

12-2011

IDENTIFICATION OF FACTORS INVOLVED IN DNA METHYLATION OF CPG-ISLAND-PROMOTERS

Yan Zhang

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations



Part of the [Cancer Biology Commons](#), and the [Molecular Genetics Commons](#)

Recommended Citation

Zhang, Yan, "IDENTIFICATION OF FACTORS INVOLVED IN DNA METHYLATION OF CPG-ISLAND-PROMOTERS" (2011). *The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access)*. 182.
https://digitalcommons.library.tmc.edu/utgsbs_dissertations/182

This Dissertation (PhD) is brought to you for free and open access by the The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digitalcommons@library.tmc.edu.

**IDENTIFICATION OF FACTORS INVOLVED IN DNA
METHYLATION OF CPG-ISLAND-PROMOTERS**

By

Yan Zhang, M.S.

APPROVED:

Jean-Pierre Issa, M.D., Supervisory Professor

Sharon Dent, Ph.D.

Peng Huang, M.D., Ph.D.

Russell R. Broaddus, M.D., Ph.D.

Xiaobing Shi, Ph.D.

APPROVED:

George M. Stancel, Ph.D.

Dean, The University of Texas Health Science Center at Houston
Graduate School of Biomedical Sciences

**IDENTIFICATION OF FACTORS INVOLVED IN DNA
METHYLATION OF CPG-ISLAND-PROMOTERS**

A

DISSERTATION

Presented to the Faculty of

The University of Texas

Health Science Center at Houston

and

The University of Texas

M. D. Anderson Cancer Center

Graduate School of Biomedical Sciences

In Partial Fulfillment

Of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Yan Zhang, M.S.

Houston, Texas

December, 2011

Dedication

I would like to dedicate this dissertation to my dear parents, Xuming Zhang and Lianchun Qu, and my lovely sister, Dong Zhang, for their long-time support and understanding. Without them, this work would not be possible.

Acknowledgments

I would like to thank the following people for their inputs in my work, education and experience during the last few years

I would like to deeply appreciate the instructions from my advisor, Dr. Jean-Pierre Issa. He guided me through my graduate training with kindness and patience. Whenever I ran into problems in experiments, he worked with me to figure out reasons and kindly recommended solutions. He also helped me gain a concept of how to think scientifically and study independently. I am greatly thankful for his openness in accepting my own ideas, his encouragement for pursuing new questions and his generosity to support travels to academic conferences. His help has provided chances for me to improve scientific skills and build a solid base for my future career.

I would like to express the appreciation of the kind advices and guidance from the members served in my advisory, examination and supervisory committees (Dr. Sharon Dent, Dr. Peng Huang, Dr. Russell Broadus, Dr. Warren Liao, Dr. Reuben Lotan, Dr. Ana Aparicio, and Dr. Xiaobing Shi).

I am also thankful to our lab members for their helpful recommendations and generosity to my project, lab work and personal life. My special thanks would be to Dr. Jingmin Shu and Dr. Jiali Si for their enlightening discussions and sharing their experiences, to Dr. Jaroslav Jelinek, Dr. Lanlan Shen, Dr. Marcos Estecio and Dr. Woonbok Chung for providing materials and suggestions, to Rong He and Saira Ahmed for their technical advice, and to all people for their friendship.

IDENTIFICATION OF FACTORS INVOLVED IN DNA METHYLATION OF CPG-ISLAND-PROMOTERS

Publication No._____

Yan Zhang, M.S.

Supervisory Professor: Jean-Pierre Issa, M.D.

Repression of many tumor suppressor genes (TSGs) in cancer is mediated by aberrantly increased DNA methylation levels at promoter CpG islands (CGI). About one-fourth of empirically defined human promoters are surrounded by or contain clustered repetitive elements. It was previously observed that a sharp transition of methylation occurs between highly methylated repetitive elements (SINE or LINE) and unmethylated CGI-promoters (e.g. P16, VHL, CDH and RIL) in normal tissues. The functions that lead to increased CGI methylation in cancer remain poorly understood. We propose that CGI-promoters contain cis-elements for triggering de novo DNA methylation. In the first part of our project, we established a site-specific integration system with enforced local transcriptional repression in colorectal cancer cells and monitored the occurrence of de novo DNA methylation in exogenous fragments containing a CGI-promoter and repetitive elements. Initial de novo methylation was seeded at specific CG sites in a repetitive element, and accelerated by persistent binding

of a KRAB-containing transcriptional repressor. Furthermore, additional repetitive elements (LINE and SINE) located adjacent to the promoter could confer DNA methylation spreading into the CGI particularly in the setting of KRAB-factor binding. However, a repressive chromatin alone was not sufficient to initiate DNA methylation, which required specific DNA sequences and was integration-site (and/or cell-line) specific. In addition, all the methylation observed showed slow and gradual accumulation over several months of culture. Overall, these results demonstrate a requirement for specific DNA sequences to trigger de novo DNA methylation, and repetitive elements as cis-regulatory factors to cooperate with strengthened transcriptional repression in promoting methylation spreading. In the second part, we re-introduced disrupted DNMT3B or DNMT1 into HCT116 DKO cells and mapped the remethylation pattern through a profiling method (DREAM). Moderate remethylation occurred when DNMT3B was re-expressed with a preference toward non-CGI and non-promoter regions. Hence, there exists a set of genomic regions with priority to be targets for DNMT3B in somatic cells.

Table of Contents

Approval sheet	i
Title page	ii
Dedication.....	iii
Acknowledgments	iv
Abstract	v
Table of Contents.....	vii
List of figures.....	xi
List of tables	xv
Chapter 1	
Introduction	1
Cytosine methylation in normal mammalian cells	1
Basic functions of DNA methyltransferases	1
DNA methylation reprogramming during development	7
DNA methylation changes in cancer	9
Recognition of methylation signals	11
Epigenetic loop of transcriptional repression	14
DNA methylation spreading and protection	15
Methylation center definition	17
Components in controlling DNA methylation — transcriptionally repressive histone signatures and their modifiers	19
Components in controlling DNA methylation — chromatin remodeling factors and nucleosomes	21
Components in controlling DNA methylation — small RNAs and long non-coding RNAs	21
Genome-wide methylation profiling approaches (next-generation sequencing) ...	24
Hypothesis and specific aims.....	27
Chapter 2	
Methods and materials	30
Human samples	30
Cell culture	30

DNA extraction and bisulfite conversion.....	30
PCR, cloning and DNA sequencing (Bisulfite cloning/sequencing)	31
Bisulfite pyrosequencing.....	31
RNA extraction and Quantitative RT-PCR.....	43
Generation of stable single clones with single integration sites.....	43
β-Gal Staining	44
Inverse PCR.....	44
Establishment of Flpin-tTS host cells	44
Flow cytometry	44
Construction of transgenes, homologous recombination and clonal selection	45
Drug treatment	45
ChIP analysis	46
Protein extraction and western blot	46
DREAM library preparation	47
DREAM library validation — Pyrosequencing and TA cloning	48
Data processing and statistics.....	48

Chapter 3

Establish a site-specific integration system with enforced transcriptional repression.....	49
Introduction	49
Results	51
Construction of flp-in host cells	51
Heterochromatic environment of the insertion sites	59
Generation of flp-in/tTS host cells	60
Discussion.....	65

Chapter 4

Repetitive elements and strengthened local heterochromatin enhanced de novo methylation in a promoter-CGI	68
Introduction	68
Results	70

Mapping the methylation patterns of RIL (PDLIM4) promoter	70
Construction of transgenes	73
Continuous culturing under selection pressure populated unmethylated clones.....	77
Unequal silencing speeds of transgenes in flp-in and flp-in/tTS cells	77
De novo DNA methylation center in transgenes.....	78
The effects of repetitive elements on DNA methylation spreading	86
Gradual accumulation of methylated cytosine over time	86
Long-distance spreading was not direction-dependent	87
Promoted CGI methylation arose from adjacent de novo methylation.....	88
A highly repressive environment can enhance de novo methylation and methylation spreading	91
Local heterochromatin in transgenes.....	92
The presence of tTS made transgenes refractory to TSA-induced derepression	97
The effects of cell lines or genomic loci on DNA methylation recruitment	97
Discussion.....	102
 Chapter 5	
The de novo methylation in somatic cells caused by re-expressing DNMT3B in a genetically engineered model	107
Introduction.....	107
Results	108
Methylation is not equally recovered in DKO cells.....	108
Re-expression of DNMT3B or DNMT1 moderately recovered global DNA methylation.....	117
Preferential representation of CGI methylation for DREAM	117
Re-expressed DNMT3B favored non-CGIs	119
Promoters were more resistant to re-expressed DNMT3B	120
Discussion.....	134
 Conclusions.....	137
Significance and future directions.....	140
References.....	144

Vita	164
-------------------	------------

List of figures

Chapter 1

Figure 1. Structural overview of the functional domains of the DNA methyltransferase family.....	4
Figure 2. A proposed model for the maintenance of DNA methylation in mammalian cells.....	6
Figure 3. Reprogramming of DNA methylation during early development in mammals..	8
.....	
Figure 4. The model for the feedback loop of epigenetic regulations.	13
Figure 5. The model of "methylation center" and protection around the CpG-promoter.	
.....	23
Figure 6. Profiling DNA methylation by DREAM analysis.....	26

Chapter 2

Figure 7. Schematic description of establishing a site-specific integration system..	53
Figure 8. Transfection of pFRT/LacZeo into cell lines..	54
Figure 9. Tiling primers to screen out possible single clones with single integration sites..	55
Figure 10. Identify and validate the single integration site of flp-in host cells. (A)	
Graphical description of inverse PCR..	57
Figure 11. DNA methylation of the environment for homologous recombination..	61
Figure 12. Chromatin signatures of the integration sites in SW48 and HCT116.....	62
Figure 13. Generation of flp-in/tTS host cells.	64

Chapter 3

Figure 14. Mapping the methylation pattern of RIL promoter by bisulfite sequencing..	72
Figure 15. Construction of transgenes.....	75
Figure 16. Continuous selection populated unmethylated clones..	76
Figure 17. Unequal GFP silencing speed in flp-in and flp-in/tTS host cells.....	81
Figure 18. Sensitive CG sites to de novo DNA methylation by bisulfite pyrosequencing.....	82
Figure 19. Methylation patterns of pINSL6 in SW48A/tTS at the second month..	84
Figure 20. The correlation of two assays in analyzing segmental methylation of pINSL6.	85
Figure 21. Gradual accumulation of transgene methylation in SW48A/tTS over time, sampled at 2, 3, 4 and 5 months.....	89
Figure 22. Methylation of two regions furthest away from pINSL6 which are inside of EGFP and hygromycin B.....	90
Figure 23. Comparison of transgene behaviors between the truncated and complete pINSL6 in SW48A/tTS.....	93
Figure 24. Methylation profiles of pINSL6 in flp-in host cells by pyrosequencing.....	94
Figure 25. Comparison of CGI methylation and GFP expression between tetO (-)- pINSL6 in SW48A/tTS, tetO(+)-pINSL6 in SW48A and tetO(+)-pINSL6 in SW48A/tTS.....	95
Figure 26. Regional methylation of transgenes in SW48A at the second month by pyrosequencing.....	96
Figure 27. ChIP-qPCR for analyzing the enrichment of histone marks in pINSL6.	99

Figure 28. The reaction of SW48A and SW48A/tTS clones to epigenetic drug treatment...100

Figure 29. Methylation profiles of pINSL6 of transgenes in flp-in host cells..... 101

Chapter 4

Figure 30. Methylation is not equally recovered in DKO cells..... 111

Figure 31. Re-expression of DNMT3B1 in DKO cells moderately recovered global
DNA methylation..... 114

Figure 32. Re-expression of DNMT1b in DKO cells moderately recovered global DNA
methylation. 116

Figure 33. Preparation of DREAM libraries..... 122

Figure 34. Overview of DREAM sequencing results..... 124

Figure 35. Scatter plots for methylation distribution of SmaI/XmaI (CCCGGG) sites..
..... 127

Figure 36. CGIs were not sensitive to re-expressed DNMT3B in DKO..... 129

Figure 37. Percentages of SmaI sites remethylated (top) and un-remethylated (bottom)
in respect to their distance to the closest transcription start sites (TSS)..... 130

Figure 38. Examples of remethylation and un-remethylation of CGIs in CGI-promoters.
..... 131

Chapter 5

Figure 39. A Model for late response of DNA methylation to transcriptional repression..
..... 139

List of tables

Table 1. PCR primers and sequencing primers used for bisulfite-sequencing and pyrosequencing.	33
Table 2. RT-PCR primers	37
Table 3. PCR primers for subcloning to constructs	38
Table 4. PCR primers for confirmation of integration	39
Table 5. Inverse-PCR primers	41
Table 6. Chip primer and probes	42
Table 7. Characterization of the single integration site in flp-in host cells.	58
Table 8. A list of genes with significant remethylation on at least one SmaI site of CGI-promoters.	132

Chapter 1

Introduction

Cytosine methylation in normal mammalian cells

DNA methylation, catalyzed by DNA methyltransferases (Dnmts), is a post-replicative mechanism that regulates expression and conveys the epigenetic memory to progeny cells. This modification transforms cytosine to 5'methyl-C, which is restricted to CpG dinucleotides in mammalian gDNA (genomic DNA). Methylated CpGs are mainly distributed within repetitive elements (LINEs, SINEs, LTR, etc) and coding regions of functional genes, while unmethylated ones are often located in regions called CpG islands (CGIs). CpG islands are defined as the patches of DNA with at least 500bp length, CG content $\geq 55\%$ and the ratio of observed CpG to expected CpG $\geq 65\%$ (1). They are not methylated in germ cells, and most of them still keep methylation-free, especially those located at the promoters of housekeeping genes and tumor suppressor genes (TSGs). About half of genes have CGIs in their promoters, and methylation of them is often associated with gene silencing. Generally speaking, vertebrates use heritable DNA methylation to lock genes in silent states in order to act for self-defense against pathogens, maintain genomic stability, and regulate mono-allelic expression of imprinted genes, and so on.

Basic functions of DNA methyltransferases

The DNA methyltransferase (DNMT) family of proteins in mammals consists of several members with respective features in catalyzing the attachment of methyl-group to the 5'-cytosine. DNMT1 is generally considered as the maintenance enzyme with higher affinity for the hemi-methylated DNA. It helps maintain established methylation patterns by working

with proliferating cell nuclear antigen (PCNA) at replication forks (2). The dynamic binding of DNMT1 with PCNA accommodates the different kinetics of replication and DNA methylation, but is not absolutely required for methylation maintenance (3). The protein in charge of recognizing hemi-methylated DNA during replication is NP95 (also called uHRF1 and ICBP90) in mammals, a SET-and RING-associated (SRA)-domain-containing protein. Its interaction with DNMT1 plays a dominant role in tethering DNA methyltransferase activity into replicating DNA (4, 5). Of note, semi-conservative reproduction of methylation maintenance does not equal to precise copying at the level of a single CG dinucleotide, instead, it is the methylation status of DNA domains that appears to be faithfully propagated between cell passages demonstrated by heterogeneous methylation patterns in clonal population of cells (2).

Different from DNMT1, DNMT3A and DNMT3B are dedicated to de novo methylation, which add methyl groups onto sites with absence of methylation at either strand. It is shown in Fig. 1 that most DNMT proteins share common C-termini that are responsible for the methyltransferase activity, while the N-terminal DNMT3A and DNMT3B consist of a cysteine-rich domain and a tryptophan-rich region (PWWP) domain and are required for directing enzymes to the major satellite repeats at pericentric heterochromatin (6). The N-terminus of DNMT1 is composed of a different set of motifs for its interaction with other proteins and there is a linker region connecting the N- and C-termini that is absent in DNMT3's structure. It is documented that sustained expression of DNMT3A and DNMT3B are required for the wave of de novo methylation during embryogenesis, and are still detectable in adult tissues though at lower levels than DNMT1 (7). DNMT3L is the homologue solely expressed in germ cells, and its structure keeps the PWWP domain and is

devoid of the functional catalytic domain and the cysteine-rich domain. Lacking the catalytic motifs, DNMT3L is considered as the cofactor for DNMT3A and DNMT3B in de novo methylation of dispersed transposons.

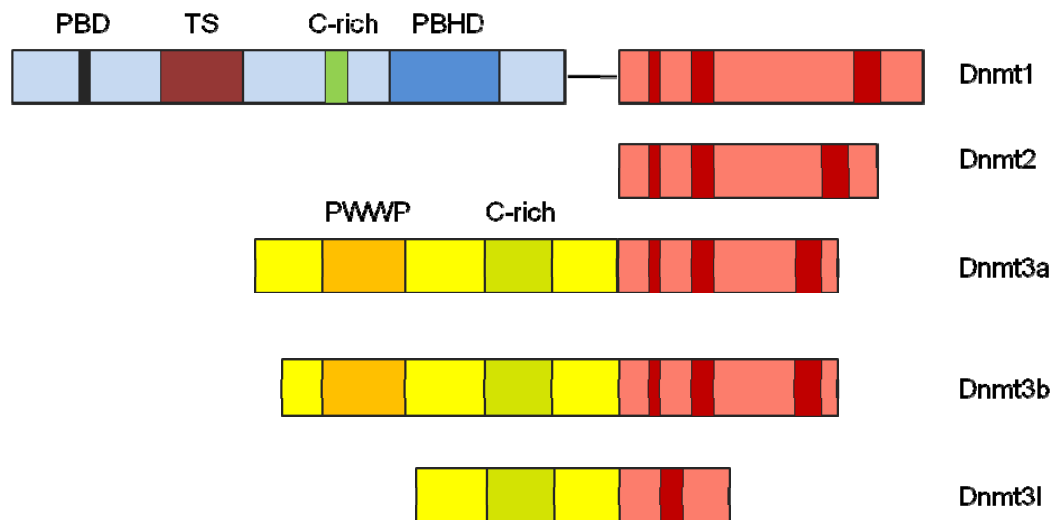


Figure 1. Structural overview of the functional domains of the DNA methyltransferase family. Most Dnmts share conserved motifs at the C-terminus. The N-terminus of Dnmt1 includes a PCNA binding domain (PBD), a targeting sequence (TS), a cysteine-rich region (C), a polybromo homology domain (PBHD); while Dnmt3a, Dnmt3b and Dnmt3L consist of a tryptophan-rich region (PWWP) and another cysteine-rich region (C-rich) (Dnmt3L only has a C-rich domain). There is a region , (GlyLys)₆ repeat between the N- and C-terminus of Dnmt1. Adapted from (7).

Further evidences are accumulating that DNMTs present functional co-operation or redundancy in both de novo and maintenance methylation. By in vitro methyltransferase assay with cell extracts and heterologous expression of DNMT1 in *Drosophila* cells, Dnmt1 was observed to account for majority of de novo methyltransferase activity in cancer cells and methylate human CGIs integrated in the fly genome (8). Individually knockout DNMT3B in HCT116, a hypermethylated colorectal cancer cell, still retained about 87% methyltransferase activity and only led to less than 3% demethylation of global 5'methyl-C and modest demethylation of pericentric satellite (9). On the contrary, both double knockout (DNMT1^{-/-}DNMT3B^{-/-} HCT116) and severe genetic DNMT1 knockout (80%-90% reduction) in cancer cells resulted in >95% depletion of genomic 5'methyl-C, including DNA repeats, IGF2(insulin-like growth factor 2) imprinting and TSGs like p16INK4A (9, 10), which suggested the vital role of DNMT1 in somatic DNA methylation and cooperation from DNMT3B in maintenance of DNA methylation. Moreover, colocalization of multiple DNMTs at the promoters of hypermethylated genes (11-13), together with a high level of hemi-methylated LINE-1 and Alu elements in mouse ES cells lacking DNMT3A and DNMT3B, provide evidences for a new model where DNMT3A and DNMT3B are suggested to undertake the responsibility for maintaining methylation especially at repeats and imprinted regions as well as methylated CGIs through strong anchoring to nucleosomes (14-16). DNMT1, on the other hand, predominantly contribute to the bulk of DNA methylation during DNA replication (Fig. 2).

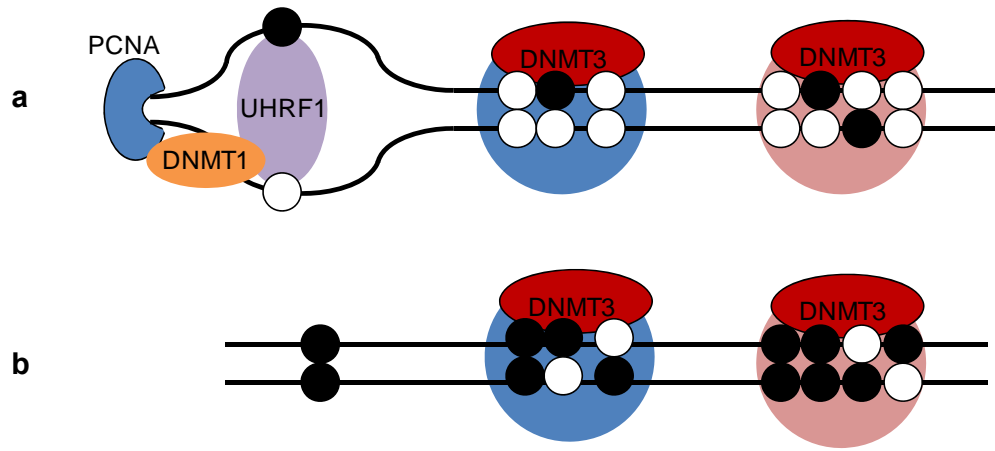


Figure 2. A proposed model for the maintenance of DNA methylation in mammalian cells. (a) During DNA replication, DNMT1 predominantly catalyzes bulk-DNA methylation with the help of PCNA and/or UHRF1, and may contribute to error correction as well. (b) De novo enzymes, DNMT3A and DNMT3B retain bound to methylated CGs especially those occupied by nucleosomes. They are compartmentalized to CGIs and repetitive elements, perform ongoing de novo methylation which may not require recognition of parental strands and catalyze the leftovers of DNMT1. Blue nucleosome, newly assembled nucleosome; pink nucleosome, parental nucleosomes (14).

DNA methylation reprogramming during development

Global DNA methylation patterns experience reprogramming during development. In mammals, the patterns inherited from parents are erased in fertilized eggs during pre-implantation stage, and then restored to normal somatic level with a new embryonic pattern after implantation, which is established by combination of de novo and maintenance methylation. But reprogramming in paternal and maternal genomes in early embryos occurs through both active and passive mechanisms. Active demethylation in the paternal genome commences immediately after fertilization and before the first DNA duplication, while demethylation of the maternal genome is passively dependent on DNA replication due to exclusion of DNMT1 from the nucleus (Fig. 3). However, methylated imprinted genes and some repetitive elements do not become unmethylated such as paternal H19 gene, paternal LINE-1 and IAP, and maternal Alus. In the short window of postimplantation, de novo methylation re-establish methylation patterns to different extent in embryonic (EM) and extraembryonic (EX) lineages. Both the demethylated and unmethylated regions, including unmethylated maternal LINE-1 and IAP, paternal Alus, and microsatellite DNA in both parental gamete genomes, are engaged in this process except for some unmethylated imprinted genes (17-19).

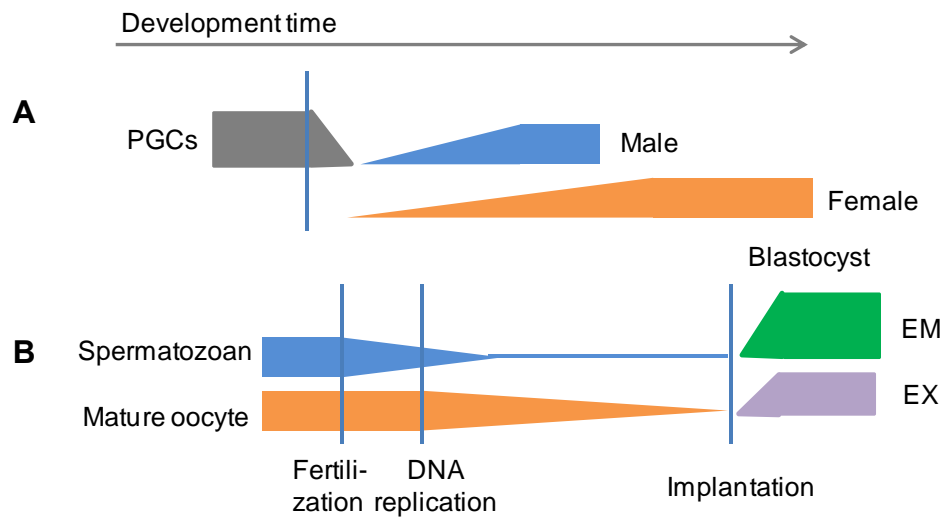


Figure 3. Reprogramming of DNA methylation during early development in mammals.

(A) Alterations of DNA methylation in gametogenesis. (B) Demethylation and re-methylation in paternal and maternal genome during early embryogenesis (19).

The second stage is during gametogenesis when demethylation and de novo methylation also occur to set up parental-specific methylation in imprinted genes. Demethylation takes place in the early development of primordial germ cells (PGCs) (Fig. 3) when PGCs enter the gonads, lasts several days and is completed by the embryonic day (E) 13 or 14 in both male and female germ cells. Remethylation in male germ cells occurs at the prospermatogonia stage (E15 to E16 and onwards), precedent to the time for female germ cells which is after birth during the maturation of oocytes (17-20).

The essential role of DNA methylation in development is demonstrated from gene knockout or mutation in mouse ES cells or embryos. Disruption of DNMT3A, DNMT3B or both in mouse ES cells or embryos led to lethality although DNA methylation did not decrease much especially for single gene knockout (17). Genetic inactivation of DNMT1 also caused lethality but there is about 80% methyl-cytosine that lost methyl-groups without sequence specificity (21, 22). Similarly, the factors, like UHRF1, G9A, LSH, could induce extensive demethylation accompanied with lethality in ES cells and embryos.

DNA methylation changes in cancer

Epigenetic abnormalities have already been proved to collaborate with genetic changes in carcinogenesis. Concurrent with chromatin remodeling, altered DNA methylation including global hypomethylation and localized hypermethylation is correlated with genomic instability and aberrantly expressed genes. Loss of imprinting (LOI) due to demethylation concurrent with bi-allelic expression of IGF2 is correlated with a higher risk of colorectal cancers or Wilm's tumors, a childhood renal cancer. Localized hypermethylation of promoter CGIs exists in various cancer types and is often associated with loss of TSG expression (RB,

p16INK4A, MLH1, MGMT, CDH1, etc). Also, methylated cytosines are susceptible to mutagenesis, like the case at TP53 coding region (23). Moreover, genomic screening of hypermethylated genes indicated a potential network of epigenetic events predisposing progenitor cells to transformation, as suggested in studies of SFRPs (secreted frizzled related genes) hypermethylation for permission of Wnt pathway in colon cancers (24).

It is proposed further that multiple hypermethylated genes in certain colorectal carcinomas could be signatures of an epigenetic phenotype for cancer-specific markers – CIMP (CpG island methylator phenotype), correlating epigenetic defects with histopathological changes in cancer (25). Generally, CIMP is more likely to be at the proximal colon with more mucinous histology, more frequency of MSI (microsatellite instability) and more chance of BRAF mutations (26). The association of CIMP with genetic instability such as MSI or CIN (chromosome instability) is gradually revealed by genome-wide research so as to predict the prognosis of certain colorectal cancers. For example, patients with *BRAFV600E* and MSI^+CIMP^+ phenotype, termed as CIMP-H, have better prognosis, in contrast to the worse outcomes for patients with *BRAFV600E* and MSI^-CIMP^+ phenotype (27, 28). More recent studies have suggested the existence of CIMP in multiple other cancer categories like glioma and ovarian cancer (29-31). In brief, altered DNA methylation is a possible early event in tumorigenesis and facilitates the colonization of precursor cells, therefore could be a kind of indicator for cancer subsets.

What causes aberrant patterns of DNA methylation in cancer is awaiting clear illustration. It is still controversial as to the initiation directly by elevated DNMT levels. It is easy to relate the level of enzyme expression with abnormal methylation states since in the early embryonic development *de novo* DNMTs are downregulated once the pattern is

established. Artificial overexpression of DNMT1 in NIH3T3 or induction of DNMT1 by overexpression of FOS was able to promote cellular transformation (32, 33). As well, the moderate increase of DNMTs protein level or activity takes place in some cancers (24). However, it cannot explain the complex involvement of DNMTs in regional hypermethylation during development and neoplasia (8, 34, 35).

Recognition of methylation signals

The identified methyl-cytosine-binding proteins (MBPs) include three families of proteins with high affinity for methylated CpGs: the MBD domain family, Kaiso and Kaiso-like proteins and the SRA domain proteins. The MBD family consists of the methyl-CpG-binding protein 2 (MeCP2), the methyl-CpG-binding-domain proteins MBD1, MBD2, MBD3 and MBD4. They are involved in complexes containing histone deacetylase and histone methyltransferase activities as transcriptional co-repressors (36). All these proteins are expressed ubiquitously in somatic cells. Mice deficient for MBD1 and MBD2 are still viable with mild and indistinct phenotypes, while aberrant MeCP2 underlies most of Rett syndrome which is a neuro-developmental disorder with exclusive frequency in human females (37). There presumably are sequence-preferences for MBPs. The X-linked gene MeCP2 has higher binding in methylated CGs with adjacent enrichment for four or more A/T bases, which may depend on first MBD2 binding. Rather, MeCP2 seems dispensable for MBD2 due to occupancy of MBD2 in half of the binding sites under MeCP2 knockout (38). MBD1 also has higher affinity for some specific methylated CGs within TCGCA and TGCGCA sequence context than any other sequence which could be stabilized by its CXXC domain (39). From current studies, the recognition of certain subsets of methylated regions

for MBPs is attributable to their MBD domains and to some extent has overlapped target sites.

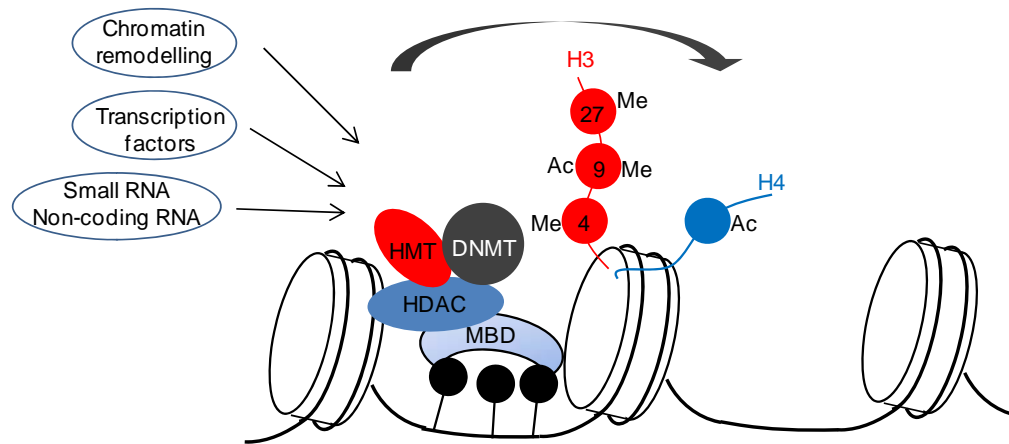


Figure 4. The model for the feedback loop of epigenetic regulations.

Epigenetic loop of transcriptional repression

There could be two aspects of transcriptional silencing with the involvement of DNA methylation (Fig. 4). First, methylated DNA could preclude the localization of transcription activators to target promoters and both are exclusive to each other. Second, methylated DNA could induce sequential recruitment of multiple layers of proteins consisting of MBPs, histone deacetylases (HDACs), histone methyltransferases (HMTs) and related accessory factors, conferring compact chromatin structures to silence genes. As well, the other way around exists that histone modifications are prior to DNA methylation occurrence in some experimental systems. Thus, these factors form a loop of epigenetic regulation affecting gene expression.

However, the dominance between histone changes and DNA methylation in gene repression is still controversial. One model describes DNA methylation as the original event, supported by the observations that in vitro methylated expression plasmids usually greatly decreases reporter gene expression (>90%) upon transient transfection into cultured mammalian cells (40) which is accompanied with deacetylated histones (41), and that loss of DNMT1 in human cells resulted in decreased H3K9me2 and H3K9me3 in pericentromeric heterochromatin(42). It seems that methylated promoters are not silenced until chromatin remodelling occurs (43).

Another pathway suggests a primary role of repressive histone methylation in repression and DNA methylation as a secondary solidifying event. Increasing evidences come from diverse systems (23, 24, 44). Further studies indicated the cooperation between DNA methylation and H3K9me can be dissected by recognizing the specific target residues of HMTs. They observed siRNA knockdown of DNMT1 impaired DNA methylation, G9a

loading at replication loci and H3K9me, but Suv39h1 activity still remained (45). The same controversy exists for H3K27me3. This mark, like H3K9me3, can be concurrent with gene silencing but without DNA methylation. It is demonstrated that Ezh2 is associated with three major DNMTs via its N-terminal portion (46), and the PcG-target genes would be susceptible to de novo methylation in cancer as they are pre-marked by H3K27me3 during early development (47, 48). However, data from Ezh2-depleted U2OS and RKO showed DNA methylation and silencing still persisted at target promoters (p16, hMLH1, MYT1 and WNT1) (49). Another possibility is that polycomb-mediated H3K27me3 and DNA hypermethylation may not have the expected direct contact and can act in parallel depending on certain cell status, as suggested in several recent papers (50, 51). There seems more results supporting the precedence of transcriptional repression and chromatin remodelling over DNA methylation.

DNA methylation spreading and protection

Heterochromatin is the condensed domains interspersed by the decondensed euchromatin. It is characterized by enrichment for methylation of H3K9 and/or H3K27, lack of acetylation of histone, occupant with HP1 and also, high levels of methylated CGs. Without the protection of boundary elements, heterochromatin state is prone to extend to adjacent areas, which is realized by the feedback loop of DNA methylation, histone modification and chromatin remodeling with the help of adaptors like HP1. As shown in some genes like estrogen receptor (ESR1), it is likely aging participates in initiation of this spreading process, so the aging-dependent promoter methylation could be associated with the risk of some cancers, although this has not been clearly clarified (52, 53). The mechanisms of

heterochromatin spreading are not defined yet, and models proposed to explain long-range spreading of silent chromatin all involve dynamic reorganization of higher chromatin structures. For instance, in the “oozing” model, sequential assembly of sets of proteins is facilitated by one particular protein binding; the “looping” model talks about long-distance fiber-fiber connections mediated by certain organizers; not to mention the factors in the “sliding” model are tracking down a chromosome with transcriptional complexes; and so on (54).

Normally, spreading can be resisted by protection from insulators (54-56). According to specific local features, some insulators work as enhancer-blocker, others as barriers. The enhancer blocking insulators are defined as sequences being placed between the enhancer and promoter while interrupting their interaction. A classical example comes from the studies of imprinted Igf2-H19 locus where there are a run of CTCF-binding sites in the ICR situated between the enhancer and the Igf2 promoter. Paternal genome expresses Igf2 because paternal-specific methylation of ICR abolishes the blocking via CTCF binding (57, 58). It is possible that CTCFs position the enhancer and promoter into separate domains by interacting and clustering as well as tethering the chromatin fiber to the nucleolar surface thereby forming loops and preventing inter-loop contacts; alternatively, some activation signals travel from the enhancer to the promoter which is stopped by half-way CTCF deposition (59). The barriers are the reason for PEV (position effect variegation) which is spreading of heritable silencing to euchromatin from heterochromatin. The cHS4 locus lying at the 5'-end of the chicken β -globin locus has complex functions of both enhancer-blocker and barrier. CTCF mediates the strong enhancer-blocking activity, and transcription factors USF1 and USF2 cooperate as barriers due to recruitment of HAT and HMT to catalyze histone deacetylation

and H3K4 methylation together with chromatin remodeling complex, thus “openness” stop the spreading of heterochromatin (60). Other transcription factors like Sp1/Sp3 could also protect histone-based silencing against spreading and lead to variable DNA methylation. For example, the RIL (PDLIM4) gene promoter is present in two forms of alleles: the long allele has additional 12-bp insert (CGGCGGCGGCTC) and a substitution of T to G 3 bases upstream of the insertion site in the short allele, which is later found as a Sp1/Sp3 binding-site so as to create a local protection in the long allele which usually has less methylated CG frequency (61).

Methylation center definition

How heterochromatin is nucleated at specific chromosomal domains remains elusive, though evidences suggested certain cis-acting sequences like silencers and repetitive elements could provide signals for nucleation (62). Because the phenomenon of DNA methylation spreading has been observed in aging and cancerous cells from CpG island borders into transcription start sites (TSSs), the “seed and spread” model of DNA methylation is proposed to explain the distinct methylation patterns and related silencing events in pre- and post-neoplasm (63, 64) (Figure 5). The methylation center is hypothesized as the region initiating DNA methylation and recruiting de novo DNMTs. Turker M.S. and colleagues identified two upstream B1 repetitive elements as cis-signals for de novo methylation of mouse Aprt gene on the X-chromosome, and methylation spreading is resisted by undetermined factors binding to one of the Sp1 sites between retrotransposons and its CGI-promoter (65-68). Other retrotransposons were also suspected methylation centers, such as human Alu, mouse LINE-1, B2 and IAP (16, 69-72). A series of genes (p16INK4A,

ER, hMLH1, VHL, CDH1, etc) are silenced concomitant with hyper-methylated CGI-promoters in colorectal, breast, prostate and hematopoietic cancers. The common feature of these genes is a CGI-promoter surrounded by consistently methylated upstream or/and downstream repetitive elements in both normal and malignant conditions. Therefore, these genes with such methylation patterns are considered as possible candidates to study methylation centers.

The concept of “methylation centers” brings up consideration of the original trigger for DNA methylation. Does DNA methylation directly deal with specific sequences, or indirectly result from other processes which possess site-preferences? Both cases may coexist during physiological and abnormal development.

It is easy to first examine the sequence preferences of DNMTs, particularly de novo DNMTs (DNMT3A, DNMT3B, and even DNMT3L in germ cells). The machinery of attracting de novo DNMTs could depend on the discrepant substrate preferences of the enzymes. For instance, both Dnmt3a and Dnmt3b favored RCGY with AT-rich flanks and disliked YCGR (34, 73). While this may be true, it cannot explain the complex involvement of DNMTs in regional de novo methylation during development and neoplasia (8, 34, 35). A hypothetical target is DNA repeats which account for the high level of global DNA methylation and about 25% of promoters are surrounded by repetitive elements (74). The suggestive evidence in mammals is that high copy number of transgene induced significant repression along with DNA methylation (75).

Many recent studies reveal components participant in chromatin and nucleosome organization are very likely more decisive than just DNMTs, and they involve all the aspects

of epigenetic regulators including histone modification, chromatin remodeling factors, micro-RNA and even nuclear skeletal molecules.

Components in controlling DNA methylation — transcriptionally repressive histone signatures and their modifiers

In epigenetic regulations, the combined status at the N-termini of histones has become the indicator for chromatin organization and transcriptional activity. Generally, H3 acetylation and methylation of histone3 at K4, K36, and K79 are linked to gene activation, whereas H3K9me3, H3K27me3 and H4K20me3 are hallmarks for repressive status (24). Each histone methylation has its own catalysis system and distribution patterns. H3K9 methylation is catalyzed by several types of histone methyltransferases (HMTs). Suv39h1 and Suv39h2 are responsible for tri-methylated H3K9 which is more associated with compact and transcriptionally inactive heterochromatin at pericentromeric areas and recently mapped to silenced euchromatic genes; mono- and di-methylation, catalyzed by G9a/GLP (EuHMTase), seem to be restricted in euchromatic regions, although H3K9me1 is mainly found around TSSs of actively transcribed genes (24, 44). Another crucial repressive histone mark H3K27me3, is catalyzed by Ezh, one of the PRC2 (polycomb repressive complex 2) components, assisted by two PcG (polycomb group) proteins of Eed and Suz12 and a member of the Jumonji C (JmjC) domain protein family, Jarid2 (76, 77). The H3K27me3 is further maintained by PRC1, also composed of several PcG proteins, leading to occlusion of transcription machinery and compaction of chromatin. The PcGs are not only involved in maintaining cell identity during development, but also in tumorigenesis, as shown by increased expression of Ezh2, Eed, Suz12 and Bmi in colon and breast cancers. The

enrichment for H3K27me3 is primarily found in imprinted genes, inactive X-chromosome and almost all the DNA hypermethylated and silenced cancer genes (44, 78).

As mentioned in the previous part, the enzymes for histone modification do interact with DNMTs. Experimentally, this interaction does not rely on the enzymatic domains, namely SET domain, to assemble DNMTs into the complex. One example is mutation of the SET domain in G9a could not affect related DNA methylation at the time of eliminating H3K9 methylation, that is, DNMT3A and DNMT3B are still located through biochemical interaction with another domain and histone methylation is not responsible for their binding (79, 80). This dispensability of histone modifications for DNA methylation also holds true for other SET-domain-containing enzymes, such as SETDB1, SUV39H1 and EZH2 (46, 81, 82). Accordingly, although DNA methylation could be set up at the early stage of development or carcinogenesis, it possibly would be a late-response step in gene silencing to consolidate local memory of repression and facilitate its heritability. This has been shown in series of micro-array analyses in cancers that many methylation events were observed in promoters where expression is already shut down in normal cells (83). The occurrence of silencing events and histone modifications prior to DNA methylation is also demonstrated in repression of specific promoters like p16INK4A silencing in primary human mammary epithelial cell (HMEC) strains during selection which is a model for early breast cancer (84). Therefore, there may be three steps in setting up epigenetic marks on silenced genes. First, transcription factors are induced to bind specific promoters and transcription is turned off, and this is a reversible silencing stage. Second, the previous active marks are removed and enzymes are recruited to add repressive modifications to residues, such as G9a for H3K9me, to create local heterochromatin, which is still a flexible stage. Finally, DNA become

methylation aided by enzyme (e.g. G9a)-containing complexes and a stable inactivation is stamped on the local regions which can be inherited over cell proliferation (85).

Components in controlling DNA methylation — chromatin remodeling factors and nucleosomes

The appropriate DNA methylation is ensured through association of accessory factors with DNMTs. The SWI/SNF chromatin remodeling complex regulates the accessibility of DNA to transcription machinery by mobilizing the octamers. Both ATRX and Lsh, members of the SNF2 helicase family, have been indicated to have effects on global methylation patterns (86, 87). Lsh, as a sequence-specific regulator, has direct associations with DNMT3a and DNMT3b instead of DNMT1, suggesting its connective roles in de novo methylation (88-90).

Heterochromatin protein 1 (HP1 α , HP1 β , and HP1 γ in mammals) is an adaptor involved in multiple facets of chromatin configuration. Its functions are related to heterochromatin formation and gene silencing, telomere capping and silencing, and positive control of gene expression (91). The plasticity of HP1 in chromatin organization is determined by its variant isoforms and binding partners, including the direct association with DNMT1 in repression of euchromatic genes and colocalized with DNMT3A and DNMT3B in heterochromatin as well (92, 93).

Components in controlling DNA methylation — small RNAs and long non-coding RNAs

MicroRNAs are a class of 19-25nt, single-stranded non-coding RNAs originated from genomic loci. It is clear that plants, worms and fission yeasts all have transcriptional gene silencing (TGS) as well as post-transcriptional silencing (PTGS) mechanisms in their gene regulation. However, the details of TGS in mammals still remain unclear. Most of the evidences for RNA-dependent DNA methylation (RdDM) come from exogenous RNA strands. Morris et al. observed gene inhibition with DNA methylation, H3K9me3 and H3K27me3 was induced by promoter-directed siRNAs in human cells (94, 95). Studies have shown the link between de novo DNA methylation and piRNAs (piwi-interacting RNAs) in germ cell development (96, 97). On the other hand, short dsRNA (double-strand RNA) targeted at the CpG island of CDH1 promoter induced TGS in cancer cells without CpG island methylated (98). Taken together, miRNAs may contribute partly to selective promoter hypermethylation in cancer cells (99), and the differences in experimental designs targeting at coding or promoter regions may be responsible for the controversial results.

The role of long non-coding RNA (lncRNA) in targeted de novo methylation is another interesting aspect and still needs further proofs. It has been suggested from studies of X chromosome inactivation and gene imprinting. In the mouse embryos, non-coding chromosome RNA expressed from Xist gene on the inactive X chromosome triggers the inactivation prior to the CGI-promoter silencing through PRC2 (100).

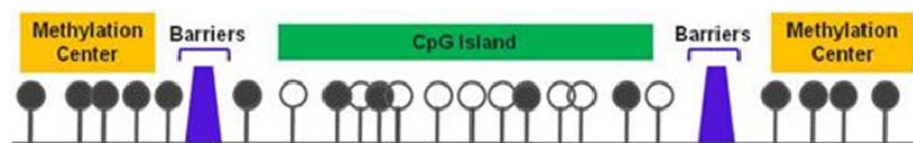


Figure 5. The model of "methylation center" and protection around the CpG-promoter.

Genome-wide methylation profiling approaches (next-generation sequencing)

The first generation of sequencing mainly refers to the automated Sanger sequencing, and more recent developed methods are considered as the next-generation, which have become commercially available platforms from Illumina/Solexa, ABI/SOLiD, Roche/454, and so on. These sequencing methodologies require multiple steps including library preparation, amplification, sequencing, imaging and data analysis, which generate millions of reads per run. Compared with micro-array technologies, they have the advantage of covering repetitive elements and assigning the features to specific alleles.

Coupled with these newer sequencing technologies, approaches to obtain genome-wide profiles of DNA methylation have come out. Basically, some require bisulfite-converted genomic DNA for template preparation like MethylC-seq and RRBS (reduced representation bisulfite sequencing) (101, 102), some are based on enrichment of methylated DNA like MeDIP-seq (methylated DNA immunoprecipitation sequencing) and MBD-seq (methylated DNA binding domain sequencing) (103, 104), and some utilize the methylation sensitive characteristics of restriction enzymes to digest genomic DNA. Each method has its own advantages and disadvantages with respect to covered regions, sequencing depth, accuracy, cost and so on. For example, MeDIP-seq and MBD-seq are not able to go deep to single base resolution but can reflect high to medium methylation of stretches of DNA and covers broader regions; the bisulfite-based methods, in contrast, allow quantitative comparison down to single nucleotides but are limited to CGIs and promoters (105, 106).

DREAM (Digital restriction enzyme analysis of methylation) is an enzyme-recognition-based approach which quantitatively maps DNA methylation with high resolutions. This technique employs restriction enzymes, such as SmaI and XmaI, to generate

“unmethylated” and “methylated” overhangs for digested DNA fragments, which can be modified to enter the subsequent pipeline of one next-generation sequencing platform (e.g. HiSeq2000 or GA II from Illumina, SOLiD from Applied Biosystem) (Fig 6). The discrimination of “unmethylated” and “methylated” CG dinucleotide results from the ability of a pair of restriction enzymes to be blocked by methylated cytosines. For example, when using sequential digestions by SmaI and XmaI, the first reaction products are fragments with blunt ends (5'-GGG and 3'-CCC) to reflect “unmethylated” CG since SmaI is blocked by methylated CG; the following digestion by XmaI leaves sticky 5'-overhangs (5'-CCGGG and 3'-C) representing “methylated” CG although XmaI has no such preference. After end repairs and adaptor ligation, specific sizes of ligation products are selected and amplified with high-fidelity, which then are ready for large-scale sequencing. Because there is no bisulfite-treatment so that the methylation ratio for a specific CG site can be calculated, the methodology of DREAM has no bias for either state and the error rate is less than 10% (unpublished data). With the decreasing cost and improved depth of sequencing, DREAM can also be very cost-effective to study epigenomes of various diseases and cancer.

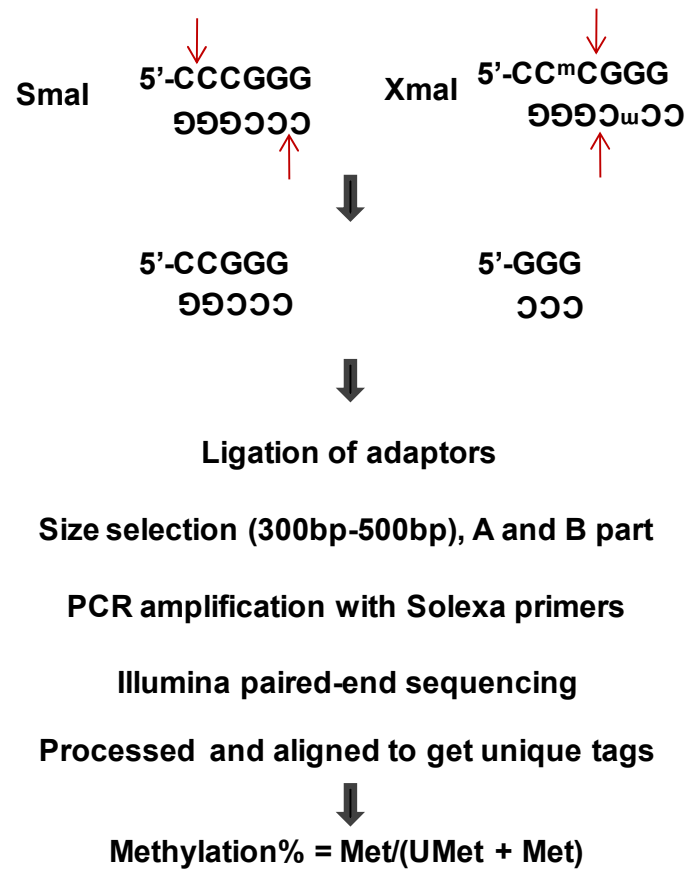


Figure 6. Profiling DNA methylation by DREAM analysis.

Hypothesis and Specific Aims

- I. CGIs are usually places precluded from DNA methylation, and only a small fraction of CGIs, particularly promoter-CGIs, is densely methylated in somatic cells (74, 107, 108). Aberrant promoter-CGI methylation has also correlation with abnormal gene expression in cancers. Repetitive elements account for almost half of genome sequences, and about 25% of promoters are surrounded by repetitive elements (74). Consistent high levels of cytosine methylation often feature upstream and downstream of CGI-promoters which become hypermethylated in malignant states, such as p16INK4A, ER, hMLH1, VHL, CDH1, etc. Intriguingly, many of the promoters contain adjacent repetitive elements (e.g. LINEs and SINEs). RIL (PDLIM4) and INSL6 identified by us also share the same features as the above genes (108, 109). Accordingly, we hypothesize that CGI-promoters contain cis-elements responsible for de novo DNA methylation.

Aim 1. Identify CG sites in sample CGI-promoters with priority for DNA methylation.

Aim 2. Assess the roles of repetitive elements in methylation spreading.

Aim 3. Assess the involvement of different local environment in DNA methylation.

We establish a site-specific integration system to ensure every constructed transgenes inserted into the same genomic locus. This study system is generated in two colorectal cancer cells and has two magnitudes of strength of repression. By stable transfection, constructs with variable fragments from sample CGI-promoters will experience de novo DNA methylation, if it could occur, so we expect differential sensitivity of CG sites to DNA methylation under the inner pressure of cell lines and integration loci.

II. Hypermethylated CGI-promoters in cancer cells suggests de novo DNA methylation still functions in somatic cells, no matter it is a bona fide process or superimposed on scarcely distributed methylation background, since DNMT3A and DNMT3B are still detectable though significantly less than during early embryogenesis. DNMT1^{-/-} DNMT3B^{-/-} HCT116 (DKO) is a relatively well-characterized cell line generated from HCT116 which is a colorectal cancer cell with high methylation environment. The DNA methyltransferase activity has been almost abolished in these double-knockout colorectal cancer cells by serial homologous recombinations (9). Even though there is still DNMT3A present and the possible alternative transcript of DNMT1, both of which account for the remnant DNA methyltransferase activity in DKO, about 95% of genomic methylation is lost and DNMT1 contributes more (~20%) to the reduction than DNMT3B (~3%) (8, 9). Simply, the DKO cells do not keep the methylation patterns in HCT116. Therefore, we propose the hypothesis that the loading sites of DNMTs are involved in de novo methylation at CGI-promoters and adjacent regions.

Aim 1. To identify the genes whose CGI-promoters are sensitive to re-expressed DNMT3B in HCT116 DKO cells (DNMT1^{-/-} DNMT3b^{-/-}).

Aim 2. To dissect the common features of remethylated sites caused by re-expressed DNMT3B.

A high-throughput sequencing method is performed to identify sets of sites or regions with distinct sensitivity for remethylation caused by re-expression of DNMT3B or DNMT1 in HCT DKO cells. We expect to observe methylation changes in CGIs vs. non-CGIs, or/and in promoters vs. non-promoters, which have undergone significant demethylation in DKO

cells. The new occurrence of methylated CG sites could, in some sense, reflect the pre-existent epigenetic signatures of the host cells.

Chapter 2

Methods and Materials

Human samples

Patient samples of colon cancer tissues and normal colon mucosa were collected from the Johns Hopkins Hospital. Samples of normal blood were obtained from the University of Texas MD Anderson Cancer Center. All specimens were given informed consent by patients prior to collection of specimens according to institutional guidelines.

Cell culture

The colorectal cancer cell lines, SW48 and HCT116, obtained from the American Type Tissue Culture Collection (ATCC), were maintained in Leibovitz's L-15 Medium and McCoy's 5a medium modified with 10% fetal bovine serum. HCT116 DKO (DNMT1^{-/-} DNMT3b^{-/-}) was generously gifted from Dr. Stephen Baylin at the Johns Hopkins University, and passages were cultured in McCoy's 5a medium modified supplemented with 0.4 mg/mL G418 and 0.05 mg/mL hygromycin B.

DNA extraction and bisulfite conversion

Genomic DNA was extracted from cell lines or tissues using standard methods. Every 10⁶ cell pellet was lysed by 600 µl cell lysis solution (25mM Tris-Cl pH 8.0, 10mM EDTA pH 8.0, 1% SDS) with 1.5 µl RNase A Solution, mixed well and incubated for 30 min at 37 °C. 200 µl of 7.5M Ammonium Acetate was then added and vortexed. After centrifugation, the supernant was transferred to a new tube for phenol-chloroform extraction, and the upper

aqueous phase was further precipitated with isopropanol. The final pellet was dissolved in 50 μ l 10 mM Tris-Cl (pH 8.5) after 70% ethanol washing. DNA concentration was measured by NanoDrop™ 1000 spectrophotometer (Thermo Scientific).

Bisulfite conversion was performed with EpiTect Bisulfite Kits (Qiagen). According to the instruction, less than 2 μ g of genomic DNA was input to every conversion reaction, and the final purified converted DNA was eluted with 20 μ l 10 mM Tris-Cl (pH 8.5).

PCR, cloning and DNA sequencing (Bisulfite cloning/sequencing)

PCR SuperMix High Fidelity (Invitrogen) was used to amplify from gDNA and amplicons were subsequently ligated to PCR4-TOPO (Invitrogen), transformed into 25 μ l One Shot TOP10 chemically competent E.coli (Invitrogen) and colonially selected on LB plates with 100 mg/ml ampicillin. Each correct TA-clone was confirmed by PCR with M13 forward (-21) primer (GTAAAACGACGGCCAG) and M13 reverse primer (TCACACAGGAAACAG CTATGAC), and the amplicon was used for Sanger sequencing. Primers for amplification of the bisulfite-converted DNA are listed in Table 1.

Bisulfite pyrosequencing

All templates used for pyrosequencing were PCR products amplified either using PCR SuperMix High Fidelity (Invitrogen) or *Taq* DNA Polymerase (NEB), which were produced through nested PCR or one-step PCR reactions with one biotin-labeled universal primer (GGGACACCGCTGATCGTTTA) and one normal primer. Please refer to Table 1. for primer informations.

Generally, 1 to 2 μ l of bisulfite-converted DNA was used for PCR reactions and the biotinylated products were subject to sequential steps to immobilize the biotinylated single-strand onto streptavidin-coated beads, including mixing PCR products with beads, agitation, capturing beads, denaturing, washing and annealing sequencing primers. The sequencing process then was automated on the PSQTMHS 96 (Qiagen) equipment, and results were analyzed and output through Pyro Q-CpG Software (Qiagen) software.

Table 1. PCR primers and sequencing primers used for bisulfite-sequencing and pyrosequencing.

Name	Step	Forward	Reverse	Sequencing	Target Sequence	Ta (°C)	Size (bp)
BSP	1°	GGTAYGATGGAAAGTAAGA ATAG	TCCAACCTCRACCAAAATAAACAC	NA	NA	50-58(53)	642
INSGFP-P1	1°	GGTAYGATGGAAAGTAAGA ATAG	TCCAACCTCRACCAAAATAAACAC			53	642
	2°	GTAAGATGGAAAGTAAGAAT AGA	U-AAAAACCCCTATCATCTCT	TAAGTTTATAAATGGAAA TA AGGTTTTTGTTTTATGAA G	GGTYGATTAATATYGTGAAA YGGGGGTTTTAGGAAAGAGGTTAT AYGTTTAT	50-61(58)	128
INSGFP-P2	1°	GGTAYGATGGAAAGTAAGA ATAG	TCCAACCTCRACCAAAATAAACAC			53	642
	2°	U- TAGGAGTAGATAGGGAGTAG G	CTCCACCAAAATAAACAC	AACCCCAAACTAATCCT	CCRCRTTATACAATAAC	50-61(58)	200
INSGFP-P3	1°	GGTAYGATGGAAAGTAAGA ATAG	TCCAACCTCRACCAAAATAAACAC			53	642
	2°	U- TGAGGGAGAAGAAGTTTTAT TAA	CCCCCTAATTTAATTCATTT	TCCCTCCACCTAAAATAC AATAAAAAATTCCTCACC A	RTTCTCCTTCTTATTTA AACRTAAAAACA	54	174
INSGFP-P4	1°	GGTAYGATGGAAAGTAAGA ATAG	TCCAACCTCRACCAAAATAAACAC			53	642
	2°	U- TTTTGGTTTTATTTTAGGAG TTAG	CCCCTTACTCACCATAATAAC	CCAACCCTACTCCCTATC	TACRCCTACRCCAAACCRAACRAC RCRCAC	50-62(58)	329
tetOINS-P1	1°	TTYGGGATTTGGTATTTAGTT G	CCGGAAACTCATACCCTAAACTC			50-58(55)	707
	2°	AGGGATTGGTATTTAGTTG GTAATT	U- ACCCATACCTCCCTATAAACCA TAA	TGGTAATTAATATGTGGT G	AAAYGTTAAGTGTGTATATGGT TATGTGYGTTTAGTTA	50-62(58)	129
tetOINS-P2	1°	TTYGGGATTTGGTATTTAGTT G	CCGGAAACTCATACCCTAAACTC			50-58(55)	707
	2°	TTGAGGTATTTGGGATGGAA ATT	U- CAAATCCTCTAAAAATCCCTATC ACA	GGATGGAAATTTGGTTTTT	TTYGGTTAGATAA	50-62(58)	99
tetOINS-P3	1°	TTYGGGATTTGGTATTTAGTT G	CCGGAAACTCATACCCTAAACTC			50-58(55)	707
	2°	GATAATTTGTTTAGTTAGGA	U-AAACTCATACCCTAAACTC	TTTATTATTGATAGGGAGT AAAT	TYGATATAYGTTTTTTATTA	50-58(58)	390
RILWH-P1	1°	ATGAAGAATTTGTTTAGGGT TAGG	U- AAACATTCAAACCTTCTTACT ACTTA	AGATGTAAGGGTTAGATA TA	YGYGTTGATATTGATTATTG	50-58.9	198
RILWH-P2	1°	GTTTAGTGTGGTGAATTTTG TAG	AACRCRAACCCAAACCCAAC			58-62(58)	487

	2°	GGTGGTTTGAGAGGAGTTTG TAA	U- TATCCACACCCACCTTTAAAAAT C	TTTTTTATAAGGTAGGATG G	GTYGTTAGGTGTGTTAGTYGGGTTT AT	50-62(60)	401
RILUP-P2	1°	TAGTGTGGTGGAATTTTGTA GA	CACTATACTAAAACTACCCCAA			52-62(55)	394
	2°	AGAAGGTTTGAATGTTTGG GGAAGTA	U- TACCCCAACTCAAATACCTCCTC ATAA	TTTTTTATAAGGTAGGATG G	YGYGTTGATATTGATTATTG	50-62(58)	286
RILCEN-P	1°	GTTTAGTGTGGTGGAATTTTG TAG	ACCCTCCCACTTCRAAAAACTCT			50-62(55)	319
	2°	AGGTGTAGATAGTTGGGTTT GGG	U-CACACCCCCACTCAACTCTC	AAGGTTAGAGTAGGATTT AG	GTYGTYGGGGTYGTTYGAAYGYG GGGATTT	50-62(55)	113
RILDN-P1	1°	ATGAAGAATTTGTTTAGGGT TAGG	U- ATCCACAATAACTCCCCCAAATA	AGATGTAAGGGTTAGATA TA	YGYGTTGATATTGATTATTG	58	333
RILDN-P2	1°	TAGTGTGGTGGAATTTTGTA GA	CRAAAACCCCAAACCTCCCTAAA			58	421
	2°	U- TTTGGGGGAGTTATTGTGGA T	CAAAACCCCAAACCTCCCTAAAT	CCCCTAACCCATCTC	CRCRAATCCTTCCRAATCCAC	60	196
P16UPF-P1	1°	ATGAAGAATTTGTTTAGGGT TAGG	U-TAAACACCTCCACTAATCA	AGATGTAAGGGTTAGATA TA	YGYGTTGATATTGATTATTG	58	96
P16UPF-P2	1°	TGAAGAATTTGTTTAGGGTT AGG	ATCTCTCCACCACCCTCC			50	548
	2°	TTTtaggttgagtgtaatg	U-CTCTATAATCCCAACATTCT	Gtagttgggattataggt at	YGTTATTAAGTTTYGTTAATTTTG	50-55	243
P16UPF-P3	1°	TGAAGAATTTGTTTAGGGTT AGG	ATCTCTCCACCACCCTCC			50	548
	2°	YGTtATGTTGGTTAGGTTT	U-ATCTCTCCACCACCCTCCA	TTAGAATGTTGGGATTATA G	AYGTGAGTTATYGTATTYGGATT T	52-58	207
P16UPR-P1	1°	ATGAAGAATTTGTTTAGGGT TAGG	U- ATTCCCTTCCCCCTTATAATTAC	AGATGTAAGGGTTAGATA TA	YGYGTTGATATTGATTATTG	58	166
P16UPR-P2	1°	TGAAGAATTTGTTTAGGGTT AGG	CACACRTACRCCACCATAACCAA			50	501
	2°	GGGGTtGTTGTGAGTTTAA TGAT	U- TACCCCAACCATAACCACTAATT	AATTTTAGTATTTTGGGAA G	TYGAGGYGGGTAGATTATTTGAGG	50-55	306
P16DNF-P1	1°	ATGAAGAATTTGTTTAGGGT TAGG	U- ACCACAAATTCCACTAATAATC C	AGATGTAAGGGTTAGATA TA	YGYGTTGATATTGATTATTG	58	114
P16DNF-P2	1°	TGAAGAATTTGTTTAGGGTT AGG	CTAAAACATAAACTCCAACCAC			58	601
	2°	GGAGGGATTATTAGTGGA TTTG	U- TTCTTTCCACAAAATCTCACTCTA	AGGTTGAGGTAGGAGAAT	YGTtGAATTYGGGAGGTTG	60	370
P16DNR-P1	1°	ATGAAGAATTTGTTTAGGGT TAGG	U- CCCTTATAATCCCTTCACTTTAA	AGATGTAAGGGTTAGATA TA	YGYGTTGATATTGATTATTG	58	466
P16DNR-P2	1°	AATTYGGTTYAGTTAGG GTAT	TCAAACRATTCTCTACCTCAAC C			58	335
	2°	GTGTGGATATTAGGAGGGA	U-	TGATAGAAATTATTAGA	YGGTYGGGYGYGGTGTtTAYGTT	58-62	241

		TTAT	CCACCCCCACTACTTTTATAT	AG	TTGTAA		
EGFP-P	1°	AAGGAAGAAGGTAATTATAA GA	U-CCAACTTATACCCCAAAAT	GGAAGAAGGTAATTATAA GA	TTYGYGTYGAGGTGAAGTTYGAGG GYGATAT	50-52	124
HYG-P	1°	AGGGTGTTAAGTTGTAAGAT T	U-TCCCCCATATAAAATCAC	AATTGTTAGTTGTTTTGTA G	TYGGTYGYGGAGGTTATGGATGYG ATYGTTGYGGTYGATTTT	50-52	174
siteD-P1	1°	GTAAGAGTAGAAATAGGGG AAGA	U-CATCTCCCTTAACTTCACAA	AATTAGTTTTTTTTTATG A	YGGAGTAAGGATGGAGAATAAGG ATATYGGTTTA	50-58	277
siteD-P2	1°	AGAGATGTAGTAGGTAGTTG GATATA	U- AATTAACCAATACTTACATAT TC	TTTTATAGTATAAGTTTAT AGGG	TAYGAAGAATGAATTTGAGAATG TTTGG	50-58	75
siteA-P1	1°	AGTGTATTTGGGGATAAGA	U-TCTTTCACCACTTAATTAATC	TTATGATAGATATTTGGTA A	AYGATGGAAATGT	50-52	78
siteA-P2	1°	AGTTTAGGTTTTTGAATGTA G	U-TTATTTACTCACCTTCTTACAT	TGTATTTAGGTTAATTTAT G	TTGTATTTGAATTYAYGTAA	50-52	243
Zeo-P1	1°	AAGTAGTGATTGGGAAAAT	U-CATTCCCCATTCAAATA	ATTGGGAAAATTTTGG	YGTATTTAATTTAATYGT	50	138
Zeo-P2	1°	GGTTGATTAATTTTTTTTATT TATGT	U- AAACTTCCAAATAATACTTCTC	TTTTTTTTATTTATGTAGA G	GTYGAGGTYGTTTYGGTTT	50-52	120

Name	Forward	Reverse	Sequencing	Target sequence	T _a (°C)	Prod. Size (bp)
RIL-8	GGGAGGGGTTGTGTAGATA	CCCCCACTCAACTCTCAAAA	na	na	52	523
RILEF	GATATATAGGGGATTGGGTTT	U-TCCTCAAAATCCRACTCAAACT	AGTTTTAGGGAGAAGGTT	yGyGTAGGGTTAyGGAATGGG	52	227
			GAGAGAATTAGGTTTTAGAT ATG	GGATyGTTTTGTyGTTAyGTAGTy GTTAG	52	227
RIL-10	TTTAAAGTGGGAAGTGGATGGATT	AAACGCCCAAACCAACCCTCC	na	na	52	353
RIL-D	GTTTATTAGGYGGAAGTTTTAGG	U-AACCAATCCAAACRCACAA	GGTTAAGGAATGGGGTA	TYGYGGGTTGGAGTYGTAGTYG GAG	54	221
RIL-G	GAGTTTTTGGGTTTTTGTAGAG	U-TCCCACCTAAACCTATCTACC	GGTTTTTTTAGATGAGGTTA	GyGGGGAGGyGGGTGyGTGAA	55	160
SFRP1 1°	GGGGAATTTGTTATATTTAAGTATTT	ATACCCCTACTCAACAAAACTACC			56	192
SFRP1 2°	GGGGAATTTGTTATATTTAAGT	ACACCCAAATCTTCCTCTA	AGTAGAAGTAGAAGAATTG T	ATGATYGGTTYGTAYGAGT	58	161
SFRP2 1°	GGTTAGGTTTTTTTGTGTTGTTTAA	U- CTCACCCCTTACTAAAAAAAATTC			57	104
SFRP2 2°	AGGTTTTTTTGTGTTGTTTAAAGA	U- CTCACCCCTTACTAAAAAAAATTC	TGTTGTTGTTTAAAGAAAT T	TTAGGGYGGGYGAGYGGYGG	59	100
SFRP4 1°	GTTGTGGTTGTATATTTTATGAG	CCATTTACACCCTAAAATTC			55	107
SFRP4 2°		U-TTACACCCTAAAATTCCTACAC	TGTGGTTGTATATTTTATGA G	GGGTTYGTAGTYGTYGYGYGYG YGTT	58	108

SFRP5 1°	GTAGAGTTGTGTTAGGGGAGTTG	ACCACTTCCCCTAACCTAAAAT			60	170
SFRP5 2°		U-CCACAAATTCCCCCTAAACA	TGTTAGGGGAGTTGTAGG	GTYGTYGGAGYGYG	60	139
P16-P1	TTTGGTAGTTAGGAAGGTTGT	U-ACCTCCCTACTCCCAACC	TGGTAGTTAGGAAGGTTGTA	TYGYGGAGGAAGGAAAYGGGG Y	55	130
P16-3-2	AGGGGTTGGTTGGTTATTAG	U-CCCTCTACCCACCTAAATC	GAGGGGGAGAGTAGGTA	GYGGGYGGYGGGGAGTAGTAT GGAGTYGGYGGYG	55	255
P16-4-2	GTTTGTAGGGGAATTGGA	U-CCTCATTCTCTTCCTTAACT	GGGAATTGGAATTAGGTA	GYGTTTYGATTTTTYGGA	55	147
P16-4-1ADD	U-GGGGAATATATTTGTATTAGATGG	CCCAACACATCTTACATTTCTT	CCCTTTTATCCCAAAC	RTTCRTAAAT	55	106
P16-4-2ADD	U-TGTGGTGTATGTTGGAATAAAT	TCTCCCAAATAAAAAAATTACAA	AATTACAAAACRTAAACAC	CRCRCCCRACCRCTTCTAAATAA TTTCRAT	55	111
LINE-1	TTTTGAGTTAGGTGTGGGATATA	U-AAAATCAAAAAATTCCTTTC	AGTTAGGTGTGGGATATAGT	TTT/CGTGGTGT/CGTT/CGTTTTT TAAGTT/CGGTTTGAAAAG	56	

Table 2. RT-PCR primers

Name	Forward	Reverse	MGB-Probes (FAM)	T_m (°C)	Amplicon (bp)
tTS	tTS-qF1: GGGCTGAAATCATTAATATTTGGATT	tTS-qR1: CCATCCTCAATGGGTGTATGC	na	60	71
ACTB	4333762F (Applied Biosystems)				
PCNA	Hs99999177_g1 (Applied Biosystems)				
DNMT1b	Hs00945901_g1 (Applied Biosystems)				
DNMT3B1	Hs01003405_m1 (Applied Biosystems)				
GAPDH	ATGGAAATCCCATCACCATCTT	CGCCCCACTTGATTTTGG	GGAGGCCGATCCAG		
EGFP	EGFP-415F: GGGCACAAGCTGGAGTACAAC	EGFP-486R: GATGCCGTTCTTCTGCTTGTC	EGFP-440T: ACAGCCACAACGTCT		71
DNMT3B-V5His	3B-V5His-qF1: CTGAAGGACTACTTTGCATGTGAAC	3B-V5His-qR1: CCGCCACTGTGCTGGATAT	3B-v5His-qP1: AGTGTGGTGAATTC		78
HA-DNMT1	HA-DNMT1b-qF1: GGGATACCCATACGATGTTCCA	HA-DNMT1b-qR1: TACGCGCCGGCATCTC	HA-DNMT1b-qP1: ATTACGCTGCGTCGAC		56

Table 3. PCR primers for subcloning to constructs

Name	Forward	Reverse	T _m (°C)	Amplicon (bp)
Zeocin	Zeo-cdsF1 (AvrII) tggaggCCTAGGcttttgcaaaaagctcccgaggagcttgtATGGccaagttg accagtgccgt (from pcDNA3.1)	Zeo-cdsR1 (BstBI) gaccTTCGAATCAgtcctgctcctcgccac	60	415
tetO	tetO-F (5'EcoRV, BstXI, SacI, AvrII, XhoI--) C inserted cggcGATATCCCAGCACAGTGGctgGAGCTCatgcCCTA GGCTCGAGccggaattgaaga (from pSIREN-tetO-shRNA(P))	tetO-R (5' NheI, AflII--) cggcGCTAGCatgcCTTAAGtatgaccgtacac	60	303
3Motif	3MF-F (5'EcoRV, BstXI--) cggcGATATCCAGCACAGTGGctcttgcctgtcacc	3MF-R (SacI) cggcGAGCTCatccagctacttg	60	140
RILWH	RIL-whole-MF1 (5'EcoRV-universal BT primer) tagtGATATCggcctgcaggccaggtccagcCTGTGCAGACAGG GAGGTG	RIL-whole-MR1 (5'SacI, BstXI,) gcgGAGCTCcgCCACTGTGCTGGGTGACTGAGGT CGTGAATG	60	1397
RILUP	RIL-whole-MF1	RIL-up-MR1(5'SacI, BstXI,) gcgGAGCTCcgCCACTGTGCTGGGGGCTGCCCCA GCTCAGA	60	356
RILCEN	RIL-cen-MF1(5'EcoRV-universal BT primer) tagtGATATCggcctgcaggccaggtccagcATGAGGAGGTATC TGAGCTG	RIL-cen-MR1(5'SacI, BstXI,) gcgGAGCTCcgCCACTGTGCTGGCCTTCAAGCCT CGGATACAC	60	663
RILDN	RIL-dn-MF1(5'EcoRV-universal BT primer) tagtGATATCggcctgcaggccaggtccagcCCCTGTGTATCCGA GGCTT	RIL-whole-MR1	60	471
Tr-pINSL6	INSSubF (NheI and XhoI) Gcatgctagctggccaggtctctcgagcccaaatgga	INSSubR (KpnI) CCGCGGTACCGTCGACTGCAGAATT	60	540

Table 4. PCR primers for confirmation of integration

Confirmation of integration						
For which	For what	Forward	Reverse	Tm	Amplicon (bp)	Break point on vector
HCT/ZEO-1D2	On vector	ZEOSET1F	ZEOSET1R	60	949(none=right)	close to 6090
	Vector-loci	ZEO1D2F: AGGCTGCAGCAGACAGTTTC	ZEO1D2R: AAGTTGGCCGCAGTGTTATC	50	>650	
SW/ZEO-1A5	On vector	ZEOSET6F	ZEOSET6R	60	1192(none=right)	450
	Vector-loci	ZEO1A5F1: CCGTAATGGGATAGGTCACG	ZEO1A5R1: TCCTGGAATGCAGCTCTTTT	60	~600bp	
SW/ZEO-1A1	On vector	ZEOSET3F	ZEOSET3R	60	1383(none=right)	4488
	Vector-loci	ZEO1A1F3: CCCGTCAGTATCGGCGGAAT	ZEO1A1R3: GCAGAAACCTGGTGCCTGAA	55	~700bp	
Integration of pFRT/LacZeo						
Forward		Reverse		Tm	Amplicon (bp)	Location
LacZ-2F: GACGTCTCGTTGCTGCATAA		LacZ-2R: GACCTGACCATGCAGAGGAT		60	468	Inside Lac
Tiling primers						
Forward		Reverse		Amplicon (bp)	Location	Tm(°c)
Zeoset1F: cccctgaacctgaacata		Zeoset1R: gaggcccttctgtcttcaa		949	5165-6113	60
Zeoset2F: cgaggagcaggactgacac		Zeoset2R: ggcagcctatgattggaatg		1251	4166-5416	60
Zeoset3F: gtgcacggcagatacacttg		Zeoset3R: ttcttcacaaagatccatgc		1383	3125-4507	60
Zeoset4F: caccgagtgatcatctg		Zeoset4R: agtaaggcggctgggatatg		1386	2001-3386	60
Zeoset5F: ctggcgtaatagcgaagagg		Zeoset5R: cagcgaccagatgatcacac		1270	758-2027	60
Zeoset6F: gggaaacgcctggtatcttt		Zeoset6R: gacagtatcgccctcaggaa		1192	782-905	60
Zeoset7F: tttgtgtccctgaatgcaa		Zeoset7R: ggcacctatctcagcgatct		1172	5968-7139	60
Zeoset8F: aagccataccaaacgacgag		Zeoset8R: agctcactcaaaggcggtaa		1235	6805-8039	60
Integration of pCDNA5FRT constructs						

Forward	Reverse	Tm	Amplicon (bp)
GFP-BGHF: tatatcatggccgacaagca	GFP-BGHR: gcgatgcaatttctcattt	60	430
INS6-GFPF1: gctaccttcgaccacctgag	INS6-GFPR1: gaacttcagggtcagcttgc	60	425
SV40-HYG-F2: cggcctctgagctattccag	SV40-HYG-R1: gcagctattacccgcagga	60	259
3MF-TETOF1: ccaagtagctgggatgagc	3MF-TETOR1: atgaccgtacacgcctacct	58	335
RIL-TETOF1: tccagtgtggtggaattctg	3MF-TETOR1	59	380 + $\left\{ \begin{array}{l} \text{RILwh 1.8kb} \\ \text{RILup 724bp} \\ \text{RILcen 1030bp} \\ \text{RILdn 839bp} \end{array} \right.$
P16-TetOF1: gccagatatacgcttgaca	P16-TetOR1: gggagtaaactcaagtgaagacg	59	103
P16-TetOF2: tgtacgggccagatatacgc	P16-TetOR2: cttgaggtacctgggatgga	62	237

Table 5. Inverse-PCR primers

Used for	Enzymes	Steps	Forward	Reverse	Tm (°C)
HCT116/zeo-1D2	BamHI+BglII	1	INVB3F: AGCCGGAAAACCTACCGGATT(3207-3227)	INVB3R: CCTCAGGAAGATCGCACTCCA (874-894)	~65
		2	INVB4F: ATGAATGGGAGCAGTGGTGA (4653-4673)	INVB4R: TAACGCCAGGGTTTTCCCAGT (697-717)	~68
SW48/zeo-1A5	EcoRI + MfeI	1	INV-7F : TGACGCGGTGCTGATTACGA (3146-3165)	INV-7R : CGCGTAAAAATGCGCTCAGG (1169-1188)	~65
		2	INVB5F : GCTGGGATTACACATGGCATGGA (3786-3808)	INV5R : TAATTCGCGTCTGGCCTTCCTGT (1055-1077)	~68
SW48/zeo-1A1	XbaI + AvrII	1	INV-7F	INV-7R	65
		2	INVB5F	INV5R	68

Table 6. Chip primer and probes

Name	Forward	Reverse	Probe	Amplicon (bp)
SiteA	TGCTGTATTTGAATCCGACGTT	GCCTTGGATGGAAGAACAAATC	ATCTGAAAGGAATCCCT	69
SiteD	AACTTCACAAACATATACACTCTGCTGAT	GACGGAGCAAGGATGGAGAA	CTAGACCGGTGTCCTT	67
TETO- INSL	GGCGTGTACGGTCATACTTAAGG	GGTGAGATAATTTTGCTCAGTTAGGA	TCTCTCGAGCCCCTTC	85
INSL-GFP	GCCGACCGCCATTGC	ACTGCAGAATTCGAAGCTTACTTAGA	TCACAGGAGATCTGC	79
ACTB	TCCCCTCCTTTTGCAGAAA	CGGCCAACGCCAAAAC	AGCGAGATTGAGGAAGA	78
GAPDH	CTGCTCCTCCTGTTTCGACAGT	TCACCTGGCGACGCAAA	AGCCGCATCTTC	52
RARB	TTGGAAGGAGAACTTGGGATCTT	AGGCTTGCTCGGCCAATC	CTGGGAACCCCCC	85
P16	GGGCGGATTTCTTTTAAACAGA	CGCCTGCCAGCAAAGG	TGAACGCACTCAAAC	57
LINE-1	ACGGAATCTCGCTGATTGCT	CGTTGCCGCCTTGCA	AGCAGTCTGAGATCAA	58
Igf2/H19 ICR	ATTCCACACGTACAGCCGATT	AAGGGAGGCCCGTCTCACT	TGCGCCATCAGGG	65

RNA extraction and Quantitative RT-PCR

Total RNA from every 10^7 cells was extracted with 1 ml TRIzol® Reagent (Invitrogen) as instructed in the manufacturer protocol, precipitated in 200 µl RNase-free isopropanol and resuspended in 30 µl RNase-free distilled water. Further elimination of remnant DNA used The DNA-free™ DNase Treatment Reagents (Ambion) and isolated RNA was measured in NanoDrop™ 1000 spectrophotometer (Thermo Scientific). The integrity of RNA was also tested using 1.5% agarose gel.

We used two-step quantitative RT-PCR to analyze RNA expression levels. The reverse-transcription was performed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) with 2 µg total RNA. The TaqMan® Gene Expression Assay (Applied Biosystems) (20 µl reaction/assay) was used to get Ct values for each target region. The delta-delta Ct normalized to both control genes (GAPDH or β-actin) and control cells, or delta Ct normalized to endogenous controls (GAPDH or β-actin) was calculated to represent the relative quantity of target gene expression in experimental samples. The “minus RT” controls were also included in every RT-PCR. Details of primers are listed in Table 2.

Generation of stable single clones with single integration sites

The general guideline to use the Flp-in system (Invitrogen) is available through the online instructions. According to the instructions, after transfection of pFRT/lacZeo into HCT116 and SW48 and stable selection of single clones with 50 µg/ml zeocin (Invitrogen), PCR and β-gal staining were performed to verify FRT integration. Then, tiling primers (Refer to primers in Table 5) were used to screen out clones with single integration, and the genomic loci were identified through inverse PCR.

β-Gal Staining

The β-Gal Staining Kit (Invitrogen) allows to detect expressed LacZ from pFRT/lacZeo in transfected cells. Reactions were performed for 20% confluent cells in each 12-well plate. Control wells were also included as for absence of LacZ. For testing the second integration by pCDNA5/FRT constructs, control cells were those positive ones from the first integration.

Inverse PCR

Genomic DNA from each single clone was digested with restriction enzymes (EcoRI and MfeI, or BamHI and BglII). Then ligation was performed using T4 DNA ligase (Takara), and the product was amplified with primers targeted at pFRT/lacZeo sequence. The amplified bands were purified (QIAquick PCR Purification Kit, Qiagen), cloned into TOPO XL vector (Invitrogen) and the adjacent sequences were identified through BLAT program. The single integration was further validated through amplifying the conjunction between the insert and the genomic locus. More details of primers are available in Table 4, Table 5 and Table 6.

Establishment of Flpin-tTS host cells

To establish tTS-containing host cells, the tetracycline-controlled transcription silencer (tTS, Clontech) was transfected and selected with 800 µg/ml G418. The presence of functional tTS was verified with RT-PCR and transient transfection of an EGFP vector carrying a tetO element under conditions of Dox (+) (2 µg/ml) or Dox (-). The clones with highest Dox (+) GFP/Dox (-) GFP ratio were selected as tTS-containing host cells.

Flow cytometry

All cells for flow cytometry were to test GFP expression. Log-phase cells from each 12-well or 6-well culture were trypsinized and suspended in 1-ml media. When dead cells needed assessment, propidium iodide (f.c. 1 µg/ml) was added to do staining. FACSCalibur (BD Biosciences) was manipulated as the manufacturer's protocol.

Construction of transgenes, homologous recombination and clonal selection

The transgenes consist of a reporter EGFP, a pINSL6 promoter, a tetO element, the fragment of interest and the backbone from vector pCDNA5/FRT (Invitrogen). All fragments were amplified using PCR SuperMix High Fidelity (Invitrogen), digested with certain restriction enzymes (NEB), subcloned to construct transgenes and sequenced for correct sequences (see primers in Table 3, Table 4 and Table 5).

To make homologous recombination mediated by FRT sites, pCDNA5/FRT constructs and POG44 (for flippase expression) were co-transfected into characterized Flp-in host cells with 1:9 ratio (w/w) and stably selected with hygromycin B (150 µg/ml) (Invitrogen) for 10 days. Correct single clones were confirmed by PCR amplification of inserts, zeocin-resistance (50 µg/ml, 10 days), and β-gal staining.

Drug treatment

For inhibition of DNA methylation and reactivation of GFP, we used 200 nM 5-Aza-2-deoxycytidine (DAC, Sigma) and/or 800 nM trichostatin A (TSA, MP Biomedicals). Cells were split 24 hours before experiment, and given one of the following treatments. (1) DAC was given every day for 96 hours, and media were replaced every day; (2) TSA was added at the last 24 hours; (3) combined treatment of the above DAC and TSA. Flowcytometry

(FACSCalibur, BD Biosciences) was performed to detect GFP expression as instructed by the manufacturer.

ChIP analysis

The procedures of chromatin immunoprecipitation (ChIP) assays were adapted from the online protocol. 1×10^6 cells were used for each immunoprecipitation. Antibodies (10 μ g) used are IgG (ab6709, Abcam), H3 (ab1791, Abcam), H3K4me3 (07-473, Millipore), H3K9ac (07-352, Millipore), H3K27me3 (07-449, Millipore), and H3K9me3 (ab8898, Abcam). Following immunoprecipitation, qPCR was performed in 7500 Real-Time PCR System (Applied Biosystems) to get Ct values. All the enrichment of histone marks was normalized to H3 (percent of H3) and the nucleosome density measured by the H3 occupancy (percent of input) was calculated against the 1/50 input control. Primers and probes are listed in Table 8.

Protein extraction and western blot

$(5 \sim 10) \times 10^6$ of cells were lysed in cold 1 ml RIPA buffer (Thermo Scientific) supplemented with protease inhibitor cocktail tablet (Roche) and PMSF (Fisher). The supernatant was transferred and aliquoted into new tubes. Each aliquot of cellular protein (50 \sim 100 μ g) was boiled for 5 min in sample-loading buffer (Biorad) and then separated through SDS-PAGE gels (Biorad) depending on the sizes of target proteins. Wet-transferred onto PVDF membranes (Biorad), the samples were either blocked with 5% nonfat milk in PBST or 1% BSA in PBST depending on the chemical properties of the antibodies. Primary antibodies used are anti-DNMT1 (Abcam), anti-DNMT3B (IMG184, Imgenex), anti-HA (Sigma), anti-

His (Invitrogen) and anti-ACTB (Sigma). The secondary antibodies were HRP-conjugated anti-rabbit (Amersham Pharmacia Biotech) or anti-mouse (Biorad). Bands development used Pierce ECL Western Blotting Substrate (Thermo Scientific) and exposed on X-Ray imaging film (Fisher).

DREAM library preparation

High quality of genomic DNA was confirmed by running 500 ng in 0.8% agarose gel before any further treatment. 5 µg of such unbroken genomic DNA was mixed with 0.2 ~ 0.5 ng spiked DNA for digestions with restriction enzymes. The spiked DNA comprised a set of PCR-amplified fragments from non-human DNA which were *in vitro* methylated at CpG sites by the *M.SssI* methylase to several magnitudes of 0% ~ 100%. The DNA mixture was first digested by *SmaI* (Fermentas) for 3 hrs followed by overnight *XmaI* (NEB) digestion. In the second day, column purified products had to be heated at 65°C for 3 min, and cooled down to free concatenated overhangs. Then Klenow fragment (3'→5' minus) (NEB) was added to the above digested DNA sample for filling the overhangs and adding “dA” to the 3'-tails. The second purification using QIAquick MinElute PCR purification kit (Qiagen) was followed by ligation of 1:50 diluted Solexa Paired Ends Adapters (PEA) (25 µM). Afterwards, selected products with the size of 250 ~ 500 bp using 2% agarose gel were divided into 2 even slices and purified separately with Qiagen Gel Extraction kit. In order for the cluster station to create clonal clustering of single molecules, PCR amplification (18 cycles) with the Solexa paired-end PCR primers was required (iProof HF master mix, BioRad). Products were further purified with Agencourt AMPure beads and two size ranges were confirmed in 2% agarose gel, or more accurately on Agilent 2100 Bioanalyzer using

Agilent DNA 1000 kit, which would tell the primer-dimer contamination as well. NanoDrop 1000 was used to determine the concentration.

DREAM library validation — Pyrosequencing and TA cloning

Two other quality validation steps were also needed. Diluted DNA from each library was amplified with PCR and examined for pyrosequencing. The target regions were within the component fragments of the spiked DNA which had ranges of methylation percentages (0% ~ 100%). The good quality was demonstrated by the close match between pyrosequencing results and the expected percentages.

The other validation is to make TA clones using 2 µl of each library. 12 clones with correct insert size were sent for sequencing and the ratio of XmaI/(SmaI + XmaI) would help estimate the average methylation of each sample. How to perform pyrosequencing and TA-cloning has been mentioned previously.

Data processing and statistics

All next-generation sequencing datasets are received after pre-analysis by DNA sequencing core facility in M.D. Anderson Cancer Center. Excel, Graphpad and R-program are major softwares used in data processing and statistical analysis.

Chapter 3

Establish a site-specific integration system with enforced transcriptional repression

Introduction

How *de novo* DNA methylation is triggered still remains a question. There are basically two strategies to study this question. One is to mimic the process of normal embryonic development or malignant progression induced by certain inner- or extra-cellular conditions. The other is to observe *de novo* methylation events in transgenes. Currently, most investigations were focused on the differentiating course of ES cells. Methylation seeding on APRT gene was observed in mouse embryonal carcinoma cell (65-68), and targeted *de novo* methylation in the enhancer and promoter regions of Oct-4 was detectable 2 days after the wild-type ES cells induced with retinoic acid (RA), followed with throughout spreading into its regulatory region, although Oct-4 does not have a CGI in its promoter (110). Compared with ES cell, seeding events are still barely studied in somatic cells, probably because these events require longer observation window than usually expected. For example, it took as many as 87 passages for P16 promoter (with a CGI) in HCT116 DKO (DNMT1^{-/-}; DNMT3B^{-/-}) to get fully remethylated (111). Nonetheless, remethylation of P16 indeed indicated that cancer cells still keep active *de novo* methylation.

There are disadvantages of both strategies. For the induction way, when the cis- or trans-elements need investigation, it would be complex to do multiple gene deletions. While for the transgene way, transfected genes generally remain unmethylated, even if introduced into cells where the endogenous ones are methylated. For the glutathione-S-transferase gene (GSTP1), hypermethylation did not occur in exogenous GSTP1 shuttles and both silencing

by mutations of transcription factor binding sites and pre-methylation were required for high levels of DNA methylation to be seen after spreading (112). The experiments using mixed stable clones to observe de novo methylation were neither ready to get replicable changes. One problem in all these experiments is the variability associated with insertion site effects. Thereafter, we generated a site-specific integration system in order to test different aspects of methylation seeding and spreading as relevant to tumor-suppressor gene silencing in cancer. By constructing two different integration sites in two colorectal cancer cell lines (SW48 and HCT116) due to their high background of DNA methylation (113, 114), and constructing two comparable groups of host cells with different local repression strength for each case through introducing a defined regulator, the tetracycline-controlled transcription silencer (tTS, a tetR DNA-binding domain fused with a KRAB domain), the trans-regulatory factors caused by transcriptional repression, genomic loci and/or cell line could be tested. By controlling the fragments of transgenes, the cis-sequences could be examined, too.

Results

Construction of flp-in host cells

Our aim was to evaluate the possibility of de novo DNA methylation in cancer cell lines through exogenous sequences without in vitro methylation. For this aim, an Flp/FRT-mediated integration system was utilized to make each transgene integrated into the same single genomic locus due to our concern for the position effects arising from multiple or non-ubiquitous chromatin environment. The basic steps are schematically described in Figure 7 where each step is indicated with Roman numerals (I, II and III).

In step I, a vector (pFRT/LacZeo) with an FRT site was introduced into cell lines and integration was screened by constructive expression of the reporter gene (LacZ) and presence of a sequence close to FRT in stable single clones (Figure 8). Two colorectal cancer cell lines, SW48 and HCT116 were chosen to make transfection because of their higher methylation background as shown before (113, 114).

The presence of FRT is not sufficient to indicate single integration. In order to identify clones with possible single integration, a screening method was designed by using tiling primers for pFRT/LacZeo. The assumption was that if single integration exists, there is one break point on the vector, so clones which show negative amplification with only one primer set are much definitely inserted into a single locus. Nonetheless, clones with negative amplification for multiple primer sets, especially those sequentially located on the vector, are also possible ones with single integration, although less likely and could not be excluded absolutely. Because primer sets are overlapped at both ends, there are cases that break is inside the overlap or only partial vector is inserted. Figure 9 shows the primer sets and representative results for some primer sets. SW48/Zeo-1A5 and SW48/Zeo-1A1 were later found as clones with single integration.

We characterized the genomic locus for clones with possible single integration through inverse PCR. Restriction enzymes with compatible ends were used to increase the frequency of digestion and generate shorter fragments for the convenience of cloning and sequencing. Figure 10 shows the general process of this method and two clones with identified sites. The site for HCT116/Zeo-1D2 was on Chr3q13.31 inside one intron of ZBTB20, and the break point on the vector was also defined by aligning the sequencing results with both the vector and the human genome (Hg18) (Blast2, NCBI and BLAT, UCSC). Therefore, two validation PCR reactions could confirm one break on the vector and the correct genomic locus. One used the tiling primer Set3 and the other used primers targeted at the linking region of the insert and the genome, so negative amplification for the former and positive for the latter proved the above finding. With the same method, another single clone, SW48/Zeo-1A5 has the integration within an exon of CD36 on Chr7q21.11. SW48/Zeo-1A1 also has the identified site on Chr2p25.1 (not shown). For convenience, we rename SW48/Zeo-1A5 as SW48A and HCT116/Zeo-1D2 as HCT116D in following results (Table 7).

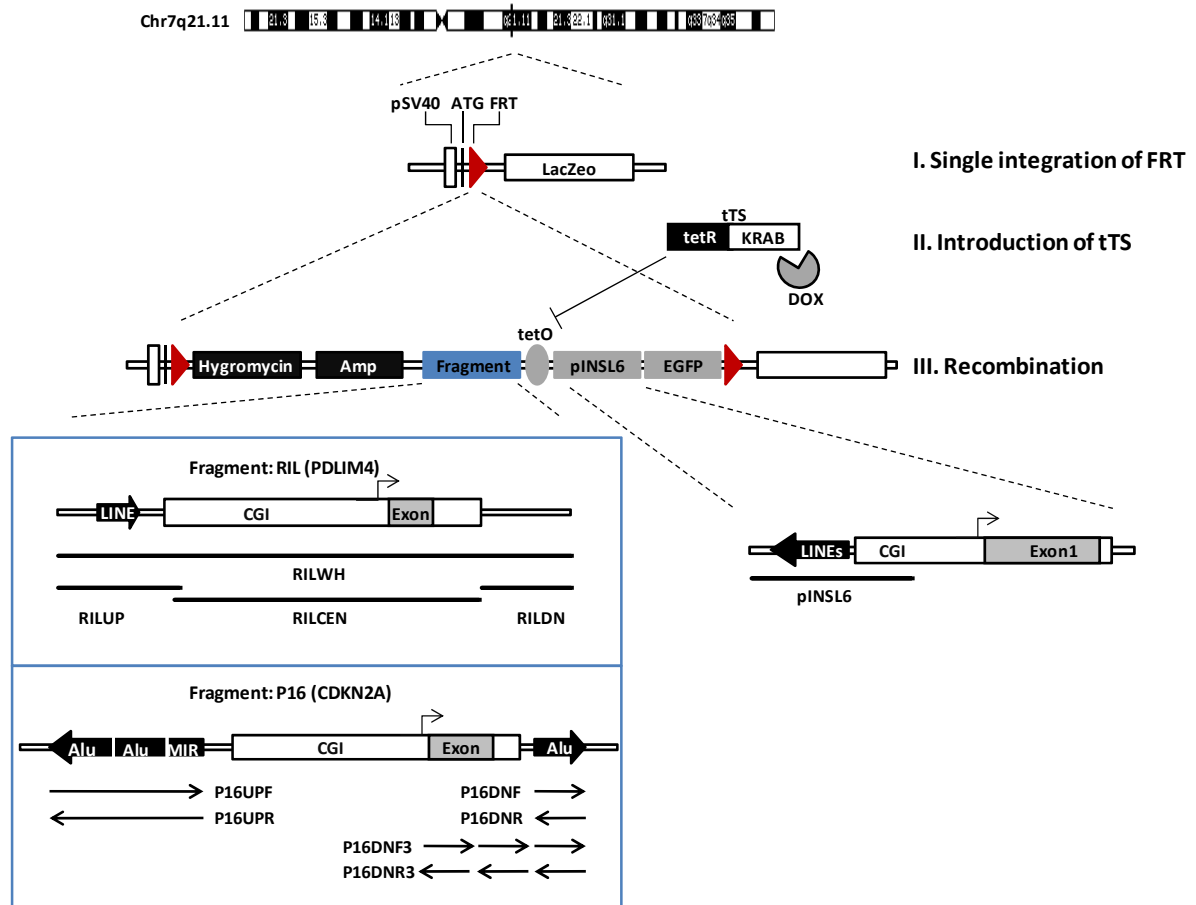


Figure 7. Schematic description of establishing a site-specific integration system. The entire system was established using three steps. I, single integration of FRT; II, expressing tTS to generate tTS-containing host cells; III, FRT-mediated homologous recombination of transgenes with subcloned fragments derived from the RIL or P16 promoter. After I and II, for each genomic locus in each cell line, a parallel of flip-in host cells were generated (with and without tTS) so as to examine the influence of local transcriptional repression on DNA methylation. pINSL6 refers to the INSL6 promoter, which contains a CGI and two short LINE elements upstream.

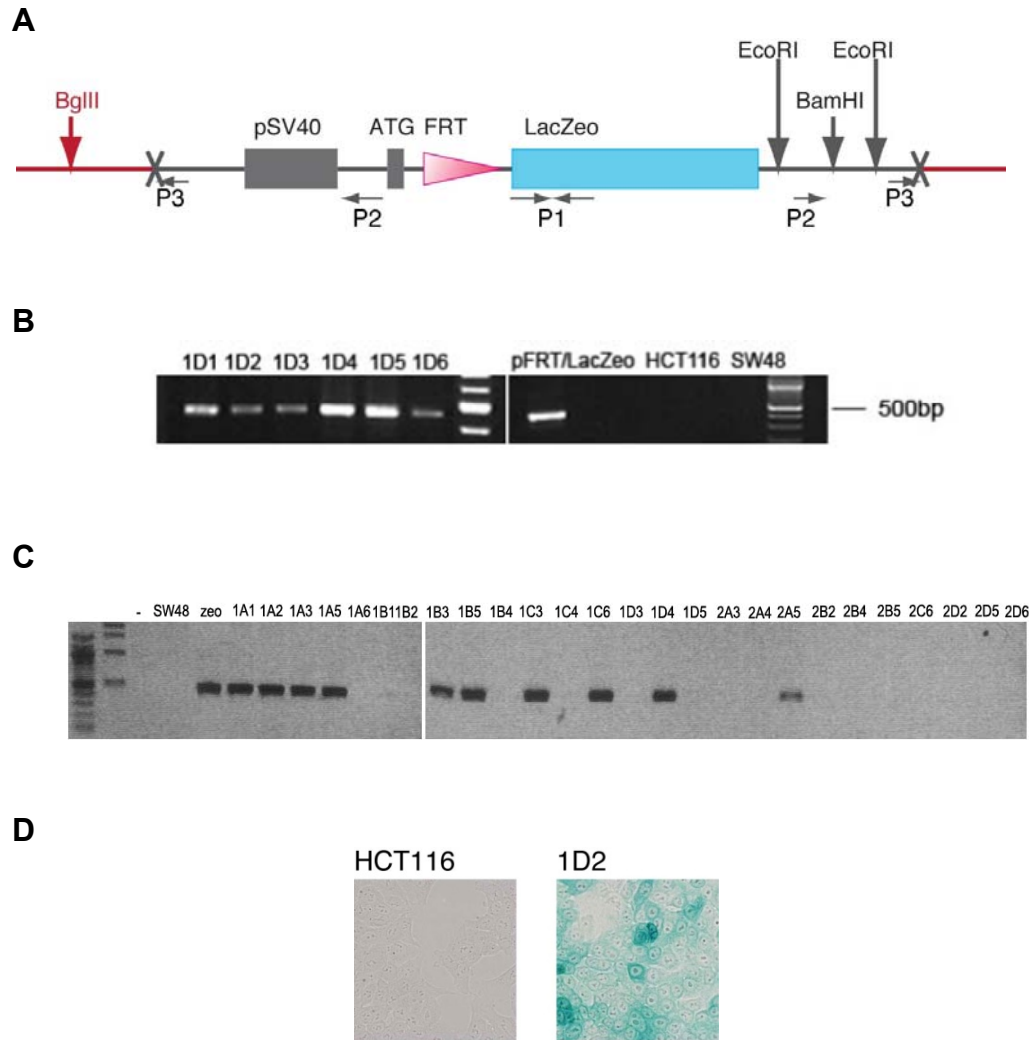
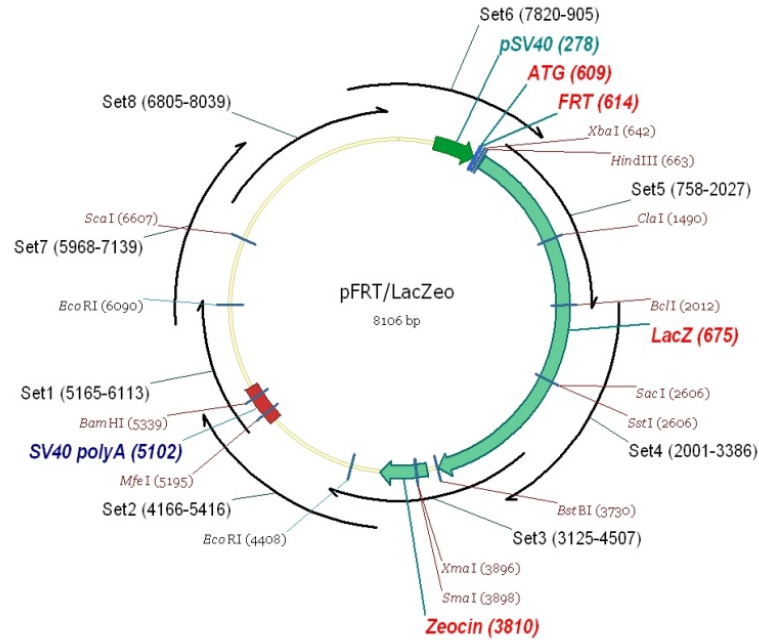


Figure 8. Transfection of pFRT/LacZeo into cell lines. (A) Linear depiction of the vector pFRT/LacZeo. Primer P1 is used to PCR one region close to the FRT site. Restriction sites for enzymes (BglII, BamHI and EcoRI) and primer P2 and P3 are also indicated for the subsequent inverse PCR. (B) and (C) Bands amplified using primer P1 to demonstrate the presence of the FRT after stable transfection in single clones selected under zeocin. (D) β -gal staining results showing the expression of LacZ in the above selected clones. All these clones in B and D are from HCT116 as representatives and those in C are from SW48.

A



B

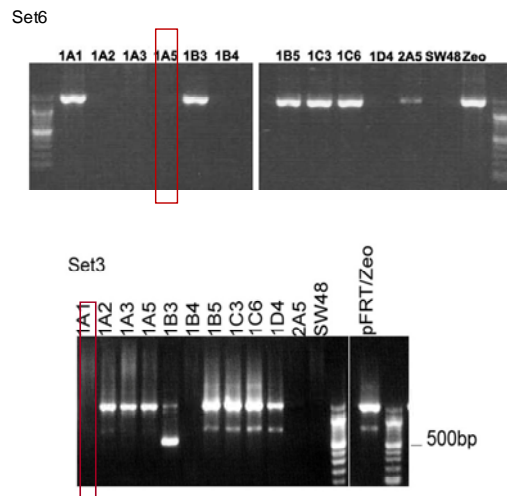
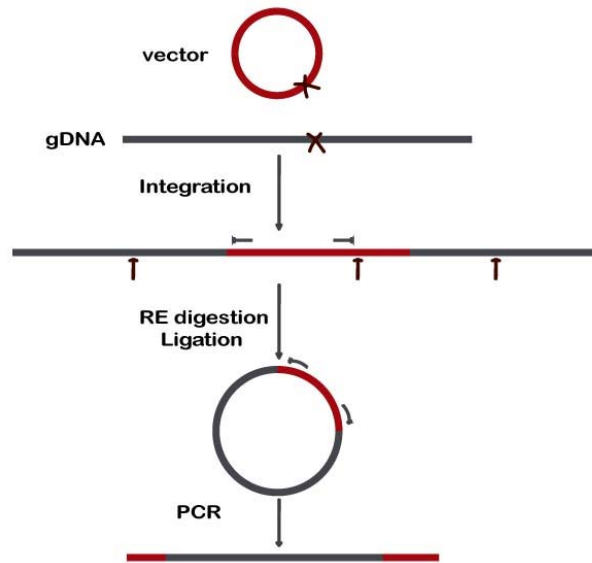


Figure 9. Tiling primers to screen out possible single clones with single integration sites.

(A) A vector graph of the circular pFRT/LacZeo. The tiling primers are shown as curved arrows with names of Set1 to Set8 that are overlapped at both ends. Some restriction sites are also indicated for inverse PCR. (B) Examples showing PCR results using primer Set6 and primer Set3 in transfected single SW48 clones. The negative amplification in only one primer set would indicated possible single integration.

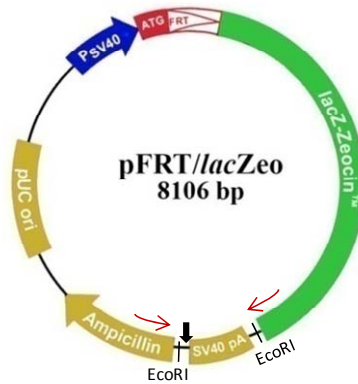
A



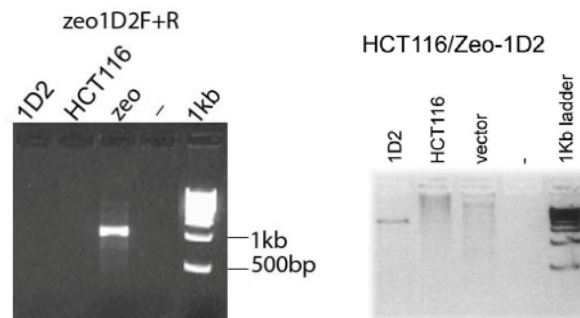
B



C



D



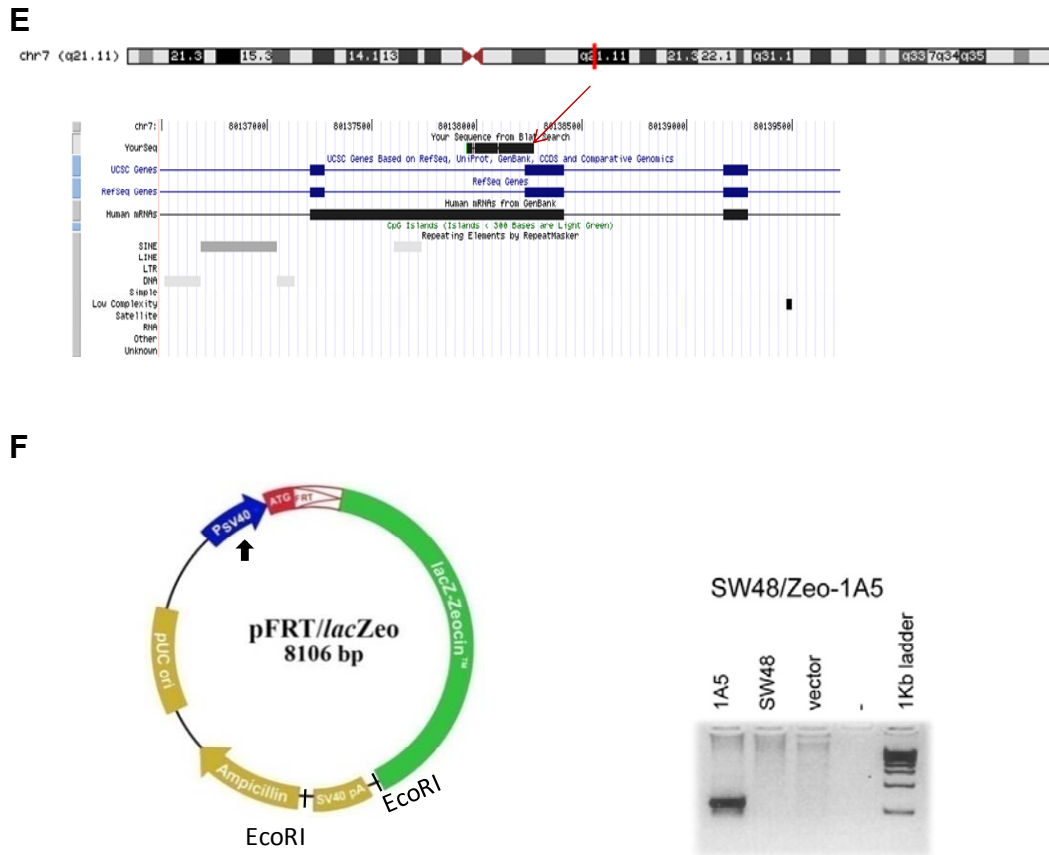


Figure 10. Identify and validate the single integration site of flp-in host cells. (A) Graphical description of inverse PCR. (B) and (E) show the genomic loci for HCT116/Zeo-1D2 and SW48/Zeo-1A5. (C) Validation of the break point on the pFRT/LacZeo vector in HCT116/Zeo-1D2 using one tiling primer Set3. (D) Validation of the insertion to the indentified genomic site by amplifying the conjunct of the vector and the site. The appearance of one band for HCT116/Zeo-1D2 confirmed the correct integration. (F) Similar validation as in (D), but for SW48/Zeo-1A5.

Table 7. Characterization of the single integration site in flp-in host cells.

Single clones	Location	Position(hg18)	Nearby genes	Orientation	Elements
HCT116D	Chr3q13.31	115911000- 115913000	ZBTB20	+	Intron
SW48A	Chr7q21.11	80138271	CD36	-	Exon
SW/zeo-1A1	Chr2p25.1	8823363	KIDINS220	-	Intron, SINE

Heterochromatin environment of the insertion sites

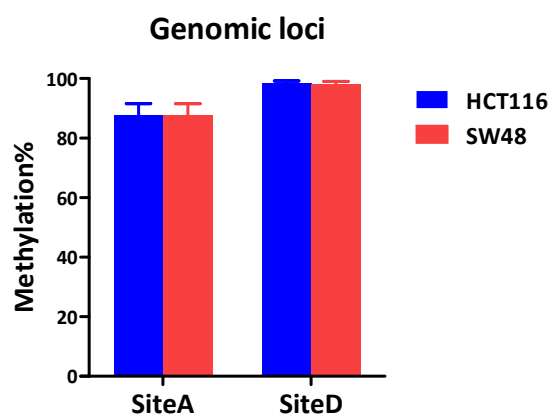
Constructing a single integration site for every transgene would equalize the experimental background, yet it does not rule out the position effect variegation (PEV) in transgene methylation which has been suggested by controversy from prior trials. We would like to use the locus with permissive environment for de novo DNA methylation to occur, and the association of DNA methylation with heterochromatin suggested the locus should not be close to an actively transcribed region. Therefore, we examined the epigenetic status around the identified loci including DNA methylation and histone marks (H3K4me3, H3K9ac and H3K9me3). SiteA refers to the locus in SW48A and SiteD is in HCT116D. Both loci were within inactive chromatin regions that had very high levels of DNA methylation (SW48, $87.6\% \pm 4.00\%$; HCT116, $98.4\% \pm 0.80\%$, mean \pm SEM) (Figure 11) and lacked active histone marks examined here (H3K4me3 and H3K9ac) in comparison to enriched levels of active marks for ACTB and GAPDH (Figure 12). Although H3K9me3 was not enriched around SiteA or SiteD, they may have other marks and still could be called as heterochromatin. The H3 occupancy at SiteA and SiteD was also examined by normalization to the corresponding input, turning out that the absence of marks was not due to unloaded H3.

It is also necessary to know if any sequence of pFRT/LacZeo could bring any significant DNA methylation to the area. Pyrosequencing showed that the inserted LacZ region did not gain much DNA methylation in any clone (SW48A, $5.4\% \pm 1.75\%$; HCT116D, $4.7\% \pm 1.65\%$), precluding the possibility of methylation recruitment caused by sequences from the first construct (Figure 11). Consequently, there are two flip-in host cells of characterized local properties (copy number, location, and transcriptional activity) available for subsequent homologous recombination.

Generation of flp-in/tTS host cells

Given that the integrated sequences contain potential promoters, we wished to introduce further transcriptional silencing into this system. We therefore inserted a defined regulator, the tetracycline-controlled transcription silencer (tTS, a tetR DNA-binding domain fused with a KRAB domain) to generate a parallel group of host cells (SW48A/tTS and HCT116D/tTS), where we could get the effects of strengthened repression. After transfection of tTS to SW48A and HCT116D, expression of tTS was analyzed for each stable single clone selected with G418 in comparison to the internal control β -actin. Since Dox can prevent tTS from binding the response element (tetO), to select the clones with functional tTS, a tetO-containing EGFP construct was transiently transfected into each tTS-expressing clone under two parallel conditions, Dox (+) and Dox (-). After 72 hours, we used flow cytometry to test GFP expression, and the ratio of Dox (+) to Dox (-) was calculated for each clone. The clones with highest ratios were chosen as flp-in/tTS host cells. Thus, for each integration event, there would be two kinds of local conditions in regard to transcriptional repression (Figure 13).

A



B

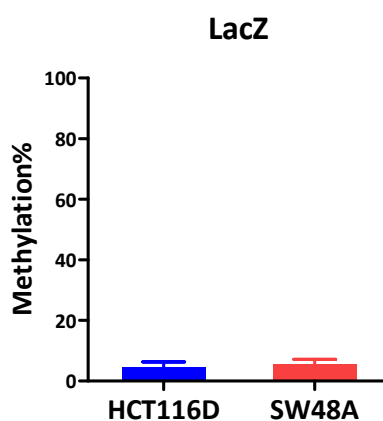
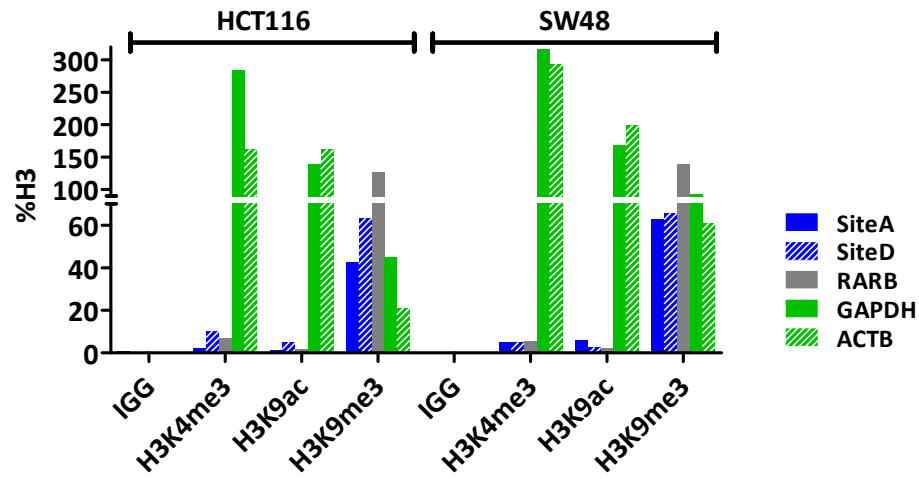


Figure 11. DNA methylation of the environment for homologous recombination. (A) Pyrosequencing results showing the endogenous methylation of the characterized integration sites (SiteA for SW48A and SW48A/tTS; SiteD for HCT116D and HCT116D/tTS). Bisulfite-treated genomic DNA from SW48 and HCT116 was used as template. Two separate pyro-assays were designed for 1-kb range to each site and average methylation was calculated. Error bar, SEM. (B) Methylation of the LacZ in flp-in host cells (SW48A and HCT116D). Pyro-assays were designed inside the coding sequence of LacZ.

A



B

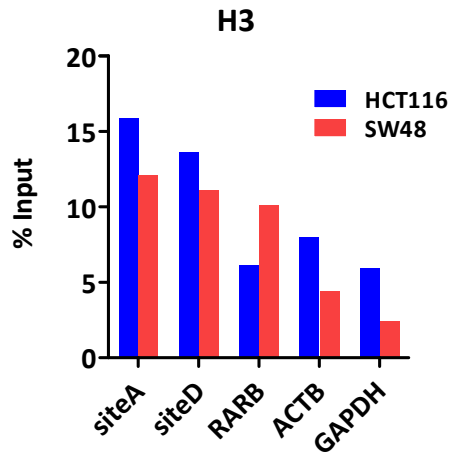
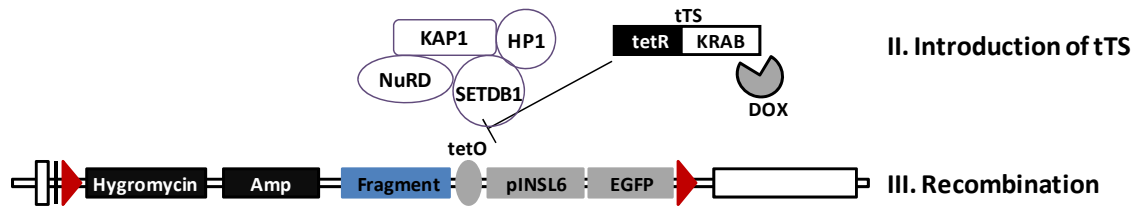
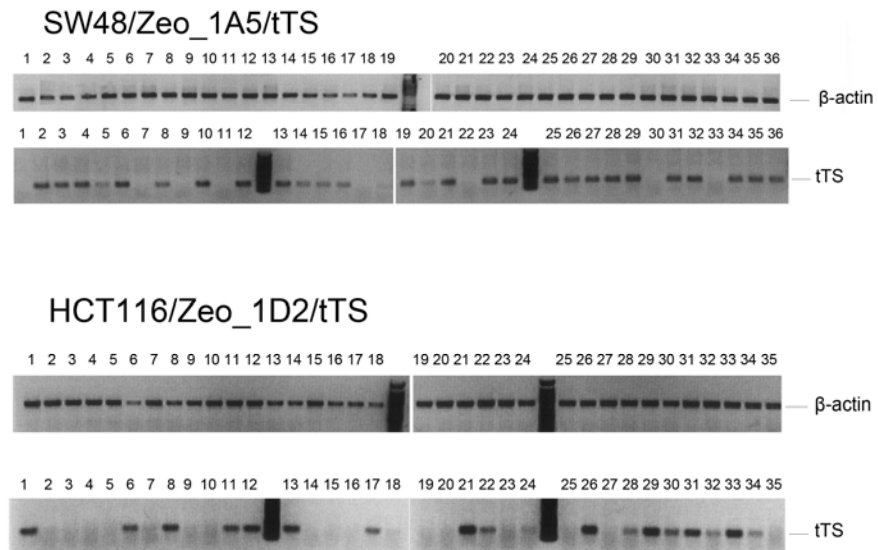


Figure 12. Chromatin signatures of the integration sites in SW48 and HCT116. ChIP-qPCR results showing (A) the enrichment for histone marks (H3K4me3, H3K9ac and H3K9me3) and (B) H3 occupancy at the integration sites. ACTB and GAPDH as the control regions for active marks, and RARB as the negative control region. In (A), average Ct-values of histone marks from technical duplicates were normalized to those of H3. In (B), normalization was calculated to the input.

A



B



C

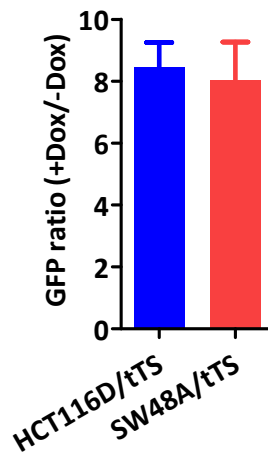


Figure 13. Generation of flp-in/tTS host cells. (A) Graphical description of the mechanism for a transcription repressor (tTS) to induce formation of robust local heterochromatin. The KRAB domain is responsible to bind KAP-1 which recruits a modification complex composed of histone methyltransferase, histone deacetylase, chromatin remodeling factors and heterochromatin protein-1. TetR is the peptide to bind the tetO element present in a promoter. Hence, tTS could be inducible to control repression under Dox treatment. (B) Stable selection of single clones of flp-in/tTS host cells by RT-PCR. RNA expression of tTS and internal control (β -actin) were analyzed for every single clone selected by G418. (C) Flowcytometry to validate the functional tTS in SW48A/tTS and HCT116D/tTS. Parallel wells of stable single clones of tTS-expressing flp-in host cells were transiently transfected with a tetO-containing EGFP construct under the presence or absence of Dox treatment. GFP fraction was measured 48hr after transfection and the clones with the highest ratio of Dox (+) to Dox (-) were used as tTS-containing flp-in host cells.

Discussion

The goal of this project is to generate a system that could investigate several aspects engaged in de novo DNA methylation, specifically on CGI-promoters. The aspects we are interested in are (1) repetitive elements adjacent to CGI-promoters that are frequently hypermethylated in cancer cells; (2) the extent of transcriptional repression in regulating DNA methylation; (3) the effect of genomic position/cell line on DNA methylation. Aspect (1) would be represented in organization of the transgenes, aspect (3) needs construction of a model system, while aspect (2) would require both of the above manipulations. Thus, we first established such a system to complying with the requirements for aspect (2) and aspect (3), of which the genomic position has been characterized and an ectopic transcriptional repressor, tTS, is expressed.

As known, transgene expression is influenced by sequences proximal to the integration site which is usually random in the genome. The influence could be either to the positive or negative trend, for example, the integration to a heterochromatin would generate silencing in a proportion of cells expected to express the transgene (115). DNA methylation of transgenes probably is subject to the same influence of position effect. Proviral transfection into mice could be inactivated by de novo DNA methylation at a later stage (116), however, the study of the glutathione-S-transferase gene (GSTP1) gene only observed sporadic methylation in GSTP1 shuttle vectors (112). A further study compared random integrated BLK sequence with homologously recombined one, and the de novo DNA methylation did show the influence out of position effect.

Therefore, in order to get reliable results, all the transgenes have to be examined under the same vicinity, and a site-specific integration system using Flp/FRT was utilized. In this

system, transgenes are transfected into host cells containing a FRT site which would mediate homologous recombination in the help of the flippase (Flp). For the flp-in host cells, we characterized the single presence of FRT and its integration site. Inverse PCR itself may not determine the copy number of the insert, but combined with tiling primers and validation experiments, we had identified two host cells from SW48 and HCT116. Due to the concern of position effect, we would like the integration site to reside in a genomic area with the property of heterochromatin. So characterization of DNA methylation and histone enrichment were examined further. Both sites in flp-in host cells (HCT116D and SW48A) have average methylation levels above 80%, and lack the most common active histone marks (H3K4me3 and H3K9ac), suggesting these sites are not actively transcribed regions at least. That is to say, there probably would be no negative effect from the vicinity on transgene methylation if these two host cells are used.

However, we were still not sure to what extent these loci would influence transgene methylation. To ensure robust heterochromatin adjacent to transgenes, in parallel to the flp-in host cells, we constructed flp-in/tTS host cells (SW48A/tTS and HCT116D/tTS) by stably expressing a defined transcriptional repressor, tTS into the above flp-in cells. The tetracycline-controlled transcription silencer (tTS) is a fusion protein composed of the tetR DNA-binding domain and a KRAB domain. It is usually used in inducible expression system, and here we are employing its sequential recruitment of the H3K9-specific histone methyltransferase (e.g. SETDB1), HP1, and the histone deacetylase (HDAC)-containing complex via KRAB-KAP1 cooperation (117, 118). Hence, even over a range of euchromatin, a highly compact heterochromatin patch can be generated and maintained for quite a few siblings. This provides an environment permissive for DNA methylation to occur. There is

another place to permit DNA methylation is the promoter used in transgenes, which would be discussed in the following Chapter.

Overall, we first established a site-specific integration system with enforced local repressive status to allow DNA methylation to occur.

Chapter 4

Repetitive elements and strengthened local heterochromatin enhanced de novo methylation in a promoter-CGI

Introduction

The triggers for de novo DNA methylation are not defined clearly so far. If DNA methylation does not happen globally which is suggested by unequal accessibility of DNMTs to genome regions, then local sequences have to signal for DNA methylation directly or indirectly (2, 23, 24). Trans-regulatory elements out of various origins are suggested in determining the preference of DNMTs. The biochemical property of DNMTs might predispose their tendency of binding specific motifs, but transcription factors, enzymes for histone modifications, chromatin remodeling factors and even RNAi are a lot more likely to set up flexible patterns of DNA methylation under various physiological and abnormal conditions. Then, DNA sequences would signal indirectly and recruit complexes of regulators prior to stable cytosine methylation. In this sense, the sites susceptible to DNA methylation would also possibly be signals to set up this epigenetic memory.

We have established the site-specific integration system to transfect constructs with sequences out of highly methylated genomic areas, especially those around CGI-promoters. DNA methylation patterns of such a few CGI-promoters (p16INK4A, MLH1, MGMT, CDH1, etc) have been mapped before showed multiple repetitive elements adjacent to a “trench” of methylation in CGIs, and they are consistently methylated in both normal and malignant cells. Presumably, repetitive elements have something to do with preferential DNA methylation. DNA repeats have long been suspected to be involved in chromatin nucleation, at least under certain circumstances in yeast, filamentous fungi and plants (119, 120). Moreover, Turker M.S. and colleagues identified two upstream B1 repetitive elements

as cis-signals for de novo methylation of mouse Aprt gene on the X-chromosome (65-68). Therefore, we investigated the methylation sensitivity of fragments from CGI-promoters including repetitive elements as constructed in transgenes.

Another concern is related with local status permissive for DNA methylation. It is assumed here that transcriptional repression is predominant to DNA methylation as shown before (83, 84). We have selected single integration sites for transgenes and imposed a repressor to construct a comparison as to repression strength. As well, we also used a promoter which is silenced endogenously, since promoters such as CMV and SV40 probably would tend to get strong transgene expression and interfere with DNA methylation. By combining several aspects into our study system, we were able to observe methylation seeding events and examined the effect of repression and repetitive elements on methylation spreading into the promoter-CGI.

Results

Mapping the methylation patterns of RIL (PDLIM4) promoter

RIL (PDLIM4) was originally identified as a new gene in normal rat fibroblasts and downregulated in H-ras transformed derivatives (121). This gene, later mapped to human chromosome 5q31.1, contains a LIM domain and a PDZ domain, implicating its possible interaction with other LIM proteins or itself in cytosolic trafficking. MCA/RDA is a powerful tool that allows PCR amplification and cloning novel CpG islands of genes that are hypermethylated in cancer. Previous results in our lab have found frequent hypermethylation of RIL in multiple tumor cell lines (almost all colon cell lines and some leukemia cell lines) and primary tumors (colon and leukemia/lymphoma), and its close association with loss of expression, which could be reversed by 5-aza-dC treatment in methylated cell lines. Re-expression of RIL led to suppression of growth and clonogenicity in vitro and sensitivity to apoptosis, which suggested RIL as a candidate TSG (109).

The CpG island (-347bp to +427bp from TSS) of RIL covers its promoter region, the first exon and part of the first intron. Long and short alleles of RIL have been characterized in RIL promoter from sequencing of large populations of normal (blood and colon) and tumor (colon, AML and CML) samples. The polymorphism is located near a CGG repeat sequence adjacent to the transcription start site. The long allele has a 12-bp insert (CGGCGGCGGCTC) and a substitution of T to G 3 bases upstream of the insertion site in the short allele. The insertion of the long allele was shown to bind with Sp1 and Sp3 transcription factors in vitro, contributing to the differential methylation of the two RIL alleles in tumors as a protective machinery (122).

We further compared the methylation of the regions surrounding the CGI in normal and tumor cells to find which area the methylation is spread from. We bisulfite-sequenced the region of -722 to +630 from TSS (25, 23 and 18 CG sites with three separate reactions) in 3 leukemia cell lines (CEM, KG1 and HEL), 4 normal blood samples, and 6 colon tumor adjacent samples. Comparing the methylation levels across the whole region, we could see two clear boundaries, upstream and downstream of the CGI, showing drops at DNA methylation. The upstream 11 CpG sites (-689 to -413) and downstream 17 CpG sites (+238 to +586) are kept densely methylated even in normal blood (average methylation density, $82.2 \pm 11.6\%$ for upstream and $73.6 \pm 17.7\%$ for downstream) and tumor adjacent colon tissues (average methylation density, $80.5 \pm 14.6\%$ for upstream and $76.2 \pm 12.1\%$ for downstream), compared with those in the CGI ($17.6 \pm 10.0\%$ in blood and $20.2 \pm 15.0\%$ in tumor adjacent colon tissues), indicating the possible protective machinery at the boundaries, and the candidate de novo methylation originated from the surrounding areas on the other hand (Figure 14). This methylation pattern of RIL promoter is very similar with those found in genes like p16, VHL and CDH1. The common feature also involves repetitive elements juxtaposed to the promoter CpG island, as in RIL a LINE (-629 to -510) located within the consistently methylated upstream region.

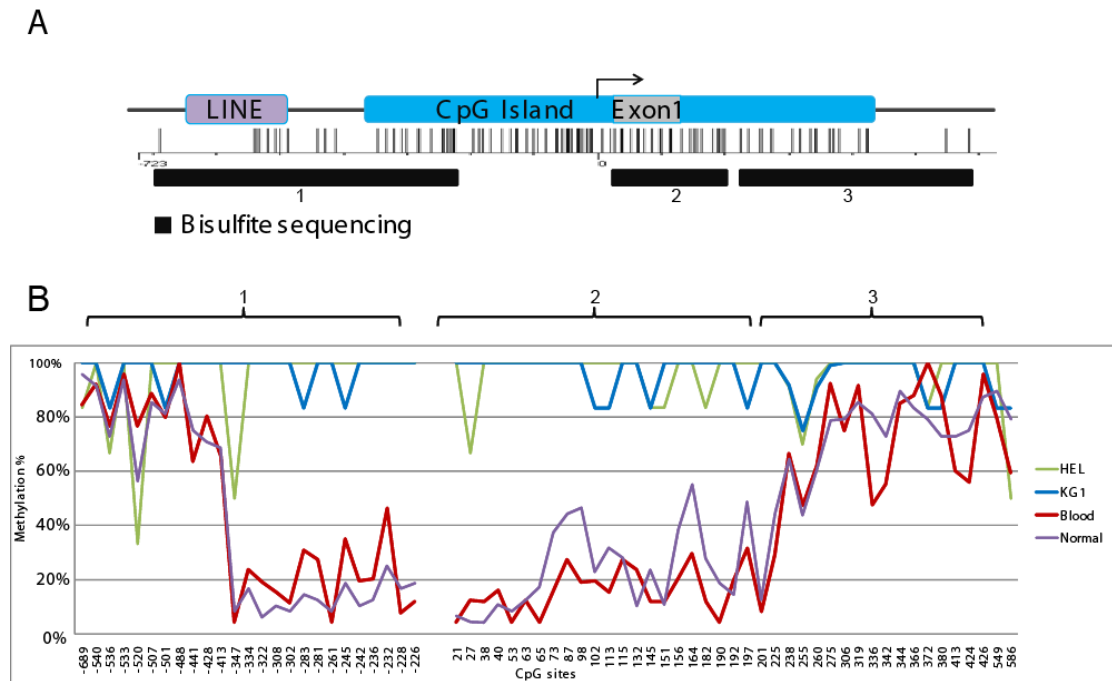
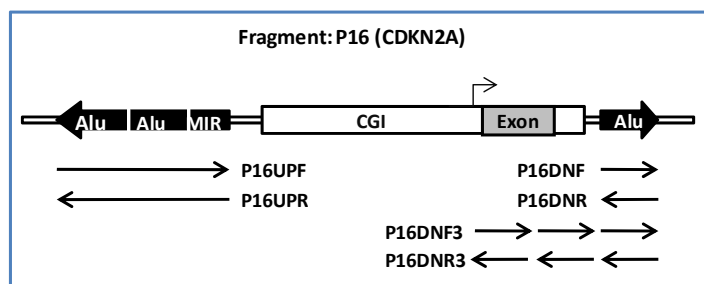
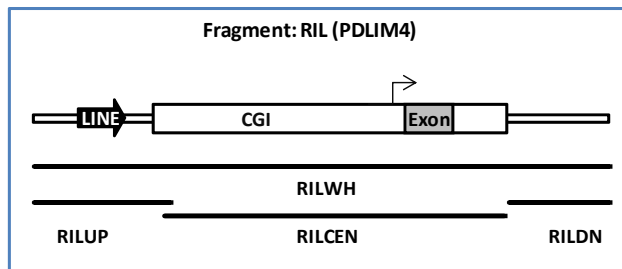
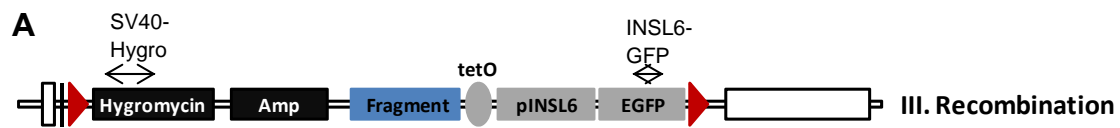


Figure 14. Mapping the methylation pattern of RIL promoter by bisulfite sequencing.

(A) Schematic RIL promoter region to be investigated. Bars, the CpG sites around -722bp to +630bp from the transcription start site (TSS, 0). Positions of bisulfite sequencing are shown in numbered rectangles. (B) The methylation patterns of RIL promoter in cancer cell lines, normal blood and normal colon tissues.

Construction of transgenes

After establishing the host cells, transgenes were introduced into the FRT site by flippase (Flp)-mediated homologous recombination. We made constructs composed of a reporter (EGFP), the Insulin-like 6 promoter (pINSL6), a tetO element, and fragments of interest (Figure 15). The INSL6 promoter is a CGI-promoter which is methylated in somatic cells including many cancer cell lines (108), and has two short LINE elements (L2, 89bp, 4 CGs; L4, 52bp, 1 CG; between L2 and L4, 13bp, 1 CG) upstream of its CGI. TetO is the response element to tTS binding in the absence of Doxycycline (Dox), and here was located upstream of the pINSL6. We chose RIL (PDLIM4) and P16 (CDKN2A) promoter regions because they have consistently methylated repetitive elements (an upstream LINE for RIL; three upstream SINEs and one downstream SINE for P16) surrounding the CGI. The fragments of interest (Figure 15) include the entire RIL promoter region (RILWH, TSS-713 to +664, 1421bp, 96 CGs), or isolated fragments containing the hypermethylated upstream (RILUP, -713 to -378, 336bp, 11 CGs), central (RILCN, -406 to +237, 643bp, 66 CGs) or downstream (RILDN, +214 to +664, 451bp, 20 CGs) regions. In the case of P16, fragments are originated from three upstream SINEs (P16UPF, forward orientation as in the genome, 980bp; P16UPR, reverse) and one downstream Alu (P16DNF, forward, 511bp; P16DNR, reverse; P16DNF3, three tandem forward copies, 1485bp; P16DNR3, three tandem reverse copies). We selected stable clones using a second marker (Hygromycin B) for only 10 days, as our observations indicated selection would confer a pressure for clones to remain unmethylated.



Control: pINSL6 only (No-frag)

B

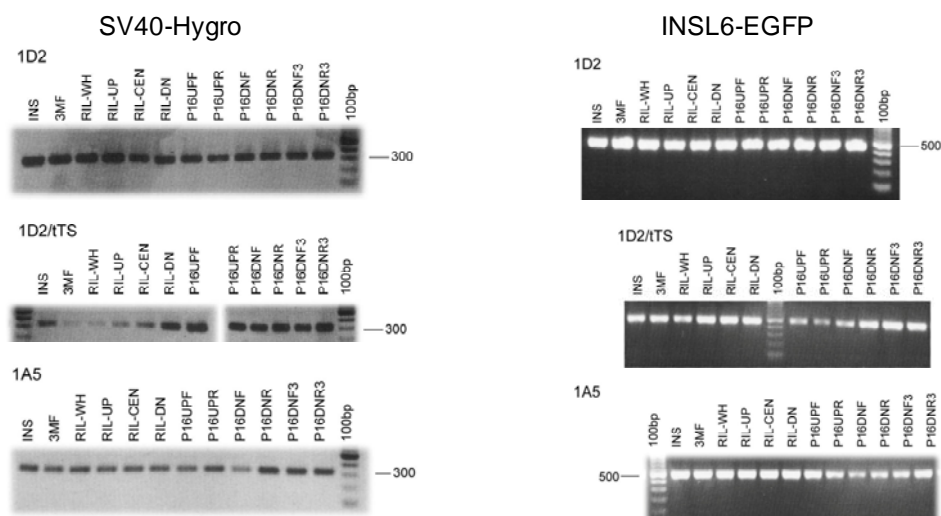


Figure 15. Construction of transgenes. (A) Graphical organization of transgenes for homologous recombination. Fragments were from RIL promoter, p16INK4A promoter and control without any fragment (No-frag). (B) PCR amplification of recombined transgenes. Primers SV40-hygro and INSL6-GFP were designed to show correct connection of the first vector pFRT/LacZeo and the recombined transgenes. Representative stable single clones were shown here.

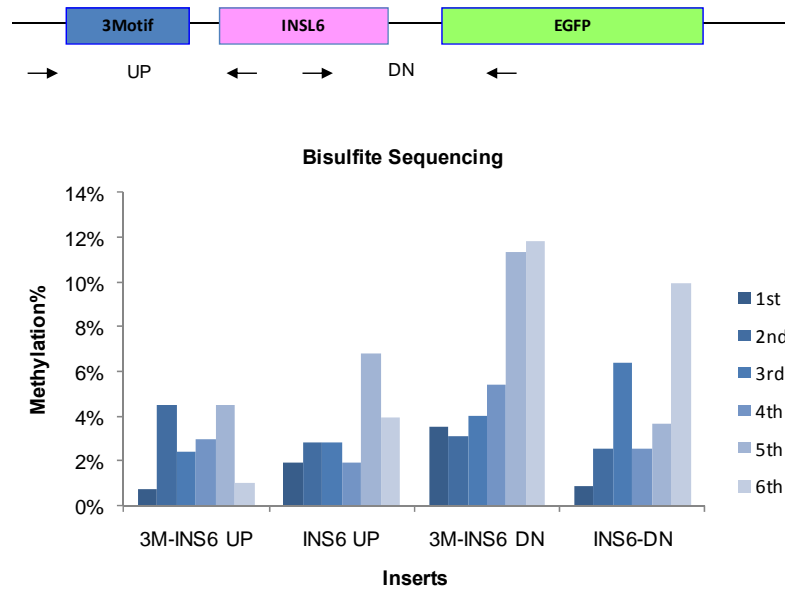


Figure 16. Continuous selection populated unmethylated clones. The vector graph indicates two constructs (without tetO) transfected into HCT116D host cells and selection with hygromycin B was present until the fourth month, and culture continued for another three month (stopped at the sixth month). Bars, the average methylation of all covered CG sites by bisulfite cloning/sequencing, and primers are shown in the vector graph. 3M-INS6, three motifs located in front of pINSL6; INS6, pINSL6. UP, the upstream primers; DN, the downstream primers.

Continuous culturing under selection pressure populated unmethylated clones

Stable single clones are usually under low concentration of selection drug to keep effective expression of transgenes. However, we have found that this contiguous culturing under selection pressure would play the opposite role in regard to our purpose of observing DNA methylation. This was seen in our preliminary examination by transfecting constructs with pINSL6 and 3Motif-pINSL6 into HCT116D. The first three months of culturing were under selection of hygromycin B, and all single clones showed no more than 8% methylation at the examined pINSL6 by bisulfite-sequencing; while another three months culturing without hygromycin B did not elevate regional methylation, and the maximum methylation of the same regions was still less than 14% (Figure 16). Replacing pINSL6 with CMV or TK promoter also seemed to be refractory to promoter methylation (not shown here). Accordingly, these results confirmed that active transcription of the promoter would not facilitate DNA methylation.

Unequal silencing speeds of transgenes in flp-in and flp-in/tTS cells

Following stable transfection of constructs, we first took a look at the changes of GFP expression along the time points from the second to the fifth month. In SW48A and HCT116D which do not express tTS, it took around four months for GFP to be undetectable in most transgenes, while in host cells without tTS the GFP fraction already went down to almost 0% at the earliest sampling time. Apparently, tTS in the host cells was functioning effectively as shown before (Figure 13) and greatly enhanced GFP suppression. Moreover, the exogenous pINSL6 itself was gradually suppressed in host cells without tTS (SW48A and HCT116D). However, culturing flp-in/tTS cells in media with Dox could not retain the same

GFP expression as in tTS (-) cells. In HCT116D/tTS, Dox restored GFP expression but to a lower level than expected; in SW48A/tTS, Dox treatment starting from 2nd month after transfection induced little GFP (Figure 17), suggesting that the KRAB-effect had become permanent.

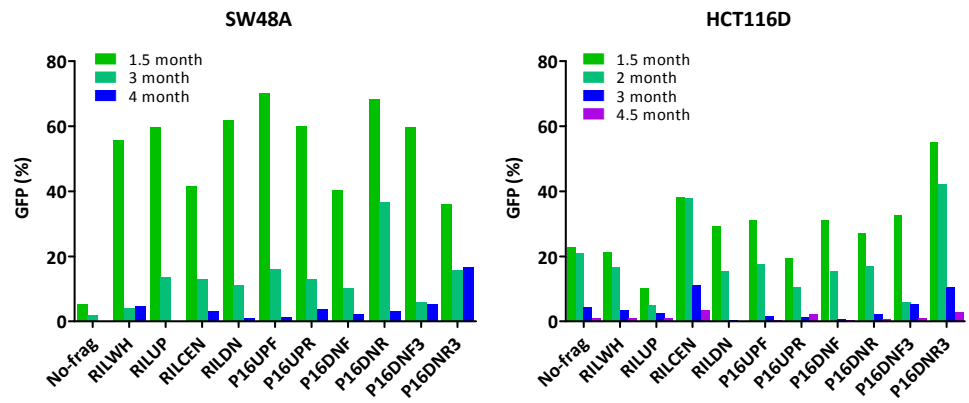
De novo DNA methylation center in transgenes

To test DNA methylation, after transfection and selection we did each analysis starting from the first month (for HCT116D and HCT116D/tTS) or the second month (for SW48A and SW48A/tTS, which grow more slowly), and repeated the same measurement at different time points till the fifth month. The methylation patterns were mapped by bisulfite pyrosequencing of several regions shown in Figure 18. Over the examined tetO-pINSL6, there are three CG sites (region A) located in the proximal LINE (L2) of the INSL6 promoter which demonstrated higher sensitivity to DNA methylation than those in the other regions. Take the construct containing only the pINSL6 (No-frag) for example, in SW48A/tTS cells of the second month, the average methylation of region A reached 73.6%, which was much higher than the other LINE (L4) (region D, 8.0%), the CGI (region C, 7.5%) or the tetO (region E, 5.1%). Therefore, these three CG sites were designated as methylation “hotspot”. The CG sites between the CGI and the hotspot (region B) achieved an intermediate methylation level of 23.3%, suggesting that it may be a “transitional” region in methylation spreading. Furthermore, bisulfite sequencing was performed in the same samples (from the 2nd month) to view the methylation status of every CG site in the examined INSL6 promoter (TSS-918 to +22, 940bp). This confirmed that the major methylated region was the hotspot (41.6% to 75.0%, mean) in all the transgenes tested. These data implied that the CG sites in

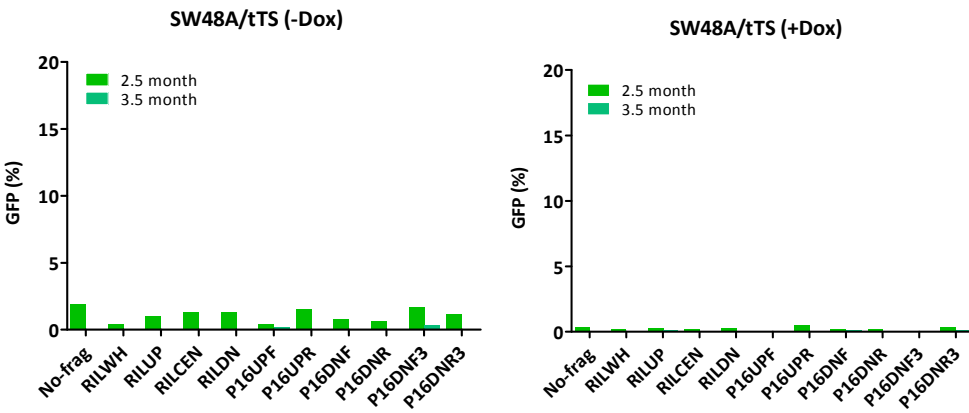
region A are rapidly “seeded” after transfection and thus may function as a methylation center (Figure 19).

Next, we evaluated the influence of upstream sequences on DNA methylation in the hotspot. In SW48A/tTS, compared with the control (No-frag), there were no significant differences in hotspot methylation for any of the additional upstream fragments, with the possible exception of P16DNF which somehow got reduced DNA methylation although this was not statistically significant (Figure 19).

A



B



C

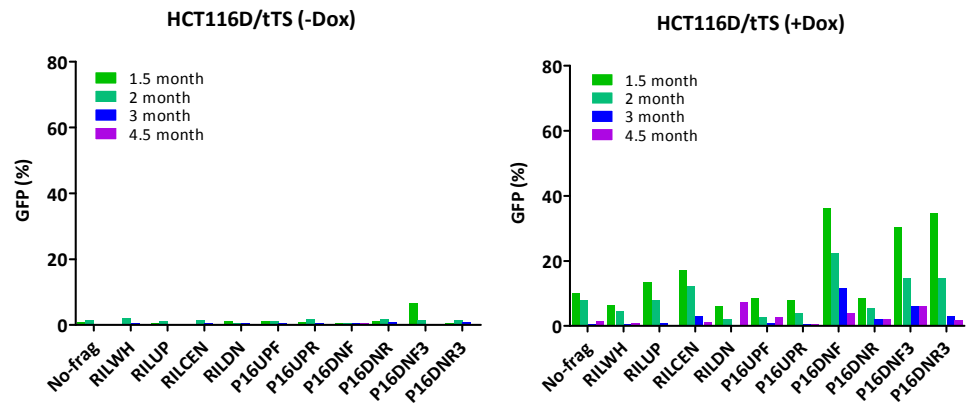
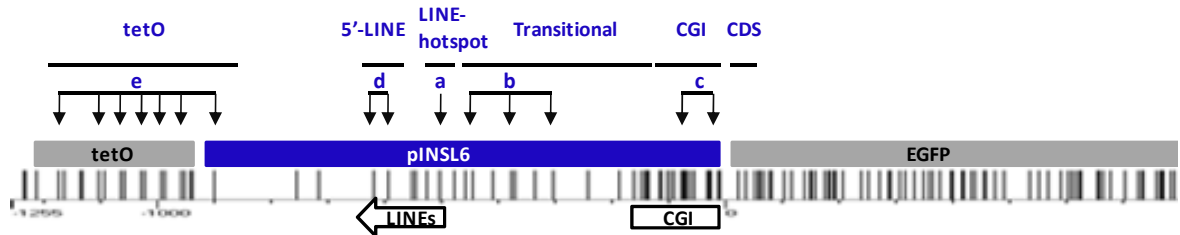


Figure 17. Unequal GFP silencing speed in flp-in and flp-in/tTS host cells. GFP expression was measured by flowcytometry. Several time points from the first month to the fifth were selected for each stable clone under long-term culture. For tTS-expressing flp-in host cells, clones were cultured in parallel wells either with or without Dox treatment. (A) SW48A and HCT116D. (B) SW48A/tTS with Dox (+) and Dox (-). (C) HCT116D/tTS with Dox (+) and Dox (-). GFP expression in long-term cultured tTS-containing clones presented partial inducibility in response to Dox. Maintaining clones under Dox could not prevent GFP silencing in the end, but could slow down the process at least in some HCT116D/tTS clones. All of the SW48A/tTS clones were refractory to the inducible GFP expression at the earliest detection time.

A



B

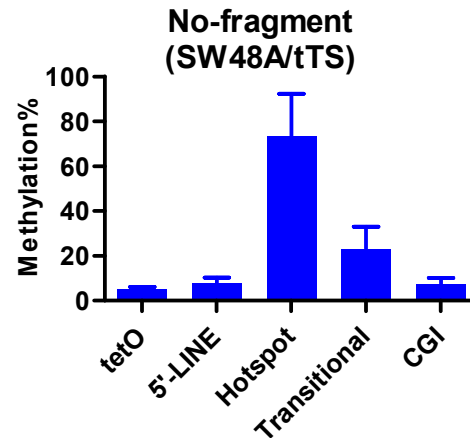
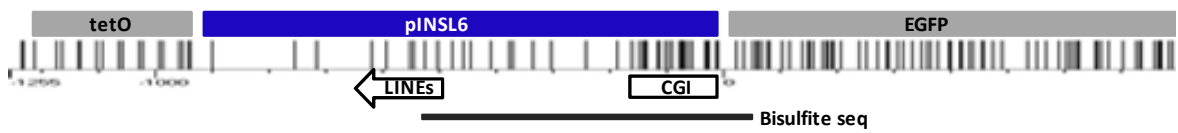


Figure 18. Sensitive CG sites to de novo DNA methylation by bisulfite pyrosequencing.

(A) The graphical distribution of CG sites in tetO-pINSL6-EGFP. The subcloned 940-bp pINSL6 consists of part of the CGI and two short LINE elements. Amplification of bisulfite-converted DNA covers 5'-end of EGFP to distinguish it from the endogenous pINSL6 (thick line). The small characters and arrows refer to the pyrosequencing target sites for regional methylation. (B) The regional methylation of CG sites in the control construct (No-frag) which does not carry any additional fragment upstream of tetO-pINSL6.

A



B



C

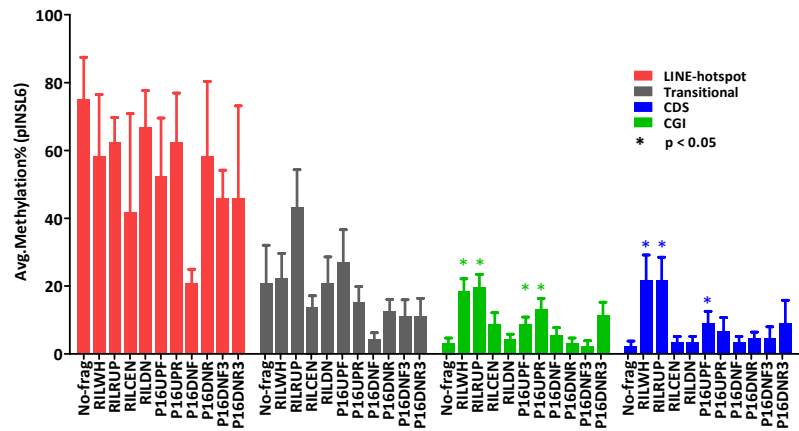


Figure 19. Methylation patterns of pINSL6 in SW48A/tTS at the second month. (A) The graphical distribution of CG sites in tetO-pINSL6-EGFP. Amplification of bisulfite-converted DNA covers 5'-end of EGFP to distinguish it from the endogenous pINSL6 (thick line). (B) Bisulfite-sequencing results displayed in circles (methylated CG sites as closed circles and unmethylated sites as open circles) and (C) calculated in four segments. No-frag, for pINSL6 only, and without any fragments from neither RIL nor P16. RILWH, RILUP, RILCEN, and RILDN represent transgenes with fragments from the RIL promoter (the entire promoter, upstream LINE, central CGI and downstream regions respectively). P16- prefix represents those with fragments from the P16 promoter (UPF and UPR, upstream SINEs in forward and reverse orientation; DNF and DNR, downstream Alu in forward and reverse direction; DNF3 and DNR3, three tandem downstream Alus in forward and reverse orientation). Thin lines stand for the segments divided based on pyrosequencing and bisulfite cloning/sequencing results. The bar graphs in (C) were average methylation levels of the subdivisions of bisulfite-sequencing-examined pINSL6 (mean \pm SEM). Asterisks, statistically significant difference in comparison with No-frag ($p < 0.05$).

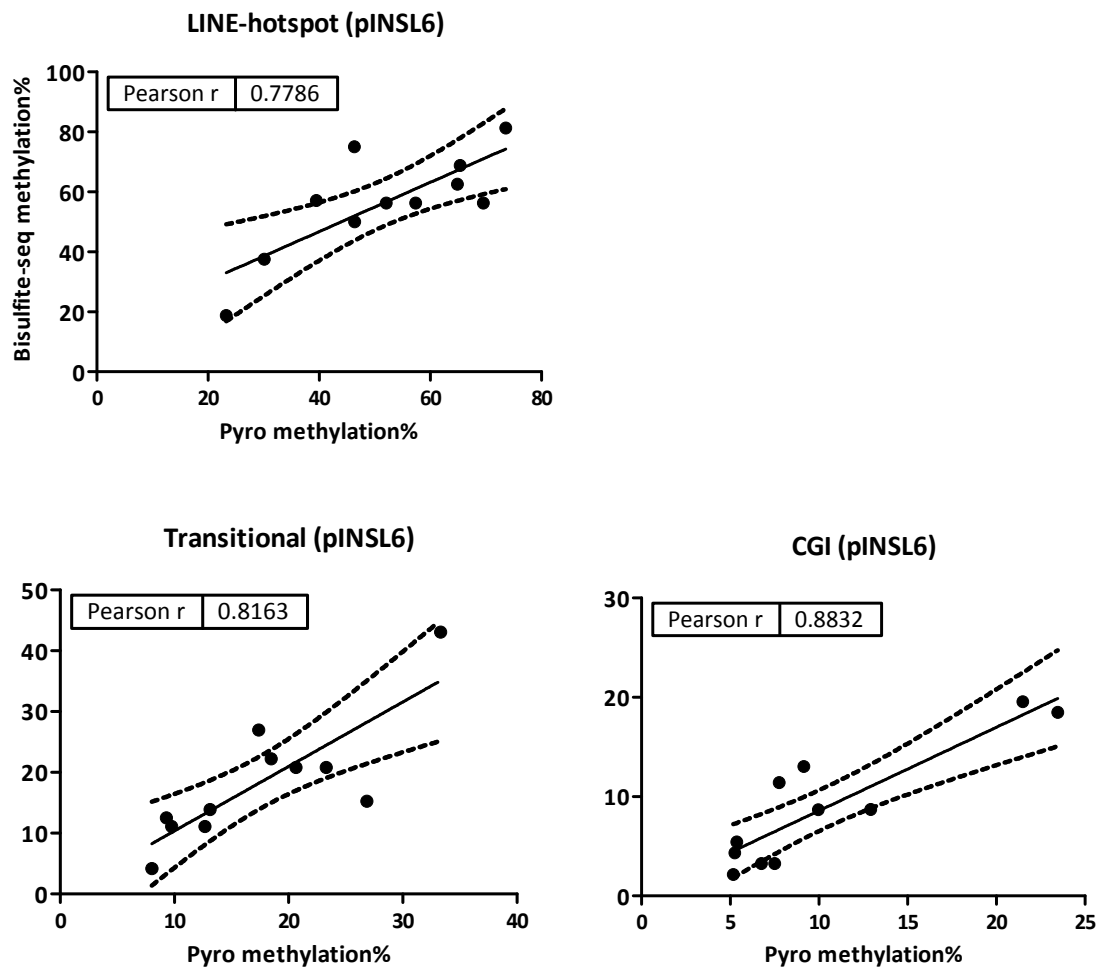


Figure 20. The correlation of two assays in analyzing segmental methylation of pINSL6.

The Pearson correlation coefficient is calculated for the average methylation levels between 11 paired bisulfite-sequencing and pyrosequencing results. The 95% confidence interval of the best-fitted line is shown in dotted curves.

The effects of repetitive elements on DNA methylation spreading

Another point out of the data in Figure 19 is the possibility that methylation was spreading from the hotspot to nearby sequences. We examined the effects of upstream sequences on spreading over time. As shown in Figure 19, upstream sequences from either the RIL promoter or the P16 promoter did bring up a distinct difference in the methylation of region C (the CGI of pINSL6). In the case of RIL, higher methylation was obtained for those constructs that included the upstream portion (18.5% for RILWH, $p < 0.01$, t-test; 19.6% for RILUP, $p < 0.01$, t-test) which is mostly a LINE element (L2), while the CGI (RILCEN, 8.9%) or the downstream portion (RILDN, 4.4%) did not affect the methylation significantly compared with the control (No-frag). A similarly enhanced methylation could be observed if the three upstream SINE elements from the P16 promoter were present in the construct (P16UPF, 8.7%, $p < 0.05$; P16UPR, 13.0%, $p < 0.05$). The transitional region between the hotspot and the CGI demonstrated intermediate methylation levels with little difference between transgenes, although RILUP and P16UPF also showed trends of higher methylation. The EGFP CDS achieved a pattern similar to the CGI. The assays of bisulfite pyrosequencing and bisulfite cloning-sequencing revealed a good correlation (correlation coefficient r was 0.78 for LINE-hotspot, 0.82 for transitional, and 0.88 for CGI) (Figure 20); we therefore used bisulfite pyrosequencing for subsequent analysis.

Gradual accumulation of methylated cytosine over time

The changes in transgene methylation were monitored over time in each single clone obtained. In figure 21, the methylation of several representative transgenes (No-frag,

RILWH, RILDN, P16UPR and P16DNF) in SW48A/tTS are plotted over four time points from the second to the fifth month. Prolonged culture time led to variable increase of methylation levels. For instance, the highest increase in hotspot methylation was 32.0% for P16UPR, 30.5% for P16DNF and 39.9% for P16UPF. In the CGI, the top changes were 39.0% for P16UPR and 29.5% for P16UPF. Together with the data from the adjacent regions, the transgenes had two distinct patterns. Some became more methylated at all early time points and tended to maintain their higher levels with rapid accumulation in most regions except the hotspot which was maximally methylated early. The most striking acceleration in the CGI were P16UPR and P16UPF (which contain three SINEs) and RILWH (which contains an L2 element). For other constructs, methylation was maintained with a slow and gradual increase in levels similar to the control. Examples of this pattern included RILDN and P16DNF.

Long-distance spreading was not direction-dependent

Given that the CGI of pINSL6 could become more methylated with repetitive elements present upstream, we wondered if regions further upstream or downstream would display a similar trend. We measured methylation levels of the distal LINE (5'LINE), the tetO element, and the fragments (336bp to 1485bp) (Figure 21), as well as the reporter EGFP (720bp) and the selection marker hygromycin B (4565bp upstream to the start of fragments) (Figure 22) in SW48A/tTS at the fifth month. Methylation was found expanded into a region up to 6kb away from “hotspot” and those with repetitive elements upstream were subjected to more methylation, which confirmed the effect of repetitive elements on methylation

spreading. Since EGFP and hygromycin B are located in two directions from TSS, spreading apparently was not dependent on the direction.

Promoted CGI methylation arose from adjacent de novo methylation

According to the “seed and spread model”, methylation spreading leads to CGI methylation once the protective boundaries are broken-down (63, 64). However, inherent cis-signals for initiating DNA methylation within CGIs independent of a hotspot might be an alternative mechanism that does not require spreading. To address this possibility, we separately constructed a truncated pINSL6 (tr-pINSL6) by removing all the sequences upstream of the CGI including both LINE elements. In the same SW48A/tTS host cell, transgenes (No-frag, RILWH and P16UPF) without the methylation hotspot (tr-pINSL6) achieved prominently lower CGI methylation at the second month (3.4% vs 7.5% for No-frag, 8.0% vs 23.5% for RILWH and 7.1% vs 12.9% for P16UPF). The difference was even more pronounced at 3 months: the truncated pINSL6 did not elevate CGI methylation from the second to third month, in contrast to the methylation increase by an average of 10% in the LINE-containing pINSL6 (Figure 23). GFP was silenced to the same level in all the constructs. We conclude that the CGI methylation arose more easily from spreading instead of de novo events. Nevertheless, the CGI itself was not absolutely methylation-free indicating DNMTs may target randomly independent of methylation centers but in a less efficient way (Figure 21).

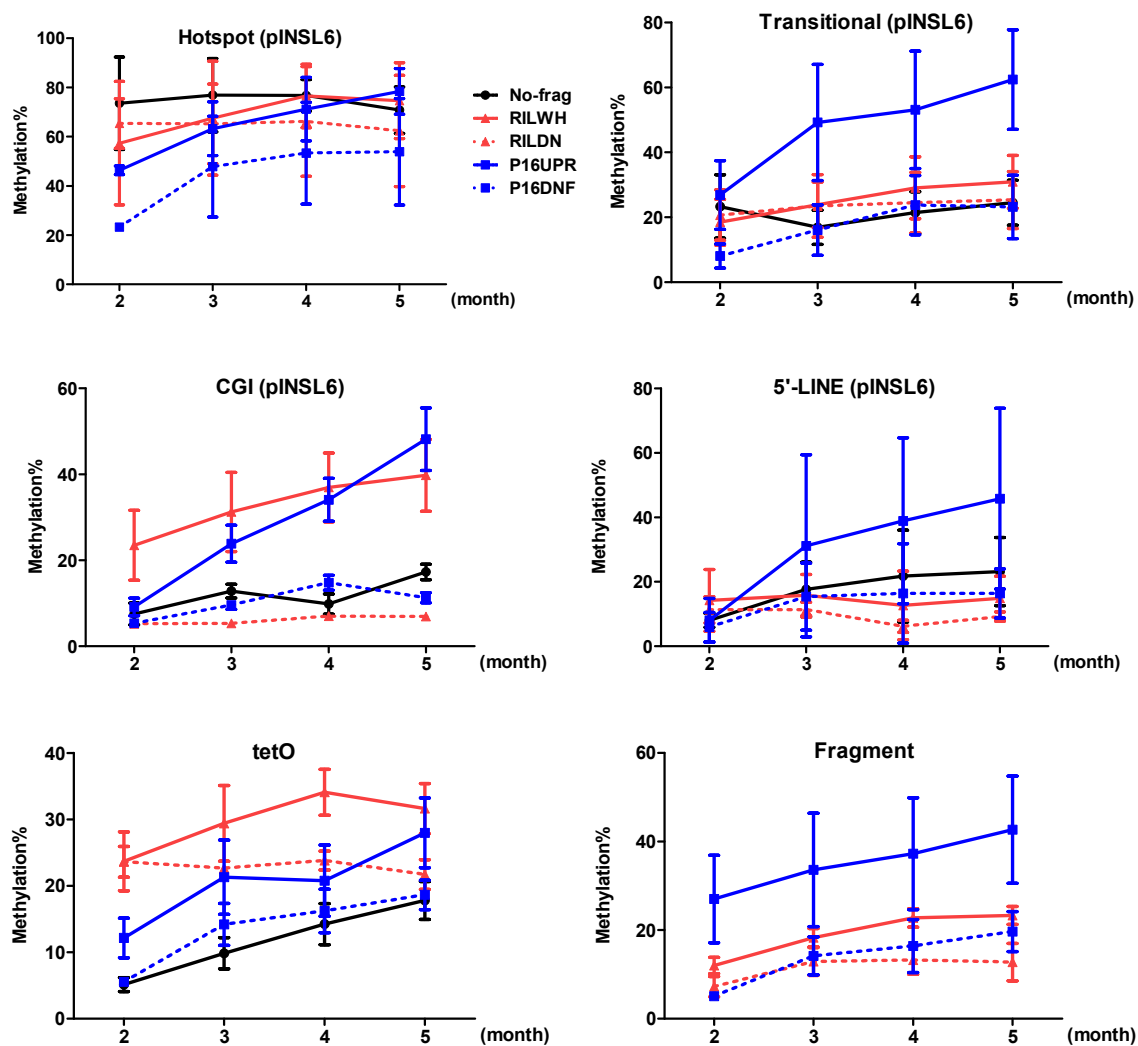


Figure 21. Gradual accumulation of transgene methylation in SW48A/tTS over time, sampled at 2, 3, 4 and 5 months. Also refer to the Figure for methylation in HCT116D/tTS and HCT116D. Trend lines represent fragments subcloned to the upstream of tetO-pINSL6.

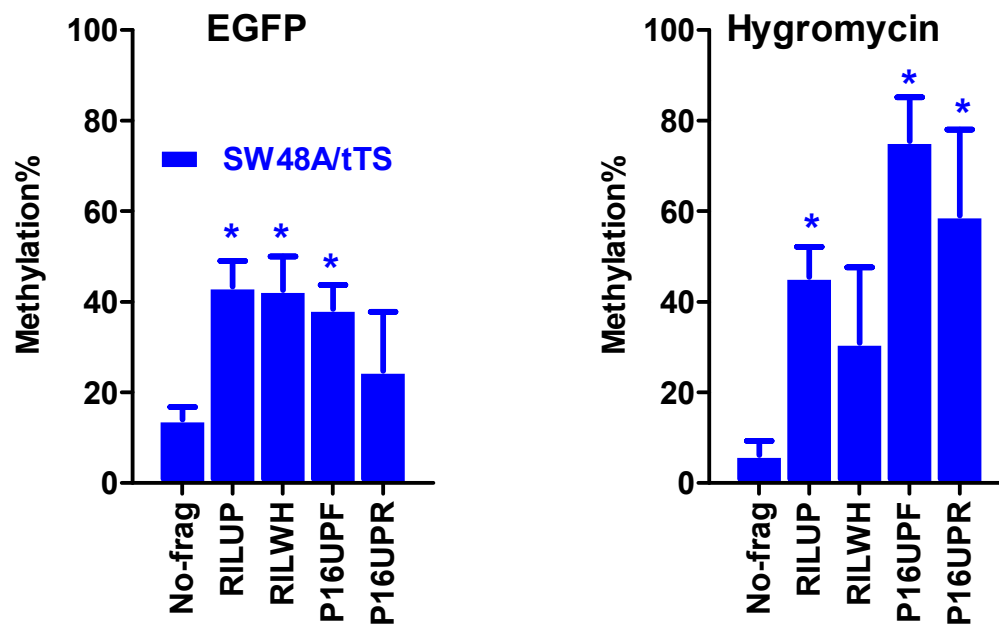


Figure 22. Methylation of two regions furthest away from pINSL6 which are inside of EGFP and hygromycin B. Methylation levels of four distinct transgenes in SW48A/tTS at the fifth month are shown here. Bars, average methylation levels (mean \pm SEM). Asterisks, significant difference in comparison to No-frag ($p < 0.05$).

A highly repressive environment can enhance de novo methylation and methylation spreading

Methylation results from SW48A/tTS, the most repressive condition have been discussed so far, and we next investigated the influence of tTS by comparison of the patterns in SW48A/tTS and SW48A. It could be seen that the repetitive elements were not the only factor in promoting methylation spreading, the presence of tTS enhanced both de novo methylation in “hotspot” and methylation spreading into the CGI and distal areas. First, the hotspot in SW48A only achieved 13.1% to 35.8% methylation at the second month and by the fifth month the methylation rose to no more than 50% (26.1% to 49.1%) (Figure 24); in contrast, SW48A/tTS host cells already induced 23.3% to 73.6% methylation at the second month and by the fifth month 42.0% to 84.2%.

A validation experiment was carried out by removing tetO (notetO-pINSL6) from the transgenes (No-frag and RILUP). In SW48A/tTS at the second month, the hotspot of notetO-pINSL6 obtained methylation levels very close to those of pINSL6 in SW48A host cells (15.6% for No-frag and 23.0% for RILUP) (Figure 25). The same trend was maintained at the third month. Removal of the tetO allowed GFP expression to be 74% for No-frag and 65% for RILUP on average, ruling out possible systemic interferences caused by cell engineering, such as non-specific effects due to tTS insertion in the course of generating SW48A/tTS host cells.

Furthermore, in the absence of tTS, the CGI methylation presented much lower levels and little variation in different constructs (5.0% to 16.1% at the fifth month), which was in vast contrast to the patterns shown in SW48A/tTS (Figure 24). The loss of influence from repetitive elements was observed in the other adjacent regions including transitional, 5'-LINE, tetO and the fragments. As well, methylation accumulated in a slower rate over time

(Figure 26). This suggests that the effect of repetitive elements on methylation spreading is limited by the extent of local repression, and both repetitive elements and strong repression are required for a CGI to become significantly methylated.

Local heterochromatin in transgenes

ChIP analysis was performed to study local enrichment in repressive histone marks. The active marks (H3K4me3 and H3K9ac) and the inactive marks (H3K9me3 and H3K27me3) were analyzed in the cells sampled from the second and sixth month. The assays for the quantitative real-time PCR were designed at the connection region of tetO-to-pINSL6 and pINSL6-to-EGFP respectively to distinguish the exogenous from the endogenous pINSL6. Compared with the control regions (GAPDH and RARB), the pINSL6 in transgenes was enriched for H3K9me3 (two to three-fold more than GAPDH and RARB), and devoid of H3K4me3 and H3K9ac, indicating a local repressive environment. By contrast, H3K27me3 was not enriched in pINSL6, therefore the repression was more related to recruitment of H3K9me3 (Figure 27). However, there were neither differences of inactive (H3K9me3 and H3K27me3) histone marks between SW48A/tTS and SW48A, nor the differences between the repetitive-elements-containing transgene (RILUP) and pINSL6 only (No-frag). Thus, at late time points examined, the enrichment for repressive marks had already reached a stable level.

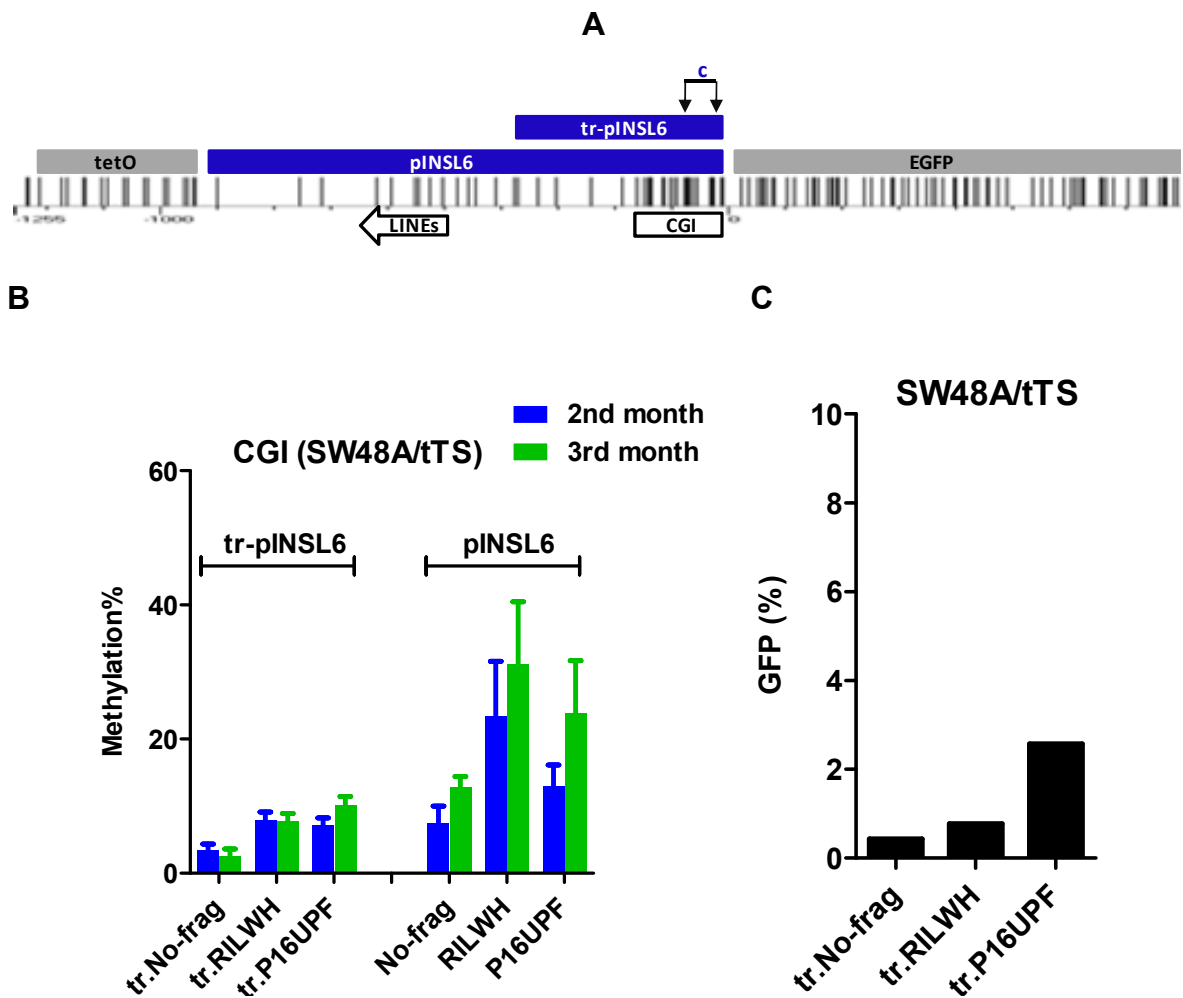


Figure 23. Comparison of transgene behaviors between the truncated and complete pINSL6 in SW48A/tTS. (A) The truncated pINSL6 (tr-pINSL6). (B) CGI methylation (the second and third month). (C) GFP expression (the second month). Refer to Figure 17 for more GFP expression profiles.

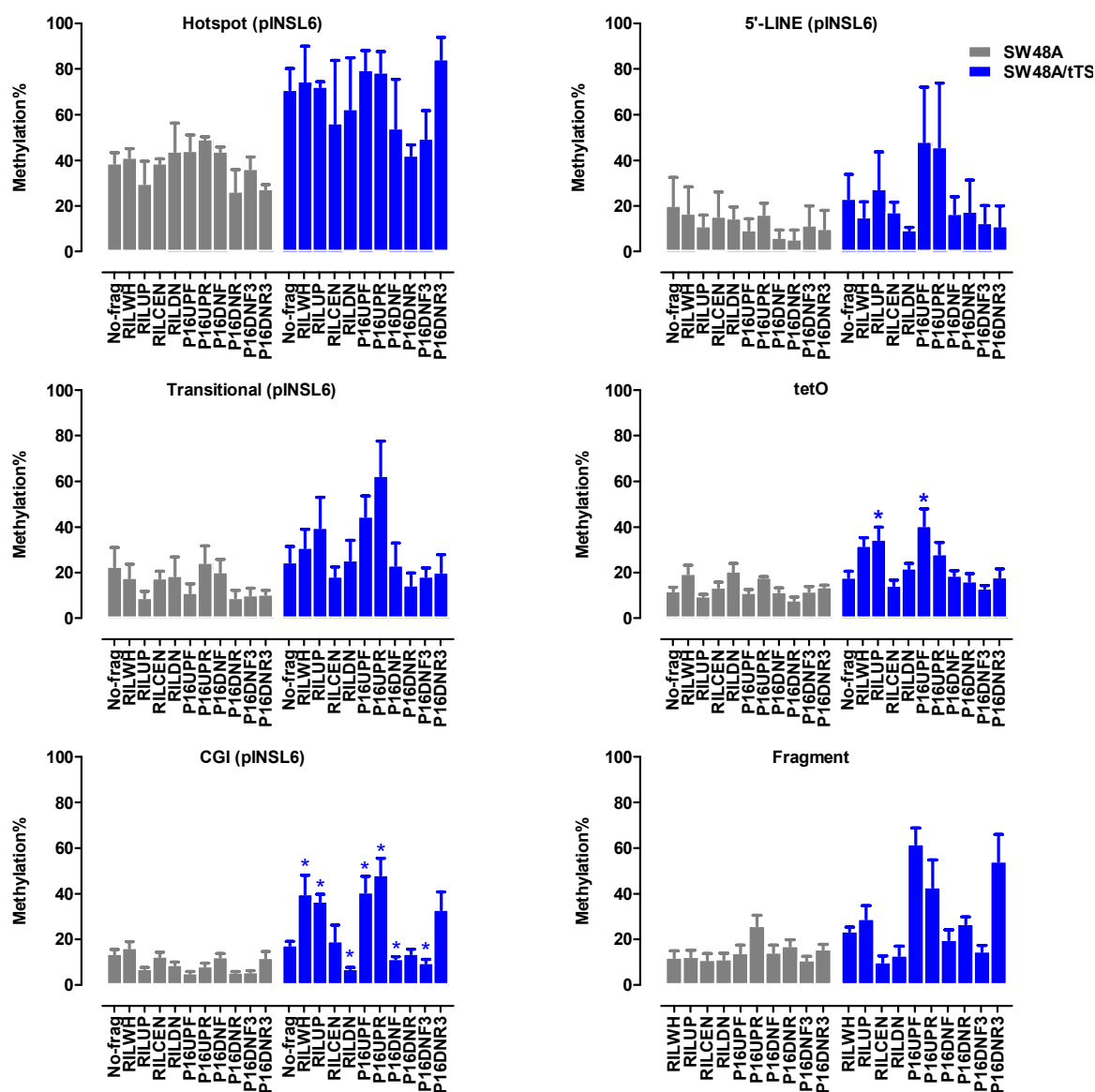


Figure 24. Methylation profiles of pINSL6 in flp-in host cells by pyrosequencing. Regional methylation detected in each transgene from SW48A/tTS and SW48A at the fifth month.

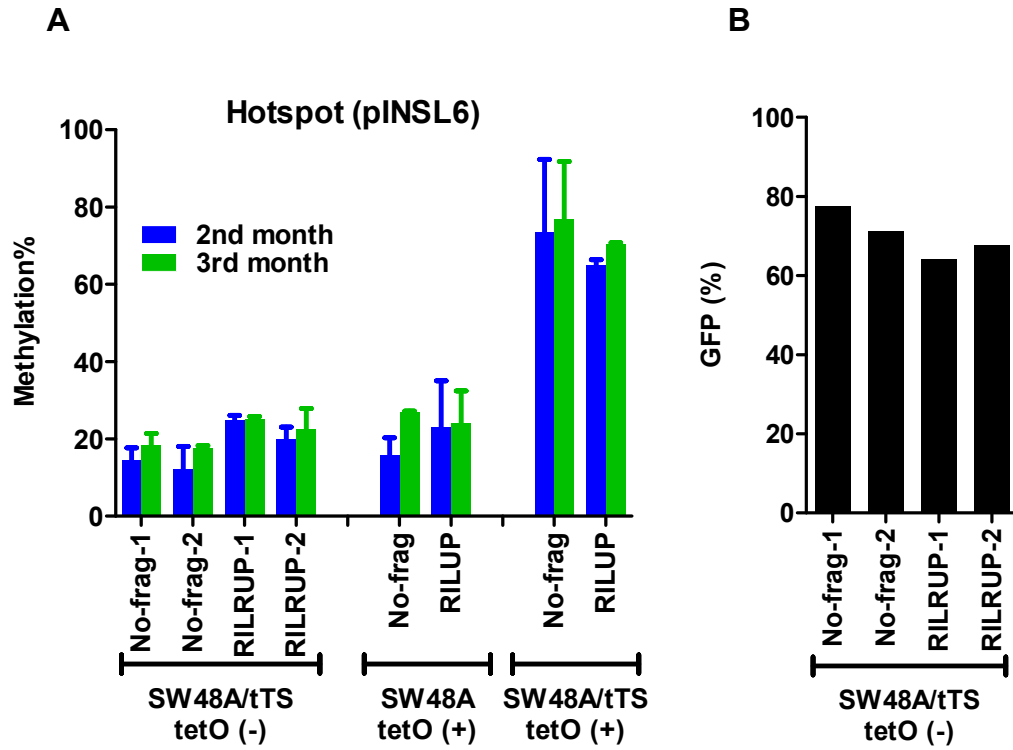


Figure 25. Comparison of CGI methylation and GFP expression between tetO (-)-pINSL6 in SW48A/tTS, tetO(+)-pINSL6 in SW48A and tetO(+)-pINSL6 in SW48A/tTS.

(A) Pyrosequencing was performed for clones at the second and third month after transfection. (B) Flow cytometry was used to detect GFP expression at the second month.

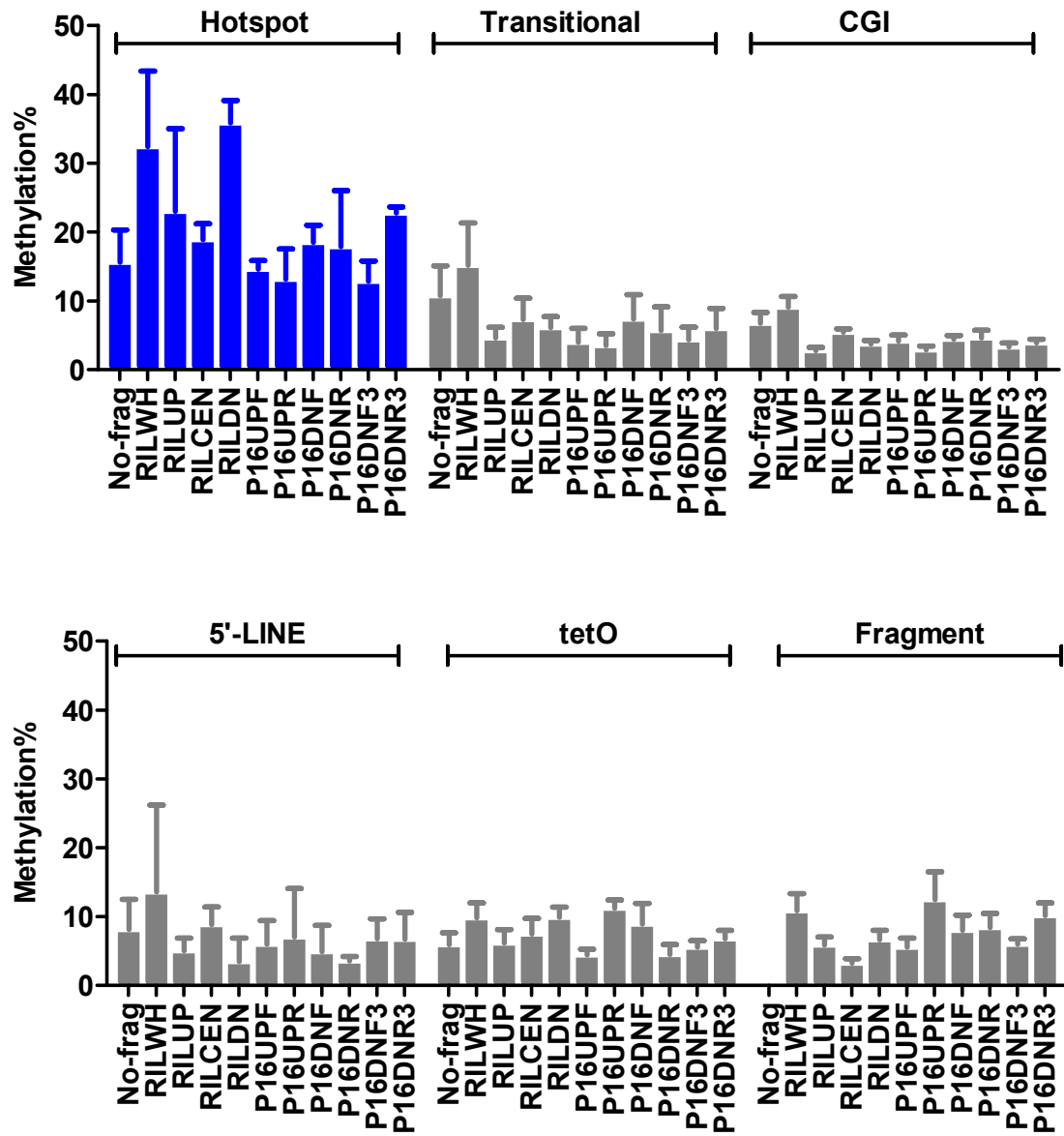


Figure 26. Regional methylation of transgenes in SW48A at the second month by pyrosequencing. Bars, average methylation levels (mean±SEM). Statistical analysis did not show significant of any fragment compared with control construct (No-frag).

The presence of tTS made transgenes refractory to TSA-induced derepression

We also tried to confirm the stronger repression related to the presence of tTS by treating clones of RILWH and No-frag from SW48A and SW48A/tTS at the fourth month. Compared with the control group, the DNA methylation inhibitor 5-aza-2'-deoxycytidine (DAC) induced global LINE-1 demethylation by 20% to 30% as well as local “hotspot” demethylation by 5% to 18%, but did not reactivate GFP expression. The HDAC inhibitor, TSA induced GFP re-expression without any demethylation. GFP expression was significantly restored by TSA for the clones from SW48A, but was not affected in SW48A/tTS, confirming the function of tTS in enforcing local repression (Figure 28).

The effects of cell lines or genomic loci on DNA methylation recruitment

Besides the host cells from SW48, we utilized a different cell line HCT116 with a different integration site (HCT116D) to assess the conservation of DNA methylation in response to transcriptional repression. In both HCT116D and HCT116D/tTS, de novo methylation behaved in the same way as in SW48A whereby the hotspot achieved quicker and higher methylation than the other regions (Figure 29). However, methylation spreading was slower, not distinctively affected by the upstream repetitive elements, and was short of correlation with the presence of tTS in HCT116D/tTS. Thus, DNA methylation of the hotspot appeared to be an intrinsic property of the CG sites, whereas methylation spreading was greatly influenced by cell line context and/or integration sites.

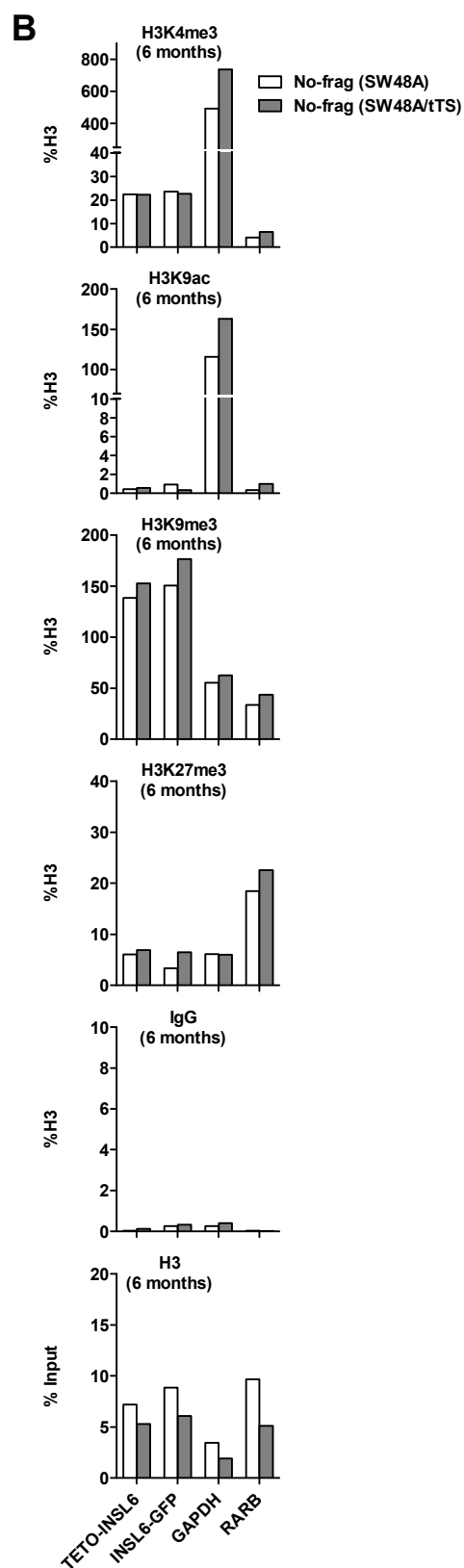
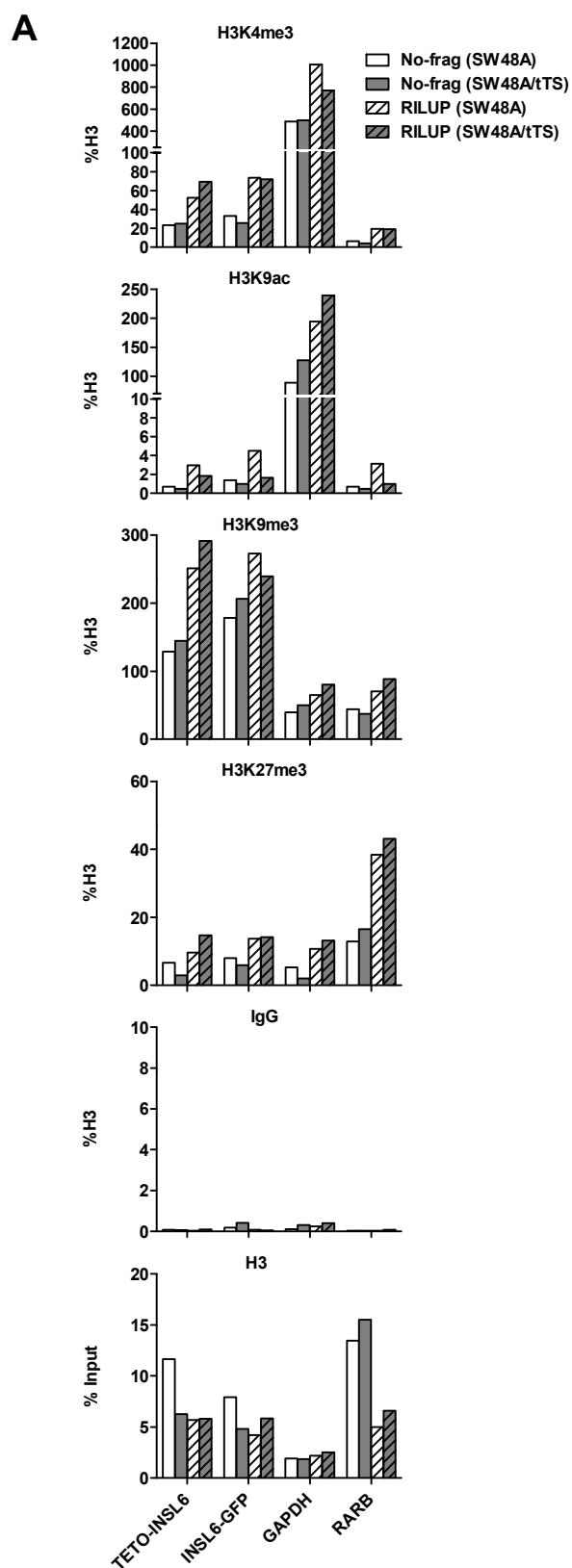


Figure 27. ChIP-qPCR for analyzing the enrichment of histone marks in pINSL6.

Antibodies against active marks (H3K4me3 and H3K9ac) and repressive marks (H3K9me3 and H3K27me3) were used to pull down sonicated chromatin. All the values were normalized to H3 and H3 occupancy was calculated as the percentage of input. ACTB and GAPDH as the control regions for active marks; RARB as the positive control for H3K27me3. TETO-INSL6 and INSL6-GFP are designed at the 5'-end and 3'-end of pINSL6 to distinguish it from the endogenous one. (A) Transgenes with no fragment or RILUP in SW48A and SW48A/tTS at the second month. (B) Transgenes with no fragment in SW48A and SW48A/tTS at the sixth month.

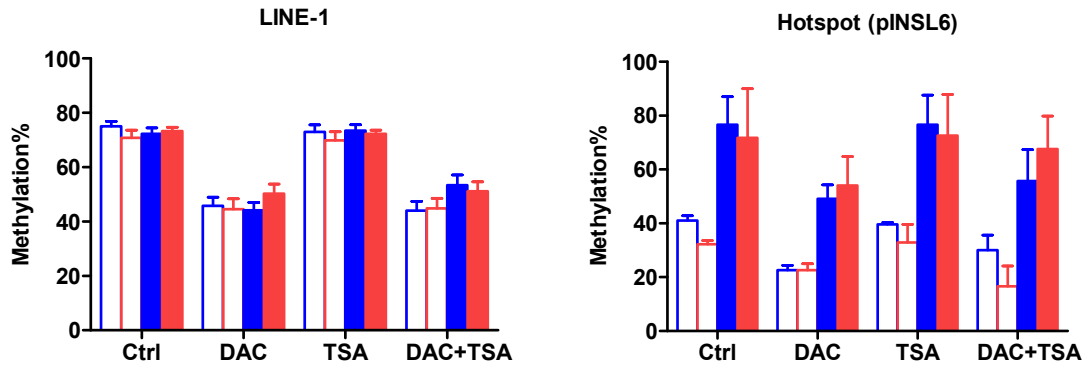
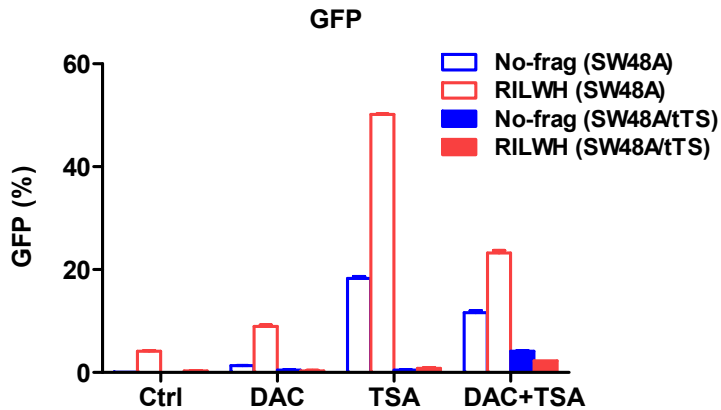
A**B**

Figure 28. The reaction of SW48A and SW48A/tTS clones to epigenetic drug treatment.

The cells (transgenes with No-frag and RILWH) of the fourth month were treated with DAC (200nM) for four days and/or TSA (800nM) during the last day and controls were cultured in regular media. (A) Global DNA methylation (LINE-1) and local methylation of the identified methylation hotspot in pINS6 were measured and compared with the respective control group. (B) GFP expression was detected by flowcytometry. All the values were averaged from biological duplicates.

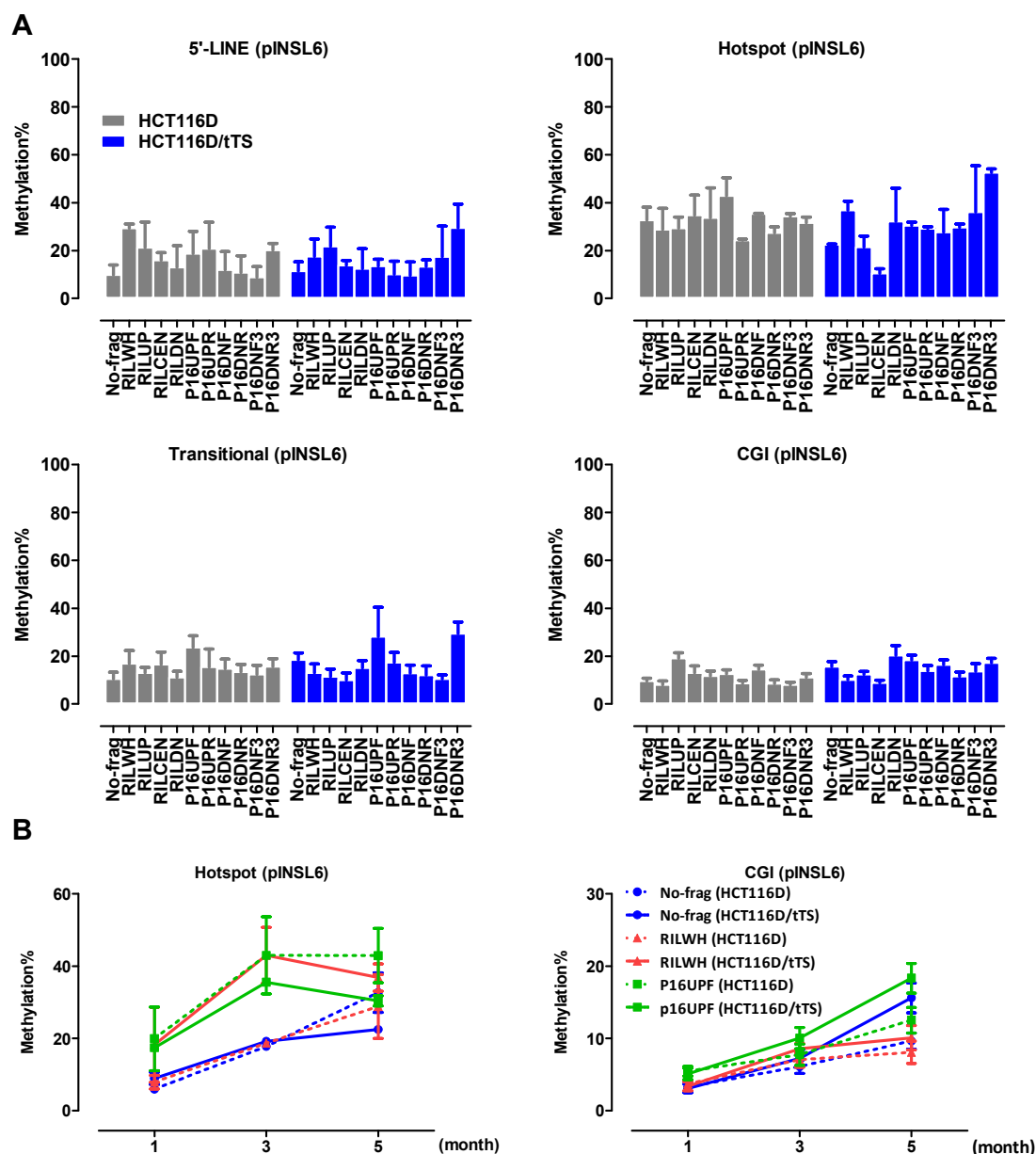


Figure 29. Methylation profiles of pINSL6 of transgenes in flp-in host cells.

Pyrosequencing was used to detect regional methylation patterns as described before. (A) Methylation in HCT116D/tTS and HCT116D, sampled at the fifth month. The x-axes represent fragments subcloned upstream of tetO-pINSL6. Bars, average methylation levels (mean \pm SEM). (B) Gradual accumulation of transgene methylation. Three time points from the first to the fifth month are shown for HCT116D/tTS and HCT116D.

Discussion

Our investigation of de novo methylation and spreading process was realized through a site-specific integration system with enforced local transcriptional repression. By studying expression, DNA methylation and histone modifications of transgenes at a single integration site, we were able to distinguish several aspects involved in DNA methylation of promoter-CGIs. We find that (1) DNA methylation originates from certain CG sites within a LINE element; (2) methylation spreading into a promoter-CGI is facilitated by transcriptional repression, presence of additional repetitive elements and is site-specific; (3) transcriptional repression is required but not sufficient to promote DNA methylation.

Repetitive elements (REs) comprise ~45% of human genome, most of which are derived from the activity of transposable elements (123). They are considered to account for global DNA methylation in normal somatic cells, while become hypomethylated in cancers increasing the risk of genomic instability (2). In mammals, almost 25% of the analyzed promoter regions contain repetitive DNA (124) and not all of them are deficient of methylation in cancer cells, such as SINE sequences located upstream of the *PI6INK2A* promoter (125) and the LINE elements at the upstream of RIL promoter (109). So we evaluated the roles of CGI-promoter-adjacent repetitive elements in DNA methylation recruitment and spreading. The first observation of our experiments is “seeding” of DNA methylation in transgenes. The exogenous 940bp INSL6 promoter consists of two short LINEs (L2, divergence 27.3%, RepeatMasker; L4, 20.0%) upstream of a CGI. There are six CGs across the repeat region, but only the proximal two CGs of the L2 with the next CG site achieved distinguished methylation from the adjacent areas at the earliest time point in almost all the constructs examined. The methylation was induced independently of cell lines

(HCT116 and SW48), genomic loci or the strength of transcriptional repression, while the extent of methylation was elevated by the presence of tTS and affected by cell line and/or loci used. Therefore, repetitive elements could serve as cis-signals for de novo DNA methylation in cancer cells, nonetheless, not all the CGs but a few could be involved. However, since only the INSL6 promoter was analyzed in the system, we do not know if there are any common sequences assigned by DNMTs which may be elucidated by genome-wide analysis. Also, the “seeding” event may be determined by dynamic nucleosome deposition as suggested by de novo methylation of the P16INK4A CGI in post-selection primary human mammary epithelial cells (HMECs) (84).

The second function of repetitive elements located further upstream is to cis-regulate methylation spreading into the adjacent regions from de novo sites, especially into CGIs. The LINE element (L2, divergence 26.9%) from RIL promoter and three concatenated upstream SINEs (MIR, divergence 24.1%; Alu, 8.7%; Alu, 9.0%) from P16 promoted striking methylation of the CGI in SW48A/tTS which represents a genomic locus with strengthened silencing in a cancer cell. In other host cells, spreading was not significant if there was any, so these REs we studied here may not be so strong as to overcome the protective machinery independently of the repressive strength, or alternatively, repetitive elements have to cooperate with transcriptional repression in order to render methylation spreading. Importantly, not all REs contribute to methylation spreading as the downstream Alu of P16 (divergence 11.9%) did not make the adjacent methylation get to the same level as the upstream REs did. Previously, some repetitive elements are empirically defined as cis-regulatory elements (124) and genome-wide analyses have shown some human and mouse promoters are derivatives from specific repetitive elements (126). Thus, the LINE of RIL or

SINEs of P16 may work as another kind of cis-signal to recruit either stronger transcription repressor or more chromatin remodeling factors thereby facilitate the access of DNMTs to the CGI.

The organization of REs may influence methylation spreading. Increased transgene copies was reported to induce more methylation and more compacted chromatin in mouse lines (127), and imprinted genes more frequently contain tandem repeat arrays in their CpG islands than randomly selected genes in both mouse and human (128). We also observed that three upstream SINEs of P16 were more effective than one Alu from the downstream in spreading, and three tandem downstream Alus seemed also to have a bit more impact. Presumably, although the compaction caused by one copy of RE may be insufficient, adding up the effects of several copies could possibly get over the threshold and make more DNMTs accessible to the adjacent regions. But this hypothesis still needs further examination.

One strategy of our experiments was to control the local repression strength by using the tetracycline-controlled transcription silencer (tTS). The tTS is usually used in inducible expression system, and here we are employing its role in sequential recruitment of the H3K9-specific histone methyltransferase (e.g. SETDB1), HP1, and the histone deacetylase (HDAC)-containing complex via KRAB-KAP1 cooperation (117, 118). Hence, even over a range of euchromatin, a highly compact heterochromatin patch can be generated and maintained for quite a few generations. On the other hand, pINSL6 is not a strong promoter because GFP was gradually silenced and the promoter was enriched for repressive histones (H3K9me3) probably through adopting the endogenous regulators targeted at pINSL6. Therefore, pINSL6 could set up a repressive background, and usage of tTS increased the repression to a higher and long-standing level. The stronger the localized repressive

heterochromatin was, the faster de novo methylation occurred and the more chance the CGI could become methylated.

The variation of position effect was another trans-regulatory aspect taken into account as well. Only tTS could not guarantee the heterochromatin patch inserted with transgenes to reach as high as the level permissive for DNA methylation. Gene body methylation in mammals has been confirmed in many studies (129). As the integration sites of the host cells are located intragenically, it is acceptable that they were densely methylated and lacked enrichment for active histone marks (H3K4me3 and H3K9ac). However, although tTS was expressed and functional in both cells (SW48A/tTS and HCT116D/tTS), the one in HCT116D/tTS served almost nothing for methylation spreading. This brings up the question, to what extent transgene methylation is subjected by a position or a large domain centered over the position, which needs detailed investigation in the future.

Finally, the disconnection of histone modifications and DNA methylation in the CGI-promoter was further revealed from several aspects. The first one is non-appearance of methylation spreading in host cells except SW48A/tTS while reporter expression was gradually suppressed and the promoter was highly enriched with H3K9me3. Second, methylation was hardly recruited into the CGI of the truncated pINSL6 devoid of the LINE element even in SW48A/tTS. Moreover, treatment with DAC only was not able to recover GFP expression along with DNA demethylation. Among the set of factors recruited by tTS binding, HP1 and the H3K9-specific HMT (e.g. SETDB1) have been shown interplayed with DNMTs (127), and the above findings confirmed DNA methylation as the consequential event of transcriptional silencing, but they also indicated transcription is more closely related to chromatin conformation and DNA methylation may be an event reflecting relatively

stronger local suppression. The disconnection of histone modifications and DNA methylation, therefore, could provide the flexibility for dynamic epigenetic machinery in regulating gene transcription.

To summarize, by controlling variable elements in a site-specific integration system, we investigated several aspects involved in triggering DNA methylation: (1) repetitive elements adjacent to CGI-promoters that are frequently hypermethylated in cancer cells; (2) the extent of transcriptional repression in regulating DNA methylation; (3) the effect of genomic position/cell line on DNA methylation. We were able to demonstrate all of them could influence methylation patterns around a CGI-promoter so as to confer flexibility in epigenetically regulating gene silencing under various circumstances.

Chapter 5

Introduction of DNMT3B lead to partially recovered methylation patterns

Introduction

In the previous sections, we utilized transgenes to investigate cis- and trans-regulation in de novo DNA methylation and spreading. Because integration sites were kept consistent, the comparison is more reliable than random integration. However, we still did not define what kinds of sites/sequences/motifs are favored by different DNMTs. In order to answer this question, we used a well-defined colorectal cancer cell with double disruption of DNMT1 and DNMT3B as another model system, and re-expressed DNMT3B or DNMT1. A newly developed method, DREAM, was employed to search for re-methylated CG sites, aided by bioinformatic tools.

HCT116 DKO (DNMT1^{-/-} DNMT3b^{-/-}) was generated by serial homologous recombinations and does not keep the methylation patterns in HCT116, because about 95% of genomic methylation is lost (8, 9). Interestingly, the P16INK4A promoter could gradually regain methylation if culture time was prolonged to over 80 passages. Although this is likely due to remnant activity of DNMT3A and alternatively spliced DNMT1, remethylation suggested signals for DNA methylation have not been lost. Hence, this genetically engineered cell line could serve as a model to assess the preference of DNMT3B or DNMT1 in remethylation, so as to shed light on the machinery of de novo methylation in cancer cells.

Results

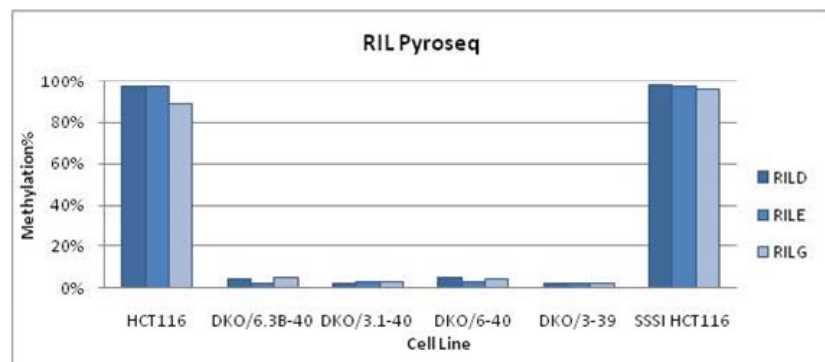
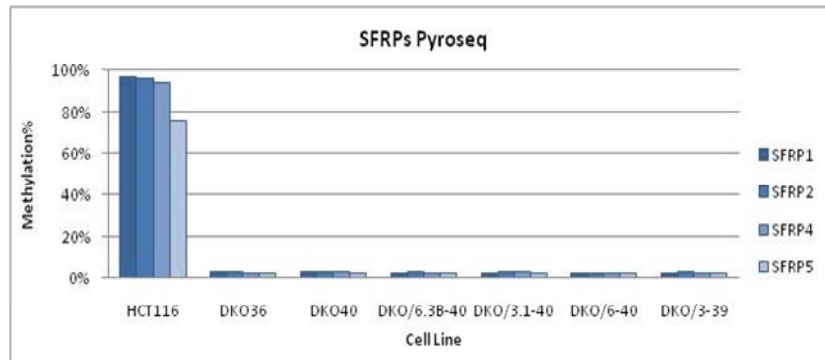
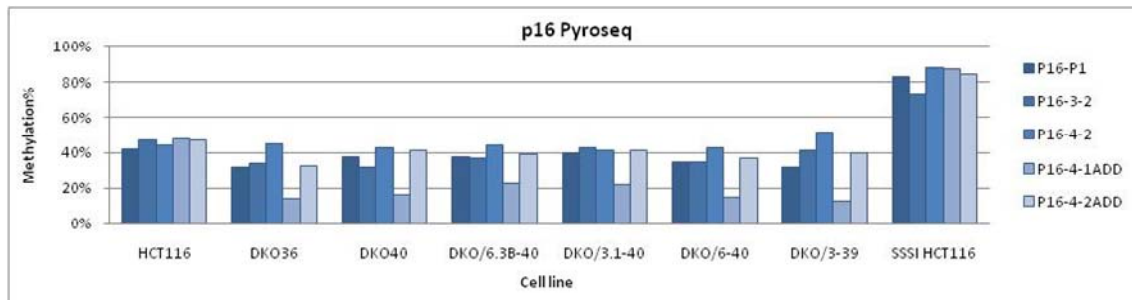
Methylation is not equally recovered in DKO cells

HCT116 DKO (DNMT1^{-/-} DNMT3B^{-/-}) was generated through serial homologous recombinations to disrupt methylation machinery in the colorectal cancer cell HCT116. Both genes were knocked down in part of their genomic sequences. DNMT1 was targeted at exon 3-5, so that global methylated cytosine was reduced by ~20% and enzyme activity was reduced by ~93%. Gene-specific methylation was not equal in DNMT1^{-/-}. For instance, a single-copy gene, p16INK4A was still methylated ~50%, and TIMP-3 was still fully methylated and silenced (130). For DNMT3B, exon 2-21 was replaced to remove active site critical to catalysis, therefore DKO cells lost ~95% methylated cytosine and almost all enzyme activity (9). The gene-specific methylation was also suffered, as shown by a set of genes identified to have increased expression in DKO due to loss of methylation through microarray screening in HCT116 and DKO (131). However, the deletion of DNMT1 was not complete, and there was alternatively spliced transcript detectable with catalytic domains (132). Further fully disruption of the exons necessary for active DNMT1 in HCT116 resulted in hemimethylation of ~20% of CG-CG dyads and caused G2 arrest and cell death (133). Due to the remaining activity of incompletely disrupted DNMT1 and DNMT3A, DKO cells exhibited partially recovered methylation and growth rate at passage 87 (93), although it became significant at least from passage 50.

Before stable re-expression of DNMT3B or DNMT1, we assessed methylation status of the DKO clones (passage 35 to passage 40) to define a starting level for our experiments. Assays for pyrosequencing were used not only for gene-specific methylation but also for global methylation. In contrast to heavy methylation of over 80%, the CGIs of SFRP1,

SFRP2, SFRP4, SFRP5 and RIL were almost methylation-free. The single-copy gene, p16INK4A had its CGI re-methylated to the same level of ~50% as in parental HCT116 cells, except for one region (named as p16-4-1ADD) that was not fully restored (~20%). We also examined mixed populations of DKO transfected with DNMT3B1 or DNMT1b, and observed similar results as in DKO cells. LINE-1 methylation could represent global DNA methylation levels. We picked 13 single clones (passage 48) of DKO cells and their methylation levels were somewhere between passage 45 and passage 50, but there were variations in clonal status (Figure 30). This suggested the heterogeneous conditions of single DKO cells in DNA methylation.

A



B

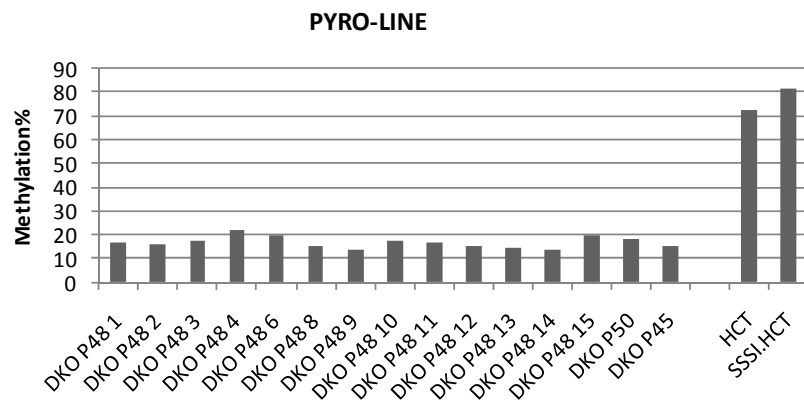
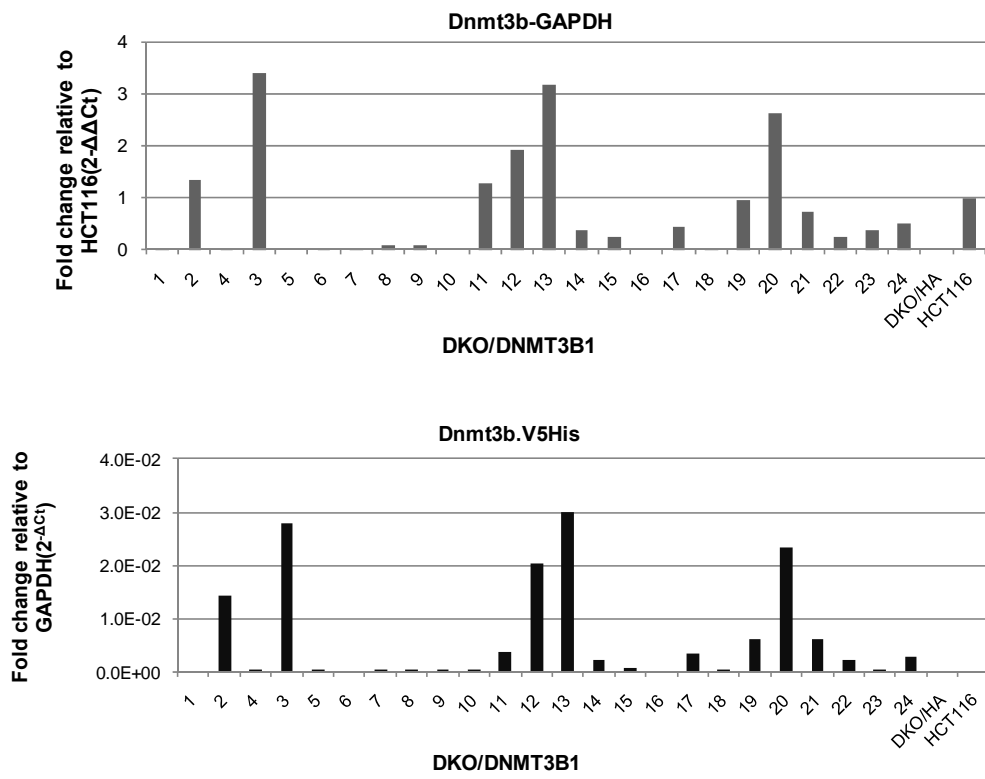
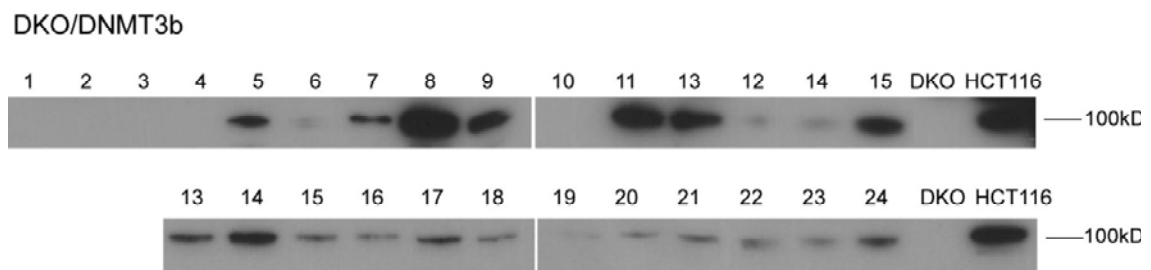


Figure 30. Methylation is not equally recovered in DKO cells. Pyrosequencing to examine average methylation levels at several CGI-promoter regions. HCT116, wild type; DKO36 and DKO40, DKO cells of passage36 and passage40; DKO/6.3B-40, DKO/3.1-40, DKO/6-40 and DKO/3-39 are mixed DKO clones transfected with pcDNA6V5His/DNMT3B, pCDNA3.1HA/DNMT1, pcDNA6V5His, and pcDNA3.1HA; SSSI HCT116, methylase sssI-treated HCT116 genomic DNA. (A) DNA methylation in the promoter-CGI of p16INK4A, SFRP1, SFRP2, SFRP4, SFRP5, and RIL. (B) Clonal variation of LINE-1 methylation in DKO (passage 48).

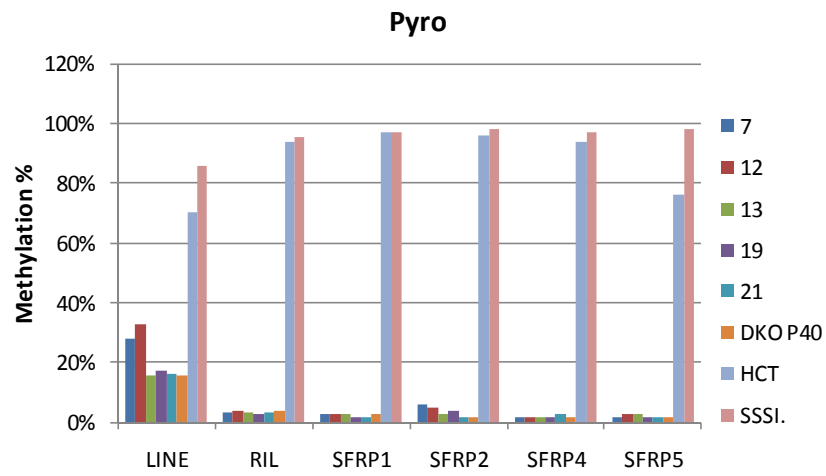
A



B



C



D

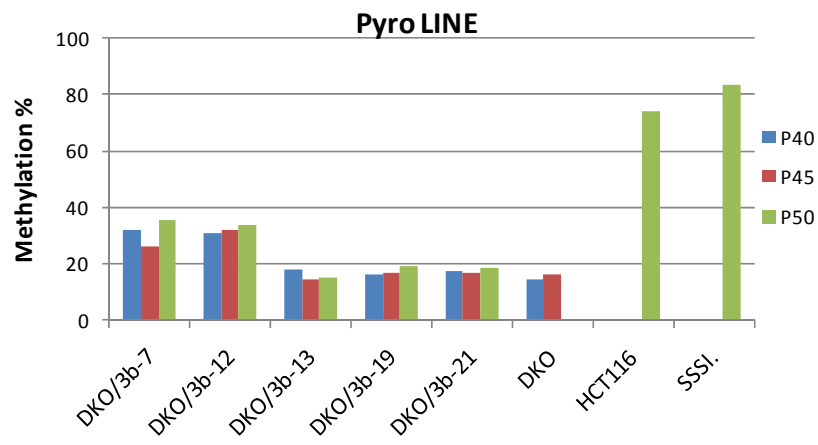
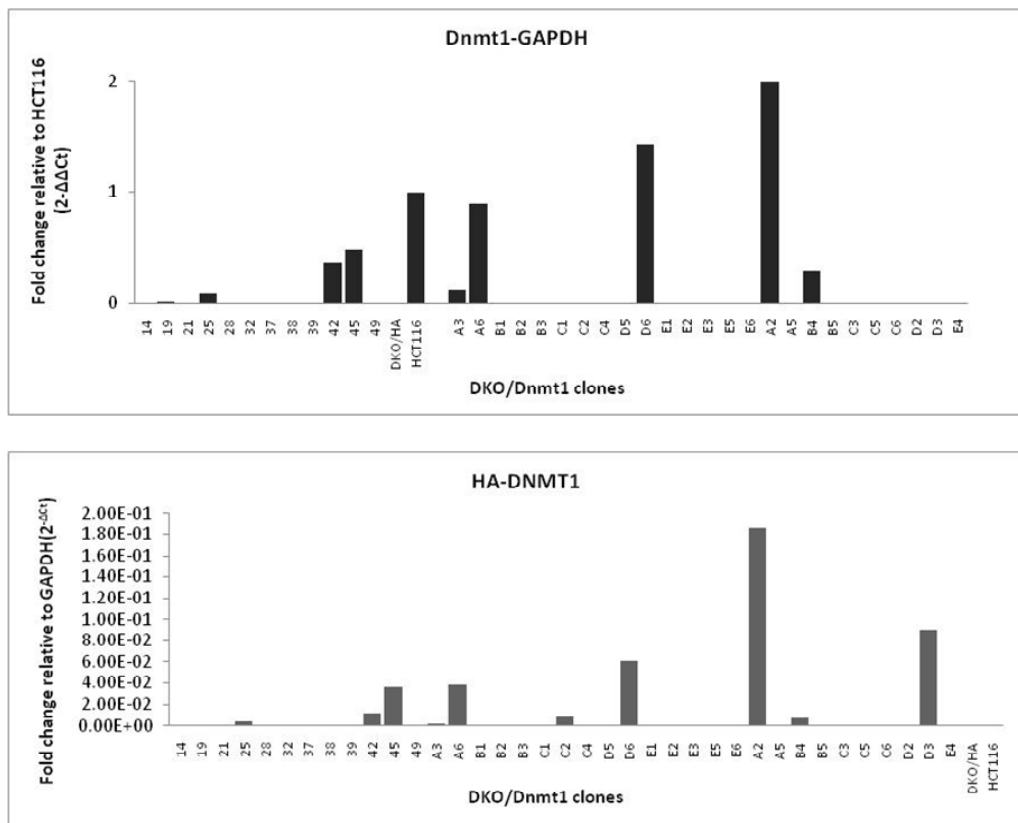


Figure 31. Re-expression of DNMT3B1 in DKO cells moderately recovered global DNA methylation. (A) Quantitative RT-PCR to screen out stable single clones with positive expression of DNMT3B1. Two Taqman assays were designed to cover the deleted exons and junction of DNMT3B-to-V5His in the transcript. Ct values were normalized to GAPDH. (B) Western blot to test the expression at the protein level. (C) Pyrosequencing to single clones with partial re-methylation of LINE-1. Gene-specific methylation was shown for the CGIs of RIL, SFRP1, SFRP2, SFRP4 and SFRP5. (D) Pyrosequencing results of LINE-1 in single clones from passage 40 to passage 50.

A



B

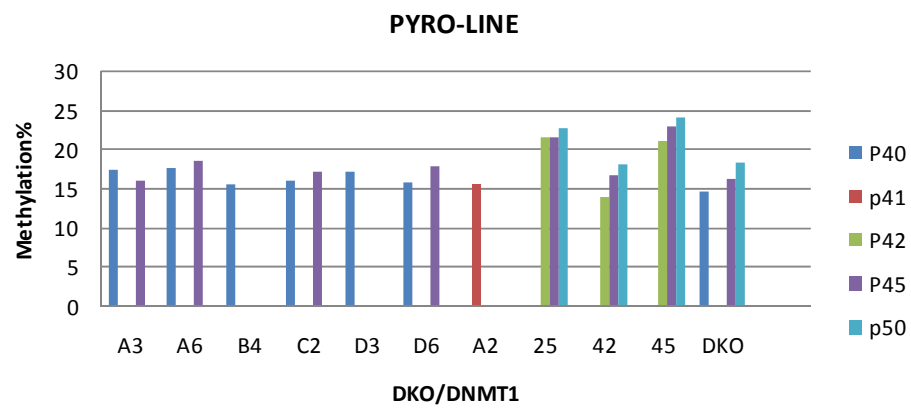


Figure 32. Re-expression of DNMT1b in DKO cells moderately recovered global DNA methylation. (A) Quantitative RT-PCR to screen out stable single clones with positive expression of DNMT1b. Two Taqman assays were designed to cover the deleted exons and junction of HA-to-DNMT1b in the transcript. Ct values were normalized to GAPDH. (B) Pyrosequencing results of LINE-1 methylation in single clones from passage 40 to passage 50.

Re-expression of DNMT3B or DNMT1 moderately recovered global DNA methylation

DKO cells (passage 36 or 37) were transfected with pCDNA6V5His/DNMT3B1 (blasticidin) or pCDNA3.1HA/DNMT1B (zeocin). The stable single clones were selected under blasticidin or zeocin respectively. The expression was examined by RT-PCR and western blotting. Besides single clones with detectable expression of DNMT3B or DNMT1, we had to select which clones regain global methylation or/and gene-specific methylation. Pyrosequencing results showed in Figure 31 and Figure 32 described most of the picked single clones did not have elevation in their methylation levels compared with the same passage number of DKO cells, and there were only two clones, 3B-7 and 3B-12, whose LINE-1 methylation increased by ~18% and ~16% (passage 40) while none of examined gene-specific methylation had any changes (RIL, SFRP1, SFRP2, SFRP4 and SFRP5). There is no need to test p16 methylation since it has already recovered methylation. As for transfection of DNMT1, two clones (3.1-25 and 3.1-45) regained LINE-1 methylation of 5%~6% (passage 45) which was not much compared with DNMT3b re-expression.

Preferential representation of CGI methylation for DREAM

We first sequenced DREAM libraries (part A and B) prepared from HCT116, DKO (passage 36), 3B-7 and 3B-12 (passage 52). The experiments for validation before sequencing are shown in Figure 33. The size of each library was checked on 2% agarose gel and effective removal of primer dimers was validated on Bioanalyzer (Agilent) by the sequencing core facility (M.D. Anderson Cancer Center). Pyrosequencing results demonstrated good correlation between theoretical and actual methylation levels of spiked DNAs. TA-cloning/sequencing of libraries also indicated good quality of each library, for

instance, out of 20 effective CCCGGG sites (from 10 correct TA-clones) for HCT116 part B (HCT116-B), there were 11 XmaI (CCC) sites and 9 SmaI (GGG) sites, whereby methylation percentage was 55%; for DKO-A and DKO-B, 0% and 13.6%, which were within the expectation range.

Here, DREAM libraries were sequenced on GAI pipeline and pre-analyzed by bioinformatic tools commercially available (ELAND) and specifically designed for DREAM in Department of Bioinformatics and Computational Biology (Dr.Shoudan Liang). Overview of the results is listed in Figure 34. Coverage of SmaI sites for each sample includes both part A and part B. Excluding tags from spiked DNA, the average methylation levels of HCT116, DKO, 3B-7 and 3B-12 were 44.0%, 5.2%, 22.8% and 17.6%, which were consistent with expected differences. Apparently, re-expression of DNMT3B induced moderate re-methylation in DKO cells.

Because DREAM utilizes restriction enzyme-based digestion, the resolution is merely restricted to regions with SmaI/XmaI sites (CCCGGG), which appear once out of every 4096 bases theoretically. Genomic distribution of this site has its own preference. Namely, the entire human genome has 378855 SmaI sites (Hg18) (220 with no specific location information on a “random” chromosome), out of which only 11.76% are in CGI (44349 sites) and 61.81% are in various repeats (234182 sites) (Figure 35). Whereas, the sites obtained from sequencing all of four samples (HCT116, DKO, 3B-7 and 3B-12) were counted by selecting those with ≥ 5 tags, which covered 20.79% (78762 sites) of genomic SmaI sites. If taking the properties of location into account, we found that 66.65% (29557 sites) of genomic CGI-SmaI sites (44349 sites) and 14.71% (49205 sites) of genomic nonCGI-SmaI sites (334506 sites) were represented by DREAM; as for location in repeats, 4.68% (10963 sites)

of genomic repeats-SmaI sites (234182 sites) and 46.86% (67799 sites) of genomic non-repeats-SmaI sites (144673 sites) were obtained. These series of numbers suggested that DREAM, as restriction enzyme-based technology, is more likely to reflect methylation of SmaI sites in CGIs and has lost much information of those in repeats.

Re-expressed DNMT3B favored non-CGIs

Re-introduction of DNMT3B into DKO cells induced moderate regain methylation in selective SmaI sites. Here, we took the common set of SmaI sites covered in all four libraries (HCT116, DKO, 3B-7 and 3B-12) and every site has more than 5 tags sequenced (≥ 5 tags). The scatter plots in Figure 36A display methylation levels in HCT116, 3B-7 and 3B-12 versus DKO. We aimed to identify sites which were greatly demethylated in DKO cells and then obtained significant increase in 3B-7 and 3B-12. Therefore, we first applied Fisher's exact test (two-tailed) to comparisons between HCT116 and DKO, 3B-7 and DKO, and 3B-12 and DKO as to methylated and unmethylated tag numbers. The p-value was adjusted by Benjamini-Hochberg method and sites with adjusted p-value (q-value) <0.05 were assigned to the set with significant difference in DNA methylation. Thus, 52579 sites for HCT116 versus DKO, 30766 sites for 3B-7 versus DKO, and 26249 sites for 3B-12 versus DKO had significantly different methylation levels. The number for the common set of significantly different sites was 21793, which was 27.67% of the total common set. If location properties considered, the percentage of CGI-SmaI sites in the significant set (12.20%) was less than that of non-CGI-SmaI sites (36.96%), while there was similar percentages of repeat-SmaI sites (24.75%) to non-repeat-SmaI sites (28.14%). Figure 36B shows the general levels of sites with significant difference in the above two-two comparisons.

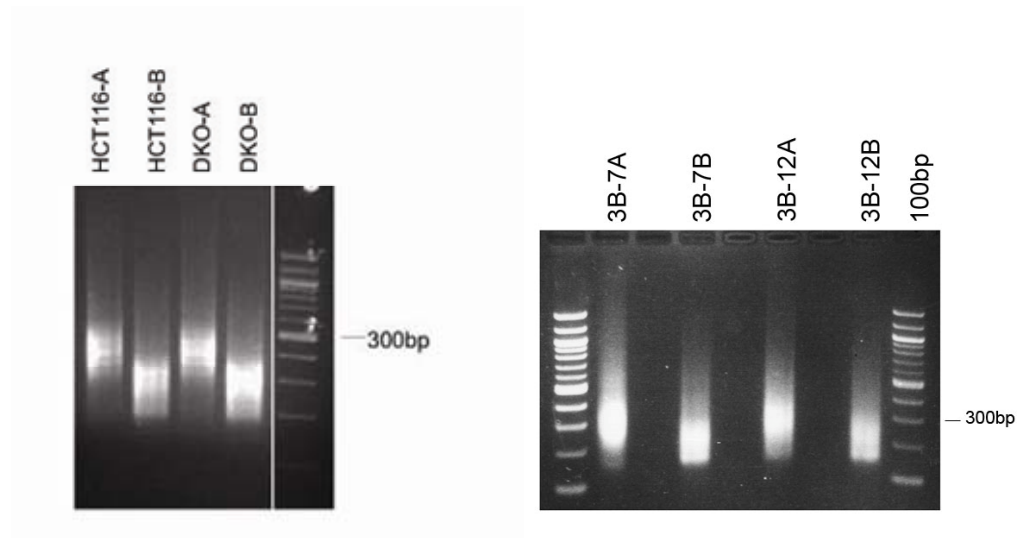
However, when definition of “methylated” and “unmethylated” required, a second criterium was applied using 20% as the cut-off for methylation levels. Among the set of sites whose methylation became significantly different in DKO ($p < 0.05$ in HCT116 vs. DKO), the portion of remethylated ones ($p < 0.05$ for all, methylation $\geq 20\%$ in HCT116, 3B-7 and 3B-12, and methylation $< 20\%$ in DKO) were found less in CGI-SmaI-sites (13.15%) than in non-CGI-SmaI sites (26.44%). Consistently, the portion of un-remethylated one ($p < 0.05$ for HCT116 vs. DKO, $p > 0.05$ for both 3B-7 vs. DKO and 3B-12 vs. DKO, methylation $\geq 20\%$ in HCT116 and methylation $< 20\%$ in DKO, 3B-7 and 3B-12) were more likely in CGI-SmaI sites (42.52%) than non-CGI-SmaI sites (13.84%) (Figure 36). This suggested CGIs were less sensitive to DNMT3B due to unclear mechanisms.

Promoters were more resistant to re-expressed DNMT3B

We took the above significantly remethylated and un-remethylated groups of SmaI sites, and analyzed their locations to find out the regions which were resistant to DNMT3B. The percentage of SmaI sites for each sliding window of 500bp was calculated in regard to the location property, and lines were plotted with relative distance to the closed transcription start site (TSS) for each percentage (Figure 37). As shown before, more CGI-SmaI sites were prone to be un-remethylated. A further finding was also as expected that promoters were more resistant for remethylation.

We listed the genes that have CGI-promoters and whose CGI-promoters achieved significant remethylation on at least one SmaI sites (Table 8). The promoter was defined as regions TSS-2000bp to TSS+500bp and the CGI-promoter has its TSS inside a CGI. Examples are also shown in Figure 39.

A



B

Pyro-seq: spiked DNA

LA	L306	LucG	EGFP	T353	T324
0%	22%	22%	50%	70%	100%

TA-seq

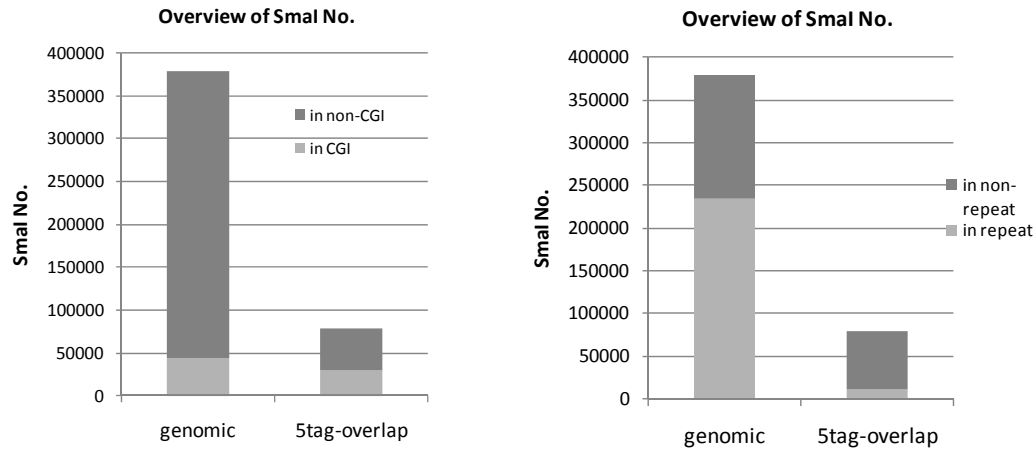
Library	No. SmaI	No. XmaI	Sites/clones	sent
HCT116-A	5	13	18/9	12
HCT116-B	9	11	20/10	12
DKO-A	22	0	22/11	12
DKO-B	19	3	22/11	12

Figure 33. Preparation of DREAM libraries. (A) Pictures of prepared libraries after size selection (200bp ~ 500bp). Part A and B are equal slices from the same gDNA. (B) Validations of prepared libraries by pyrosequencing of spiked DNAs and TA-cloning/sequencing.

HCT116		
Avg.Met%	44.00	
no. Smal sites	151,638	
reference Smal sites	7	
total sites covered	151,645	
total tag number	40,629,732	
total reference tags	2,287,296	
	no. sites (exclude ref.)	total tags
sites(>0tag)	151,638	38,342,436
sites(>=5tags)	118,422	38,266,227
sites(>=10tags)	96,768	38,119,632
sites(>=20tags)	75,259	37,822,520
DKO		
Avg.Met%	5.20	
no. Smal sites	140,330	
reference Smal sites	7	
total sites covered	140,337	
total tag number	35,977,471	
total reference tags	1,826,673	
	no. sites (exclude ref.)	total tags
sites(>0tag)	140,330	34,150,798
sites(>=5tags)	82,841	34,019,190
sites(>=10tags)	58,428	33,862,611
sites(>=20tags)	51,641	33,776,346
3B.7		
Avg.Met%	22.81	
no. Smal sites	162,895	
reference Smal sites	7	
total sites covered	162,902	
total tag number	53,956,332	
total reference tags	3,393,334	
	no. sites (exclude ref.)	total tags
sites(>0tag)	162,895	50,562,998
sites(>=5tags)	145,528	50,520,152
sites(>=10tags)	126,320	50,387,957
sites(>=20tags)	102,815	50,058,143
3B.12		
Avg.Met%	17.66	
no. Smal sites	164,679	
reference Smal sites	7	
total sites covered	164,686	
total tag number	52,422,055	
total reference tags	2,680,276	
	no. sites (exclude ref.)	total tags
sites(>0tag)	164,679	49,741,779
sites(>=5tags)	153,789	49,713,419
sites(>=10tags)	135,166	49,581,363
sites(>=20tags)	103,135	49,130,066

Figure 34. Overview of DREAM sequencing results. Libraries prepared from HCT116, DKO, and two single clones of DKO/DNMT3B (3B-7 and 3B-12) were sequenced by GAI (Illumina). Every table shows combined sequencing results from part A and B which were sequenced separately. Average methylation percentage was calculated as (total methylated tags) / (total unmethylated tags + total methylated tags), and all these tag numbers did not count in those from spiked DNA. Covered numbers of SmaI/XmaI site (CCCGGG) and corresponding tag numbers were also counted based on different cut-offs as to each SmaI/XmaI site.

A



B

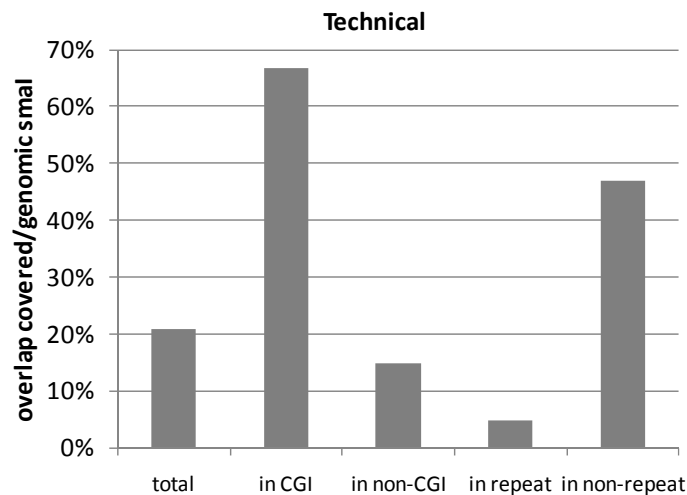


Figure 35. Technical preferences of DREAM. (A) Absolute numbers of SmaI/XmaI site in theoretical human genome (Hg18) and those obtained in all the sequenced libraries of HCT117, DKO, 3B-7 and 3B-12 (sites with ≥ 5 tags). Depending on the properties of SmaI/XmaI sites, total numbers are also shown either for in CGI vs. non-CGI or for in repeat vs. non-repeat. (B) Percentage of covered SmaI/XmaI sites as relative to genomic sites. Results are also listed according to their location properties.

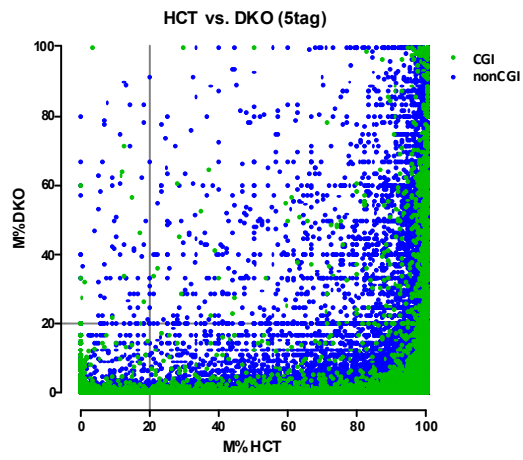
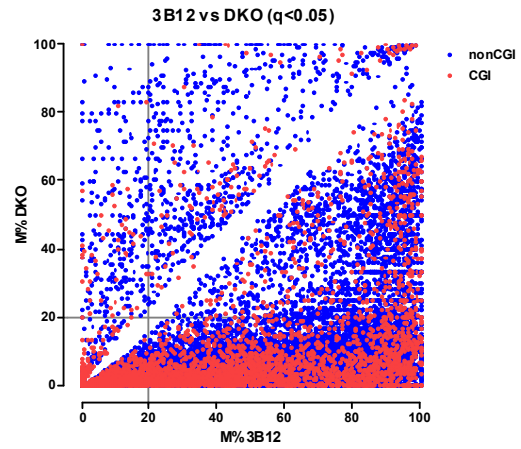
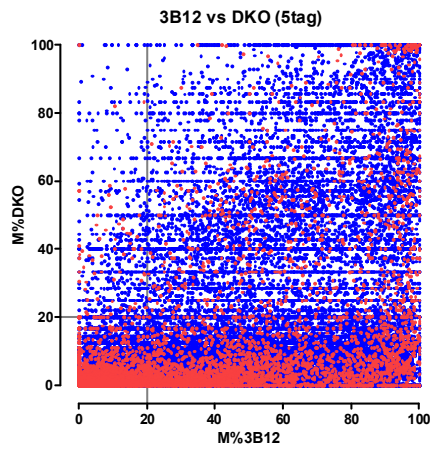
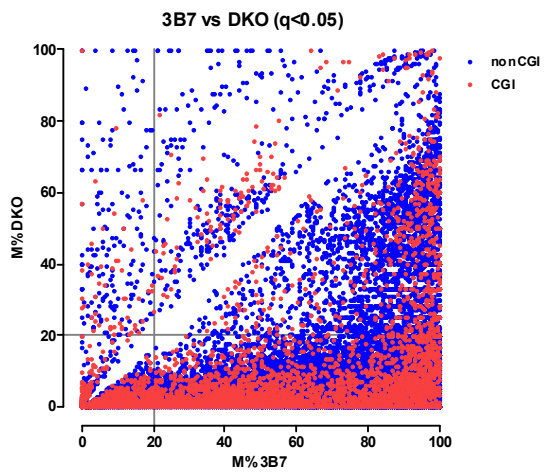
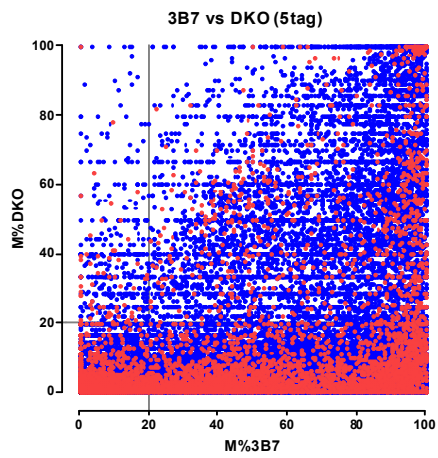
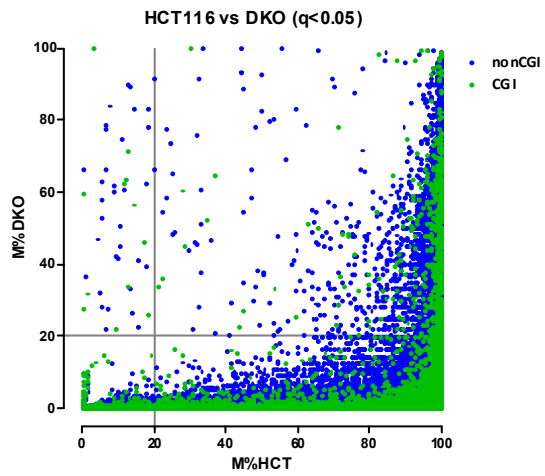
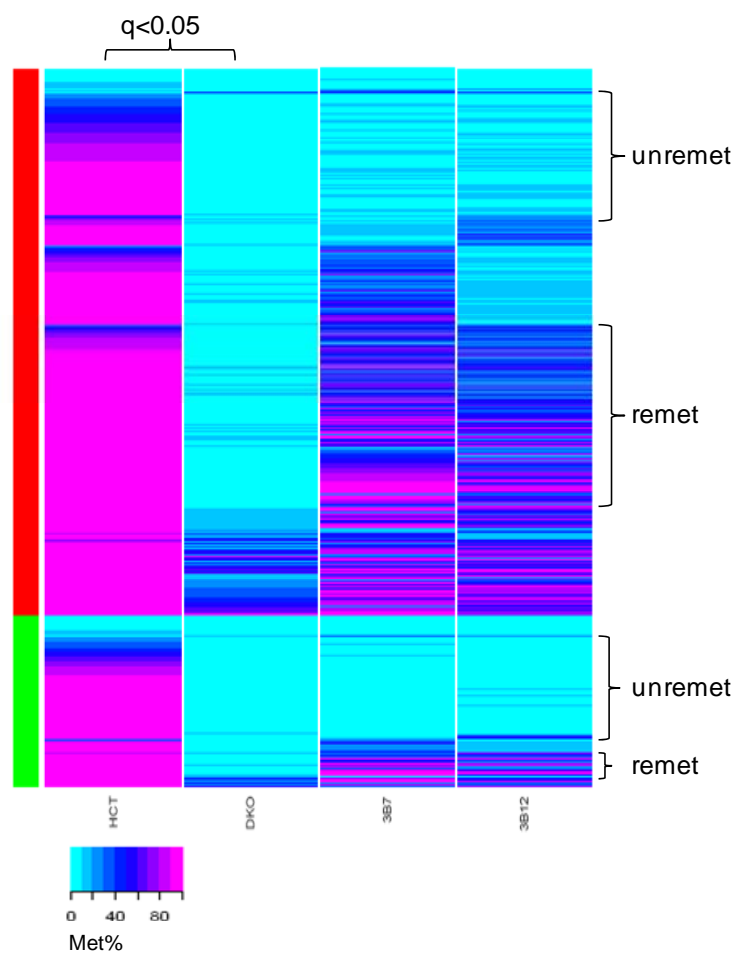
A**B**

Figure 35. Scatter plots for methylation distribution of SmaI/XmaI (CCCGGG) sites.

(A) The common set of SmaI sites obtained from all four libraries. Only sites with ≥ 5 tags were taken into account. HCT116, 3B-7 and 3B-12 are plotted versus DKO. Green dots or red dots stand for CGI-SmaI sites, and blue dots for non-CGI-SmaI sites. (B) SmaI sites with significant difference in DNA methylation ($p < 0.05$, adjusted, Benjamini and Hochberg method) between two-two comparison of HCT116, 3B-7 and 3B-12 are plotted versus DKO.

A



B

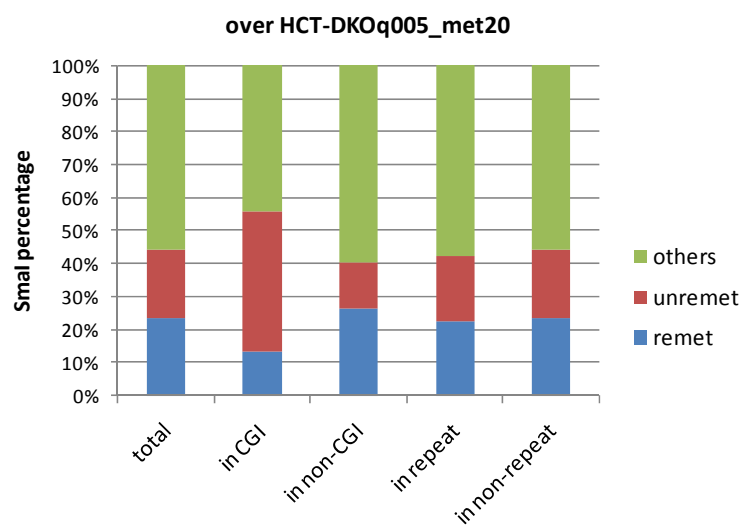


Figure 36. CGIs were not sensitive to re-expressed DNMT3B in DKO. (A) Heat map to show the methylation ranges of SmaI sites among those whose methylation was significantly different in DKO cells. Sites were classified into non-CGI-SmaI sites (red bar) and CGI-SmaI sites (green bar). Methylation levels are indicated with colors from 0% to 100%. The sets of interest are also pointed out, which are sites remethylated and un-remethylated in 3B-7 and 3B-12. (B) Calculation of the percentages of each categories with respect to location and changes.

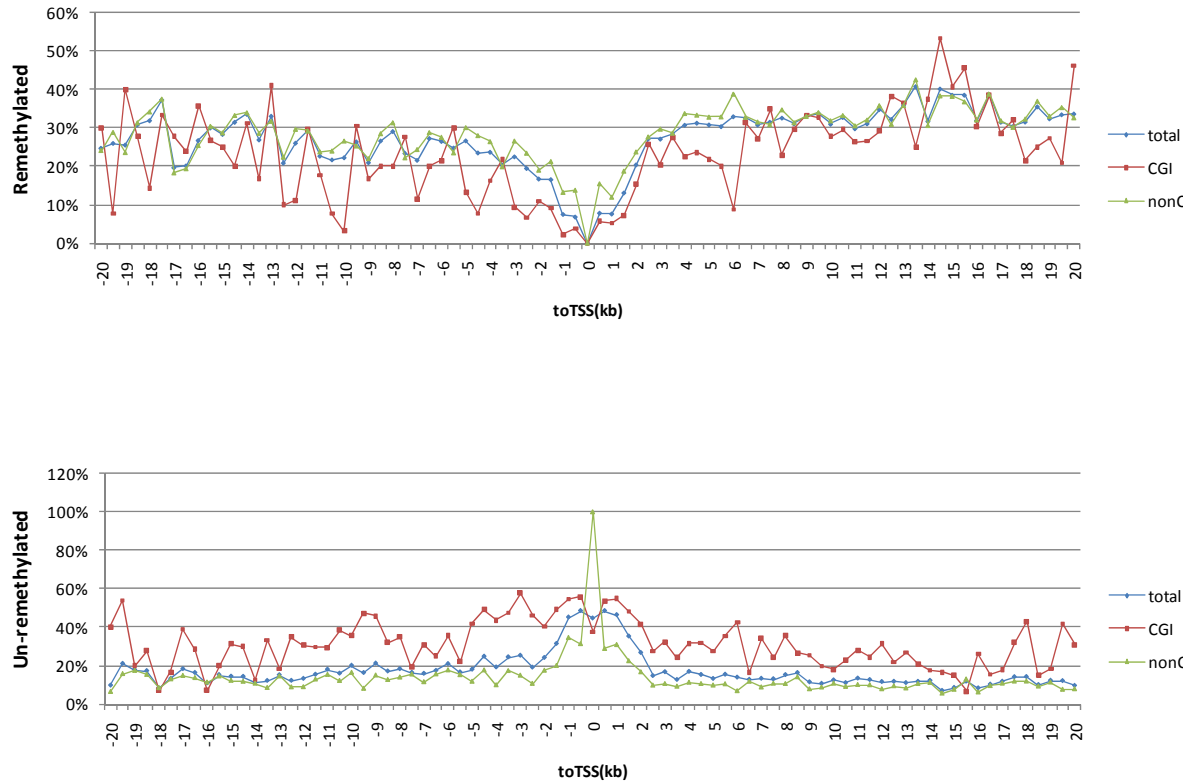


Figure 37. Percentages of SmaI sites remethylated (top) and un-remethylated (bottom) in respect to their distance to the closest transcription start sites (TSS). CGI- and non-CGI- SmaI sites are plotted for each case. Both p -value < 0.05 and methylation level (20%) were considered to calculate the percentages.

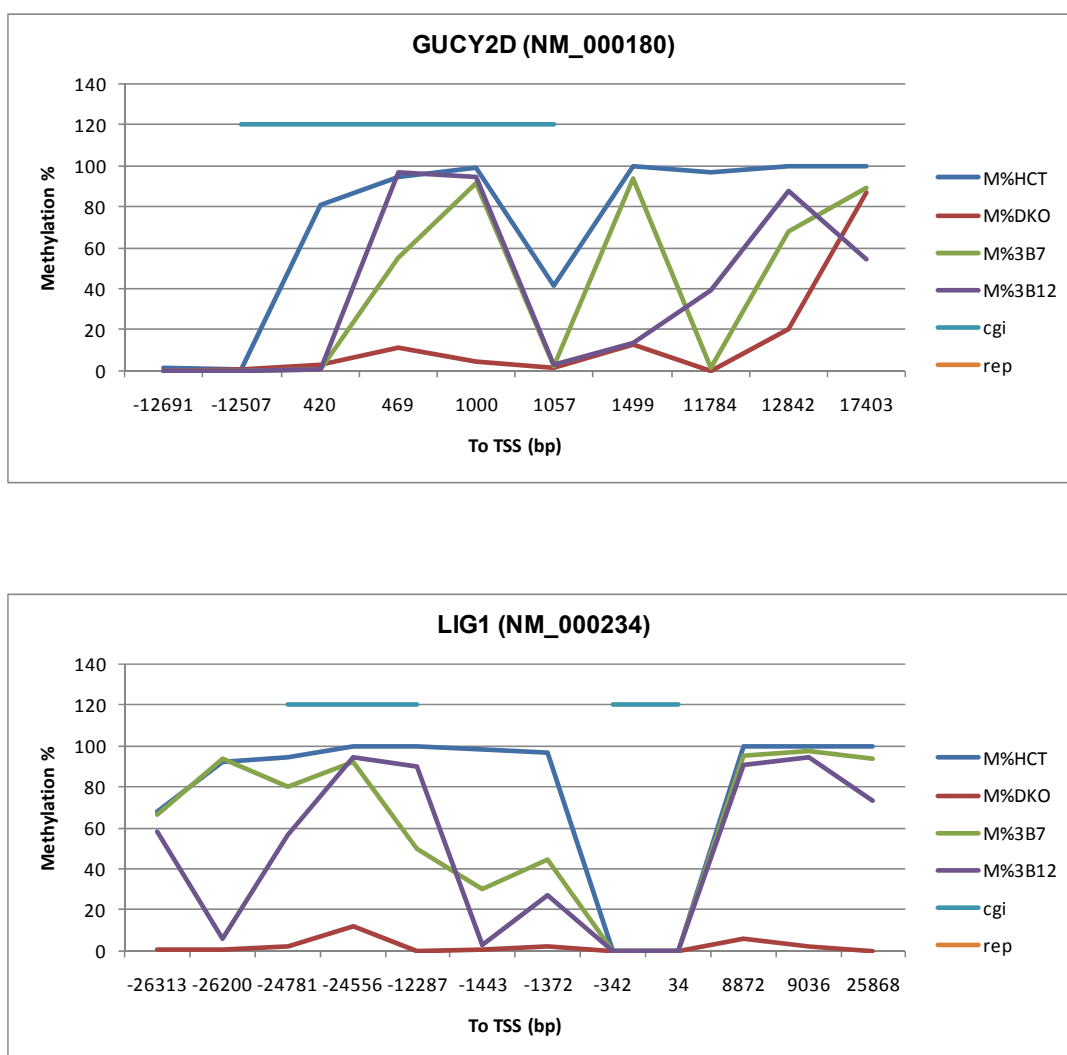


Figure 38. Examples of remethylation and un-remethylation of CGIs in CGI-promoters.

Two genes (GUCY2D and LIG1) are shown here. Promoter is the region of TSS-2000bp to TSS+500bp; CGI-promoters have the TSS inside the CGIs.

Table 8. A list of genes with significant remethylation on at least one SmaI site of CGI-promoters.

RefSeq_GeneA	Name_GeneA		RefSeq_GeneA	Name_GeneA		RefSeq_GeneA	Name_GeneA
NM_000093	COL5A1		NM_004794	RAB33A		NM_032451	SPIRE2
NM_000180	GUCY2D		NM_005028	PIP4K2A		NM_032638	GATA2
NM_000234	LIG1		NM_005048	PTH2R		NM_054020	CATSPER2
NM_000475	NR0B1		NM_005052	RAC3		NM_058197	CDKN2A
NM_000660	TGFB1		NM_005398	PPP1R3C		NM_080823	SRMS
NM_000697	ALOX12		NM_005519	HMX2		NM_133261	GIPC3
NM_000799	EPO		NM_005586	MDFI		NM_133457	EMID2
NM_000843	GRM6		NM_005662	VDAC3		NM_138437	GPRASP2
NM_000863	HTR1B		NM_005961	MUC6		NM_138632	TRIOBP
NM_000961	PTGIS		NM_006262	PRPH		NM_138689	PPP1R14B
NM_001001852	PIM3		NM_006315	PCGF3		NM_138769	RHOT2
NM_001002034	FAM109B		NM_006794	GPR75		NM_145004	ADAM32
NM_001003841	SLC6A19		NM_006936	SUMO3		NM_145282	LOC153328
NM_001003938	HBM		NM_006953	UPK3A		NM_147164	CNTFR
NM_001010887	ACER2		NM_006984	CLDN10		NM_152310	ELOVL3
NM_001014980	FAM132A		NM_007056	SFRS16		NM_152331	ACOT4
NM_001025290	DPPA5		NM_009586	SIM2		NM_152355	ZNF441
NM_001025436	SPAG16		NM_012458	TIMM13		NM_152380	TBX15
NM_001029864	KIAA1755		NM_013267	GLS2		NM_152640	DCP1B
NM_001031734	FDX1L		NM_014037	SLC6A16		NM_153221	CILP2
NM_001044370	MPPED1		NM_014350	TNFAIP8		NM_153230	FBXO39
NM_001045476	WDR38		NM_014379	KCNV1		NM_153270	KLHL34
NM_001063	TF		NM_014625	NPHS2		NM_153334	SCARF2
NM_001080433	CCDC85A		NM_014898	ZFP30		NM_153449	SLC2A14
NM_001080473	MFSD2B		NM_015356	SCRIB		NM_172057	KCNH2
NM_001080509	TSPAN11		NM_015456	COBRA1		NM_172389	NFATC1
NM_001085401	C6orf201		NM_015688	FAM184B		NM_173516	PNLDC1
NM_001085480	FAM162B		NM_016002	SCCPDH		NM_173540	FUT11
NM_001098517	CADM1		NM_016170	TLX2		NM_173547	TRIM65
NM_001098519	LRRC43		NM_016463	CXXC5		NM_173560	RFX6
NM_001103167	ZGLP1		NM_016529	ATP8A2		NM_173563	C6orf146
NM_001111034	ACP5		NM_016564	CEND1		NM_175619	ZAR1
NM_001115016	KIAA1310		NM_016592	GNAS		NM_177478	FTMT
NM_001122636	GALNT9		NM_016831	PER3		NM_178861	RNF113B
NM_001127612	PAX6		NM_017516	RAB39		NM_181689	NNAT
NM_001127688	BEX4		NM_017614	BHMT2		NM_181723	EFHA2
NM_001129895	HGC6.3		NM_017655	GIPC2		NM_181756	ZNF233
NM_001130011	TEX101		NM_017767	SLC39A4		NM_182539	TCTE1
NM_001134431	TOR2A		NM_017891	C1orf159		NM_183374	CYP26C1
NM_001135197	CCDC36		NM_017918	CCDC109B		NM_198389	PDPN
NM_001142575	IMPDH1		NM_018346	RSAD1		NM_198494	ZNF642
NM_001142634	SPDYA		NM_018355	ZNF415		NM_198496	VWA2
NM_001144382	PLCL2		NM_018467	USE1		NM_198537	YJEFN3

NM_001145030	C3orf77		NM_018487	TMEM176A		NM_198571	C7orf52
NM_001145132	C5orf52		NM_018488	TBX4		NM_198580	SLC27A1
NM_001145650	ZNF529		NM_018942	HMX1		NM_199037	SCN1B
NM_001145657	RAP1GAP		NM_018951	HOXA10		NM_199054	MKNK2
NM_001146054	SNCA		NM_018962	DSCR6		NM_199235	COLEC11
NM_001159330	ARHGAP27		NM_019112	ABCA7		NM_201379	PLEC
NM_001160226	CNR1		NM_020062	SLC2A4RG		NM_203425	C17orf82
NM_001161625	NXNL2		NM_020134	DPYSL5		NM_206833	CTXN1
NM_001164405	BHLHA9		NM_020404	CD248		NM_206860	TACC2
NM_001166412	SMOC2		NM_020804	PACSIN1		NM_206894	ZNF790
NM_001167676	LOC100128071		NM_021116	ADCY1		NM_207437	DNAH10
NM_001172675	ZNF347		NM_021649	TICAM2		NR_001458	MIR155HG
NM_001485	GBX2		NM_021831	AGBL5		NR_002781	TSPYL3
NM_001801	CDO1		NM_022405	SLC6A20		NR_002814	LOC374443
NM_002237	KCNG1		NM_022829	SLC13A3		NR_003366	ANKRD20B
NM_002402	MEST		NM_023070	ZNF643		NR_004382	MESTIT1
NM_002496	NDUFS8		NM_024017	HOXB9		NR_024349	LOC284023
NM_002508	NID1		NM_024560	ACSS3		NR_024418	LOC389332
NM_002531	NTSR1		NM_024600	TMEM204		NR_024523	LOC641367
NM_002754	MAPK13		NM_024663	NPEPL1		NR_026807	C6orf155
NM_003012	SFRP1		NM_024677	NSUN7		NR_026880	MGC12916
NM_003253	TIAM1		NM_024718	C9orf86		NR_028334	FLJ40504
NM_003269	NR2E1		NM_024857	ATAD5		NR_028343	ZNF415
NM_003353	UCN		NM_025268	TMEM121		NR_028501	ECEL1P2
NM_003550	MAD1L1		NM_030613	ZFP2		NR_029837	MIR219-2
NM_004210	NEURL		NM_031909	C1QTNF4		NR_030638	MIR941-2
NM_004324	BAX		NM_031925	TMEM120A		NR_030641	MIR943
NM_004473	FOXE1		NM_031947	SLC25A2		NR_030646	MIR1225
NM_004539	NARS		NM_032134	QRICH2		NR_031602	MIR1237
						NR_031730	MIR1909

Discussion

DREAM (Digital restriction enzyme analysis of methylation) is a recently developed methodology coupling specific amplification of fragmented genomic DNA by restriction enzymes with the next-generation sequencing technology. This method has its distinct advantage to quantitatively measure the methylation levels of single CGs genome-widely due to paired usage of methylation-sensitive and methylation-insensitive restriction endonucleases (e.g. SmaI and XmaI for CCCGGG).

On the other hand, the recognition site of CCCGGG restricts the representation of DREAM to certain locations. In human genome (Hg18), 11.76% of SmaI sites are in CGI and 61.81% are in various repeats. However, the access of endonucleases to genomic areas of variable condensity would limit complete digestion, and size selection during library preparation must have caused loss of some fragments of over 500-bp length due to sparsely located SmaI sites. Additionally, it is not guaranteed by the sequencing platform that all of single molecules could be clustered and sequenced, because amplification process could not be avoided using next-generation sequencing. Therefore, instead of representing non-CGI or repeats, DREAM results obtained from four samples showed higher percentage of site coverage in CGIs and loss of information of repeats. Therefore, it would benefit our aim to study DNA methylation in CGI-promoters.

From the datasets of significantly remethylated SmaI sites, we could infer two preferences of re-expressed DNMT3B in DKO cells. (1) CGIs, no matter whose locations are, were less vulnerable to DNA methylation; (2) Promoters which became demethylated were more resistant to DNMT3B. For the first one, the percentage of remethylated CGIs was less than that of remethylated non-CGIs, which, in some sense, was consistent with previous

findings about CGI-methylation. Also, the more immunity of promoters to remethylation proved the regulatory roles of transcription in recruiting and preventing DNA methylation. DNA methylation of CGIs at annotated TSSs (promoter) has different patterns from that of CGIs in gene bodies (intragenic) or between annotated genes (intergenic). In human and mouse genomes, about 70% annotated gene promoters have CGIs and almost half of CGIs have transcription starts whereas the other half are either inside or between transcription units (“orphan CGIs”) (134). Promoter-CGIs of normal cells are typically nonmethylated, possibly because the deposition of transcription factors can be facilitated by the higher frequency of CG presence which is contained in most transcription factor binding sequences (135), and/or the presence of RNAPII at CGIs is irrespective of gene activity (136). Other mechanisms are likely involved cooperatively in protecting CGIs from DNA methylation, such as active demethylation systems induced by Tet protein family (137, 138), and active chromatin marks (e.g. H3K4me3) which would interfere with DNMT activity (139). For “orphan CGIs” of normal cells, 17% of them have been shown methylated and intragenic CGIs have more tendency to become methylated (20%-45%) than intergenic CGIs (134, 140). The resistancy of some of them could result from active transcription of non-coding RNAs, or alternative splicing of genes in a manner dispensible of DNA methylation (141, 142). In contrast, cancers possess specific aberrant patterns of CGI-methylation, and genome-wide results have not shown distinct preferences of DNA methylation in promoter-CGIs, intragenic, or intergenic CGIs such as for colorectal cancers (134). But DNA methylation profiles suggested the resemblance of cancer-specific methylation styles and normal developmental patterns (143).

Our study of remethylation location does not disprove any of the above findings, because quite a few genes are reactivated due to loss of memory of inhibition through cytosine demethylation (131). We are trying to uncover the properties of re-methylated sites using the baseline in DKO cells by the assumption that DNA methylation machinery in any cases complies with that in normal process.

Conclusions

Our project is to identify the elements involved in triggering de novo DNA methylation of CGI-promoters. Considering epigenetic regulation would not work without the cooperation between cis-elements and trans-elements, we designed two kinds of strategies to investigate both aspects. One was to observe de novo methylation in transgenes, and the other was to see if there was preference for DNMTs.

For the first strategy, we established a site-specific integration system that could allow comparison of several aspects in influencing de novo DNA methylation, specifically for CGI-promoters, under the same genomic background. The aspects of interest are (1) repetitive elements adjacent to CGI-promoters that are frequently hypermethylated in cancer cells; (2) the extent of transcriptional repression in regulating DNA methylation; (3) the effect of genomic position/cell line on DNA methylation. The system was constructed in two cancer cells with single integration sites to study the third aspect, and also a comparable cell lines with robust silencing induced by a defined transcriptional repressor were generated in parallel to study the second aspect. A set of fragments derived from two CGI-promoters (RIL and P16INK4A) surrounded by repetitive elements (SINE and LINE) were separately introduced into this system for the first aspect. We examined expression, DNA methylation and chromatin signatures of transgenes over a period of time, and found that (1) DNA methylation originates from certain CG sites within a LINE element; (2) methylation spreading into a promoter-CGI is facilitated by transcriptional repression, presence of additional repetitive elements and is site-specific; (3) transcriptional repression is required but not sufficient to promote DNA methylation. These observations are consistent with previous findings that DNA methylation serves as epigenetic locking step subsequent to gene

silencing. We further reveal that moderate gene silencing could not be sufficient for DNA methylation to occur, and triggering DNA methylation requires strong repression as well as additional influence via certain DNA sequence features, such as repetitive elements. The conclusion is described as a model in Figure 40.

For the other strategy, we re-introduced DNMT3B into a characterized cell line, HCT116 DKO (DNMT1-/- DNMT3B-/-) to see the distribution of remethylated CG sites. By profiling genome-wide patterns of DNA methylation in HCT116, DKO and two clones with re-expressed DNMT3B, we were able to substract sets of CCCGGG sites either remethylated or un-remethylated when DNMT3B was present in DKO, and find that on the baseline of methylation in DKO, CGIs are less vulnerable for DNMT3B and promoters for most genes in DKO cells are more immune to remethylation. Combined together, CGI-promoters are, once again, shown not sensitive to DNA methylation.

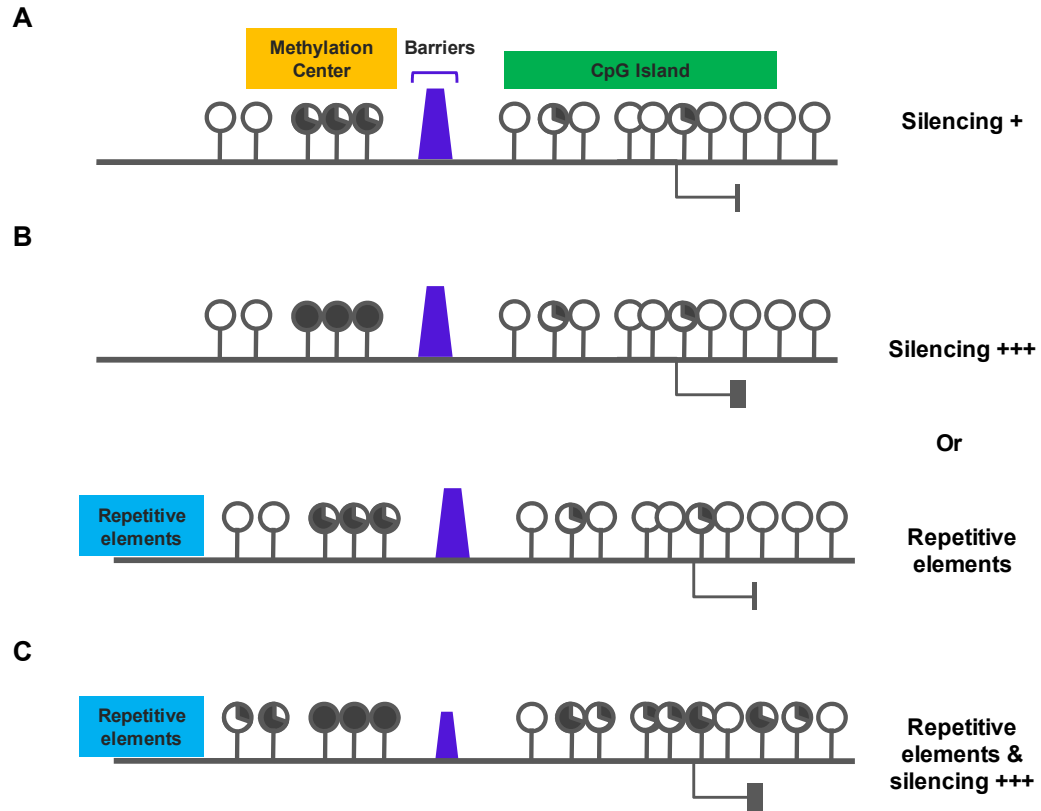


Figure 39. A Model for late response of DNA methylation to transcriptional repression.

(A) The promoter-CGI is protected from methylation by unknown mechanisms (“barriers”), while cytosines become methylated at neighbouring CG sites (“methylation center”) under moderate transcriptional repression. (B) Neither strong silencing nor presence of repetitive elements nearby can individually impose methylation spreading into the promoter-CGI, although strong silencing is able to enhance methylation center. (C) Both factors have to cooperate in order to overcome the “barrier” and facilitate CGI-methylation.

Significance and future directions

CpG-Islands (CGIs) play essential roles in regulation of gene transcription. More than half of annotated promoters in vertebrates are associated with CGIs, and quite a few of them belong to housekeeping genes, tissue-specific genes, tumor-suppressor genes and development-related genes (144). In normal cells, promoter CGIs are usually distinct due to lack of DNA methylation in vast contrast to the heavily methylated genomes. The specific mechanisms predisposing promoter CGIs against DNA methylation still remain unclear. There are several proposed elements involved in the process. First, the innate location of CGIs. Nearly half of CGIs are over TSSs (transcription start sites), and the rest are located between or within transcription units (134). CGIs often lack TATA box and are characterized by heterogeneous TSSs, leading to dispersed transcription initiations and increasing the possibility for transcription factors to bind (145). In fact, genome-wide evidences have demonstrated higher CG contents of transcription factor binding sites or recognition sites (145). Second, boundary elements. It was observed factors like CTCF, Sp1 and USF1 could serve as barriers or enhancer blockers to prevent DNA methylation spreading from the adjacent regions. This has been well illustrated in the example of H19/Igf2 imprinting locus (54). Third, DNA demethylation of promoter-CGIs. The discoveries of active demethylases suggest another possibility of CGIs to resist DNA methylation. The ability of these enzymes to convert 5'-methylcytosines to 5'-hydroxymethyl-cytosines posed that the CGIs are under sporadic de novo methylation which would be quickly removed by these enzymes, like Tet1 and Tet2. One domain of Tet1 is related with the CXXC-domain which was suggested to recognize CGIs, and recent findings further consolidated the directed demethylation by Tet1

to CGIs (137, 146). Fourth, characteristic chromatin signatures of promoter-CGIs. The active marks, H3K4me3 and histone acetylation, stands particularly for active transcription, which have long been indicated with inverse correlation with DNA methylation. While even with repressive marks, the CGIs may not always the accompanied features of silenced CGIs. Either DNA methylation occurs slower than the enrichment for inactive marks (84), or would not appear at all (112).

In our experiments, we do not have sufficient evidences to prove any of the above protection mechanisms, but rather we are aiming to first investigate possible participants for methylation spreading into CGIs. In the site-specific integration system used here, we manipulated exogenous fragments, transcriptional strength, integration sites and/or cell lines, and found that significant CGI-methylation may have to achieve seeded DNA methylation at adjacent sites. Methylation spreading could become apparent once the structural features and the transcription activity of CGI-promoters had established a local environment facilitating the access of DNMTs to CGIs. In this sense, not every promoter-CGIs are under the same chances to be sensitive to DNA methylation. Therefore, as illustrated in figure 39, inactive transcription could be easily reverted, while under the permission of sequence features together with strong repression, once DNA methylation significantly occurs within CGIs, gene silencing would be locked and propagated to the next few generations through semi-conservative inheritance of DNA methylation. Furthermore, we would infer that the variable patterns of DNA methylation in normal and abnormal development may be predetermined by gene structures, and also reflect the upstream signaling during the early time points.

However, the conclusions drawn from our experiments must be testified by further studies. Using the same site-specific integration system, we could expand the objects in order

to test the applicability of our conclusions. (1) exogenous fragments from other repetitive elements, either active or inactive, could be included; (2) instead of tetO and tTS, other transcription factors and respective binding/recognition sites could be compared in parallel; (3) more cell lines and integration sites could be constructed.

Our studies of cis-regulatory and trans-regulatory aspects for de novo DNA methylation would facilitate our understanding of methylated genes which are less sensitive to epigenetic drug treatment. Because if genes are robustly silenced by transcription factors like tTS, part of which is derived from zinc-finger proteins, and possess a very condensed local chromatin conformation, although their CGIs are methylated, treatments with epigenetic drugs (TSA or DAC) would possibly not be capable of making local chromatin decondensed enough for gene reactivation. If these genes are critical for reverting cancer cell destiny, then epigenetic drugs could not prevent relapse by only recovering partial epigenetic profiles. In our experiments, we observed the inability of TSA or DAC to reactivate transgenes with enforced transcriptional repression by tTS, thereby the above possibility could exist. Thus, if such situations influence therapeutical outcomes, further profiling insensitive genes and identification of new chemical would be necessary.

We have already found that both cis- and trans- regulatory aspects are necessary for de novo DNA methylation. However, it is still unclear what the common features of the cis-sequences are for de novo methylation, and what kinds of chromatin signatures predetermine DNA methylation. In order to investigate these questions, further studies could be applied:

- (1) Genome-wide analysis of histone marks (H3K4me3, H3K27me3 and H3K9me3) in HCT116, DKO and DKO/DNMT3B clones by ChIP-seq. Combined with gene expression profiles by microarrays and DNA methylation profiles obtained by

DREAM, the pre-existent features of transcription and epigenetics could be identified for genes/regions sensitive or refractory to DNA methylation.

- (2) The above datasets could also be combined with publicly available information of binding motifs for various transcription factors or enzymes to obtain cis-signals for DNA methylation.
- (3) On the other hand, the identified sequence features could be introduced into the site-specific system used in this project for functional validation.

References

1. Takai, D., and P. A. Jones. 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.
2. Bird, A. 2002. DNA methylation patterns and epigenetic memory. *Genes & Development* 16:6-21.
3. Schermelleh, L., A. Haemmer, F. Spada, N. R  sing, D. Meilinger, U. Rothbauer, M. C. Cardoso, and H. Leonhardt. 2007. Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Research* 35:4301-4312.
4. Bostick, M., J. K. Kim, P.-O. Est  ve, A. Clark, S. Pradhan, and S. E. Jacobsen. 2007. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science* 317:1760-1764.
5. Probst, A. V., E. Dunleavy, and G. Almouzni. 2009. Epigenetic inheritance during the cell cycle. *Nat Rev Mol Cell Biol* 10:192-206.
6. Chen, T., N. Tsujimoto, and E. Li. 2004. The PWWP Domain of Dnmt3a and Dnmt3b Is Required for Directing DNA Methylation to the Major Satellite Repeats at Pericentric Heterochromatin. *Mol. Cell. Biol.* 24:9048-9058.
7. Goll, M. G., and T. H. Bestor. 2005. EUKARYOTIC CYTOSINE METHYLTRANSFERASES. *Annual Review of Biochemistry* 74:481-514.
8. Jair, K.-W., K. E. Bachman, H. Suzuki, A. H. Ting, I. Rhee, R.-W. C. Yen, S. B. Baylin, and K. E. Schuebel. 2006. De novo CpG Island Methylation in Human Cancer Cells. *Cancer Res* 66:682-692.

9. Rhee, I., K. E. Bachman, B. H. Park, K.-W. Jair, R.-W. C. Yen, K. E. Schuebel, H. Cui, A. P. Feinberg, C. Lengauer, K. W. Kinzler, S. B. Baylin, and B. Vogelstein. 2002. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* 416:552-556.
10. Brown, K. D., and K. D. Robertson. 2007. DNMT1 knockout delivers a strong blow to genome stability and cell viability. *Nat Genet* 39:289-290.
11. Gun-Do Kim, J. N., Nicole Kelesoglu, Richard J. Roberts and Sriharsa Pradhan 2002. Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. *EMBO Journal* 21.
12. Di Croce, L., V. A. Raker, M. Corsaro, F. Fazi, M. Fanelli, M. Faretta, F. Fuks, F. L. Coco, T. Kouzarides, C. Nervi, S. Minucci, and P. G. Pelicci. 2002. Methyltransferase Recruitment and DNA Hypermethylation of Target Promoters by an Oncogenic Transcription Factor. *Science* 295:1079-1082.
13. Jharna Datta, K. G. S. M. S. S. T. S. T. J. 2003. Biochemical fractionation reveals association of DNA methyltransferase (Dnmt) 3b with Dnmt1 and that of Dnmt 3a with a histone H3 methyltransferase and Hdac1. *Journal of Cellular Biochemistry* 88:855-864.
14. Jones, P. A., and G. Liang. 2009. Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet* 10:805-811.
15. Jeong, S., G. Liang, S. Sharma, J. C. Lin, S. H. Choi, H. Han, C. B. Yoo, G. Egger, A. S. Yang, and P. A. Jones. 2009. Selective Anchoring of DNA Methyltransferases 3A and 3B to Nucleosomes Containing Methylated DNA. *Mol. Cell. Biol.* 29:5366-5376.

16. Liang, G., M. F. Chan, Y. Tomigahara, Y. C. Tsai, F. A. Gonzales, E. Li, P. W. Laird, and P. A. Jones. 2002. Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements. *Mol. Cell. Biol.* 22:480-491.
17. Okano, M., D. W. Bell, D. A. Haber, and E. Li. 1999. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* 99:247-257.
18. Yoder, J. A., C. P. Walsh, and T. H. Bestor. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics* 13:335-340.
19. Reik, W., W. Dean, and J. Walter. 2001. Epigenetic Reprogramming in Mammalian Development. *Science* 293:1089-1093.
20. Li, L., P. Zheng, and J. Dean. Maternal control of early mouse development. *Development* 137:859-870.
21. Li, E., T. H. Bestor, and R. Jaenisch. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69:915-926.
22. Lei, H., S. P. Oh, M. Okano, R. Juttermann, K. A. Goss, R. Jaenisch, and E. Li. 1996. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* 122:3195-3205.
23. Jones, P. A., and S. B. Baylin. 2002. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 3:415-428.
24. Ting, A. H., K. M. McGarvey, and S. B. Baylin. 2006. The cancer epigenome--components and functional correlates. *Genes Dev.* 20:3215-3231.
25. Schuebel, K., W. Chen, and S. B. Baylin. 2006. CIMP: origin for promoter hypermethylation in colorectal cancer? *Nat Genet* 38:738-740.

26. Weisenberger, D. J., K. D. Siegmund, M. Campan, J. Young, T. I. Long, M. A. Faasse, G. H. Kang, M. Widschwendter, D. Weener, D. Buchanan, H. Koh, L. Simms, M. Barker, B. Leggett, J. Levine, M. Kim, A. J. French, S. N. Thibodeau, J. Jass, R. Haile, and P. W. Laird. 2006. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* 38:787-793.
27. Walther, A., E. Johnstone, C. Swanton, R. Midgley, I. Tomlinson, and D. Kerr. 2009. Genetic prognostic and predictive markers in colorectal cancer. *Nat Rev Cancer* 9:489-499.
28. Hinoue, T., D. J. Weisenberger, C. P. E. Lange, H. Shen, H.-M. Byun, D. Van Den Berg, S. Malik, F. Pan, H. Noushmehr, C. M. van Dijk, R. A. E. M. Tollenaar, and P. W. Laird. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Research*.
29. Noushmehr, H., D. J. Weisenberger, K. Diefes, H. S. Phillips, K. Pujara, B. P. Berman, F. Pan, C. E. Pelloso, E. P. Sulman, K. P. Bhat, R. G. W. Verhaak, K. A. Hoadley, D. N. Hayes, C. M. Perou, H. K. Schmidt, L. Ding, R. K. Wilson, D. Van Den Berg, H. Shen, H. Bengtsson, P. Neuvial, L. M. Cope, J. Buckley, J. G. Herman, S. B. Baylin, P. W. Laird, and K. Aldape. Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma. *Cancer Cell* 17:510-522.
30. Kim, M., T. I. Long, K. Arakawa, R. Wang, M. C. Yu, and P. W. Laird. DNA Methylation as a Biomarker for Cardiovascular Disease Risk. *PLoS ONE* 5:e9692.
31. Houshdaran, S., S. Hawley, C. Palmer, M. Campan, M. N. Olsen, A. P. Ventura, B. S. Knudsen, C. W. Drescher, N. D. Urban, P. O. Brown, and P. W. Laird. DNA Methylation Profiles of Ovarian Epithelial Carcinoma Tumors and Cell Lines. *PLoS ONE* 5:e9359.

32. Wu, J., J. P. Issa, J. Herman, D. E. Bassett, B. D. Nelkin, and S. B. Baylin. 1993. Expression of an exogenous eukaryotic DNA methyltransferase gene induces transformation of NIH 3T3 cells. *Proceedings of the National Academy of Sciences* 90:8891-8895.
33. Bakin, A. V., and T. Curran. 1999. Role of DNA 5-Methylcytosine Transferase in Cell Transformation by fos. *Science* 283:387-390.
34. Oka, M., N. Rodic, J. Graddy, L.-J. Chang, and N. Terada. 2006. CpG Sites Preferentially Methylated by Dnmt3a in Vivo. *J. Biol. Chem.* 281:9901-9908.
35. Linhart, H. G., H. Lin, Y. Yamada, E. Moran, E. J. Steine, S. Gokhale, G. Lo, E. Cantu, M. Ehrich, T. He, A. Meissner, and R. Jaenisch. 2007. Dnmt3b promotes tumorigenesis in vivo by gene-specific de novo methylation and transcriptional silencing. *Genes & Development* 21:3110-3122.
36. Klose, R. J., and A. P. Bird. 2006. Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences* 31:89-97.
37. Bienvenu, T., and J. Chelly. 2006. Molecular genetics of Rett syndrome: when DNA methylation goes unrecognized. *Nat Rev Genet* 7:415-426.
38. Klose, R. J., S. A. Sarraf, L. Schmiedeberg, S. M. McDermott, I. Stancheva, and A. P. Bird. 2005. DNA Binding Selectivity of MeCP2 Due to a Requirement for A/T Sequences Adjacent to Methyl-CpG. *Molecular Cell* 19:667-678.
39. Clouaire, T., J. I. de las Heras, C. Merusi, and I. Stancheva. Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA. *Nucleic Acids Research* 38:4620-4634.

40. Nan, X., H.-H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, and A. Bird. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393:386-389.
41. Eden, S., T. Hashimshony, I. Keshet, H. Cedar, and A. W. Thorne. 1998. DNA methylation models histone acetylation. *Nature* 394:842-842.
42. Espada, J., E. Ballestar, M. F. Fraga, A. Villar-Garea, A. Juarranz, J. C. Stockert, K. D. Robertson, F. Fuks, and M. Esteller. 2004. Human DNA Methyltransferase 1 Is Required for Maintenance of the Histone H3 Modification Pattern. *J. Biol. Chem.* 279:37175-37184.
43. Kass, S. U., N. Landsberger, and A. P. Wolffe. 1997. DNA methylation directs a time-dependent repression of transcription initiation. *Current biology : CB* 7:157-165.
44. Martin, C., and Y. Zhang. 2005. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* 6:838-849.
45. Esteve, P.-O., H. G. Chin, A. Smallwood, G. R. Feehery, O. Gangisetty, A. R. Karpf, M. F. Carey, and S. Pradhan. 2006. Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* 20:3089-3103.
46. Vire, E., C. Brenner, R. Deplus, L. Blanchon, M. Fraga, C. Didelot, L. Morey, A. Van Eynde, D. Bernard, J.-M. Vanderwinden, M. Bollen, M. Esteller, L. Di Croce, Y. de Launoit, and F. Fuks. 2006. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439:871-874.
47. Schlesinger, Y., R. Straussman, I. Keshet, S. Farkash, M. Hecht, J. Zimmerman, E. Eden, Z. Yakhini, E. Ben-Shushan, B. E. Reubinooff, Y. Bergman, I. Simon, and H. Cedar. 2007. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* 39:232-236.

48. Rajasekhar, V. K., and M. Begemann. 2007. Concise Review: Roles of Polycomb Group Proteins in Development and Disease: A Stem Cell Perspective. *Stem Cells* 25:2498-2510.
49. McGarvey, K. M., E. Greene, J. A. Fahrner, T. Jenuwein, and S. B. Baylin. 2007. DNA Methylation and Complete Transcriptional Silencing of Cancer Genes Persist after Depletion of EZH2. *Cancer Res* 67:5097-5102.
50. Gal-Yam, E. N., G. Egger, L. Iniguez, H. Holster, S. m. Einarsson, X. Zhang, J. C. Lin, G. Liang, P. A. Jones, and A. Tanay. 2008. Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proceedings of the National Academy of Sciences* 105:12979-12984.
51. Kondo, Y., L. Shen, A. S. Cheng, S. Ahmed, Y. Boumber, C. Charo, T. Yamochi, T. Urano, K. Furukawa, B. Kwabi-Addo, D. L. Gold, Y. Sekido, T. H.-M. Huang, and J.-P. J. Issa. 2008. Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation. *Nat Genet* 40:741-750.
52. Issa, J.-P. 1999. Aging, DNA methylation and cancer. *Critical Reviews in Oncology/Hematology* 32:31-43.
53. Issa, J.-P. J., Y. L. Ottaviano, P. Celano, S. R. Hamilton, N. E. Davidson, and S. B. Baylin. 1994. Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat Genet* 7:536-540.
54. Talbert, P. B., and S. Henikoff. 2006. Spreading of silent chromatin: inaction at a distance. *Nat Rev Genet* 7:793-803.
55. Bushey, A. M., E. R. Dorman, and V. G. Corces. 2008. Chromatin Insulators: Regulatory Mechanisms and Epigenetic Inheritance. *Molecular Cell* 32:1-9.

56. Valenzuela, L., and R. T. Kamakaka. 2006. Chromatin Insulators. *Annual Review of Genetics* 40:107-138.
57. Bell, A. C., and G. Felsenfeld. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 405:482-485.
58. Hark, A. T., C. J. Schoenherr, D. J. Katz, R. S. Ingram, J. M. LeVorse, and S. M. Tilghman. 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* 405:486-489.
59. Gaszner, M., and G. Felsenfeld. 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* 7:703-713.
60. West, A. G., M. Gaszner, and G. Felsenfeld. 2002. Insulators: many functions, many mechanisms. *Genes & Development* 16:271-288.
61. Bumber, Y. A., Y. Kondo, X. Chen, L. Shen, Y. Guo, C. Tellez, M. R. H. Est cio, S. Ahmed, and J.-P. J. Issa. 2008. An Sp1/Sp3 Binding Polymorphism Confers Methylation Protection. *PLoS Genet* 4:e1000162.
62. Grewal, S. I. S., and D. Moazed. 2003. Heterochromatin and Epigenetic Control of Gene Expression. *Science* 301:798-802.
63. Toyota, M., and J.-P. J. Issa. 2005. Epigenetic Changes in Solid and Hematopoietic Tumors. *Seminars in Oncology* 32:521-530.
64. Turker, M. S. 2002. Gene silencing in mammalian cells and the spread of DNA methylation. *Oncogene* 21.
65. Turker, M. S. 1999. The establishment and maintenance of DNA methylation patterns in mouse somatic cells. *Seminars in Cancer Biology* 9:329-337.

66. Yates, P. A., R. W. Burman, P. Mummaneni, S. Krussel, and M. S. Turker. 1999. Tandem B1 Elements Located in a Mouse Methylation Center Provide a Target for de Novo DNA Methylation. *J. Biol. Chem.* 274:36357-36361.
67. Mummaneni, P., P. L. Bishop, and M. S. Turker. 1993. A cis-acting element accounts for a conserved methylation pattern upstream of the mouse adenine phosphoribosyltransferase gene. *J. Biol. Chem.* 268:552-558.
68. Yates, P. A., R. Burman, J. Simpson, O. N. Ponomoreva, M. J. Thayer, and M. S. Turker. 2003. Silencing of Mouse Aprt Is a Gradual Process in Differentiated Cells. *Mol. Cell. Biol.* 23:4461-4470.
69. Hasse, A., and W. A. Schulz. 1994. Enhancement of reporter gene de novo methylation by DNA fragments from the alpha-fetoprotein control region. *J. Biol. Chem.* 269:1821-1826.
70. Magewu, A. N., and P. A. Jones. 1994. Ubiquitous and tenacious methylation of the CpG site in codon 248 of the p53 gene may explain its frequent appearance as a mutational hot spot in human cancer. *Mol. Cell. Biol.* 14:4225-4232.
71. Graff, J. R., J. G. Herman, S. Myohanen, S. B. Baylin, and P. M. Vertino. 1997. Mapping Patterns of CpG Island Methylation in Normal and Neoplastic Cells Implicates Both Upstream and Downstream Regions in de Novo Methylation. *J. Biol. Chem.* 272:22322-22329.
72. Walsh, C. P., J. R. Chaillet, and T. H. Bestor. 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 20:116-117.

73. Handa, V., and A. Jeltsch. 2005. Profound Flanking Sequence Preference of Dnmt3a and Dnmt3b Mammalian DNA Methyltransferases Shape the Human Epigenome. *Journal of Molecular Biology* 348:1103-1112.
74. Weisenberger, D. J., M. Velicescu, J. C. Cheng, F. A. Gonzales, G. Liang, and P. A. Jones. 2004. Role of the DNA Methyltransferase Variant DNMT3b3 in DNA Methylation. *Molecular Cancer Research* 2:62-72.
75. Garrick, D., S. Fiering, D. I. K. Martin, and E. Whitelaw. 1998. Repeat-induced gene silencing in mammals. *Nat Genet* 18:56-59.
76. Peng, J. C., A. Valouev, T. Swigut, J. Zhang, Y. Zhao, A. Sidow, and J. Wysocka. 2009. Jarid2/Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells. *Cell* 139:1290-1302.
77. Li, G., R. Margueron, M. Ku, P. Chambon, B. E. Bernstein, and D. Reinberg. Jarid2 and PRC2, partners in regulating gene expression. *Genes & Development* 24:368-380.
78. Sparmann, A., and M. van Lohuizen. 2006. Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* 6:846-856.
79. Epsztejn-Litman, S., N. Feldman, M. Abu-Remaileh, Y. Shufaro, A. Gerson, J. Ueda, R. Deplus, F. Fuks, Y. Shinkai, H. Cedar, and Y. Bergman. 2008. De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nat Struct Mol Biol* 15:1176-1183.
80. Dong, K. B., I. A. Maksakova, F. Mohn, D. Leung, R. Appanah, S. Lee, H. W. Yang, L. L. Lam, D. L. Mager, D. Schubeler, M. Tachibana, Y. Shinkai, and M. C. Lorincz. 2008. DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J* 27:2691-2701.

81. Lehnertz, B., Y. Ueda, A. A. H. A. Derijck, U. Braunschweig, L. Perez-Burgos, S. Kubicek, T. Chen, E. Li, T. Jenuwein, and A. H. F. M. Peters. 2003. Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Current biology : CB* 13:1192-1200.
82. Li, H., T. Rauch, Z.-X. Chen, P. E. Szab³, A. D. Riggs, and G. P. Pfeifer. 2006. The Histone Methyltransferase SETDB1 and the DNA Methyltransferase DNMT3A Interact Directly and Localize to Promoters Silenced in Cancer Cells. *Journal of Biological Chemistry* 281:19489-19500.
83. Keshet, I., Y. Schlesinger, S. Farkash, E. Rand, M. Hecht, E. Segal, E. Pikarski, R. A. Young, A. Niveleau, H. Cedar, and I. Simon. 2006. Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat Genet* 38:149-153.
84. Hinshelwood, R. A., J. R. Melki, L. I. Huschtscha, C. Paul, J. Z. Song, C. Stirzaker, R. R. Reddel, and S. J. Clark. 2009. Aberrant de novo methylation of the p16INK4A CpG island is initiated post gene silencing in association with chromatin remodelling and mimics nucleosome positioning. *Human Molecular Genetics* 18:3098-3109.
85. Cedar, H., and Y. Bergman. 2009. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* 10:295-304.
86. Gibbons, R. J., T. L. McDowell, S. Raman, D. M. O'Rourke, D. Garrick, H. Ayyub, and D. R. Higgs. 2000. Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat Genet* 24:368-371.
87. Dennis, K., T. Fan, T. Geiman, Q. Yan, and K. Muegge. 2001. Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes & Development* 15:2940-2944.

88. Zhu, H., Geiman, TM, Xi, S, Jiang, Q, Schmidtmann, A, Chen, T, Li, E, Muegge, K. 2006. Lsh is involved in de novo methylation of DNA. *EMBO J* 25.
89. De La Fuente, R., C. Baumann, T. Fan, A. Schmidtmann, I. Dobrinski, and K. Muegge. 2006. Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. *Nat Cell Biol* 8:1448-1454.
90. Myant, K., and I. Stancheva. 2008. LSH Cooperates with DNA Methyltransferases To Repress Transcription. *Mol. Cell. Biol.* 28:215-226.
91. Fanti, L., and S. Pimpinelli. 2008. HP1: a functionally multifaceted protein. *Current Opinion in Genetics & Development* 18:169-174.
92. Smallwood, A., P.-O. Esteve, S. Pradhan, and M. Carey. 2007. Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev.* 21:1169-1178.
93. Bachman, K. E., M. R. Rountree, and S. B. Baylin. 2001. Dnmt3a and Dnmt3b Are Transcriptional Repressors That Exhibit Unique Localization Properties to Heterochromatin. *J. Biol. Chem.* 276:32282-32287.
94. Morris, K. V., S. W. L. Chan, S. E. Jacobsen, and D. J. Looney. 2004. Small Interfering RNA-Induced Transcriptional Gene Silencing in Human Cells. *Science* 305:1289-1292.
95. Weinberg, M. S., L. M. Villeneuve, A. L. I. Ehsani, M. Amarzguioui, L. Aagaard, Z.-X. Chen, A. D. Riggs, J. J. Rossi, and K. V. Morris. 2006. The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. *RNA* 12:256-262.
96. Aravin, A. A., and D. Bourc'his. 2008. Small RNA guides for de novo DNA methylation in mammalian germ cells. *Genes Dev.* 22:970-975.

97. Aravin, A. A., R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K. F. Toth, T. Bestor, and G. J. Hannon. 2008. A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. *Molecular Cell* 31:785-799.
98. Ting, A. H., K. E. Schuebel, J. G. Herman, and S. B. Baylin. 2005. Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. *Nat Genet* 37:906-910.
99. Ting, A. H., H. Suzuki, L. Cope, K. E. Schuebel, B. H. Lee, M. Toyota, K. Imai, Y. Shinomura, T. Tokino, and S. B. Baylin. 2008. A Requirement for DICER to Maintain Full Promoter CpG Island Hypermethylation in Human Cancer Cells. *Cancer Res* 68:2570-2575.
100. Wutz, A., and R. Jaenisch. 2000. A Shift from Reversible to Irreversible X Inactivation Is Triggered during ES Cell Differentiation. *Molecular Cell* 5:695-705.
101. Lister, R., R. C. O'Malley, J. Tonti-Filippini, B. D. Gregory, C. C. Berry, A. H. Millar, and J. R. Ecker. 2008. Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell* 133:523-536.
102. Meissner, A., T. S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, A. Gnirke, R. Jaenisch, and E. S. Lander. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454:766-770.
103. Serre, D., B. H. Lee, and A. H. Ting. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Research* 38:391-399.
104. Down, T. A., V. K. Rakyan, D. J. Turner, P. Flicek, H. Li, E. Kulesha, S. Graf, N. Johnson, J. Herrero, E. M. Tomazou, N. P. Thorne, L. Backdahl, M. Herberth, K. L. Howe,

- D. K. Jackson, M. M. Miretti, J. C. Marioni, E. Birney, T. J. P. Hubbard, R. Durbin, S. Tavaré, and S. Beck. 2008. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotech* 26:779-785.
105. Harris, R. A., T. Wang, C. Coarfa, R. P. Nagarajan, C. Hong, S. L. Downey, B. E. Johnson, S. D. Fouse, A. Delaney, Y. Zhao, A. Olshen, T. Ballinger, X. Zhou, K. J. Forsberg, J. Gu, L. Echipare, H. O'Geen, R. Lister, M. Pelizzola, Y. Xi, C. B. Epstein, B. E. Bernstein, R. D. Hawkins, B. Ren, W.-Y. Chung, H. Gu, C. Bock, A. Gnirke, M. Q. Zhang, D. Haussler, J. R. Ecker, W. Li, P. J. Farnham, R. A. Waterland, A. Meissner, M. A. Marra, M. Hirst, A. Milosavljevic, and J. F. Costello. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotech* 28:1097-1105.
106. Bock, C., E. M. Tomazou, A. B. Brinkman, F. Muller, F. Simmer, H. Gu, N. Jager, A. Gnirke, H. G. Stunnenberg, and A. Meissner. Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotech* 28:1106-1114.
107. Illingworth, R., A. Kerr, D. DeSousa, H. Järnagren, P. Ellis, J. Stalker, D. Jackson, C. Clee, R. Plumb, J. Rogers, S. Humphray, T. Cox, C. Langford, and A. Bird. 2008. A Novel CpG Island Set Identifies Tissue-Specific Methylation at Developmental Gene Loci. *PLoS Biol* 6:e22.
108. Shen, L., Y. Kondo, Y. Guo, J. Zhang, L. Zhang, S. Ahmed, J. Shu, X. Chen, R. A. Waterland, and J.-P. J. Issa. 2007. Genome-Wide Profiling of DNA Methylation Reveals a Class of Normally Methylated CpG Island Promoters. *PLoS Genet* 3:e181.
109. Bumberg, Y. A., Y. Kondo, X. Chen, L. Shen, V. Gharibyan, K. Konishi, E. Estey, H. Kantarjian, G. Garcia-Manero, and J.-P. J. Issa. 2007. RIL, a LIM Gene on 5q31, Is Silenced

by Methylation in Cancer and Sensitizes Cancer Cells to Apoptosis. *Cancer Res* 67:1997-2005.

110. Athanasiadou, R., D. de Sousa, K. Myant, C. Merusi, I. Stancheva, and A. Bird. Targeting of De Novo DNA Methylation Throughout the Oct-4 Gene Regulatory Region in Differentiating Embryonic Stem Cells. *PLoS ONE* 5:e9937.

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

112. Song, J. Z., C. Stirzaker, J. Harrison, J. R. Melki, and S. J. Clark. 2002. Hypermethylation trigger of the glutathione-S-transferase gene (GSTP1) in prostate cancer cells. *Oncogene* 21:1048-1061.

113. Estecio, M. R. H., V. Gharibyan, L. Shen, A. E. K. Ibrahim, K. Doshi, R. He, J. Jelinek, A. S. Yang, P. S. Yan, T. H. M. Huang, E. H. Tajara, and J.-P. J. Issa. 2007. LINE-1 Hypomethylation in Cancer Is Highly Variable and Inversely Correlated with Microsatellite Instability. *PLoS ONE* 2:e399.

114. Richards, K. L., B. Zhang, K. A. Baggerly, S. Colella, J. C. Lang, D. E. Schuller, and R. Krahe. 2009. Genome-Wide Hypomethylation in Head and Neck Cancer Is More Pronounced in HPV-Negative Tumors and Is Associated with Genomic Instability. *PLoS ONE* 4:e4941.

115. Williams, A., N. Harker, E. Ktistaki, H. Veiga-Fernandes, K. Roderick, M. Tolaini, T. Norton, K. Williams, and D. Kioussis. 2008. Position effect variegation and imprinting of transgenes in lymphocytes. *Nucleic Acids Research* 36:2320-2329.

116. Fradin, A., J. L. Manley, and C. L. Prives. 1982. Methylation of simian virus 40 Hpa II site affects late, but not early, viral gene expression. *Proceedings of the National Academy of Sciences* 79:5142-5146.
117. Ayyanathan, K., M. S. Lechner, P. Bell, G. G. Maul, D. C. Schultz, Y. Yamada, K. Tanaka, K. Torigoe, and F. J. Rauscher. 2003. Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. *Genes & Development* 17:1855-1869.
118. Fahrner, J. A., and S. B. Baylin. 2003. Heterochromatin: stable and unstable invasions at home and abroad. *Genes & Development* 17:1805-1812.
119. Selker, E. U. 1999. Gene Silencing: Repeats that Count. *Cell* 97:157-160.
120. Martienssen, R. A., and V. Colot. 2001. DNA Methylation and Epigenetic Inheritance in Plants and Filamentous Fungi. *Science* 293:1070-1074.
121. Kiess, M., B. Scharm, A. Aguzzi, A. Hajnal, R. Klemenz, I. Schwarte-Waldhoff, and R. Schafer. 1995. Expression of ril, a novel LIM domain gene, is down-regulated in HRAS-transformed cells and restored in phenotypic revertants. *Oncogene* 10:61-68.
122. Bumber, Y. A., Y. Kondo, X. Chen, L. Shen, Y. Guo, C. Tellez, M. R. H. Estecio, S. Ahmed, and J.-P. J. Issa. 2008. An Sp1/Sp3 Binding Polymorphism Confers Methylation Protection. *PLoS Genetics* 4:e1000162.
123. Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, and A. H. Schulman. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8:973-982.

124. Jordan, I. K., I. B. Rogozin, G. V. Glazko, and E. V. Koonin. 2003. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends in Genetics* 19:68-72.
125. Weisenberger, D. J., M. Velicescu, J. C. Cheng, F. A. Gonzales, G. Liang, and P. A. Jones. 2004. Role of the DNA Methyltransferase Variant DNMT3b3 in DNA Methylation. NIH grants CA 82422 and 1RO1 CA 83867 (P. A. J.), NIH Training Grant T32 DE07211-11 (M. V.), and NIH Training Grant in Basic Research in Oncology T32 CA09659 (D. W.). Note: D. J. Weisenberger and M. Velicescu contributed equally to this work. *Molecular Cancer Research* 2:62-72.
126. Feschotte, C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 9:397-405.
127. Garrick, D., S. Fiering, D. I. K. Martin, and E. Whitelaw. 1998. Repeat-induced gene silencing in mammals. *Nature Genet.* 18:56-59.
128. Hutter, B., V. Helms, and M. Paulsen. 2006. Tandem repeats in the CpG islands of imprinted genes. *Genomics* 88:323-332.
129. Suzuki, M. M., and A. Bird. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 9:465-476.
130. Rhee, I., K.-W. Jair, R.-W. C. Yen, C. Lengauer, J. G. Herman, K. W. Kinzler, B. Vogelstein, S. B. Baylin, and K. E. Schuebel. 2000. CpG methylation is maintained in human cancer cells lacking DNMT1. *Nature* 404:1003-1007.
131. Schuebel, K. E., W. Chen, L. Cope, S. C. Glockner, H. Suzuki, J.-M. Yi, T. A. Chan, L. V. Neste, W. V. Crielinge, S. v. d. Bosch, M. van Engeland, A. H. Ting, K. Jair, W. Yu,

- M. Toyota, K. Imai, N. Ahuja, J. G. Herman, and S. B. Baylin. 2007. Comparing the DNA Hypermethylome with Gene Mutations in Human Colorectal Cancer. *PLoS Genet* 3:e157.
132. Spada, F., A. Haemmer, D. Kuch, U. Rothbauer, L. Schermelleh, E. Kremmer, T. Carell, G. Langst, and H. Leonhardt. 2007. DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *J. Cell Biol.* 176:565-571.
133. Chen, T., S. Hevi, F. Gay, N. Tsujimoto, T. He, B. Zhang, Y. Ueda, and E. Li. 2007. Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nat Genet* 39:391-396.
134. Illingworth, R. S., U. Gruenewald-Schneider, S. Webb, A. R. W. Kerr, K. D. James, D. J. Turner, C. Smith, D. J. Harrison, R. Andrews, and A. P. Bird. Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. *PLoS Genet* 6:e1001134.
135. Landolin, J. M., D. S. Johnson, N. D. Trinklein, S. F. Aldred, C. Medina, H. Shulha, Z. Weng, and R. M. Myers. Sequence features that drive human promoter function and tissue specificity. *Genome Research* 20:890-898.
136. Takeshima, H., S. Yamashita, T. Shimazu, T. Niwa, and T. Ushijima. 2009. The presence of RNA polymerase II, active or stalled, predicts epigenetic fate of promoter CpG islands. *Genome Research* 19:1974-1982.
137. Wu, H., A. C. D'Alessio, S. Ito, K. Xia, Z. Wang, K. Cui, K. Zhao, Y. Eve Sun, and Y. Zhang. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* 473:389-393.

138. Tahiliani, M., K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, and A. Rao. 2009. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* 324:930-935.
139. Ooi, S. K. T., C. Qiu, E. Bernstein, K. Li, D. Jia, Z. Yang, H. Erdjument-Bromage, P. Tempst, S.-P. Lin, C. D. Allis, X. Cheng, and T. H. Bestor. 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448:714-717.
140. Maunakea, A. K., R. P. Nagarajan, M. Bilenky, T. J. Ballinger, C. D'Souza, S. D. Fouse, B. E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V. M. Heine, D. H. Rowitch, X. Xing, C. Fiore, M. Schillebeeckx, S. J. M. Jones, D. Haussler, M. A. Marra, M. Hirst, T. Wang, and J. F. Costello. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466:253-257.
141. Mercer, T. R., M. E. Dinger, and J. S. Mattick. 2009. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 10:155-159.
142. Kornblihtt, A. R. 2006. Chromatin, transcript elongation and alternative splicing. *Nat Struct Mol Biol* 13:5-7.
143. Irizarry, R. A., C. Ladd-Acosta, B. Wen, Z. Wu, C. Montano, P. Onyango, H. Cui, K. Gabo, M. Rongione, M. Webster, H. Ji, J. B. Potash, S. Sabunciyan, and A. P. Feinberg. 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 41:178-186.
144. Zhu, J., F. He, S. Hu, and J. Yu. 2008. On the nature of human housekeeping genes. *Trends in genetics* : TIG 24:481-484.

145. Juven-Gershon, T., J.-Y. Hsu, J. W. M. Theisen, and J. T. Kadonaga. 2008. The RNA polymerase II core promoter -- the gateway to transcription. *Current Opinion in Cell Biology* 20:253-259.
146. Wu, H., A. C. D'Alessio, S. Ito, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, and Y. Zhang. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes & Development* 25:679-684.

Vita

Yan Zhang was born in Dalian, Liaoning Province, China on September 16, 1978, the daughter of Xuming Zhang and Lianchun Qu and the sister of Dong Zhang. After completing her education in the Affiliated High School of Liaoning Normal University, Dalian, China in 1997, she entered China Medical University in Shenyang, China. She received her Bachelor's degree of Clinical Medicine in July, 2002. She then continued to study in Fudan University, Shanghai and in 2005, earned her degree of Master of Medical Sciences with a major in pathogenic biology in State Key Laboratory of Oncogenes and Related Genes at Shanghai Cancer Institute. After a short-term study in McGill University, Canada, she joined The University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences in August of 2006. From June 2007, she started her project in the laboratory of Dr. Jean-Pierre Issa in the Department of Leukemia at The University of Texas M.D. Anderson Cancer Center to study cancer epigenetics.

Permanent address:

Room 402, Suzhou Street

Dalian, Liaoning, China 116023