-adenosyl-methionine (SAM)) were

added to the nuclei. For chromatin probing, 90 μl of nuclei suspension containing 10^6 nuclei were aliquoted for each dose of M.CviPI used to probe chromatin. Individual 100 μl methylation reactions were carried out at three M.CviPI concentrations. An untreated sample (0 U M.CviPI) serves as a background control to assess the frequency of false positive GC methylation, i.e. due to non-conversion of C by bisulfite, mutation during PCR amplification or sequencing errors. Nuclei were also probed with 30 U and 100 U M.CviPI, in a final volume of 100 μl to obtain total enzyme concentrations of 0.3 U/μl and 1 U/μl, respectively. This 3.33-fold difference in concentration allows one to assess different degrees of chromatin accessibility. After pre-warming the nuclei at 37°C for 5 min, 0 U (enzyme diluent only), 30 U and 100 U doses of M.CviPI were added and nuclei methylation reactions were carried out for 15 min at 37°C. Reactions are terminated by the addition of an equal volume of 2× Methylation Stop Buffer (MSB: 100 mM NaCl, 10 mM EDTA, 1% SDS) followed by treatment with 100 μg/ml proteinase K at 50°C for 16 h. DNA was isolated as described for the pyrosequencing methylation analysis.

Bisulfite Genomic Sequencing

Bisulfite conversion of M.CviPI-modified DNA was carried out as described in Darst et al. (2010), routinely achieving bisulfite conversion efficiencies of >98%. Briefly, 2 μg of DNA were denatured in alkaline denaturation solution for 15 min at room temperature, followed by a second 5 min denaturation step at 95°C. Denatured DNA was then treated with 200 μl of saturated sodium metabisulfite solution at pH 5.0 in the dark at 50°C for 4-6 h. Bisulfite-converted DNA was desalted, desulfonated and purified using the EZ bisulfite DNA clean-up kit™ (Zymo Research Corporation). Three different

fragments were generated to analyze the *EPM2AIP1/MLH1* bidirectional promoter. Primer sequences used for MAPit analysis are listed in Table 4-3. To minimize potential for PCR bias, at least three 20 μl hot-start PCR reactions were performed with HotStar Taq® Plus DNA polymerase (Qiagen) and gene specific primers that anneal to deaminated DNA. Thermocycling parameters included a hot-start step at 95°C for 5 min followed by 39 cycles of 94°C for 1 min, primer-specific annealing temperature for 1 min and 72°C for 2 min. When using M.CviPI to probe chromatin structure, primers were designed to avoid not only potential endogenous methylation sites (CG) but also sites corresponding to the specific recognition sequence of M.CviPI (GC). When this was not feasible, degenerate bases were incorporated at positions of potential methylation to avoid PCR bias towards either methylated or unmethylated molecules (Chapter 2 for details). PCR products were gel purified by agarose gel electrophoresis and recovered from gel slices by the QIAEX II gel extraction kit (Qiagen). To obtain individual molecules for sequencing, single PCR products were cloned using either the TOPO TA Cloning Kit (Invitrogen) or pGEM-T easy vector system (Promega). Plasmid DNA was amplified directly from transformed *E. coli* TOP10 cells using TempliPhi rolling circle amplification (GE Healthcare). Cloned inserts in the resulting single-stranded DNA were sequenced using BigDye sequencing mix and analyzed on a conventional automated 3730 sequencer (Applied Biosystems) at the core facility in the Interdisciplinary Center for Biotechnology Research (ICBR) at the University of Florida.

Table 4-3. Primer pair sequences for MAPit analysis

Gene	Primer sequence
EPM2AIP2	CPO1840 (a1) 5'- CAa TCA AAT TTC TCA ACT CTa Taa a
	CPO1841 (a2) 5'- AtT GGT ATA tAA AGT ttt ttT tAt ttt AG
MLH1	CPO1842 (a1) 5'- Taa Ata TaA ACA aAT aAT TTC Taa AAT aAA Ta
	CPO2137 (a2) 5'- ttt AtA GAG TTG AGA AAT TTG AtT
EPM2AIP1/MLH1 in-between fragment	CPO1892 (a2) 5'- TTA AaT CRT ARC CCT TAA aTa
	CPO1893 (a1) 5'- TTt TTT TGG GYG TtA Ttt AtA

a2 deamination forward primer. Lower case t denotes C to T transitions
a1 deamination reverse primer. Lower case a denotes G to A transitions

Data Analysis

Sequenced molecules were analyzed with MethylViewer (Pardo et al., 2010) as described in detail in Chapter 3. Briefly, sequences were aligned with Sequencer 4.10.1 (Gene Codes). The alignment was exported in FASTA format and used as the input file for analysis with MethylViewer as described in Chapter 3.

Results

Epigenetically-Distinct Patterns of Chromatin Accessibility and DNA Methylation Revealed Simultaneously by MAPit Single-Molecule Footprinting

For our studies on the contribution of chromatin structure to the derepression and resilencing phenomena observed after 5-aza-dC treatment of cancer cells, we focused in the *EPM2AIP1* bidirectional promoter. *MLH1* is a member of the DNA mismatch repair system (MMR) that contains a promoter CG island (Bronner et al., 1994). *MLH1* is frequently epigenetically silenced by DNA hypermethylation in various cancer types, including colon cancer (Esteller et al., 1998; Veigl et al., 1998). The function of the *EPM2AIP1* gene product is not known. Figure 4-1A shows a schematic representation of the *EPM2AIP1/MLH1* bidirectional promoter.

We first analyzed transcript levels from the two genes in the *EPM2AIP1*/*MLH1* bidirectional promoter in the colon cancer cell lines HCT166 and RKO by qRT-PCR.

High levels of transcript from both genes were observed in HCT116 cells, whereas no transcript was detected from either gene in the RKO line (Figure 4-1B and C). Quantitative analysis of CG methylation by pyrosequencing showed that four CG sites near the *MLH1* TSSa were densely methylated in RKO but not in HCT116 cells (Figure 4-1D). Therefore, low and high amounts of DNA methylation at the bidirectional *EPM2AIP1*/*MLH1* promoter in HCT116 and RKO cells, respectively, correlate with abundant and non-detectable levels of transcript.

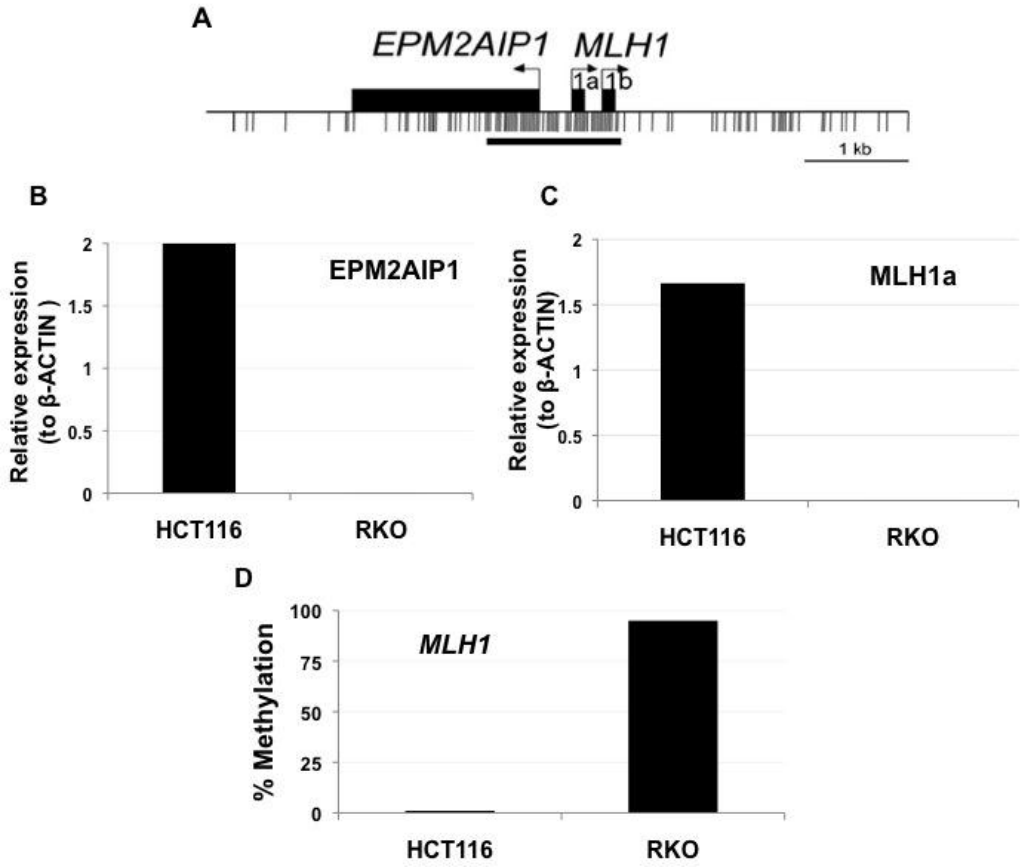

Figure 4-1. *EPM2AIP1/MLH1* bidirectional promoter is differentially expressed in HCT116 and RKO colon cancer cell lines. A) Schematic of the *EPM2AIP1*/*MLH1* bidirectional promoter to scale (Lin et al., 2007). Bent arrows, TSSs; thick black bar below promoter, CGI ; thick black bars above, coding regions; vertical hashes, CG sites; bar at bottom left shows a scale of 1 kb. *EPM2AIP1* B) and *MLH1* C) relative expression levels. Transcript levels in HCT116 and RKO cells were determined by quantitative comparison to a standard curve and normalized to those of β-ACTIN. D) Quantitative DNA methylation analysis of the *MLH1* promoter by pyrosequencing in HCT116

and RKO cell lines. Percent methylation was averaged over the four CG sites in the *MLH1* promoter indicated by a black line in Figure 4-4).

Next, we tested the dependence of silencing of *EPM2AIP1/MLH1* expression on DNA methylation by treating RKO cells with the global DNA demethylating agent 5-aza-dC (Figure 4-2A). Strong 5-aza-dC-dependent increases in the level of EPM2AIP1, MLH1a and MLH1b transcripts suggested that erasure of DNA methylation by the cytosine analog led to transcriptional derepression (Figure 4-2B). These results are in agreement with previous reports that *MLH1* is epigenetically silenced by DNA hypermethylation (Bronner et al., 1994; Esteller et al., 1998; Lin et al., 2007; Veigl et al., 1998).

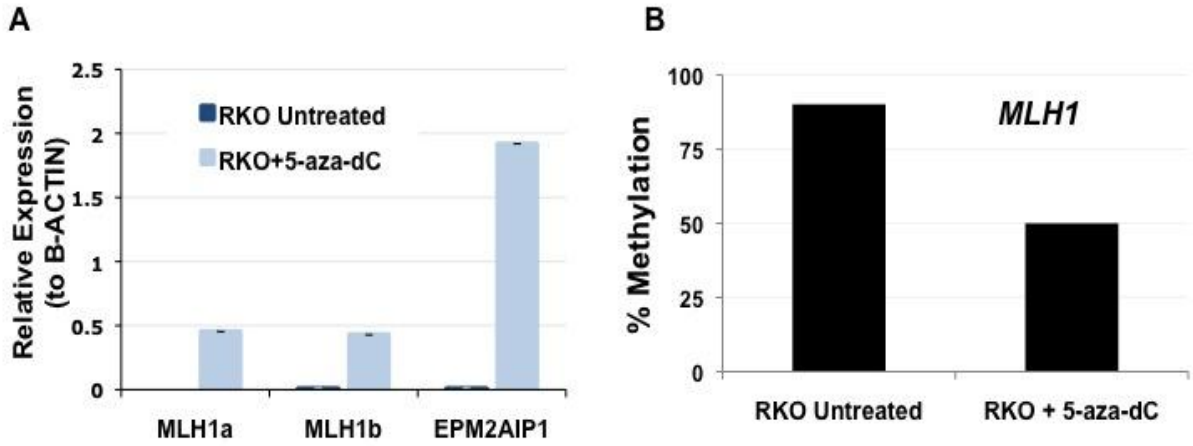

Figure 4-2. *EPM2AIP1* and *MLH1* transcriptional silencing is dependent on DNA methylation. A) Derepression of *EPM2AIP1* and *MLH1* expression by 5-aza-dC treatment of RKO cells. Transcript levels were determined by quantitative comparison to a standard curve and plotted as normalized to β-ACTIN levels. Error bars represent three technical replicates. B) Quantitative DNA methylation analysis of the *MLH1* promoter by pyrosequencing of DNA isolated from RKO cells ± 5-aza-dC. The average percent methylation of four CG sites in the *MLH1* promoter is shown.

Nucleosome positioning over promoters plays a crucial role in setting the transcriptional state of genes. To determine how CG methylation and chromatin accessibility correlate directly on individual copies of the *EPM2AIP1* and *MLH1*

promoters, we probed their chromatin structure in HCT116 and RKO cells by MAPit, as we described previously (Kilgore et al., 2007; Pardo, 2011). MAPit uses recombinant M.CviPI to probe chromatin accessibility when exogenously supplied to purified nuclei. This enzyme methylates accessible GC site cytosines to G-m^{5}C (Xu et al., 1998a) that can be read by bisulfite genomic sequencing (BGS) (Frommer et al., 1992). As nucleosomes and non-histone factors bound to DNA hinder DNMT access to DNA (Barrett and Spelsberg, 1998; Kilgore et al., 2007; Kladde and Simpson, 1996), the resultant GC methylation pattern is used to infer chromatin structure. M.CviPI and endogenous DNMTs in differentiated cells modify different sites, i.e. respectively GC and CG. Therefore, MAPit provides a single molecule readout of both methylation patterns in the same experiment. Moreover, the chromatin structure of loci dense in m^{5}CG can be mapped.

HCT116 and RKO cells were grown under the same conditions and were processed for MAPit analysis as described in the Materials and Methods and in more detail in Chapter 2. Due to constraints on product size with PCR from deaminated DNA, we divided the bidirectional *EPM2AIP1*/*MLH1* promoter region into two BGS amplicons. Respectively, the upstream (682 bp) and downstream (775 bp) amplicons correspond to the promoters for *EPM2AIP1* and *MLH1*, which encompasses both TSSa and TSSb (Lin et al., 2007; Veigl et al., 1998). We design a third amplicon (653 bp) bridging the *EPM2AIP1* and *MLH1* amplicons, which expands over the shared bidirectional promoter region and encompasses *EPM2AIP1* TSS and *MLH1* TSSa.

MAPit revealed striking differences in endogenous CG methylation pattern at the bidirectional promoter between HCT116 cells, in which both genes are transcribed, and

in RKO cells, where they are epigenetically silenced (Figure 4-3, *EPM2AIP1*; Figure 4-4, *MLH1*). Consistent with the pyrosequencing results in Figure 4-1D, both promoters had very low levels (< 1%) of endogenous DNA methylation (m^5CG; black filled circles) in HCT116 cells (Figures 4-3B and 4-4B, respectively). By contrast, dense endogenous DNA methylation (>98%) was present at the silenced *EPM2AIP1* and *MLH1* promoters in RKO cells (Figures 4-3C and 4-4C, respectively).

Examination of the chromatin accessibility aspect of the MAPit data also revealed a strikingly different organization between cells in which the *EMPM2AIP1/MLH1* bidirectional promoter is expressed (HCT116) *versus* silenced (RKO). To aid visualization of GC sites accessed by M.CviPI, a 2:2 convention was adopted whereby areas or patches of ≥2 consecutive G-m^5C sites were shaded red, and were broken by ≥2 consecutively unmethylated GC sites. Methylated GCG sites (G-m^5CG) were omitted from assignment of accessible areas. Thus, spans of protection against M.CviPI GC methylation were not shaded and correspond to footprinted DNA-bound proteins or nucleosomes (Pardo, 2011). Applying this 2:2 convention, the low-level CG methylation observed within the *EPM2AIP1* promoter amplicon from HCT116 cells was accompanied by a large accessible area of ~300 bp (red shaded regions) that encompassed the TSS in the majority of molecules (Figure 4-3B; +30 and 100 U M.CviPI). Upstream of this TSS and the hypersensitive region were footprints of various sizes. Downstream of the *EPM2AIP1* TSS and hypersensitive region in HCT116 cells, there were 1-2 nucleosome-sized (~147 bp) or larger footprints. Interestingly, these footprints occupied highly variable positions in each molecule, indicating a dynamic chromatin organization possibly due to nucleosome sliding.

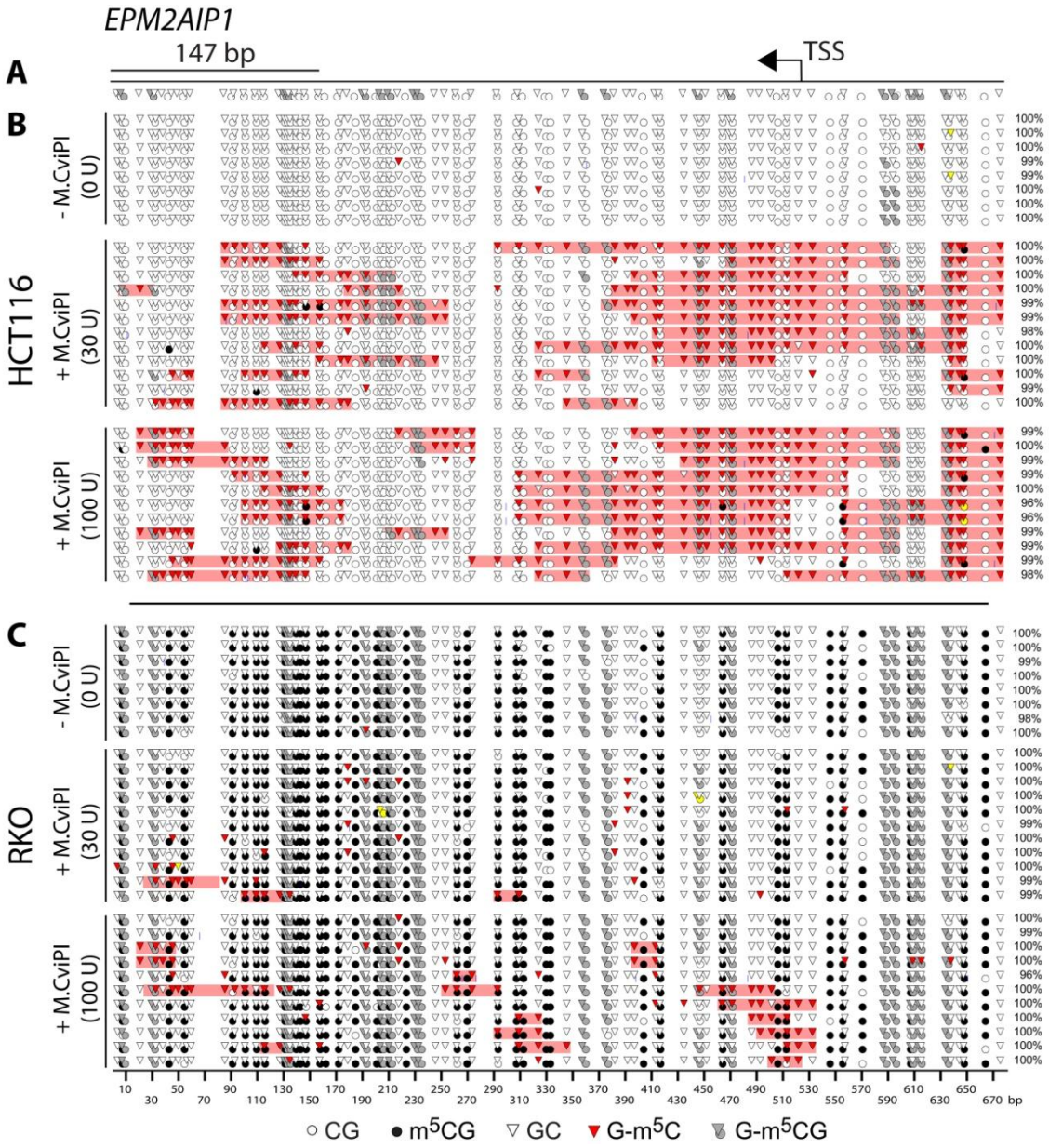

Figure 4-3. Starkly contrasting chromatin accessibility and CG methylation at the transcriptionally-active and silenced *EPM2AIP1* promoter. A) Map (to scale) of the 682 bp region analyzed by MAPit. Circles, CG sites; inverted triangles, GC sites probed by M.CviPI; Gray-filled symbols, overlapping GC and CG site, i.e. GCG. Bar labeled 147 bp at the top left is at scale and represents the size of DNA incorporated in a nucleosome core particle. MAPit analysis of the *EPM2AIP1* promoter transcribed in HCT116 B) and silenced in RKO C) cells. Nuclei were probed with the indicated amounts of M.CviPI. Each row of symbols represents DNA methylation status of each site along a single DNA molecule. Symbols are defined in the key at the bottom of C. Red shaded rectangles mark areas of accessibility to M.CviPI as defined in the text. Yellow filled symbols are sequenced residues that could not be unequivocally assigned as unmethylated or methylated. Bisulfite conversion efficiencies are

shown at the right of each molecule. Map coordinates (in bp) of the BGS amplicon are shown at the bottom of the molecules in C.

Similarly, the chromatin structure of the *MLH1* promoter in HCT116 cells showed two accessible areas that either encompassed or were located just upstream of TSSa and TSSb (Figure 4-4B; +30 and 100 U M.CviPI). In a significant fraction of molecules, clear footprints of 20 to 40 bp were present within the region of GC accessibility near TSSb, which were either absent or not as evident within the TSSa hypersensitive region. As observed at the *EPM2AIP1* TSS (Figure 4-3B), areas of protection against methylation by M.CviPI of nucleosome-size or larger mapped downstream of both accessible regions associated with *MLH1* TSSa and TSSb. In particular, a single nucleosome was apparently weakly positioned downstream of TSSa. That is, the nucleosome occupied variable positions within a few helical turns of each other on different molecules. The protection observed downstream of TSSb cannot be unequivocally assigned to a nucleosome, because downstream of the protection an accessible linker was not mapped. Nevertheless, the chromatin structures associated with *MLH1* TSSa and TSSb as well as *EPM2AIP1* TSS in HCT116 cells, namely, localization within or nearby nucleosome depleted regions (NDRs), were strongly indicative of actively transcribed genes in diverse organisms (Heintzman et al., 2007; Jiang and Pugh, 2009b; Lin et al., 2007; Ozsolak et al., 2007). In addition, clear footprints within the NDRs at the *EPM2AIP1* TSS and *MLH1* TSSb may be attributable to core promoter assembly of the RNA polymerase II preinitiation complex or other factors, as previously footprinted by exogenously added DNMTs (Gal-Yam et al., 2006; Kladde et al., 1996; Lin et al., 2007; Xu et al., 1998b).

Lin et al. (2007) observed many of these same chromatin structural features at the actively-transcribed *EPM2AIP1*/*MLH1* bidirectional promoter (Discussion). In that study, the CG-methylating DNMT M.SssI was used as the single molecule footprinting probe. Therefore, chromatin structure of densely CG methylated and thus epigenetically-silenced promoters could not be probed directly in RKO cells. Instead, these cells were treated with 5-aza-dC to reactivate the bidirectional promoter prior to probing nuclei with M.SssI. Single molecules were analyzed with and without selective amplification of unmethylated molecules by methylation-specific PCR (MSP; (Herman and Baylin, 2003).

To avoid these complications, we directly isolated nuclei from RKO cells that epigenetically-silence the *EPM2AIP1*/*MLH1* bidirectional promoter and probed their chromatin structure with M.CviPI. In stark contrast to the open chromatin structure of HCT116 cells (Figures 4-3B and 4-4B), the entire *EPM2AIP1*/*MLH1* bidirectional promoter in RKO cells was highly inaccessible to M.CviPI (Figures 4-3C and 4-4C; +30 and 100 U M.CviPI). The pattern of GC methylation in the fraction of molecules accessed by M.CviPI was reminiscent of short internucleosomal linkers, and suggested that nucleosomes were randomly organized in both regions in RKO cells. Nucleosome positioning cannot be assessed on the remaining molecules lacking GC accessibility. The striking absence of GC methylation from large expanses of chromatin in many molecules was consistent with hypermethylation of the CGI (Figures 4-3C and 4-4C; 0 M.CviPI); however, it was unexpected and suggested higher-order compaction of *EPM2AIP1*/*MLH1* chromatin in RKO cells. MAPit data for the fragment bridging the *EPM2AIP1* and *MLH1* amplicons is shown in Figure 4-5. Chromatin structures observed

around the TSS of *EPM2AIP1* and the TSSa of *MLH1* with this amplicon match the general chromatin structure observed with the individual amplicons for *EPM2AIP1* and *MLH1* (compare Figure 4-3 *EPM2AIP1*, Figure 4-4 *MLH1* and Figure 4-5 shared promoter).

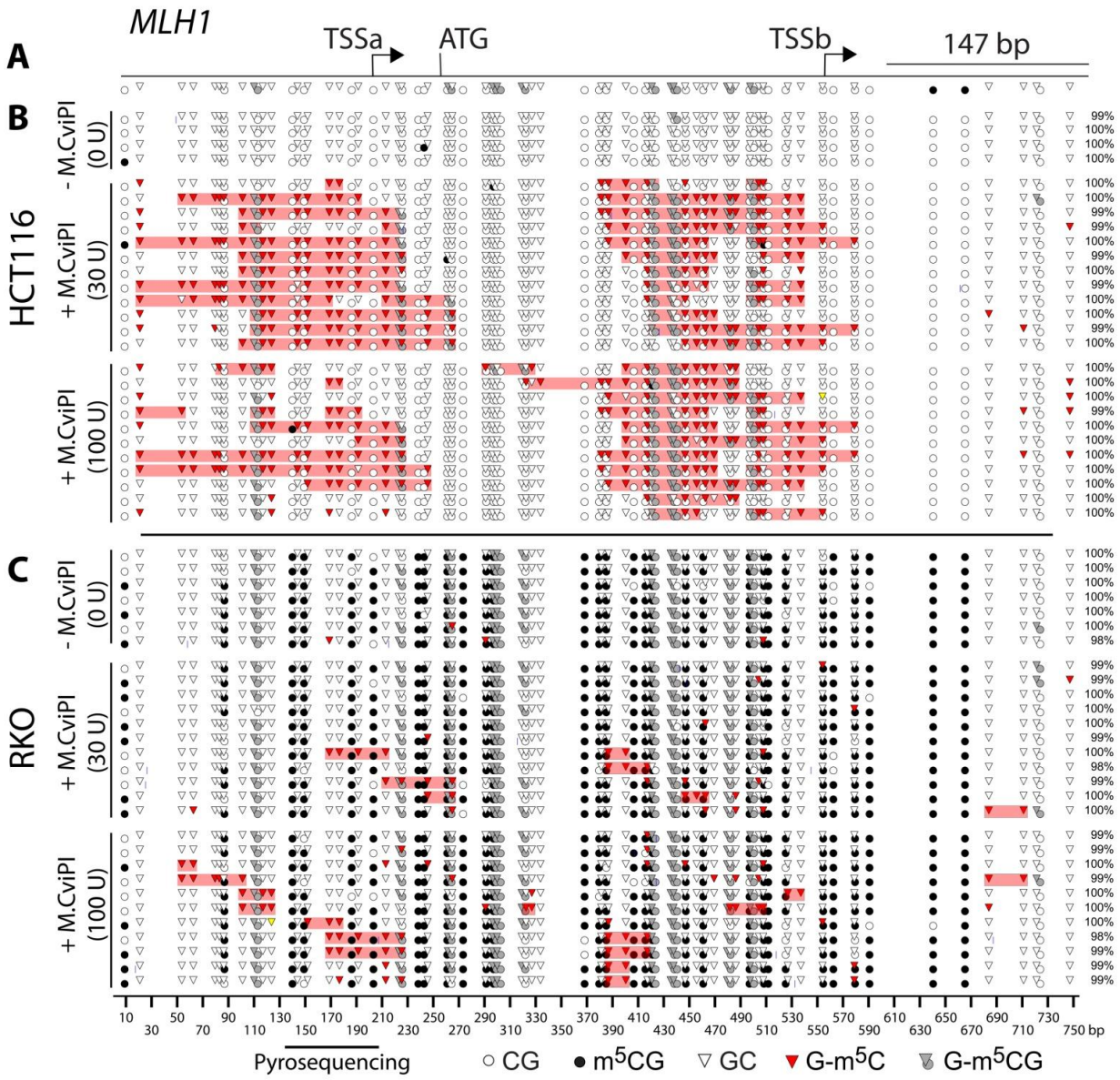

Figure 4-4. Starkly contrasting chromatin structures at the transcriptionally-active and silenced *MLH1* promoter. A) Map (to scale) of the 755 bp region analyzed by MAPit. MAPit analysis of the *MLH1* promoter expressed in HCT116 cells B) and silenced in RKO cells C). Symbols in all panels are as defined in Figure 4-3. Black bar at the bottom indicates four CG sites analyzed by

pyrosequencing. Map coordinates (in bp) of the BGS amplicon are shown at the bottom of the molecules in C.

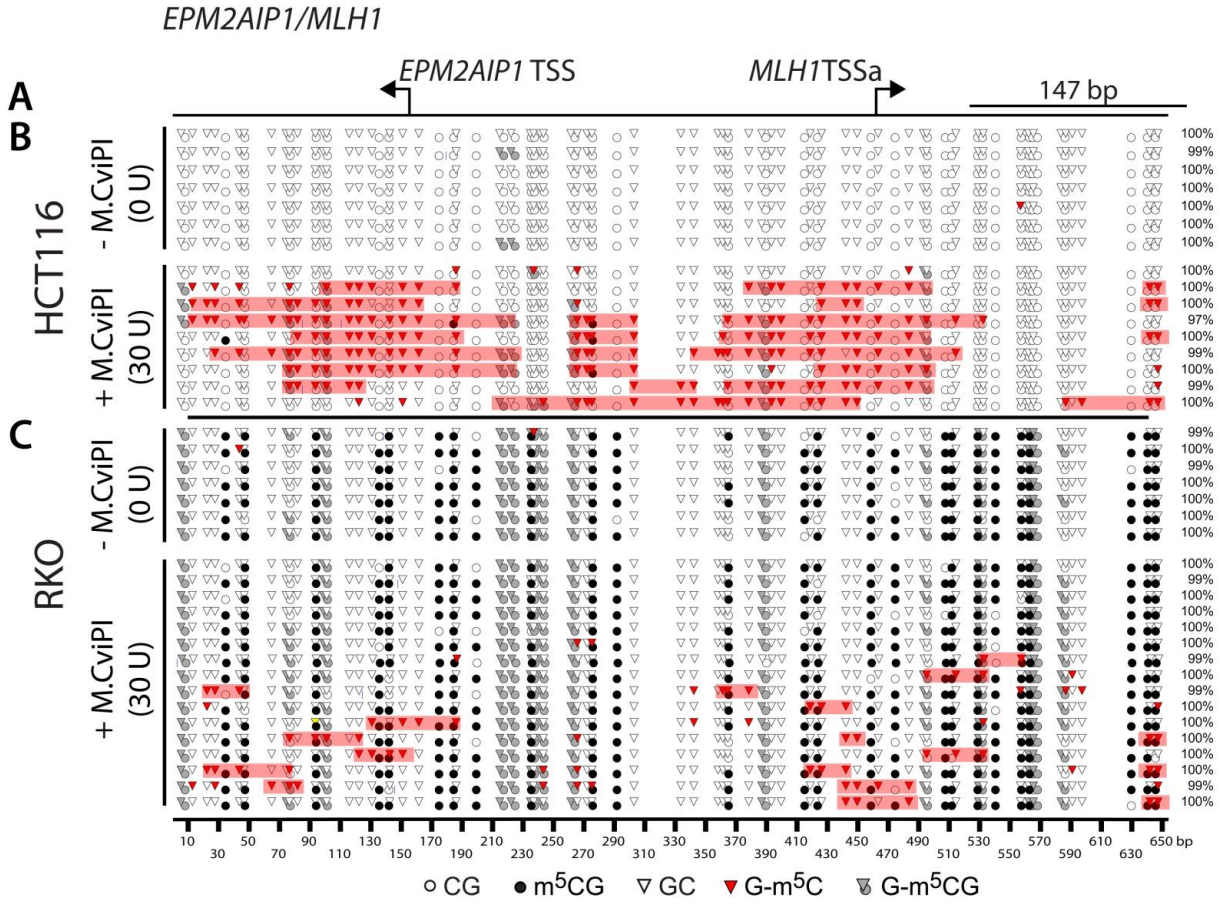

Figure 4-5. Starkly contrasting chromatin structures at the transcriptionally-active and silenced *EPM2AIP1/MLH1* bidirectional promoter. A) Map (to scale) of the 653 bp region analyzed by MAPit. MAPit analysis of the shared promoter region between the *EPM2AIP1* and *MLH1a* TSS. Genes are expressed in HCT116 cells B) and silenced in RKO cells C). Symbols in all panels are as defined in Figure 4-3. Map coordinates (in bp) of the BGS amplicon are shown at the bottom of the molecules in C.

DNA Methylation is Not a Driver of Resilencing of the *EPM2AIP1/MLH1* Bidirectional Promoter after 5-aza-dC Withdrawal

DNA demethylation and gene derepression upon administration of 5-aza-dC are well documented in both cell lines and patients. 5-aza-dC has successfully been used in the clinic to treat patients with MDS, resulting in significant demethylation and gene

reactivation. Reacquisition of DNA hypermethylation and resilencing of TSGs is also a common event observed after treatment cessation (Issa and Kantarjian, 2009).

To date the mechanisms of action of 5-aza-dC remain ill defined (Jones and Liang, 2009). Recent work has extended studies of 5-aza-dC beyond its effects on DNA methylation to other layers of epigenetic regulation, mainly histone modifications (Kagey et al., 2010; Wong et al., 2011). By contrast, few studies have addressed the effects of 5-aza-dC treatment on chromatin structure. To do so, we investigated the interplay between chromatin structure (CG methylation and chromatin accessibility) after transcriptional derepression by 5-aza-dC and a subsequent time course of resilencing TSG expression after drug withdrawal as schematized in Figure 4-6A. More specifically, RKO cells were initially incubated with 5-aza-dC according to a high or low dose regimen (Materials and Methods) and then changes in *EPM2AIP1/MLH1* gene expression (qRT-PCR) and chromatin structure (MAPit) monitored over time after drug removal. An important advantage of using MAPit in such an experiment is that 5-aza-dC-mediated demethylation rarely goes to completion; therefore, simultaneous mapping of chromatin accessibility and CG methylation to each molecule eliminates concern that changes may map to separate subpopulations of molecules.

We first treated RKO cells with 500 nM 5-aza-dC (low dose) for 3 d and then maintained the cells in drug-free media for 30 d. After the drug was removed, cell growth occurred in two phases; an initial phase of both slow growth and elevated cell death that lasted until day 7, followed by a second phase of increased growth rate and lower cell death (Figures 4-6B and C). As expected, *MLH1* expression was derepressed after treatment with 5-aza-dC (Figure 4-7A); however, by 22 d post-drug, the level of

MLH1a transcript approached the initial level of untreated RKO cells. Thus, a residual low-level *MLH1* expression was stable for at least 11 d (22 – 11) of continuous culture in the absence of 5-aza-dC.

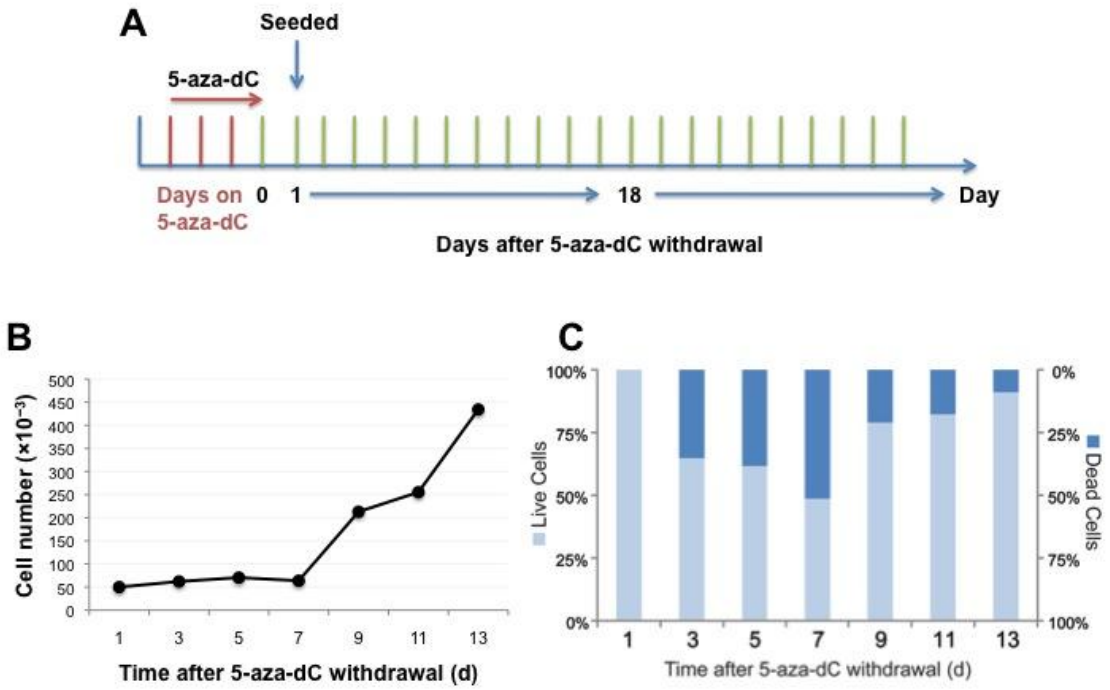

Figure 4-6. Low dose 5-aza-dC treatment temporarily reduces growth of RKO colorectal cancer cells. A) Schematic of the experimental design. RKO cells were treated with 500 nM 5-aza-dC for 3 d. Twenty four hours after drug withdrawal cells were seeded at 5×10^4 cells per well in 6-well plates and subsequently grown in drug-free media for the indicated times. B) RKO cell growth over time after removal of 5-aza-dC. C) Percentages of live and dead cells after 5-aza-dC treatment and subsequent resilencing time course as determined by trypan blue staining.

To determine if the changes in gene expression during the resilencing time course had an epigenetic basis, we assayed CG methylation at the *MLH1* promoter by pyrosequencing of genomic DNA isolated from a separate aliquot of cells from the same experiment. Accumulated levels of CG methylation in the pyrosequencing amplicon decreased from ~90% in the initial untreated RKO cells to ~40% after the first day of 5-aza-dC removal (Figure 4-7B). However, CG methylation did not significantly increase

for the duration of the resilencing time course. Instead, reacquisition of DNA methylation after 5-aza-dC removal was very slow, and only trended towards ~50% CG methylation in the cell population even after 30 d of drug-free growth (Figure 4-7B).

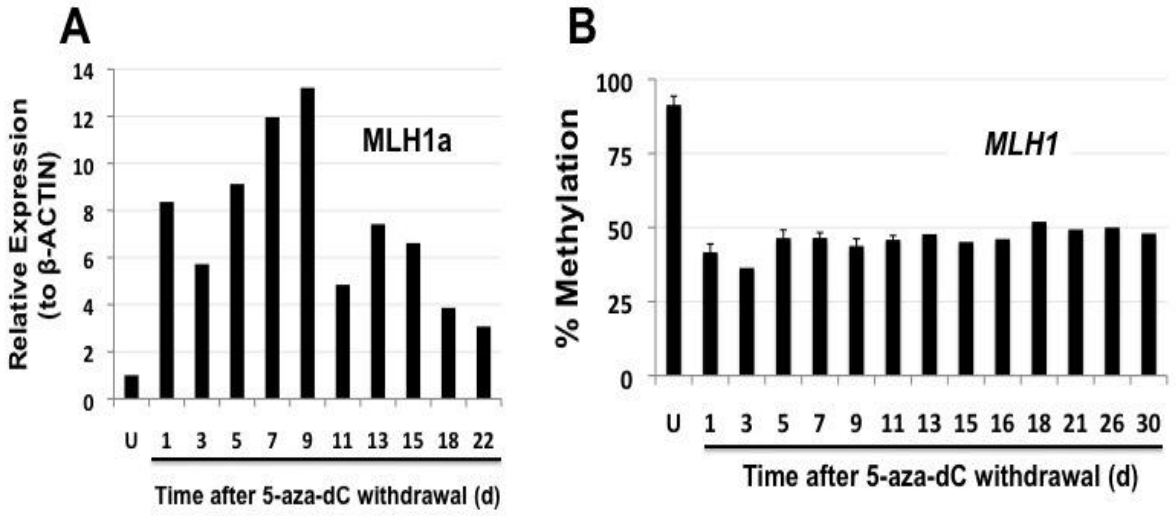

Figure 4-7. *MLH1* gene resilencing after 5-aza-dC withdrawal does not correspond well to changes in CG methylation. RKO cells were manipulated as in Figure 4-5. A) Changes in *MLH1* relative expression during the resilencing time course. MLH1a transcript levels were normalized to those for β-ACTIN. Values were determined by the $\Delta\Delta C_T$ method. U, untreated RKO cells. B) Changes in DNA methylation at the *MLH1* promoter in RKO cells ± 5-aza-dC. Percent methylation was measured quantitatively by pyrosequencing and the average over the four CG sites indicated in Figure 4-4C is shown. Error bar, N = 3 experimental replicates. U, untreated RKO cells.

In contrast to our results, previous studies have reported faster reacquisition of m^5CG after 5-aza-dC removal, and thus concluded that DNA methylation is a required for TSG resilencing (Kagey et al., 2010; Wong et al., 2011). We noted that those studies employed higher 5-aza-dC doses, longer treatment times or both. To resolve this discrepancy, we repeated the 5-aza-dC treatment using a high dose of 10 μM 5-aza-dC for 3 d. This protocol resulted in a maximal level of MLH1a transcript at day 5 after drug withdrawal, which then decreased by day 7 (Figure 4-8A). Thus, the time to onset of *MLH1* resilencing was earlier with the high-dose 5-aza-dC treatment protocol than with

the low dose protocol used in Figure 4-7 (peak MLH1a transcript at day 9 post-drug removal; resilencing by day 11).

As expected, treating RKO cells with a high dose of 5-aza-dC resulted in more efficient *MLH1* promoter demethylation, from 80% m^5CG in untreated cells to 28% at day 7 after drug withdrawal (Figure 4-8B). Remarkably, while little, if any, remethylation had occurred at the *MLH1* promoter after removal of the low 5-aza-dC concentration (Figure 4-7B), at the high dose, m^5CG reaccumulated to 70% by day 12 (Figure 4-8B). Therefore, for both doses of 5-aza-dC, resilencing of *MLH1* promoter expression, as gauged by levels of MLH1a transcript, preceded DNA remethylation by a significant number of days. We speculate that the more efficient remethylation of DNA in response to removal of the high 5-aza-dC dose might be due to selective survival and/or growth advantage of a DNA demethylation-resistant subpopulation of cells.

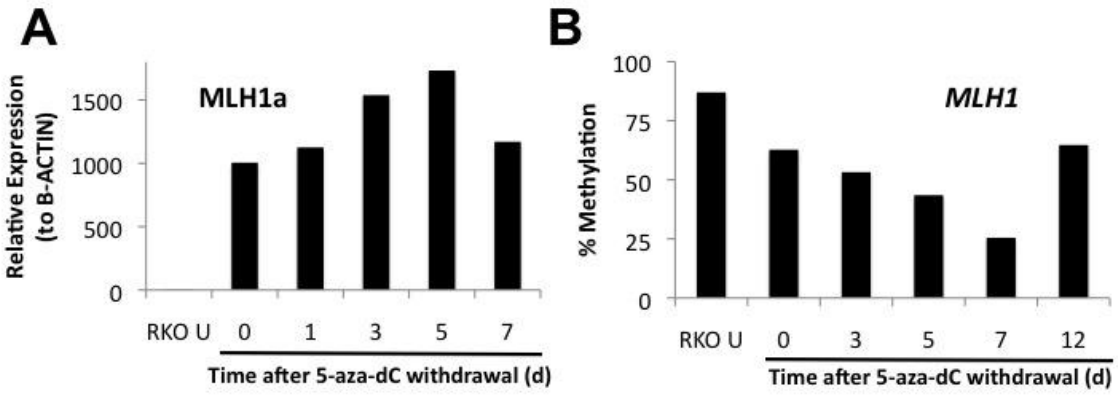

Figure 4-8. Earlier onset of *MLH1* promoter resilencing and m^5CG reacquisition after high-dose 5-aza-dC treatment and withdrawal. RKO cells were treated with 10 μM 5-aza-dC for 3 d and subsequently grown in drug-free media. A) Changes in *MLH1* relative expression during the resilencing time course. MLH1a transcript levels were normalized to β-ACTIN. Values were determined by the $\Delta\Delta C_T$ method. B) Changes in *MLH1* DNA methylation over the resilencing time course. Percent methylation was quantitatively measured by pyrosequencing and the average over four CG sites is shown.

To exclude the possibility that the slow rate of DNA remethylation observed upon withdrawal of the low dose of 5-aza-dC was unique to *MLH1*, we analyzed two other TSGs known to be epigenetically silenced in RKO cells, *TIMP3* and *CDH1*. *TIMP3* encodes tissue inhibitor of metalloproteinases 3, a member of a family of inhibitors of the matrix metalloproteinases (MMP). MMPs are endopeptidases involved in degradation of the extracellular matrix (Apte et al., 1994). TIMP3 can also exert tumor suppressor activity by inhibiting vascular endothelial factor (VEGF)-mediated angiogenesis (Masson et al., 2010; Qi et al., 2003). *CDH1* codes for classic cadherin 1 (also referred to as E-cadherin), a Ca^{2+}-dependent transmembrane protein involved in cell-cell adhesion (Berx et al., 1995; Frixen et al., 1991). *CDH1* downregulation is strongly correlated with invasive potential and poor prognosis in a variety of human cancers, including those of the breast, prostate and colon (Berx and Van Roy, 2001; Grady et al., 2000; Graff et al., 1995).

For these additional target genes, the low-dose 5-aza-dC treatment regimen caused changes in DNA methylation comparable to those seen at the *EPM2AIP1*/*MLH1* bidirectional promoter. *TIMP3* methylation decreased from 80% to 30% and, as before, DNA methylation slowly increased over the resilencing time course, achieving 40% m^5CG by day 11 (Figure 4-9B). DNA methylation of *CDH1* decreased from ~90% to 50% after incubation of RKO cells with 5-aza-dC, and again slowly accumulated to 60% over the 21 d time course (Figure 4-9D). These data, by analyzing the resilencing time course of three different loci, suggest that resilencing is a common occurrence subsequent to 5-aza-dC withdrawal. Moreover, as transcriptional resilencing precedes

reaccumulation of m^5CG at all three promoters (*MLH1*, *TIMP3* and *CDH1*), DNA methylation does not appear to be the molecular driver of resilencing.

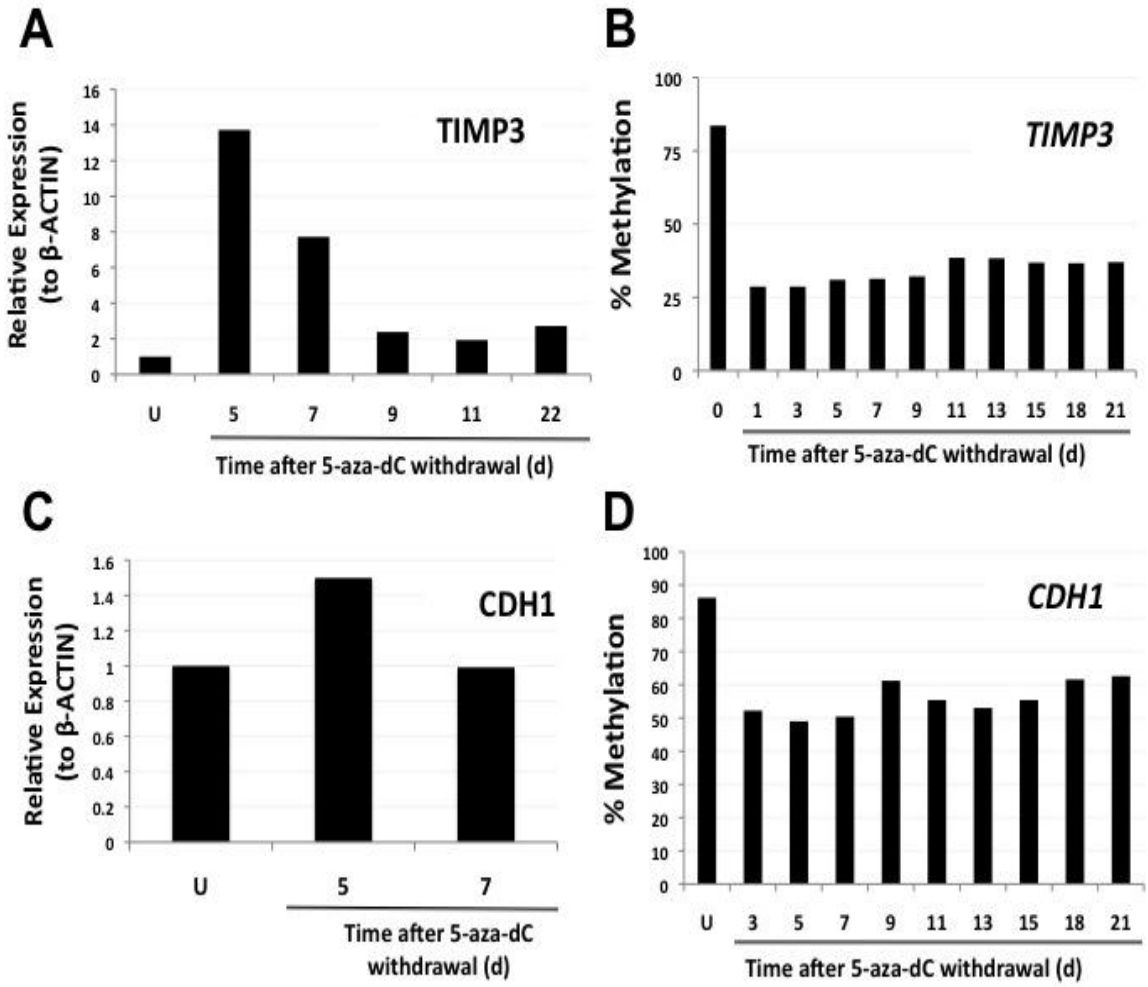

Figure 4-9 *TIMP3* and *CDH1* promoter resilencing and m^5CG reacquisition after 5-aza-dC treatment and withdrawal. RKO cells were treated as in Figure 4-5. *TIMP3* A) and *CDH1 C)* relative expression changes during the resilencing time course. Transcript levels were normalized to β-ACTIN. Values were determined by the $\Delta\Delta C_T$ method. Changes in DNA methylation at *TIMP3* B) and *CDH1 D)* during resilencing. Shown is the percent methylation quantitatively measured by pyrosequencing and averaged over eight and six CG sites at *TIMP3* and *CDH1*, respectively. U, untreated RKO cells.

Demethylation of hypermethylated TSGs is required for 5-aza-dC-mediated derepression.

Although our results so far suggest that DNA methylation is not an upstream driver of resilencing, we wanted to test if removal of m^5CG by 5-aza-dC addition is required for

transcriptional derepression of a highly methylated gene. RKO cells were thus treated for 24 hr with 100 nM of the HDAC inhibitor TSA in order to accumulate positive histone acetylation found at actively-transcribed genes. Significant changes in either the level of MLH1a transcript or CG methylation were not observed (Figure 4-10). This suggests that, at least for a heavily methylated promoter, CG methylation acts as a maintenance mechanism for the transcriptional memory of the gene.

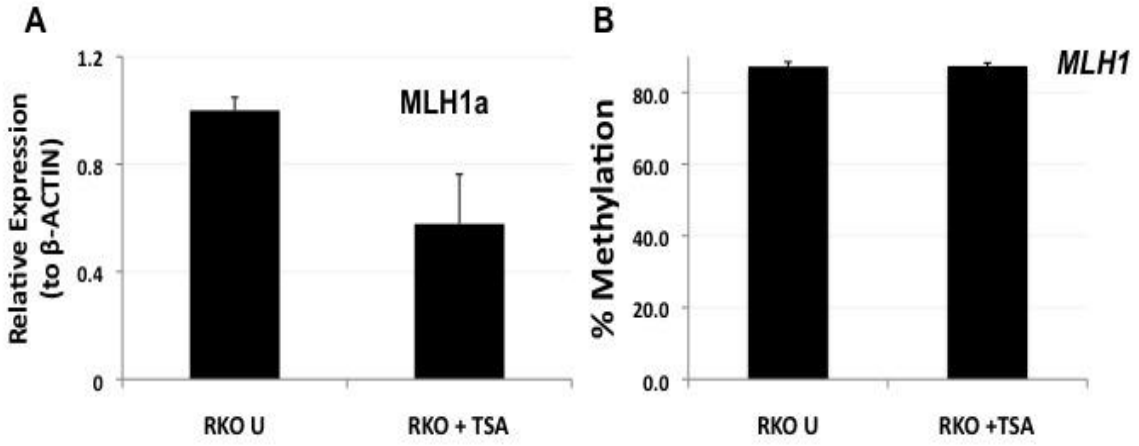

Figure 4-10 DNA demethylation is required for MLH1 transcriptional derepression. A) *MLH1* relative expression after TSA of RKO cells. Transcript levels were normalized to β-ACTIN. Values were determined by the $\Delta\Delta C_T$ method. B) Quantitative DNA methylation analysis of the *MLH1* promoter by pyrosequencing of DNA isolated from RKO cells ± TSA. The average percent methylation of four CG sites in the *MLH1* promoter is shown. Error bar, N = 3 technical replicates.

Changes in Chromatin Structure Correlated with TSG Derepression and Resilencing After 5-Aza-dC Treatment.

A previous study examined the effects of 5-aza-dC addition and its subsequent withdrawal on a hypermethylated reporter gene that had been integrated at an unknown genomic location (Si et al., 2010). They showed that histone H3 occupancy at the transgene promoter increased rapidly and correlated with its resilencing. Since we found that downregulation of transcription the *MLH1*, *TIMP3* and *CDH1* promoters preceded remethylation after removing 5-aza-dC, we sought to determine if changes in

chromatin accessibility also occurred during resilencing. We also wanted to map any structural changes at high resolution and their timing relative to when transcript levels began to decline.

We therefore performed MAPit single molecule footprinting on nuclei purified at different times after 5-aza-dC withdrawal (Figures 4-11 and 4-12). First, we analyzed the *MLH1* promoter when derepression levels of expression were peaking, at day 8 post-drug removal (cf. Figure 4-6). At this time point, the chromatin structure of some copies of the *MLH1* promoter resembled that of an actively transcribed promoter, with the NDRs of TSSa and TSSb in more or less the same locations as were observed in HCT116 cells (compare Figure 4-11 to 4-4B). Of particular interest was the loss of the footprint within the NDR of TSSb in HCT116 cells in Figure 4-4B, and appearance of a new footprint of comparable size within the TSSa NDR (Figure 4-11B).

As expected, levels of endogenous methylation were reduced at the *MLH1* promoter by treatment with 5-aza-dC (Figure 4-11B). However, as demethylation by 5-aza-dC seldom goes to completion, single molecule MAPit analysis enabled us to distinguish the chromatin structures of promoters that were demethylated from those that were not. The 7 molecules in Figure 4-11B that showed significant DNA demethylation also displayed one or both M.CviPI-accessible NDR(s). Interestingly, none of these 7 molecules was completely demethylated, and 6 of 7 retained methylation of 2-4 CG sites just upstream of TSSb. In contrast, molecules that retained dense CG methylation showed no significant GC accessibility and are likely transcriptionally silent, i.e. they lacked NDRs characteristic of the active promoter found in HCT116 cells (cf. Figure 4-4B).

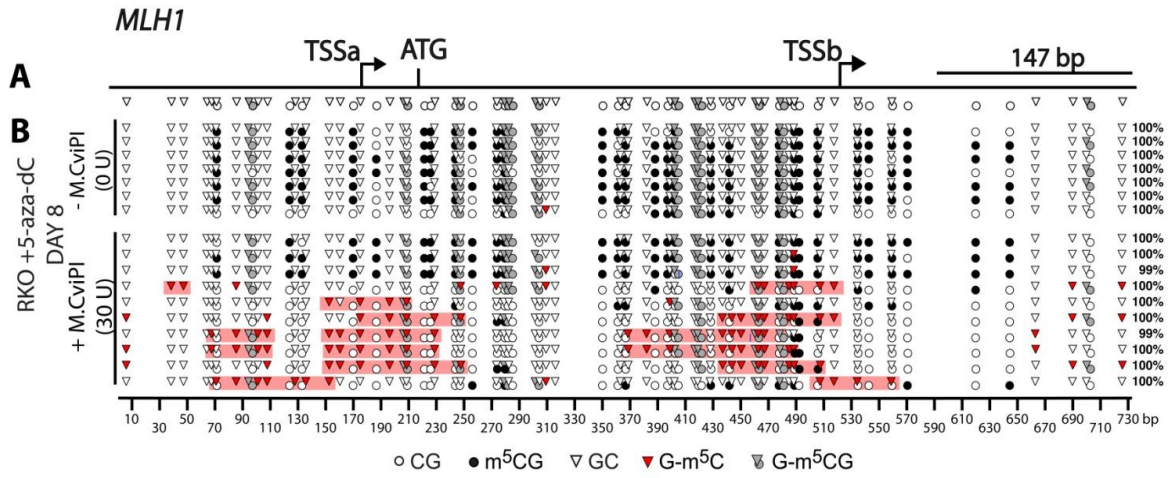

Figure 4-11. Mixed chromatin populations are found in derepressed RKO cells after 5-aza-dC treatment. A) Map (to scale) of the 755 bp region analyzed by MAPit. B) MAPit analysis of the derepressed *MLH1* promoter region in RKO cells at day 8 after treatment with 5-aza-dC. Symbols are defined as in Figure 4-3. Map coordinates (in bp) of the BGS amplicon are shown at the bottom of the molecules in B.

Next, we analyzed the chromatin organization at the *EPM2AIP1* promoter by MAPit of RKO cells that had recovered from 5-aza-dC treatment for various times (Figure 4-12). We analyzed cells at three different times, when expression of the MLH1a promoter was submaximal (day 5), peaking (day 8) and declining (day 13) (cf. Figure 4-7A). In Figure 4-12A, by day 5, small areas of accessibility of 40-70 bp had appeared, consistent with internucleosomal linkers that became accessible as nucleosome movement began (presumably towards producing the open chromatin conformation ultimately observed at day 8). Accessible GC sites initially appeared in an area downstream of the promoter that coincided with the smaller downstream accessible area observed in the expressed promoter of HCT116 cells (compare Figures 4-12A and 4-3B; map units 1-350). At peak expression at day 8, some 5-aza-dC-treated RKO cells showed a chromatin organization that resembled that of the active promoter seen in HCT116 cells, i.e. a NDR encompassing the TSS (Figure 4-12C). Again, as observed

just upstream of the TSS in HCT116 cells, in 5-aza-treated RKO cells, this NDR appeared to be occupied by a factor that created a footprint (Figure 4-12C; +M.CviPI, last two molecules). At day 13 (Figure 4-12D), when expression began to decline, chromatin accessibility downstream of the TSS NDR (map units 50 to 190 bp) had almost disappeared and few molecules retained the TSS NDR. These results indicate that the open chromatin organization of the *EPM2AIP1* gene was being lost as it was resilenced in the population of cells.

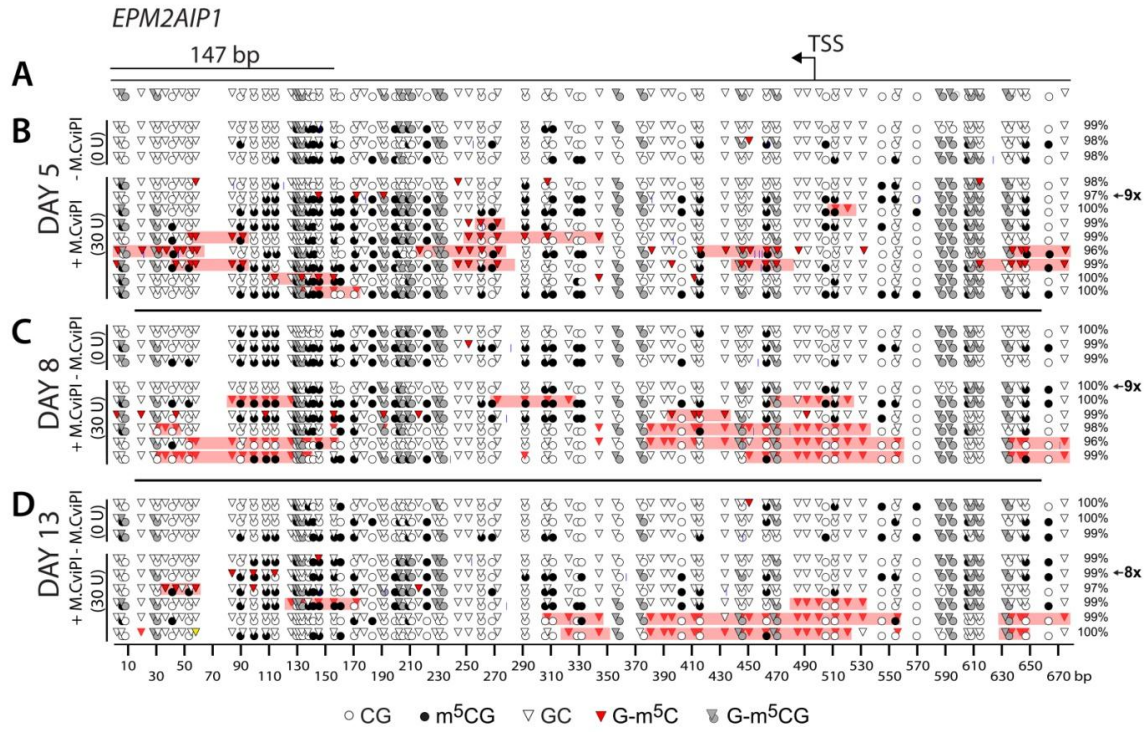

Figure 4-12. Dynamic changes in chromatin structure at the *EPM2AIP1* promoter correlate with expression changes during gene resilencing. A) Map (to scale) of the 682 bp region analyzed by MAPit. MAPit of RKO cells that had recovered from 5-aza-dC treatment at day 5 B), day 8 C) and day 13. Multipliers on the right side represent the number of molecules with the same chromatin accessibility and levels of endogenous methylation. Symbols are as defined in Figure 4-3. Map coordinates (in bp) of the BGS amplicon are shown at the bottom of the molecules in D.

RKO cells treated with 5-aza-dC respond to a second treatment with the nucleoside analog

One concern regarding the above gene resilencing time courses is selection of a subpopulation of 5-aza-dC-resistant cells that eventually predominates the population, perhaps due to a growth advantage. Our results, however, do not support this premise, because the initial level of demethylation was essentially maintained for the duration of the time course in the face of mounting resilencing. Nevertheless, to further exclude selection of 5-aza-dC-resistant subpopulation of cells, 22 d after 5-aza-dC removal, we re-subjected the RKO cells to the same low dose treatment regimen (3 d +5-aza-dC; 1 d drug-free recovery), and measured levels of expression and m^5CG of *MLH1*, *TIMP3* and *CDH1* (Figure 4-13).

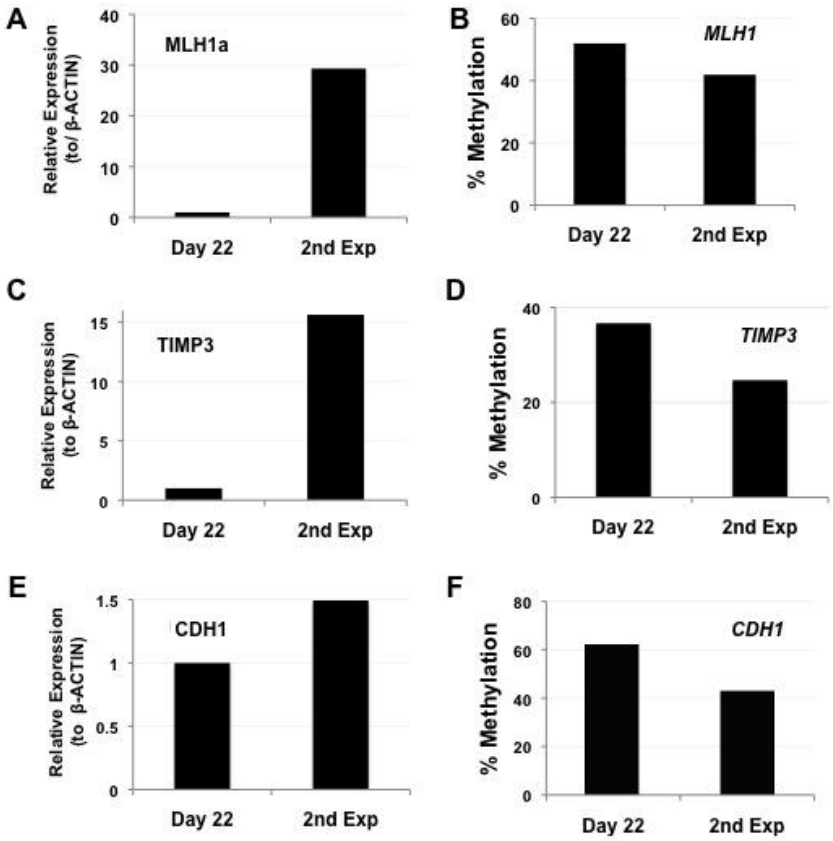

Figure 4-13. RKO cells respond to a second treatment with 5-aza-dC. RKO cells were subjected to a second 5-aza-dC treatment regimen on day 22 after recovery

from the first. After 24 h of removing 5-aza-dC for the second time, *MLH1* A), *TIMP3* C) and *CDH1* E) expression levels were measured. Transcript levels were normalized to β-ACTIN. Values were determined by the $\Delta\Delta C_T$ method. Also, quantitative DNA methylation analysis by pyrosequencing was measured in the same experiment for *MLH1* B), *TIMP3* D) and *CDH1* F).

After re-treatment with 5-aza-dC, all three genes derepressed levels of transcription to similar degrees caused by the initial exposure (Figure 4-13). In contrast, there was only a modest 10% reduction in m^5CG at each gene as compared to the initial level at day 22. These small reductions in m^5CG in the face of dramatic increases in expression, at least of *MLH1* and *TIMP3*, suggest that 5-aza-dC-mediated derepression is either indirect or involves a component(s) independent of DNA-methylation.

Intermediate levels of CG methylation in clonal isolates argues against mixed populations of cells with either fully methylated and unmethylated alleles

The persistent intermediate levels of DNA methylation observed during the resilencing time course could formally be attributable to a mixture of cells harboring alleles that were either densely methylated or completely unmethylated. To address this possibility, we established independent clonal cell populations from the initial 5-aza-dC treated RKO cell population. Clones were established 18 d after 5-aza-dC withdrawal by limited dilution and were independently propagated in 5-aza-dC-free media. On day 35, we analyzed levels of MLH1a expression and m^5CG in each established clone (Figure 4-14A). Most clones (75%) maintained intermediate levels of m^5CG from 20 to 60%. Only 15% (3 of 20) of the clones, showed endogenous amounts of m^5CG above 60%. Interestingly none of the 20 clones reacquired m^5CG levels of ~95% as observed in the stock RKO population, i.e. not treated with 5-aza-dC. In contrast, 15% (3 of 20) of the

clones maintained low DNA methylation levels of <20% (Figure 4-14A and C). These results indicate that the majority of RKO cells in the original population responded to the demethylating effects of 5-aza-dC.

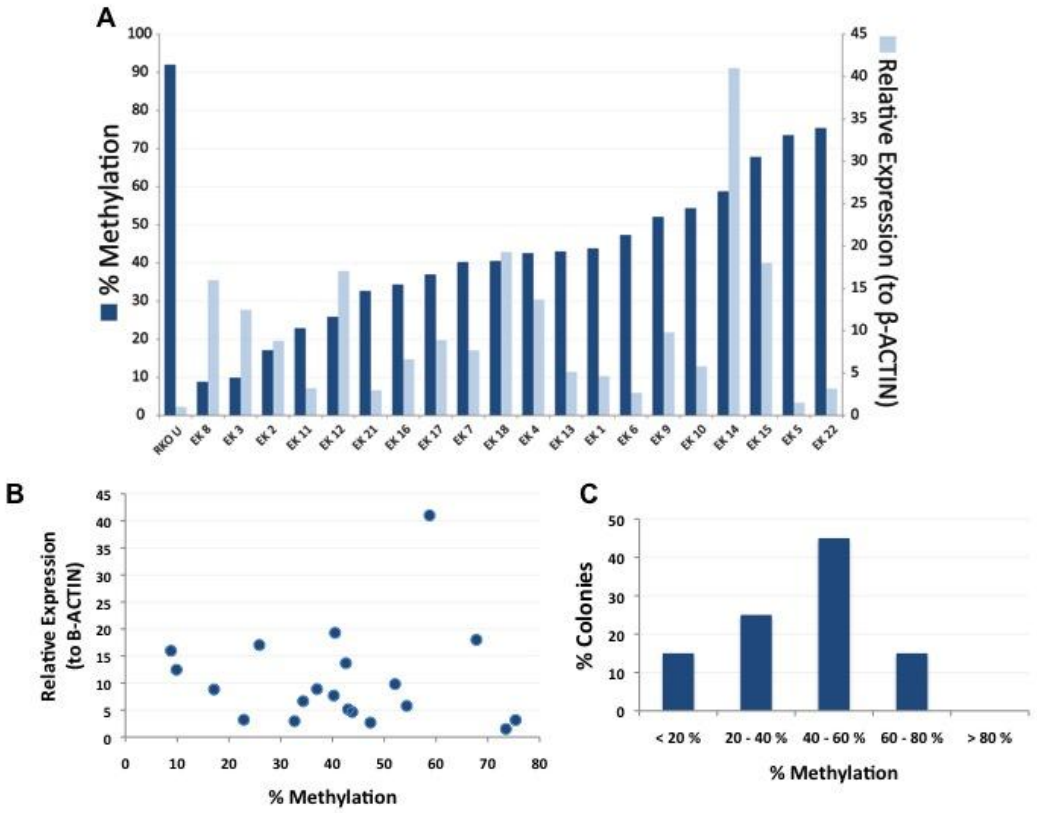

Figure 4-14 Maintenance of intermediate levels of DNA methylation and gene expression in clonal cell populations. A) DNA methylation levels and relative expression for clones derived from RKO cells treated with a low dose of 5-aza-dC. Clonal populations (labeled EK #) were sorted in ascending order according to their m^5CG level as determined by pyrosequencing (light blue bars). The corresponding relative expression of MLH1a is represented by dark blue bars. B) Insignificant correlation between m^5CG level and MLH1a expression in the populations of clonal cells. C) Distribution of clones with various ranges of DNA methylation.

Each of the clonal RKO cell populations retained residual *MLH1* expression, even after 35 d of culture in the absence of 5-aza-dC. None of the clones showed as low an expression level as that observed in the original 5-aza-dC-untreated RKO cell line (Figure 4-14A). However 91% of the clones showed resilencing relative to the

expression level observed at the time point of peak expression, day 8 after 5-aza-dC was removed. In contrast, 9% of the clonal populations actually exhibited higher expression of MLH1a than this peak expression (Figure 4-14B). Taken together, the observed expression patterns of the clonally derived populations argue strongly against the presence of a substantial subpopulation of 5-aza-dC-unresponsive cells in the original RKO cell line.

DNA methylation and expression levels were also not strongly correlated among the individual RKO cell clones (Figure 4-14B). Both DNA methylation and expression levels were intermediate to those seen in the untreated population and after 5-aza-dC treatment. Taken together, these results argue strongly against a 5-aza-dC-resistant subpopulation of cells. Moreover, our data suggest that, upon removal of 5-aza-dC, DNA methylation does not appear to drive resilencing of the genes we tested.

Discussion

The role of DNA hypermethylation of promoters in epigenetic silencing of TSGs is a well-established epigenetic phenomenon contributing to cancer initiation and progression (Esteller, 2008; Esteller and Herman, 2002; Feinberg et al., 2006) (McCabe et al., 2009; Sincic and Herceg, 2011). DNA demethylating agents like 5-aza-dC were therefore among the first epigenetic altering drugs to be used in the clinic to reactivate silenced TSGs and inhibit cancer cell growth (Cortez and Jones, 2008; Yang et al., 2010). However, after completing the course of 5-aza-dC treatment, frequently TSG reactivation is not maintained and TSG expression gradually returns to the pre-therapy silenced state. Enhancing the efficacy of DNA demethylating drugs as therapeutic agents requires further elucidation of mechanisms involved in the incipient reactivation

as well as the initiation and progression of the ensuing resilencing (Issa and Kantarjian, 2009; Oki et al., 2007).

To date, research on epigenetic regulation has focused almost exclusively on the contributions made by the interplay between DNA methylation and histone modifications (Kagey et al., 2010; Wong et al., 2011). Our goal herein was to develop a better understanding of the changes in chromatin structure known to be associated with epigenetic dysregulation of TSGs (Esteller, 2007a, 2008; Jones and Baylin, 2007; Kondo and Issa, 2010; Si et al., 2010; Vaissiere et al., 2008). To do so, we investigated the temporal coordination of DNA methylation and chromatin accessibility in RKO colon cancer cells during TSG reactivation in response to 5-aza-dC and resilencing subsequent to drug removal.

A key advance of our studies was the use of MAPit methylation protection to simultaneously map CG methylation and chromatin accessibility to individual DNA molecules. This single molecule capability avoided averaging the contributions of diverse molecules at potentially different stages of TSG reactivation and resilencing. We focused on the well-characterized *EPM2AIP1*/*MLH1* bidirectional promoter, previously reported to be epigenetically silenced in colon cancer and to undergo resilencing after removal of 5-aza-dC (Veigl et al., 1998).

To obtain baseline chromatin structures of the *EPM2AIP1* and *MLH1* promoters, we first performed MAPit on nuclei isolated from 5-aza-dC-untreated cells. Single promoter molecules from HCT116 cells, which actively transcribe both promoters, exhibited a chromatin structure characteristic of expressed genes (Ozsolak et al., 2007; Schones et al., 2008). More specifically, active promoter copies showed accessible

areas of ≥150 bp in size in the vicinity of TSSs. The *EPM2AIP1* promoter shows an open area of ~250 bp around and upstream of the TSS, consistent with eviction of a single nucleosome, i.e. 147 bp nucleosomal DNA plus two linkers of ~50 bp each. Occasionally, this nucleosome-depleted region (NDR) extends all the way to the *MLH1* TSSa as determined by MAPit analysis of the intervening region. Chromatin structure of the *MLH1* promoter is most consistent with two NDRs, one associated with each TSS, flanking a single, loosely positioned nucleosome. These NDRs were found only in the context of DNA hypomethylation (Figures 4-3B and 4-4B) and likely serves to facilitate recruitment and assembly of the RNA pol II preinitiation complex. In support of this model, subnucleosomal footprints were frequently observed within NDRs at the *EPM2AIP1* TSS and *MLH1* TSSb. The mechanism by which NDRs at the *EPM2AIP1*/*MLH1* bidirectional promoter are created, e.g. direct nucleosome disassembly, sliding or combination thereof, is unknown and requires further study.

In contrast, the silenced *EPM2AIP1* and *MLH1* promoters in RKO cells showed dense CG methylation and dramatically reduced accessibility, probably due to high nucleosome occupancy (Figures 4-3C and 4-4C). Small areas of GC accessibility observed in some molecules suggest the presence of weakly positioned or “fuzzy” nucleosomes in a fraction of RKO cells represented by these molecules. The absence of accessible linkers on the remaining molecules precludes mapping nucleosome positions on them. However, the highly-closed structure of these molecules is consistent with higher-order compaction of arrays of nucleosomes, which can limit accessibility of linker DNA as well (reviewed in (Li and Reinberg, 2011)). Linker DNA may also be protected by linker histones and other non-histone proteins (e.g. MBD-containing

proteins) known to associate with epigenetically silenced regions (Dhasarathy and Wade, 2008; Georgel et al., 2003; Wade, 2001).

The presence of NDRs as seen in HCT116 cells (Figures 4-3B and 4-4B) does not indicate active transcription *per se*. Nevertheless, transcriptional derepression of the silenced *MLH1* promoter in RKO cells after 5-aza-dC treatment occurred in concert with appearance of NDRs on a fraction of molecules. Remarkably, and in contrast to previous studies (Lin et al., 2007), we observed that the majority of molecules obtained after 5-aza-dC treatment of RKO cells were only partially depleted of CG methylation yet acquired areas of accessibility to M.CviPI. We cannot formally rule out complete demethylation of least a fraction of molecules and rapid *de novo* by 5 d after 5-aza-dC removal, the first assay point of Figure 4-12A; however, this possibility is undermined by the absence of significant accrual of m^5CG at three different TSGs over an additional 16 d of growth (Figures 4-7B and 4-9B, D). This slow rate of m^5CG accrual also argues against a significant selective growth advantage of 5-aza-dC-resistant cells having taken over the culture population. Regardless, it is of particular note that the region of highest m^5CG depletion co-localized with the *EPM2AIP1* TSS (Figure 4-12). This begs the question of whether chromatin remodeling associated with creating a NDR upon addition of 5-aza-dC constitutes a component of active demethylation rather than passive demethylation during replication. Alternatively, it is possible that the NDR itself is unable to be methylated by endogenous DNMTs. Taking the expression data into account, our results strongly suggest that transcription can occur in the presence of substantial levels of promoter methylation.

CGI hypermethylation of TSG promoters is often ascribed as the molecular lesion that underlies epigenetic gene silencing in cancer. There are, however, examples where silencing occurs in the absence of increased CG methylation (Banelli et al., 2000; Kondo et al., 2008; Markus et al., 2007). Reinforcing the idea that m^5CG is not required for gene silencing, we observed that transcriptional resilencing after withdrawal of a low dose of 5-aza-dC occurred in the absence of significant DNA remethylation (Figures 4-7B and 4-9B, D). Even after administration of the high 5-aza-dC dose, promoter resilencing occurred well before DNA remethylation (Figure 4-8).

Although gene resilencing in our system was not dependent on significant overall DNA remethylation, it is important to consider the alternate model that DNA methylation may orchestrate promoter resilencing. Classic studies have shown that Sp1 cannot bind to sites containing m^5CG *in vitro* and *in vivo* (Lee et al., 1998; Li et al., 1994). More recently it has also been found that modification of a few specific CG sites in the *Myogenin* gene in mouse cells interferes with transcription factor binding and promoter activation (Palacios et al., 2010). In this scenario, DNA methylation density does not have to change dramatically to drive the TSG resilencing that we observed. That stated, it seems unlikely that all three loci we investigated would be directly regulated by a common transcription factor, whose binding is blocked by specific retention of m^5CG (or remethylation) at each of its binding sites during the 5-aza-dC treatment protocol.

We (Figure 4-9) and others (Ou et al., 2007) have shown that the HDAC inhibitor TSA does not lead to derepression or demethylation of densely methylated TSGs. This indicates that inhibition of histone deacetylation is not sufficient to reactivate such promoters. By contrast, epigenetically silenced genes with low levels of promoter DNA

methylation can be reactivated by TSA (Ou et al., 2007). These observations suggest that maintenance of the silenced state at genes with dense m^5CG involve multiple, redundant epigenetic pathways that may each reinforce each other and vary in a gene specific manner. Along these lines, 5-aza-dC addition leads not only to gene derepression concomitant with NDR formation (Figures 4-11 and 4-12), but also to increases in positive histone marks (H3K4me3 and histone acetylation) and reduction of silent histone marks (H3K9me3 and H3K27me3) (Si et al., 2010; Wong et al., 2011). Taking these and our data together, TSG derepression apparently requires DNA demethylation, reprogramming of histone modifications and reorganization of chromatin to an open accessible state.

Apparently, based on our results, once 5-aza-dC has overridden multiple repressive epigenetic layers, TSGs can be resilenced without a prerequisite of dense DNA methylation. Our data therefore point to a mechanism of resilencing after 5-aza-dC removal that is driven by chromatin architecture, e.g. loss of NDRs, which may also involve histone marks. During resilencing, as might be expected over time, both H3K4me2/3 and histone acetylation decline, whereas H3K9me3 and/or H3K27me3 accumulate in a gene specific manner. Controversy remains on how histone marks contribute to the resilencing process (Kagey et al., 2010; Si et al., 2010; Wong et al., 2011). Our model does not rule out the possibility that slow accrual of m^5CG over very long times may indeed reinforce and maintain long-term epigenetic silencing.

CHAPTER 5
SUMMARY AND FUTURE DIRECTIONS

Epigenetic regulation is crucial for the proper orchestration of gene expression in non-diseased cells. It is widely accepted that the different layers of epigenetic regulation (DNA methylation, histone modifications and nucleosome positioning) act in concert to exercise proper control of gene expression. The last decade has witnessed an unprecedented increase in epigenetic investigation that has resulted in the cataloguing patterns of DNA methylation, histone marks and DNA methylation in normal and disease states. While many strides have been made, a detailed mechanistic understanding of how the various layers of epigenetic regulation influence each other and are functionally coordinated to elicit and maintain cell- and tissue-specific expression programs has not emerged.

Transcriptional silencing of TSGs was one of the first epigenetic events to be directly associated with cancer development and progression. A principal goal of our studies was to elucidate how different layers of epigenetic regulation collaborate to drive TSGs from a highly-expressing to permanently-silenced state. To address this, we first developed the novel MAPit technique, enabling an integrative view of both endogenous CG methylation and chromatin accessibility in a single experimental platform. Moreover, the single molecule resolution afforded by the technique avoids population averaging and definitively addresses if epigenetic changes are map to the same or different subpopulations of molecules. To support these studies computationally, we developed the MethylViewer program for rapid and accurate analysis of complex MAPit datasets in collaboration with bioinformaticists at the University of Leeds.

Lastly, we studied epigenetic progression of TSG resilencing in cultured colon cancer cells after administration of the therapeutic demethylating agent, 5-aza-dC. This resilencing phenomenon is of great clinical interest, because it is a major obstacle to long-term remission in therapies aimed to reverse DNA hypermethylation. We observed for the first time progressive changes in chromatin accessibility occurring at TSG promoters during resilencing. More specifically, we found that, at least for the three TSGs monitored in this study, resilencing was not accompanied by significant reacquisition of DNA methylation following its removal by 5-aza-dC. Instead, we observed that resilencing of TSG expression tracked most closely with changes in chromatin accessibility. Therefore, our data support a model in which downregulation of transcription leads to changes in promoter chromatin structure, or changes in chromatin organization drive resilencing. We emphasize that the scenario of TSG resilencing may not parallel the *de novo* epigenetic events that established TSG silencing.

Simultaneous Single-Molecule mapping of Protein-DNA Interactions and DNA Methylation by MAPit

Herein, we established the proof of principle for using MAPit to map nucleosome positioning at high resolution in transcriptionally-active *versus* -repressed TSG promoters, and directly correlate this information to DNA methylation (Chapter 2). We expect that the development of MAPit for the use in mammalian systems, which constituted a significant part of this study, will facilitate our understanding of the interplay between DNA methylation and nucleosome positioning. By distinguishing if different types of epigenetic changes map to the same or different molecules, MAPit could eventually aid navigation of the epigenetic complexity in heterogeneous cell populations characteristic of human tumors. Interrogating heterogeneous molecular

populations by MAPit can provide crucial information as to whether, The use of M.CviPI, a GC-methylating enzyme that we cloned and expressed previously, was a crucial advance that allowed adaptation of MAPit to mammalian cultured cells. For the first time we could unequivocally map two distinct epigenetic events, GC accessibility and endogenous CG methylation, onto the same DNA molecule. Furthermore, we could detect subpopulations of molecules with distinct chromatin architectures over the population of cells.

Future advancement of MAPit poses exciting opportunities as well as challenges. While the studies herein were confined to cultured mammalian cells, we see no reason why MAPit could not be optimized for profiling epigenetic diversity in primary human or animal samples. This would constitute a remarkable advance as diseased human tissue, particularly tumors, is incredibly heterogeneous, and thus far only studied by population-averaged techniques to query chromatin structure. This is especially important as it is believed that the inherent cellular heterogeneity of tumors underlies therapeutic failure for many tumor types. Thus, the ability to deconvolve tumor epigenetic heterogeneity could increase the likelihood of identifying tumor-tailored epigenetic therapies.

A second conceivable development would be the melding of MAPit and chromatin immunoprecipitation. This would allow for the correlation of specific DNA binding factors to characteristic chromatin structures or specific chromatin accessibility patterns within the context of DNA methylation in a single assay. A third development termed “pulse-chase MAPit” capitalizes on the benefits of using DNMT probes rather than nucleases to define chromatin states whereby the interrogated molecules are not subject to

cleavage or degradation. Since probed molecules remain intact, one could utilize the diverse substrate specificities of DNMTs to mark chromatin with one probe, apply a manipulation of interest, then mark chromatin with a second probe thus tracing the remodeling of a single molecule from an initial to a final state. Finally, since the output of MAPit data is obtained by BGS, any application of MAPit could be optimized for large-scale BGS platforms such as Illumina or 454 sequencing. With recent advances in high-throughput bisulfite sequencing techniques, the first genome-wide or target-enriched MAPit chromatin studies are around the corner.

MethylViewer: A Computational Analysis and Editing for Bisulfite Sequencing and Methyltransferase Accessibility Protocol for Individual Templates (MAPit) Projects

MAPit generates enormous amounts of sequencing data, the analysis of which is labor intensive, time consuming and prone to human error. Realizing its potential applications, we were motivated to seek an appropriate collaboration to develop MethylViewer, a stand-alone freeware program for analysis of MAPit datasets and rapid generation of publication-quality images. The program is very versatile, allowing simultaneous analysis of methylation status of up to four different recognition sites, or processing of traditional BGS datasets as well. MethylViewer can also be used to design primers for MAPit analysis, which is more challenging than designing primers for traditional BGS. Finally, MethylViewer provides a binary output of the modification map that can be used to input into other programs to generate heat maps or for statistical calculations. All MAPit data presented in this study were analyzed with MethylViewer and figures of these analyses were obtained by using its graphical output feature.

Interplay between DNA Methylation and Chromatin Structure during TSG Resilencing in Colon Cancer Cells

Elucidating the molecular mechanisms that drive epigenetic gene silencing is crucial to a better understanding of epigenetic disease. Elucidating how epigenetic TSG silencing occurs in cancer is anticipated to aid identification of biomarkers of early stage carcinogenesis, and the improvement of therapeutic strategies for cancer treatment.

The DNA demethylating agents Vidaza and Decitabine have proven useful in the treatment of myelodysplastic syndromes. Attempts are in progress to expand their use to solid tumors are likely to be hampered as the mechanism of action of these drugs is not completely understood. In addition, one of the main drawbacks of DNA demethylating treatments is the puzzling but frequent observation, in both cell lines and patients, of resilencing of TSGs that were initially reactivated by the drug treatment.

We used a cell line-based system to study TSG resilencing after 5-aza-dC treatment. These studies yielded provocative results. First, we found that DNA methylation density does not appear to be a significant driver of the resilencing process, which contradicts other reports (Wong et al., 2011). MAPit analysis allowed us to determine that TSG resilencing after drug removal correlates better with gene promoter chromatin structure. Chromatin accessibility changed quickly during the resilencing time course as opposed to levels of endogenous methylation that remained rather constant over the full duration of each experiment.

It should be noted that completing these studies would not have been possible without the use of the GC-methylating enzyme M.CviPI. First, the cellular population following 5-aza-dC treatment and withdrawal was highly heterogeneous. Thus, chromatin probing methods based on population averaging would have obscured many

of the epigenetic features identified at the single molecule level by MAPit. Second, since our TSGs of interest were endogenously hypermethylated, previous uses of MAPit utilizing M.SssI (a CG specific DNMT) would have precluded designation of a particular methylation event as endogenous or probe mediated.

In support of our findings that nucleosome organization corresponds more closely to TSG resilencing than DNA methylation, Si et al. (2010) also found that increased nucleosome occupancy rather than CG methylation was associated with gene resilencing following 5-aza-dC withdrawal. The authors found that total histone H3 levels were enriched in cells that resilenced *versus* those that maintained expression. As conventional ChIP was used to identify these differences, cells needed to be sorted into silenced and expressing subpopulations, which necessitated the use of an artificial reporter transgene rather than an endogenous locus that we studied.

It is of high interest to further evaluate how histone modifications change during the resilencing time course. In principle, this data can be integrated with CG methylation and chromatin accessibility data obtaining by MAPit, by probing nuclei with M.CviPI followed by ChIP (i.e. MAPit-ChIP), yielding further temporal insights into the resilencing process. Identification of mechanistic driver(s) in the process will require evaluation of the extent to which knockdown of specific candidate chromatin remodelers and post-translational modifiers by RNAi disrupt TSG resilencing. These studies are expected to uncover important mechanistic information and potentially identify new targets for therapeutic interventions and/or point to combinatorial therapeutic strategies that could enhance the efficacy of clinically-used DNA methylation inhibitor.

LIST OF REFERENCES

Adams, A., Gottschling, D.E., Kaiser, C.A., and Stearns, T. (1997). Methods in yeast genetics: A Cold Spring Harbor Laboratory course manual.

Ai, L.B., Kim, W.J., Demircan, B., Dyer, L.M., Bray, K.J., Skehan, R.R., Massoll, N.A., and Brown, K.D. (2008). The transglutaminase 2 gene (TGM2), a potential molecular marker for chemotherapeutic drug sensitivity, is epigenetically silenced in breast cancer. Carcinogenesis *29*, 510-518.

Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature *446*, 572-576.

Almer, A., and Horz, W. (1986). Nuclease hypersensitive regions with adjacent positioned nucleosomes mark the gene boundaries of the PHO5 / PHO3 locus in yeast. EMBO.J. *5*, 2681-2687.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of Molecular Biology *215*, 403-410.

Apte, S.S., Mattei, M.G., and Olsen, B.R. (1994). Cloning of the cDNA-enconding human tissue inhibitor of encoding human tissue inhibitor of metalloproteinases-3 (TIMP3-3) and mapping of the TIMP3 gene to chromosome-22. Genomics *19*, 86-90.

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature *455*, 818-U812.

Bachman, K.E., Park, B.H., Rhee, I., Rajagopalan, H., Herman, J.G., Baylin, S.B., Kinzler, K.W., and Vogelstein, B. (2003). Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. Cancer Cell *3*, 89-95.

Bai, L., and Morozov, A.V. (2010). Gene regulation by nucleosome positioning. Trends in Genetics *26*, 476-483.

Bakker, J., Lin, X.H., and Nelson, W.G. (2002). Methyl-CpG binding domain protein 2 represses transcription from hypermethylated pi-class glutathione S-transferase gene promoters in hepatocellular carcinoma cells. Journal of Biological Chemistry *277*, 22573-22580.

Ball, M.P., Li, J.B., Gao, Y., Lee, J.H., LeProust, E.M., Park, I.H., Xie, B., Daley, G.Q., and Church, G.M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nature Biotechnology *27*, 361-368.

Banelli, B., Casciano, I., and Romani, M. (2000). Methylation-independent silencing of the p73 gene in neuroblastoma. Oncogene *19*, 4553-4556.

Barrett, T.J., and Spelsberg, T.C. (1998). Steroid receptors at the nexus of transcriptional regulation. J.Cell Biochem.Suppl *30-31:185-93*, 185-193.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Baylin, S.B., and Ohm, J.E. (2006). Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nature Reviews Cancer *6*, 107-116.

Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., and Schreiber, S.L. (2004). Global nucleosome occupancy in yeast. Genome Biol *5*, R62.

Berx, G., Staes, K., Vanhengel, J., Molemans, F., Bussemakers, M.J.G., Vanbokhoven, A., and Vanroy, F. (1995). Cloning and characterization of the human invasion suppressor gene e-cadherin (CDH1). Genomics *26*, 281-289.

Berx, G., and Van Roy, F. (2001). The E-cadherin/catenin complex: an important gatekeeper in breast cancer tumorigenesis and malignant progression. Breast Cancer Res. *3*, 289-293.

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA-encoding DNA methyltransferase of mouse cells - the carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. Journal of Molecular Biology *203*, 971-983.

Bestor, T.H. (2000). The DNA methyltransferases of mammals. Hum.Mol.Genet. *9*, 2395-2402.

Bestor, T.H., and Bourc'his, D. (2004). Transposon silencing and imprint establishment in mammalian germ cells. Cold Spring Harbor symposia on quantitative biology *69*, 381-387.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6-21.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209-213.

Bird, A.P., and Wolffe, A.P. (1999). Methylation-induced repression - Belts, braces, and chromatin. Cell *99*, 451-454.

Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J., and Lengauer, T. (2005). BiQ analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. Bioinformatics (Oxford, England) *21*, 4067-4068.

Boeger, H., Griesenbeck, J., Strattan, J.S., and Kornberg, R.D. (2004). Removal of promoter nucleosomes by disassembly rather than sliding *in vivo*. Mol Cell *14*, 667-673.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311-322.

Brenner, C., Deplus, R., Didelot, C., Loriot, A., Vire, E., De Smet, C., Gutierrez, A., Danovi, D., Bernard, D., Boon, T.*, et al.* (2005). Myc represses transcription through recruitment of DNA methyltransferase corepressor. Embo Journal *24*, 336-346.

Bronner, C.E., Baker, S.M., Morrison, P.T., Warren, G., Smith, L.G., Lescoe, M.K., Kane, M., Earabino, C., Lipford, J., Lindblom, A.*, et al.* (1994). Mutation in the DNA mismatch repair gene homolog *HMLH1* is associated with hereditary nonpolyposis colon-cancer. Nature *368*, 258-261.

Burgers, W.A., Fuks, F., and Kouzarides, T. (2002). DNA methyltransferases get connected to chromatin. Trends in Genetics *18*, 275-277.

Cairns, B.R. (2009). The logic of chromatin architecture and remodelling at promoters. Nature *461*, 193-198.

Cameron, E.E., Bachman, K.E., Myohanen, S., Herman, J.G., and Baylin, S.B. (1999). Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. Nat.Genet. *21*, 103-107.

Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. Current opinion in genetics & development *14*, 155-164.

Carr, I.M., Valleley, E.M., Cordery, S.F., Markham, A.F., and Bonthron, D.T. (2007). Sequence analysis and editing for bisulphite genomic sequencing projects. Nucleic Acids Res *35*, e79.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P.*, et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

Chadee, D.N., Taylor, W.R., Hurta, R.A., Allis, C.D., Wright, J.A., and Davie, J.R. (1995). Increased phosphorylation of histone H1 in mouse fibroblasts transformed with oncogenes or constitutively active mitogen-activated protein kinase kinase. J.Biol.Chem. *270*, 20098-20105.

Chan, S.H., Zhu, Z., Van Etten, J.L., and Xu, S.Y. (2004a). Cloning of CviPII nicking and modification system from chlorella virus NYs-1 and application of Nt.CviPII in random DNA amplification. Nucleic Acids Res *32*, 6187-6199.

Chan, S.H., Zhu, Z.Y., Van Etten, J.L., and Xu, S.Y. (2004b). Cloning of CviPII nicking and modification system from chlorella virus NYs-1 and application of Nt.CviPII in random DNA amplification. Nucleic Acids Research *32*, 6187-6199.

Chen, T.P., Hevi, S., Gay, F., Tsujimoto, N., He, T., Zhang, B.L., Ueda, Y., and Li, E. (2007). Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. Nature Genet. *39*, 391-396.

Chen, T.P., Ueda, Y., Dodge, J.E., Wang, Z.J., and Li, E. (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Molecular and Cellular Biology *23*, 5594-5605.

Chiurazzi, P., Pomponi, M.G., Pietrobono, R., Bakker, C.E., Neri, G., and Oostra, B.A. (1999). Synergistic effect of histone hyperacetylation and DNA demethylation in the reactivation of the FMR1 gene. Human Molecular Genetics. *8*, 2317-2323.

Choy, M.K., Movassagh, M., Goh, H.G., Bennett, M.R., Down, T.A., and Foo, R.S.Y. (2010). Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. BMC genomics *11*.

Chuang, L.S.H., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G.L., and Li, B.F.L. (1997). Human DNA (cytosine-5) methyltransferase PCNA complex as a target for p21(WAF1). Science *277*, 1996-2000.

Clapier, C.R., and Cairns, B.R. (2009). The Biology of Chromatin Remodeling Complexes. Annual Review of Biochemistry *78*, 273-304.

Clark, S.J., Harrison, J., Paul, C.L., and Frommer, M. (1994). High sensitivity mapping of methylated cytosines. Nucl.Acids Res. *22*, 2990-2997.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature *452*, 215-219.

Cortez, C.C., and Jones, P.A. (2008). Chromatin, cancer and drug therapies. Mutat. Res.-Fundam. Mol. Mech. Mutagen. *647*, 44-51.

Darst, R.P., Pardo, C.E., Ai, L., Brown, K.D., and Kladde, M.P. (2010). Bisulfite sequencing of DNA. Curr Protoc Mol Biol *Chapter 7*, Unit 7 9 1-17.

Datta, J., Ghoshal, K., Denny, W.A., Gamage, S.A., Brooke, D.G., Phiasivongsa, P., Redkar, S., and Jacob, S.T. (2009). A New Class of Quinoline-Based DNA Hypomethylating Agents Reactivates Tumor Suppressor Genes by Blocking DNA Methyltransferase 1 Activity and Inducing Its Degradation. Cancer Research *69*, 4277-4285.

Davey, C., Pennings, S., and Allan, J. (1997). CpG methylation remodels chromatin structure in vitro. Journal of Molecular Biology *267*, 276-288.

Davey, C.S., Pennings, S., Reilly, C., Meehan, R.R., and Allan, J. (2004). A determining influence for CpG dinucleotides on nucleosome positioning in vitro. Nucleic Acids Research *32*, 4322-4331.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev *25*, 1010-1022.

Dechassa, M.L., Sabri, A., Pondugula, S., Kassabov, S.R., Chatterjee, N., Kladde, M.P., and Bartholomew, B. (2010). SWI/SNF has intrinsic nucleosome disassembly activity that is dependent on adjacent nucleosomes. Mol Cell.

Dhasarathy, A., and Wade, P.A. (2008). The MBD protein family-Reading an epigenetic mark? Mutat. Res.-Fundam. Mol. Mech. Mutagen. *647*, 39-43.

Di Croce, L., Raker, V.A., Corsaro, M., Fazi, F., Fanelli, M., Faretta, M., Fuks, F., Lo Coco, F., Kouzarides, T., Nervi, C.*, et al.* (2002). Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor. Science *295*, 1079-1082.

Diala, E.S., and Hoffman, R.M. (1982). DNA methylation levels in normal and chemically-transformed mouse 3T3-cells. Biochemical and Biophysical Research Communications *104*, 1489-1494.

Diekmann, S. (1987). DNA METHYLATION CAN ENHANCE OR INDUCE DNA CURVATURE. EMBO Journal *6*, 4213-4217.

Dong, L., Wang, W., Wang, F., Stoner, M., Reed, J.C., Harigai, M., Samudio, I., Kladde, M.P., Vyhlidal, C., and Safe, S. (1999). Mechanisms of transcriptional activation of *bcl-2* gene expression by 17β-estradiol in breast cancer cells. The Journal of biological chemistry *274*, 32099-32107.

Duan, R., Porter, W., Samudio, I., Vyhlidal, C., Kladde, M., and Safe, S. (1999). Transcriptional activation of c-*fos* protooncogene by 17β-estradiol: mechanism of aryl hydrocarbon receptor-mediated inhibition. Molecular endocrinology (Baltimore, Md *13*, 1511-1521.

Ducasse, M., and Brown, M.A. (2006). Epigenetic aberrations and cancer. Molecular Cancer *5*.

Dumont, N., Crawford, Y.G., Sigaroudinia, M., Nagrani, S.S., Wilson, M.B., Buehring, G.C., Turashvili, G., Aparicio, S., Gauthier, M.L., Fordyce, C.A.*, et al.* (2009). Human mammary cancer progression model recapitulates methylation events associated with breast premalignancy. Breast Cancer Res. *11*.

Eden, A., Gaudet, F., Waghmare, A., and Jaenisch, R. (2003). Chromosomal instability and tumors promoted by DNA hypomethylation. Science *300*, 455-455.

El-Osta, A., Kantharidis, P., Zalcberg, J.R., and Wolffe, A.P. (2002). Precipitous release of methyl-CpG binding protein 2 and histone deacetylase 1 from the methylated human multidrug resistance gene (MDR1) on activation. Molecular and Cellular Biology *22*, 1844-1857.

Esteller, M. (2002). CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. Oncogene *21*, 5427-5440.

Esteller, M. (2007a). Cancer epigenomics: DNA methylomes and histone-modification maps. Nature Reviews Genetics *8*, 286-298.

Esteller, M. (2007b). Epigenetic gene silencing in cancer: the DNA hypermethylome. Human molecular genetics *16*, R50-R59.

Esteller, M. (2008). Molecular origins of cancer: Epigenetics in cancer. New England Journal of Medicine *358*, 1148-1159.

Esteller, M., Corn, P.G., Baylin, S.B., and Herman, J.G. (2001). A gene hypermethylation profile of human cancer. Cancer Research *61*, 3225-3229.

Esteller, M., and Herman, J.G. (2002). Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. Journal of Pathology *196*, 1-7.

Esteller, M., Levine, R., Baylin, S.B., Ellenson, L.H., and Herman, J.G. (1998). MLH1 promoter hypermethylation is associated with the microsatellite instability phenotype in sporadic endometrial carcinomas. Oncogene *17*, 2413-2417.

Fahrner, J.A., Eguchi, S., Herman, J.G., and Baylin, S.B. (2002). Dependence of histone modifications and gene expression on DNA hypermethylation in cancer. Cancer Research *62*, 7213-7218.

Fatemi, M., Pao, M.M., Jeong, S., Gal-Yam, E.N., Egger, G., Weisenberger, D.J., and Jones, P.A. (2005). Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. Nucleic Acids Res *33*, e176.

Fedor, M.J., Lue, N.F., and Kornberg, R.D. (1988). Statistical positioning of nucleosomes by specific protein- binding to an upstream activating sequence in yeast. J.Mol.Biol. *204*, 109-127.

Feinberg, A.P. (2008). Epigenetics at the epicenter of modern medicine. Jama-Journal of the American Medical Association *299*, 1345-1350.

Feinberg, A.P., Ohlsson, R., and Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. Nature Reviews Genetics *7*, 21-33.

Feinberg, A.P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature *301*, 89-92.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E.*, et al.* (2010). Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A *107*, 8689-8694.

Feng, T.Y., Tu, J., and Kuo, T.T. (1978). Characterization of deoxycytidylate methyltransferase in *Xanthomonas-oryzae* infected with bacteriophage-XP12. Eur. J. Biochem. *87*, 29-36.

Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I.K., Sharon, E., Lubling, Y., Widom, J., and Segal, E. (2008). Distinct Modes of Regulation by Chromatin Encoded through Nucleosome Positioning Signals. Plos Computational Biology *4*.

Fraga, M.F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K.*, et al.* (2005). Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nature Genet. *37*, 391-400.

Fraga, M.F., and Esteller, M. (2005). Towards the human cancer epigenome - A first draft of histone modifications. Cell Cycle *4*, 1377-1381.

Frixen, U.H., Behrens, J., Sachs, M., Eberle, G., Voss, B., Warda, A., Lochner, D., and Birchmeier, W. (1991). E-cadherin-mediated cell cell-adhesion prevents invasiveness of human carcinoma-cells. Journal of Cell Biology *113*, 173-185.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proceedings of the National Academy of Sciences of the United States of America *89*, 1827-1831.

Fuks, F. (2005). DNA methylation and histone modifications: teaming up to silence genes. Current opinion in genetics & development *15*, 490-495.

Fuks, F., Burgers, W.A., Godin, N., Kasai, M., and Kouzarides, T. (2001). Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. Embo Journal *20*, 2536-2544.

Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Research *31*, 2305-2312.

Gal-Yam, E.N., Jeong, S., Tanay, A., Egger, G., Lee, A.S., and Jones, P.A. (2006). Constitutive nucleosome depletion and ordered factor assembly at the GRP78 promoter revealed by single molecule footprinting. PLoS Genet *2*, e160.

Gardinergarden, M., and Frommer, M. (1987). CPG ISLANDS IN VERTEBRATE GENOMES. Journal of Molecular Biology *196*, 261-282.

Gardinergarden, M., and Frommer, M. (1994). Transcripts and cpg islands associated with the proopiomelanocortin gene and other neurally expressed genes. Journal of Molecular Endocrinology *12*, 365-382.

Gaudet, F., Hodgson, J.G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J.W., Leonhardt, H., and Jaenisch, R. (2003). Induction of tumors in mice by genomic hypomethylation. Science *300*, 489-492.

Georgel, P.T., Horowitz-Scherer, R.A., Adkins, N., Woodcock, C.L., Wade, P.A., and Hansen, J.C. (2003). Chromatin compaction by human MeCP2: assembly of novel secondary chromatin structures in the absence of DNA methylation. J.Biol.Chem. *278*, Jun 4 Epub.

Ghoshal, K., Datta, J., Majumder, S., Bai, S.M., Kutay, H., Motiwala, T., and Jacob, S.T. (2005). 5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, and nuclear localization signal. Molecular and Cellular Biology *25*, 4727-4741.

Goelz, S.E., Vogelstein, B., Hamilton, S.R., and Feinberg, A.P. (1985). Hypomethylation of DNA from benign and malignant human-colon neoplasms. Science *228*, 187-190.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annual Review of Biochemistry *74*, 481-514.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X.Y., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNA(AsP) by the DNA methyltransferase homolog Dnmt2. Science *311*, 395-398.

Gottschling, D.E. (1992). Telomere-proximal DNA in *Saccharomyces cerevisiae* is refractory to methyltransferase activity in vivo Proc Natl Acad Sci USA *89*, 4062-4065.

Goyal, R., Reinhardt, R., and Jeltsch, A. (2006). Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. Nucleic Acids Research *34*, 1182-1188.

Grady, W.M., Willis, J., Guilford, P.J., Dunbier, A.K., Toro, T.T., Lynch, H., Wiesner, G., Ferguson, K., Eng, C., Park, J.G.*, et al.* (2000). Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. Nature Genet. *26*, 16-17.

Graff, J.R., Herman, J.G., Lapidus, R.G., Chopra, H., Xu, R., Jarrard, D.F., Isaacs, W.B., Pitha, P.M., Davidson, N.E., and Baylin, S.B. (1995). E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas. Cancer Research *55*, 5195-5199.

Grandjean, V., Yaman, R., Cuzin, F., and Rassoulzadegan, M. (2007). Inheritance of an Epigenetic Mark: The CpG DNA Methyltransferase 1 Is Required for De Novo Establishment of a Complex Pattern of Non-CpG Methylation. Plos One *2*.

Grossman, L. (1981). Enzymes Involved in the Repair of Damaged DNA. Archives of Biochemistry and Biophysics *211*, 511-522.

Gruenbaum, Y., Stein, R., Cedar, H., and Razin, A. (1981). Methylation of CpG sequences in eukaryotic DNA. Febs Letters *124*, 67-71.

Grunau, C., Schattevoy, R., Mache, N., and Rosenthal, A. (2000). MethTools - a toolbox to visualize and analyze DNA methylation data. Nucleic Acids Research *28*, 1053-1058.

Hagerman, P. (1990). Pyrimidine 5-methyl groups influence the magnitude of DNA curvature. Biochem. *29*, 1980-1983.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57-70.

Harris, T.J.R., and McCormick, F. (2010). The molecular pathology of cancer. Nature Reviews Clinical Oncology *7*, 251-265.

Hatziapostolou, M., and Iliopoulos, D. (2011). Epigenetic aberrations during oncogenesis. Cellular and Molecular Life Sciences *68*, 1681-1702.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S.*, et al.* (2010). Distinct Epigenomic Landscapes of Pluripotent and Lineage-Committed Human Cells. Cell Stem Cell *6*, 479-491.

Hayatsu, H. (1976). Bisulfite modification of nucleic acids and their constituents. Prog.Nucl.Acid Res. *16*, 75-124.

Hayatsu, H., and Shiragami, M. (1979). Reaction of bisulfite with the 5-hydroxymethyl group in pyrimidines and in phage DNAS. Biochemistry *18*, 632-637.

Hayatsu, H., Shiraishi, M., and Negishi, K. (2008). Bisulfite modification for analysis of DNA methylation. Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al *Chapter 6*, Unit 6 10.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y.T., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C.X., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genet. *39*, 311-318.

Hellman, A., and Chess, A. (2007). Gene body-specific methylation on the active X chromosome. Science *315*, 1141-1143.

Henderson, I.R., and Jacobsen, S.E. (2007). Epigenetic inheritance in plants. Nature *447*, 418-424.

Herman, J.G., and Baylin, S.B. (2003). Mechanisms of disease: Gene silencing in cancer in association with promoter hypermethylation. New England Journal of Medicine *349*, 2042-2054.

Hermann, A., Gowher, H., and Jeltsch, A. (2004a). Biochemistry and biology of mammalian DNA methyltransferases. Cellular and Molecular Life Sciences *61*, 2571-2587.

Hermann, A., Goyal, R., and Jeltsch, A. (2004b). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. Journal of Biological Chemistry *279*, 48350-48359.

Hetzl, J., Foerster, A.M., Raidl, G., and Scheid, O.M. (2007). CyMATE: a new tool for methylation analysis of plant genornic DNA after bisulphite sequencing. Plant Journal *51*, 526-536.

Hinshelwood, R.A., Melki, J.R., Huschtscha, L.I., Paul, C., Song, J.Z., Stirzaker, C., Reddel, R.R., and Clark, S.J. (2009). Aberrant de novo methylation of the p16(INK4A) CpG island is initiated post gene silencing in association with chromatin remodelling and mimics nucleosome positioning. Human molecular genetics *18*, 3098-3109.

Hodges-Garcia, Y., and Hagerman, P.J. (1992). Cytosine methylation can induce local distortions in the structure of duplex DNA. Biochem. *31*, 7595-7599.

Holz-Schietinger, C., and Reich, N.O. (2010). The Inherent Processivity of the Human de Novo Methyltransferase 3A (DNMT3A) Is Enhanced by DNMT3L. Journal of Biological Chemistry *285*, 29091-29100.

Hoose, S.A., and Kladde, M.P. (2006). DNA methyltransferase probing of DNA-protein interactions. Methods Mol Biol *338*, 225-244.

Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., and Eden, A. (2008). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene *27*, 404-408.

Huang, Y., Pastor, W.A., Shen, Y.H., Tahiliani, M., Liu, D.R., and Rao, A. (2010). The Behaviour of 5-Hydroxymethylcytosine in Bisulfite Sequencing. Plos One *5*.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands - 'A rough guide'. Febs Letters *583*, 1713-1720.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z.J., Montano, C., Onyango, P., Cui, H.M., Gabo, K., Rongione, M., Webster, M.*, et al.* (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nature Genet. *41*, 178-186.

Iskow, R.C., McCabe, M.T., Mills, R.E., Torene, S., Pittard, W.S., Neuwald, A.F., Van Meir, E.G., Vertino, P.M., and Devine, S.E. (2010). Natural Mutagenesis of Human Genomes by Endogenous Retrotransposons. Cell *141*, 1253-U1268.

Issa, J.P., and Kantarjian, H.M. (2009). Targeting DNA methylation.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129-U1151.

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

Jeltsch, A. (2006). On the Enzymatic Properties of Dnmt1 Specificity, Processivity, Mechanism of Linear Diffusion and Allosteric Regulation of the Enzyme. Epigenetics *1*, 63-66.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jessen, W.J., Dhasarathy, A., Hoose, S.A., Carvin, C.D., Risinger, A.L., and Kladde, M.P. (2004). Mapping chromatin structure *in vivo* using DNA methyltransferases. Methods *33*, 68-80.

Jessen, W.J., Hoose, S.A., Kilgore, J.A., and Kladde, M.P. (2006). Active *PHO5* chromatin encompasses variable numbers of nucleosomes at individual promoters. Nat Struct Mol Biol *13*, 256-263.

Jiang, C., and Pugh, B.F. (2009a). A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome. Genome Biol *10*, R109.

Jiang, C., and Pugh, B.F. (2009b). Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet *10*, 161-172.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941-945.

Johnson, L.M., Cao, X.F., and Jacobsen, S.E. (2002). Interplay between two epigenetic marks: DNA methylation and histone H3 lysine 9 methylation. Current Biology *12*, 1360-1367.

Jones, P.A., and Baylin, S.B. (2002). The fundamental role of epigenetic events in cancer. Nat.Rev.Genet. *3*, 415-428.

Jones, P.A., and Baylin, S.B. (2007). The epigenomics of cancer. Cell *128*, 683-692.

Jones, P.A., and Liang, G.N. (2009). OPINION Rethinking how DNA methylation patterns are maintained. Nature Reviews Genetics *10*, 805-811.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat.Genet. *19*, 187-191.

Kagey, J.D., Kapoor-Vazirani, P., McCabe, M.T., Powell, D.R., and Vertino, P.M. (2010). Long-term Stability of Demethylation after Transient Exposure to 5-Aza-2 '-Deoxycytidine Correlates with Sustained RNA Polymerase II Occupancy. Molecular Cancer Research *8*, 1048-1059.

Kaplan, N., Moore, I., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E. (2010). Nucleosome sequence preferences influence in vivo nucleosome organization. Nat Struct Mol Biol *17*, 918-920.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J.*, et al.* (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-U129.

Keogh, M.C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J.*, et al.* (2005). Cotranscriptional Set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell *123*, 593-605.

Kilgore, J.A., Hoose, S.A., Gustafson, T.L., Porter, W., and Kladde, M.P. (2007). Single-molecule and population probing of chromatin structure using DNA methyltransferases. Methods *41*, 320-332.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C.X., Singer, M.A., Richmond, T.A., Wu, Y.N., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. Nature *436*, 876-880.

Kladde, M.P., and Simpson, R.T. (1994). Positioned nucleosomes inhibit Dam methylation in vivo Proc Natl Acad Sci USA *91*, 1361-1365.

Kladde, M.P., and Simpson, R.T. (1996). Chromatin structure mapping in vivo using methyltransferases. Methods Enzymol *274*, 214-233.

Kladde, M.P., Xu, M., and Simpson, R.T. (1996). Direct study of DNA-protein interactions in repressed and active chromatin in living cells. EMBO J. *15*, 6290-6300.

Kladde, M.P., Xu, M., and Simpson, R.T. (1999). DNA methyltransferases as probes of chromatin structure in vivo. Methods Enzymol. *304*, 431-447.

Kleinjan, D.A., Seawright, A., Childs, A.J., and van Heyningen, V. (2004). Conserved elements in Pax6 intron 7 involved in (auto)regulation and alternative transcription. Developmental Biology *265*, 462-477.

Klimasauskas, S., Gerasimaite, R., Vilkaitis, G., and Kulakauskas, S. (2002). N4,5-dimethylcytosine, a novel hypermodified base in DNA. Nucleic Acids Res *2 (Supple.),*, 73-74.

Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends in biochemical sciences *31*, 89-97.

Klose, R.J., and Yi, Z. (2007). Regulation of histone methylation by demethylimination and demethylation. Nat. Rev. Mol. Cell Biol. *8*, 307-318.

Knoepfler, P.S., Bergstrom, D.A., Uetsuki, T., Dac-Korytko, I., Sun, Y.H., Wright, W.E., Tapscott, S.J., and Kamps, M.P. (1999). A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx- Meis1/Prep1. NAR *27*, 3752-3761.

Komashko, V.M., Acevedo, L.G., Squazzo, S.L., Iyengar, S.S., Rabinovich, A., O'Geen, H., Green, R., and Farnham, P.J. (2008). Using ChIP-chip technology to reveal common principles of transcriptional repression in normal and cancer cells. Genome research *18*, 521-532.

Kondo, Y., and Issa, J.P.J. (2010). DNA methylation profiling in cancer. Expert Reviews in Molecular Medicine *12*.

Kondo, Y., Shen, L., Cheng, A.S., Ahmed, S., Boumber, Y., Charo, C., Yamochi, T., Urano, T., Furukawa, K., Kwabi-Addo, B*., et al.* (2008). Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation. Nature Genet. *40*, 741-750.

Korber, P., Luckenbach, T., Blaschke, D., and Horz, W. (2004). Evidence for histone eviction in *trans* upon induction of the yeast *PHO5* promoter. Mol Cell Biol *24*, 10965-10974.

Kouidou, S., Agidou, T., Kyrkou, A., Andreou, A., Katopodi, T., Georgiou, E., Krikelis, D., Dimitriadou, A., Spanos, P., Tsilikas, C.*, et al.* (2005). Non-CpG cytosine methylation of p53 exon 5 in non-small cell lung carcinoma. Lung cancer (Amsterdam, Netherlands) *50*, 299-307.

Kouzarides, T. (1999). Histone acetylases and deacetylases in cell proliferation. Curr.Opin.Genet.Dev. *9*, 40-48.

Kouzarides, T. (2000). Acetylation: a regulatory modification to rival phosphorylation? EMBO J. *19*, 1176-1179.

Kouzarides, T. (2002). Histone methylation in transcriptional control. Current opinion in genetics & development *12*, 198-209.

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

Kriaucionis, S., and Heintz, N. (2009). The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. Science *324*, 929-930.

Kwak, H.I., Gustafson, T., Metz, R.P., Laffin, B., Schedin, P., and Porter, W.W. (2007). Inhibition of breast cancer growth and invasion by single-minded 2s. Carcinogenesis *28*, 259-266.

Lai, A.Y., Fatemi, M., Dhasarathy, A., Malone, C., Sobol, S.E., Geigerman, C., Jaye, D.L., Mav, D., Shah, R., Li, L.*, et al.* (2010). DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas. The Journal of experimental medicine *207*, 1939-1950.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Langst, G., and Becker, P.B. (2004). Nucleosome remodeling: one mechanism, many phenomena? Biochimica et biophysica acta *1677*, 58-63.

Latham, T., Gilbert, N., and Ramsahoye, B. (2008). DNA methylation in mouse embryonic stem cells and development. Cell and Tissue Research *331*, 31-55.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J.*, et al.* (2010a). Dynamic changes in the human methylome during differentiation. Genome research *20*, 320-331.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Sung, K.W.K., Rigoutsos, I., Loring, J.*, et al.* (2010b). Dynamic changes in the human methylome during differentiation. Genome research *20*, 320-331.

Lee, C.K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. Nature Genet. *36*, 900-905.

Lee, J.S., Lee, C.H., and Chung, J.H. (1998). Studying the recruitment of Sp1 to the beta-globin promoter with an in vivo method: protein position identification with nuclease tail (PIN*POINT). Proc.Natl.Acad.Sci.U.S.A *95*, 969-974.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high- resolution atlas of nucleosome occupancy in yeast. Nature Genet. *39*, 1235-1244.

Lehnertz, B., Ueda, Y., Derijck, A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T.P., Li, E., Jenuwein, T., and Peters, A. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Current Biology *13*, 1192-1200.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707-719.

Li, B.Y., Adams, C.C., and Workman, J.L. (1994). Nucleosome binding by the constitutive transcription factor Sp1. J.Biol.Chem. *269*, 7756-7763.

Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. Nature Reviews Genetics *3*, 662-673.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.

Li, G., and Widom, J. (2004). Nucleosomes facilitate their own invasion. Nat Struct Mol Biol *11*, 763-769.

Li, G.H., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. Current opinion in genetics & development *21*, 175-186.

Li, L.C., and Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. Bioinformatics (Oxford, England) *18*, 1427-1431.

Liang, G.G., Chan, M.F., Tomigahara, Y., Tsai, Y.C., Gonzales, F.A., Li, E., Laird, P.W., and Jones, P.A. (2002). Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Molecular and Cellular Biology *22*, 480-491.

Lieb, J.D., and Clarke, N.D. (2005). Control of transcription through intragenic patterns of nucleosome composition. Cell *123*, 1187-1190.

Lima, S.C., Hernandez-Vargas, H., and Herceg, Z. (2010). Epigenetic signatures in cancer: Implications for the control of cancer in the clinic. Current Opinion in Molecular Therapeutics *12*, 316-324.

Lin, J.C., Jeong, S., Liang, G., Takai, D., Fatemi, M., Tsai, Y.C., Egger, G., Gal-Yam, E.N., and Jones, P.A. (2007). Role of nucleosomal occupancy in the epigenetic silencing of the MLH1 CpG island. Cancer Cell *12*, 432-444.

Lin, T., Ponn, A., Hu, X., Law, B.K., and Lu, J. (2010). Requirement of the histone demethylase LSD1 in Snail-mediated transcriptional repression during epithelial-mesenchymal transition. Oncogene *29*, 4896-4904.

Liotta, L., and Petricoin, E. (2000). Molecular profiling of human cancer. Nature Reviews Genetics *1*, 48-56.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M.*, et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315-322.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997a). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997b). Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. Nature *389*, 251-260.

Lund, A.H., and van Lohuizen, M. (2004). Epigenetics and cancer. Genes & Development *18*, 2315-2335.

Macleod, D., Ali, R.R., and Bird, A. (1998). An alternative promoter in the mouse major histocompatibility complex class II I-A beta gene: Implications for the origin of CpG islands. Molecular and Cellular Biology *18*, 4433-4443.

Markus, J., Garin, M.T., Bies, J., Galili, N., Raza, A., Thirman, M.J., Le Beau, M.M., Rowley, J.D., Liu, P.P., and Wolff, L. (2007). Methylation-independent silencing of the tumor suppressor INK4b (p15) by CBF beta-SMMHC in acute myelogenous leukemia with inv(16). Cancer Research *67*, 992-1000.

Masson, D., Rioux-Leclercq, N., Fergelot, P., Jouan, F., Mottier, S., Theoleyre, S., Bach-Ngohou, K., Patard, J.J., and Denis, M.G. (2010). Loss of expression of TIMP3 in clear cell renal cell carcinoma. European Journal of Cancer *46*, 1430-1437.

Matsuo, K., Silke, J., Gramatikoff, K., and Schaffner, W. (1994). The CpG-specific methylase SssI has topoisomerase activity in the presence of Mg2+. Nucleic.Acids.Res. *22*, 5354-5359.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C*., et al.* (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358-362.

McCabe, M.T., Brandes, J.C., and Vertino, P.M. (2009). Cancer DNA Methylation: Molecular Mechanisms and Clinical Implications. Clinical Cancer Research *15*, 3927-3937.

Metz, R.P., Kwak, H.I., Gustafson, T., Laffin, B., and Porter, W.W. (2006). Differential transcriptional regulation by mouse single-minded 2s. Journal of Biological Chemistry *281*, 10839-10848.

Miele, V., Vaillant, C., d'Aubenton-Carafa, Y., Thermes, C., and Grange, T. (2008). DNA physical properties determine nucleosome occupancy from yeast to fly. Nucleic Acids Research *36*, 3746-3756.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nature Genet. *37*, 1090-1097.

Nabilsi, N.H., Broaddus, R.R., and Loose, D.S. (2009). DNA methylation inhibits p53-mediated survivin repression. Oncogene *28*, 2046-2050.

Nan, X., Meehan, R.R., and Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic.Acids.Res. *21*, 4886-4892.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386-389.

Nan, X.S., Campoy, F.J., and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell *88*, 471-481.

Nathan, D., and Crothers, D.M. (2002). Bending and flexibility of methylated and unmethylated EcoRI DNA. Journal of Molecular Biology *316*, 7-17.

Nelson, M., Burbank, D.E., and Van Etten, J.L. (1998). Chlorella viruses encode multiple DNA methyltransferases. Biol.Chem. *379*, 423-428.

Nelson, M., Zhang, Y., and Van Etten, J.L. (1993). DNA methyltransferases and DNA site-specific endonucleases encoded by chlorella viruses. EXS. *64*, 186-211.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell *11*, 709-719.

Nguyen, C., Liang, G.M., Nguyen, T.T., Tsao-Wei, D., Groshen, S., Lubbert, M., Zhou, J.H., Benedict, W.F., and Jones, P.A. (2001). Susceptibility of nonpromoter CpG islands to de novo methylation in normal and neoplastic cells. Journal of the National Cancer Institute *93*, 1465-1472.

Nur, I., Szyf, M., Razin, A., Glaser, G., Rottem, S., and Razin, S. (1985). Procaryotic and eukaryotic traits of dna methylation in *Spiroplasmas* (Mycoplasmas). Journal of Bacteriology *164*, 19-24.

Ogawa, O., Eccles, M.R., Szeto, J., McNoe, L.A., Yun, K., Maw, M.A., Smith, P.J., and Reeve, A.E. (1993). Relaxation of insulin-like growth factor-II gene imprinting implicated in wilms-tumor. Nature *362*, 749-751.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Oki, Y., Aoki, E., and Issa, J.P.J. (2007). Decitabine - Bedside to bench. Critical Reviews in Oncology Hematology *61*, 140-152.

Oki, Y., Jelinek, J., Shen, L., Kantarjian, H.M., and Issa, J.P.J. (2008). Induction of hypomethylation and molecular response after decitabine therapy in patients with chronic myelomonocytic leukemia. Blood *111*, 2382-2384.

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K.Q., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D*., et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-U713.

Ordway, J.M., Bedell, J.A., Citek, R.W., Nunberg, A.N., and Jeddeloh, J.A. (2005). MethylMapper: a method for high-throughput, multilocus bisulfite sequence analysis and reporting. Biotechniques *39*, 464-+.

Ou, J.N., Torrisani, J., Unterberger, A., Provencal, N., Shikimi, K., Karimi, M., Ekstrom, T.J., and Szyf, M. (2007). Histone deacetylase inhibitor Trichostatin A induces global and gene-specific DNA demethylation in human cancer cell lines. Biochemical Pharmacology *73*, 1297-1307.

Ozsolak, F., Song, J.S., Liu, X.S., and Fisher, D.E. (2007). High-throughput mapping of the chromatin structure of human promoters. Nature Biotechnology *25*, 244-248.

Palacios, D., Summerbell, D., Rigby, P.W.J., and Boyes, J. (2010). Interplay between DNA Methylation and Transcription Factor Availability: Implications for Developmental Activation of the Mouse Myogenin Gene. Molecular and Cellular Biology *30*, 3805-3815.

Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes & Development *10*, 1991-2002.

Pardo, C., Hoose, S.A., Pondugula, S., and Kladde, M.P. (2009). DNA methyltransferase probing of chromatin structure within populations and on single molecules. Methods Mol Biol *523*, 41-65.

Pardo, C.E., Carr, I.M., Hoffman, C.J., Darst, R.P., Markham, A.F., Bonthron, D.T., and Kladde, M.P. (2010). MethylViewer: computational analysis and editing for bisulfite sequencing and methyltransferase accessibility protocol for individual templates (MAPit) projects. Nucleic Acids Res *39*, e5.

Pardo, C.E., Darst, R.P., Nabilsi, N.H., Delmas, A.L., Kladde, M.P. (2011). Simultaneous Single-Molecule Mapping of Protein-DNA Interactions and DNA Methylation by MAPit, Vol 95 (Jonh Wiley & Sons, Ltd).

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of Histone H3 Lysine 27 Methylation in X Inactivation. Science *300*, 131-135.

Polach, K.J., and Widom, J. (1995). Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. J Mol Biol *254*, 130-149.

Polach, K.J., and Widom, J. (1996). A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. J Mol Biol *258*, 800-812.

Pondugula, S., and Kladde, M.P. (2008). Single-molecule analysis of chromatin: Changing the view of genomes one molecule at a time. Journal of cellular biochemistry *105*, 330-337.

Pondugula, S., Neef, D.W., Voth, W.P., Darst, R.P., Dhasarathy, A., Reynolds, M.M., Takahata, S., Stillman, D.J., and Kladde, M.P. (2009). Coupling phosphate homeostasis to cell cycle-specific transcription: mitotic activation of Saccharomyces cerevisiae PHO5 by Mcm1 and Forkhead proteins. Mol Cell Biol *29*, 4891-4905.

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. Nature Biotechnology *28*, 1057-1068.

Pradhan, S., Bacolla, A., Wells, R.D., and Roberts, R.J. (1999). Recombinant Human DNA (Cytosine-5) Methyltransferase. I. expression, purification, and comparison of de novo and maintenance methylation. J.Biol.Chem. *274*, 33002-33010.

Pratt, K., and Hattman, S. (1981). Deoxyribonucleic acid methylation and chromatin organization in Tetrahymena thermophila. Mol.Cell Biol *1*, 600-608.

Pratt, K., and Hattman, S. (1983). Nucleosome phasing in Tetrahymena macronuclei. J.Protozool. *30*, 592-598.

Qi, J.H., Ebrahem, Q., Moore, N., Murphy, G., Claesson-Welsh, L., Bond, M., Baker, A., and Anand-Apte, B. (2003). A novel function for tissue inhibitor of metalloproteinases-3 (TIMP3): inhibition of angiogenesis by blockage of VEGF binding to VEGF receptor-2. Nature Medicine *9*, 407-415.

Radman-Livaja, M., and Rando, O.J. (2010). Nucleosome positioning: How is it established, and why does it matter? Developmental Biology *339*, 258-266.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc.Natl.Acad.Sci.U.S.A *97*, 5237-5242.

Rauch, T.A., Wu, X.W., Zhong, X., Riggs, A.D., and Pfeifer, G.P. (2009). A human B cell methylome at 100-base pair resolution. Proceedings of the National Academy of Sciences of the United States of America *106*, 671-678.

Renbaum, P., Abrahamove, D., Fainsod, A., Wilson, G., Rottem, S., and Razin, A. (1990). Cloning, characterization, and expression in *Escherichia coli* of the gene coding for the CpG DNA from *Spiroplasma* sp strain MQ-1 (M. Sss I). Nucleic Acids Res *18*, 1145-1152.

Renbaum, P., and Razin, A. (1992). Mode of action of the Spiroplasma CpG methylase M.SssI. FEBS Lett *313*, 243-247.

Rhee, I., Jair, K.W., Yen, R.W.C., Lengauer, C., Herman, J.G., Kinzler, K.W., Vogelstein, B., Baylin, S.B., and Schuebel, K.E. (2000). CpG methylation is maintained in human cancer cells lacking DNMT1. Nature *404*, 1003-1007.

Richard-Foy, H., and Hager, G.L. (1987). Sequence specific positioning of nucleosomes over the steroid inducible MMTV promoter. EMBO.J. *6*, 2321-2328.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E.*, et al.* (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by Noncoding RNAs. Cell *129*, 1311-1323.

Robertson, K.D. (2001). DNA methylation, methyltransferases, and cancer. Oncogene *20*, 3139-3155.

Robertson, K.D. (2005). DNA methylation and human disease. Nat Rev Genet *6*, 597-610.

Robertson, K.D., Keyomarsi, K., Gonzales, F.A., Velicescu, M., and Jones, P.A. (2000). Differential mRNA expression of the human DNA methyltransferases (DNMTs) 1,3a and 3b during the G(0)/G(1) to S phase transition in normal and tumor cells. Nucleic Acids Research *28*, 2108-2113.

Robertson, K.D., and Wolffe, A.P. (2000). DNA methylation in health and disease. Nature Reviews Genetics *1*, 11-19.

Roca, J., Gartenberg, M.R., Oshima, Y., and Wang, J.C. (1992). A hit-and-run system for targeted genetic manipulations in yeast. Nucl.Acids Res. *20*, 4671-4672.

Rodriguez-Paredes, M., and Esteller, M. (2011). Cancer epigenetics reaches mainstream oncology. Nature Medicine *17*, 330-339.

Rohde, C., Zhang, Y., Jurkowski, T.P., Stamerjohanns, H., Reinhardt, R., and Jeltsch, A. (2008). Bisulfite sequencing Data Presentation and Compilation (BDPC) web servera useful tool for DNA methylation analysis. Nucleic Acids Research *36*.

Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. Nat Rev Mol Cell Biol *7*, 437-447.

Samudio, I., Vyhlidal, C., Wang, F., Stoner, M., Chen, I., Kladde, M., Barhoumi, R., Burghardt, R., and Safe, S. (2001). Transcriptional activation of deoxyribonucleic acid polymerase alpha gene expression in MCF-7 cells by 17 beta-estradiol. Endocrinology *142*, 1000-1008.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Satchwell, S.C., Drew, H.R., and Travers, A.A. (1986). SEQUENCE PERIODICITIES IN CHICKEN NUCLEOSOME CORE DNA. Journal of Molecular Biology *191*, 659-675.

Sawan, C., Vaissiere, T., Murr, R., and Herceg, Z. (2008). Epigenetic drivers and genetic passengers on the road to cancer. Mutat. Res.-Fundam. Mol. Mech. Mutagen. *642*, 1-13.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proceedings of the National Academy of Sciences of the United States of America *103*, 1412-1417.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.

Schreiber, S.L., and Bernstein, B.E. (2002). Signaling network model of chromatin. Cell *111*, 771-778.

Segal, E., Fondufe-Mittendorf, Y., Chen, L.Y., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P.Z., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772-778.

Segal, E., and Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. Current Opinion in Structural Biology *19*, 65-71.

Selker, E.U. (1990). Premeiotic instability of repeated sequences in *Neurospora-crassa*. Annual review of genetics *24*, 579-613.

Sharif, J., Muto, M., Takebayashi, S.I., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K.*, et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908-U925.

Sharma, S., Kelly, T.K., and Jones, P.A. (2010). Epigenetics in cancer. Carcinogenesis *31*, 27-36.

Shen, L.L., Kantarjian, H., Guo, Y., Lin, E., Shan, J.Q., Huang, X.L., Berry, D., Ahmed, S., Zhu, W., Pierce, S.*, et al.* (2010). DNA Methylation Predicts Survival and Response to Therapy in Patients With Myelodysplastic Syndromes. Journal of Clinical Oncology *28*, 605-613.

Shiraishi, M., Sekiguchi, A., Chuu, Y.H., and Sekiya, T. (1999). Tight interaction between densely methylated DNA fragments and the methyl-CpG binding domain of the rat MeCP2 protein attached to a solid support [In Process Citation]. Biol.Chem. *380*, 1127-1131.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V.R. (2008). Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. PLoS Biol *6*, e65.

Si, J., Boumber, Y.A., Shu, J., Qin, T., Ahmed, S., He, R., Jelinek, J., and Issa, J.P.J. (2010). Chromatin Remodeling Is Required for Gene Reactivation after Decitabine-Mediated DNA Hypomethylation. Cancer Research *70*, 6968-6977.

Sincic, N., and Herceg, Z. (2011). DNA methylation and cancer: ghosts and angels above the genes. Curr. Opin. Oncol. *23*, 69-76.

Singal, R., and Grimes, S.R. (2001). Microsoft (R) Word (TM) Macro for analysis of cytosine methylation by the bisulfite deamination reaction. Biotechniques *30*, 116-120.

Singh, J., and Klar, A.J.S. (1992). Active genes in yeast display enhanced *in vivo* accessibility to foreign DNA methylases: a novel *in vivo* probe for chromatin structure of yeast. Genes Dev. *6*, 186-196.

Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature *415*, 810-813.

Soppe, W.J.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. Embo Journal *21*, 6549-6559.

Stancheva, I. (2005). Caught in conspiracy: cooperation between DNA methylation and histone H3K9 methylation in the establishment and maintenance of heterochromatin. Biochem. Cell Biol. *83*, 385-395.

Stewart, D.J., Issa, J.P., Kurzrock, R., Nunez, M.I., Jelinek, J., Hong, D., Oki, Y., Guo, Z., Gupta, S., and Wistuba, II (2009). Decitabine Effect on Tumor Global DNA Methylation and Other Parameters in a Phase I Trial in Refractory Solid Tumors and Lymphomas. Clinical Cancer Research *15*, 3881-3888.

Stewart, F.J., Panne, D., Bickle, T.A., and Raleigh, E.A. (2000). Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. J Mol Biol *298*, 611-622.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. Nature *458*, 719-724.

Suzuki, H., Gabrielson, E., Chen, W., Anbazhagan, R., van Engeland, M., Weijenberg, M.P., Herman, J.G., and Baylin, S.B. (2002). A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. Nature Genet. *31*, 141-149.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nature Reviews Genetics *9*, 465-476.

Taby, R., and Issa, J.P.J. (2010). Cancer Epigenetics. Ca-a Cancer Journal for Clinicians *60*, 376-392.

Tahiliani, M., Koh, K.P., Shen, Y.H., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L.*, et al.* (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science *324*, 930-935.

Takai, D., and Jones, P.A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proceedings of the National Academy of Sciences of the United States of America *99*, 3740-3745.

Tamaru, H., Zhang, X., McMillen, D., Singh, P.B., Nakayama, J., Grewal, S.I., Allis, C.D., Cheng, X.D., and Selker, E.U. (2003). Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in Neurospora crassa. Nature Genet. *34*, 75-79.

Tarakhovsky, A. (2010). Tools and landscapes of epigenetics. Nature Immunology *11*, 565-568.

Tillo, D., Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Field, Y., Lieb, J.D., Widom, J., Segal, E., and Hughes, T.R. (2010). High Nucleosome Occupancy Is Encoded at Human Regulatory Sequences. Plos One *5*.

Ting, A.H., McGarvey, K.M., and Baylin, S.B. (2006). The cancer epigenome - components and functional correlates. Genes & Development *20*, 3215-3231.

Ushijima, T., and Asada, K. (2010). Aberrant DNA methylation in contrast with mutations. Cancer Science *101*, 300-305.

Ushijima, T., Watanabe, N., Okochi, E., Kaneda, A., Sugimura, T., and Miyamoto, K. (2003). Fidelity of the methylation pattern and its variation in the genome. Genome research *13*, 868-874.

Vaissiere, T., Sawan, C., and Herceg, Z. (2008). Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutation Research-Reviews in Mutation Research *659*, 40-48.

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J.A., Costa, G., McKernan, K.*, et al.* (2008). A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome research *18*, 1051-1063.

Veigl, M.L., Kasturi, L., Olechnowicz, J., Ma, A.H., Lutterbaugh, J.D., Periyasamy, S., Li, G.M., Drummond, J., Modrich, P.L., Sedwick, W.D.*, et al.* (1998). Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. Proceedings of the National Academy of Sciences of the United States of America *95*, 8698-8702.

Vilkaitis, G., and Klimasauskas, S. (1999). Bisulfite sequencing protocol displays both 5-methylcytosine and N-4-methylcytosine. Analytical biochemistry *271*, 116-119.

Vilkaitis, G., Suetake, I., Klimasauskas, S., and Tajima, S. (2005). Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. The Journal of biological chemistry *280*, 64-72.

Vogel, K., Horz, W., and Hinnen, A. (1989). The two positively acting regulatory proteins PHO2 and PHO4 physically interact with PHO5 upstream activation regions. Mol.Cell Biol. *9*, 2050-2057.

Vogelauer, M., Wu, J., Suka, N., and Grunstein, M. (2000). Global histone acetylation and deacetylation in yeast. Nature *408*, 495-498.

Vogelstein, B., and Kinzler, K.W. (2004). Cancer genes and the pathways they control. Nature Medicine *10*, 789-799.

Vyhlidal, C., Samudio, I., Kladde, M.P., and Safe, S. (2000). Transcriptional activation of transforming growth factor α by estradiol: requirement for both a GC-rich site and an estrogen response element half-site. J Mol Endocrinol *24*, 329-338.

Wade, P.A. (2001). Methyl CpG binding proteins: coupling chromatin architecture to gene regulation. Oncogene *20*, 3166-3173.

Wang, Y., and Leung, F.C.C. (2004). An evaluation of new criteria for CpG islands in the human genome as gene markers. Bioinformatics (Oxford, England) *20*, 1170-1177.

Warnecke, P.M., Stirzaker, C., Song, J., Grunau, C., Melki, J.R., and Clark, S.J. (2002). Identification and resolution of artifacts in bisulfite sequencing. Methods *27*, 101-107.

Watt, P.M., Kumar, R., and Kees, U.R. (2000). Promoter demethylation accompanies reactivation of the HOXII proto-oncogene in leukemia. Genes Chromosomes & Cancer *29*, 371-377.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature Genet. *39*, 457-466.

Widom, J. (2001). Role of DNA sequence in nucleosome stability and dynamics. Quarterly Reviews of Biophysics *34*, 269-324.

Wilson, G.G., and Murray, N.E. (1991). Restriction and modification systems. Annual review of genetics *25*, 585-627.

Wong, C.M., Wong, C.C.L., Ng, Y.L., Au, S.L.K., Ko, F.C.F., and Ng, I.O.L. (2011). Transcriptional Repressive H3K9 and H3K27 Methylations Contribute to DNMT1-Mediated DNA Methylation Recovery. Plos One *6*.

Wu, H., Coskun, V., Tao, J.F., Xie, W., Ge, W.H., Yoshikawa, K., Li, E., Zhang, Y., and Sun, Y.E. (2010). Dnmt3a-Dependent Nonpromoter DNA Methylation Facilitates Transcription of Neurogenic Genes. Science *329*, 444-448.

Wu, J.C., and Santi, D.V. (1987). Kinetic and catalytic mechanism of HhaI methyltransferase. Journal of Biological Chemistry *262*, 4778-4786.

Xu, G.L., and Bestor, T.H. (1997). Cytosine methylation targetted to pre-determined sequences. Nat.Genet. *17*, 376-378.

Xu, M., Kladde, M.P., Van Etten, J.L., and Simpson, R.T. (1998a). Cloning, characterization and expression of the gene coding for a cytosine-5-DNA methyltransferase recognizing GpC. Nucleic Acids Research *26*, 3961-3966.

Xu, M., Simpson, R.T., and Kladde, M.P. (1998b). Gal4p-mediated chromatin remodeling depends on binding site position in nucleosomes but does not require DNA replication. Molecular and Cellular Biology *18*, 1201-1212.

Xu, Y.H., Manoharan, H.T., and Pitot, H.C. (2007). CpG PatternFinder: a Windows (R)-based utility program for easy and rapid identification of the CpG methylation status of DNA. Biotechniques *43*, 334-+.

Yang, X.J., Lay, F., Han, H., and Jones, P.A. (2010). Targeting DNA methylation for epigenetic therapy. Trends in Pharmacological Sciences *31*, 536-546.

Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in S-cerevisiae. Science *309*, 626-630.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916-919.

Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nat Struct Mol Biol *16*, 847-U870.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. *13*, 1924-1935.

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes & Development *15*, 2343-2360.

Zhang, Z.H., and Pugh, B.F. (2011). High-Resolution Genome-wide Mapping of the Primary Structure of Chromatin. Cell *144*, 175-186.

Zhao-Xia Chen, J.R.M.C.-L.H.A.D.R.F.C. (2005). Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. Journal of cellular biochemistry *95*, 902-917.

Zhu, J., He, F.H., Hu, S.N., and Yu, J. (2008). On the nature of human housekeeping genes. Trends in Genetics *24*, 481-484.

Zhu, W.G., Dai, Z., Ding, H., Srinivasan, K., Hall, J., Duan, W., Villalona-Calero, M.A., Plass, C., and Otterson, G.A. (2001). Increased expression of unmethylated CDKN2D by 5-aza-2'-deoxycytidine in human lung cancer cells. Oncogene *20*, 7787-7796.

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nature Genet. *39*, 61-69.

BIOGRAPHICAL SKETCH

Carolina E Pardo was born in Santafé de Bogotá, Colombia. She was raised in Quito, Ecuador where she attended The American School of Quito from kindergarten until graduation from high school. She then went back to Bogotá, Colombia where she joined La Universidad de Los Andes from where she obtained her undergraduate degree in biology. After graduation she worked with the Whale Heart Satellite Tracking Group conducting marine mammal research. She went back to Los Andes University and got her master's degree in biological sciences at the Human Genetics Laboratory. During this time she also taught at the undergraduate level at Los Andes University. In 2005 Carolina was accepted to Texas A&M Univesrtity in College Station, Texas, to the Biochemistry and Biophysics program. She Joined Dr. Michael Kladde's laboratory and then moved with Dr. Kladde to the University of Florida where she obtained her PhD degree. Carolina will continue working in science.