

12-2012

## Tet1: A Unique Dna Demethylase For Maintenance Of Dna Methylation Pattern

CHUNLEI JIN

Follow this and additional works at: [https://digitalcommons.library.tmc.edu/utgsbs\\_dissertations](https://digitalcommons.library.tmc.edu/utgsbs_dissertations)



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

---

### Recommended Citation

JIN, CHUNLEI, "Tet1: A Unique Dna Demethylase For Maintenance Of Dna Methylation Pattern" (2012). *Dissertations and Theses (Open Access)*. 324.  
[https://digitalcommons.library.tmc.edu/utgsbs\\_dissertations/324](https://digitalcommons.library.tmc.edu/utgsbs_dissertations/324)

This Dissertation (PhD) is brought to you for free and open access by the MD Anderson UTHealth Houston Graduate School at DigitalCommons@TMC. It has been accepted for inclusion in Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact [digcommons@library.tmc.edu](mailto:digcommons@library.tmc.edu).

**TET1: A UNIQUE DNA DEMETHYLASE FOR MAINTENANCE  
OF DNA METHYLATION PATTERN**

By

Chunlei Jin, M.B.B.S, M.S.

APPROVED:

---

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

---

Michelle Barton, Ph.D.

---

Ralf Krahe, Ph.D.

---

Min Gyu Lee, Ph.D.

---

Bin Wang, Ph.D.

---

APPROVED:

---

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

**TET1: A UNIQUE DNA DEMETHYLASE FOR MAINTENANCE  
OF DNA METHYLATION PATTERN**

A

DISSERTATION

Presented to the Faculty of

The University of Texas

Health Science Center at Houston

and

The University of Texas

MD Anderson Cancer Center

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Chunlei Jin, M.B.B.S., M.S.

Houston, Texas

December, 2012

## DEDICATION

To my wife, Shuangshuang Wu, and my parents, Bingyun Jin and Caimei Zhang, for all your love and support throughout my life.

## ACKNOWLEDGEMENTS

I am foremost grateful for my mentor, Dr. Jean-Pierre Issa. His support, guidance and inspiration made this dissertation possible. More importantly, the independence, innovation and balancing between life and work that I have learned from him will continue to benefit my future career and all the life.

I also deeply appreciate the help and instruction from Dr. Michelle Barton. In addition to the diligent participation in my advisory, examining, and supervisory committees, she also kindly served as my co-advisor, giving me the opportunity to study in her laboratory and instructing me during my last one and half years in the school.

In addition, I am also grateful for my current and previous committee members, Dr. Ralf Krahe, Dr. Min Gyu Lee, Dr. Bin Wang, Dr. Bingliang Fang, Dr. Zhen Fanand Dr. Peng Huang, for guiding me through my research project. I hope to have learned some of their collective wisdom, and to be as supportive and nurturing to my future students as they have been with me.

I also thank all members in Dr. Issa' and Dr. Barton's laboratories, for their kind cooperation and creating such great research environment.

Last but not least, I deeply thank my parents Bingyun Jin and Caimei Zhang for their unconditional love and support for three decades.

# TET1: A UNIQUE DNA DEMETHYLASE FOR MAINTENANCE OF DNA METHYLATION PATTERN

Publication No. \_\_\_\_\_

Chunlei Jin, M.B.B.S., M.S.

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

DNA methylation at the C5 position of cytosine (5-methylcytosine, 5mC) is a crucial epigenetic modification of the genome and has been implicated in numerous cellular processes in mammals, including embryonic development, transcription, X chromosome inactivation, genomic imprinting and chromatin structure. Like histone modifications, DNA methylation is also dynamic and reversible. However, in contrast to well defined DNA methyltransferases, the enzymes responsible for erasing DNA methylation still remain to be studied. The ten-eleven translocation family proteins (TET1/2/3) were recently identified as Fe(II)/2-oxoglutarate (2OG)-dependent 5mC dioxygenases, which consecutively convert 5mC into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine and 5-carboxylcytosine both in vitro and in mammalian cells. Based on their potent oxidative activities on 5mC, TET proteins have shown great potential as the long-sought DNA demethylases that induce DNA demethylation through multiple potential mechanisms. Here, we show that overexpression of TET1 catalytic domain alone (TET1-CD) but not full length TET1 (TET1-FL) induces global DNA demethylation in HEK293T cells. Genome-wide mapping of 5hmC further reveals a unique regulation pattern of 5mC by TET1-FL, where its 5hmC production is relatively inhibited as local basal DNA methylation level increases. By contrast, overexpression of TET1-CD exhibited a strong positive correlation between 5hmC production and

basal DNA methylation level. In support of it, we interestingly found that through CXXC domain TET1 specifically binds hypomethylated but not hypermethylated CpG-rich regions. Moreover, overexpression of TET1-FL specifically decreased DNA methylation levels to certain extent only in hypomethylated CpG sites (methylation level  $\leq 10\%$ ). To further investigate the effect of TET1, we also developed a lentiviral shRNA mediated TET1 knockdown in HEK293T cells, which originally have a comparable TET1 expression level as human embryonic stem cells. Knockdown of TET1 significantly induced an increase of DNA methylation in the pre-methylated edges, but not unmethylated edges and center regions of CpG islands (CGIs), indicating that TET1 can efficiently maintain the DNA hypomethylated state of CGIs by inhibiting the spreading of de novo DNA methylation from the pre-methylated edges of CGIs. Finally, with the use of the inducible TET1-CD overexpression system in HEK293T cells, we found that knockdown or inhibition of APEX1, the key player of DNA base excision repair pathway (BER), did not impair DNA demethylation induced by TET1-CD overexpression, suggesting that TET-mediated DNA demethylation is independent on BER. Moreover, our results also suggest that the replication-dependent passive pathway is not the primary mechanism for TET-mediated DNA demethylation. In conclusion, our results demonstrated that TET1 is a unique DNA demethylase which cannot significantly change DNA methylation levels, but rather specifically maintains the DNA hypomethylation state in CpG-rich regions by removing aberrant de novo DNA methylation. Moreover, neither BER nor DNA replication is required for TET-mediated DNA demethylation. Future studies focusing on potential 5-carboxylcytosine decarboxylases are necessary to elucidate the underlying mechanism.

## TABLE OF CONTENTS

<b>APPROVAL SHEET</b> .....	<b>i</b>
<b>TITLE PAGE</b> .....	<b>ii</b>
<b>DEDICATION</b> .....	<b>iii</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>iv</b>
<b>ABSTRACT</b> .....	<b>v</b>
<b>TABLE OF CONTENTS</b> .....	<b>vii</b>
<b>LIST OF ILLUSTRATIONS</b> .....	<b>xi</b>
<b>LIST OF TABLES</b> .....	<b>xiv</b>
<b>CHAPTER 1 INTRODUCTION</b> .....	<b>1</b>
<b>1.1 DNA methylation in mammalian cells</b> .....	<b>1</b>
1.1.1 Distribution pattern of DNA methylation .....	1
1.1.2 Biological functions of DNA methylation .....	3
1.1.3 Maintenance of DNA methylation during cell division .....	5
1.1.4 Mechanisms controlling de novo DNA methylation.....	6
1.1.5 Immunity of CGIs to DNA methylation .....	10
1.1.6 Aberrant DNA methylation in cancer .....	12
<b>1.2 Active DNA demethylation and potential mechanisms in mammals</b> .....	<b>13</b>
1.2.1 Evidences for active DNA demethylation.....	14
1.2.2 Potential mechanisms of active DNA demethylation and DNA demethylases.....	17
<b>1.3 TET1: a novel 5-methylcytosine dioxygenase</b> .....	<b>22</b>
1.3.1 TET-mediated 5mC oxidation leads to DNA demethylation.....	22

1.3.2 Studies on the biological functions of TET1 .....	27
<b>1.4 Hypothesis, specific aims and rationales .....</b>	<b>29</b>
<b>CHAPTER 2 MATERIAL AND METHODS.....</b>	<b>32</b>
2.1 Cell culture .....	32
2.2 Gene cloning and plasmid construction for TET1 overexpression .....	32
2.3 Site-directed mutagenesis .....	33
2.4 Plasmid transfection .....	33
2.5 Fluorescence-activated cell sorting (FACS).....	36
2.6 DNA extraction and bisulfite-pyrosequencing/sequencing .....	36
2.7 Protein extraction and western blot assay.....	43
2.8 DNA dot blot against 5hmC .....	43
2.9 RNA extraction, cDNA synthesis, and quantitative PCR.....	44
2.10 <i>HpaII</i> -digestion DNA methylation assay .....	44
2.11 Digital restriction enzyme analysis of methylation (DREAM) .....	47
2.12 Hydroxymethylated DNA immunoprecipitation-sequencing (hMeDIP-Seq) .....	49
2.13 ChIP-quantitative PCR (ChIP-qPCR) .....	50
2.14 Lentiviral shRNA-mediated TET1 knockdown.....	51
2.15 siRNA knockdown .....	51
2.16 Cell growth curve .....	54
2.17 Tetracycline-induced TET1-CD overexpression.....	54
2.18 BrdU-containing DNA immunoprecipitation .....	54
2.19 Data processing and statistics .....	55

<b>CHAPTER 3 RESULTS .....</b>	<b>56</b>
<b>3.1 Different regulation of DNA methylation by TET1-FL and TET1-CD overexpression ..</b>	<b>56</b>
3.1.1 TET1 ORF Cloning and overexpression plasmids construction .....	56
3.1.2 Overexpression of TET1-FL and TET1-CD in HEK293T cells .....	58
3.1.3 Overexpression of TET1-CD but not TET1-FL induces DNA demethylation in some genomic loci .....	61
3.1.4 TET1-CD overexpression induces global DNA demethylation without distribution bias ....	63
3.1.5 TET1-FL overexpression fails to induce significant DNA demethylation in all categorized genomic regions.....	67
3.1.6 Differential regulation of 5hmC distribution pattern by TET1-FL and TET1-CD overexpression .....	69
3.1.7 TET1-FL shows a unique regulation pattern of 5mC.....	77
3.1.8 TET1-FL specifically binds hypomethylated CGIs through its CXXC domain .....	80
3.1.9 Proposed model.....	82
<b>3.2 TET1 maintains the DNA hypomethylated state of CGIs in HEK293T cells .....</b>	<b>84</b>
3.2.1 TET1-FL overexpression specifically decreases DNA methylation in hypomethylated CGIs .....	85
3.2.2 TET1 knockdown leads to an increase of DNA methylation in hypomethylated CGIs.....	87
3.2.2.1 Screening of TET1 expression in various human cell lines .....	87
3.2.2.2 Lentiviral shRNA-mediated TET1 knockdown in HEK293T cells .....	89
3.2.2.3 TET1 knockdown inhibits cell growth in HEK293T cells.....	92
3.2.2.4 Increase of DNA methylation in the pre-methylated edges of hypomethylated	

CGIs after TET1 knockdown .....	94
<b>3.3 Mechanistic study of TET-mediated oxidative DNA demethylation .....</b>	<b>99</b>
3.3.1 Tetracycline-induced TET1-CD overexpression system in HEK293T cells.....	99
3.3.2 Long term TET1-CD overexpression and possible dedifferentiation in HEK293T cells....	103
3.3.3 TET-mediated DNA demethylation is independent of BER.....	106
3.3.4 TET-mediated DNA demethylation appears to be independent on DNA replication.....	109
<b>CHAPTER 4 DISCUSSION, SUMMARY, AND FUTURE DIRECTIONS.....</b>	<b>114</b>
<b>4.1 Discussion .....</b>	<b>114</b>
<b>4.2 Summary .....</b>	<b>121</b>
<b>4.3 Future directions .....</b>	<b>122</b>
<b>BIBLIOGRAPHY.....</b>	<b>124</b>
<b>VITA .....</b>	<b>144</b>

## LIST OF ILLUSTRATIONS

Figure 1. Dynamics of DNA methylation during embryogenesis in mouse.....	15
Figure 2. Potential mechanisms for active DNA demethylation in mammals.....	18
Figure 3. Domain structure of human TET family proteins .....	23
Figure 4. TET-mediated 5-methylcytosine oxidation reactions and potential mechanisms for DNA demethylation.....	26
Figure 5. The principle of digital restriction enzyme analysis of methylation (DREAM).....	48
Figure 6. The flow chart for hydroxymethylated DNA immunoprecipitation-sequencing (hMeDIP-Seq).....	52
Figure 7. Construction of expression plasmids encoding TET1-FL or TET1-CD .....	57
Figure 8. Both overexpression of TET1-CD and TET1-FL produce 5hmC in genomic DNA .....	60
Figure 9. 5hmC behaviors in a same pattern as 5mC in bisulfite-pyrosequencing .....	62
Figure 10. DNA demethylation of some genomic loci induced by TET1-CD but not TET1-FL overexpression .....	64
Figure 11. Global DNA demethylation induced by TET1-CD overexpression in HEK293T cells.....	66
Figure 12. TET1-CD overexpression induces DNA demethylation evenly in all categorized genomic regions.....	68
Figure 13. TET1-FL overexpression fails to induce global DNA demethylation in HEK293T cells.....	70
Figure 14. No significant DNA demethylation induced by TET1-FL overexpression in all categorized genomic regions.....	71
Figure 15. Validation of the specificity of 5hmC antibody in hMeDIP experiment.....	72

Figure 16. Overexpression of TET1-FL and TET1-CD generate different 5hmC distribution pattern.....	76
Figure 17. Compromised 5hmC production by TET1-FL as basal DNA methylation level increases .....	79
Figure 18. The CXXC domain specifically targets TET1-FL towards hypomethylated CGIs.....	81
Figure 19. Model for the self-inhibited catalytic function of TET1-FL for DNA demethylation ..	83
Figure 20. Decreased DNA methylation in high-tag-number CGI sites by TET1-FL overexpression in HEK293T cells.....	86
Figure 21. TET1 expression level varies among different human cell lines .....	88
Figure 22. Construction of lentiviral shRNA-mediated TET1 knockdown in HEK293T cells .....	90
Figure 23. Inhibited cell proliferation by TET1 knockdown in HEK293T cells.....	93
Figure 24. ChIP-qPCR analysis of endogenous TET1 target genes in HEK293T cells.....	95
Figure 25. Bisulfite-sequencing analysis of CGI promoters of BCL2L11, PASC1, PSEN2 and TTC9 .....	96
Figure 26. TET1 knockdown leads to spreading of de novo methylation at the pre-methylated edges of CGIs.....	98
Figure 27. Construction of tetracycline-induced TET1-CD overexpression system in HEK293T cells.....	101
Figure 28. Characterization of single cell clones with inducible overexpression of TET1-CD and DNA demethylation in specific genomic loci .....	102
Figure 29. Long-term overexpression of TET1-CD and potential dedifferentiation in HEK293T cells.....	104

Figure 30. DNA demethylation induced by TET1-CD overexpression is independent of APEX1 .....	107
Figure 31. Mimosine treatment inhibits doxycycline-induced TET1-CD overexpression.....	111
Figure 32. TET1-CD overexpression-induced DNA demethylation may be only through active pathway.....	112
Figure 33. Dynamics of inducible TET1-CD expression, genomic 5hmC and DNA methylation levels in doxycycline-treated HEK293T D1 clone cells .....	113

## LIST OF TABLES

Table 1. PCR primers and sequencing primers for TET1 plasmid construction .....	34
Table 2. PCR and sequencing primers used in bisulfite-pyrosequencing assay .....	38
Table 3. PCR primers used in bisulfite-sequencing assay. ....	42
Table 4. PCR primers for RT-qPCR of gene expression analysis.....	45
Table 5. PCR primers for <i>HpaII</i> sensitivity assay and ChIP-qPCR .....	46
Table 6. ShRNA and siRNA used for knockdown experiment .....	53
Table 7. Summary of tags and CpG sites detected in DREAM.....	65
Table 8. Summary of tags and peaks detected in hMeDIP-seq .....	74

# CHAPTER 1 INTRODUCTION

## 1.1 DNA methylation in mammalian cells

### 1.1.1 Distribution pattern of DNA methylation

In mammalian genomic DNA, a methyl group from S-adenosyl methionine can be catalytically added to the C5 position of cytosine by DNA methyltransferases, forming the so-called fifth base 5-methylcytosine (5mC). Such methylation of cytosine predominantly occurs at CpG dinucleotides with an exception in mouse and human embryonic stem cells (ESCs) where about 25% of 5mC locate in non-CpG dinucleotides (Lister et al., 2009; Ramsahoye et al., 2000). Interestingly, since 5mC can be spontaneously deaminated to thymine, CpG dinucleotides are increasingly mutated during the long course of evolution, demonstrated by the low percents of CpG dinucleotides in the human genome (only 21% of the expected frequency) (Lander et al., 2001).

DNA methylation is a pervasive modification of the genomic DNA in mammals. Analysis of the total base composition of DNA in human tissues and cell lines reveals that 5mC accounts for up to 1% of total bases and 60-90% of all CpG dinucleotides are methylated (Ehrlich et al., 1982; Tucker, 2001). The distribution pattern of DNA methylation in genomic DNA is initially established by the de novo DNA methyltransferases DNMT3A and DNMT3B during embryogenesis and gametogenesis (Kaneda et al., 2004; Okano et al., 1999), then maintained by maintenance DNA methyltransferases DNMT1 during cell division (Hermann et al., 2004). DNMT3L is enzymatically inactive but acts as a general stimulatory factor for de novo methylation by DNMT3A and DNMT3B (Hata et al., 2002; Suetake et al., 2004).

DNA methylation shows an uneven distribution pattern in mammalian genome. Methylated CpGs are predominately distributed within the repetitive sequences which mainly constitute transposone elements and heterochromatin, while unmethylated ones are often clustered in specific regions called CpG islands (CGIs), which accounts for only 1-2% of the total genome (Bird et al., 1985; Bird, 1986; Weber et al., 2005). CGIs are short interspersed DNA sequences that deviate significantly from the average genomic pattern by being GC-rich, CpG-rich, and predominantly unmethylated (Deaton and Bird, 2011). Additionally, compared with the flanking regions of genes, the gene body regions are often methylated at a relatively higher level (Weber et al., 2005). Thus, this DNA methylation pattern suggests that hypermethylation may be the default DNA methylation state in mammals.

On the other hand, the mammalian DNA methylation pattern is also dynamic. The paternal genome of zygotes undergoes a global demethylation immediately after fertilization (Mayer et al., 2000; Oswald et al., 2000), while the maternal genome is gradually demethylated during cleavage divisions in a replication-dependent manner (Howlett and Reik, 1991). After implantation, de novo methylation occurs in the diploid genome to establish DNA methylation pattern again (Okano et al., 1999). Moreover, primordial germ cells are subjected to another round of global DNA demethylation during late embryonic development, but differentiated somatic cells generally maintain their methylation patterns (Hajkova et al., 2002; Lee et al., 2002). Additionally, de novo methylation also occurs in somatic cells. For example, a fraction of CGIs become abnormally methylated in certain tissues during aging or in cancer cells (Baylin and Herman, 2000; Issa, 2000). Collectively, the dynamic pattern of DNA methylation suggests an important role of DNA methylation for development and diseases in mammals.

Taken together, the DNA methylation pattern in mammalian genome varies over space and time. More importantly, that distribution pattern lays the foundation for its biological functions and also suggests a complex regulatory mechanism in controlling de novo methylation and demethylation processes in cells.

### **1.1.2 Biological functions of DNA methylation**

Over the past decades, DNA methylation has been identified as a crucial epigenetic modification which plays an essential role in compacting chromatin structure and silencing gene expression (Bird, 2002). Different from histone modifications, DNA methylation is relatively stable and thus provides a more enduring epigenetic regulatory effect in mammalian cells. With its effects in regulation of chromatin structure and gene expression, DNA methylation so far has been reported to be involved in many important biological processes, including X chromosome inactivation, genomic imprinting and transposable element silencing.

Both X chromosome inactivation and genomic imprinting are epigenetic processes by which one allele is inactivated to realize mono-allelic expression. X-inactivation causes XX female cells to have same X chromosome gene products as XY males and imprinting makes certain genes be expressed in a parent-of-origin-specific manner. The DNA methylation found in inactive X chromosome and imprinted alleles is clearly required for their stable inactivation, as these silenced genes could be effectively relieved after demethylating treatment (El Kharroubi et al., 2001; Li et al., 1993; Mohandas et al., 1981; Venolia et al., 1982). Moreover, deficiency of DNA methylation machinery results in reduced methylation in inactive X chromosome and imprinted genes in cells and also leads to reactivation of those originally inactivated genes to varying extent

(Li et al., 1993; Miniou et al., 1994; Sado et al., 2000).

Transposable elements are DNA sequences that can change their relative position (self-transpose) and make up much of intergenic regions of genomic DNA. Silencing of those transposable elements is critical for mammalian cells to maintain the integrity of genome and also normal gene transcription program (Bird, 1995; Hedges and Deininger, 2007). The evidences that transposable elements are generally heavily methylated and both demethylating treatment and deficiency of DNMT1 lead to derepression of transposable elements solidly support the requirement of DNA methylation for transposable element silencing (Liu et al., 1994; Walsh et al., 1998; Woodcock et al., 1997).

How does DNA methylation work to achieve its repressive regulation of gene transcription? To date four different mechanisms have been supposed: (1) DNA methylation directly inhibits the DNA binding of transcriptional factors which specifically target CG-containing sequences (Comb and Goodman, 1990; Prendergast et al., 1991); (2) DNA methylation directly blocks the binding of H3K4 (histone H3 lysine 4) methyltransferase to DNA and thus interfere with the formation of active chromatin structure (Birke et al., 2002; Lee and Skalnik, 2005; Okitsu and Hsieh, 2007); (3) DNA methylation increases nucleosome occupancy at gene promoters and then repress the binding of transcriptional factors and RNA polymerase II (Davey et al., 1997; Ozsolak et al., 2007; Patel et al., 1997); and (4) the recruited methyl-CpG-binding proteins (MBPs) attract transcriptional repressor and also histone deacetylases which further contribute to the formation of inactive chromatin structure (Harikrishnan et al., 2005; Jones et al., 1998). It is noteworthy that although DNA methylation appears to serve as a lock for permanent gene silencing, DNA demethylation alone is insufficient to reactivate methylation-silenced gene unless chromatin

remodeling occurs simultaneously, suggesting that DNA methylation cooperates with other factors to realize gene silencing (Si et al., 2010).

### **1.1.3 Maintenance of DNA methylation during cell division**

As an epigenetic modification, DNA methylation pattern is faithfully maintained between cell generations through a DNMT1-mediated semiconservative mechanism. In this long-established model, DNMT1 is recruited to DNA replication foci by interacting with proliferating cell nuclear antigen (PCNA) and Np95/UHRF1, and then methylates newly synthesized CpGs opposite to methylated ones in the parent DNA strand (Bostick et al., 2007; Chuang et al., 1997). It is notable that the specific binding of Np95/UHRF1 to methylated CpGs through its SET and RING finger-associated (SRA) domain ensures the precise loading of DNMT1 to replicating methylated DNA regions (Avvakumov et al., 2008; Bostick et al., 2007; Sharif et al., 2007). Consistently, Np95-deficient mouse embryonic stem cells (mESCs) and embryos show loss of global and local DNA methylation and reactivation of retrotransposons and imprinted genes (Sharif et al., 2007).

On the other hand, however, accumulating evidences have also indicated that the classical DNMT1-dependent mechanism may be not the only one for maintenance of DNA methylation in mammals. In human colorectal cancer cells where DNMT1 gene is disrupted through homologous recombination, CGI methylation can be stably maintained during cell division with only a 20% decrease in global DNA methylation level (Rhee et al., 2000). Moreover, mESCs lacking Dnmt3A and Dnmt3B also show a gradual loss of DNA methylation in various repetitive elements and single-copy genes during cell division (Chen et al., 2003; Liang et al., 2002). Furthermore, through the interaction with specific histone modification enzymes (e.g. G9A,

EZH2) which are associated with inactive chromatin structure, DNMT3 but not DNMT1 can be efficiently recruited to heavily methylated CGIs and repetitive sequences (Jeong et al., 2009; Schlesinger et al., 2007). Therefore, those evidences strongly suggest that methylation activity by de novo enzymes DNMT3A and DNMT3B is also required for DNA methylation maintenance, particularly in densely methylated genomic regions. Actually, given the rapid generation of hemimethylated sites in methylated CGIs or repetitive elements during DNA replication, it is reasonable that a more efficient maintenance process by DNMT3 and/or DNMT1 but not DNMT1 alone is necessary to fully methylate those densely hemimethylated CpGs (Jones and Liang, 2009).

Taken together, the Np95/UHRF1-mediated localization of DNMT1 to DNA replication fork, the recruitment of DNMT3A and DNMT3B to densely methylated regions and the cooperation among those DNMTs are all required for the faithful maintenance of DNA methylation during cell division.

#### **1.1.4 Mechanisms controlling de novo DNA methylation**

As mentioned in previous part, the biological functions of DNA methylation are closely associated with its distribution patterns, which are initially set up by DNMT3A- and DNMT3B-catalyzed de novo DNA methylation. To clearly understand the complex pattern of DNA methylation in mammalian genome, some mechanisms of how de novo methylation is targeted and regulated have been revealed.

A best studied mechanism controlling de novo DNA methylation is DNMT3 recognizing specific histone modifications and histone-modifying enzymes, most of which are interestingly

associated with inactive heterochromatic state. For example, histone H3 tail unmethylated at lysine 4 (H3K4me0) can be specifically recognized by the PHD-like domain of DNMT3L (Ooi et al., 2007), which further recruits DNMT3A to induce de novo DNA methylation in imprinted genes in germ cells (Chen et al., 2005; Hata et al., 2002; Suetake et al., 2004). Alternatively, DNMT3A may directly bind H3K4me0 through its ADD domain, followed by de novo methylation of associated DNA (Otani et al., 2009; Zhang et al., 2010b). In agreement with it, mouse oocytes deficient in KDM1B, a histone H3K4 demethylase, show not only a substantial increase in H3K4 methylation but also loss of DNA methylation marks at four out of seven imprinted genes examined (Cicccone et al., 2009). Moreover, two genome-wide studies have also reported that H3K4 methylation is strongly negatively correlated with DNA methylation, further confirming H3K4me0 as a trigger for de novo methylation (Meissner et al., 2008; Weber et al., 2007).

Histone methyltransferases have also been found to interact with DNMT3 and in turn lead to de novo DNA methylation in associated genomic regions. Suv39h, a histone methyltransferase for H3K9, can interact with DNMT3A as well as DNMT1 both in vitro and in vivo (Fuks et al., 2003). Moreover, *Suv39h*<sup>-/-</sup> mESCs show a loss of Dnmt3b enrichment and DNA methylation at pericentric repeats, suggesting a critical role of Suv39h in Dnmt3b-dependent de novo methylation of those sequences (Lehnertz et al., 2003). Another H3K9 methyltransferase G9a has also been found to interact with both Dnmt3a and Dnmt3b independently of its histone methyltransferase activity, which may further induce de novo methylation in some early-embryonic genes (e.g. *Oct4*) in mESCs (Epsztejn-Litman et al., 2008). Consistently, a significant reduction of DNA methylation at retrotransposons, major satellite repeats and densely

methyated CpG-rich promoters has also been observed in mESC lacking G9a (Dong et al., 2008). Surprisingly, heterochromatin protein 1 which is not histone methyltransferase but specifically binds H3K9me3 can also directly recruit DNMT3 to induce de novo methylation of associated DNA (Smallwood et al., 2007). In addition to those H3K9 methyltransferases, Ezh2, another histone methyltransferase specific for H3K27, also shows a similar role in interacting with DNMTs and inducing de novo methylation at associated DNA (Vire et al., 2006). Polycomb group (PcG) proteins and DNA methylation system both are essential for the heritable repression of gene activity in mammals. Since Ezh2 is also an important component of Polycomb repressive complex 2, such function of Ezh2 forms a direct connection between these two key epigenetic repression systems (Vire et al., 2006).

The histone modifications mentioned above are all associated with transcriptional repression. Interestingly, H3K36me3 which is mainly located in actively transcribed gene bodies also exhibits a distribution pattern positively correlated with DNA methylation, suggesting that H3K36me3 may also work as a recruitment platform for DNMTs (Ball et al., 2009; Hodges et al., 2009; Mikkelsen et al., 2007). In support of it, a specific interaction between H3K36me3 and the PWWP domain of DNMT3A was recently revealed, and more importantly this interaction can further increase the activity of DNMT3A for methylation of nucleosomal DNA in vitro (Dhayalan et al., 2010).

In addition to histone modification and histone-modifying enzymes, chromatin remodeling proteins have also been reported to regulate de novo methylation. Lsh (lymphoid-specific helicase) is an important component of SNF/SWI chromatin remodeling complexes that induce sliding of the nucleosomes along the DNA by an ATP-dependent disruption of DNA-histone interaction

(Becker and Horz, 2002; Peterson, 2002). A genome-wide loss of DNA methylation has been reported in mice with depletion of Lsh (Dennis et al., 2001). Moreover, the deficiency of Lsh in mouse embryonic fibroblasts inhibits the acquisition of DNA methylation but does not affect the maintenance of previously methylated episomes, suggesting a specific role of Lsh in the establishment of novel methylation patterns (Zhu et al., 2006). Furthermore, an interaction of Lsh with Dnmt3a and Dnmt3b but not with Dnmt1 was also detected in mESCs, strongly supporting that Lsh can directly control de novo DNA methylation (Dennis et al., 2001; Myant and Stancheva, 2008). Additionally, SNF/SWI helicase ATRX ( $\alpha$ -thalassemia/MR, X-linked) seems to also have a regulatory function in de novo methylation, as its mutations significantly change the DNA methylation pattern in various repetitive sequences including the rDNA arrays, a Y-specific satellite and subtelomeric repeats (Gibbons et al., 2000).

Lastly, RNAs are also supposed to be involved in targeted de novo DNA methylation in mammalian cells. It has been well reported that siRNAs (small interfering RNAs) in plants can guide de novo DNA methyltransferases to set up sequence-specific DNA methylation (Zhang and Zhu, 2011). Although fewer studies have been done in mammals, it is highly possible that RNA-directed de novo methylation and transcriptional silencing is conserved in mammals. By studying human tissue culture cells, Morris et al. (2004) reported that promoter-directed siRNA induces de novo DNA methylation in promoters of both exogenous and endogenous genes as well as transcriptional repression. Additionally, MILI and MIWI2 (two mouse homologs of *Drosophila* Piwi proteins) in mouse fetal germ cells were found to be required for the DNA methylation in the regulatory regions of the retrotransposons, suggesting an essential role of piRNA in establishing de novo DNA methylation of retrotransposons (Aravin et al., 2008; Kuramochi-Miyagawa et al.,

2008).

### **1.1.5 Immunity of CGIs to DNA methylation**

CGI is usually identified as a patch of DNA with at least 200 bp length, CG content  $\geq 50\%$  and the ratio of observed CpG to expected CpG  $\geq 60\%$  (Gardiner-Garden and Frommer, 1987). In contrast to the heavily methylated non-CGI regions, CGIs are generally unmethylated in genomic DNA, constituting the most striking feature of DNA methylation pattern in mammals (Bird, 2002). Given that about 70% of annotated gene promoters coincide with a CGI, maintenance of methylation-free state in most CGIs should be critical for normal transcription program in cells (Saxonov et al., 2006). Then how are CGIs protected from the de novo DNA methylation wave during embryogenesis and gametogenesis?

The function of CGIs as transcription promoters is believed to mainly contribute to their methylation-free state. Although about half of CGIs are considered as orphan CGIs which locate either within or between characterized transcripts, accumulating evidences from gene-specific or genome-wide studies so far have indicated that most, perhaps all, of them are actually associated with transcription initiation (Illingworth et al., 2010; Maunakea et al., 2010). One pathway by which CGI promoters can resist de novo DNA methylation attack is the binding of transcription factors, which may markedly decrease the local accessibility of DNMT3. For example, the binding of Sp1 transcription factor plays a critical role in protecting mouse adenine phosphoribosyltransferase gene promoter from DNA methylation, as deletion or mutation of its binding sites results in a significant acquisition of de novo DNA methylation (Brandeis et al., 1994; Macleod et al., 1994). The other pathway may be related with the activity of transcription itself, though the underlying mechanism is still unclear. A genome-wide study has shown that up

to 90% of genes containing a CGI promoter are actively transcribed during early embryogenesis and gametogenesis (Sequeira-Mendes et al., 2009). Moreover, like the binding of transcription factors, the recruitment of RNA polymerase II at CGI promoters has also been reported to be involved in the immunity of CGIs from DNA methylation (Takeshima et al., 2009).

Interestingly, CGIs cover the initiation sites for not only transcription but also DNA replication (Delgado et al., 1998; Sequeira-Mendes et al., 2009). Such co-localization of CGIs with DNA replication origins may therefore provide another mechanism for the protection of CGIs from DNA methylation. It is highly possible that due to potential intermediates in DNA replication process, the de novo DNA methyltransferases may fail to be assembled into replication initiation complex or keep inactive at early replication stage, and as a consequence methylation-free state is established in CGIs (Antequera and Bird, 1999; Deaton and Bird, 2011).

Another potential mechanism underlying the immunity of CGIs to DNA methylation is by CGIs-related histone modifications. As mentioned in previous part, histone H3 tails with unmethylated lysine 4 (H3K4me0) can recruit DNMT3A and DNMT3B either directly or via DNMT3L to induce de novo DNA methylation in associated DNA. However, the H3K4 in CGIs is typically methylated into H3K4me3, which therefore leads to a poor chromatin binding of DNMT3L, DNMT3A and DNMT3B (Ooi et al., 2007; Zhang et al., 2010b). Moreover, the in vitro catalytic activities of DNMT3A and DNMT3B on chromatin with this histone modification are also much lower compared with that on chromatin with H3K4me0 (Zhang et al., 2010b). Additionally, the rarity of other histone modifications which can interact with DNMTs (including H3K9me3 and H3K27me3) in CGIs also favors the maintenance of their methylation-free state.

Taken together, several potential mechanisms have been proposed to protect CGIs from de

novo DNA methylation. Moreover, these mechanisms may simultaneously function to achieve an optimal effect. For example, both binding of transcription factors and active chromatin signature H3K4me3 are often associated with high transcription activity in a CGI promoter. In this case, each mechanism alone may be not sufficient to completely maintain a CGI at methylation-free state.

### **1.1.6 Aberrant DNA methylation in cancer**

Consistent with its critical regulatory effects on many important cellular processes (e.g. chromatin stability, genomic imprinting), DNA methylation has been found to be widely involved in tumor initiation, promotion, aggressiveness, and metastasis (Robertson, 2005). Global DNA hypomethylation and gene-specific hypermethylation both feature the aberrant DNA methylation pattern in cancers. The loss of DNA methylation in repetitive sequences primarily explains global DNA hypomethylation (Riggs and Jones, 1983; Yoder et al., 1997). Importantly, it also results in abnormal reactivation of transposable elements, disruption of normal transcriptional program and genomic instability, which is a hallmark of tumor cells (Robertson and Wolffe, 2000; Shames et al., 2007). In contrast to the global DNA hypomethylation, gene-specific hypermethylation has been found to often occur in CGIs and leads to an effective silencing of related genes transcription. These affected genes are mainly involved in cell-cycle regulation, DNA repair, apoptosis, cell signaling and tumor invasion, such as *P16*, *hMLH1* and *RASSF1A* (Robertson, 2005). Obviously, such gene-specific hypermethylation can provide tumor cells with an advantage in cell growth and invasion. Additionally, in almost all types of tumors the abnormally hypermethylated CGIs may form a CpG-island-methylator-phenotype (CIMP), which is defined as 3-5 fold increase in

methylation frequency (Issa, 2004). In addition to the reactivation of transposable elements and silencing of tumor suppressor genes, aberrant DNA methylation in cancer is also associated with loss of imprinting (Jelinic and Shaw, 2007). The resultant high expression of such gene product can also provide cells with a growth advantage. For example, loss of imprinting of IGF2/H19 locus has been detected a wide range of tumor types, including lung, liver, colon and kidney cancer (Jelinic and Shaw, 2007; Robertson, 2005).

Although the mechanism of its origin in cancer still remains poorly understood, aberrant DNA methylation has been considered as an important target for epigenetic diagnosis and therapy. The aberrant hypermethylation can be detected in the earliest precursor lesions of tumors, suggesting it may play a critical role in tumor transformation and be used as a biomarker for early cancer detection (Cravo et al., 1994). Moreover, in clinical, DNMTs inhibitors 5-azacytisine and 5-aza-2'-deoxycytidine have been approved for treatment of myelodys plastic syndromes and acute myeloid leukemia through inducing DNA demethylation, differentiation and reactivation of silenced tumor suppressor genes. And clinical trials about 5-aza-2'-deoxycytidine for solid tumors therapy are also ongoing.

## **1.2 Active DNA demethylation and potential mechanisms in mammals**

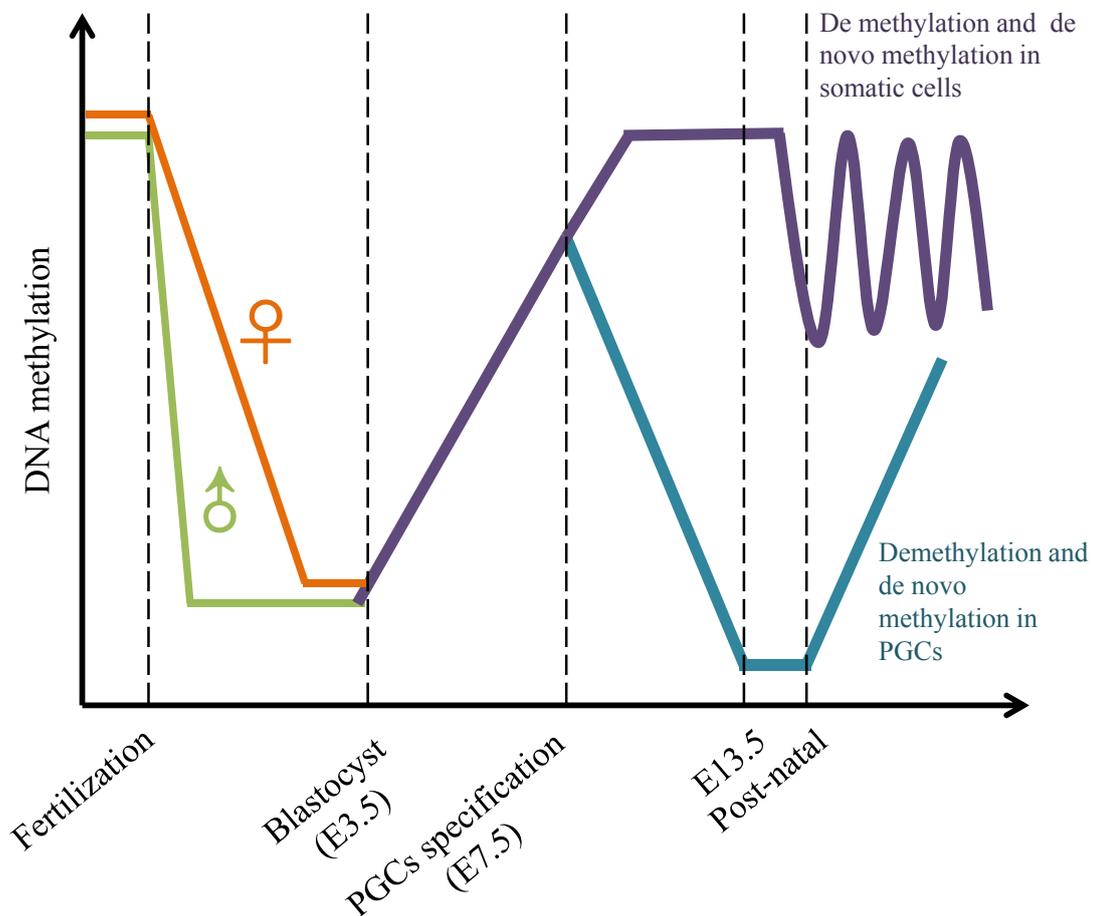
Given its role in long-term and heritable silencing of gene transcription, DNA methylation in mammals have been thought to be a relative stable epigenetic modification compared with histone modifications. However, a number of studies so far have reported that during specific development stages or in certain cell contexts global or gene-specific DNA methylation will be removed through either passive or active mechanism. Passive DNA demethylation is a DNA

replication-dependent process where the newly synthesized DNA strand failed to be methylated due to loss of DNMTs activity, while active DNA demethylation removes the methyl group of 5mC in a replication-independent and enzymatic manner. Based on a number of studies where DNMTs deficiency or DNMTs inhibitor treatment significantly disrupts DNA methylation maintenance, passive DNA demethylation has been widely accepted. By contrast, mainly because of the uncertainty of DNA demethylases the mechanism of active DNA demethylation in mammals still remains elusive, despite a clear mechanism revealed in plants (Ooi and Bestor, 2008; Wu and Zhang, 2010).

### **1.2.1 Evidences for active DNA demethylation**

Although the mechanisms remains elusive, the evidences for active DNA demethylation in mammals have been accumulated in the past decades. Based on the range of affected genomic regions, these reported DNA demethylation events can be generally divided into two groups: genome-wide and gene-specific DNA demethylation.

In 2000, genome-wide active DNA demethylation was first demonstrated in the paternal pronuclei of mouse zygotes (Figure 1A) (Mayer et al., 2000; Oswald et al., 2000). With the use of anti-5mC immunostaining assay, paternal pronuclei were found to be globally demethylated shortly (about 6-8 hours) after fertilization. Moreover, such DNA demethylation occurs before the first cell division of zygotes and cannot be blocked by treatment of aphidicolin (a DNA replication inhibitor), strongly indicating active DNA demethylation as the underlying mechanism. A more detailed change of DNA methylation pattern came from bisulfite-sequencing analysis, which confirmed the loss of DNA methylation but also showed some genomic regions (mainly



**Figure 1. Dynamics of DNA methylation during embryogenesis in mouse**

During mouse development, the paternal genome (♂) of zygote is rapidly and globally demethylated immediately after fertilization, whereas the maternal genome (♀) undergoes a DNA replication-dependant demethylation during the first cleavage stages. After implantation, a new DNA methylation pattern is re-established by de novo DNA methylation in the diploid genome. Primordial germ cells (PGCs) were specified at E7.5 and begin to migrate toward genital ridge, during which another round of global DNA demethylation occurs. De novo DNA methylation also occurs in PGCs when they enter meiosis. By contrast, most somatic cells maintain their DNA methylation pattern. However, certain adult somatic cells may undergo further demethylation or de novo methylation in specific genomic loci.

including imprinting control regions, retrotransposons, centric and pericentric region) are resistant to those demethylation activities (Oswald et al., 2000; Santos et al., 2002). By contrast, maternal pronuclei resist DNA demethylation at the same time, but interestingly undergo a gradual and passive DNA demethylation in the inaccessibility of DNMT1 during subsequent cell division in cleavage-stage embryos (Carlson et al., 1992; Howlett and Reik, 1991). That genome-wide DNA demethylation in zygotic paternal pronuclei could lead to transcriptional reactivation of paternal genes, and thus may be essential for early embryonic development.

Another well-reported genome-wide active DNA demethylation occurs in mouse primordial germ cells (PGCs) (Figure 1B). PGCs are germ cell precursors and derived from a few posterior epiblast cells which are induced by Bmp4 and Bmp8 signals from neighbor cells at embryonic day 7.5 (E7.5) (Ohinata et al., 2009). PGCs begin to migrate at E8.5 and reach the genital ridge at E11.5, followed by a rapid global DNA demethylation (Hajkova et al., 2002; Lee et al., 2002). The intact function of Dnmt1 and limited cell division taken during that period support a mechanism of active DNA demethylation. Most strikingly, different from that in the paternal pronuclei of zygotes, the DNA demethylation in PGCs also occurs in imprinted genes, which is necessary for establishing a new DNA methylation pattern in germ cells (Lee et al., 2002).

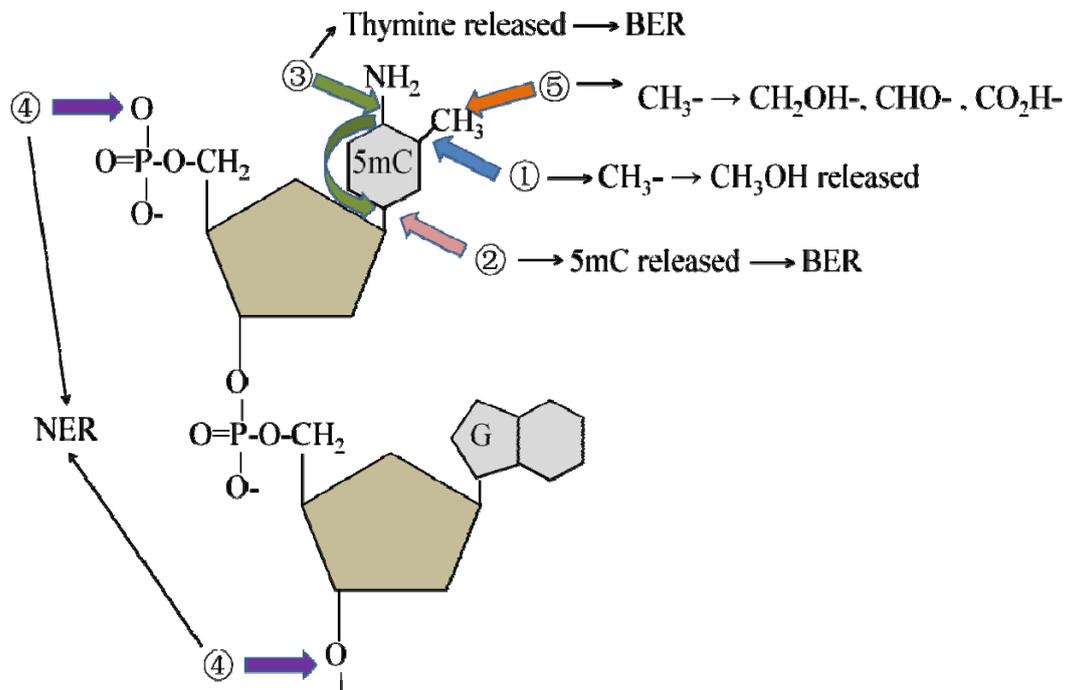
Based on those above evidences, genome-wide active DNA demethylation in mammals appears to only take place at specific stages of early development. By contrast, considerable studies supported the occurrence of gene-specific active DNA demethylation in differentiated somatic cells in response to specific stimulus (Wu and Zhang, 2010). For example, in an in vivo model of hippocampal neurogenesis of adult mice, electro-convulsive treatment rapidly induces DNA demethylation at the regulatory region of brain-derived neurotrophic factor (Bdnf) and the

brain-specific promoter B of fibroblast growth factor-1 (Fgf-1) without a significant global DNA demethylation (Ma et al., 2009). Moreover, this activity-dependent active DNA demethylation contributes to the upregulation of those growth factor families and thus plays an essential role in neurogenesis (Ma et al., 2009). In addition, Bruniquel and Schwartz (2003) also reported that a significant DNA demethylation in the promoter-enhancer of the interleukin-2 (Il2) gene also occurs as early as 20 minutes after activation of T cells, and is necessary and sufficient for the enhancement of Il2 promoter-driven transcription. The absence of DNA replication further supports that active demethylation is the underlying mechanism (Bruniquel and Schwartz, 2003).

### **1.2.2 Potential mechanisms of active DNA demethylation and DNA demethylases**

If there are DNMTs that methylate DNA, DNA demethylases must also exist to remove methylation mark and fully achieve the biological functions of DNA methylation. This belief leads to intense efforts to identify potential DNA demethylases in mammals. So far, several mechanisms for active DNA demethylation have been proposed, including 1) hydrolysis; 2) base excision repair (BER) with the use of 5mC DNA glycosylases; 3) deamination of 5mC coupled with BER; 4) nucleotide excision repair (NER); and 5) oxidative demethylation (Figure 2). These proposed mechanisms have also been well reviewed by Zhu (2009) and Wu and Zhang (2010).

The mechanism of hydrolysis by which the methyl group of 5mC is directly released in a form of methanol represents the simplest way to actively remove DNA methylation modification. However, given the breakage of the strong carbon-carbon bond is thermodynamically unfavorable, it seems unlikely to achieve active demethylation by hydrolysis. Actually, the first finding that via this mechanism methyl-CpG-binding domain protein 2 (MBD2) induces demethylation in vitro at



**Figure 2. Potential mechanisms for active DNA demethylation in mammals**

Multiple possible mechanisms have been reported for active DNA demethylation: (1) the methyl group is directly hydrolyzed and released as methanol; (2) potential 5mC glycosylases remove 5mC base, followed by base excision repair (BER); (3) 5mC is first deaminated into T, then G/T mismatch glycosylases remove T, followed by BER; (4) methylated CpG dinucleotides and/or flanking regions are removed by nucleotide excision repair pathway (NER); (5) methyl group is consecutively oxidized into hydroxymethyl, formyl and carboxyl group, followed by BER or other mechanisms to achieve DNA demethylation.

the absence of any co-factors, has never been successfully replicated by other laboratories (Bhattacharya et al., 1999; Bird, 2002). Moreover, both normal DNA methylation pattern and global DNA demethylation in the paternal pronuclei of zygotes have also been observed in MBM2-null mice (Hendrich et al., 2001; Santos et al., 2002). Thus, the mechanism of hydrolysis for active DNA demethylation is highly debatable.

Another mechanism for active DNA demethylation is by BER with the use of 5mC DNA glycosylases. In this mechanism, the 5mC base is first recognized and cleaved by 5mC DNA glycosylases, followed by removal of deoxyribose at the apurinic/apyrimidinic (AP) sites by AP endonuclease. The resultant gap is finally filled with an unmethylated cytosine by DNA polymerase and DNA ligase. Although it has been solidly proved to underlie active DNA demethylation in plants (Zhu, 2009), this mechanism in animal still remains undetermined. Both thymine DNA glycosylase (TDG) and MBD4 show 5mC DNA glycosylase activity in vitro, but such activity against 5mC is 30-40-fold lower than their G/T mismatch repair activity (Zhu et al., 2000a; Zhu et al., 2000b). Additionally, deficiency of TDG leads to embryonic lethality in mice, whereas MBD4-null mice are viable and also show normal DNA methylation pattern (Millar et al., 2002). Thus, future studies are needed to further clarify this mechanism in mammals.

Although TDG and MBD4 have very weak 5mC DNA glycosylase activity, their strong G/T mismatch repair activities are well utilized in another mechanism to achieve active demethylation. Here 5mC is first converted to T through activation induced deaminase (AID)/apolipoprotein B RNA-editing catalytic component-1 (APOBEC-1)-catalyzed deamination. The resultant T is then cleaved by TDG/MBD4 and replaced with unmethylated cytosine via BER (Cortellino et al., 2011; Rai et al., 2008). In support of it, interactions between TDG/MBD4 and AID has been detected in

mouse (Cortellino et al., 2011) and zebra fish embryos (Rai et al., 2008), respectively. Interestingly, a recent study reported that DNMT3A and DNMT3B can also deaminate 5mC to T and thus participate in estradiol-induced active demethylation in estrogen receptor target gene TFF1 (Metivier et al., 2008). This surprising function of DNMTs awaits further confirmation. In contrast to above supportive evidences, however, AID-null and APOBEC1-null mice exhibit a normal development and reproductive ability, raising some doubts on the significance of AID and APOBEC1 in active DNA demethylation (Morrison et al., 1996; Revy et al., 2000). Compared that in the wild type counterparts, the DNA methylation levels in PGCs from AID-null embryos increase but are still much lower than that in mESCs and somatic cells, indicating that AID is not critical for the global demethylation in PGCs (Popp et al., 2010).

Active DNA demethylation could also be induced through NER. Different from BER which fixes a single base of DNA damage caused by oxidation, alkylation or deamination, NER repairs chemical- or radiation-induced bulky, helix-distorting lesions such as pyrimidine dimers and 6,4 photoproducts. The primary evidences supporting NER as a mechanism of active demethylation came from a study on growth arrest and DNA-damage-inducible protein 45 alpha (GADD45A), a protein with diverse biological functions including DNA damage response and NER (Barreto et al., 2007; Zhan, 2005). In that study, overexpression of GADD45A activated methylation-silenced reporter plasmids and also induced global DNA demethylation, while its knockdown regulated gene expression and resulted in DNA hypermethylation (Barreto et al., 2007). Moreover, XPG, another essential player of NER, interacts with GADD45A and is also required for the DNA demethylation induced by GADD45A overexpression (Barreto et al., 2007). However, another independent group repeated these same experiments but failed to observe a similar function for

GADD45A at all (Jin et al., 2008). Moreover, Gadd45a-null mice also show neither global nor loci-specific hypermethylation phenotype (Engel et al., 2009). Thus, that NER works as a mechanism for active demethylation has still not been conclusively proved.

In addition to above possible mechanisms, oxidative reaction may also constitute another pathway for active DNA demethylation in mammals. AlkB, a 2-oxyoglutarate (2OG)-dependent dioxygenase in bacteria, is involved in DNA damage repair by removing the methyl groups of 1-methyladenine and 3-methylcytosine in a form of formaldehyde (Falnes et al., 2002). Interestingly, in thymidine salvage pathway thymine 7-hydroxylase in fungi consecutively oxidizes thymidine into iso-orotate, which can be further converted to uridine by a decarboxylase (Smiley et al., 2005; Warn-Cramer et al., 1983). Thus, although the breakage of a carbon-carbon bond is thermodynamically unfavorable, these studies prove the existence of enzymes with such catalytic function. In 2009, under an effort to identify human homologous proteins for JBP1/2 (trypanosome base J-binding proteins, with a similar catalytic function as thymine 7-hydroxylase), Rao group published a groundbreaking finding that human ten-eleven translocation (TET) family proteins are also 2-oxyoglutarate (2OG)-dependent dioxygenases and capable to catalyze the conversion of 5mC to 5-hydroxymethylcytosine (5hmC) in vitro and in vivo (Tahiliani et al., 2009). Since then, much effort has been devoted to the study of the role of TET in active DNA demethylation in mammals, and lots of supportive evidences have so far been accumulated. Those studies will be detailedly reviewed in next section “TET1: a novel 5-methylcytosine dioxygenase”.

Taken together, there are multiple potential mechanisms for active DNA demethylation in mammals. However, due to conflicting results, unsuccessful replication, or lack of decisive

biochemical and genetic evidences, none of them (including TET-mediated oxidative demethylation) has been conclusively proved. Possible biological redundancy may partially explain the lack of decisive genetic support from previous knockout experiments. Additionally, it is also proposed that these mechanisms exist to promote active demethylation but each one may only function in its specific cell context (Wu and Zhang, 2010).

### **1.3 TET1: a novel 5-methylcytosine dioxygenase**

#### **1.3.1 TET-mediated 5mC oxidation leads to DNA demethylation**

TET1, the founding member of TET protein family, was initially identified as a rare fusion partner of mix-lineage leukemia gene (*MLL*) in several acute myeloid leukemia cases containing the t(10;11)(q22;q23) (Lorsbach et al., 2003; Ono et al., 2002). Through BLAST analysis of the NCBI and Celera databases, Lorsbach et al. (2003) further identified two other human homologous proteins (TET2 and TET3), as well as three mouse homologous Tet proteins (Tet1, Tet2 and Tet3). All three TET proteins have conserved Cys-rich and DSBH domains in C-terminal, and TET1 and TET3 also have an N-terminal CXXC domain, which is a specific DNA binding domain for CpG motif (Figure 3).

The biological functions of TET proteins remain unknown until 2009, when Rao group first identified human TET1 as a novel Fe(II)/ 2OG-dependent 5mC hydroxylase which catalytically converts 5mC to 5hmC (Tahiliani et al., 2009). Then all three mouse Tet proteins were also successfully proved to convert 5mC into 5hmC through their conserved C-terminal catalytic domain which consists of Cys-rich and DSBH domains (Ito et al., 2010). Moreover, more recent studies further found that TET proteins can consecutively convert 5mC into 5hmC,



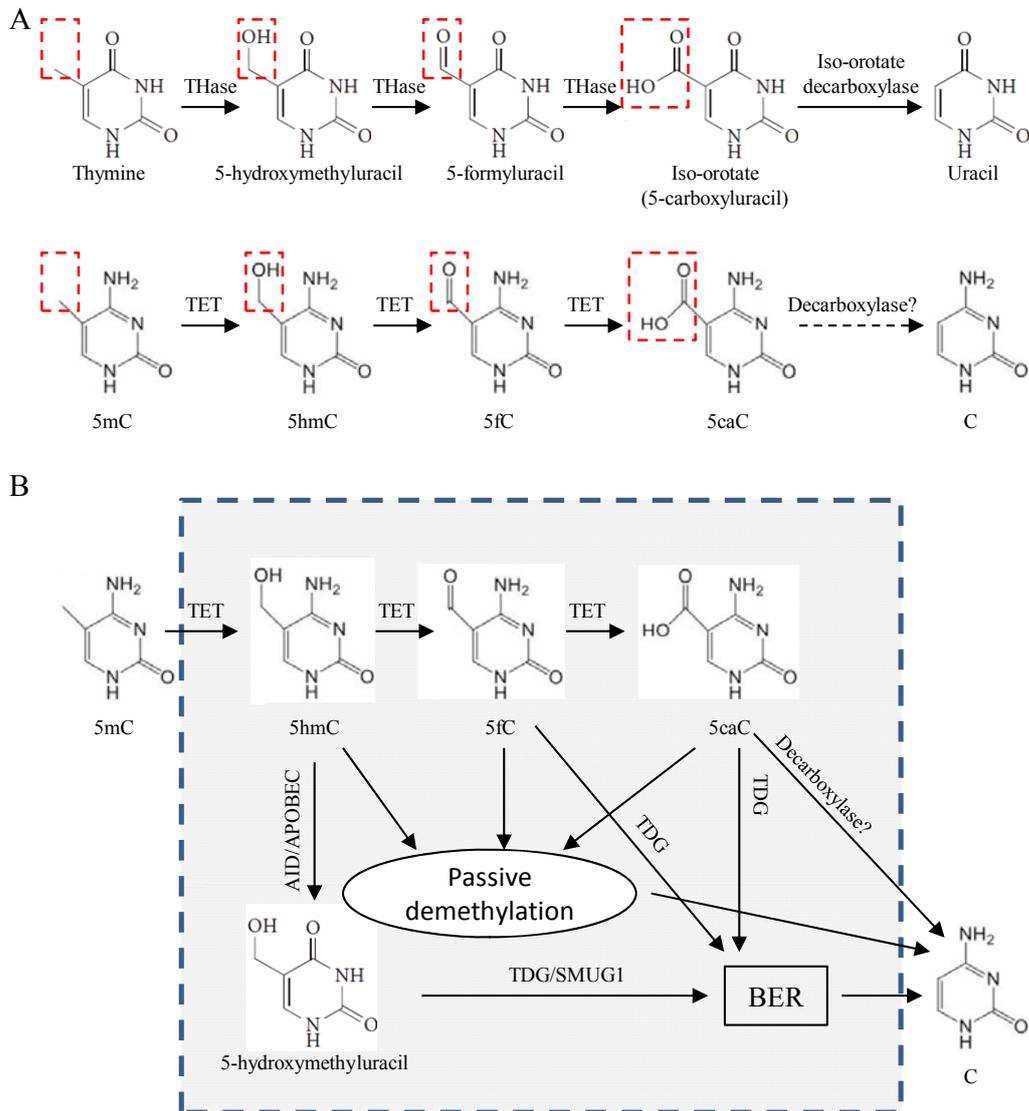
**Figure 3. Domain structure of human TET family proteins**

All three TET proteins have C-terminal conserved Cys-rich and DSBH domains, which form the catalytic domain for 5-methylcytosine dioxygenase activity. TET1 and TET3 also have an N-terminal CXXC domain, which is a specific DNA binding domain for the CpG motif.

5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), which is very similar to thymine 7-hydroxylase-mediated oxidative reaction in thymidine salvage pathway (He et al., 2011; Ito et al., 2011) (Figure 4A). Meanwhile, a relative abundance of those 5mC derivatives, especially 5hmC which has been long thought to be a rare DNA damage product in mammalian DNA, was also detected in mESCs, Purkinje neurons and various mouse tissues (Ito et al., 2011; Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). These breakthrough findings not only demonstrated a unique enzymatic activity of TET proteins to directly modify 5-methyl group of 5mC, but also raised an intriguing possibility that TET-mediated 5mC oxidation may constitute a novel DNA demethylation pathway in mammals. So far considerable evidences have been accumulated in support of this hypothesis. Overexpression of the catalytic domain of TET1 (TET1-CD) in HEK293T cells induces a significant DNA demethylation in exogenous nonreplicable 5mC-containing DNA probes as well as many endogenous genomic loci, while 5hmC-containing DNA probes can also be demethylated even at the absence of TET1-CD overexpression, providing direct evidence that TET-mediated 5mC oxidation can promote active DNA demethylation (Guo et al., 2011). The frequent mutations of TET2 in various myeloid malignancies are predominantly associated with lower 5hmC level and hypermethylation phenotype in patient bone marrow genomic DNA, which also supports a potential demethylating effect for TET2 (Figuroa et al., 2010; Ko et al., 2010). Additionally, Tet3-mediated 5mC oxidation was demonstrated to produce asymmetric marking of 5hmC in the paternal pronuclei of mouse zygotes, which highly coincides with the loss of 5mC immunoreactivity in paternal pronuclei after fertilization and may further lead to DNA demethylation at specific genomic loci, including the regulatory regions of Oct4 and Nanog genes (Gu et al., 2011; Iqbal et al., 2011;

Wossidlo et al., 2011).

Although these above evidences have shown that TET-mediated 5mC oxidation forms a novel pathway for DNA demethylation in mammalian cells, the detailed mechanism is still unclear (Figure 4B). 5hmC is poorly recognized by DNMT1 and genomic 5hmC, 5fC and 5caC undergo cell-division-dependent dilution in mouse embryo, suggesting a DNA replication-dependent passive demethylation by TET proteins (Inoue et al., 2011; Inoue and Zhang, 2011; Valinluck and Sowers, 2007). However, it has also been reported that overexpression of TET1-CD in HEK293T cells significantly induces DNA demethylation of exogenous nonreplicable 5mC-containing DNA probes as well as many endogenous genomic loci, strongly indicating that TET can promote active DNA demethylation (Guo et al., 2011; Zhang et al., 2010a). Moreover, inhibition of either Apurinic-aprimidinic endonuclease-1 (APEX1) or Poly(ADP-ribose) -polymerase-1 (PARP1), two key components of BER, completely blocks the DNA demethylation in not only exogenous DNA probes but also endogenous genomic loci, indicating that TET-mediated DNA demethylation may be only through BER-involved active pathway. Consistently, recent studies also reported that 5fC and 5caC in CpG dinucleotides can be efficiently excised by TDG, supporting that TDG-initiating BER may underlie TET-mediated DNA demethylation (He et al., 2011; Maiti and Drohat, 2011). On the contrary, however, both 5fC and 5caC in mouse zygotes were shown to be relatively stable and exhibited replication-dependent dilution instead of being quickly removed through such TDG-catalyzed process (Inoue et al., 2011). Moreover, given Tet3 induces a rapid global DNA demethylation in the paternal pronuclei of zygotes (Gu et al., 2011; Wossidlo et al., 2011), BER is also unlikely to act in Tet3-mediated DNA demethylation as it would put tremendous pressure on the repair machinery of zygotes (Wu and Zhang, 2010).



**Figure 4. TET-mediated 5-methylcytosine oxidation reactions and potential mechanisms for DNA demethylation**

(A) Top panel describes thymine hydroxylase (THase)-mediated oxidative reaction in thymidine salvage pathway. Thymine is converted to 5-hydroxymethyluracil, 5-formyluracil, and iso-orotate by THase in three consecutive oxidation reactions. Iso-orotate is then converted to uracil by iso-orotate decarboxylase. Lower panel describes 5-methylcytosine (5mC) oxidation reactions induced by TET family proteins. TET proteins consecutively convert 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), which may be further converted into unmethylated cytosine (C) by a potential decarboxylase like iso-orotate decarboxylase. (B) Potential mechanisms for DNA demethylation induced by TET-mediated 5mC oxidation. AID: activation-induced (cytidine) deaminase; APOBEC: apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like; TDG: thymine-DNA glycosylase; SMUG1: single-strand selective monofunctional uracil DNA glycosylase 1; BER: base excision repair.

Interestingly, it has been supposed that like that in thymidine salvage pathway, a potential decarboxylase may directly convert 5caC into unmethylated cytosine to realize active demethylation (Ito et al., 2011). Therefore, so far the mechanism for TET-mediated DNA demethylation has been still elusive.

Collectively, based on their potent oxidative activities on 5mC, TET proteins have shown great potential as the long-sought DNA demethylases that induce DNA demethylation through an undetermined mechanism, and future studies comparing the effects of different TET proteins on DNA demethylation and elucidating the underlying mechanisms are necessary to significantly improve our understanding of this novel DNA demethylation pathway.

### **1.3.2 Studies on the biological functions of TET1**

Tet1 is highly expressed in mESCs and rapidly down-regulated once differentiation is induced, suggesting a potential role of Tet1 in the pluripotency and self-renewal of mESCs (Tahiliani et al., 2009). In order to well define its function, the genome-wide distribution of Tet1 chromatin binding sites in mESCs has been investigated by several independent groups using chromatin immunoprecipitation coupled with high throughput DNA sequencing (ChIP-seq). Wu et al. (2011) reported that Tet1 preferentially binds CpG-rich sequences at promoters of both transcriptionally active (including some pluripotency-related genes) and Polycomb-repressed genes. Interestingly, Tet1 shows dual functions of transcriptional regulation by promoting transcription of pluripotency factors as well as participating in the repression of Polycomb-targeted developmental regulators, suggesting a critical role of Tet1 in reaching balance between pluripotency maintenance and lineage commitment (Wu et al., 2011). Consistent with it, another study from the same group showed that knockdown of Tet1 could impair the self-renewal of mESCs through

down-regulation of *Nanog* and also results in a bias towards trophoctoderm differentiation in pre-implantation embryos (Ito et al., 2010). On the other hand, other two laboratories also confirmed the chromatin binding profile of Tet1 and its dual functions in transcriptional regulation in mESCs (Williams et al., 2011; Xu et al., 2011). Interestingly, the regulation of Tet1 on gene transcription has also been found to be possibly independent on its catalytic activity (Williams et al., 2011). Additionally, these studies further revealed that 5hmC mainly locates in transcription start sites (TSS) and gene body and such distribution can be regulated by Tet1 in mESCs (Williams et al., 2011; Xu et al., 2011). However, in these studies knockdown of *Tet1* did not induce any change in cell morphology and proliferation, which contradicts the previously reported result that knockdown of Tet1 impairs the self-renewal of mESCs (Ito et al., 2010). To precisely identify the role of Tet1 in pluripotency of mESCs and in vivo development, *Tet1*-null mice and mESCs were also developed (Dawlaty et al., 2011). Despite significant decrease of 5hmC content in genomic DNA and skewed differentiation toward trophoctoderm in vitro, deficiency of Tet1 does not induce significant change of key pluripotency genes expression and cell morphology in mESCs (Dawlaty et al., 2011). More importantly, although some pups have a relatively smaller body size at birth, the Tet1-null mice are viable, fertile and grossly normal, indicating that Tet1 may be not essential for embryonic and postnatal development (Dawlaty et al., 2011). However, the result that intercross of Tet1-null males and females produces less progenies, suggesting that Tet1 deficiency may impair gametogenesis (Dawlaty et al., 2011; Wu and Zhang, 2011). Therefore, based on an overall consideration of those available studies, Tet1 seems to be not essential for the pluripotency and self-renewal of mESCs.

Interestingly, it was also reported that in the dentate gyrus of adult mice, shRNA-mediated

knockdown of Tet1 leads to inhibition of neuronal activity-induced active DNA demethylation in promoters of some genes including *Bdnf* and *Fgf1*, indicating Tet1 may be involved in hippocampal neurogenesis (Guo et al., 2011). Future studies with Tet1-null mice focusing on the neural development and functions will provide more confirmative evidences. In addition, TET1 was initially identified as a fusion gene in AML (Lorsbach et al., 2003; Ono et al., 2002), raising a question that if TET1 is a tumor suppressor gene. However, by contrast to TET2 which is one of the most frequently mutated genes in hematological malignancies, TET1 mutation so far has been rarely detected in cancers, and thus few knowledge now has been obtained for the role of TET1 in cancer.

Taken together, TET1 preferentially binds CpG-rich gene promoters and regulates gene transcription through or not its catalytic function in mESCs. Tet1 may be not required for the pluripotency and self-renewal of mESCs, and moreover the study on *Tet1*-null mice also basically excludes a critical role of TET1 for embryonic and postnatal development. Additionally, TET1 may be involved in neurogenesis and human cancers, but more supportive evidences are still required. Further studies, particular those with the use of *Tet1*-null mice, are necessary to improve our understanding of the biological functions of TET1.

#### **1.4 Hypothesis, specific aims and rationales**

**Hypothesis:** TET1 is an important epigenetic regulator of DNA methylation with a potent demethylating function.

**To test this hypothesis, three specific aims were carried out:**

1. To globally study whether full length TET1 overexpression induces DNA demethylation by its

5mC dioxygenase enzymatic activity.

2. To investigate the changes of DNA methylation after TET1 knockdown in human cell lines.
3. To study the mechanism of TET-mediated DNA demethylation.

**Rationale:** TET1 has been proven as a 5mC dioxygenase, but its function in DNA demethylation is still unknown. Previous studies revealed that overexpression of TET1-CD is capable to induce DNA demethylation in exogenous methylated DNA probes or endogenous genomic loci in HEK293T cells or adult mouse neurons, but no similar study has been done with full-length TET1 (TET1-FL) overexpression (Guo et al., 2011; Zhang et al., 2010a). We hypothesized that other domains in TET1, especially the CXXC domain possibly responsible for its binding to CpG-rich genomic regions, may play an important role in regulation of its catalytic activity on 5mC, and thus make TET1-FL quite different from TET1-CD in regard of DNA demethylating activity. In addition, the previous results that DNA methylation levels increase in Tet1-bound CpG-rich regions after knockdown of Tet1 in mESCs only suggest a possible role of Tet1 in maintaining DNA hypomethylation state but do not directly support a demethylating activity for TET1 (Wu et al., 2011). Moreover, the impairment of mESCs maintenance and differentiation induction by Tet1 knockdown add many other complexities to explain that increase of DNA methylation in mESCs (Ito et al., 2010). Thus, ESCs seems to be not suitable for studying the effect of TET1 knockdown on DNA methylation. Lastly, as for the mechanism of TET-mediated DNA demethylation, TDG-initiating BER seems to actively replace 5hmU (product from deamination of 5hmC), 5fC and 5caC with unmethylated cytosine (He et al., 2011; Maiti and Drohat, 2011), while the replication-dependent dilution of 5hmC, 5fC and 5caC in mouse zygotes suggests a passive

demethylation mechanism (Inoue et al., 2011; Inoue and Zhang, 2011). Thus, both passive and active demethylation pathways have not been conclusively proven.

## **CHAPTER 2 MATERIAL AND METHODS**

### **2.1 Cell culture**

Human embryonic kidney cell lines HEK293T and HEK293FT were obtained from the American Type Tissue Culture Collection and Invitrogen, respectively, and maintained in Dulbecco's Modified Eagle Medium (DMEM) modified with 10% fetal bovine serum and 100 µg/ml streptomycin-penicillin. The selected HEK293T single clone cells with TET1 knockdown or inducible TET1-CD overexpression were maintained in Dulbecco's Modified Eagle Medium modified with 10% fetal bovine serum and 1~1.5 µg/mL puromycin.

### **2.2 Gene cloning and plasmid construction for TET1 overexpression**

To clone human full length TET1 open-reading frame (TET1-FL ORF), total RNA was extracted from SY5Y cells (gift from Dr. Howard B. Gutstein, M.D. Anderson Cancer Center) using TRIzol® Reagent (Invitrogen). Given the big size of TET1-FL ORF, reverse transcript was performed with TET1-specific primer and AccuScript PfuUltra II RT-PCR Kit (Stratagene) according to manufacturer's instructions. The completeness of the resultant cDNA, was firstly tested by PCR with primers targeting various TET1 exons. Subsequently, PCR amplification of TET1-FL ORF was done with AccuPrime™ Taq DNA Polymerase High Fidelity (Invitrogen). The PCR product was then purified from gel electrophoresis and cloned into pCR®-XL-TOPO® vector (Invitrogen). Subsequently, to construct TET1-FL overexpression plasmid, TET1-FLORF was transferred into pIRES-hrGFP II vector (Stratagene) which has a C-terminal 3×Flag tag and GFP reporter. TET1-CD ORF amplified from the above TET1-FL ORF clone was also cloned into

the same expression vector. The sequences of all clones were validated by Sanger DNA sequencing. All primers used above are listed on Table 1.

### **2.3 Site-directed mutagenesis**

The catalytically mutant TET1-FL and TET1-CD (H1671D, Y1673A) (mTET1-FL and mTET1-CD) and CXXC domain-mutated TET1-FL (C594A) were generated by site-directed mutagenesis with a homemade kit. In brief, the wild type plasmids were amplified for 20-25 cycles with primers containing the desired mutation and PfuUltra HF DNA polymerase (Stratagene). The resultant PCR products were digested with FastDigest® DpnI (Fermentas) to remove the original plasmids, and then transformed TOP10 chemical competent cells (Invitrogen). The sequences of those mutant plasmids were confirmed by DNA sequencing. The primers were designed with online QuikChange Primer Design soft (Stratagene) and listed on Table 1.

### **2.4 Plasmid transfection**

For (m)TET1-FL and (m)TET1-CD overexpression, pIRES-(m)TET1-FL-Flag-GFP and pIRES-(m)TET1-CD-Flag-GFP were transfected into HEK293T cells using FuGene® HD Transfection Reagent (Roche). The transfection complex was prepared by mixing transfection reagent with plasmids at a ratio of 4:1( $\mu$ l: $\mu$ g) in OPTI-MEM® Reduced Serum Medium (Invitrogen). 18-24 hour after transfection, the transfection complex was removed and cells were culture in regular medium.

**Table 1. PCR primers and sequencing primers for TET1 plasmid construction**

	Primer name	Forward (5'-3')	Reverse (5'-3')	Tm (°C)	Amplicon size (bp)
<b>Cloning primers</b>	TET1-GSP	TATATACTGCAAGTTGCTAATACTTGAATG	NA	NA	NA
	TET1-EX12-F/R	GCGCGAGTTGGAAAGTTTG	GCGCAGGAAACAGAGTCATT	56	181
	TET1-EX34-F/R	GAGGGAAAAGAAGCCCAAAG	TTTTGTTCTTCCCCATGACC	56	111
	TET1-EX1112-F/R	CCGAATCAAGCGGAAGAATA	ACCAGGAGAAGCCTGGAGAT	56	215
	ACTB-EX12-F/R	ACAGAGCCTCGCCTTTGC	CACGATGGAGGGGAAGAC	56	160
	TET1-FL ORF -F/R	ATTATGGACTCTGTAGCTATGTCTCGA	ATCCTACAGACCCAATGGTTATAGG	56	6,434
	TET1-CD ORF -F/R	ACCATGGCTAAAGATTCTGAACTGC	TAAAACGACGGCCAGTGAAT	56	2,294
	<b>Sequencing primers</b>	Primer name	Sequence (5'-3')		
XLTET1-S-F2		ATGATACCAGTGGTTCCCCA			
XLTET1-S-F3		CATTAGCCCCTGAGAGAGGA			
XLTET1-S-F4		AGAATTCGGCAAGACATTGG			
XLTET1-S-F5		GCAACCCATACCCAAATTGA			
XLTET1-S-F6		AAACACCCTTACCGGAGTCA			
XLTET1-S-F7		TGGCTACACGATTAGCTCCA			
XLTET1-S-F8		TTCAGAAAGAAGCAGCACTCC			
XLTET1-S-R2		TGGGGAACCACTGGTATCAT			
XLTET1-S-R3		TCCTCTCTCAGGGGCTAATG			
XLTET1-S-R4		CCAATGTCTTGCCGAATTCT			
XLTET1-S-R5		TCAATTTGGGTATGGGTTGC			
XLTET1-S-R6		TGACTCCGGTAAGGGTGTTT			

**Table 1. PCR primers and sequencing primers for TET1 cloning, plasmid sequencing and site-directed mutagenesis (continued)**

Sequencing primers	Primer name	Sequence (5'-3')			
	XLTET1-S-R7	TGGAGCTAATCGTG TAGCCA			
	XLTET1-S-R8	GGAGTGCTGCTTCTTTCTGAA			
Site-directed mutagenesis primers	Primer name	Forward (5'-3')	Reverse (5'-3')	Tm (°C)	Amplicon size (bp)
	TET1-CD-Mutant-F/R	GACTTCTGTGCTCATCCCTACAGGGCCA TTCACAACATGAATAA	TTATTCATGTTGTGAATGGCCCTGTA GGGATGAGCACAGAAGTC	56	9,953
	TET1-CXXC-Mutant-F/R	AGCGATGTGGGGTCGCTGAACCCTGCC AGC	GCTGGCAGGGTTCAGCGACCCAC ATCGCT	55	11,983

## **2.5 Fluorescence-activated cell sorting (FACS)**

To collect GFP positive cells, the HEK293T cells transfected with pIRES-(m)TET1-FL-Flag-GFP or pIRES-(m)TET1-CD-Flag-GFP were collected with GIBCO® 0.25% trypsin (Invitrogen), spun and resuspended in DMEM only supplemented with 100 µg/ml streptomycin-penicillin at a density of  $5 \times 10^6$  cells/ml. These cells were then filtered through Falcon® Cell Strainers (40 µm, BD Biosciences), and sorted with Becton Dickinson FACS Calibur Flow Cytometer at the Flow Cytometry Core Facility at the University of Texas M.D. Anderson Cancer Center.

## **2.6 DNA extraction and bisulfite-pyrosequencing/sequencing**

Genomic DNA was extracted from cell lines according to the following protocol.  $10^6$ - $10^7$  cell pellet was lysed by 450 µl cell lysis solution (25mM Tris-Cl pH 8.0, 10mM EDTA pH 8.0, 1% SDS) with 2 µl RNase A solution (20 mg/µl), mixed well and incubated at 37 °C for 1 hour. Then 150 µl of 10M ammonium acetate was added and vigorously mixed, followed by incubation in ice for 5 minutes and centrifugation for 15 min (3,000 rpm, 4°C). The supernatant was then transferred to a new tube, and mixed with 600 µl isopropanol. After centrifugation for 20 min (14,000 rpm, 4°C), the resultant pellet was washed with 70% ethanol and then resuspended in 50-200 µl TE buffer (10 mM Tris pH 8.0, 1 mM EDTA). DNA concentration was measured by NanoDrop™ 1000 spectrophotometer (Thermo Scientific).

Bisulfite conversion of genomic DNA was done with EpiTect bisulfite kits (Qiagen) according to manufacturer's instructions. Briefly, 2 µg genomic DNA was used for each conversion reaction, and the converted DNA was finally eluted from the column with 40 µl of EB buffer (10 mM Tris

pH 8.5)

For bisulfite-pyrosequencing, we generally used two-step PCR for amplification. In the first step, 1 µl of each bisulfite-converted DNA sample was used in each reaction (25 µl total volume). In the second step, 0.1-0.3 µl of the 1<sup>st</sup> step PCR product was used as template, and biotinylated forward or reverse primers was used for generation of biotin-labeled PCR product. Instead of the biotinylated forward or reverse primers, the 5' tailed forward or reverse primers and a biotinylated universal primer (5'-GGGACACCGCTGATCGTTTA-3') can also be used to label PCR product (Colella et al., 2003). With the Pyrosequencing Vacuum Prep Tool (Biotage) the biotin-labeled DNA strands were captured by streptavidin sepharose beads (GE Healthcare). Then they are annealed to sequencing primers and sequenced by the PSQ HS 96 Pyrosequencing system (Biotage). Lastly, the results were analyzed with Pyro Q-CpG Software (Qiagen) software.

For bisulfite-sequencing, a similar two-step PCR as that in bisulfite-pyrosequencing was used but no biotinylated primers were used in the 2<sup>nd</sup> step PCR. The final PCR product was then cloned into pCR4-TOPO vector (Invitrogen) and transformed TOP10 chemical competent cells (Invitrogen). After ~14 hours of incubation at 37 °C, individual clones were pick up and amplified with PCR using pCR4 forward primers (5'-TCTGGAATTGTGAGCGGATA-3') and reverse primer (5'-GTTTTCCCAGTCACGACGTT-3'). Those PCR products were then sequenced with M13-RV primer at the DNA sequencing core facility at the University of Texas M.D. Anderson Cancer Center or Beckman Coulter Genomic Service. The bisulfite-pyrosequencing/sequencing primers used above were designed with PSQ Assay Design soft (Biotage) or MethPrimer soft ((Li and Dahiya, 2002), and listed on Table 2 and 3, respectively.

**Table 2. PCR and sequencing primers used in bisulfite-pyrosequencing assay**

Gene name	Step	Forward (5'-3')	Reverse (5'-3')	Sequencing (5'-3')	Target (5'-3')	Tm (°C)	Amplicon size (bp)
<i>LINE1</i>	1	TTTGAGTTAGGTGTG GGATATA	Biotin-AAAATCAAAAATTC CCTTTC	AGTTAGGTGTGG GATATAGT	TTT/CGTGGTGT/CGT T/CGTTTTTTAAGTT/ CGGTTTGAAAAG	56	
<i>PGRB</i>	1	TGTGGGTGGTATTTTT AATGAGA	CCCCCTCACTAAAACCCTA AA	GGGATTTGAGATT TT	YGGAGATGATTGTY GTTYGTAGTAYGGA GTTAG	60	
	2	GAGAATTAGTTTTATT TGTTATTTGAGTGA	Biotin- CAACCCATTCCCAAAAAA ATC				
<i>RASSF1A</i>	1	GGGGGAGTTTGAGTT TATTGA	Biotin- CTACCCCTTAACTACCCCTT CC	GGGTYGTATTYG GTTGGAG	YGTGTAAAYGYGTT GYGTATYGYGYGGG GTAT	56	
<i>OCT4</i>	1	GGGTTAGAGGTTAAG GTTAGTG	AAATCCCAAACCAAATATC	GTAAGTTTTTATT TTATTAGG	TTTTYGGTTTGGGG YGTTTTTTTTTTTTAT GGYGGGAT	52	
	2	GGGTTAGAGGTTAAG GTTAGTG	GGGACACCGCTGATCGTTT AAATCCCAAACCAAATATC*				
<i>BCL2L11</i>	1	GATTGGGAGAGGAAG AAAAGTTG	CAAACCCCAAACTAAATT AATCC	GGGAGGAGAGTT TAAAGA	TTYGTTTYGYGTTT TYGYG	56	199
	2	GTAGGAGGAGGGTGT TTGAGTTT	GGGACACCGCTGATCGTTT ACAAACCCCAAACTAAAT TAATCC*				

Gene name	Step	Forward (5'-3')	Reverse (5'-3')	Sequencing (5'-3')	Target (5'-3')	Tm (°C)	Amplicon size (bp)
<i>PACS1</i>	1	AGATGGGTTTAGGGG TAGTTTGA	CCCCCAACCATAAAAATC TA	GGGTAGTAGGGT AGGGTTA	GGTYGGTAGGYGGA GAGTYGGTTTTT YG	56	230
	2	TGTAGGGGTAGTAGG GTAGGGTTA	GGGACACCGCTGATCGTTT ACCCCCAACCATAAAAAT CTA*				
<i>PSEN2</i>	1	TTTLAGGTGGGGTTTT AGTGGA	CCCTACCAACACTCTCTCT CTCT	AGGTGGGGTTTT AGTG	GAYGAGGGAAYGYG GYGTYG	56	367
	2	TTTLAGGTGGGGTTTT AGTGGA	GGGACACCGCTGATCGTTT AACCCCCCTCACCTACT CT*				
<i>TTC9</i>	1	AGGTTGAGGAGGGAG GAGG	ACCCCCAACCCTTTCTCT	TGGAGTAGTTTTT GGTAGTA	GYGGGAGAATGGG AGTGYGGGGYG	56	301
	2	AGGAGGTGGAGTAG TTTTTGGA	GGGACACCGCTGATCGTTT AACCCCCAACCCTTTCTC T*				
<i>BHLHA9</i>	1	TGGGAGGTAAGAGGT TTTTTAAGA	CCCCAACTAAACCCTACA AA	TTGAGGATATTTG GAGTG	TTAYGGTTYGGTYG YGYGYGGG	56	197
	2	TGGGAGGTAAGAGGT TTTTTAAGA	GGGACACCGCTGATCGTTT ACCCAACTAAACCCTAC AAA*				
<i>LRRC56</i>	1	GGGGAGATTTTTTAT TTGGGAAA	CAAACCCCTCAAATACCAA CTTC	TTTTTATTTGGGA AAGGT	GGGYGYGAGTTTTT AGTAGAGATYG	56	159

Gene name	Step	Forward (5'-3')	Reverse (5'-3')	Sequencing (5'-3')	Target (5'-3')	Tm (°C)	Amplicon size (bp)
	2	GGGGAGATTTTTTAT TTGGGAAA	GGGACACCGCTGATCGTTT AACCCCTCAAATACCAACT TCCT*				
<i>OPLAH</i>	1	GGGGTTTTTGAGGGA GAGATT	CCCCAAAATCCAAAATCC	TGATGTTTTTATG TTAGTA	TYGYGTTTTTYGG GTTTTGGATTGGGTY G	60	258
	2	GGGGTTTTTGAGGGA GAGATT	GGGACACCGCTGATCGTTT ACCCCCCAAATCCAAA AT*				
<i>SFMBT1</i>	1	GTGATTGGTTAGGATA AGATGGTA	CTCCAATTCCCAACTATCC TATA	GTTAGGATAAGAT GGTATGA	GYGGAGAAGYGGT YGGATTTTAGAT	54	274
	2	TTGGTTAGGATAAGAT GGTATGAG	GGGACACCGCTGATCGTTT ACTCCAATTCCCAACTATC CTATA*				
<i>NFATC1</i>	1	GGTTTATTATTGGAGA AAATTAGT	TAACAACCCCAAATCCT	AGAAAATTAGTTA GTGAAAG	GGTYGYGGGAGAA GTTYGGGGAYG	60	148
	2	TTGGAGAAAATTAGT TAGTGAA	GGGACACCGCTGATCGTTT ATAACAACCCCAAATCCT T*				
<i>IVNSIABP</i>	1	AATTTATTGGGGTTTT TTATAT	AATTACTACCAAATCCCAA CTT	TTTCACAAATCTC CCTC	CCA/GAACAAAATAA TA	54	116
	2	GGGACACCGCTGATC GTTTATATTGGGGTTT TTTATATTAGGT*	AATTACTACCAAATCCCAA CTT				

Gene name	Step	Forward (5'-3')	Reverse (5'-3')	Sequencing (5'-3')	Target (5'-3')	Tm (°C)	Amplicon size (bp)
<i>SHC2</i>	1	GTGTTTATTTAATGGG TAAAG	TCCAAAAAAACCCTAATC	GTTGTATTTTTTA GGGAGG	TTYGGGGTAGTGT YG	60	185
	2	TGAGGAGAGAGTAGT TTTATTAT	GGGACACCGCTGATCGTTT ATCCAAAAAAACCCTAATC *				
<i>DPYSL5</i>	1	GAAATGTATTTTTTAA AGGTTAGT	AAACCTCCAAACTACAAC	ATCTCAACAAAAT CTACCC	RAACRCAAAACCCR CA	54	296
	2	GGGACACCGCTGATC GTTTAGAAATGTATTT TTAAAGGTTAGT*	CCCATCTCAACAAAATCT				
<i>NANOG</i>	1	GGTTTTTTAATTTATT GGGATTATAGG	CCCAACAACAATACTTCT AAATTCA	GGATTATAGGGGT GGGT	TATYGYGTTYGGTT	56	155
	2	GGTTTTTTAATTTATT GGGATTATAGG	GGGACACCGCTGATCGTTT ACCCAACAACAATACTTC TAAATTCA*				

\*To label single DNA strand with biotin, the reverse or forward primer in the second step PCR contains a universal 5' tail (GGGACACCGCTGATCGTTTA), and a biotinylated universal primer (biotin- GGGACACCGCTGATCGTTTA) was also mixed with the tailed primer at a ratio of 4:1.

**Table 3. PCR primers used in bisulfite-sequencing assay.**

Primer name	Step	Forward (5'-3')	Reverse (5'-3')	Tm (°C)	Amplicon size (bp)
BCL2L11-edge	1	GGGTTGAAAGTTGTTGTTATTAGATG	TTAAAACTTAACTCCCAACTTAAACC	56	277
	2	TTTGATTTGTTTATTGTGTTGTGTT	TTAAAACTTAACTCCCAACTTAAACC		
BCL2L11-center	1	GATTGGGAGAGGAAGAAAAGTTG	CAAACCCCAAACTAAATTAATCC	56	199
	2	GTAGGAGGAGGGTGTGTTGAGTTT	CAAACCCCAAACTAAATTAATCC		
PACS1-edge	1	AGATGGGTTTAGGGTAGTTTGA	CCCCCAACCATAAAAATCTA	56	230
	2	TGTAGGGGTAGTAGGGTAGGGTTA	CCCCCAACCATAAAAATCTA		
PACS1-center	1	GGGGGAAGTTTGGGAGTTAG	AATAACCTAAACCAACTTAAAAAAC	56	225
	2	GGGAAGTTTGGGAGTTAGAT	AATAACCTAAACCAACTTAAAAAAC		
PSEN2-edge	1	AGTGTTTTTTAATGTGAGAATAAT	CCACTAAAACCCACCTAAAATC	56	249
	2	GAGGGGATGTGGATTTAAAATTATAA	CCACTAAAACCCACCTAAAATC		
PSEN2-center	1	TTTTAGGTGGGGTTTTAGTGGA	CCCTACCCAACACTCTCCTCTCT	56	367
	2	TTTTAGGTGGGGTTTTAGTGGA	ACCCCCCCTCACCTACTCT		
TTC9-edge	1	GGATTTTTTGAGGAAGGGTATAGA	CTCCTTTTTATAAATCCAAATTATC	56	265
	2	TTGAGGAAGGGTATAGAATTGTTTT	CTCCTTTTTATAAATCCAAATTATC		
TTC9-center	1	AGGTTGAGGAGGGAGGAGG	ACCCCCAACCCTTTCTCT	56	301
	2	AGGAGGTTGGAGTAGTTTTTGGTA	ACCCCCAACCCTTTCTCT		
KAZN-edge	1	GGTGGGTGTATTTAGTATTTTTTTTATTTA	CCCAAATAACCCCAATACC	60	385
	2	GTATTTTTTTTATTTAGAGGATGGT	CCCAAATAACCCCAATACC		
MUM1-edge	1	TTAGTAGGAGGAGGTTGTAGTGAGTT	ACCCTTTATCCTACAATCAAACC	60	438
	2	GAGGAGGTTGTAGTGAGTTGAGTT	ACCCTTTATCCTACAATCAAACC		
RFX6-edge	1	TTAATAGTGTATTTGAAAGTAGGGATGG	TTACCCAAAACTACACACAACAAT	56	559
	2	GAAAGTAGGGATGGGGTTAGTTATAGT	TTACCCAAAACTACACACAACAAT		
VAX2-edge	1	GTTGGTTTTTTTTAGTAAATGGTGG	TCAACAAAATCAATAACCAAATAC	60	655
	2	TTTTTTTAGTAAATGGTGGGGATT	TCAACAAAATCAATAACCAAATAC		

## **2.7 Protein extraction and western blot assay**

Protein extraction was performed using RIPA buffer (25mM Tris pH 7.6, 150mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS) supplemented with 1× protease inhibitor cocktail solution (Roche). The concentrations of proteins were tested with BCA Protein Assay Reagent (Thermal Scientific). In western blot assay, 25-50 µg aliquot of each protein sample was boiled for 5 min in Laemmli sample buffer (Biorad) supplemented with 5%β-mercaptoethanol and then separated through SDS-PAGE gels (Biorad) with different concentrations depending on the sizes of target proteins. The primary antibodies used included anti-Flag (Stratagene), anti-TET1 (GeneTex), anti-APEX1 (Abcam), anti-Lamin B (Abcam) and anti-ACTB (GeneTex), while secondary antibodies included dperoxidase labelled anti-rabbit and anti-mouse antibodies (GE Healthcare). The bands development was done with Amersham® ECL Plus Western Blotting Detection Reagents (GE Healthcare) and exposed to X-Ray imaging film (Fisher).

## **2.8 DNA dot blot against 5hmC**

1 M NaOH and 200 mM EDTA pH8.2 were added to each sample of genomic DNA to a final concentration of 0.4 mM NaOH/10 mM EDTA. Then DNA was denatured at 100°C for 10 min, followed by rapid chilling on ice. 2 µl denatured DNA was then spotted onto the positively charged nylon membrane (Roche) with regular 20 µl pipet, keeping the diameter of each dot to <4 mm. After the membrane became dry, rinsed it in 2×SSC buffer (0.3 M NaCl, 30 mMNa<sub>3</sub>C<sub>6</sub>H<sub>5</sub>O<sub>7</sub>) followed by complete air dry. Wrap the dry membrane in UV-transparent plastic wrap, and place DNA-side-down on a UV transilluminator for 3 min to immobilize the DNA. Subsequently, the membrane was probed with anti-5hmC antibody (Active Motif) in a same way as that in western blot assay.

## 2.9 RNA extraction, cDNA synthesis, and quantitative PCR

Total RNA from cell lines was isolated with TRIzol® Reagent (Invitrogen) and treated with Turbo DNA-free™ kit (Ambion) according to manufacturer's instructions. First-strand cDNA was synthesized from 2 µg total RNA using the High Capacity cDNA Archive Kit (Applied Biosystems). Power SYBR® Green PCR Master Mix (Applied Biosystems) was used for real-time PCR to quantify gene expression. The primers were designed to span an intron or cross and intron/exon boundary. All reactions were run in triplicate. PCR reaction comprised a 10 min activation step at 95°C, followed by 40 cycles of 95°C for 15 s, and 60°C for 1 min. The results were analyzed by PRISM® 7500 Sequence Detection System (Applied Biosystems) to get threshold cycle (Ct) value for each reaction. The average Ct value of target genes for each sample was subtracted from the average Ct value of endogenous control (GAPDH) to obtain a normalized Ct value ( $\Delta C_t$ ). Finally, the mRNA expression of objective gene was shown by  $2^{\Delta C_t}$  value. The primers used are listed in Table 4.

## 2.10 *HpaII*-digestion DNA methylation assay

500 ng genomic DNA was incubated with 10 units *HpaII* (NEB) or in a mock reaction without *HpaII* at 37°C for 8 hr or overnight, followed by 80°C inactivation for 20 min. The DNA from *HpaII* digestion or mock treatment was tested by real-time PCR with Power SYBR® Green PCR Master Mix (Applied Biosystems) and primers flanking specific *HpaII* digestion sites. The real-time PCR reaction was same to that in “2.9 RNA extraction, cDNA synthesis and quantitative PCR”. Each reaction was performed in triplicate. DNA methylation of a CCGG site was calculated by  $2^{C_t(\text{mock})-C_t(\text{HpaII})} \times 100\%$ . The primers used are listed in Table 5.

**Table 4. PCR primers for RT-qPCR of gene expression analysis**

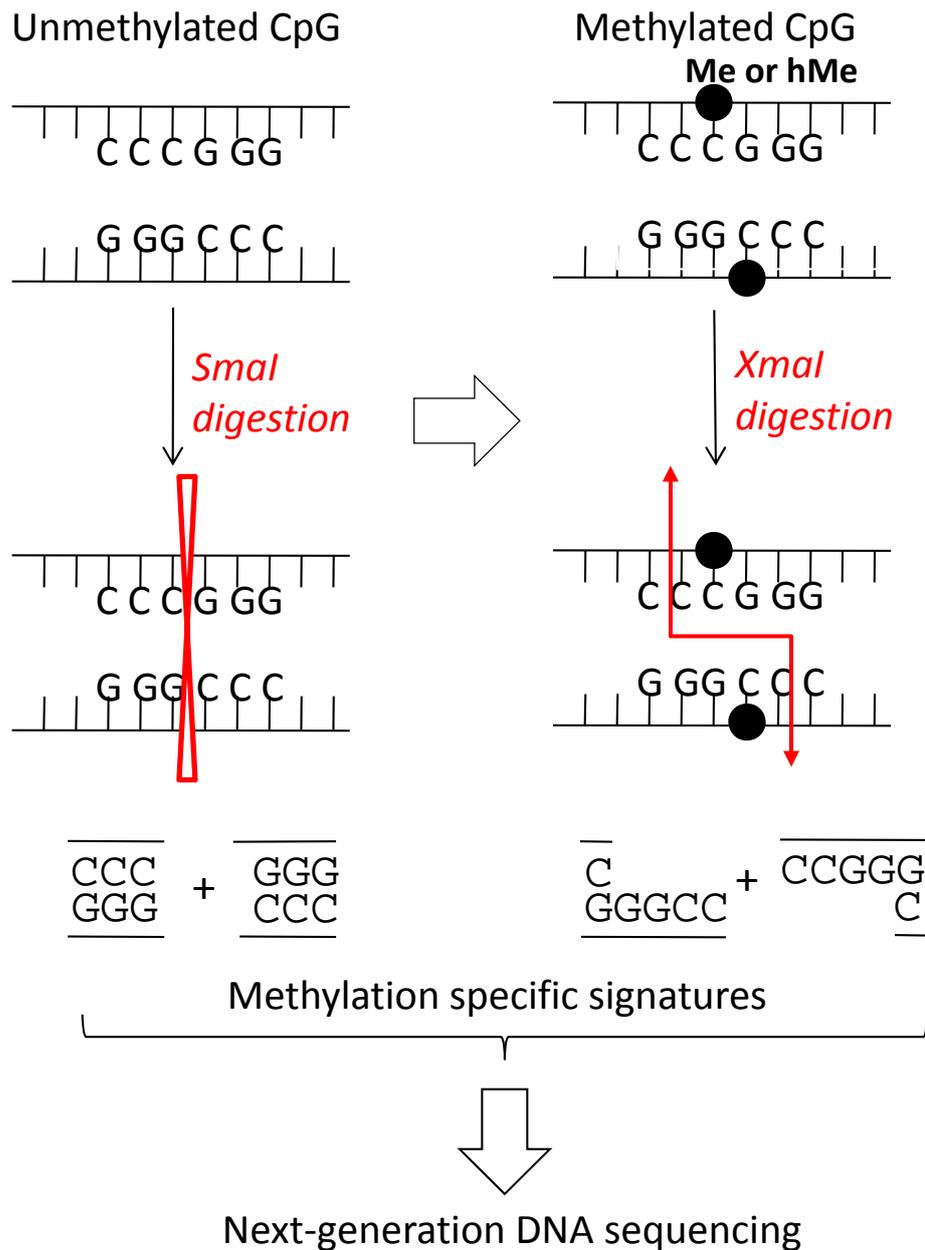
<b>Primer name</b>	<b>Forward (5'-3')</b>	<b>Reverse (5'-3')</b>	<b>Tm (°C)</b>	<b>Amplicon size (bp)</b>
TET1-CD-RT-F/R	GATTCTGAACTGCCACCTG	TCCATGATTCCCTGACAGC	60	116
NANOG-RT-F/R	CAAAGGCAAACAACCCACTT	TCTGCTGGAGGCTGAGGTAT	60	158
OCT4-RT-F/R	AAGCGATCAAGCAGCGACTAT	GGAAAGGGACCGAGGAGTACA	60	127
SOX2-RT-F/R	GCGCCCTGCAGTACAATC	GCTGGCCTCGGACTTGAC	60	140
KLF4-RT-F/R	CCCAATTACCCATCCTTCCT	ACGATCGTCTTCCCCTCTT	60	127
hTERT-RT-F/R	GCCGATTGTGAACATGGACTAC	GTAGTTGAGCACGCTGAACAGT	60	109
CD44-RT-F/R	ACACATATTGCTTCAATGCTTCAG	CAGGATTCGTTCTGTATTCTCCTT	60	153
APEX1-RT-F/R	GCTGCCTGGACTCTCCTC	GCTGTTACCAGCACAAACGA	60	180
GADPH-RT-F/R	TTGTCAAGCTCATTTCCTGGT	CTTACTCCTGGAGGCCATGT	60	90

**Table 5. PCR primers for *HpaII*-digestion methylation assay and ChIP-qPCR**

<b>Primer name</b>	<b>Forward (5'-3')</b>	<b>Reverse (5'-3')</b>	<b>Tm (°C)</b>	<b>Amplicon size (bp)</b>
UBE2B- <i>HpaII</i> -F/R	CTCAGGGGTGGATTGTTGAC	TGTGGATTCAAAGACCACGA	60	177
IVNS1ABP- <i>HpaII</i> -F/R	CTACTGGGGCCCTTTATA	CTAAATCCCTCACCTTCAAT	60	170
NPAS1- <i>HpaII</i> -F/R	ATGACTGACCCAAGTCTCT	AGTTCAGCAAGGCCTAGAG	60	152
PARRES1- <i>HpaII</i> -F/R	GGAAATAAGGCCGTGGTG	GAAACGTGGAAAAGGAGTGG	60	96
BCL2L11-CP-F/R	GCGCAACGTTCTCTCTCACT	GCATGTGCAAAGCAGGTAAA	60	116
PACS1-CP-F/R	GACCGAATCCCGGAAAAG	GGGTCCTGCCCTCAAATC	60	107
PSEN2-CP-F/R	CCCAGTGGACGAGGGAAC	GCTCCAGCGGAGTTTACG	60	122
TTC9-CP-F/R	GAGCGCACGAGTTCAAAAG	CCCGGTGGTATTTGCCTAT	60	81
BHLHA9-CP-F/R	GCGCATCCTAGACTACAACG	CGATCTTGGAGAGCCTCTTG	60	86
LRRC56-CP-F/R	GGTGTGTGTTTCCTGGTCCT	CTGACCTTCGGTTGGAAGTG	60	80
OPLAH-CP-F/R	GGTCCCTCAGCACGAAAGAG	GGTCGAAGGTCGAGCAAAG	60	96
SFMBT1-CP-F/R	GCTCCCTCTGAGACCTGAAA	CTGCGGAAGCCTAGAGACAT	60	96
KAZN-CP-F/R	CGCTGGTGGCAAAGTTCT	ACGCCGAGGAACTCACTCT	60	93
MUM1-CP-F/R	GTCTGACTGCAGGACAAAGG	AGTGCTGCCTGAGGAGACC	60	93
RFX6-CP-F/R	CGGAGCTGGAAGACACCTT	GGAGCTGCACACAGCAGTC	60	84
VAX2-CP-F/R	AGGCTCCAGGGAGAGTGG	CCCTTACCTCGCACCAGTAT	60	90

## 2.11 Digital restriction enzyme analysis of methylation (DREAM)

5 µg genomic DNA from (m)TET1-FL or (m)TET1-CD overexpressed HEK293T cells was first spiked with 0.5 ng of a set of specific calibrators made of PCR-amplified fragments from non-human DNA and in vitro methylated at CpG sites by the M.SssI methylase to 0, 25, 50, 75 and 100% methylation levels. The DNA mixture was then sequentially digested by 5 µl SmaI (3 hours at 37°C, Fermentas) and 50 U XmaI (~16 hours at 37°C, NEB). After purified with QIAquick PCR purification kit (Qiagen), the digested DNA was heated at 65 °C for 3min followed by snap cooling to create free concatenated CCGG overhangs. Klenow fragment (3'→5' minus) (NEB) and CGA mix (dCTP, dGFP, dATP, 10 Mm each) were then added to the DNA in order to fill the overhangs and add "A" tail to 3' end. The resultant DNA was purified again and then ligated with Solexa Paired Ends Adapters (PEA1: 5'phosphate-GATCGGAAG AGCGGTTTCAGCAGGAATGCCGFG-3', PEA2: 5'-ACACTCTTTCCTACACGACGCTCTT-CCGATOT-3') using Quick T4 DNA ligase (Enaymatics). Subsequently, the DNA ligated with adapters was separated through 2% agarose gel. Two gel slices with the size of 250~375bp and 375~500 bp, respectively, were cut off and purified separately with QIAquick Gel Extraction Kit (Qiagen). PCR amplification (18 cycles) of the gel-extracted DNA was performed using Solexa paired-end PCR primers and iProof HF master mix (BioRad). PCR products were further purified with Agencourt AMPure PCR Purification Kit (Beckman Coulter) and tested for concentration with NanoDrop™ 1000 spectrophotometer (Thermo Scientific). Finally, the PCR production from gel slice of 250~375 bp was sent for next generation sequencing at the DNA sequencing core facility at the University of Texas M.D. Anderson Cancer Center. The principle of DREAM is depicted in Figure 5.



**Figure 5. The principle of digital restriction enzyme analysis of methylation (DREAM)**

The genomic DNA is sequentially digested with a pair of enzymes recognizing the same restriction site (CCCGGG) containing a CpG dinucleotides. The first enzyme, *SmaI*, cuts only at unmethylated CpG and leaves blunt ends. The second enzyme, *XmaI* is not blocked by methylation and leaves a short 5' overhang. The enzymes thus create methylation specific signatures at ends of digested DNA fragments. These are then deciphered by next generation DNA sequencing. The methylation level for each sequenced restriction site is calculated based on the numbers of DNA molecules with the methylated or unmethylated signatures:  $\text{methylation \%} = (\text{number of methylated signatures}) / (\text{number of methylated signatures} + \text{number of unmethylated signatures}) \times 100$ .

## **2.12 Hydroxymethylated DNA immunoprecipitation-sequencing**

### **(hMeDIP-Seq)**

3 µg Genomic DNA from (m)TET1-FL or (m)TET1-CD overexpressed HEK293T cells was diluted TE buffer and sonicated with Bioruptor (Diagenode). The desirable fragment size is 100~500 bp. The resultant DNA was purified with QIAquick PCR purification kit (Qiagen) and tested for concentration with NanoDrop™ 1000 spectrophotometer (Thermo Scientific). 500 ng purified sonicated DNA was first spiked with 20 pg of a set of specific calibrators made of PCR-amplified fragments from non-human DNA with different dCTP components, then performed end repair with T4 DNA polymerase, Klenow DNA polymerase and T4 PNK (NEB), added with “A” bases to the 3’ end of the DNA fragments with Kenowexo (3’ to 5’ minus, NEB), and ligated with Solexa Paired ends Adapters to DNA fragments with Quick T4 DNA ligase (Enaymatics). Subsequently, the prepared DNA was purified again, diluted in TE buffer and denatured at 95 °C for 10 min, followed by snap chilling in ice. For immunoprecipitation, the denatured DNA was incubated with 1 µl anti-5hmC antibody (Active Motif) or control rabbit IgG (Millipore) at 4 °C with rotation overnight. These DNA-antibody complexes were then precipitated by Dynal® Protein G magnetic beads (Invitrogen) at 4°C with rotation for 2 hours. After extensive washing and elution at 65 °C for 15 min, DNA was separated from beads. For library size selection, the resultant DNA was separated on 2% agarose gel and gel slice of 275-325 bp was cut off followed by gel extraction with QIAquick Gel Purification Kit (Qiagen). To amplify the gel-extracted DNA, PCR was performed for 10-15 cycles with Solexa paired-end PCR primers and Phusion™ High-Fidelity DNA Polymerase (NEB). The PCR product was purified with Agencourt AMPure PCR Purification Kit (Beckman Coulter), and sent for next generation

DNA sequencing at the DNA sequencing core facility at the University of Texas M.D. Anderson Cancer Center. A flowchart for hMeDIP-Seq is depicted in Figure 6.

### **2.13 ChIP-quantitative PCR (ChIP-qPCR)**

Cells were firstly fixed with fresh 1% formaldehyde at room temperature for 10 min and quenched with 125 mM glycine. The cells then were washed with cold PBS and resuspended in SDS lysis buffer (50 mM Tris pH 8.1, 10mM EDTA, 1% SDS), followed by sonication to get an average fragment size of 200-500 bp. The resultant chromatin samples were diluted by 10 folds in ChIP dilution buffer (16.7 mM Tris pH 8.1, 1.2 mM EDTA, 167 mM NaCl, 1.1% Triton X-100, 0.01% SDS) and incubated with Dynal® Protein G magnetic beads (Invitrogen) overnight at 4°C with rotation for pre-clearing. At the same time, antibodies against proteins of interest or control IgG were pre-crosslinked with Dynal® Protein G magnetic beads at 4°C overnight. After that the cleared chromatin samples were incubated with the antibody-bead complex at 4°C with rotation overnight. The immunoprecipitated chromatin-antibody-bead complexes were then extensively washed with RIPA washing buffer (0.5 M EDTA, 5M LiCl, 1M HEPES-KOH pH 7.6, 10% NP-40, 10% Na-Deoxycholate) and TE buffer containing 50 mM NaCl, resuspended in Elution buffer (50 mM Tris pH 8.1, 10mM EDTA, 1% SDS), and heated at 65 °C for 15 min to separate chromatin from beads. To reverse crosslink, the isolated chromatin was incubated at 65 °C overnight, followed by digestion with RNase A and Proteinase K. Finally, the resultant DNA was purified with Qiaquick PCR Purification Kit (Qiagen).

To test the occupancy of proteins of interest in certain genomic regions, the purified immunoprecipitated DNA and 10% input control DNA were tested by real-time PCR with Power

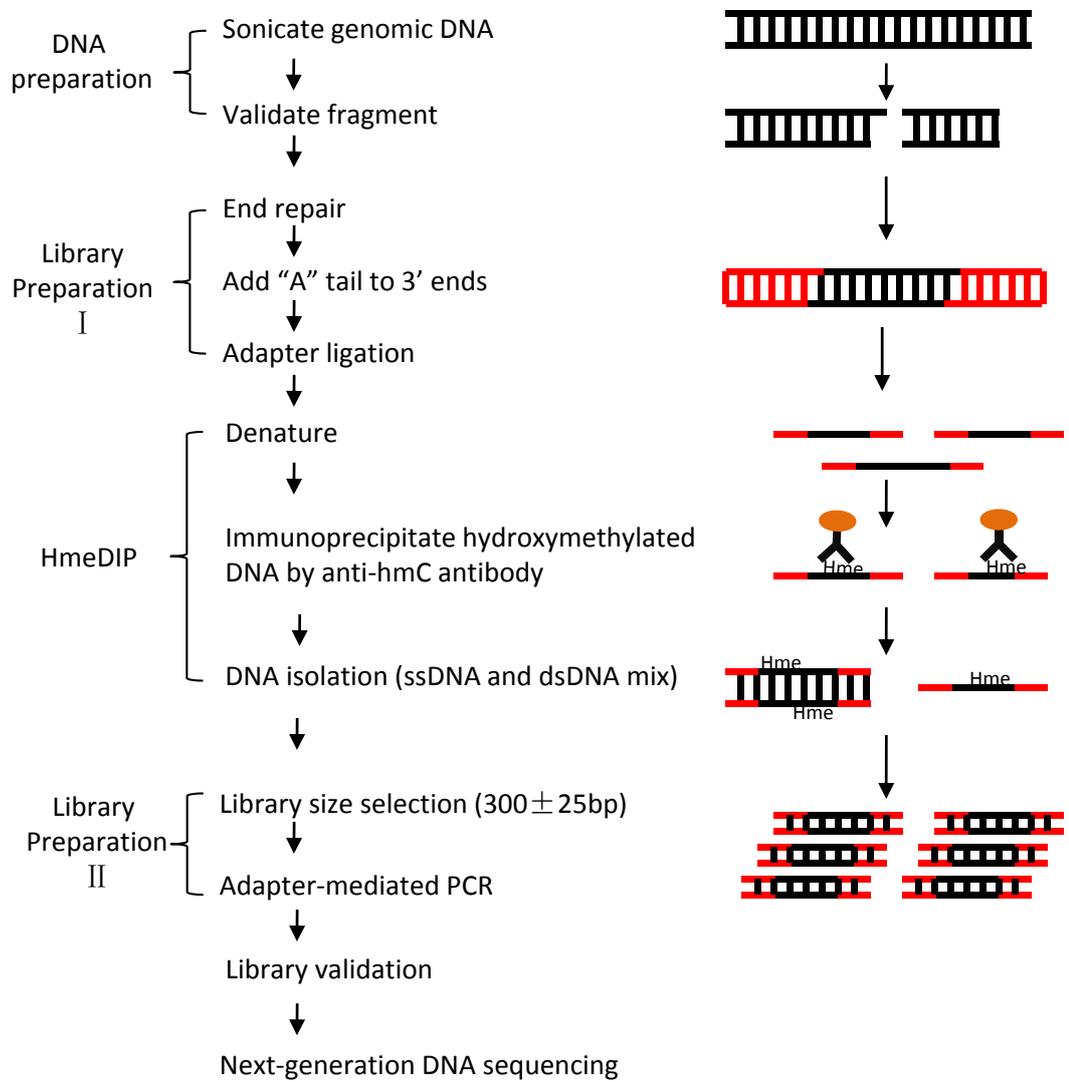
SYBR® Green PCR Master Mix (Applied Biosystems) and primers targeting those specific regions. PCR reaction program was similar to that in “2.9 RNA extraction, cDNA synthesis and quantitative PCR”. The primers used are listed in Table 5.

## **2.14 Lentiviral shRNA-mediated TET1 knockdown**

Four different TET1 shRNA in pTRIPZ vectors (OpenBiosystems) were transferred into *MulI* and *XhoI* sites of the pGIPZ vectors (OpenBiosystems). To produce lentiviral particles, pGIPZ-shTET1 and package plasmids psPAX2 and pMD2.G (Addgene) were transfected into HEK293FT cells using Lipofectamine® 2000 Transfection Reagents (Invitrogen). Two days after transfection, the viral medium was collected and filtered with 0.45 µm PVDF filters (Millipore). The filtered viral medium was added with polybrene (Sigma) at a final concentration of 8 µg/ml, and then transduced target cells-HEK293T cells. Cultured HEK293T cells for two days after transduction, then splitted them into new dishes at an appropriate density (~30% confluence) and began puromycin selection (1.5 µg/ml, Sigma). To get optimal knockdown efficiency, single cell cloning was done by serial dilution. All sequences of shRNA used see Table 6.

## **2.15 siRNA knockdown**

Two different siRNA for human APEX1 and one non-silencing control siRNA were bought from Sigma (Table 6). For transfection, 75 pmol siRNA was diluted in 250 µl of OPTI-MEM® Reduced Serum Medium (Invitrogen), and then mixed with 4 µl Lipofectamine™ RNAiMAX Transfection Reagent (Invitrogen). To improve knockdown efficiency, a second transfection was performed 3 days after the first transfection.



**Figure 6. The flow chart for hydroxymethylated DNA immunoprecipiation-sequencing (hMeDIP-Seq)**

The genomic DNA is fragmented by sonication, and then adapters are ligated to the ends of the resultant DNA fragments. Through an immunoprecipitation process with the use of anti-hmC antibody, hmC-containing DNA is specifically isolated and in turn amplified by adaptor-mediated PCR. Those PCR products are then tested by next generation DNA sequencing on the Illumina platform, and finally aligned to genomic DNA.

**Table 6. ShRNA and siRNA used for knockdown experiment**

<b>Name</b>	<b>Sequence (5'-3')</b>	<b>Oligo/siRNA ID</b>	<b>Company</b>
shControl	CTCGCTTGGGCGAGAGTAAG	RHS4743	OpenBiosystems
shTET1#1	CGAATCAAGCGGAAGAATA	V2THS_141063	OpenBiosystems
shTET1#2	CTGAGAATATACCAAGTAA	V2THS_201600	OpenBiosystems
shTET1#3	CTTCGATAATTAAGATCAA	V2THS_202091	OpenBiosystems
shTET1#4	CTTTGCTAGTGCAGTGTAT	V2THS_203196	OpenBiosystems
siControl	/	SIC001	Sigma
siAPEX1#1	/	SASI_Hs01_00122789	Sigma
siAPEX1#2	/	SASI_Hs01_00122792	Sigma

## **2.16 Cell growth curve**

$3 \times 10^4$  HEK293T cells were seeded in 6-well culture plate in DMEM containing 10% FBS and 100  $\mu\text{g/ml}$  streptomycin-penicillin. Culture medium was replaced every two days. The cell numbers were counted daily for total four days with the use of Z2 Cell and Particle Counter (Beckman Coulter).

## **2.17 Tetracycline-induced TET1-CD overexpression**

pTRIPZ is specifically designed for inducible lentiviral shRNA-mediated knockdown (OpenBiosystems). Tetracycline-induced TET1-CD overexpression plasmid was conveniently constructed based on non-silencing pTRIPZ-shRNA plasmid by replacing its GFP cDNA fragment with (m)TET1-CD-Flag ORF. The procedures for viral particle production, transduction of HEK293T cells were similar to shRNA-mediated TET1 knockdown protocol mentioned previously. HEK293T cells transduced were then selected with puromycin (1.5  $\mu\text{g/ml}$ , Sigma). Finally, single cell cloning was done by serial dilution.

## **2.18 BrdU-containing DNA immunoprecipitation**

To distinguish nascent DNA from parent DNA, 50 $\mu\text{M}$  BrdU (BD Pharmingen) was added into HEK293T cell culture for 24 hour. To isolate BrdU-containing nascent DNA, a BrdU-containing DNA-IP was done with a similar manner to that in hMeDIP mentioned above. In brief, sonicated genomic DNA was first spiked with two specific calibrators made of PCR-amplified fragments from non-human DNA with or without BrdUTP (eEnzyme). BrdU-containing DNA fragments were then immunoprecipitated by anti-BrdU antibody (Sigma) and Dynabeads® M-280 sheep anti-mouse IgG (Invitrogen). The isolated DNA was finally analyzed with bisulfite-

pyrosequencing assay as in “2.6 DNA extraction and bisulfite-pyrosequencing/sequencing”.

## **2.19 Data processing and statistics**

All next generation DNA sequencing datasets were received after pre-analysis by Drs. Shoudan Liang and Yue Lu in the Center for Cancer Epigenetics and DNA sequencing core facility in M.D. Anderson Cancer Center. Statistical analyses were performed by student's t-test and Wilcoxon paired signed-rank test. Excel and Graphpad Prism were major softwares used in data processing and statistical analysis.

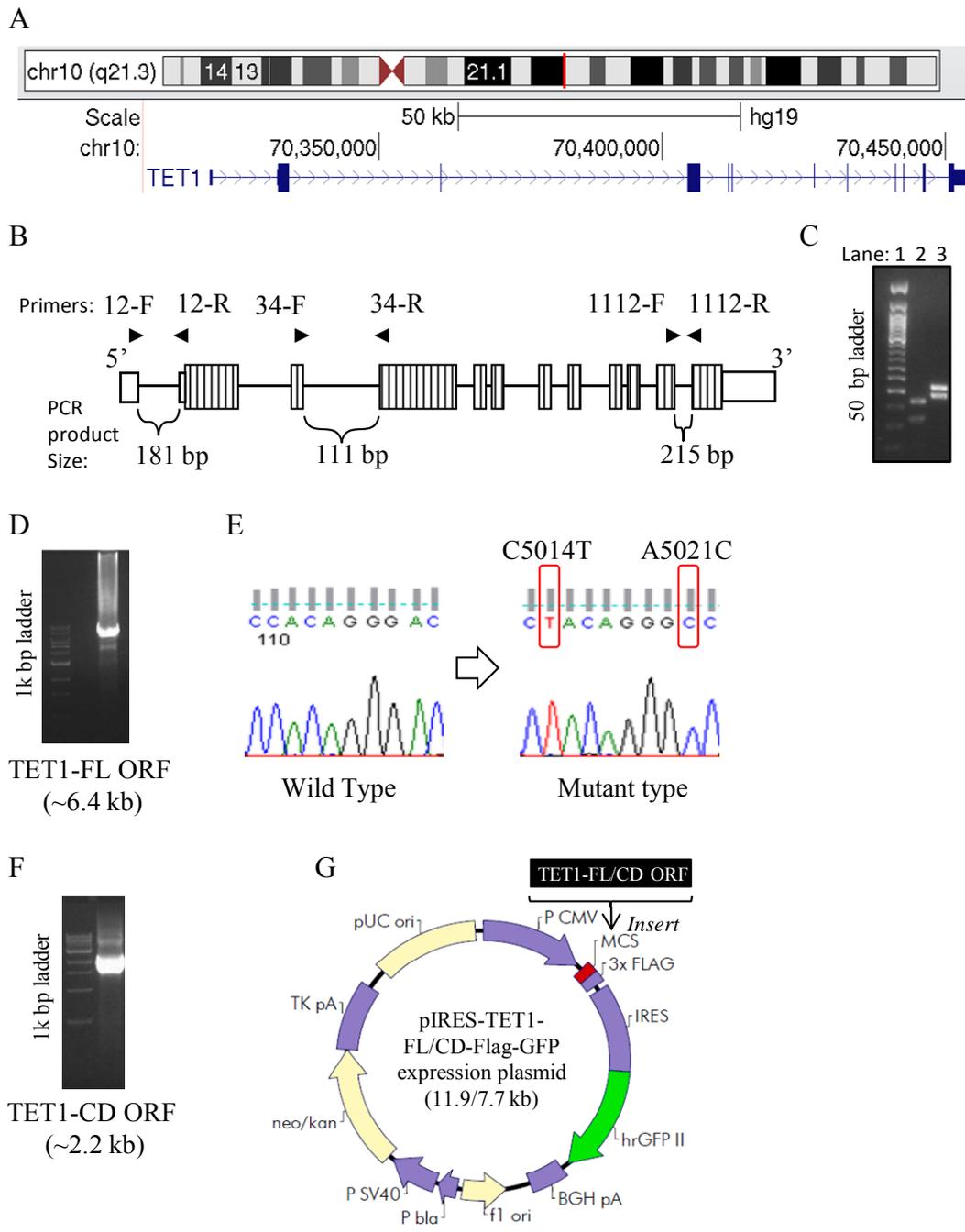
## **CHAPTER 3 RESULTS**

### **3.1 Different regulation of DNA methylation by TET1-FL and TET1-CD overexpression**

#### **3.1.1 TET1 ORF Cloning and overexpression plasmids construction**

Human TET1 gene locates in chromosome 10q21 (Chr10:70,320,413-70,454,239) and has only one annotated transcript (ENST00000373644), which comprises 12 exons and encodes 2136 amino acids (Figure 7A). We cloned TET1-FL ORF from human neuroblastoma cell line SY5Y, as this cell line has been reported to express normal TET1 protein (Tahiliani et al., 2009). Considering the long length of TET1-FL ORF, cDNA synthesis was performed using AccuScript PfuUltra II RT-PCR Kit with either oligo-dT or gene-specific primer, which targets the 3'untranslated-region (UTR) of TET1 mRNA. Once the reverse transcription reactions were finished, we first checked the completeness of TET1 cDNA by PCRs with three primer pairs targeting 1<sup>st</sup> and 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup>, and 11<sup>th</sup> and 12<sup>th</sup> exons of TET1, respectively (Figure 7B). As Figure 7C shows, the PCR products with all three primer pairs showed bands of right size in DNA electrophoresis, proving the success of our TET1 cDNA synthesis. Moreover, both oligo-dT and gene-specific primer worked well in the reverse transcription reactions.

Next we amplified TET1-FL ORF with above TET1 cDNA (synthesized with gene-specific primer) using AccuPrime™ Taq High Fidelity DNA Polymerase. After optimization of thermal cycling conditions, we successfully got a PRC product with the right size of 6.4 kb (Figure 7D). To accommodate such big insert, pCR®-XL-TOPO® vector instead of regular cloning vector (e.g. pCR®4-TOPO®) was used and generated considerable single clones. By using restriction digestion assay (*KpnI* and *AvrII* enzyme), four right clones was selected and sent to



**Figure 7. Construction of expression plasmids encoding TET1-FL or TET1-CD**

(A) Location of TET1 gene in genomic DNA and exons of TET1. (B-C) PCR analysis of the completeness of TET1 cDNA synthesized by reverse transcription with oligo-dT primer (C, lane 2) or TET1-GSP (gene specific primer) (C, lane 3). To easily test the completeness of synthesized TET1 cDNA, different intron-spanning PCR primers were designed (B). Primers TET1-EX34-F/R and ACTB EX12-F/R (not shown here) were used by PCR on lane 2, while TET1-EX12-F/R and TET1-EX1112-F/R were used on lane 3. (D) Amplification PCR of TET1-FL ORF from the confirmed full length TET1-cDNA. (E) DNA sequencing confirms two point mutations (C5014T, A5021C) in catalytically mutant TET1-FL (H1672Y, D1674A). (F) Amplification PCR of TET1-CD ORF from the confirmed TET1-FL ORF plasmids. (G) Plasmid map for pIRES-TET1-FL-Flag-GFP and pIRES-TET1-CD-Flag-GFP.

bi-directional DNA sequencing. As expected, the sequencing results showed each clones have 9 ~ 16 errors, which most possibly came from the reverse transcription reaction. Except 1/3 of total errors occurring at wobble position of genetic codes, the other errors were repaired by re-combination among those four clones and re-cloning of error-containing short regions from original TET1 cDNA. Finally, we successfully got a clone containing completely right sequence of TET1-FL ORF.

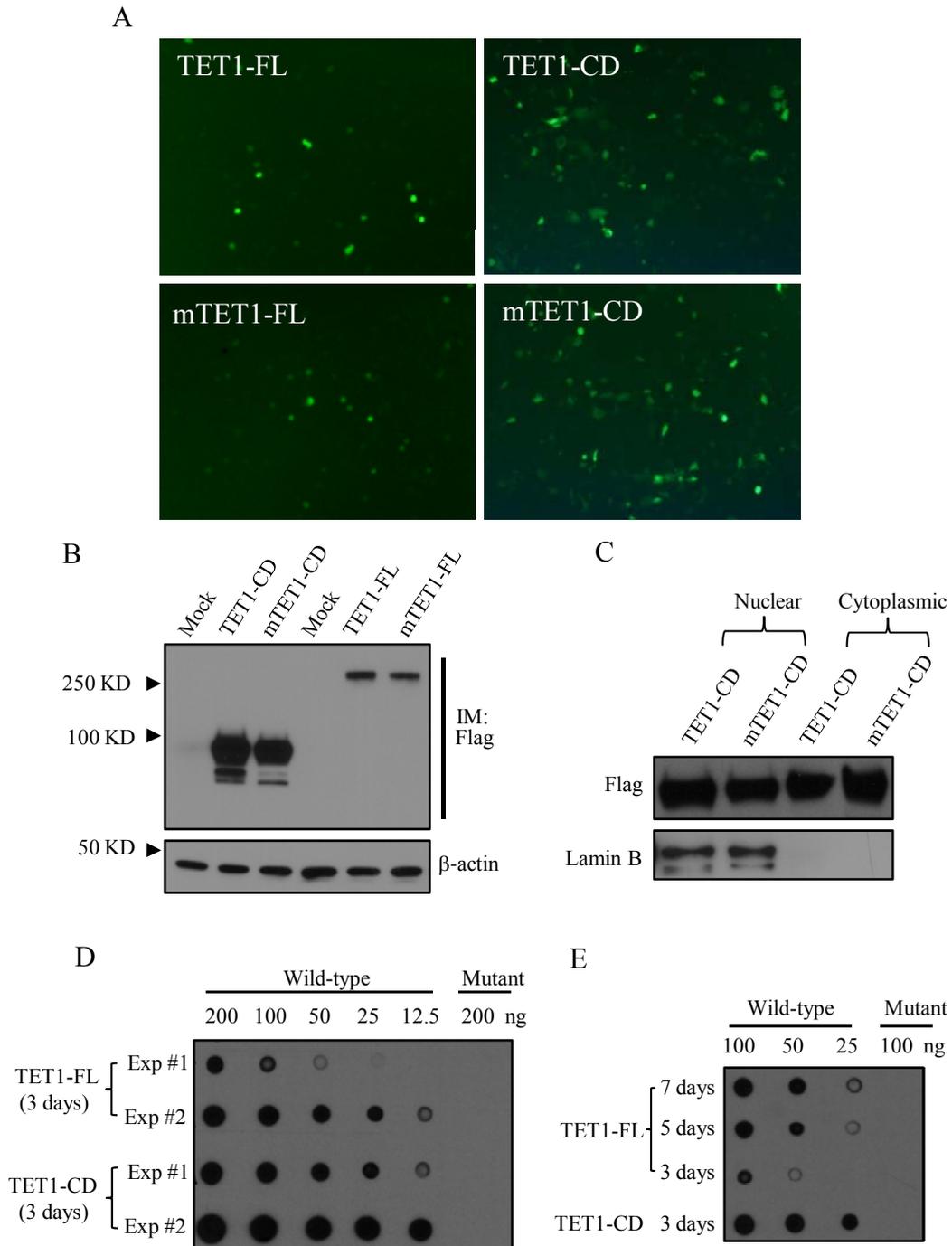
To construct catalytically mutant TET1-FL (mTET1-FL) ORF, we next introduced two point mutations (H1672Y, D1674A) (Tahiliani et al., 2009) into the available wild type TET1-FL ORF using our homemade site-directed mutagenesis kit (Figure 7E). With both clones of TET1-FL ORF and mTET1-FL ORF, we then conveniently constructed two clones containing TET1-CD ORF and mTET1-CD ORF (Figure 7F), respectively. To construct expression plasmids, we transferred those four different ORFs from pCR®-XL-TOPO® vector to the EcoRI and NotI sites of pIRES-hrGFP vector which contains a 3×Flag tag and GFP reporter (Figure 7G). Thus, we finally got four different expression plasmids which were labeled as pIRES-TET1-FL-Flag-GFP, pIRES-mTET1-FL-Flag-GFP, pIRES-TET1-CD-Flag-GFP and pIRES-mTET1-CD-Flag-GFP, respectively.

### **3.1.2 Overexpression of TET1-FL and TET1-CD in HEK293T cells**

To overexpress (m)TET1-FL and (m)TET1-CD in HEK293T cells, we next transiently transfected the cells with those above four expression plasmids, followed by FACS to collect GFP positive cells 3 days after transfection (Figure 8A). Western blot assay with anti-Flag antibody confirmed the overexpression of all four transgenes (Figure 8B). Compared with (m)TET1-CD,

(m)TET1-FL showed a much lower overexpression level, which may be caused by the low transgene copy number and low transcriptional activity due to their much bigger size. Given (m)TET1-CD has two less nuclear localization signals (NLS) than (m)TET1-FL, we also tested whether the ectopically expressed (m)TET1-CD proteins could be efficiently transferred to nuclei. By separately isolating nuclear and cytoplasmic proteins, we found that (m)TET1-CD proteins were enriched in both fractions, indicating that the only NLS in DSBH domain is enough for the nuclear localization of (m)TET1-CD (Figure 8C).

We next studied the catalytic functions of these overexpressed TET1 proteins. By using DNA dot-blot assay with specific anti-5hmC antibody, a dramatic production of 5hmC was expectedly detected in the genomic DNA from GFP positive cells transfected TET1-FL or TET1-CD, but not those transfected with mTET1-FL or mTET1-CD (Figure 8D). Compared with TET1-CD, TET1-FL overexpression resulted in a much lower 5hmC production. Considering TET1-FL transfection yielded a lower level of transgene expression than TET1-CD (Figure 8B), we then asked whether prolonging the culture time of TET1-FL-overexpressed cells can achieve a comparable 5hmC level as that by TET1-CD overexpression. The 5hmC level by TET1-FL overexpression significantly increased when cells were cultured for up to 7 days after transfection, but it was still much lower than that by TET1-CD, suggesting that other reasons than the lower level of TET1-FL expression may mainly explain this (Figure 8E). Actually, because of its smaller size and lack of potential DNA binding domain (i.e. CXXC domain) TET1-CD may be able to access more extensive genomic regions (e.g. tightly packed heterochromatic regions), which essentially leads to its high production of 5hmC. Taken together, the marked production of 5hmC by TET1-FL and TET1-CD but not their catalytically mutant controls functionally validates the



**Figure 8. Both overexpression of TET1-CD and TET1-FL produce 5hmC in genomic DNA**

(A) Fluorescent images of HEK293T cells 2 days after wild-type and catalytically mutant TET1-CD (TET1-CD, mTET1-CD) or TET1-FL (TET1-FL, mTET1-FL) transfection. (B) Western blot analysis of overexpression of Flag-tagged (m)TET1-FL and (m)TET1-CD 3 days after transfection. (C) Presence of (m)TET1-CD overexpression in both nuclear and cytoplasmic fractions. (D-E) DNA dot-blot analysis of genomic 5hmC levels in HEK293T cells overexpressing wild-type and catalytically mutant TET1-CD or TET1-FL 3, 5, or 7 days after transfection.

model of (m)TET1-FL and (m)TET1-CD overexpression we constructed and thus allows our subsequent studies on the potential demethylating effects of TET1-FL and TET1-CD.

### **3.1.3 Overexpression of TET1-CD but not TET1-FL induces DNA demethylation in some genomic loci**

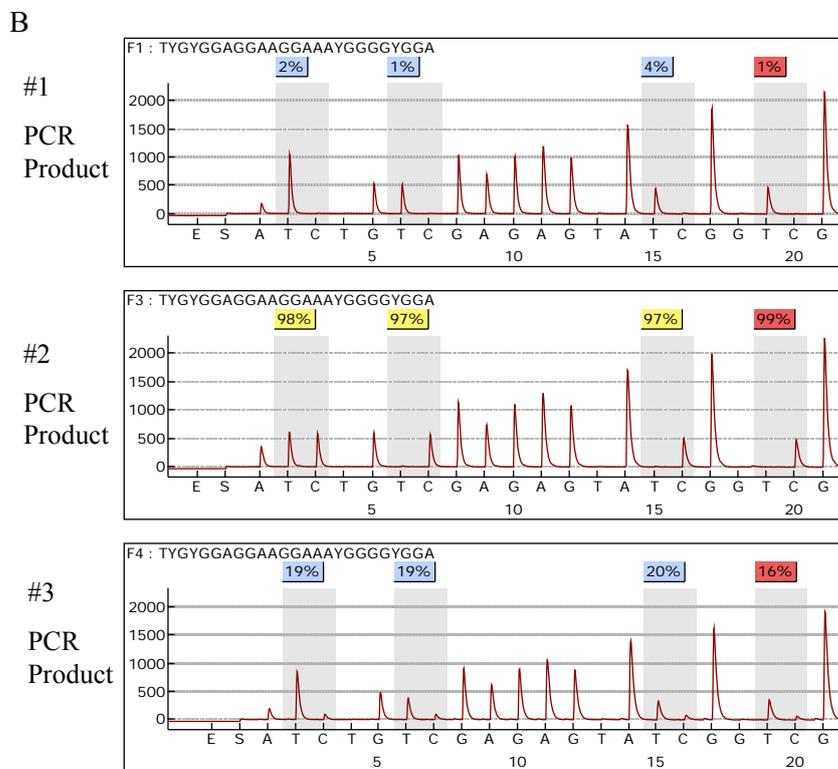
We next examined whether TET1-FL-mediated 5mC oxidation further leads to DNA demethylation in genomic DNA. Bisulfite-sequencing has been extensively used for quantitative analysis of DNA methylation in specific CpG sites. In this assay 5mC is distinguished from unmethylated cytosine relies on the fact that reaction with sodium bisulfite promotes deamination of unmethylated cytosine but not 5mC to yield U. 5hmC, the primary product of TET-mediated 5mC oxidation, has been reported to react with bisulfite to yield cytosine 5-methylenesulfonate which is also resistant to deamination (Hayatsu et al., 1970), suggesting that 5hmC may be not distinguished from 5mC by bisulfite-sequencing. To confirm it, we synthesized two different oligonucleotides using PCR with either dCTP or 5hm-dCTP and then tested them with bisulfite-pyrosequencing. As shown in Figure 9, 5hmC was indistinguishable from 5mC in bisulfite-pyrosequencing. Actually, a similar result has also been acquired by a recent study from Rao group (Huang et al., 2010). Thus, the DNA methylation levels for 5hmC-modified CpGs can still be accurately analyzed by bisulfite-sequencing. Actually, a more recent study further reported that 5fC also behaviors as 5mC in bisulfate conversion reaction, but 5caC acts as unmodified cytosine (He et al., 2011). Therefore, considering that 5mC oxidation reactions by TET proteins also produce some 5caC, TET-mediated DNA demethylation extent may be overvalued to certain extent in bisulfite-based DNA methylation analysis. But even so, we found only overexpression of

- A
- PCR Template: one fraction of p16 gene promoter
- ```

acaccaaacacccccgattcaatttgcagtttaggaaggt
tgtatcgcgagggaaggaaacggggcgggggcggatttc
ttttaacagagtgaacgcactcaaacacgcctttgctg
gcagggcggggagcgcgactgagagcagggagggcggg
  
```
- ↓
- Generate different substracts using PCR amplification with different dNTP mixtures:

|    | dATP, dGTP, dTTP | dCTP  | 5hm-dCTP |
|----|------------------|-------|----------|
| #1 | 10 mM each       | 10 mM | 0 mM     |
| #2 | 10 mM each       | 0 mM  | 10 mM    |
| #3 | 10 mM each       | 5 mM  | 5 mM     |

- ↓
- Bisulfite-pyrosequencing



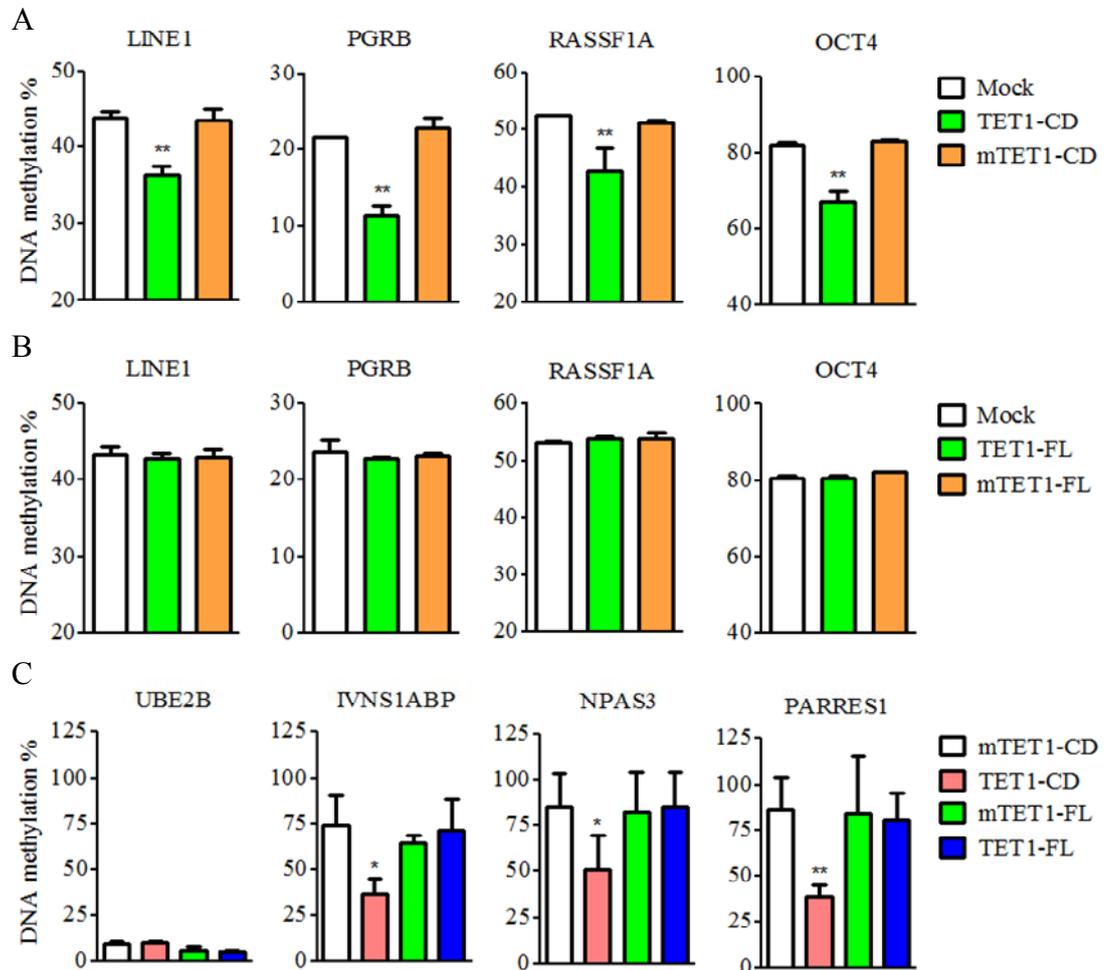
**Figure 9. 5hmC behaviors in a same pattern as 5mC in bisulfite-pyrosequencing**

(A) Generation of different oligonucleotides containing different modified cytosine by PCR with different dCTP component. As shown in the table, only dCTP was used in #1 PCR; only 5hm-dCTP was used in #2 PCR; and both dCTP and 5hm-dCTP (1:1) were used in #3 PCR. The PCR template is from human *p16* gene promoter. The PCR primer targeted sequences are underlined. And the four CpG dinucleotides in red were finally tested for DNA methylation levels by bisulfite-pyrosequencing. (B) Example graphic result of bisulfite-pyrosequencing for oligonucleotides from #1, #2 and #3 PCR, respectively.

TET1-CD but not TET1-FL was observed to induce significant DNA demethylation at some randomly selected genomic loci, including long interspersed non-repetitive element-1 (*LINE1*) and promoters of *RASSF1A*, *OCT4* and *PGRB* genes (Figure 10A, 10B). We also used *HpaII* enzyme-based DNA methylation assay. Since *HpaII* digestion is blocked by 5mC and its all oxidative derivatives (He et al., 2011), this method offers a relatively accurate way to detect DNA demethylation induced by TET. Consistent with the bisulfite-pyrosequencing results, only overexpression of TET1-CD showed a significant decreased of DNA methylation in the promoters of *IVNSIABP*, *NPAS3* and *PARRES1* (Figure 10C). Taken together, TET1-FL exhibited a different functional pattern from TET1-CD in regard of DNA demethylation, which suggests that previously reported demethylating effect of TET1-CD cannot be simply explained as that of TET1.

#### **3.1.4 TET1-CD overexpression induces global DNA demethylation without distribution bias**

The results from other laboratories and ours have showed that overexpression of TET1-CD can induce DNA demethylation in some endogenous genomic loci (Guo et al., 2011; Zhang et al., 2010a). To further characterize the effect of TET1-CD on DNA methylation, we next globally examined DNA methylation changes in HEK293T cells overexpressing (m)TET1-CD. With the use of DREAM, the DNA methylation levels of 34,322 and 33,395 CpG sites were quantified in cells transfected with TET1-CD and mTET1-CD, respectively (Table 7). Pair-wise comparison of 32,803 common CpGs between TET1-CD and mTET1-CD transfection revealed that overexpression of TET1-CD is capable to induce a pronounced global DNA demethylation (Figure 11A). It is notable that 2957 CpGs (~9.0% of total common CpGs) were demethylated by



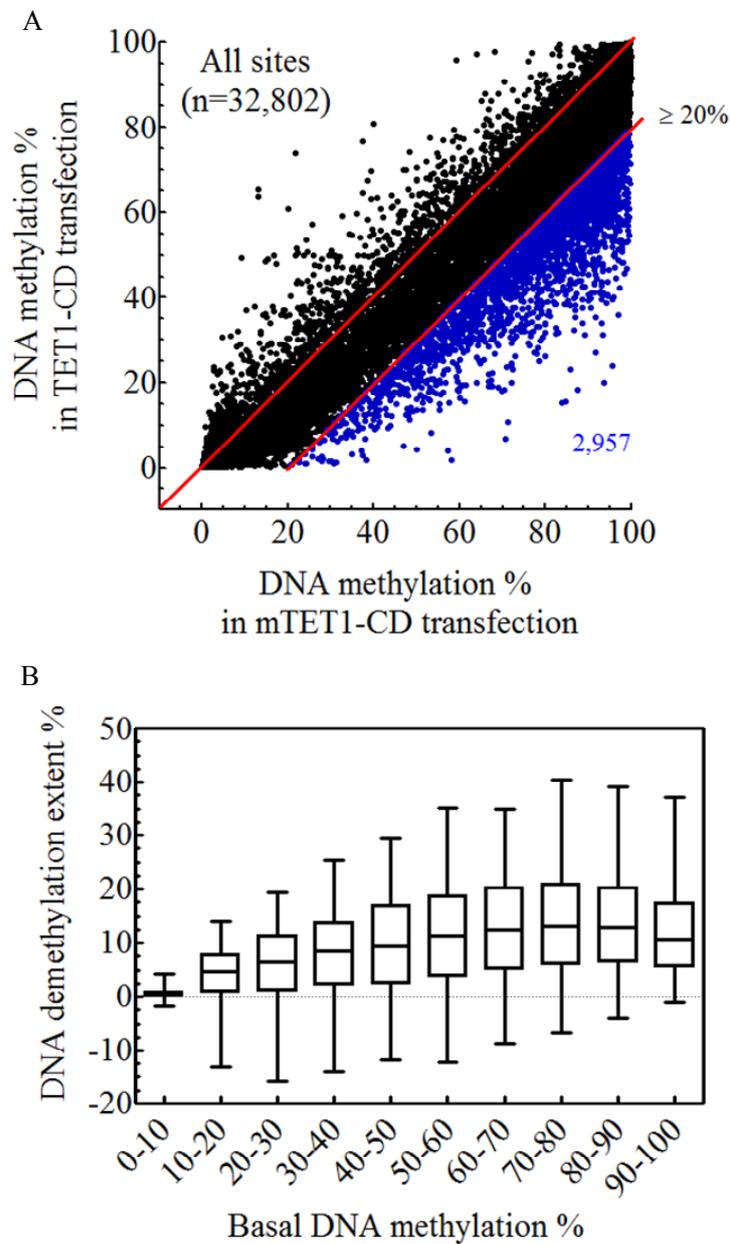
**Figure 10. DNA demethylation of some genomic loci induced by TET1-CD but not TET1-FL overexpression**

(A-B) Bisulfite-pyrosequencing analysis of DNA methylation levels of *LINE1*, *RASSF1A*, *OCT4* and *PGRB* genes 3 or 7 days after (m)TET1-CD or (m)TET1-FL transfection, respectively. (C) *HpaII* sensitivity assay of *UBE2B*, *IVNS1ABP*, *NPAS3* and *PARRES1* genes 3 or 7 days after (m)TET1-CD or (m)TET1-FL transfection, respectively. The hypomethylated *UBE2B* gene was used as mark for complete digestion of *HpaII*. Error bars represent SD of 2-3 independent experiments. \*  $p < 0.05$ , \*\*  $p < 0.01$  by student's t-test compared to corresponding mutant control.

**Table 7. Summary of tags and CpG sites detected in DREAM**

| <b>Transfection</b> | <b>Total number of mapped tags</b> | <b>Number of methylated tags</b> | <b>Number of unmethylated tags</b> | <b>Total number of CpG sites covered</b> | <b>Number of common CpG sites</b> |
|---------------------|------------------------------------|----------------------------------|------------------------------------|------------------------------------------|-----------------------------------|
| TET1-CD             | 16,020,800                         | 3,746,398<br>(23.4%)             | 12,274,402<br>(76.6%)              | 34,322                                   | 32,802*                           |
| mTET1-CD            | 12,088,163                         | 3,619,375<br>(30.0%)             | 8,468,788<br>(70.0%)               | 33,395                                   |                                   |
| TET1-FL             | 32,634,341                         | 13,244,935<br>(40.6%)            | 19,389,406<br>(59.4%)              | 42,988                                   | 40,937&                           |
| mTET1-FL            | 35,025,843                         | 13,433,683<br>(38.4%)            | 21,592,160<br>(61.6%)              | 43,936                                   |                                   |

Cutoff:  $\geq 20$  tags. \* Common covered CpG sites between TET1-CD and mTET1-CD transfection. & Common covered CpG sites between TET1-FL and mTET1-FL transfection.



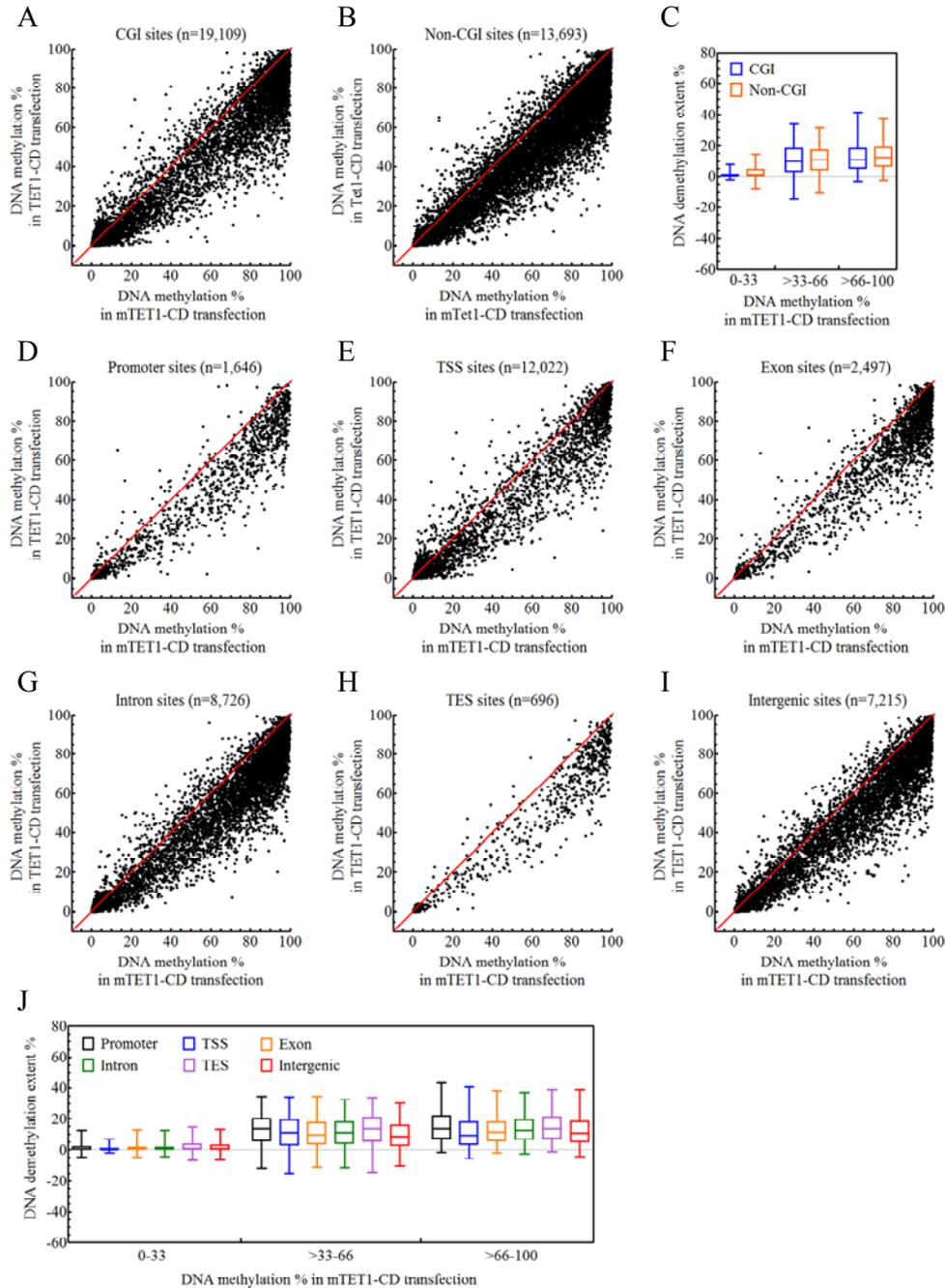
**Figure 11. Global DNA demethylation induced by TET1-CD overexpression in HEK293T cells**

(A) Pair-wise comparison of DNA methylation levels at 32,803 common CpGs between TET1-CD and mTET1-CD transfections. Blue dots and number represent the CpG sites with demethylation extent of  $\geq 20\%$ . (B) Boxplot shows demethylation efficiency indicated by demethylation extent in HEK293T cells overexpressing TET1-CD as a function of basal methylation level. The methylation levels are grouped in 10% intervals from 0-100% methylation. Boxes mark the interquartile range, and whiskers the 2.5th and 97.5th percentiles.

more than 20% (Figure 11A). We next examined the demethylation efficiency of TET1-CD by calculating demethylation extent for each CpG and grouping CpGs by their basal methylation levels (i.e. methylation level in mTET1-CD-overexpressed cells) in 10 intervals from 10% to 100% methylation. Interestingly, TET1-CD-induced demethylation occurred at CpGs of all methylation levels, and the demethylation efficiency was generally positively correlated with the basal methylation level (Figure 11B). Moreover, to determine whether TET1-CD-induced DNA demethylation has distributional bias, we also compared its demethylation efficiency between CpG islands (CGIs) and non-CGIs, as well as among promoters, transcription start sites (TSSs), exons, introns, transcription end sites (TESs) and intergenic regions. Interestingly, we clearly found that TET1-CD-induced DNA demethylation was not biased toward any certain regions (Figure 12). Therefore, TET1-CD overexpression really induced a genome-wide DNA demethylation in HEK293T cells, suggesting that TET proteins-mediated 5mC oxidation could serve as an efficient pathway for global DNA demethylation in mammalian cells.

### **3.1.5 TET1-FL overexpression fails to induce significant DNA demethylation in all categorized genomic regions**

Despite a significant 5hmC production, TET1-FL overexpression still failed to induce DNA demethylation in randomly selected genomic loci where TET1-CD induced. Moreover, considering the presence of CXXC domain, we proposed that TET1-FL may specifically function in certain genomic regions in contrast to the extensive accessibility of TET1-CD. Thus, we next also performed DREAM to analyze the genome-wide effect of TET1-FL overexpression on DNA methylation. The DNA methylation levels of 43,936 and 42,988 CpG were quantified in cells transfected with TET1-FL and mTET1-FL, respectively (Table 7). Different from that between



**Figure 12. TET1-CD overexpression induces DNA demethylation evenly in all categorized genomic regions**

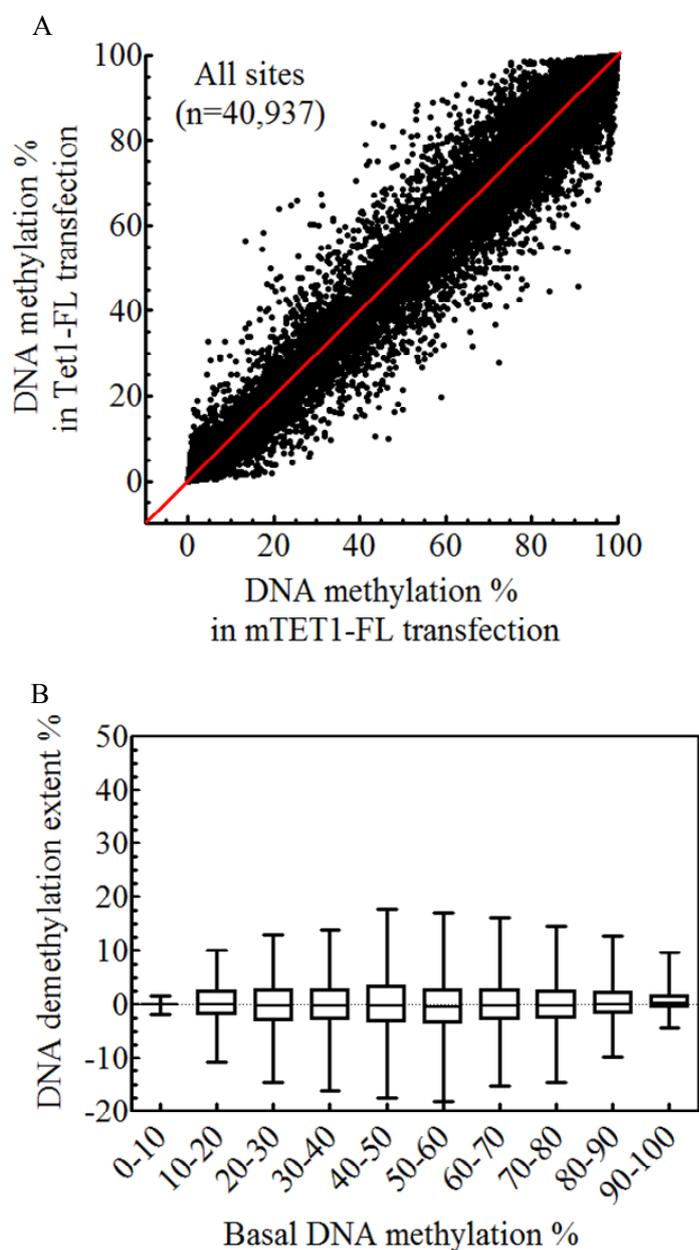
(A and B, D-I) DNA methylation change in CGI (A), non-CGI (B), promoter (D), TSS (E), exon (F), intron (G), TES (H) and intergenic region (I) after overexpression of TET1-CD. The CGI annotation is obtained from the UCSC website. Promoter: -5kb to -1kb relative to TSS; TSS: -1kb to 0.5kb relative to TSS; TES: -0.5kb to 1kb relative to TES; intergenic: 1kb from TES to -5kb of downstream gene. (C, J) Boxplots show demethylation efficiency of TET1-CD in CGIs and non-CGIs (C), and promoter, TSS, exon, intron, TES and intergenic region (J). The efficiency is indicated by demethylation extent as a function of basal methylation level, which is grouped in 33% intervals from 0-100% methylation. Boxes mark the interquartile range, and whiskers the 2.5th and 97.5th percentiles.

TET1-CD and mTET1-CD, pair-wise comparison of all 40,937 common CpGs between TET1-FL and mTET1-FL transfection did not show a significant DNA demethylating effect for TET1-FL (Figure 13A). Consistently, the analysis of DNA demethylation extent as a function of basal methylation level also failed to detect a significant demethylation efficiency of TET1-FL at all methylation levels (Figure 13B). Given Tet1 specifically binds CpG-rich regions in mESC (Wu et al., 2011; Xu et al., 2011), we then examined whether TET1-FL overexpression selectively induces DNA demethylation in CGI. However, by separately analyzing DNA methylation changes in CGIs and non-CGIs, we still failed to find significant DNA demethylation induced by TET1-FL overexpression in either region (Figure 14A-C). Moreover, the deep analysis of DNA methylation changes in promoters, TSSs, exons, introns, TESs and intergenic regions further excluded the possibility that TET1-FL may specifically induce significant DNA demethylation in certain regions (Figure 14C-J). Therefore, we concluded that unlike that of TET1-CD, overexpression of TET1-FL cannot induce significant DNA demethylation in any categorized genomic regions, suggesting that TET1 may be at least not an efficient DNA demethylase as previously predicted.

### **3.1.6 Differential regulation of 5hmC distribution pattern by TET1-FL and TET1-CD overexpression**

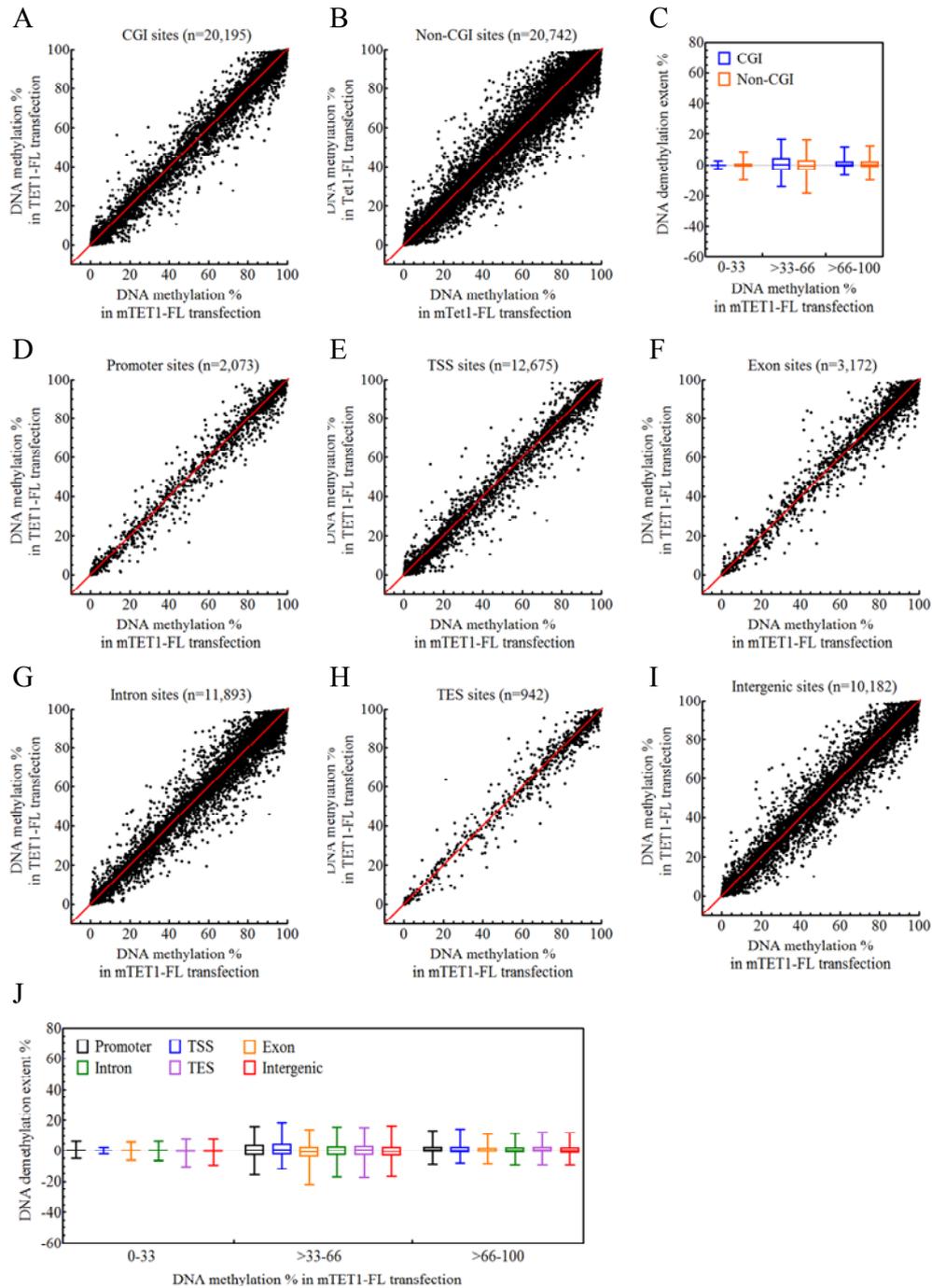
As the primary product of TET-catalyzed 5mC oxidation reaction, 5hmC also serves as a critical intermediate for TET-induced DNA demethylation (Guo et al., 2011; He et al., 2011; Ito et al., 2011; Valinluck and Sowers, 2007). To more specifically determine the different effects of TET1-FL and TET1-CD on DNA demethylation, we next performed genome-wide mapping of 5hmC in HEK293T cells with hMeDIP-seq. The specificity of 5hmC antibody under our DIP conditions was demonstrated by its specific immunoprecipitation of 5hmC-containing, but no

5mC or C-containing DNA probe (Figure 15).



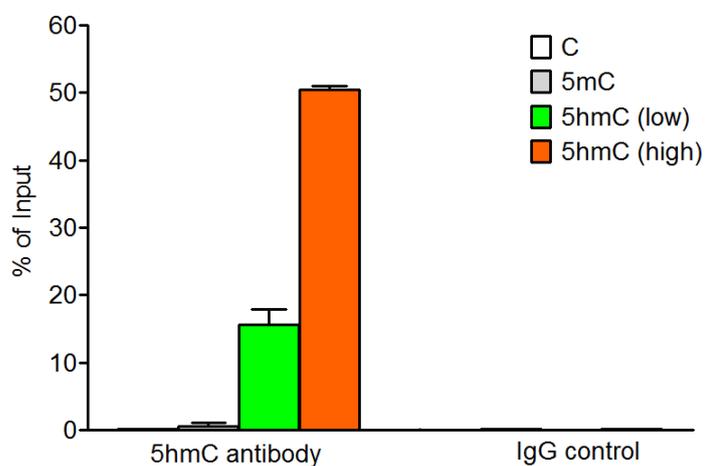
**Figure 13. TET1-FL overexpression fails to induce global DNA demethylation in HEK293T cells**

(A) Pair-wise comparison of DNA methylation levels at 40,937 common CpGs between TET1-FL and mTET1-FL transfections. (B) Boxplot shows demethylation efficiency indicated by demethylation extent in HEK293T cells overexpressing TET1-FL as a function of basal methylation level. The methylation levels are grouped in 10% intervals from 0-100% methylation. Boxes mark the interquartile range, and whiskers the 2.5th and 97.5th percentiles.



**Figure 14. No significant DNA demethylation induced by TET1-FL overexpression in all categorized genomic regions**

(A and B, D-I) DNA methylation change in CGIs (A), non-CGIs (B), promoter (D), TSS (E), exon (F), intron (G), TES (H) and intergenic region (I) after overexpression of TET1-FL. The identification of each genomic region see Figure 12. (C, J) Boxplots show demethylation efficiency of TET1-FL in CGIs and non-CGIs (C), and promoter, TSS, exon, intron, TES and intergenic region (J). The efficiency is indicated by demethylation extent as a function of basal methylation level, which is grouped in 33% intervals from 0-100% methylation. Boxes mark the interquartile range, and whiskers the 2.5th and 97.5th percentiles.



**Figure 15. Validation of the specificity of 5hmC antibody in hMeDIP experiment**

Four different DNA probes were prepared by PCR with different dCTP composition (C, only dCTP; 5mC, dCTP:5me-dCTP=1:1; 5hmC(low), dCTP:5hme-dCTP=9:1; 5hmC(high), dCTP:5hme-dCTP=1:1). All four probes were simultaneously spiked into HEK293T genomic DNA and subjected to hMeDIP procedure. The immunoprecipitated DNA probes were then analyzed by qPCR. Error bars represent SD from 3 independent experiments.

We first analyzed the genome-wide 5hmC distribution in control HEK293T cells which overexpressed mTET1-CD or mTET1-FL. Since neither mTET1-CD nor mTET1-FL is capable to catalyze 5mC oxidation, these two kinds of control HEK293T cells may exhibit a similar 5hmC distribution pattern, which should also be shared by non-transfected HEK293T cells. A total of 61,096 and 75,482 5hmC peaks were respectively identified in mTET1-CD- and mTET1-FL-overexpressed cells (Table 8). As expected, these 5hmC peaks in corresponding control cells showed an almost same genomic distribution, where 7-8% of peaks are located at promoters and TSSs, 5-6% at exons and 40-41% at introns (Figure 16B). Moreover, the 5hmC distribution patterns across gene body, exon and exon-intron boundary were also almost same between those two kinds of control cells, where 5hmC is significantly enriched around TSSs but evenly distributed at exons and introns at a relatively low level (Figure 16C-E). The highly similar 5hmC distributions in mTET1-CD- and mTET1-FL-overexpressed cells indirectly confirmed the reproducibility of our hMeDIP-seq assay.

We next examined the genome-wide 5hmC distribution in TET1-CD- and TET1-FL-overexpressed HEK293T cells. Consistent with the high 5hmC production detected in DNA dot-blot assay, a total of 314,577 5hmC peaks, which is almost 5 times more than that in mTET1-CD-overexpressed control HEK293T cells, were identified in TET1-CD-overexpressed cells (Table 8). By contrast, TET1-FL-overexpressed HEK293T cells showed only 111,627 5hmC peaks, just ~1.5 times more than that in mTET1-FL-overexpressed control cells (Table 8). Besides the difference in 5hmC peak number, we also interestingly found that the overall peak distributions in TET1-FL- and TET1-CD-overexpressed cells became different, in contrast to the high similarity between their control cells. Compared with TET1-CD-overexpressed cells,

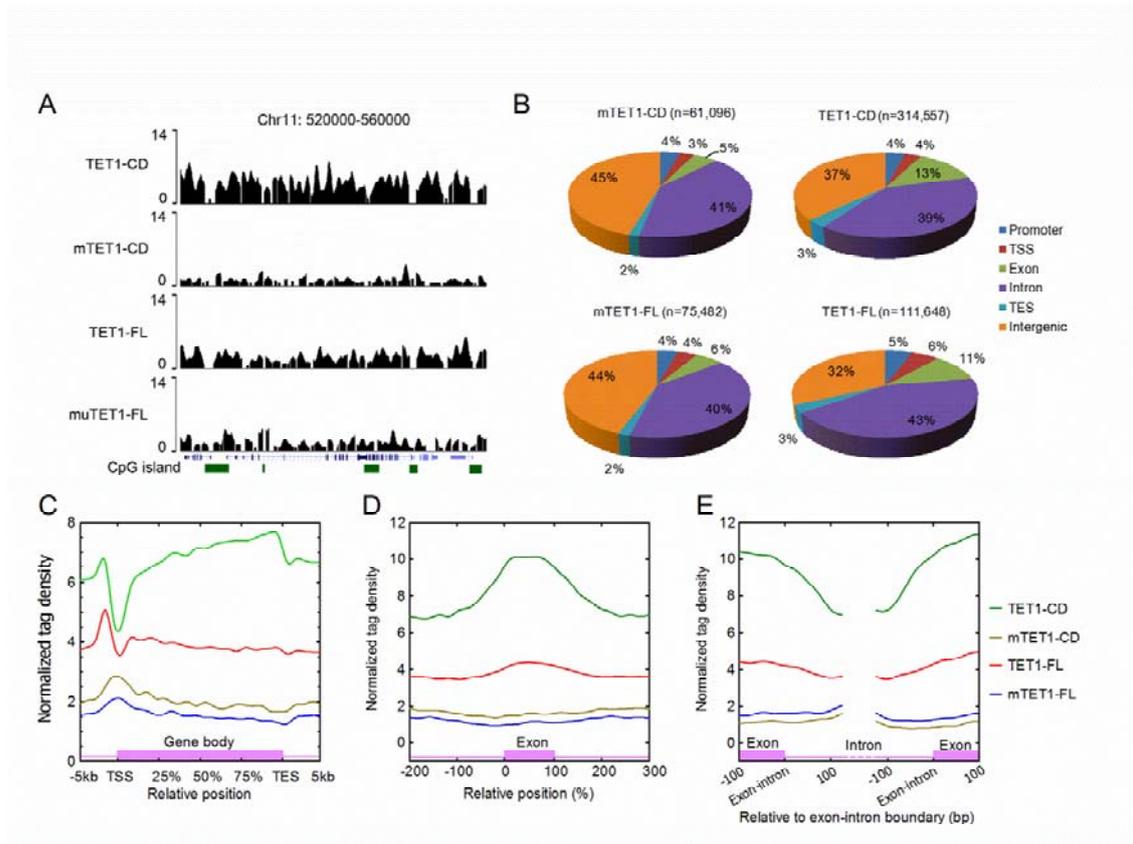
**Table 8. Summary of tags and peaks detected in hMeDIP-seq**

| Tags                        | TET1-CD    |            | mTET1-CD   |            | TET1-FL    |            | mTET1-FL   |            |
|-----------------------------|------------|------------|------------|------------|------------|------------|------------|------------|
|                             | IP         | Input      | IP         | Input      | IP         | Input      | IP         | Input      |
| Total number of tags        | 34,478,733 | 37,480,391 | 37,935,813 | 43,966,561 | 38,582,533 | 49,624,627 | 32,173,502 | 42,399,817 |
| Tags that don't pass filter | 8,440,889  | 9,043,239  | 8,788,483  | 11,704,316 | 7,553,867  | 14,509,619 | 5,552,724  | 8,975,966  |
| Tags that do not match      | 583,448    | 689,111    | 963,407    | 749,250    | 753,689    | 819,551    | 831,814    | 743,068    |
| Tags that multiply match    | 5,954,995  | 5,047,547  | 5,491,576  | 5,768,488  | 5,897,866  | 5,997,573  | 4,899,714  | 5,895,509  |
| Usable tags                 | 19,381,405 | 22,552,692 | 22,364,300 | 25,596,637 | 24,110,566 | 28,129,510 | 20,517,097 | 26,591,968 |
| Unique usable tags          | 18,434,399 | 18,561,419 | 19,270,918 | 20,706,856 | 21,036,618 | 26,298,816 | 18,366,199 | 24,381,049 |
| Peaks identified            | 314,577    | /          | 61,096     | /          | 111,627    | /          | 75,482     | /          |

The number of background tags in each ChIP sample is estimated by the noise rate from CCAT ( $\text{background\#} = \text{total\#} \times \text{noise rate}$ ). The noise rates are 0.516287, 0.811387, 0.702136 and 0.77295 for TET1-CD, mTET1-CD, TET1-FL and mTET1-FL, respectively.

TET1-FL-overexpressed cells showed relatively higher percents of 5hmC peaks in promoters and TSSs, and introns (11% vs. 8% and 43% vs. 39%, respectively), but lower percents in exons and intergenic regions (11% vs. 13% and 32% vs. 37%, respectively) (Figure 16B). Considering TET1-FL but not TET1-CD contains a CXXC DNA binding domain, these differences in 5hmC peak number and distribution may be closely associated with their different DNA binding feature. In support of it, a specific enrichment of Tet1 at promoters and TSSs has been observed in mESCs (Williams et al., 2011; Xu et al., 2011), and it may reasonably explain the higher percents of 5hmC peak in promoters and TSSs in TET1-FL-overexpressed cells. To further demonstrate the different functional pattern of TET1-CD and TET1-FL in 5hmC production, we further compared their 5hmC distribution across gene bodies, exons and exon-intron boundaries. Strikingly, contrary to their control cells where 5hmC is significantly enriched around TSSs, both TET1-CD- and TET1-FL-overexpressed cells showed the lowest 5hmC tag density at the same region (Figure 16C). More surprisingly, in TET1-CD-overexpressed cells 5hmC tag density gradually and sharply increased from TSSs toward TESs followed by a sudden decline around TESs, while TET1-FL-overexpressed cells exhibited a peak of 5hmC tag density at promoters but low level through gene bodies as control cells (Figure 16C). Consistent with it, we further found that the 5hmC tag density more dramatically increased at exons in TET1-CD-overexpressed cells than TET1-FL-overexpressed cells (Figure 16D and 16E). Thus, these findings further confirm the different pattern of TET1-FL and TET1-CD in 5hmC production.

Taken together, these results reveal that TET1-FL and TET1-CD differentially regulated global 5hmC distribution in HEK293T cells. In contrast to TET1-CD overexpression which dramatically increased 5mC peak number and changed 5hmC distribution pattern, overexpression



**Figure 16. Overexpression of TET1-FL and TET1-CD generate different 5hmC distribution pattern**

(A) Examples of hMeDIP-Seq profiles at a genomic region in HEK293T cells overexpressing (m)TET1-CD or (m)TET1-FL. (B) Genomic distribution of 5hmC peaks detected in HEK293T cells overexpressing (m)TET1-CD or (m)TET1-FL. (C-E) Normalized 5hmC tag density distribution across gene body (C), exon (D) and exon-intron boundary (E). Each gene body and exon is normalized to 0%-100% in (C and D). In (E), normalized tag density is plotted from -100 bp to 100 bp relative to exon-intron boundaries.

of TET1-FL regulated 5hmC distribution in a much more moderate manner so that the original 5hmC distribution pattern mostly remained in TET1-FL-overexpressed cells. Moreover, those data also complement our global DNA methylation analysis results and strongly support that TET1 is at least not an efficient DNA demethylase as previously predicted.

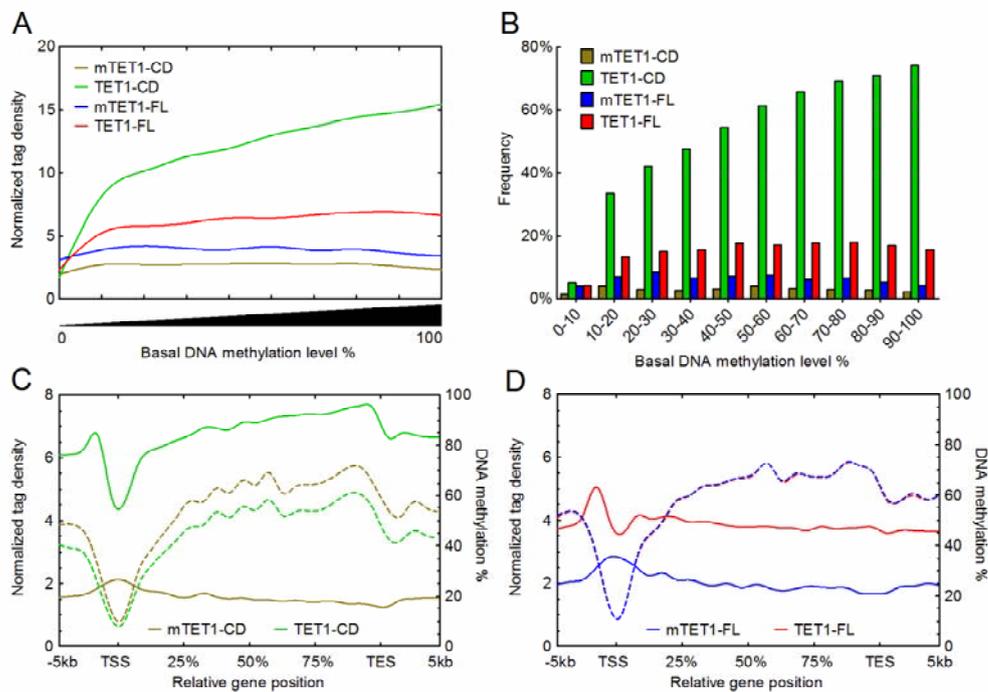
### **3.1.7 TET1-FL shows a unique regulation pattern of 5mC**

The requirement of pre-existing 5mC for TET-catalyzed 5hmC production implies a positive correlation between 5hmC and 5mC distribution in genomic DNA. However, multiple studies have reported that 5hmC and 5mC actually have distinct genomic distributions in mESC (Ficz et al., 2011; Williams et al., 2011; Wu et al., 2011). For example, contrary to 5mC, 5hmC is significantly enriched at TSSs but generally not detectable at repetitive elements and minor satellite repeats in mESC (Williams et al., 2011). Although its underlying mechanism remains unclear, this discrepant distribution of 5hmC and 5mC suggests that TET1-catalyzed 5hmC production may not always positively correlate with local 5mC level. More importantly, this potential functional pattern of TET1 may inevitably impair its capability to induce DNA demethylation. Therefore, we next combined our genome-wide DNA methylation and 5hmC distribution analysis results and analyzed the correlation between 5mC and 5hmC level in both TET1-CD- and TET1-FL-overexpressed cells. In the analysis of 5hmC tag density relative to basal DNA methylation level, the 5hmC tag density in TET1-CD-overexpressed cells significantly and progressively increased along with the increase of basal methylation level, demonstrating a strong positive correlation between 5hmC and 5mC level (Figure 17A). By contrast, the 5hmC tag density in TET1-FL-overexpressed cells remained at a constant and low level regardless of the increase of basal methylation level, as that in control cells overexpressing mTET1-CD and

mTET1-FL (Figure 16A). Similarly, the analysis of the frequency of 5hmC peak relative to basal DNA methylation level also revealed a strong positive correlation in TET1-CD- but not TET1-FL-overexpressed cells (Figure 17B). Thus, different from TET1-CD, TET1-FL displays a unique regulation pattern of 5mC where its catalytic activity appears to be gradually inhibited as the basal DNA methylation level increases.

To further determine whether the unique regulation pattern of 5mC by TET1-FL contributes to its failure to induce DNA demethylation, we next profiled 5hmC tag density together with DNA methylation level across gene body. The basal DNA methylation levels in both kinds of control cells expectedly reached bottom at TSSs and then gradually increased towards TESs, followed by a dramatic drop around TESs (Figure 17C and 17D). Strikingly, the 5hmC distribution profile in TET1-CD-overexpressed cells highly resembles the basal DNA methylation pattern, further confirming the positive correlation between 5mC and 5hmC level in the setting of TET1-CD overexpression (Figure 17C). More importantly, as the 5hmC density significantly increased towards TESs, TET1-CD-induced demethylation extent (i.e. the gap between methylation levels of mTET1-CD-overexpressed cells and TET1-CD-overexpressed cells) also accordingly increased, demonstrating the requirement of high 5hmC production for significant DNA demethylation. On the contrary, however, due to its unique regulation pattern of 5mC, TET1-FL overexpression failed to produce more 5hmC as basal DNA methylation level increased along gene bodies, which ultimately led to its failure to induce DNA demethylation (Figure 17D).

Taken together, by analyzing the correlation between 5hmC distribution and basal DNA methylation level, we found a unique regulation pattern of 5mC by TET1-FL, where TET1 fails to



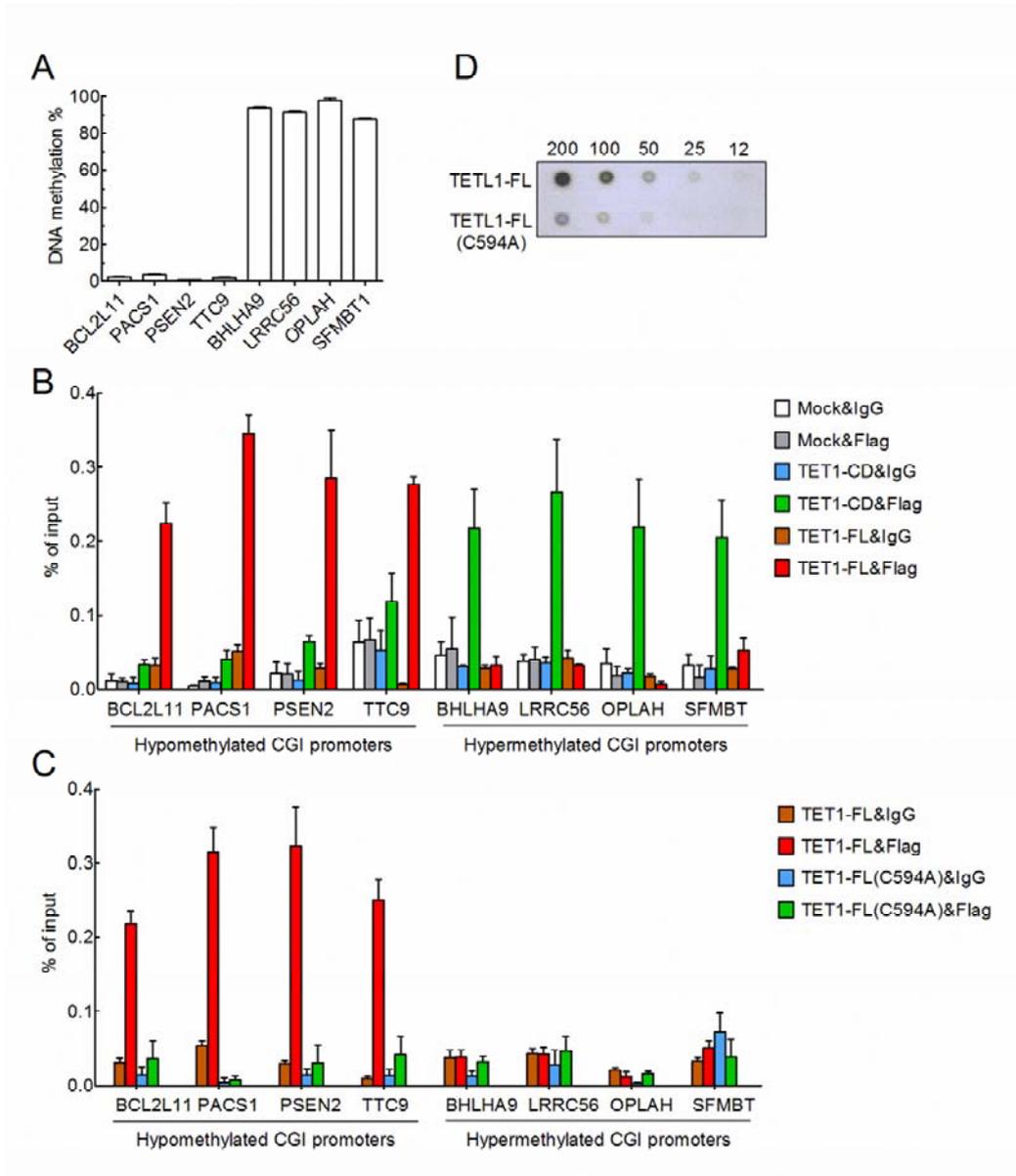
**Figure 17. Compromised 5hmC production by TET1-FL as basal DNA methylation level increases**

(A) Normalized 5hmC tag density as a function of basal DNA methylation level. The methylation levels are grouped in 10% intervals from 0-100% methylation. (B) Frequency distribution of 5hmC peaks-covering Smal sites to differentially methylated sites. The methylation levels are grouped in 10% intervals from 0-100% methylation. (C and D) Correlation of normalized 5hmC tag density with basal DNA methylation level across gene body in HEK293T cells overexpressing TET1-CD (C) or TET1-FL (D). Solid line represents normalized 5hmC tag density, and dotted line basal DNA methylation level.

catalyze more 5hmC production as basal methylation level increases. This regulation pattern therefore well explains the failure of TET1-FL to induce significant DNA demethylation.

### **3.1.8 TET1-FL specifically binds hypomethylated CGIs through its CXXC domain**

Having demonstrated that the unique regulation pattern of 5mC by TET1-FL contributes to its failure to induce DNA demethylation, we next investigated the mechanism underlying that unique functional pattern of TET1-FL. Given a high 5hmC yield required for significant DNA demethylation depends on not only high level of substrate 5mC but also rich amount of TET1 bound, we hypothesized that TET1 may gradually lose its binding to genomic regions as their methylation levels increase, which makes it always fail to produce enough 5hmC to achieve significant DNA demethylation. To address it, we first tested TET1-FL occupancy at eight hypomethylated (*BCL2L11*, *PACSI*, *PSEN2* and *TTC9*) or hypermethylated (*BHLHA9*, *LRRC56*, *OPALH* and *SFMBT1*) CGI promoters with CHIP-qPCR. The different DNA methylation states of these promoters were initially identified based on our genome-wide DNA methylation analysis results and also validated by bisulfite-pyrosequencing assay (Figure 18A). In support of our hypothesis, TET1-FL is highly enriched at the hypomethylated CGI promoters but dramatically excluded from hypermethylated CGI promoters (Figure 18B), consistent with a recently published result that in mESC Tet1-bound CGIs are associated with lower 5mC levels than CGIs not bound by Tet1 (Wu et al., 2011). To well understand the effect of TET1-CD on DNA demethylation, we also examined the occupancy of TET1-CD in those hypomethylated and hypermethylated CGI promoters. Surprisingly, TET1-CD which lacks CXXC domain extensively bound both kinds of CGI promoters with a preference for the hypermethylated CGIs (Figure 18B). Although the



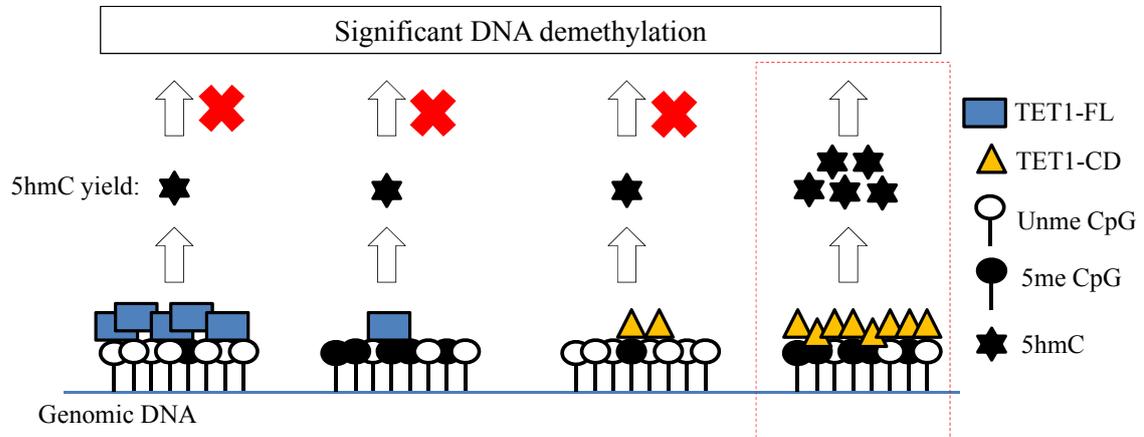
**Figure 18. The CXXC domain specifically targets TET1-FL towards hypomethylated CGIs**  
 (A) Bisulfite-pyrosequencing analysis of eight CGI promoters in untreated HEK293T cells. (B) Flag ChIP-qPCR analysis of TET1-CD and TET1-FL occupancy at differentially methylated CGI promoters in (A). (C) Flag ChIP-qPCR analysis of TET1-FL and TET1-FL (C594) occupancy at differentially methylated CGI promoters in (A). (D) DNA dot-blot analysis of genomic 5-hydroxymethylcytosine (5hmC) level in HEK293T cells overexpressing TET1-FL or TET1-FL (C594). Error bars represent SD from 2-3 independent experiments.

underlying mechanism is unknown, this distinct binding feature of TET1-CD strongly supports its potent function for genome-wide DNA demethylation.

To further determine whether the binding of TET1-FL to hypomethylated CGI promoters is dependent on its CXXC domain, we next constructed a mutant TET1-FL that contains a single substitution mutation in CXXC domain (C594A) (Xu et al., 2011) and then overexpressed it in HEK293T cells. As Figure 17C shows, TET1-FL(C594A) completely lost the enrichment at hypomethylated CGI promoters, confirming the CXXC-dependent TET1 binding. As a result of the loss of DNA binding, TET1-FL(C594A) overexpression also produced much lower 5hmC yield compared with that of normal TET1-FL (Figure 18D). Therefore, these results reveal that TET1-FL specifically binds hypomethylated but not hypermethylated CGI promoters via its CXXC domain. More importantly, it is highly possible that the binding feature of TET1-FL contributes to its unique regulation pattern of 5mC and then causes its failure to induce significant DNA demethylation.

### **3.1.9 Proposed model**

Based on our available results, we now conclude that the CXXC and 5mC dioxygenase catalytic domains in TET1 form an interesting but conflicting domain combination: CXXC domain specific targets TET1-FL towards hypomethylated but not hypermethylated CGI regions, whereas its catalytic domain requires 5mC as substrate for 5hmC production as well as DNA demethylation. Here we further propose a model to well describe the regulatory effect of TET1-FL on DNA methylation (Figure 19). In hypermethylated CpG-rich regions the substrate 5mC is enriched, but just due to their high methylation levels few TET1-FL binds them, which thus leads to a low production of 5hmC and then poor DNA demethylation; on the contrary



**Figure 19. Model for the self-inhibited catalytic function of TET1-FL for DNA demethylation**

Through its CXXC domain, TET1-FL preferentially binds to hypomethylated CpG-rich regions, but is under represented in hypermethylated regions. Due to the poor 5mC substrates, TET1-FL produces few 5hmC in hypomethylated regions. In hypermethylated regions, TET1-FL still produce few 5hmC because of the poor binding of TET1-FL. Thus in both cases, no significant DNA demethylation can be induced by TET1-FL. On the contrary, through an unknown mechanism, TET1-CD which lacks the CXXC domain preferentially binds hypermethylated region, which thus enables TET1-CD to induce high production of 5hmC as well as significant DNA demethylation.

through CXXC domain TET1-FL preferentially binds hypomethylated CpG-rich regions, but its 5hmC production and DNA demethylation activity are still very poor because of the low level of 5mC in those regions (Figure 19). Strikingly, by contrast with TET1-FL, through unknown mechanism TET1-CD is more enriched at hypermethylated CpG-rich regions than those hypomethylated, which well fits its catalytic function on 5mC to efficiently induce DNA demethylation in those methylated regions (Figure 19). In short, the CXXC domain specifically targets TET1 to hypomethylated regions and consequently prevents the catalytic domain from inducing DNA demethylation in moderately or highly methylated regions. Therefore, this model reveals a self-inhibited catalytic function of TET1 in regulating DNA methylation, and reasonably explains the observed failure of TET1-FL overexpression to induce significant DNA demethylation.

### **3.2 TET1 maintains the DNA hypomethylated state of CGIs in HEK293T cells**

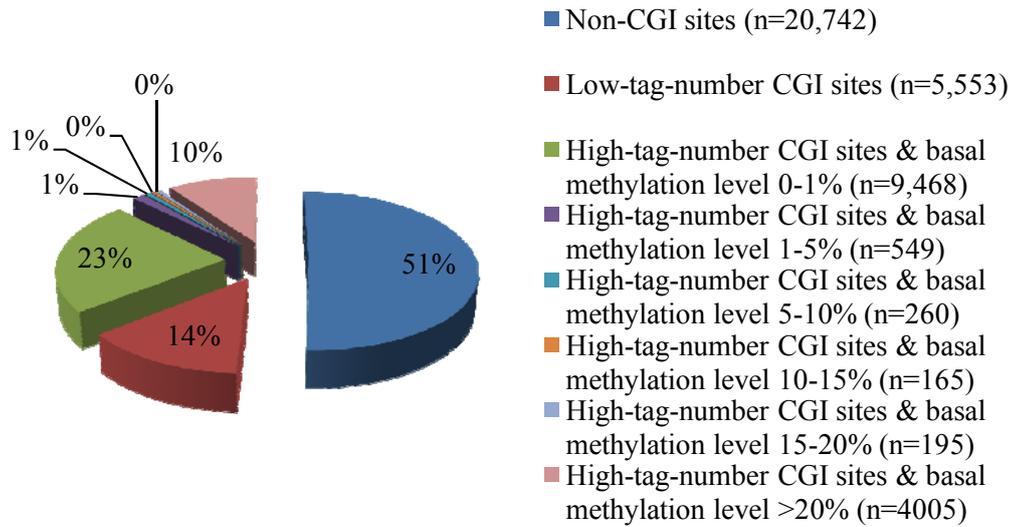
Overexpression of TET1-CD induces a significant global DNA demethylation, supporting that TET-mediated 5mC oxidation can serve as an efficient pathway for DNA demethylation. By contrast, TET1-FL overexpression fails to induce significant DNA demethylation and CXXC domain specifically targets TET1 to hypomethylated CGIs. Thus, these available evidences raise an intriguing hypothesis that TET1 plays as a unique DNA demethylase not intended to change DNA methylation levels, but rather specifically functioning in hypomethylated CGIs to maintain their DNA hypomethylation state by removing stochastic de novo DNA methylation. To address this hypothesis, we next mainly studied two questions: whether TET1-FL overexpression specifically decreases DNA methylation in hypomethylated CGI regions and how DNA

methylation changes in hypomethylated CGIs after TET1 knockdown.

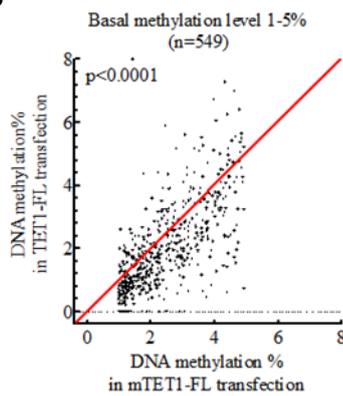
### **3.2.1 TET1-FL overexpression specifically decreases DNA methylation in hypomethylated CGIs**

Given the preferential binding of TET1-FL to hypomethylated CGIs, TET1-FL overexpression may specifically decrease DNA methylation level to some extent in hypomethylated CGI regions, though previous analysis has shown that TET1-FL overexpression failed to induce significant DNA demethylation. To confirm it, we deeply analyzed the DREAM data for (m)TET1-FL transfections with a focus on hypomethylated CGI sites. Considering the basal methylation level is very low and the potential DNA methylation change should be also relatively weak, we further narrowed the hypomethylated CGI sites by elevating the cutoff for tag number from 20 to 100. As Figure 20A shows, those high-tag-number CGI sites were divided into six groups based on their basal methylation levels, among which we focused on the groups with lower methylation level of 1-5%, 5-10%, 10-15% and 15-20%. Interestingly, we found much more sites showing decreased methylation level after TET1-FL overexpression in groups of 1-5% and 5-10% but not 10-15% and 15-20% (Figure 20B). With Wilcoxon paired signed-rank test, statistic significance was also obtained in groups of 1-5% and 5-10%, indicating TET1-FL overexpression specifically decrease DNA methylation in hypomethylated CGI sites. Therefore, in contrast to its failure to induce significant DNA demethylation in moderately or highly methylated regions, TET1-FL overexpression really decreased DNA methylation to certain extent in hypomethylated CGI sites, which strongly supports the potential role of TET1 in maintaining the DNA hypomethylated state in CGIs.

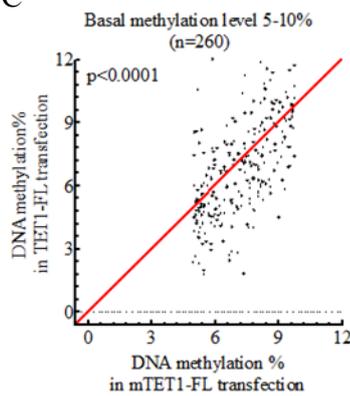
A



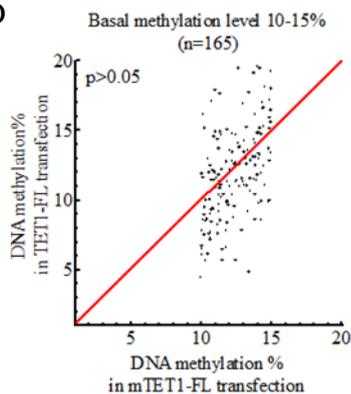
B



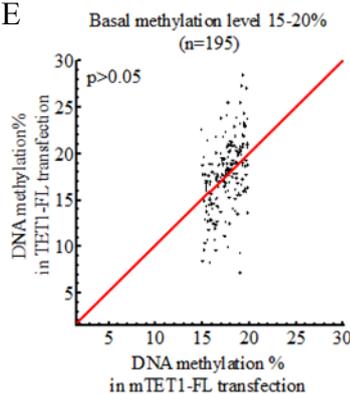
C



D



E



**Figure 20. Decreased DNA methylation in high-tag-number CGI sites by TET1-FL overexpression in HEK293T cells**

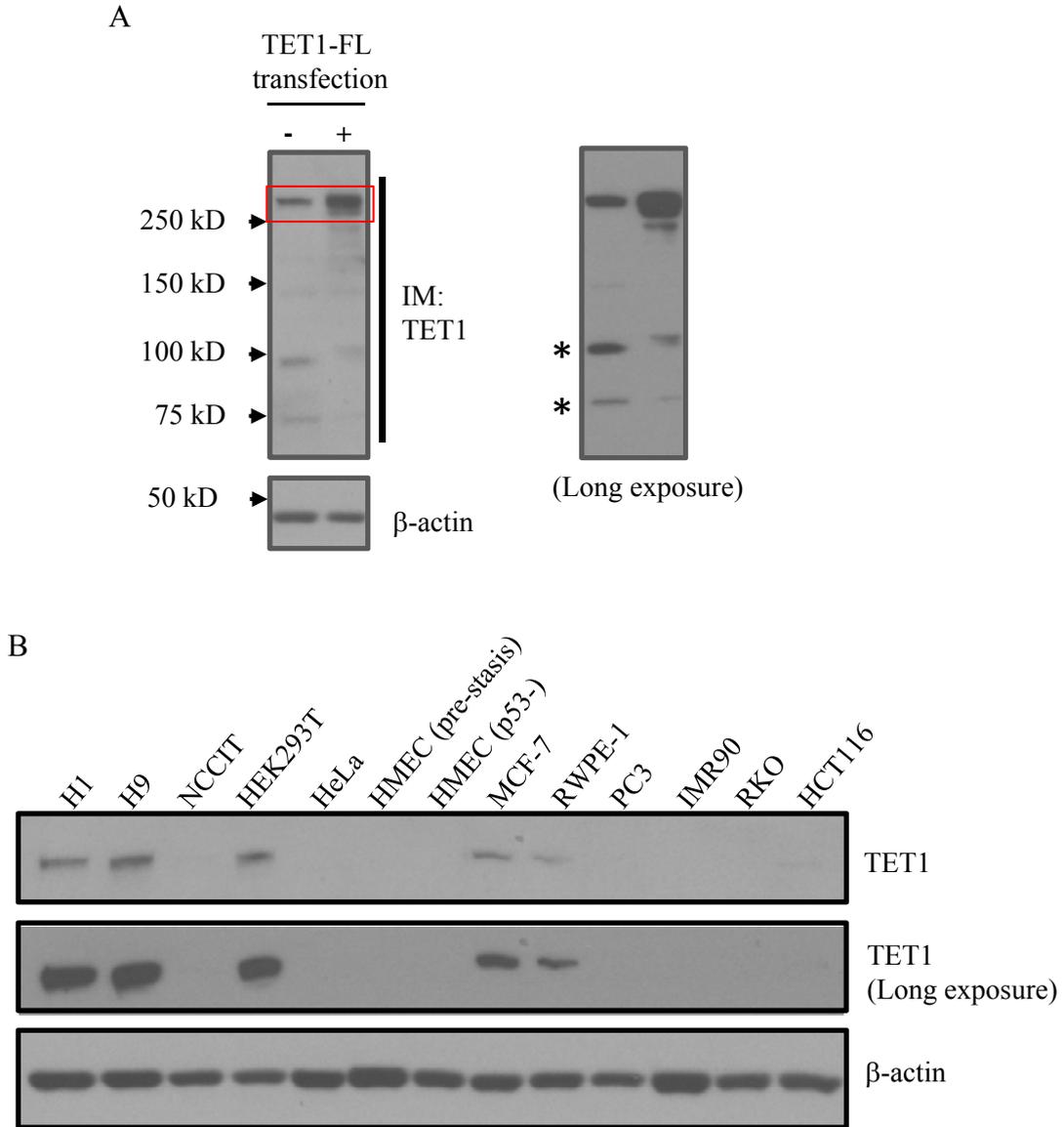
(A) Percentage of *Sma*I sites classified by CGI, tag number and basal methylation level. High-tag-number,  $\geq 100$  tags; low-tag-number,  $< 100$  tags. (B-E) DNA methylation change in high-tag-number CGI sites with basal methylation level of 1-5% (B), 5-10% (C), 10-15% (D) and 15-20% (E). p values obtained from Wilcoxon paired signed-rank test.

### **3.2.2 TET1 knockdown leads to an increase of DNA methylation in hypomethylated CGIs**

To further demonstrate that TET1 specifically maintains the DNA hypomethylation state in hypomethylated CGIs, we next studied whether DNA methylation increases in hypomethylated CGIs after TET1 knockdown. Although previous studies have reported that depletion of Tet1 induces increased DNA methylation levels in Tet1-bound regions in mESC, the underlying mechanism remains unknown (Wu et al., 2011; Xu et al., 2011). Moreover, the impairment of mESCs maintenance and differentiation induction by Tet1 knockdown add many other complexities to explain that increase of DNA methylation in mESCs (Ito et al., 2010). Thus, ESCs seems to be not suitable for studying the effect of TET1 knockdown on DNA methylation.

#### **3.2.2.1 Screening of TET1 expression in various human cell lines**

To construct stable TET1 knockdown, an optimal target cell line with high TET1 expression level is required. Tet1 has been only reported to be highly expressed in mESCs but significantly down-regulated in response to withdrawal of leukemia inhibitory factor, suggesting low expression of Tet1 in differentiated somatic cells (Tahiliani et al., 2009). Moreover, little is known about the TET1 expression level in human cell lines. With western blot assay we next screened TET1 expression level in various human cell lines, including ESCs, non-tumor and tumor cell lines. The specificity of anti-human TET1 antibody was validated with our TET1-FL overexpressed cell samples (Figure 21A). Compared with other cell lines, both human ES cell line H1 and H9 showed a much higher TET1 expression level (Figure 21B), consistent with the reported high expression of Tet1 in mESCs. NCCIT is a teratocarcinoma cell line and can differentiate into derivatives of all three embryonic germ layers, but surprisingly no detectable



**Figure 21. TET1 expression level varies among different human cell lines**

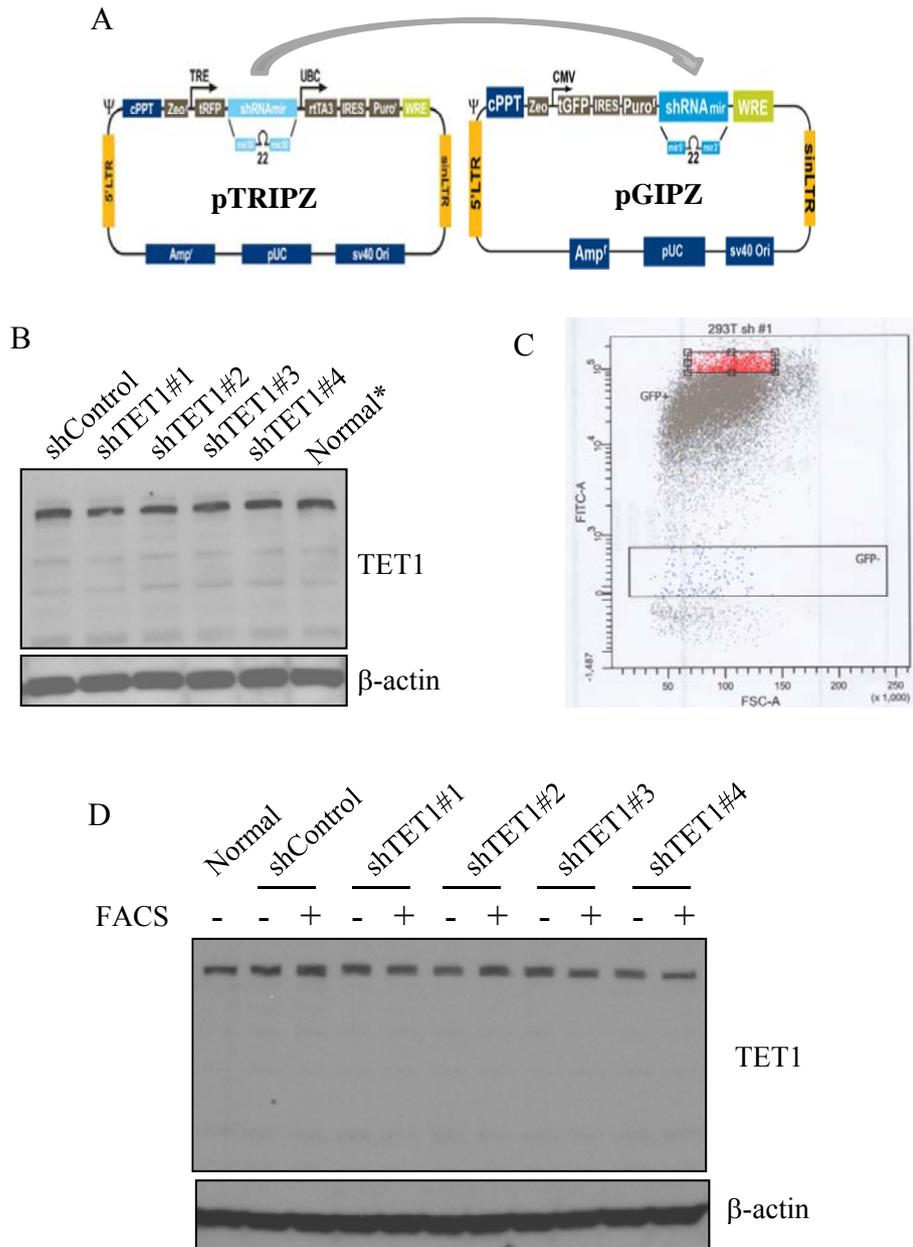
(A) The specificity of anti-human TET1 antibody was validated with our TET1-FL overexpressed cell samples; \* unspecific band. (B) TET1 expression level is variable among different human cell lines. Both human ES cell line H1 and H9 showed a much higher TET1 expression level. Moreover, HEK293T cells also show a comparable TET1 expression level as H1 and H9 cell lines.

TET1 protein level was observed, suggesting TET1 is at least dispensable for the pluripotency of NCCIT cells (Figure 21B). Strikingly, HEK293T cells exhibited a comparable TET1 protein level as human ESCs, suggesting TET1 may play an important role in this cell line (Figure 21B). Additionally, there was no detectable TET1 protein level in HeLa, HMEC (pre-stasis), HMEC (p53-), PC3, IMR90, RKO and HCT116 (Figure 21B). Interestingly, compared with breast non-tumor cell line HME, MCF7 has considerable TET1 expression. On the contrary, prostate non-tumor cell line RWPE-1 has considerable TET1 expression but cancer cell line PC3 does not have (Figure 21B). Taken together, although here no conclusion can be made regarding the expression level of TET1 in cancers, HEK293T is obviously a good target cell line for our subsequent TET1 knockdown experiments.

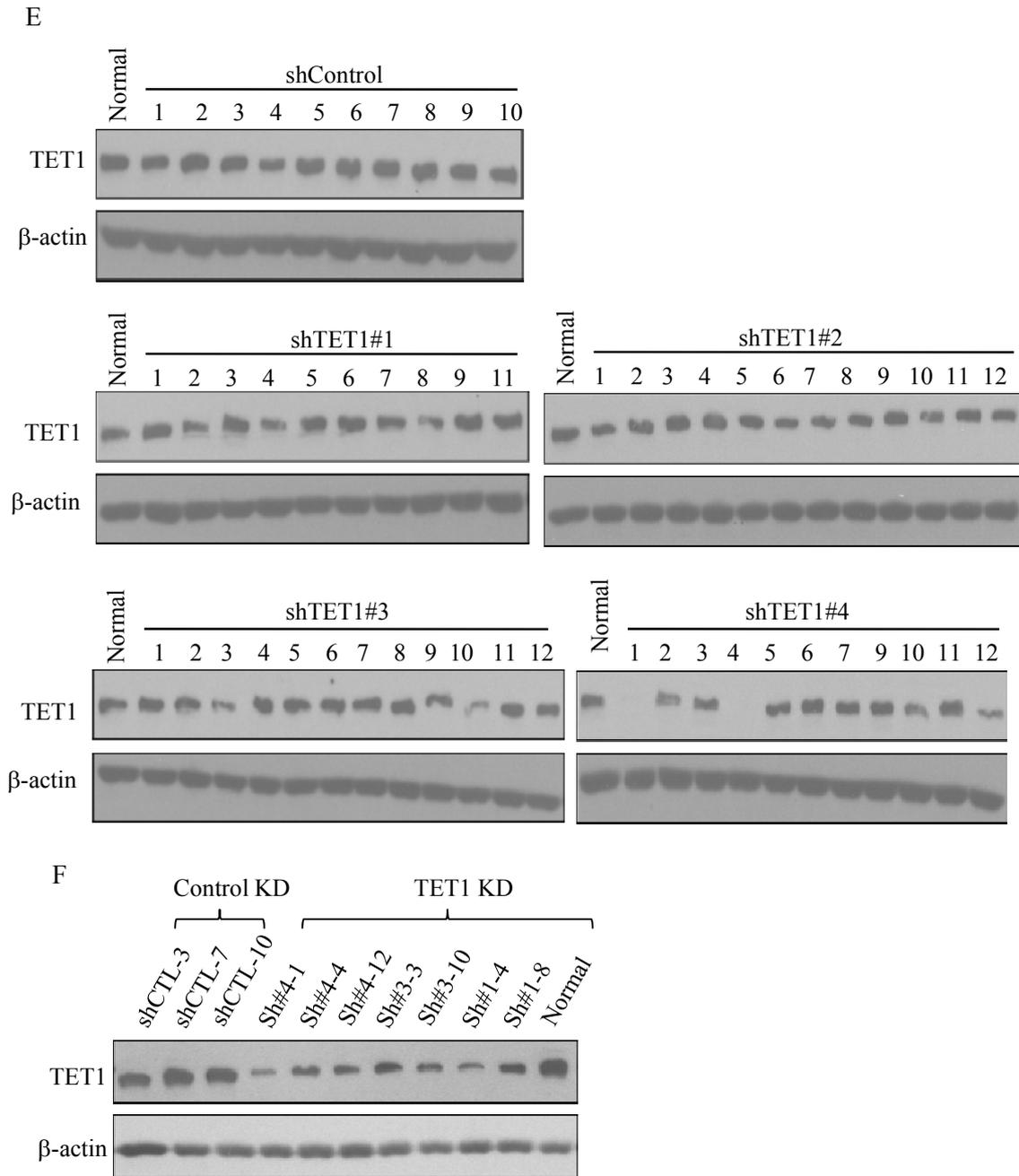
### **3.2.2.2 Lentiviral shRNA-mediated TET1 knockdown in HEK293T cells**

To construct stable TET1 knockdown in HEK293T cells, we screened four different shRNAs which were based on lentiviral pGIPZ vectors. These TET1 shRNAs were initially inserted in tetracycline-inducible shRNA pTRIPZ vector, but transferred to pGIPZ by us for a continuous knockdown effect (Figure 22A). According to standard lentiviral transduction protocol, we first got five stable cell pool populations, one with control shRNA, the other four with different TET1 shRNAs. However, compared with control shRNA cell pool, none of those four TET1 shRNA cell pools showed significant TET1 knockdown effect (Figure 22B).

Given the GFP reporter is driven by the same promoter as shRNA in pGIPZ vector, we proposed that cells with high GFP intensity may also highly express TET1 shRNA which in turn leads to marked knockdown effect. To confirm this strategy for improving knockdown effect, we



**Figure 22. Construction of lentiviral shRNA-mediated TET1 knockdown in HEK293T cells**  
 (A) Vector maps for pTRIPZ and pGIPZ. The TET1 shRNAs were transferred from pTRIPZ to pGIPZ vector for continuous expression. (B) Compared with control shRNA cell pool, none of those four TET1 shRNA cell pools showed significant TET1 knockdown effect in western blot assay. (C) The shControl or shTET1 transfected cell pools were sorted to collect a smaller cell population showing the highest 5% GFP intensity. (D) No significant different of TET1 protein level between parent cell pools and the sorted 5% cell populations was observed in western blot assay.



**Figure 22. Construction of lentiviral shRNA-mediated TET1 knockdown in HEK293T cells**  
 (E) Screen TET1 expression level in single cell clones transfected with shControl or shTET1 by Western blot assay. (F) Western blot assay confirmed TET1 knockdown effect in some single cell clones compared with shControl transfected cells.

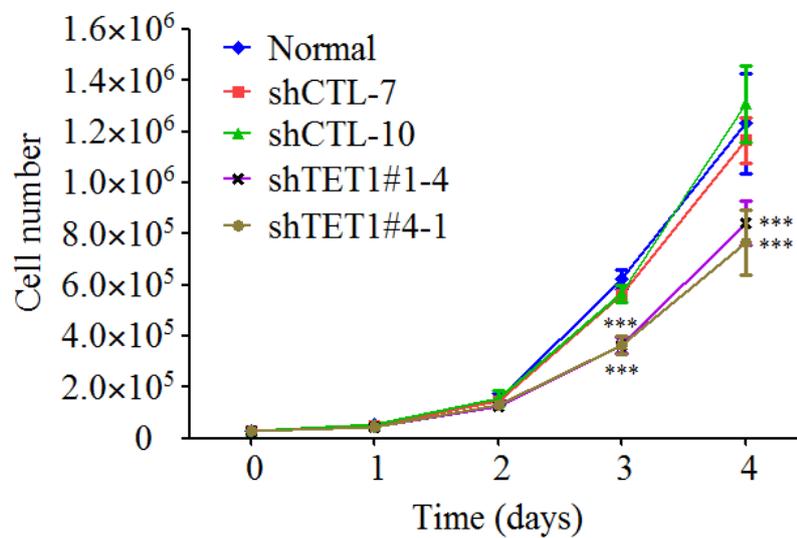
next sorted the available five cell pools to collect a smaller cell population showing the highest 5% GFP intensity (Figure 22C). However, with western blot assay no significant different of TET1 protein level between parent cell pools and sorted 5% cell populations was observed, indicating no positive correlation between GFP intensity and shRNA knockdown effect in our cells (Figure 22D). It could be explained by that, GFP intensity may be not only determined by its promoter transcription activity and copy number, but may also significantly fluctuate in different phases of cell cycle.

The above failure then prompted us to develop single cell clones. By serial dilution 10-12 single cell clones were established for each shRNA group. As expected, the TET1 protein levels in control shRNA single cell clones were comparable to that in normal HEK293T cell (Figure 22E). Fortunately, several TET1 shRNA clones from shTET1#1, 3 and 4 groups showed a significant knockdown effect when compared with normal cells (Figure 22E). We next further confirmed the TET1 knockdown effect by directly comparing those TET1 shRNA clones with control shRNA clones (Figure 22F). Based on this result, we finally chose shTET1#1-4 and shTET1#4-1, and shControl-10 clones for our subsequent experiments.

### **3.2.2.3 TET1 knockdown inhibits cell growth in HEK293T cells**

Tet1 knockdown has been reported to impair the maintenance of mESCs and lead to skewed differentiation of pre-implantation embryos (Ito et al., 2010). HEK293T has a comparable TET1 expression level as human ES cell lines, suggesting that TET1 may play an important role there. Thus, before investigating the possible change of DNA methylation in hypomethylated CGI, we first investigated whether TET1 knockdown significantly affects cell morphology and

proliferation. With the use of our HEK293T single cell clones with stable TET1 or control knockdown, we interestingly found that TET1 knockdown did not induce any change in cell morphology, but significantly inhibited cell growth in both shTET1#1-4 and shTET1#4-1 clones (Figure 23). These findings suggest that TET may be involved in the regulation of cell cycle progression, and further studies on the detailed change of cell cycle and related cell cycle regulators (e.g. p21) will provide more insight into that.



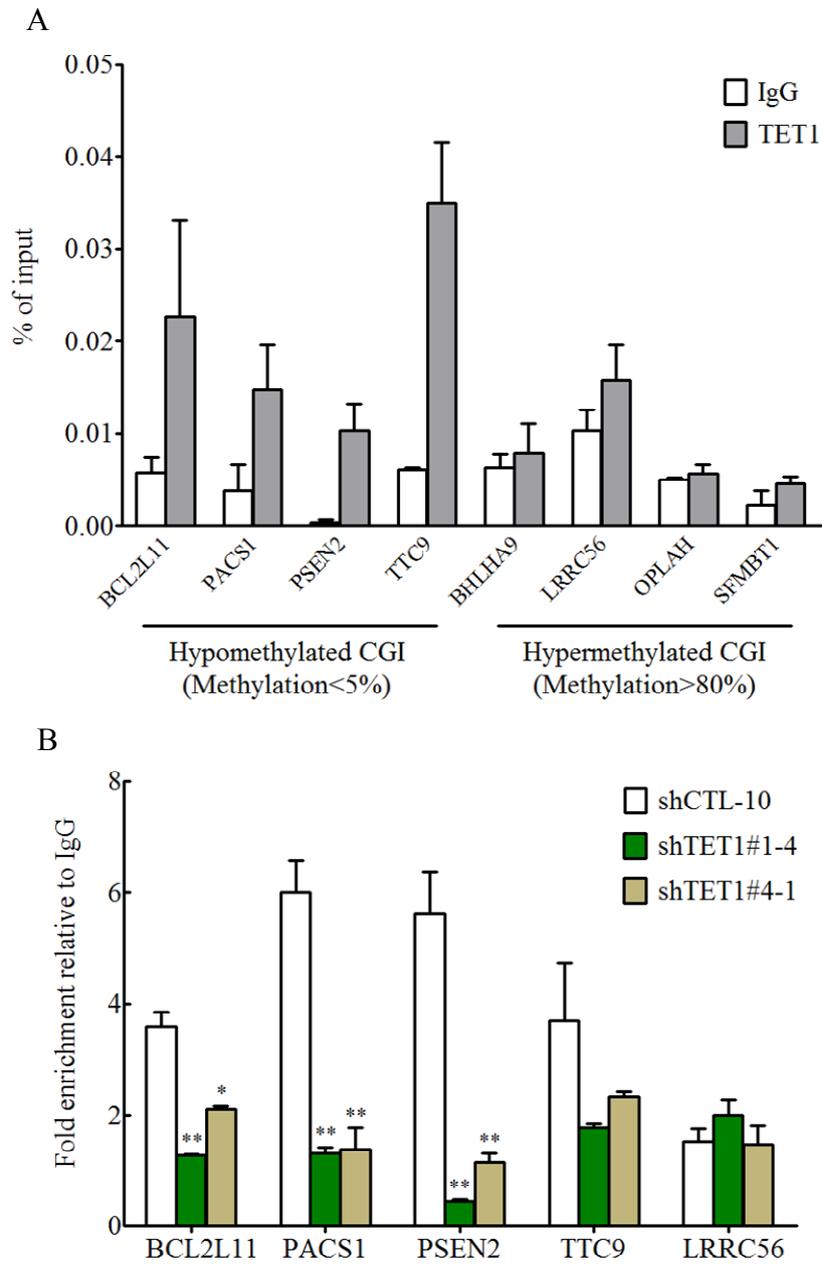
**Figure 23. Inhibited cell proliferation by TET1 knockdown in HEK293T cells**

Growth curves of mock, control and TET1 knockdown HEK293T cells were determined by counting the cell numbers every day. Error bars represent SD from 4 independent experiments. \*\*\*  $p < 0.001$  by two-way ANOVA/Bonferroni post-test compared to shCTL-10 clone cells.

### **3.2.2.4 Increase of DNA methylation in the pre-methylated edges of hypomethylated CGIs after TET1 knockdown**

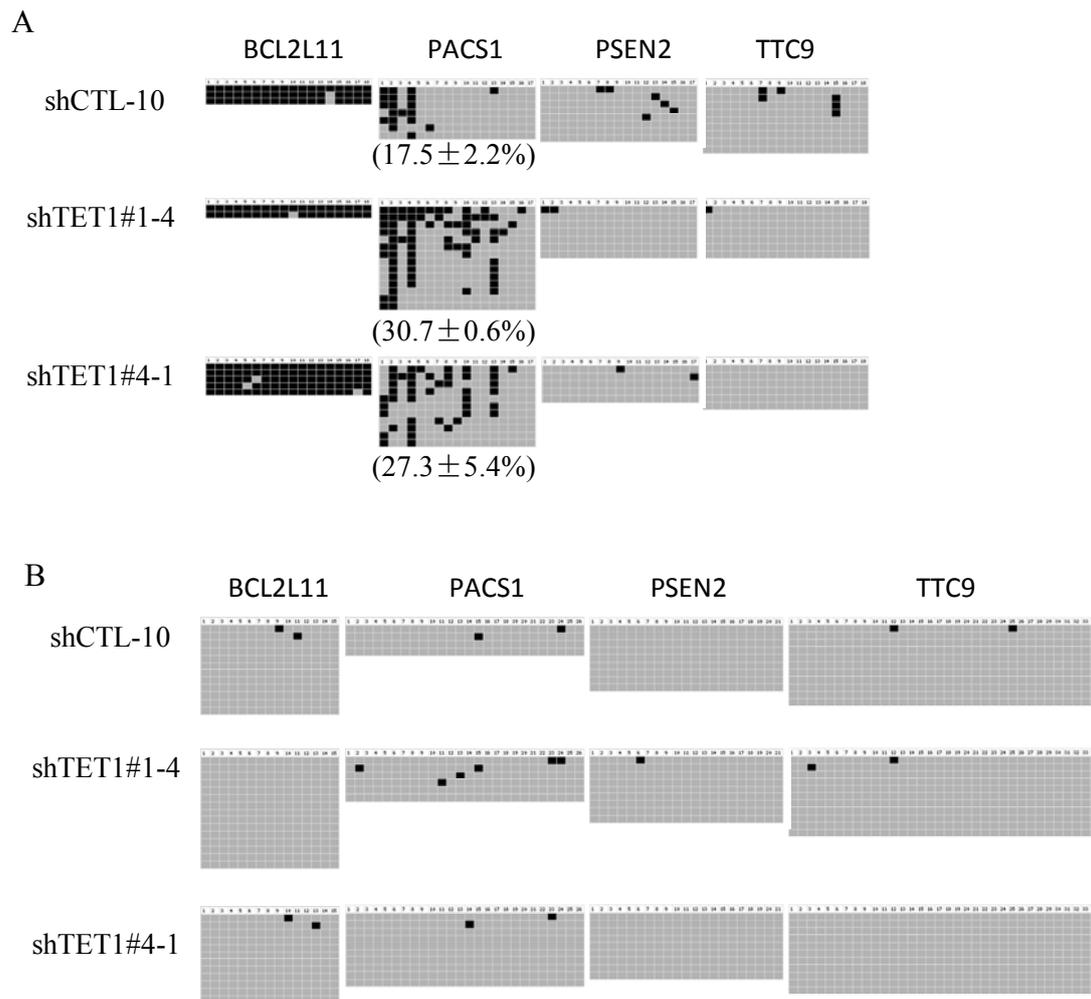
We next asked how DNA methylation changes in hypomethylated CGIs after TET1 knockdown. To this end, we first went to identify some target regions for endogenous TET1 in HEK293T cells. In our TET1 overexpression experiment, we have proved that ectopically expressed TET1-FL binds to the hypomethylated CGI promoters of BCL2L11, PACS1, PSEN2 and TTC9 but not the hypermethylated CGI promoters of BHLHA9, LRRC56, OPALH and SFMBT1 (Figure 18). Consistently, endogenous TET1 in normal HEK293T cells is also preferentially enriched at those hypomethylated CGI promoters but not hypermethylated CGI promoters (Figure 24A). Moreover, the TET1 enrichment signals at those hypomethylated CGIs specifically decreased in shTET1#1-4 and shTET1#4-1 clone cells, further confirming the binding of TET1 to these hypomethylated CGI regions (Figure 24B).

Subsequently, we studied the possible DNA methylation changes in those hypomethylated CGIs after TET1 knockdown. With the established control and TET1 knockdown clones cells, both edge and central regions of each CGI were tested by bisulfite-sequencing. Among these eight tested regions in control knockdown clone cells, most regions are expectedly hypomethylated, while the edge region of BCL2L11 were completely methylated and that of PACS1 CGI showed considerable pre-existing DNA methylation at 5'upstream (Figure 25). More interestingly, both TET1 knockdown clones cells exhibited a significant increase of DNA methylation in the edge region of PACS1 CGI but not in other tested regions, indicating that TET1 knockdown may selectively induce DNA methylation gain in the edges of certain CGIs (Figure 25). Given the pre-existing DNA methylation plays a critical role as a seed for de novo DNA methylation to



**Figure 24. ChIP-qPCR analysis of endogenous TET1 target genes in HEK293T cells**

(A) Endogenous TET1 preferentially binds hypomethylated but not hypermethylated CGI promoters in normal untransfected HEK293T cells. (B) TET1 knockdown specifically decreases the enrichment of TET1 in hypomethylated CGI promoters. Error bars represent SD from 2-3 independent experiments. \*  $p < 0.05$ , \*\*  $p < 0.01$  by Student's t-test compared to shCTL-10 clone cells.



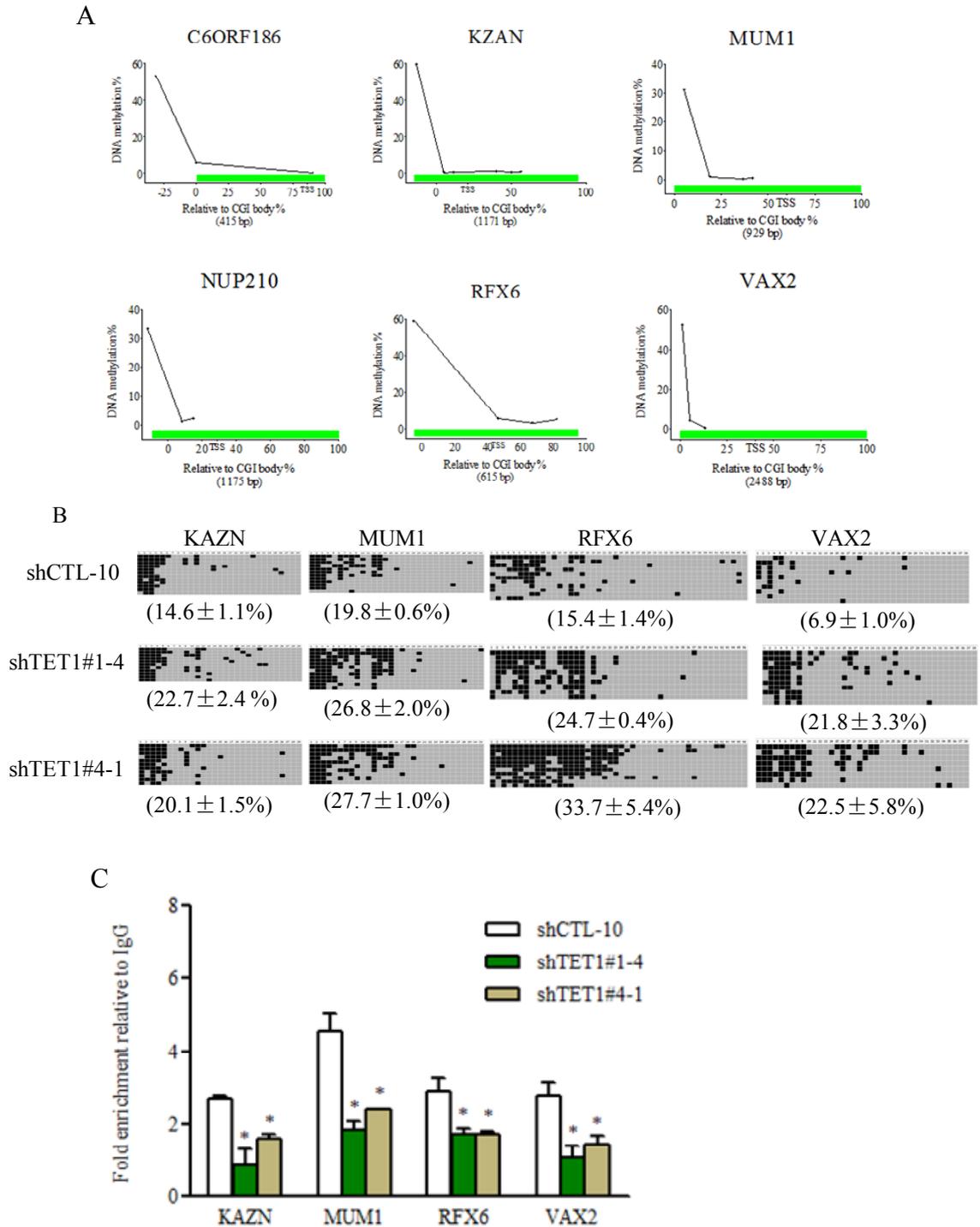
**Figure 25. Bisulfite-sequencing analysis of CGI promoters of BCL2L11, PACS1, PSEN2 and TTC9**

(A-B) Both the edge (A) and center regions (B) of CGI in BCL2L11, PACS1, PSEN2 and TTC9 promoter were tested for DNA methylation changes after TET1 knockdown in HEK293T cells. Mean  $\pm$  SD from 2 independent experiments.

spread into nearby unmethylated regions, and among all tested regions only the edge of PACS1 CGI contains pre-existing DNA methylation, it may be further proposed that through its 5mC oxidation-mediated DNA demethylating ability TET1 rapidly removes the stochastic de novo methylation near the pre-methylated edges of CGIs and thus efficiently block the spreading of DNA methylation into the unmethylated central regions of CGIs.

To accumulate more supportive evidences, we next looked for more hypomethylated CGI promoters with pre-methylated edges. Based on our DREAM results we initially chose six such CGIs promoters (*C6ORF186*, *KAZN*, *MUM1*, *NUP210*, *RFX6* and *VAX2*) (Figure 26A). The DNA methylation states of these CGI edges were then successfully validated by bisulfite-sequencing except those of *C6ORF186* and *NUP210*, due to the high recombination rate of their sequences in bacterial transformation (Figure 26B). More importantly, compared with control knockdown cells, both TET1 knockdown clones cells show significant increase of DNA methylation in all CGI promoter edges of *KAZN*, *MUM1*, *RFX6* and *VAX2*, strongly supporting the inhibitory effect of TET1 on the spreading of aberrant de novo methylation into hypomethylated CGIs which contain pre-methylated edges (Figure 26B). Finally, the binding of TET1 at those four CGI promoters was also confirmed by ChIP-qPCR (Figure 26C).

Taken together, these above results show that in hypomethylated CGIs TET1 can efficiently block the spreading of DNA methylation from pre-methylated edges into the unmethylated center regions, which should be through its rapidly removing the aberrant de novo methylation at pre-methylated edges with its 5mC oxidation activity. Thus, combined with the results that TET1 overexpression specifically decreased DNA methylation in hypomethylated CGI regions, those findings strongly support that TET1 specifically functions as a unique DNA demethylase to



**Figure 26. TET1 knockdown leads to spreading of de novo methylation at the pre-methylated edges of CGIs**

A) Six CGIs with pre-methylated edge were chosen based on global DNA methylation analysis results. Green bar represents each CGI. (B) Bisulfite-sequencing analysis of the pre-methylated edges of those CGI after TET1 knockdown. Mean  $\pm$  SD from 2 independent experiments. (C) TET1 knockdown decreases TET1 enrichment at those CGIs by ChIP-qPCR. Error bars represent SD from 2 independent experiments. \*  $p < 0.05$  by Student's t-test compared to shCTL-10 clone cells.

maintain the DNA hypomethylation state of CGIs by removing aberrant de novo DNA methylation.

### **3.3 Mechanistic study of TET-mediated oxidative DNA demethylation**

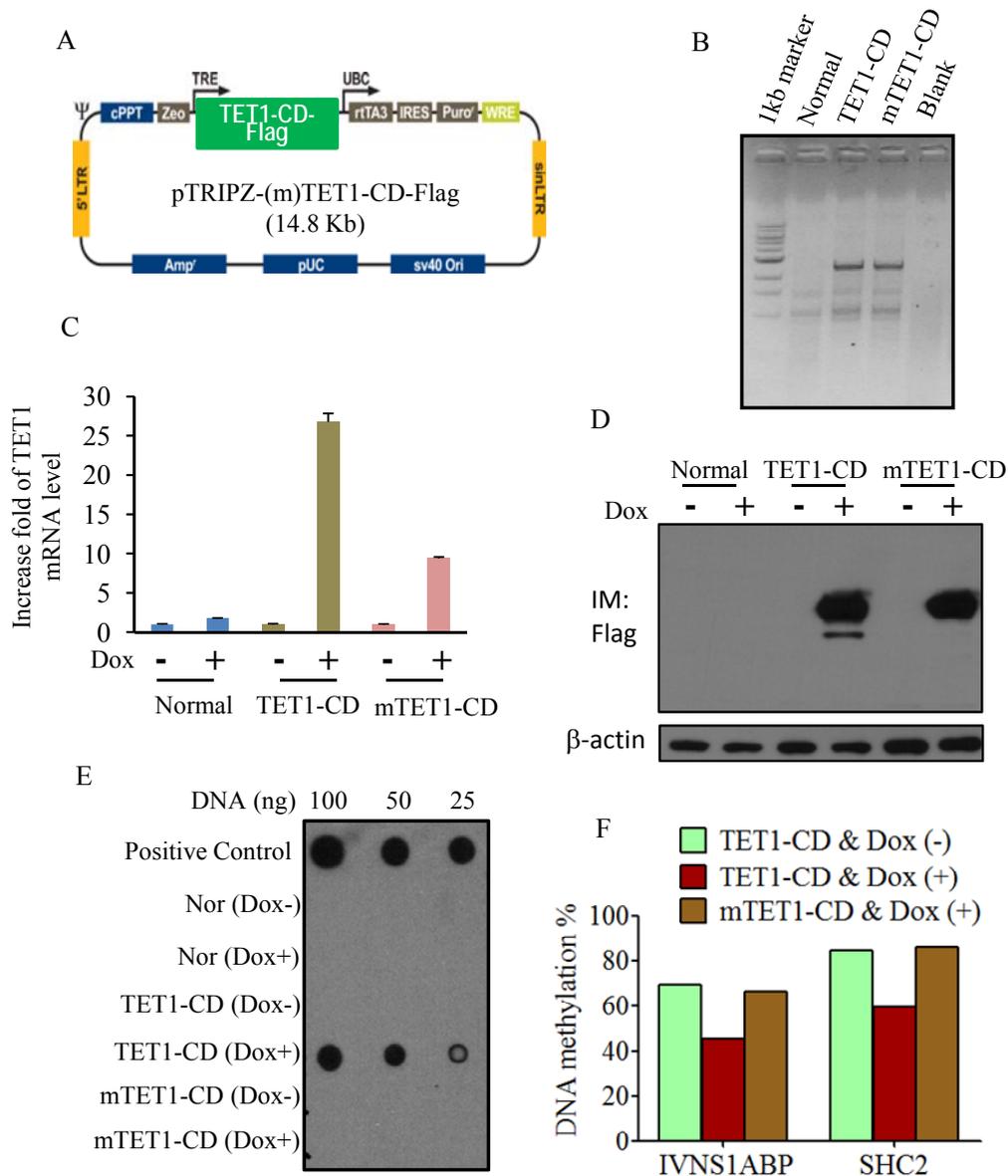
#### **3.3.1 Tetracycline-induced TET1-CD overexpression system in HEK293T cells**

Despite the accumulating evidences that TET-mediated 5mC oxidation leads to DNA demethylation, the underlying mechanism is still unclear. To study it, we developed a tetracycline-induced TET1-CD overexpression system in HEK293T cells. Compared with transient transfection of TET1-CD (or TET2-CD) which were used in previous studies (Guo et al., 2011; He et al., 2011), our inducible overexpression system undoubtedly provides a more convenient, precise and versatile model for our mechanistic study.

The tetracycline-induced (m)TET1-CD overexpression plasmids were constructed by cloning (m)TET1-CD-Flag ORF into lentiviral pTRIPZ vector, which is initially designed for tetracycline-induced shRNA expression. To make two unique restriction sites flanking red fluorescent protein (RFP) coding region, one AgeI restriction sites in RFP coding region was mutated by site-directed mutagenesis. Subsequently, (m)TET1-CD-Flag ORF was easily transferred into the AgeI and MluI sites of non-silencing pTRIPZ control vector (Figure 27A).

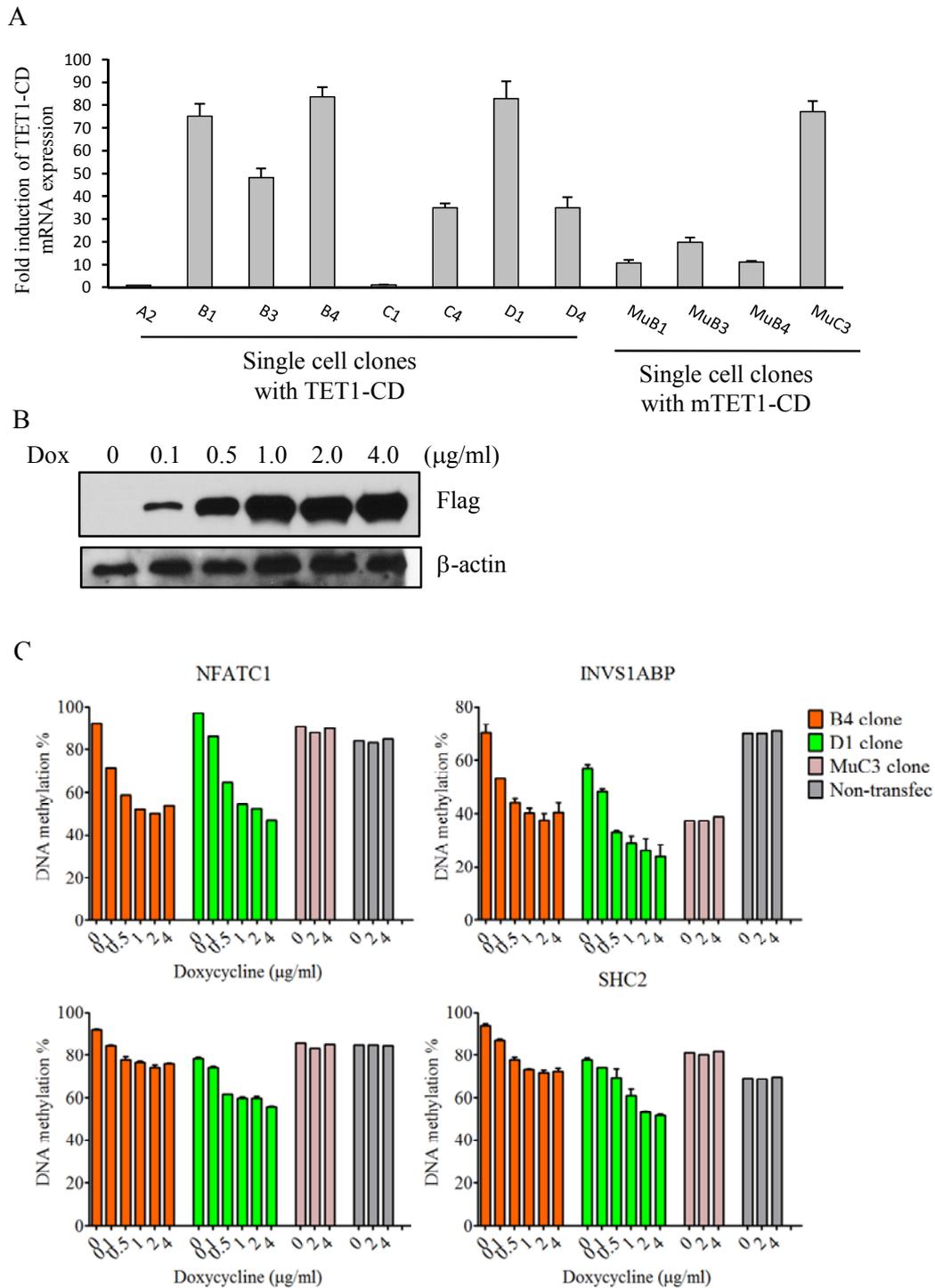
We next transfected the above pTRIPZ-(m)TET1-CD-Flag plasmids into HEK293T according to standard lentiviral transduction protocol. Stable cell pool populations were obtained after one week of puromycin selection. The integration of TET1-CD ORF into genomic DNA was confirmed by PCR genotyping in both pTRIPZ-TET1-CD-Flag and pTRIPZ-mTET1-CD-Flag transduced HEK293T (Figure 27B), and the inducible TET1-CD overexpression was also

validated in those two stable cell pool populations by doxycycline treatment, which significantly increased mRNA and protein level of TET1-CD (Figure 27C-D). To further confirm the function of those overexpressed TET1-CD, we next tested the 5hmC production and also DNA methylation change in IVNS1ABP and SHC2 genes at the presence of doxycycline treatment. As Figure 27E-F show, doxycycline treatment induced marked production of 5hmC and also significant DNA demethylation in TET1-CD but not mTET1-CD transduced HEK293T cells. Moreover, given the heterogeneity of (m)TET1-CD overexpression level in stable cell pool population, subcloning for optimal single cell clones was further performed. As expected, the increase fold of TET1-CD mRNA induced by doxycycline treatment was markedly variable among different clones (Figure 28A), confirming the necessity of single cell subcloning. With this screening of mRNA increase folds, two TET1-CD-overexpressed clones (B4 and D1 clones) and one mTET1-CD-overexpressed clone (MuC3 clone) with the highest increase foldsof (m)TET1-CD mRNA were successfully identified and would be used in subsequent experiments. Moreover, in D1 clone cells the doxycycline-dose-dependent TET1-CD inducible overexpression was also validated (Figure 28B). Consistent with it, doxycycline-dose-dependent DNA demethylation was also observed in both B4 and D1 clones but not MuC3 clone (Figure 28C).Taken together, these above results solidly confirmed the eligibility of our inducible TET1-CD overexpression system for the mechanistic study of TET-mediated DNA demethylation.



**Figure 27. Construction of tetracycline-induced TET1-CD overexpression system in HEK293T cells.**

(A) (m)TET1-CD-Flag ORF was inserted into the AgeI and MluI sites of non-silencing pTRIPZ control vector to realize inducible overexpression of (m)TET1-CD. (B) PCR genotyping confirmed the stable intergration of (m)TET1-CD ORF in genomic DNA in HEK293T cells. (C-D) Inducible TET1-CD overexpression at both mRNA (C) and protein levels (D) with or without 3-day doxycycline treatment (Dox, 1 $\mu$ g/ml, dissolved in sterile water). Error bars represent SD from 2 independent experiments. (E) DNA dot blot shows marked production of 5hmC only in TET1-CD transfected cell pool 3 days after Dox treatment (1 $\mu$ g/ml). Positive control DNA sample comes from TET1-CD transiently transfected and GFP(+) cells (see Figure 8). (F) Bisulfite-pyrosequencing assay for DNA methylation changes of IVNS1ABP and SHC2 before and after Dox treatment (1 $\mu$ g/ml).

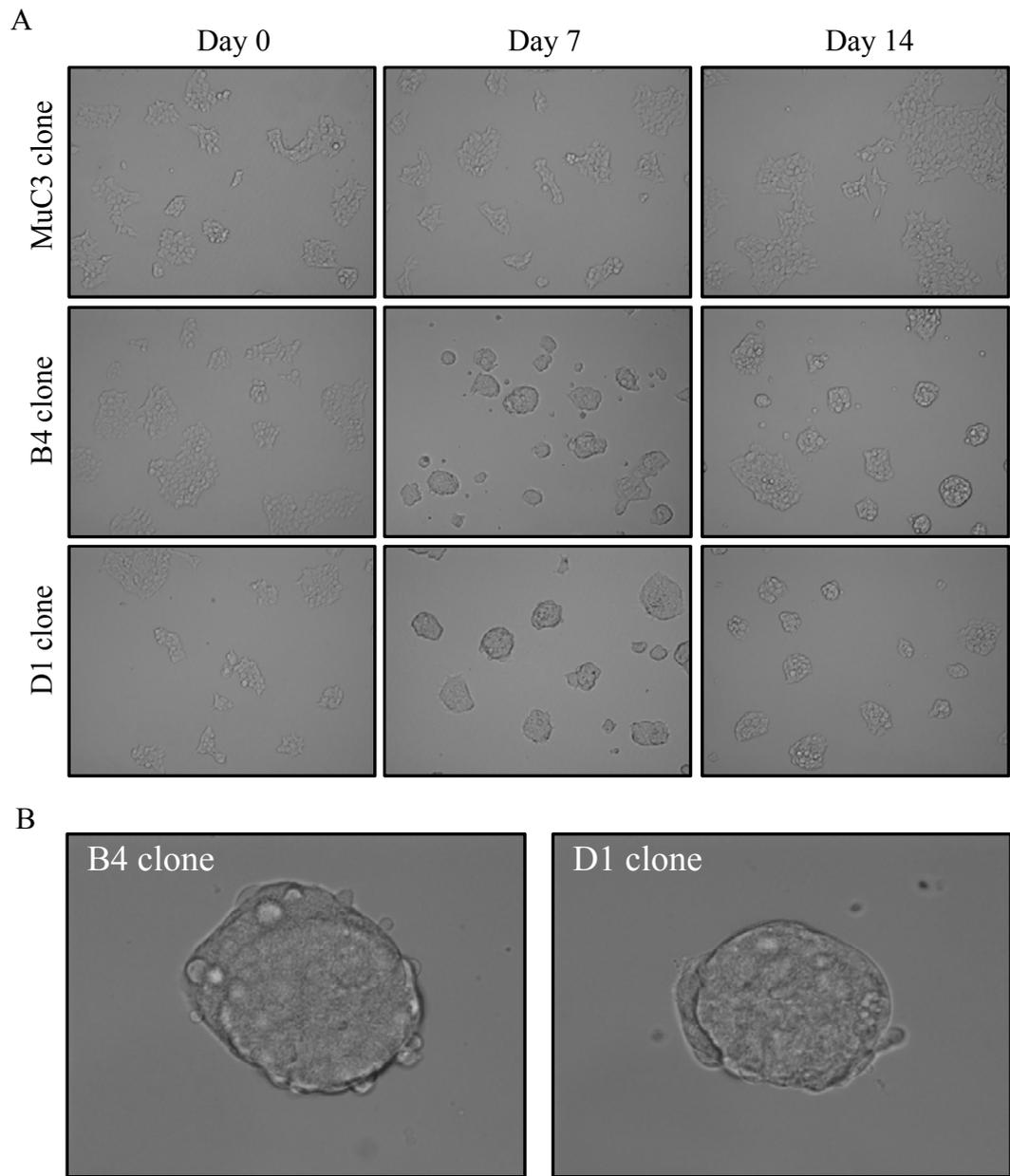


**Figure 28. Characterization of single cell clones with inducible overexpression of TET1-CD and DNA demethylation in specific genomic loci**

(A) Screening of inducible TET1-CD overexpression level among various HEK293T single cell clones transfected with (m)TET1-CD. Cells were treated with doxycycline 1  $\mu\text{g/ml}$  for 3 days. (B) Doxycycline-dose-dependent TET1-CD inducible overexpression in D1 single cell clone. Cells were treated with various dosages of doxycycline for 1 day. (C) Doxycycline-dose-dependent DNA demethylation in both B4 and D1 single cell clones but not MuC3 clone. Error bars represent SD from 2 independent experiments.

### 3.3.2 Long term TET1-CD overexpression and possible dedifferentiation in HEK293T cells

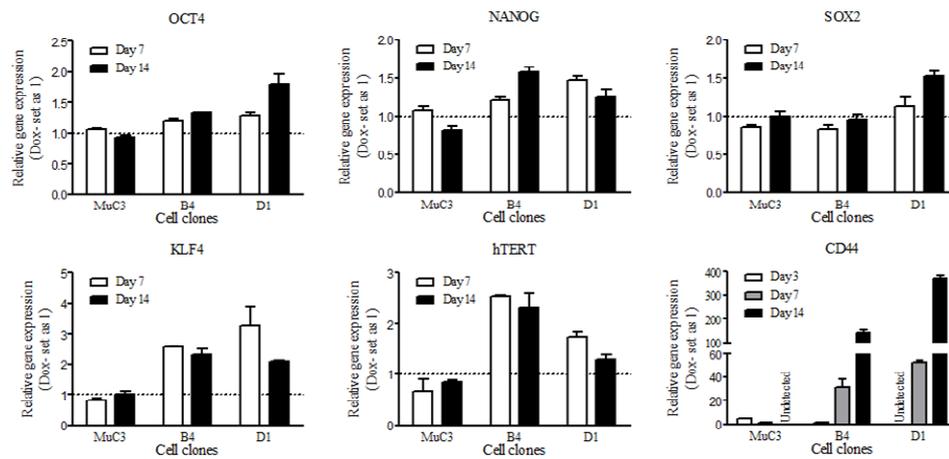
By overexpression of certain key transcription factors (e.g. OCT4, SOX2, C-MYC and KLF4), differentiated somatic cells can be dedifferentiated into induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka, 2006). Before investigating the mechanism of TET-mediated DNA demethylation, we tried to examine whether TET1-CD overexpression can also dedifferentiate somatic cells, given that DNA demethylation may reactivate some stemness genes which are silenced by DNA methylation in somatic cells and inefficient DNA demethylation is one of the causes of the low efficiency in iPSCs generation (Huangfu et al., 2008). Interestingly, a significant morphological change occurred in B4 and D1 but not MuC3 clone cells 7 days after doxycycline treatment (Figure 29A). Lots of cell clusters from doxycycline-treated D1 and B4 cells became more tightly packaged and have a round shape (Figure 29B). Thus, we next tested the expression of some important stemness genes, including *OCT4*, *NANOG*, *SOX2*, *KLF4*, *hTERT* and *CD44*. As Figure 29C shows, except *SOX2* all genes tested showed increased mRNA level to variable extent. Strikingly, *CD44* mRNA level was markedly increased by up to 150 and 380 folds in B4 and D1 clone cells, respectively. Since CD44 is a membrane protein, we next performed flow cytometric analysis and found that ~40% of treated D1 cells became CD44(+) (Figure 29D). To determine if those increased gene expression is related with TET1-CD-mediated DNA demethylation, we further tested the DNA methylation change in *OCT4* and *NANOG* gene promoters. Consistent with their increased gene expressions, doxycycline treatment also induced significant DNA demethylation in those gene promoters (Figure 29E). Thus, these above evidences suggest that TET1-CD may serve as a useful factor for the generation of iPSCs.



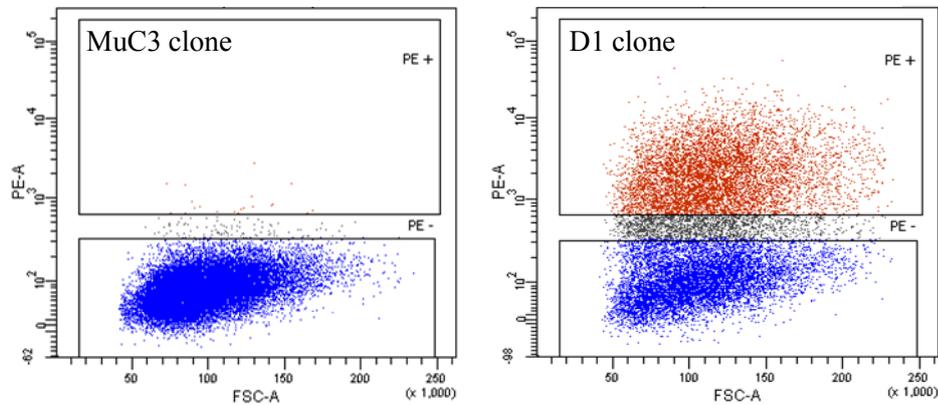
**Figure 29. Long-term overexpression of TET1-CD and potential dedifferentiation in HEK293T cells**

(A) Significant morphological change resulted from overexpression of TET1-CD but not mTET1-CD. MuC3, B4 and D1 single cell clones were cultured in DMEM medium with 10% FBS, 1% P/S, and 2  $\mu\text{g/ml}$  doxycycline. (B) Example of cell clusters observed in B4 and D1 cell clones treated with 2  $\mu\text{g/ml}$  doxycycline for 7 days.

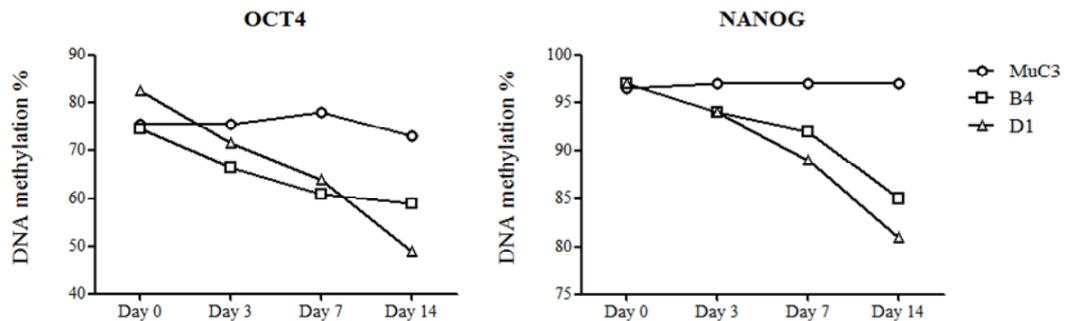
C



D



E



**Figure 29. Long-term overexpression of TET1-CD and potential dedifferentiation in HEK293T cells**

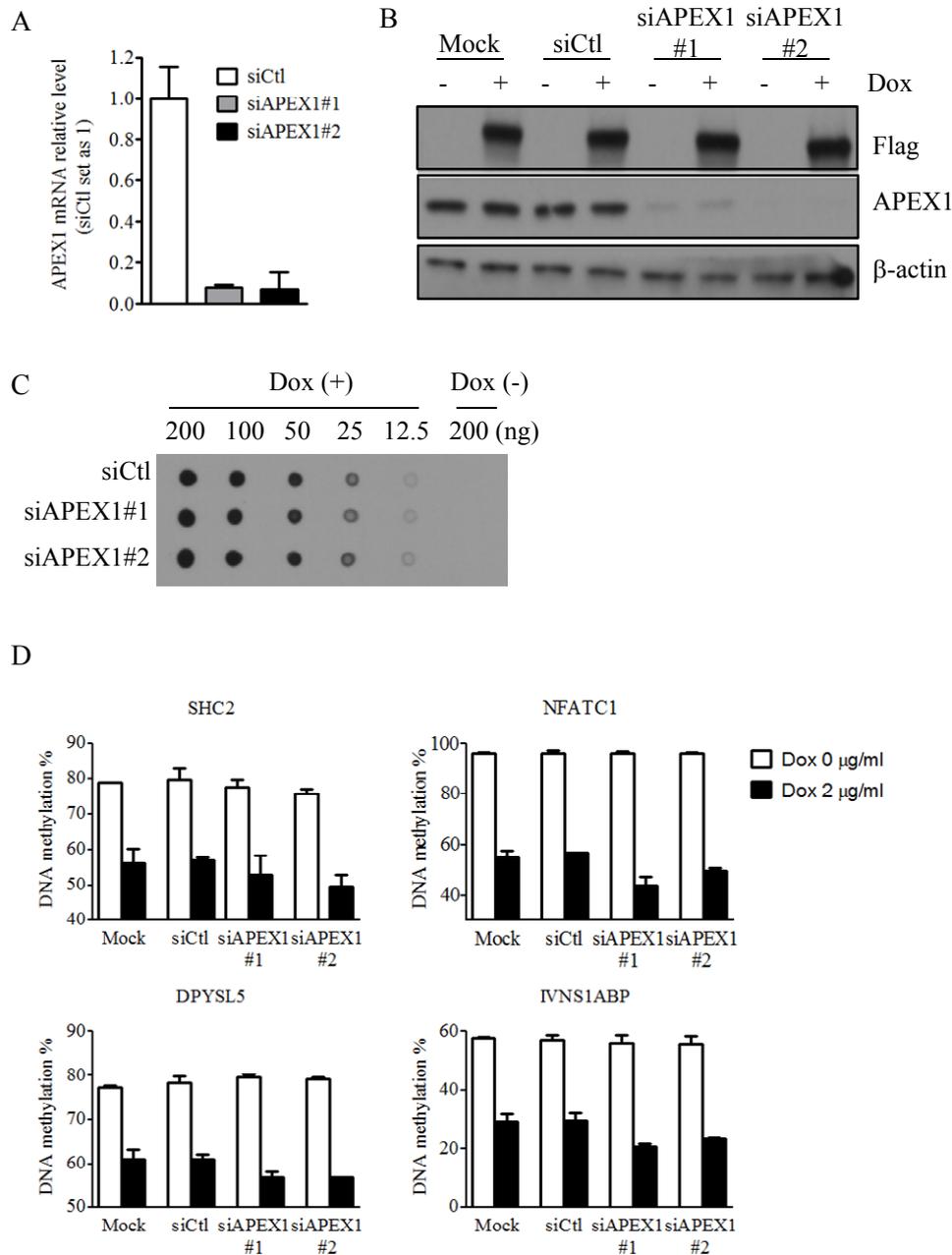
(C) Real-time qPCR analysis of gene expressions of stem-related genes after doxycycline treatment (2  $\mu\text{g/ml}$ ). Data are expressed as the mean  $\pm$  SEM from triplicate tests. (D) Flow cytometric analysis of CD44 (+) MuC3 and D1 clone cells 3 weeks after doxycycline treatment (2  $\mu\text{g/ml}$ ). (E) Bisulfite-pyrosequencing analysis of DNA methylation change at gene promoters of *OCT4* and *NANOG* after (m)TET1-CD overexpression induced by doxycycline treatment (2  $\mu\text{g/ml}$ ).

To investigate whether overexpression of TET1-CD alone is efficient to generate iPSCs, we next cultured doxycycline-treated D1 and B4 cells in iPSC medium (KnockOut™ ESC/iPSC Media Kit, Invitrogen) with mouse embryonic fibroblast cell as feeder cells. However, those cells grew very slowly and most died in two weeks. A possible reason is that long term TET1-CD overexpression not only reactivated some stemness genes, but also significantly impaired the genomic stability by demethylating hypermethylated repetitive sequences. Additionally, HEK293T may be not a good cell line for iPSCs generation. Thus, the use of TET1-CD for generation of iPSCs from HEK293T cells is questionable and future study with TET1-FL and other somatic cells (e.g. human dermal fibroblasts) may be more promising.

### **3.3.3 TET-mediated DNA demethylation is independent of BER**

In mammals BER is responsible for repairing damaged bases and DNA single-strand breaks induced by reactive oxygen species and alkylating agents (Fortini and Dogliotti, 2007). Moreover, BER has also been often reported to be involved in the mechanisms of active DNA demethylation (Wu and Zhang, 2010). To determine the potential mechanism for TET-mediated oxidative DNA demethylation, we thus first investigated whether BER is also required for TET-mediated demethylation with the use of our tetracycline-induced TET1-CD overexpression system.

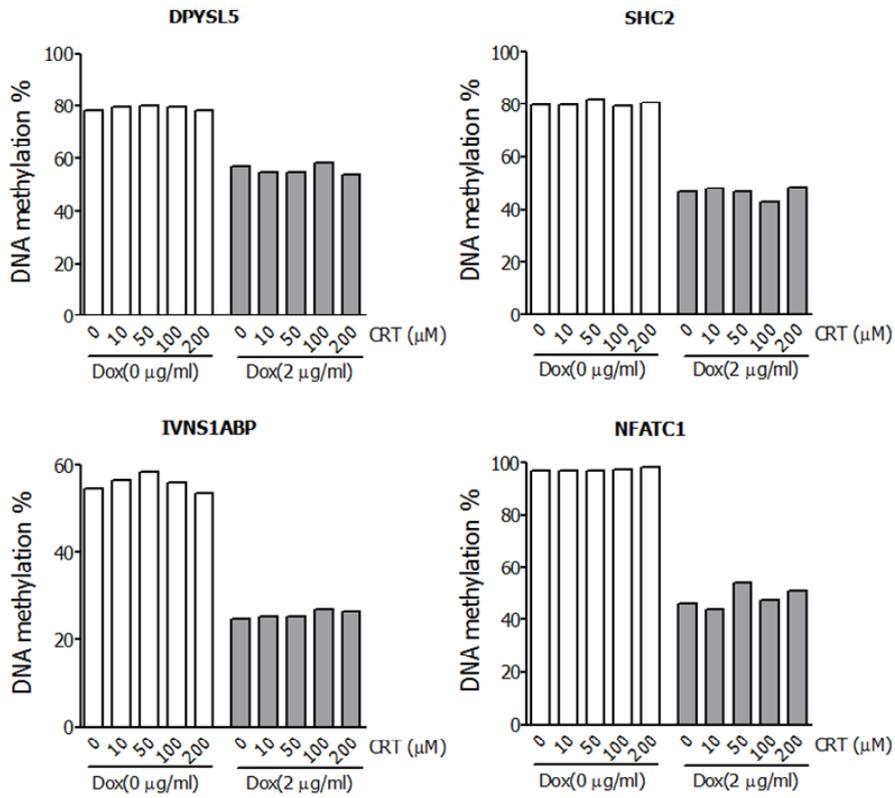
APEX1 is a major AP endonuclease in mammalian cells, plays a central role in BER pathway (Doetsch and Cunningham, 1990). Two different siRNAs for APEX1 were transfected into D1 clone cells and both showed a significant knockdown effect on APEX1 in both mRNA and protein level (Figure 30A-B). Importantly, siRNA treatment also did not affect cell



**Figure 30. DNA demethylation induced by TET1-CD overexpression is independent of APEX1**

(A) Real-time qPCR analysis of siRNA-mediated APEX1 knockdown in HEK293T D1 clone cells. (B) Western blot assay of APEX1 and inducible TET1-CD overexpression in the absence of siAPEX1 and doxycycline (2mg/ml) treatment. (C) DNA dot blot assay of 5hmC production by doxycycline (2mg/ml)-induced TET1-CD overexpression at the presence of siControl or siAPEX1 treatment. (D) Bisulfite-pyrosequencing assay of effect of siAPEX1 on DNA demethylation by doxycycline (2mg/ml)-induced TET1-CD overexpression. Error bars represent SD from 3 independent experiments.

E



**Figure 30. DNA demethylation induced by TET1-CD overexpression is independent on APEX1**

(E) Bisulfite-pyrosequencing assay of effect of APEX1 inhibitor-CRT (CRT0044876, dissolved in DMSO) on DNA demethylation by doxycycline-induced TET1-CD overexpression.

proliferation and inducible TET1-CD expression (Figure 30B). Moreover, no significant change in 5hmC production was observed in doxycycline-treated D1 clone cells after APEX1 knockdown (Figure 30C). However, we surprisingly found that APEX1 knockdown also did not antagonize the DNA demethylation induced by TET1-CD overexpression in DPYSL5, IVNS1ABP, NFATC1 and SHC2 genes, but even showed some potentiating effect (Figure 30D), suggesting that APEX1 and BER are not involved in TET-mediated DNA demethylation. To further confirm it, we next treated D1 clone cells with APEX1 inhibitor-CRT0044876. Consistently, CRT0044876 treatment also did not inhibit the DNA demethylation induced by TET1-CD overexpression (Figure 30E). Thus, our findings strongly suggest that BER is not required for TET-mediated DNA demethylation.

### **3.3.4 TET-mediated DNA demethylation appears to be independent on DNA replication**

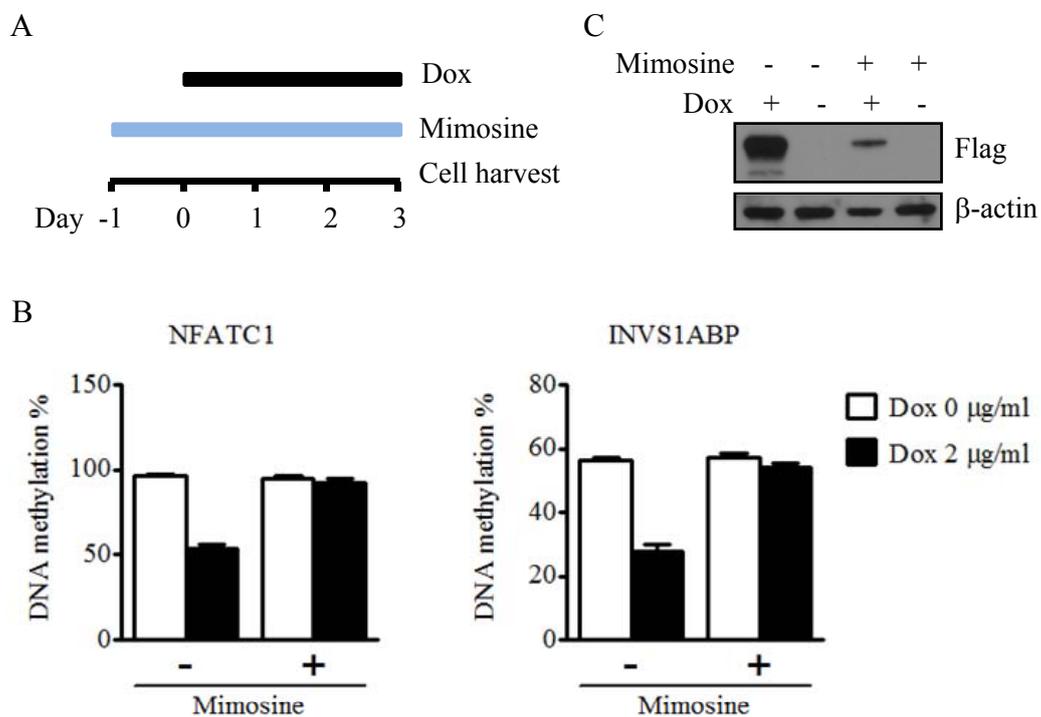
Considering that 5hmC is poorly recognized by DNMT1 (Valinluck and Sowers, 2007), we next asked whether TET-mediated oxidative DNA demethylation is through a replication-dependent passive pathway. HEK293T D1 clone cells were treated with mimosine which can block mammalian cells in late G1 phase and is relatively non-toxic compared with other chemical inhibitors (Jackman and O'Connor, 2001). Mimosine treatment (400  $\mu$ M) resulted in a significant cell proliferation inhibition and also completely abolished DNA demethylation of NFATC1 and INVS1ABP gene promoters by doxycycline-induced TET1-CD overexpression (Figure 31A, 31B). Unfortunately, however, we also found that mimosine treatment also markedly inhibited doxycycline-induced TET1-CD expression (Figure 31C). These results indicated that the strategies of cell cycle arrest may inevitably repress the inducible TET1-CD

overexpression in our model, thus basically impair the use of this model for subsequent mechanistic study. Actually, enforcedly blocking DNA replication and cell division could result in an extensive change of cellular environment and transcriptome, which would in turn inhibit not only the possible passive demethylation but also the potential active demethylation mechanism. Therefore, use of non-dividing cells instead of enforced inhibition of DNA replication in dividing cells appears to be more reasonable to study the passive mechanisms of DNA demethylation.

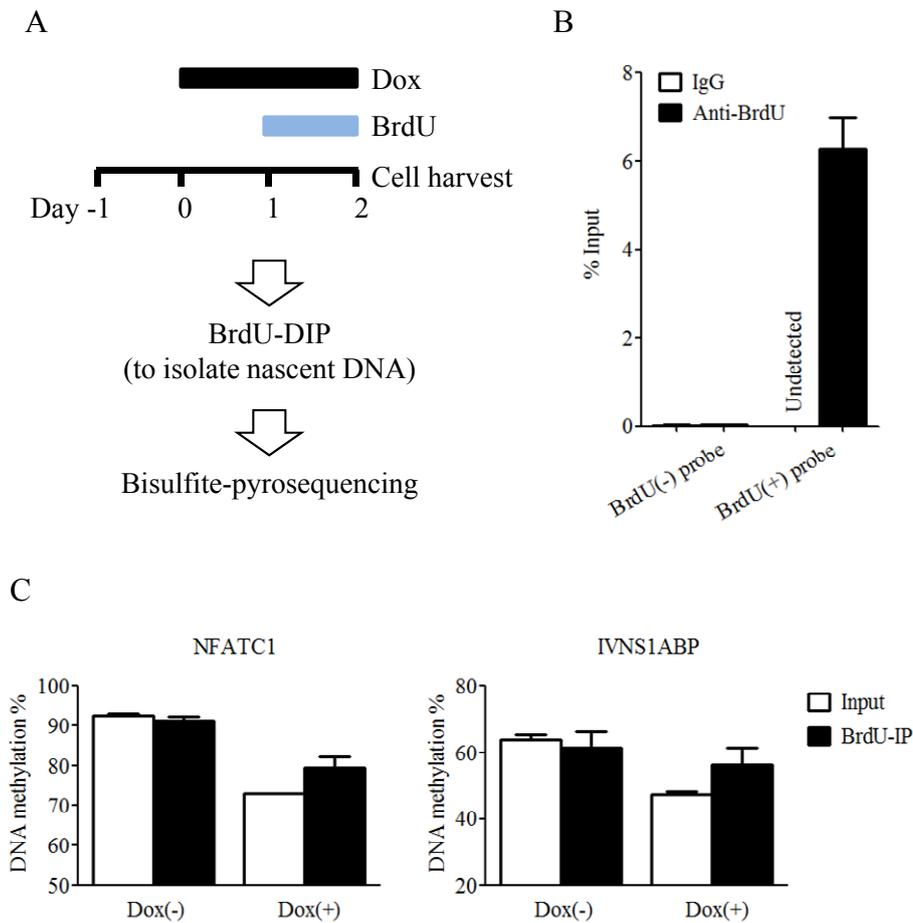
We next tried another strategy to investigate the possibility that replication-dependent passive pathway underlies TET-mediated oxidative DNA demethylation. In this strategy, BrdU was used to mark nascent DNA, which was subsequently isolated by DIP with anti-BrdU antibody, followed by bisulfite-pyrosequencing assay (Figure 32A). The high specificity of BrdU-DIP was validated with the use of DNA probes containing BrdU or not (Figure 32B). Surprisingly, our results showed that BrdU-IP (nascent) DNA was less demethylated than input control DNA at both *IVNS1ABP* and *NFATC1* genes, suggesting that passive DNA demethylation is at least not the primary mechanism for TET-mediated oxidative DNA demethylation (Figure 32C).

In addition, dynamic study of DNA demethylation and 5hmC contents in HEK293T D1 clone cells seemed to also exclude the involvement of passive pathway in TET-mediated DNA demethylation. The cells were treated with doxycycline just in the first three days and then cultured in regular medium (Figure 33A). The inducible TET1-CD expression completely disappeared at day 6, while 5hmC content reached peak at day 3, decreased by ~8 folds at day 6 and then become almost undetected at day 9 (Figure 33B, C). However, we interestingly found that DNA methylation levels in *IVNS1ABP* and *NFATC1* genes bottomed at day 6 and then slightly increased at day 9 and later (Figure 33 D). Given 5hmC content was continuously diluted

during that period, these evidences also suggest that TET-mediated DNA demethylation may not occur through passive pathway. Taken together, those above evidences all suggest that replication-dependent passive pathway is at least not the primary mechanism for TET-mediated oxidative DNA demethylation.

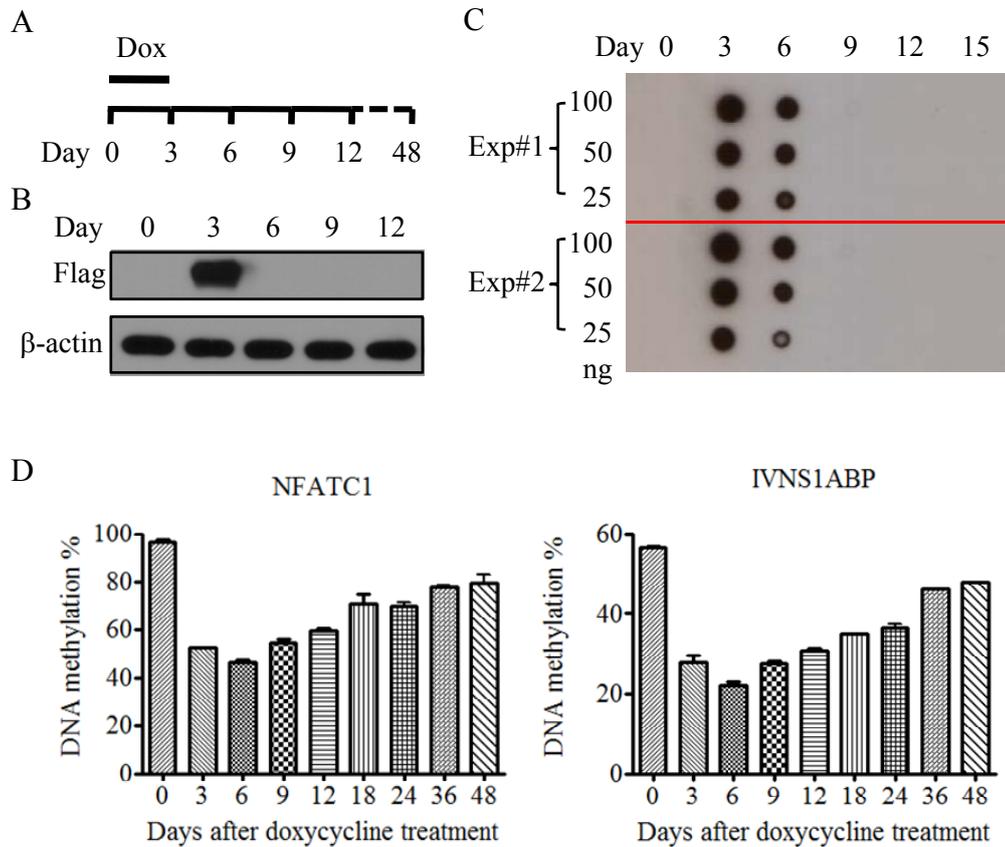


**Figure 31. Mimosine treatment inhibits doxycycline-induced TET1-CD overexpression**  
 (A) Mimosine (400 µM) and doxycycline treatment (2 µg/ml) schedule in HEK293T D1 clone cells. (B) Bisulfite-pyrosequencing analysis of DNA methylation levels after mimosine and/or doxycycline treatment. Error bars represent SD from 2 independent experiments. (C) Inhibition of doxycycline-induced TET1-CD overexpression by mimosine treatment.



**Figure 32. TET1-CD overexpression-induced DNA demethylation may be only through active pathway**

(A) Schematic of the strategy of BrdU-DIP for mechanistic study. HEK293T D1 clone cells were cultured with BudU (50  $\mu$ M) at the presence or absence of doxycycline (2  $\mu$ g/ml). (B) Validation of BrdU-DIP efficiency with the use of DNA probe containing BrdU. Error bar represent SD from 3 independent experiments. (C) Bisulfite-pyrosequencing analysis of DNA methylation with the input and immunoprecipitated DNA. Error bars represent SD from 2 independent experiments.



**Figure 33. Dynamics of inducible TET1-CD expression, genomic 5hmC and DNA methylation levels in doxycycline-treated HEK293T D1 clone cells**

(A) Doxycycline treatment (2  $\mu$ g/ml) schedule in HEK293T D1 clone cells. (B) Western blot assay of the dynamics of TET1-CD protein level. (C) DNA dot blot assay of the dynamics of genomic 5hmC levels. (D) Bisulfite-pyrosequencing analysis of the dynamics of DNA methylation levels after doxycycline treatment. Error bars represent SD from 2 independent experiments.

## **CHAPTER 4 DISCUSSION, SUMMARY, AND FUTURE DIRECTIONS**

### **4.1 Discussion**

#### **TET1 is a unique DNA demethylase to maintain the DNA hypomethylated state in CGIs**

Although the evidences for TET family proteins as DNA demethylases have been accumulated, the specific role of TET1 in DNA methylation pattern still remains unclear (Figuroa et al., 2010; Guo et al., 2011; He et al., 2011; Ito et al., 2011; Tahiliani et al., 2009; Zhang et al., 2010a). Here we surprisingly found that TET1-FL overexpression in HEK293T cells failed to induce significant DNA demethylation in any categorized genomic regions, in spite of a global DNA demethylation observed in TET1-CD-overexpressed cells. Genome-wide mapping of 5hmC further revealed a unique regulation pattern of 5mC by TET1, where its 5hmC production is relatively inhibited as local basal DNA methylation levels increases. By contrast, TET1-CD overexpression showed a strong positive correlation between 5hmC yield and local basal DNA methylation levels. Importantly, such different regulation patterns of 5mC by TET1-FL and TET1-CD well explains their different capability to induce DNA demethylation. Thus, based on these genome-wide analyses results, we not only confirm that 5mC oxidation mediated by TET proteins (or other potential 5mC dioxygenases) can serve as an efficient mechanism for global DNA demethylation, but also demonstrate that TET1 is at least not a potent DNA demethylase which can actively induce global DNA demethylation when overexpressed. For example, both overexpressions of the deaminase/ glycosylase pair AID/MBD4 and Gadd45a, two previously reported DNA demethylase candidates, have been found to elicit a significant global DNA demethylation (Barreto et al., 2007; Rai et al., 2008).

As a CpG DNA binding domain, the CXXC domain in TET1 is believed to play an important role in its demethylating function. In addition to a strong binding to unmethylated CpG dinucleotides as other CXXC domains (e.g. MLL-CXXC and DNMT1-CXXC domains), the CXXC domain of TET1 has also been reported to have considerable binding for methylated CpG dinucleotides in in vitro GST pull-down assay (Xu et al., 2011; Zhang et al., 2010a). Thus, this distinct binding feature of the CXXC domain appears to well fit with the catalytic activity of TET1 to induce DNA demethylation. However, a variety of evidences from other and our studies suggest that it may be not the case. On one hand, through CXXC domain TET1 is specifically enriched at CpG-rich regions in genomic DNA, like CGIs (Williams et al., 2011; Wu et al., 2011; Xu et al., 2011). Given the hypomethylated state in most CGIs, such location feature markedly restricts the catalytic function of TET1 to induce DNA demethylation. On the other hand, we also found that compared with that by TET1-CD, local 5hmC production by TET1-FL was significantly inhibited as basal DNA methylation levels increases, suggesting a poor binding of TET1 to sporadically distributed and methylated CpG dinucleotides in cells. Moreover, our ChIP-qPCR results also reveal that TET1 specifically binds hypomethylated but not hypermethylated CGI promoters through its CXXC domain. This discrepancy of methyl-CpG binding of the CXXC domain between in vitro GST-pull down assay and ChIP assay may be due to the potential interference from other domains in TET1 and also the complex chromatin environment in cells. Particularly, both recruitment of endogenous methyl-CpG binding proteins (e.g. MeCP2) and densely packed heterochromatin state in hypermethylated genomic regions may dramatically inhibit the binding of TET1 to methylated DNA. Therefore, the CXXC domain and 5mC dioxygenase catalytic domain in TET1 appears to form an interesting but conflicting domain

combination, where CXXC domain specifically targets TET1 to hypomethylated regions and consequently prevents its catalytic domain from inducing DNA demethylation in moderately or highly methylated regions.

The immunity of CGIs to DNA methylation is an important characteristic of mammalian DNA methylation pattern, and aberrant methylation in CGIs has been associated with various human diseases, such as cancer (Bird, 2002; Herman and Baylin, 2003). By excluding DNMTs from CGIs several factors have been reported to protect CGIs from de novo methylation, including binding of transcription factors, high transcription activity, and active chromatin mark H3K4me3 (Deaton and Bird, 2011). TET1 preferentially binds hypomethylated but not hypermethylated CGIs through its CXXC domain, strongly suggesting a potential role of TET1 in hypomethylated CGIs. In support of it, we found that overexpression of TET1-FL specifically decreased the DNA methylation level in hypomethylated CGIs, despite no significant DNA demethylation in moderate or highly methylated regions. More importantly, knockdown of TET1 in HEK293T cells induce a significant increase of DNA methylation only in the pre-methylated edges of hypomethylated CGIs, indicating that TET1 maintains the DNA hypomethylation state of CGIs by inhibiting the aberrant spreading of de novo DNA methylation from pre-methylated CGI edges. Thus, TET1 presents a novel DNA demethylase-based mechanism for the maintenance of hypomethylated state in CGIs. Different from previously reported mechanisms by which DNMTs are always excluded from CGIs, this new mechanism utilizes a special DNA demethylase to remove aberrant and stochastic de novo DNA methylation at the methylated CGI edges and consequently maintain the DNA hypomethylated states in CGIs. Undoubtedly, the cooperation of these two mechanisms can provide a more “solid” protection for hypomethylated CGIs against

DNA methylation attack. We note that two recent studies reported increased DNA methylation levels in Tet1-bound regions after depletion of Tet1 in mESCs (Wu et al., 2011; Xu et al., 2011). However, they only suggest a possible role of Tet1 in maintaining the DNA hypomethylation state but do not directly prove a demethylating activity for TET1 (Wu et al., 2011). Moreover, the concomitant impairment of mESCs self-renewal and maintenance also makes it more complex to explain the underlying mechanism for those increase of DNA methylation (Ito et al., 2010).

The unique function of TET1 to maintain the DNA hypomethylation state of CGIs further suggests a potential involvement of TET1 in the pathogenesis of human cancer. It has been well known that aberrant hypermethylation of CGIs constitutes an important epigenetic feature in human cancer and it provides cancer cells with an advantage in cell growth and invasion by silencing various tumor suppressor genes (Robertson, 2005). However, how such DNA hypermethylation occurs in cancer is still unclear. Our findings that TET1 protects hypomethylated CGIs against DNA methylation attack raise a possibility that the aberrant hypermethylation of CGIs in cancer may be closely associated with the mutation or dysregulation of TET1. In support of it, *TET1* has been found to be a fusion partner of MLL in acute myeloid leukemia (Lorsbach et al., 2003; Ono et al., 2002). Moreover, a recent study which generated exome sequences for a set of 72 human colon tumor-normal pairs further revealed that *TET1* was frequently mutated in colon cancer (6/72) (Seshagiri et al., 2012). Additionally, a common and significant decrease of TET1 expression was also reported in human colorectal cancers (Kudo et al., 2012). Therefore, by disrupting the normal DNA hypomethylation state of CGIs, the mutation or down-regulation of *TET1* gene may serve as an essential step in tumorigenesis and tumor progression.

Lastly, our study on TET1 also provides some new insights into our understanding of the functions of TET2 and TET3. Due to the lack of CXXC domain, TET2 may regulate DNA methylation more extensively than TET1. As a result, loss or mutation of TET2 could lead to a more extensive change in DNA methylation pattern. Consistently, the mutation of TET2 has been reported to be associated with a DNA hypermethylation phenotype in acute myeloid leukemia (Figueroa et al., 2010). Further studies on the mechanism and distribution of genomic DNA binding of TET2 are necessary to comprehensively elucidate its role in normal and cancer cells. As for TET3, a number of evidences have shown that it induces the global DNA demethylation in paternal pronuclei of mouse zygotes (Gu et al., 2011; Wossidlo et al., 2011). As the DNA binding domain, the CXXC domain in TET3 must play an essential role in the determination of genomic loci for DNA demethylation. More importantly, in contrast to the function of TET1 to maintain DNA hypomethylated state of CGIs, TET3 actively induces DNA demethylation, strongly suggesting a difference between the CXXC domains of TET1 and TET3.

### **Mechanism for TET-mediated 5mC oxidative DNA demethylation**

With their 5mC dioxygenases catalytic function, TET proteins consecutively oxidize 5mC to 5hmC, 5fC and 5caC. Subsequently, these 5mC derivatives are replaced with unmethylated cytosine to finally complete the process of active DNA demethylation. In contrast to the demonstrated 5mC oxidation reactions by TET, how those 5mC derivatives become unmethylated cytosine has been still unclear. Here we successfully established a tetracycline-induced TET1-CD overexpression system in HEK293T cells, which provided great convenience for our mechanistic study. Like transient TET1-CD transfection, this induced TET1-CD overexpression resulted in

significant DNA demethylation in several genomic loci. BER is the most often reported mechanism for active DNA demethylation in mammal (Wu and Zhang, 2010). However, we found that neither knockdown of APEX1 with siRNA or inhibition of APEX1 with CRT0044876 affected the DNA demethylation induced by TET1-CD overexpression, providing strong evidences that TET-mediated active DNA demethylation is independent on BER. Indeed, one previous study reported that CRT0044876 treatment completely antagonized DNA demethylation in HEK293T cells transiently transfected with TET1-CD (Guo et al., 2011; Zhang et al., 2010a). Compared with that study, our experiment used a more consistent TET1-CD overexpression system, and bisulfite-pyrosequencing rather than *HpaII* sensitivity assay to more accurately test DNA methylation change. Additionally, another previous study reported that 5fC and 5caC in CpG dinucleotides can be recognized and excised by TDG, suggesting that TDG-initiating BER may underlie TET-mediated DNA demethylation (He et al., 2011; Maiti and Drohat, 2011). But actually, no experiments on the role of BER in TET-mediated DNA demethylation were further done in those studies. Therefore, although BER has been reported or proposed to be involved in TET-mediated active DNA demethylation, our study strongly suggests that such TET-mediated DNA demethylation is independent on BER. Moreover, given Tet3 induces global DNA demethylation in the paternal pronuclei of mouse zygotes within several hours after fertilization (Gu et al., 2011; Wossidlo et al., 2011), BER is also unlikely to contribute to Tet3-mediated DNA demethylation as it would put tremendous pressure on the repair machinery of zygotes (Wu and Zhang, 2010).

BER is not required for TET-mediated DNA, but by which mechanism do TET proteins induce active DNA demethylation? By removing carboxyl group from 5caC, a potential

decarboxylase could directly convert 5caC into unmethylated cytosine and thus realize TET-mediated active DNA demethylation. This possibility is supported by thymidine salvage pathway where thymine 7-hydroxylase in fungi oxidizes thymidine into iso-orotate, and then a decarboxylase directly converts iso-orotate into uridine by removing the carboxyl group (Smiley et al., 2005; Warn-Cramer et al., 1983). A genome-wide siRNA screens to identify that potential decarboxylase may be necessary for future mechanistic study of TET-mediated DNA demethylation.

On the other hand, in addition to active demethylation pathway, TET-mediated 5mC oxidation may also lead to DNA demethylation through a replication-dependent passive way, as the derivatives of 5mC (5hmC, 5fC and 5caC) may be poorly recognized by DNMT1 (Inoue et al., 2011; Inoue and Zhang, 2011; Valinluck and Sowers, 2007). However, although a cell division-dependent dilution of genomic 5hmC, 5fC, and 5caC has been observed during early development of mouse embryo (Inoue et al., 2011; Inoue and Zhang, 2011), whether the passive pathway underlies TET-mediated DNA demethylation has not been directly demonstrated. To confirm the possible role of passive pathway in the mechanism of TET-mediated DNA demethylation, we tried various strategies, including cell cycle arrest with mimosine, isolating nascent DNA with BrdU-DIP, and dynamic study of genomic 5hmC as well as DNA methylation. Mimosine treatment successfully blocked cell cycle but also unexpectedly inhibited the inducible overexpression of TET1-CD, thus no DNA demethylation was induced by doxycycline treatment at all. The other two strategies overcame such disadvantage, but surprisingly showed that passive pathway may be at least not the predominant mechanism for TET-mediated DNA demethylation. These unexpected results directly reach to a possibility that 5hmC (and 5fC and 5caC) may be

recognized by DNMT1 as well as 5mC during DNA replication. Actually, among these three derivatives, only 5hmC has been previously observed to be poor recognized by DNMT1 (Inoue et al., 2011; Inoue and Zhang, 2011; Valinluck and Sowers, 2007). But even in that previous study, the reduced selectivity of DNMT1 to 5hmC was only detected in an in vitro DNMT1 methylation protection assay, which used a significantly different reaction situation from that of DNA replication process in cells. Therefore, to further confirm our conclusions, future studies to directly test the selectivity of DNMT1 to 5hmC, 5fC and 5caC during DNA replication in cells are necessary.

## **4.2 Summary**

In conclusion, our study showed that the CXXC and 5mC dioxygenase catalytic domains in TET1 form an interesting domain combination: the CXXC domain specifically targets TET1 towards hypomethylated but not hypermethylated CGI regions, whereas the catalytic domain requires 5mC as substrate for 5hmC production as well as DNA demethylation. As a result, overexpression of TET1-CD but not TET1-FL induced significant global DNA demethylation. On the other hand, TET1 specifically maintains the DNA hypomethylation state in CpG-rich regions by its 5mC dioxygenase enzymatic activity, which unveils a novel DNA demethylase-based mechanism for the maintenance of DNA hypomethylated state in CGIs. Cooperated with other mechanisms through which DNMTs are excluded from CGIs to in order to maintain their DNA hypomethylation state, this novel mechanism provides a more “solid” protection for hypomethylated CGIs against DNA methylation attack. Therefore, our study for the first time revealed that TET1 works as a unique DNA demethylase which does not actively induce DNA

demethylation, but rather specifically maintains the DNA hypomethylation state in CpG-rich regions by its 5mC dioxygenase enzymatic activity.

As for the mechanism of TET-mediated DNA demethylation, we found that knockdown or inhibition of APEX1 did not impair the DNA demethylation induced by TET1-CD overexpression, providing strong evidences that TET-mediated active DNA demethylation is independent of BER. Moreover, through two different strategies, we also found that passive pathway may be at least not the primary mechanism for TET-induced DNA demethylation. Thus, although BER-involved active pathway and replication-dependent passive pathway have been highly proposed, our available results suggest that other possible mechanisms (e.g. 5caC decarboxylases) may underlie the DNA demethylation induced by TET proteins.

### **4.3 Future directions**

As discussed above, we identified TET1 works as a unique DNA demethylase which specifically maintains the DNA hypomethylation state of CGIs by its 5mC dioxygenase enzymatic activity. However, we still do not know how TET1 globally maintains the DNA methylation pattern. Additionally, combined with the accumulating evidences about TET1 mutations in cancer, our finding that TET1 specifically maintains the DNA hypomethylation state of CGIs also suggests a potential involvement of TET1 in the pathogenesis of human cancer. Future studies on the role of TET1 in tumorigenesis and tumor progression may provide a new therapeutic target for cancer. Moreover, in contrast to the function of TET1 to maintain the DNA hypomethylation state of CGIs, TET3 actively induces DNA demethylation. To know whether and how the CXXC domain of TET3 regulates its special function is both interesting and important. On the other hand,

our available results significantly contradict two main hypothesized mechanisms for TET-mediated DNA demethylation: BER-dependent active pathway and DNA replication-dependent passive pathway. Which mechanism really underlies TET-mediated DNA demethylation is still unclear. Therefore, to answer those questions, the following future studies could be necessary:

1. To globally study the effect of TET1 knockdown on DNA methylation pattern by genome-wide bisulfite-sequencing;
2. To investigate the role of TET1 in tumorigenesis and tumor progression: e.g., study whether depletion of *TET1* induces malignant transformation in vitro or tumorigenesis in knockout mice, and also determine whether concomitant *TET1* knockout promotes tumorigenesis in available mouse tumorigenesis model, like the colon tumorigenesis in APC knockout mice.
3. To investigate the role of the CXXC domain in the function of TET3: e.g., solve the crystal structure of CXXC domain by computer modeling, determine its binding performance on (un)methylated CpG motif both in vitro and in vivo, and compare them with those of TET1 CXXC domain.
4. To further study the mechanism of TET-mediated DNA demethylation: test the selectivity of DNMT1 to 5hmC, 5fC and 5caC during DNA replication in cells and identify potential 5caC decarboxylases with genome-wide siRNA screens.

## **BIBLIOGRAPHY**

Antequera, F., and Bird, A. (1999). CpG islands as genomic footprints of promoters that are associated with replication origins. *Curr Biol* 9, R661-667.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* 31, 785-799.

Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* 455, 822-825.

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

Barreto, G., Schafer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Doderlein, G., Maltry, N., Wu, W., Lyko, F., and Niehrs, C. (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* 445, 671-675.

Baylin, S.B., and Herman, J.G. (2000). DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet* 16, 168-174.

Becker, P.B., and Horz, W. (2002). ATP-dependent nucleosome remodeling. *Annu Rev Biochem* 71, 247-273.

Bhattacharya, S.K., Ramchandani, S., Cervoni, N., and Szyf, M. (1999). A mammalian protein with specific demethylase activity for mCpG DNA. *Nature* 397, 579-583.

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

- Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* 40, 91-99.
- Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. *Nature* 321, 209-213.
- Bird, A.P. (1995). Gene number, noise reduction and biological complexity. *Trends Genet* 11, 94-100.
- Birke, M., Schreiner, S., Garcia-Cuellar, M.P., Mahr, K., Titgemeyer, F., and Slany, R.K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Res* 30, 958-965.
- Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317, 1760-1764.
- Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1994). Sp1 elements protect a CpG island from de novo methylation. *Nature* 371, 435-438.
- Bruniquel, D., and Schwartz, R.H. (2003). Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. *Nat Immunol* 4, 235-240.
- Carlson, L.L., Page, A.W., and Bestor, T.H. (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. *Genes Dev* 6, 2536-2541.
- Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., and Li, E. (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol* 23, 5594-5605.
- Chen, Z.X., Mann, J.R., Hsieh, C.L., Riggs, A.D., and Chedin, F. (2005). Physical and functional

interactions between the human DNMT3L protein and members of the de novo methyltransferase family. *J Cell Biochem* 95, 902-917.

Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G., and Li, B.F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* 277, 1996-2000.

Ciccione, D.N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Xu, G., Li, E., and Chen, T. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* 461, 415-418.

Colella, S., Shen, L., Baggerly, K.A., Issa, J.P., and Krahe, R. (2003). Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. *Biotechniques* 35, 146-150.

Comb, M., and Goodman, H.M. (1990). CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. *Nucleic Acids Res* 18, 3975-3982.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., *et al.* (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 146, 67-79.

Cravo, M., Fidalgo, P., Pereira, A.D., Gouveia-Oliveira, A., Chaves, P., Selhub, J., Mason, J.B., Mira, F.C., and Leitao, C.N. (1994). DNA methylation as an intermediate biomarker in colorectal cancer: modulation by folic acid supplementation. *Eur J Cancer Prev* 3, 473-479.

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

Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.W., Page, D.C., and Jaenisch, R. (2011). Tet1 is dispensable for maintaining pluripotency

and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* 9, 166-175.

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

Delgado, S., Gomez, M., Bird, A., and Antequera, F. (1998). Initiation of DNA replication at CpG islands in mammalian chromosomes. *EMBO J* 17, 2426-2435.

Dennis, K., Fan, T., Geiman, T., Yan, Q., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev* 15, 2940-2944.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem* 285, 26114-26120.

Doetsch, P.W., and Cunningham, R.P. (1990). The enzymology of apurinic/apyrimidinic endonucleases. *Mutat Res* 236, 173-201.

Dong, K.B., Maksakova, I.A., Mohn, F., Leung, D., Appanah, R., Lee, S., Yang, H.W., Lam, L.L., Mager, D.L., Schubeler, D., *et al.* (2008). DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J* 27, 2691-2701.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res* 10, 2709-2721.

El Kharroubi, A., Piras, G., and Stewart, C.L. (2001). DNA demethylation reactivates a subset of imprinted genes in uniparental mouse embryonic fibroblasts. *J Biol Chem* 276, 8674-8680.

Engel, N., Tront, J.S., Erinle, T., Nguyen, N., Latham, K.E., Sapienza, C., Hoffman, B., and Liebermann, D.A. (2009). Conserved DNA methylation in *Gadd45a*(<sup>-/-</sup>) mice. *Epigenetics* 4,

98-99.

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H., and Bergman, Y. (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nat Struct Mol Biol* *15*, 1176-1183.

Falnes, P.O., Johansen, R.F., and Seeberg, E. (2002). AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* *419*, 178-182.

Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* *473*, 398-402.

Figueroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., *et al.* (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* *18*, 553-567.

Fortini, P., and Dogliotti, E. (2007). Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair (Amst)* *6*, 398-409.

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

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. *J Mol Biol* *196*, 261-282.

Gibbons, R.J., McDowell, T.L., Raman, S., O'Rourke, D.M., Garrick, D., Ayyub, H., and Higgs,

D.R. (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat Genet* 24, 368-371.

Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.F., Liu, W., Xie, Z.G., Shi, L., He, X., Jin, S.G., *et al.* (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 477, 606-610.

Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 145, 423-434.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* 117, 15-23.

Harikrishnan, K.N., Chow, M.Z., Baker, E.K., Pal, S., Bassal, S., Brasacchio, D., Wang, L., Craig, J.M., Jones, P.L., Sif, S., and El-Osta, A. (2005). Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. *Nat Genet* 37, 254-264.

Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* 129, 1983-1993.

Hayatsu, H., Wataya, Y., Kai, K., and Iida, S. (1970). Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry* 9, 2858-2865.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., *et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303-1307.

Hedges, D.J., and Deininger, P.L. (2007). Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat Res* 616, 46-59.

Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A., and Bird, A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev* 15, 710-723.

Herman, J.G., and Baylin, S.B. (2003). Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* 349, 2042-2054.

Hermann, A., Goyal, R., and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem* 279, 48350-48359.

Hodges, E., Smith, A.D., Kendall, J., Xuan, Z., Ravi, K., Rooks, M., Zhang, M.Q., Ye, K., Bhattacharjee, A., Brizuela, L., *et al.* (2009). High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Res* 19, 1593-1605.

Howlett, S.K., and Reik, W. (1991). Methylation levels of maternal and paternal genomes during preimplantation development. *Development* 113, 119-127.

Huang, Y., Pastor, W.A., Shen, Y., Tahiliani, M., Liu, D.R., and Rao, A. (2010). The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One* 5, e8888.

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E., and Melton, D.A. (2008). Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat Biotechnol* 26, 795-797.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R., James, K.D., Turner, D.J.,

- Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* 6.
- Inoue, A., Shen, L., Dai, Q., He, C., and Zhang, Y. (2011). Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res* 21, 1670-1676.
- Inoue, A., and Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* 334, 194.
- Iqbal, K., Jin, S.G., Pfeifer, G.P., and Szabo, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc Natl Acad Sci U S A* 108, 3642-3647.
- Issa, J.P. (2000). CpG-island methylation in aging and cancer. *Curr Top Microbiol Immunol* 249, 101-118.
- Issa, J.P. (2004). CpG island methylator phenotype in cancer. *Nat Rev Cancer* 4, 988-993.
- Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466, 1129-1133.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300-1303.
- Jackman, J., and O'Connor, P.M. (2001). Methods for synchronizing cells at specific stages of the cell cycle. *Curr Protoc Cell Biol Chapter 8*, Unit 8 3.
- Jelinic, P., and Shaw, P. (2007). Loss of imprinting and cancer. *J Pathol* 211, 261-268.
- Jeong, S., Liang, G., Sharma, S., Lin, J.C., Choi, S.H., Han, H., Yoo, C.B., Egger, G., Yang, A.S.,

and Jones, P.A. (2009). Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. *Mol Cell Biol* 29, 5366-5376.

Jin, S.G., Guo, C., and Pfeifer, G.P. (2008). GADD45A does not promote DNA demethylation. *PLoS Genet* 4, e1000013.

Jones, P.A., and Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet* 10, 805-811.

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

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* 429, 900-903.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., *et al.* (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 468, 839-843.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929-930.

Kudo, Y., Tateishi, K., Yamamoto, K., Yamamoto, S., Asaoka, Y., Ijichi, H., Nagae, G., Yoshida, H., Aburatani, H., and Koike, K. (2012). Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer Sci* 103, 670-676.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., *et al.* (2008). DNA methylation of retrotransposon genes is

regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* 22, 908-917.

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

Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., Ogura, A., and Ishino, F. (2002). Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development* 129, 1807-1817.

Lee, J.H., and Skalnik, D.G. (2005). CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J Biol Chem* 280, 41725-41731.

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

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature* 366, 362-365.

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

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

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

Liu, W.M., Marais, R.J., Rubin, C.M., and Schmid, C.W. (1994). Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. *Nucleic Acids Res* 22, 1087-1095.

Lorsbach, R.B., Moore, J., Mathew, S., Raimondi, S.C., Mukatira, S.T., and Downing, J.R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* 17, 637-641.

Ma, D.K., Jang, M.H., Guo, J.U., Kitabatake, Y., Chang, M.L., Pow-Anpongkul, N., Flavell, R.A., Lu, B., Ming, G.L., and Song, H. (2009). Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. *Science* 323, 1074-1077.

Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes Dev* 8, 2282-2292.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* 286, 35334-35338.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., *et al.* (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466, 253-257.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature* 403, 501-502.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X.,

Bernstein, B.E., Nusbaum, C., Jaffe, D.B., *et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766-770.

Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G., *et al.* (2008). Cyclical DNA methylation of a transcriptionally active promoter. *Nature* 452, 45-50.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553-560.

Millar, C.B., Guy, J., Sansom, O.J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P.D., Bishop, S.M., Clarke, A.R., and Bird, A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* 297, 403-405.

Miniou, P., Jeanpierre, M., Blanquet, V., Sibella, V., Bonneau, D., Herbelin, C., Fischer, A., Niveleau, A., and Viegas-Pequignot, E. (1994). Abnormal methylation pattern in constitutive and facultative (X inactive chromosome) heterochromatin of ICF patients. *Hum Mol Genet* 3, 2093-2102.

Mohandas, T., Sparkes, R.S., and Shapiro, L.J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* 211, 393-396.

Morris, K.V., Chan, S.W., Jacobsen, S.E., and Looney, D.J. (2004). Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* 305, 1289-1292.

Morrison, J.R., Paszty, C., Stevens, M.E., Hughes, S.D., Forte, T., Scott, J., and Rubin, E.M. (1996). Apolipoprotein B RNA editing enzyme-deficient mice are viable despite alterations in lipoprotein metabolism. *Proc Natl Acad Sci U S A* 93, 7154-7159.

Myant, K., and Stancheva, I. (2008). LSH cooperates with DNA methyltransferases to repress transcription. *Mol Cell Biol* 28, 215-226.

Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., and Saitou, M. (2009). A signaling principle for the specification of the germ cell lineage in mice. *Cell* 137, 571-584.

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

Okitsu, C.Y., and Hsieh, C.L. (2007). DNA methylation dictates histone H3K4 methylation. *Mol Cell Biol* 27, 2746-2757.

Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., and Hayashi, Y. (2002). LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Res* 62, 4075-4080.

Ooi, S.K., and Bestor, T.H. (2008). The colorful history of active DNA demethylation. *Cell* 133, 1145-1148.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D., *et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448, 714-717.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* 10, 475-478.

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep* 10, 1235-1241.

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

Patel, S.A., Graunke, D.M., and Pieper, R.O. (1997). Aberrant silencing of the CpG island-containing human O6-methylguanine DNA methyltransferase gene is associated with the loss of nucleosome-like positioning. *Mol Cell Biol* 17, 5813-5822.

Peterson, C.L. (2002). Chromatin remodeling enzymes: taming the machines. Third in review series on chromatin dynamics. *EMBO Rep* 3, 319-322.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 463, 1101-1105.

Prendergast, G.C., Lawe, D., and Ziff, E.B. (1991). Association of Myn, the murine homolog of max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. *Cell* 65, 395-407.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* 135, 1201-1212.

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

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., *et al.* (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* 102,

565-575.

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

Riggs, A.D., and Jones, P.A. (1983). 5-methylcytosine, gene regulation, and cancer. *Adv Cancer Res* *40*, 1-30.

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

Robertson, K.D., and Wolffe, A.P. (2000). DNA methylation in health and disease. *Nat Rev Genet* *1*, 11-19.

Sado, T., Fenner, M.H., Tan, S.S., Tam, P., Shioda, T., and Li, E. (2000). X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev Biol* *225*, 294-303.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev Biol* *241*, 172-182.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* *103*, 1412-1417.

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., *et al.* (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* *39*, 232-236.

Sequeira-Mendes, J., Diaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gomez, M. (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells.

PLoS Genet 5, e1000446.

Seshagiri, S., Stawiski, E.W., Durinck, S., Modrusan, Z., Storm, E.E., Conboy, C.B., Chaudhuri, S., Guan, Y., Janakiraman, V., Jaiswal, B.S., *et al.* (2012). Recurrent R-spondin fusions in colon cancer. *Nature* 488, 660-664.

Shames, D.S., Minna, J.D., and Gazdar, A.F. (2007). DNA methylation in health, disease, and cancer. *Curr Mol Med* 7, 85-102.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., *et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450, 908-912.

Si, J., Bumber, Y.A., Shu, J., Qin, T., Ahmed, S., He, R., Jelinek, J., and Issa, J.P. (2010). Chromatin remodeling is required for gene reactivation after decitabine-mediated DNA hypomethylation. *Cancer Res* 70, 6968-6977.

Smallwood, A., Esteve, P.O., Pradhan, S., and Carey, M. (2007). Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev* 21, 1169-1178.

Smiley, J.A., Kundracik, M., Landfried, D.A., Barnes, V.R., Sr., and Axhemi, A.A. (2005). Genes of the thymidine salvage pathway: thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and iso-orotate decarboxylase from *Neurospora crassa*. *Biochim Biophys Acta* 1723, 256-264.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J Biol Chem* 279, 27816-27823.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer,

L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930-935.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663-676.

Takekoshi, H., Yamashita, S., Shimazu, T., Niwa, T., and Ushijima, T. (2009). The presence of RNA polymerase II, active or stalled, predicts epigenetic fate of promoter CpG islands. *Genome Res* 19, 1974-1982.

Tucker, K.L. (2001). Methylated cytosine and the brain: a new base for neuroscience. *Neuron* 30, 649-652.

Valinluck, V., and Sowers, L.C. (2007). Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res* 67, 946-950.

Venolia, L., Gartler, S.M., Wassman, E.R., Yen, P., Mohandas, T., and Shapiro, L.J. (1982). Transformation with DNA from 5-azacytidine-reactivated X chromosomes. *Proc Natl Acad Sci U S A* 79, 2352-2354.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M., *et al.* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439, 871-874.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 20, 116-117.

Warn-Cramer, B.J., Macrander, L.A., and Abbott, M.T. (1983). Markedly different ascorbate dependencies of the sequential alpha-ketoglutarate dioxygenase reactions catalyzed by an essentially homogeneous thymine 7-hydroxylase from *Rhodotorula glutinis*. *J Biol Chem* 258,

10551-10557.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 37, 853-862.

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

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473, 343-348.

Woodcock, D.M., Lawler, C.B., Linsenmeyer, M.E., Doherty, J.P., and Warren, W.D. (1997). Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *J Biol Chem* 272, 7810-7816.

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun* 2, 241.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* 473, 389-393.

Wu, H., and Zhang, Y. (2011). Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev* 25, 2436-2452.

Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. *Nat Rev*

Mol Cell Biol *11*, 607-620.

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., *et al.* (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol Cell* *42*, 451-464.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* *13*, 335-340.

Zhan, Q. (2005). Gadd45a, a p53- and BRCA1-regulated stress protein, in cellular response to DNA damage. *Mutat Res* *569*, 133-143.

Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., and Shi, Y.G. (2010a). TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res* *20*, 1390-1393.

Zhang, H., and Zhu, J.K. (2011). RNA-directed DNA methylation. *Curr Opin Plant Biol* *14*, 142-147.

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., and Jeltsch, A. (2010b). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res* *38*, 4246-4253.

Zhu, B., Zheng, Y., Angliker, H., Schwarz, S., Thiry, S., Siegmann, M., and Jost, J.P. (2000a). 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. *Nucleic Acids Res* *28*, 4157-4165.

Zhu, B., Zheng, Y., Hess, D., Angliker, H., Schwarz, S., Siegmann, M., Thiry, S., and Jost, J.P. (2000b). 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA

glycosylase associated with the chicken embryo DNA demethylation complex. *Proc Natl Acad Sci U S A* 97, 5135-5139.

Zhu, H., Geiman, T.M., Xi, S., Jiang, Q., Schmidtman, A., Chen, T., Li, E., and Muegge, K. (2006). Lsh is involved in de novo methylation of DNA. *EMBO J* 25, 335-345.

Zhu, J.K. (2009). Active DNA demethylation mediated by DNA glycosylases. *Annu Rev Genet* 43, 143-166.

## VITA

Chunlei Jin was born in Wenzhou, Zhejiang, China on October 7, 1980 to parents Bingyun Jin and Caimei Zhang. He is an alumnus of the Medical School of Zhejiang University, where he obtained a Bachelor of Medicine and Bachelor of Surgery degree in 2003. After one year of clinical internship in People's Hospital of Zhejiang Province and Zhejiang Child's Hospital from 2002 to 2003, he began his graduate study of neuroscience in the Department of Pharmacology in the same school, and earned an M.S. degree in 2006. He then worked as a clinical researcher for new drug development at Hangzhou Jiuyuan Gene Engineering Co., Ltd. from 2006 to 2008. From 2008 August, he began to pursue a Ph.D. degree in the University of Texas, Graduate School of Biomedical Sciences at Houston.

Permanent address:

Room 8C, Hongying Building, Danan Road

Wenzhou, Zhejiang, 325000, China