Purdue University
Purdue e-Pubs

Open Access Dissertations | Theses and Dissertations

January 2016

APPROACHING DNA METHYLATION AT THE NANOSCALE

Yi Cui
Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_dissertations

Recommended Citation
Cui, Yi, "APPROACHING DNA METHYLATION AT THE NANOSCALE" (2016). *Open Access Dissertations*. 1282.
https://docs.lib.purdue.edu/open_access_dissertations/1282

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Yi Cui

Entitled

APPROACHING DNA METHYLATION AT THE NANOSCALE

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Joseph Irudayaraj
Chair

Jenna Rickus

David Umulis

Jean-Christophe Rochet

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Joseph Irudayaraj

Approved by: Bernard Engel 4/8/2016

Head of the Departmental Graduate Program Date

APPROACHING DNA METHYLATION AT THE NANOSCALE

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Yi Cui

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2016

Purdue University

West Lafayette, Indiana

I would like to dedicate this work to my ever beloved grandparents, as well as to my wife and family for their unconditional support and endless love.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Dr. Joseph Irudayaraj for providing me with the opportunities and resources to conduct the cutting edge research of biological engineering. His guidance, encouragement, and knowledge have cultivated me to be an independent scientist and well prepared me for the next journey in life. His philosophy of scientific research has been a continuous inspiration, motivating me to aim higher, strive harder, and achieve better. The years working with him are my invaluable experience and life-long treasure.

I would also like to thank my committee members, Dr. Jenna Rickus, Dr. David Umulis, and Dr. Jean-Christophe Rochet, for their insightful comments and support.

I want to thank my fellow labmates and colleagues, Dr. Jiji Chen, Dr. Kyuwan Lee, Dr. Jing Liu, Dr. Basudev Chowdhury, Dr. Ulhas Kadam, Dr. Il-Hoon Cho, Dr. Asia Naz, Dr. Xiaolei Wang, Dr. Samrat Roy Choudhury, Dr. Zhongwu Zhou, Dr. Chaoqing Dong, Dr. Wen Ren, Dr. Seunghyun Lee, Yi Fang, Nur Damayanti, Pushpak Bhandari, and Wenjie Liu, for all the unforgettable memories and cheering moments we have had over the past years.

Last but not least, I want to give my heartfelt gratitude to my family for their accompanying. All the days and nights, overcome struggles and obstacles, shared happiness and joy, have made this dissertation a product of love.

Declaration of Collaborative Work

Chapter 1 contains both figures and text generated and written by myself.

Chapter 2 contains both figures and text from a published manuscript entitled: "Real-time dynamics of methyl-CpG-binding domain protein 3 and its role in DNA de-methylation by fluorescence correlation spectroscopy" [2013, *Epigenetics*, 8 (10): 1089-1100]. This article is a product of joint efforts between Dr. Il-Hoon Cho (Eulji University, Republic of Korea), Dr. Basudev Chowdhury (Purdue University), and me. Dr. Cho and Dr. Chowdhury conceptualized and optimized the enhanced ELISA method that involves information presented in Figures 2.3, 2.5, and 2.6. The final paper was composed and written by me.

Chapter 3 contains both figures and text from a published manuscript entitled: "Dissecting the behavior and function of MBD3 in DNA methylation homeostasis by single-molecule spectroscopy and microscopy" [2015, *Nucleic Acids Research*, 43 (6): 3046-3055]. This project was completed by myself.

Chapter 4 contains both figures and text from an unpublished manuscript entitled: "Regulatory landscape and clinical implication of MBD3 in human malignant glioma" [2016, submitted]. This article is a product of joint efforts between Dr. Jian Li (XiangYa Hospital of Central South University, China), Dr. Sara E. Wirbisky, Dr. Jennifer Freeman (Purdue University), and me. Dr. Wirbisky and Dr. Freeman helped me in designing and performing the microarray experiments (Figures 4.2C); Dr. Li contributed to the *in vivo* and clinical studies (Figures 4.1C, 4.9C, 4.10C-D, 4.11, and 4.12). The final paper was composed and written by me.

Chapter 5 contains both figures and text from a published manuscript entitled:

"Decitabine nanoconjugate sensitizes human glioblastoma cells to temozolomide" [2015, *Molecular Pharmaceutics*, 12 (4): 1279-1288]; and from an unpublished manuscript entitled: "PLGA-PEG nano-delivery system for epigenetic therapy" [2016, in revision]. These two articles are products of joint efforts between Dr. Asia Naz (University of Karachi, Pakistan), Dr. Christopher J. Collins, Dr. David H. Thompson (Purdue University), and me. Dr. Naz contributed to the chemical synthesis and optimization parts (Figures 5.2 and 5.7B). Dr. Naz, Dr. Collins, and Dr. Thompson worked together for the compound characterization (Figures 5.3, 5.4, and 5.5; Table 5.2). The *in vivo* mouse experiments were performed with help from the Purdue Center for Cancer Research Biological Evaluation Core Facility (Dr. Bennett Elzey's group). These two papers were composed and written by Dr. Naz and me (equal contribution).

Chapter 6 contains both figures and text from a published manuscript entitled: "Quantitative real-time kinetics of optogenetic proteins CRY2 and CIB1/N using single-molecule tools" [2014, *Analytical Biochemistry*, 458C: 58-60]; and from an unpublished manuscript entitled: "Optogenetic manipulation of site-specific subtelomeric DNA methylation" [2015, in revision]. These two articles are products of joint efforts between Dr. Samrat Roy Choudhury and me. Dr. Choudhury contributed to the molecular cloning and pyrosequencing (Figures 6.9A and 6.11; Tables 6.1 and 6.2). The content included in this chapter was mainly composed and written by myself.

Chapter 7 contains both figures and text generated and written by myself.

For all the conducted projects, I have received continuous support and mentoring from my supervisor Dr. Joseph Irudayaraj and my graduate committee members.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

LIST OF ABBREVIATIONS

5caC	5-carboxylcytosine
5fC	5-formalcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
AZA	5-azacitydine
BBB	Blood-brain barrier
BER	Base excision repair
Cas9	CRISPR-associated endonuclease 9
CGI	CpG island
CIB1	Cryptochrome interacting basic-helix-loop-helix protein 1
CIITA	MHC class II transactivator
CNS	Central nervous system
Co-IP	Co-immunoprecipitation
CRISPR	Clustered regularly interspaced short palindromic repeats
CRY2	Cryptochrome 2
DAC	5-aza-2'-deoxycitydine
DNMT	DNA methyltransferase

ELISA	Enzyme-linked immunosorbent assay
EPR	Enhanced permeation and retention effect
FCS	Fluorescence correlation spectroscopy
FLCS	Fluorescence lifetime correlation spectroscopy
FLIM	Fluorescence lifetime imaging microscopy
FRET	Förster resonance energy transfer
GBM	Glioblastoma multiforme
GPC	Gel permeation chromatography
HRP	Horseradish peroxidase
IHC	Immunohistochemistry
IP injection	Intraperitoneal injection
lncRNA	Long non-coding RNA
MBD	Methyl-CpG-binding domain protein
MHC	Major histocompatible complex
miRNA	Micro RNA
MSP	Methylation-specific PCR
NMR	Nuclear magnetic resonance spectroscopy
NuRD	Nucleosome remodeling deacetylase
OS	Overall survival
PCB	Phycocyanobilin
PCH	Photon counting histogram
PEG	Polyethylene glycol

PFS	Progression-free survival
PHYB	Phytochrome B
PIE	Pulsed interleaved excitation
PIF6	Phytochrome interaction factor 6
PLGA	Poly(lactic-co-glycolic acid)
qRT-PCR	Quantitative real-time polymerase chain reaction
RES	Reticuloendothelial system
siRNA	Small interfering RNA
SNR	Signal-to-noise ratio
SPAD	Single-photon avalanche diode
TALE	Transcription activator-like effector
TCSPC	Time-correlated single-photon counting
TDG	Thymine DNA glycosylase
TEM	Transmission electron microscopy
TET	Ten-eleven-translocation protein
TRF1	Telomeric repeat binding factor 1
WHO grade	World health organization tumor grade
ZFP	Zinc finger protein

ABSTRACT

Cui, Yi. Ph.D., Purdue University, May 2016. Approaching DNA Methylation at the Nanoscale. Major Professor: Joseph Irudayaraj.

Epigenetics involves a variety of biochemical modifications occurring on chromatin that are able to regulate and fine-tune genetic activities without altering the underlying DNA sequence. So far, four types of epigenetic modulation have been extensively studied: DNA methylation, histone post-translational modification, non-coding RNA, and nuclear organization. A fast-growing body of evidence suggests that epigenetic mechanism plays a fundamental role in physiology and pathology. Of the elucidated epigenetic processes, DNA methylation is the one under intense research due to its direct impact on gene expression, which dynamically bridges the microscale genotype and macroscale phenotype. However, to date our understanding of DNA methylation has primarily come from ensemble and end-point measurements using a population of cells or bulky samples. Unlike genetic aberrations (*e.g.*, amplification, deletion, translocation, and mutation) that are rare to occur and resistant to reverse, events relating to epigenetic modifications take place in a time- and context-dependent manner, thus necessitating nanoscale tools to dissect epigenetic dynamics at a finer spatiotemporal resolution. In my Ph.D. work, a group of advanced single-molecule fluorescence tools, in addition to methodologies in bioengineering, nanotechnology, molecular biology, and bioinformatics, is implemented to approach

the DNA methylation-related activities. Furthermore, part of my research aims to translate the new discoveries from single-cell experiments to biomedical applications. Another focus of my dissertation is on effective manipulation of DNA methylation by novel techniques including nanomaterials and optogenetics. Taken together, these efforts lay the groundwork for better understanding of DNA methylation and provide ample avenues for improving epigenetic therapies.

After the introduction chapter, the second and third chapters of my dissertation center on elucidating the behavior of MBD3 protein within a single living cell, in order to infer the real-time dynamics of DNA methylation. Taking use of fluorescence correlation spectroscopy (FCS), fluorescence lifetime imaging (FLIM), Förster resonance energy transfer (FRET), and photon counting histogram (PCH), the *in situ* binding preference, dissociation kinetics, and stoichiometric transition of MBD3 were probed with unprecedented precision. Due to the substitution of two amino acids in its DNA-binding motif, MBD3 has evolved to differ from other MBD homologues in physicochemical properties as well as in biological functions. Consequently, its altered binding affinity and specificity with DNA methylation lead to unique molecular behaviors that can be sensed by single-molecule techniques. In this line of research, 1) the intracellular diffusion of MBD3 was explored to monitor active DNA demethylation and 2) the cell cycle-dependent activity of MBD3 was characterized. In these attempts my study establishes the dissociation of MBD3-DNA as a marking event in active DNA methylation, which might be involved in early oncogenesis. Moreover, the role of MBD3 in the DNA maintenance methylation is revealed for the first time – *i.e.,* it constitutes a key balancing force to sustain the DNA methylation homeostasis.

Considering its importance in DNA methylation, MBD3 was proposed to participate in brain cancer because the central nervous system possesses abundant DNA methylation marks and active methyl metabolism. Hence in the fourth chapter of my dissertation the regulatory landscape and clinical implication of MBD3 in human malignant glioma is studied. Utilizing whole-genome transcriptome microarray, the MBD3-mediated epigenetic regulation in glioma was profiled. The data uncover that MBD3 stands in the midst of an exquisite anti-glioma network via influencing a large number of transmembrane proteins and nuclear factors. By analyzing the differentially expressed genes upon depletion of MBD3, the MHC class II molecules were identified to be a group of downstream effectors activated by MBD3, which provides important insight into the anti-glioma immunity. Furthermore, MBD3 functions to impact a wide range of non-coding RNAs including the oncogenic miR-17-92 cluster whose dysfunction could lead to an aggressive glioma proliferation and metastasis. From a pilot clinical study, the expression level of MBD3 was found to positively correlate with the patient prognosis and thus could be explored as a novel epigenetic biomarker.

Reactivation of tumor supper genes using DNA demethylating agents has proven to be effective in treating hematopoietic malignancies, but efficient delivery of these drugs into solid tumors is a challenging task. In the fifth chapter a polymeric nano-vector comprising PLGA and PEG is developed to improve the solubility, bioavailability, and therapeutic potency of the FDA-approved DNA demethylating agents – AZA and DAC. With equimolar drug content, the nano-conjugated AZA showed better anti-proliferative effect than free drug in treating xenografted breast cancer. This formulation can also be used to deliver a safe dosage of DAC for sensitizing the alkylation-resistant glioblastoma cells to

temozolomide. Moreover, the improved cell uptake and drug release was characterized by single-molecule tools. The results demonstrate that PLGA-PEG nano-micelles could rapidly enrich in cancer cells and resist to efflux. Notably, the pH-dependent hydrolysis of PLGA enables a targeted drug release in acidic cancer cells featuring overactive glycolysis. Hence, this line of research describes a novel strategy to enhance epigenetic therapy and provides a cost-effective option for future pharmaceutical applications.

The sixth chapter of my dissertation presents a light-controllable technique for loci-specific epigenome editing. The optogenetic proteins CRY2-CIB and PHYB-PIF are photosensitive pairs whose associations can be induced by blue and red light, respectively. Under proper illumination, CRY2 or PHYB is rapidly photoactivated to couple with CIB or PIF, and their optogenetic associations are reversible in the absence of the triggering light. The first half of this chapter elucidates the feasibility of using single-molecule fluorescence tools to characterize real-time optogenetic binding kinetics. In the second half, a pair of DNMT3A-CRY2-mCherry and TRF1-CIB1-EGFP fusion proteins was constructed to achieve selective *de novo* DNA methylation at subtelomeric CpGs. By varying the illumination conditions, the subtelomeric methylation can be incrementally manipulated. These optogenetic modules provide a powerful way to study DNA methylation and can be conveniently adapted for multiplexing epigenome editing and *in vivo* applications.

CHAPTER 1. INTRODUCTION

1.1 Epigenetics and DNA Methylation

According to the definition by the NIH "Roadmap Epigenomics Project," "Epignetics is an emerging frontier of science that involves the study of changes in the regulation of gene activity and expression that are not dependent on gene sequence." Epigenetic modifications, through modulating the physicochemical properties of chromatin, dynamically switch on and off the transcription of relevant genes. One of the longest studied epigenetic mechanisms is DNA methylation, *i.e.*, the transfer of a methyl group from S-adenosylmethionine (SAM) to 5-carbon of the pyrimidine ring in cytosine (Figure 1.1). DNA methylation is intimately involved in life and diseases, and the DNA methylome of a particular cell or organism can be inherited over generations. Therefore, understanding of DNA methylation constitutes a cornerstone for modern biology and medicine.

1.1.1 Overview

For a typical human cell, 3 billion base pairs (bp) of DNA are compacted in its nucleus, which however only encode approximately 21,000 genes. This limited repertoire of genes is shared by numerous differentiated cells and expressed in a precisely controlled manner during the lifespan of every person. For a long period in history, it has baffled scientists why monozygotic twins could have distinct personalities and susceptibilities

to certain diseases even though they boasted the identical genomic information. Genetics alone is far from sufficient to explain the biological diversity and phenotypic variations thereto. In 1940s, C.H. Waddington first coined the term “epigenotype” to describe the causal interactions between the genotype and phenotype during development.[1] However, it was vastly unknown on what mechanism bridged a fixed set of genetic materials to the developmental divergence among somatic cells in a metazoan organism. First discovered by T.B. Johnson and R.D. Coghill in 1925,[2] the presence of 5-methylcytosine (5mC) was later proposed to inactivate one X chromosome in female genome and in 1970s 5mC was established to be a negative regulator for gene expression.[3, 4] Since then, DNA methylation and other epigenetic modifications have filled the gap to explain the time- and context-dependent gene expression in a single cell. Over the past decades, researches have accumulated to reach the conclusion that DNA methylation is indispensable for virtually all biological processes.

In mammalian cells, the enzymatic reaction of DNA methylation preferably occurs at a CpG dinucleotide that is not uniformly distributed over the genome. A considerable portion of CpGs concentrate in short clusters and stretches called CpG islands [CGI, a 500-2,000 bp region having at least 50% of CG content and 60% of an observed-to-expected CpG ratio, *i.e.*, number of CpGs × length of sequence/(number of C × number of G)] near or within the transcription start site of genes (*e.g.*, promoters). Although 70%-80% of scattered CpGs are methylated, most CGIs in promoters are devoid of DNA methylation to keep the genes poised for activation. Once heavily methylated under a diseased condition, the promoter region would be condensed by a series of repressor factors and become inaccessible to transcription activators.

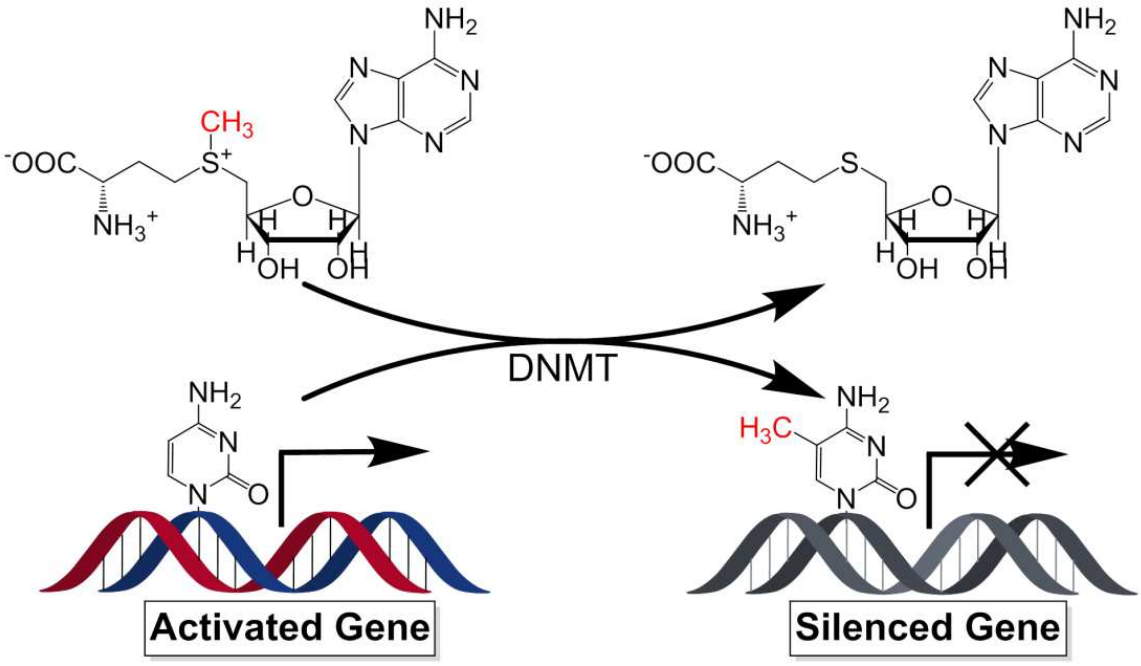


Figure 1.1 Diagram of DNA methylation. A methyl group is transferred from SAM to cytosine by DNMT enzyme to repress a gene.

In the epigenetic network, DNA methylation is not an isolated modulator but oftentimes connect a wide range of partners to achieve transcriptional inhibition. For instance, 5mC site could recruit methyl-CpG-binding domain protein 2 (MBD2) and its associated Mi-2/NuRD complex containing histone deacetylases (HDACs) to temporarily turn down a gene.[5] In comparison, 5mC occupied by methyl-CpG-binding protein 2 (MeCP2) would attract the histone H3 lysine 9 (H3K9) methyltransferase SETDB1 and heterochromatin protein 1 (HP1).[6] In this scenario, DNA methylation acts in cohort with histone modifications and chromatin-remodeling factors to form heterochromatin for a long-term, steady gene silencing.

1.1.2 DNA Methylation in Life and Diseases

For a healthy cell where most of its CGIs in promoters are unmethylated, the CpGs in other regions such as gene bodies and intergenic repetitive elements are heavily methylated to maintain the chromosomal integrity and stability. In the embryogenesis and development of a diploid organism, DNA methylation also plays a crucial role. All human genes have two alleles with one from the maternal germ cell and the other from the paternal source. Genomic imprinting is an epigenetic process by which some genes are expressed in a parent-of-origin-specific manner. If only the allele from the mother needs to be transcribed, the other one from the farther has to be imprinted, *i.e.*, silenced by DNA methylation. Under a similar circumstance, the female somatic cells have two X chromosomes but only one copy is transcriptionally active whereas the other one is often epigenetically inactivated. This X-chromosome inactivation is controlled by an exquisite mechanism that primarily depends on the DNA methylation mediated heterochromatin formation. Another critical process involving DNA methylation lies in tissue differentiation and lineage fidelity. For a mature tissue or organ to carry out its normal function, not all the genes in genome are required. Differentiated from a pluripotent stem cell to a particular cell type, a large number of unneeded genes are switched off by DNA methylation and this epigenetic memory would be faithfully passed down to keep the lineage identity. Despite that DNA methylation at the gene promoter directly suppress that gene, intragenic DNA methylation is also able to fine-tune the activity of RNA polymerase and the alternative splicing of mRNAs.[7, 8]

Due to its functional versatility and importance, DNA methylation is implicated in a myriad of pathological conditions. Mistakes in genomic imprinting have pleiotropic con-

sequences such as placental deficiencies, fetal undergrowth or overgrowth, increased risk of childhood cancers, neurological defects and developmental retardation. The Angelman (AS) and Prader-Willi (PWS) syndromes are two distinct, reciprocally imprinted disorders that share the same chromosomal location. Loss or inactivation of at least one 15q11-13 gene occurs in AS and PWS, with maternal deficiencies giving rise to AS and paternal defects causing PWS. Thus for the genes involved in AS and PWS, both of their alleles are not functional because one is silenced by imprinting and the other one is mutated. Other common imprinting disorders include: Beckwith-Wiedemann syndrome (11p15), Russell-Silver syndrome (11p15), Albright's hereditary osteodystrophy (20q13), and NOEY2 dysfunction (1p31).[9]

Recently, in a plethora of studies aberrant DNA methylation has been defined as a driving force for oncogenesis. The best-characterized epigenetic events in relation to cancer are: 1) global DNA hypomethylation and 2) intense hypermethylation at the promoters of certain tumor suppressor genes.[10] The epigenome of a cancer cell could be depleted by 20%-60% of its DNA methylation, mostly at coding regions, enhancers, introns, and repetitive sequences. In the cancerous DNA demethylation, one destructive effect arises via activation of the endogenous retrovirus and transposable elements present in mammalian genomes, which in turn potently compromise the expression of growth regulatory genes, foster genome-wide mutations, and enforce fatal chromosomal recombinations. *De novo* methylation at the promoter region of tumor suppressor genes is another critical step facilitating cancer initiation. Although the majority of promoters CGIs are free of DNA methylation, an anomalous gain-of-methylation for a key tumor suppressor gene could unduly release the precancerous cells from the restrictions on proliferation, metabolism,

adherence, and metastasis. Notably, the pattern of promoter hypermethylation in some genes is highly tissue-specific and thus can be explored as novel biomarkers. For example, upon DNA hypermethylation, the deregulation of BRCA1 is a hallmark of breast and ovarian tumors, and the repression of hMLH1 predominantly occurs in endometrial, colorectal, and gastric neoplasm.[11] In contrast, hypermethylation of the MGMT gene can bring about a prognostic benefit for glioma patients who receive the standard alkylating chemotherapy (*e.g.,* temozolomide) because inactivation of MGMT alleviates the intrinsic activity of reversing the therapeutic guanine alkylation.

1.1.3 DNA Methylation-Related Enzymes

Conserved along cell division, modified during cell differentiation, and transgressed in cell transformation, the patterns of genome-wide DNA methylation actively dictate cell fate and are under a rigorous control by methyltransferases and demethylases. Thus the DNA methylation homeostasis is stringent on a delicate balance between these two forces of enzymes and their co-factors.

The addition of a methyl mark to cytosine is mainly catalyzed by DNA methyltransferases (DNMTs) in human cells, and to date three genuine DNMTs have been biochemically characterized with the exception that DNMT2 was later found to be a RNA methyltransferase.[12] The catalytic domain of DNMTs is highly conserved, featuring a ~400 amino acids sequence containing motifs I-X in the C-terminus.[13] The N-terminal regulatory domain distinguishes DNMT1 from DNMT3A/B: a proliferating cell nuclear antigen (PCNA)-binding domain (PBD), a pericentric heterochromatin-targeting sequence (TS), a CXXC zinc finger motif, and two bromo adjacent homology (BAH) domains are con-

tained in DNMT1; in comparison, the N-terminal domain of DNMT3A and DNMT3B is composed of a conserved Pro-Trp-Trp-Pro PWWP domain and a plant homeodomain (PHD, also known as ADD, *i.e.,* ATRX-DNMT3-DNMT3L) in their N-terminus (Figure 1.2A). DNMT1 was discovered in 1988 and DNMT3A/B were discovered ten years later.[14, 15] In the early years, based on their binding preference for hemi-methylated or unmethylated DNA, these DNMTs were arbitrarily categorized as maintenance methyltransferase (DNMT1) and *de novo* methyltransferase (DNMT3). DNMT1 is the most abundant DNMT throughout the life cycle of a differentiated cell whereas DNMT3 largely enriches in early development. DNMT1 starts to peak its expression in S-phase and possesses a substrate preference for hemi-methylated DNA at replication loci.[16] During the establishment of tissue-specific methylomes from an epigenetically naive stem cell, DNMT3A and DNMT3B are responsible for adding numerous *de novo* methylation marks to those imprinted genes, satellite sequences, long and short interspersed nuclear elements (LINEs/SINEs), and heterochromatin.[17] However, in recent years studies have shown that actually DNMT1 and DNMT3 are not mutually independent but act in concert for the methylation homeostasis.[18] DNMT3 can support DNMT1 in the maintenance methylation of repetitive elements while overexpression of DNMT1 would result in a marked gain of *de novo* methylation.[19, 20]

In the production of germ cells, the DNA methylation landscape in the gametes, especially the paternal copy, undergoes a rapid and massive demethylation. However, the first active DNA demethylase in human was not found until 2009.[21] Before that, the conventional view in the removal of DNA methylation inclined to: 1) DNA demethylation needs to be achieved via a passive dilution (*i.e.,* the hemi-methylated sites are stopped to

become fully methylated after DNA replication) or 2) the methylated cytosine should experience deamination to become thymine and then be replaced by a new cytosine through the base excision repair (BER), yet both of which were deemed as utterly inefficient processes. Since the recent discovery of the ten-eleven-translocation (TET) family protein-catalyzed conversion of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), the full picture of DNA demethylation has come to our sight: the oxidation of 5mC essentially facilitates the passive and active demethylation processes. The existence of 5hmC at hemi-methylated sites (*i.e.*, 5'-5mC-G-3':3'-G-5hmC-5' pairs) can remarkably accelerate the passive demethylation while 5fC and 5caC are better substrates to be processed by thymine DNA glycosylase (TDG) in BER.[22] TET family proteins (TET1-3) are a group of Fe(II)/α-ketoglutarate (α-KG)-dependent oxygenases and harbor a double-stranded β-helix (DBSH) core motif where high-valet Fe(IV) is generated to oxidize the methyl group on cytosine. TET1 and TET3 also contain a CXXC domain in their N-terminus to facilitate binding with CG-rich sequences (Figure 1.2B). TET3 is predominantly enriched in zygotes and responsible for the active demethylation of the paternal epigenome while TET1 and TET2 are more expressed in primordial germ cells. In some tissues (*e.g.*, brain, heart, lung, liver, and kidney) TET proteins are intrinsically expressed, exerting critical influences on the fidelity of differentiation.[23, 24]

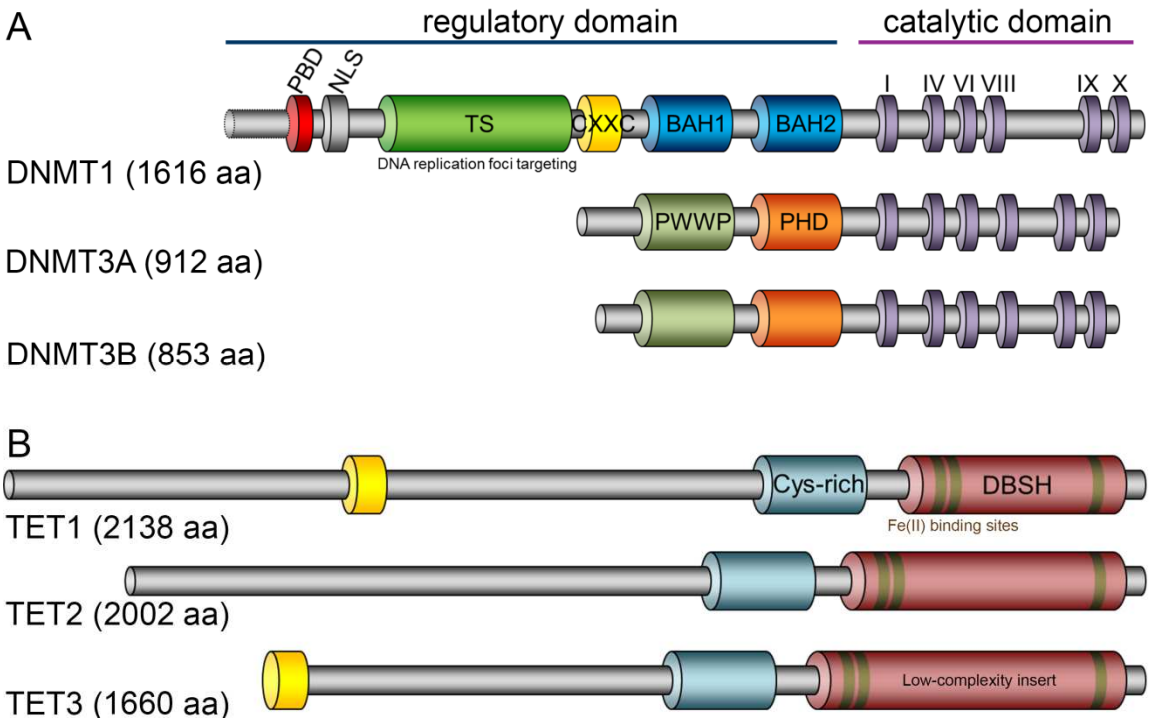


Figure 1.2 Schematic illustration of human methyltransferases and demethylases.

Aberrant expression or mutations of DNMTs and TETs are extensively implicated in various diseases, particularly in hematopoietic malignancies. The pathological translocation of mixed lineage leukemia protein (MLL)-TET1 [t(10;11)(q22;q23)] is a common event in acute myeloid leukemia (AML).[25] TET2 is frequently mutated in lymphoma, myelodysplastic syndrome (MDS), myelofibrosis, and different types of leukemia.[24] Mutations in DNMT3A are usually correlated with poor prognosis for patients having AML or MDS.[26, 27] The biallelic DNMT3B mutation is the main cause for the rare "immunodeficiency, centromeric region instability, and facial anomalies syndrome" (ICF).[28] In addition, abnormal DNA methylation induced by dysfunctional DNMTs or TETs could substantially promote solid tumor initiation and progression.

1.1.4 DNA Methylation-Binding Proteins

Besides the methylation-editing enzymes, methylated cytosine can also be recognized by a group of reader proteins including Kaiso family proteins, MBD family proteins, and the SET- and Ring finger-associated (SRA) domain proteins (Figure 1.3). These proteins either function as a structural component of chromatin or serve as a coordinator to connect other factors. Kaiso and Kaiso-like proteins interact with DNA methylation through three Kruppel-like C2H2 zinc finger motifs but they require two consecutive mCpGs for an efficient binding.[29, 30] In cooperation with HDACs and chromatin-remodeling factors, Kaiso turns out to be a strong transcription repressor. The first MBD protein MeCP2 was discovered in 1992,[31] followed by the discovery of the other MBD members in the next several years. Till today at least ten MBD-containing proteins have been identified in human being. The core DNA-binding motif of MBD proteins contains about 70 residues folding into an α/β sandwich structure that is embedded into the major groove of target DNA sites.[32] The binding specificity and affinity between MBD and DNA methylation is subjected to changes of some key residues such as Val-20, Arg-22, Lys-30, Tyr-34, Arg-44, and Ser-45.[32-34] For instance, the human MBD2 and MBD3 are highly homologous, but MBD3 has a much weaker *in vitro* binding affinity with 5mC owing to the K30H and Y34F substitution.[35] Over the past years the SRA domain-containing protein UHRF1 (also known as NP95 or ICBP90) has drawn considerable attention since the finding that UHRF1 could co-localize with hemi-methylated DNA and DNMT1 during the S-phase in cell cycle.[36-38] After locating a hemi-methylated CpG site, UHRF1 would quickly move in and flip the methylated cytosine out of the double helix, a characteristic conformation subsequently recruiting DNMT1 to transfer a methyl group to

the unmethylated CpG on the complementary strand. This discovery reveals that UHRF1 is a vital part in the DNA maintenance methylation.

Figure 1.3 Solved structures of major DNA methylation-binding proteins. This figure was generated using PyMOL with structural information retrieved from the RCSB PDB database: Kaiso (PDB ID: 4F6N), MBD (PDB ID: 2YK8), SRA (PDB ID: 2ZO1).

As mentioned above, DNA methylation-binding proteins have evolved to gain a diversity of functions to interpret the epigenetic code on DNA. To effectively regulate a gene, DNA methylation often needs to co-work with other epigenetic modulators. For instance, a number of DNA methylation-binding proteins (*e.g.*, Kaiso, MeCP2, MBD2, MBD3, and SRA) directly or indirectly associate with HDACs; MBD1 usually co-exists with H3K9 trimethylation and HP1; and both MBD2 and MBD3 are core subunits of the nucleosomal remodeling deacetylase complex Mi-2/NuRD.[39] Interestingly, under certain circumstances some MBD proteins (*e.g.*, MBD2, MBD3, and MBD4) can interact with unmethylated DNA or even demethylate their binding sequences.[40-42] In this regard, these proteins might also be involved in gene activation.

A multitude of diseases and disorders are linked to DNA methylation-binding proteins. The best-known one is Rett syndrome, a neurodevelopment disorder caused by mu-

tations in MeCP2.[43] MBD3 is important for development and deletion of MBD3 in mouse model is embryonic lethal.[44] Except for MBD2, mutations in all other MBD proteins have been identified in autism spectrum disorder.[39] In cancers, abnormal expression of one or more MBD proteins has been frequently observed, such as overexpression of MeCP2 and MBD2 in breast cancer,[45] overexpression of MBD2 and MBD4 in brain glioblastoma,[46, 47] up-regulation of MBD1 and down-regulation of MBD2 in prostate cancer,[48] and down-regulation of MBD3 in gastrointestinal and endometrial cancers.[49, 50]

1.2 Nanoscale Detection and Manipulation of DNA Methylation

At present our understanding of epigenetic regulation has primarily come from ensemble, static, and end-point tests utilizing a population of cells or bulky samples, largely overlooking the cellular heterogeneity and real-time intermediates. In order to dissect and engineer a single cell, achieving a finer experimental scale than that in the conventional biological approaches is imperative. Advanced single-molecule fluorescence techniques open a unique window for us to inspect epigenetic molecules and events with marvelous resolutions. In the meantime, nanobiotechnology provides us tremendous opportunities to edit epigenome with unprecedented convenience and accuracy.

1.2.1 Overview

In eukaryotic cells, the basic unit for epigenetic regulation is nucleosome that contains a histone octamer core wrapped by 147 bp of DNA (Figure 1.4). Epigenetic modifications function through influencing the intertwined DNA-histone interface to control gene transcription. However, the extremely small size of a single nucleosome – approxi-

mately 11 nm – poses a daunting task to detect or manipulate a specific type of epigenetic mark on it.

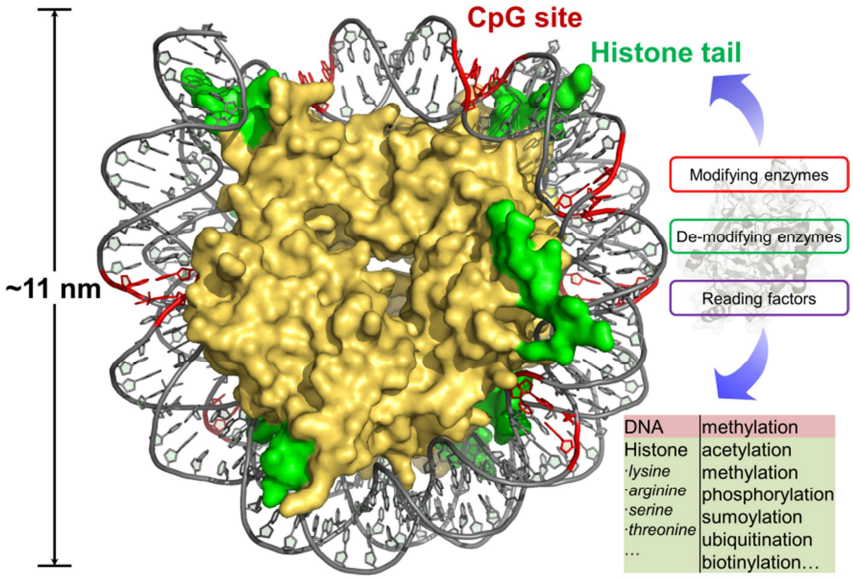


Figure 1.4 Basic unit for epigenetic modifications to occur – a single nucleosome.

Capturing and controlling epigenetic events in single living cells require the spatial and temporal resolution of the methodology to meet the scales of the ongoing cellular and molecular activities (Figure 1.5). Yet the intracellular environment is extremely crowded. Given a human cell with an average volume of ~1 picoliter, 3×10^9 bp DNA (equivalent to 2 meters if fully stretched) is packed in its nucleus whose diameter is merely 10-15 microns. Even DNA methylation only constitutes ~1% of the human genome, which could generate a fairly high *in situ* density of 5mC at particular regions such as satellite DNA repeats and heterochromatin sites. To overcome the bottleneck of microscopic observation, nanoscale imaging and molecular engineering are becoming increasingly important. On a different note, the genome-wide content of DNA methylation is dynamically fluctu-

ating along with different cell behaviors, posing another layer of complexity for single-cell studies.

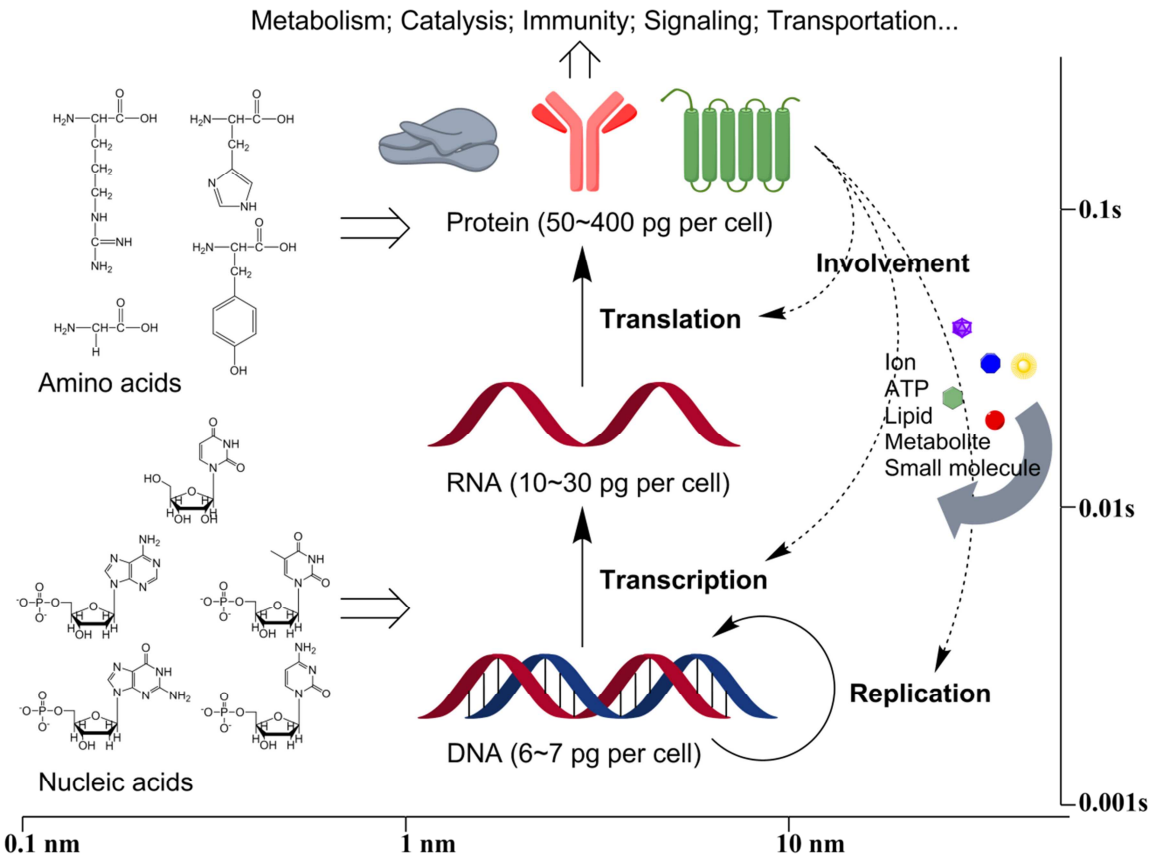


Figure 1.5 Spatiotemporal scales required for single-cell and single-molecule studies.

1.2.2 Single-Molecule Fluorescence Techniques

About 340 years have passed since Robert Hooke and Antonie van Leeuwenhoek popularized the modern microscopy. The 20th century has embraced an exciting number of breakthroughs that rendered optical microscope an irreplaceable part in life science research. The invention of fluorescence microscope in 1910s has spurred a revolutionary leap, after which fluorescent labeling has quickly become the norm for observing micro-

organisms and subcellular structures. Derived from the principle of fluorescence, the further developments of confocal microscopy (1960s), fluorescence fluctuation microscopy (1970s), total internal reflection fluorescence microscopy (1980s), multi-photon microscopy (1990s), light-sheet fluorescence microscopy (1990s), and super-resolution microscopy (2000s) have been enormously advancing the spatiotemporal limits of optical dissection. The capability of single-molecule fluorescence techniques to provide *in situ*, real-time information of intracellular dynamics is an unparalleled advantage. In this thesis, a toolset of ultrafast pulsed lasers and single-photon avalanche diode (SPAD) detectors are integrated into a scanning confocal time-resolved microscope system (PicoQuant) for probing the DNA methylation-related activities. The mainly used techniques are depicted in Figure 1.6 and introduced below.

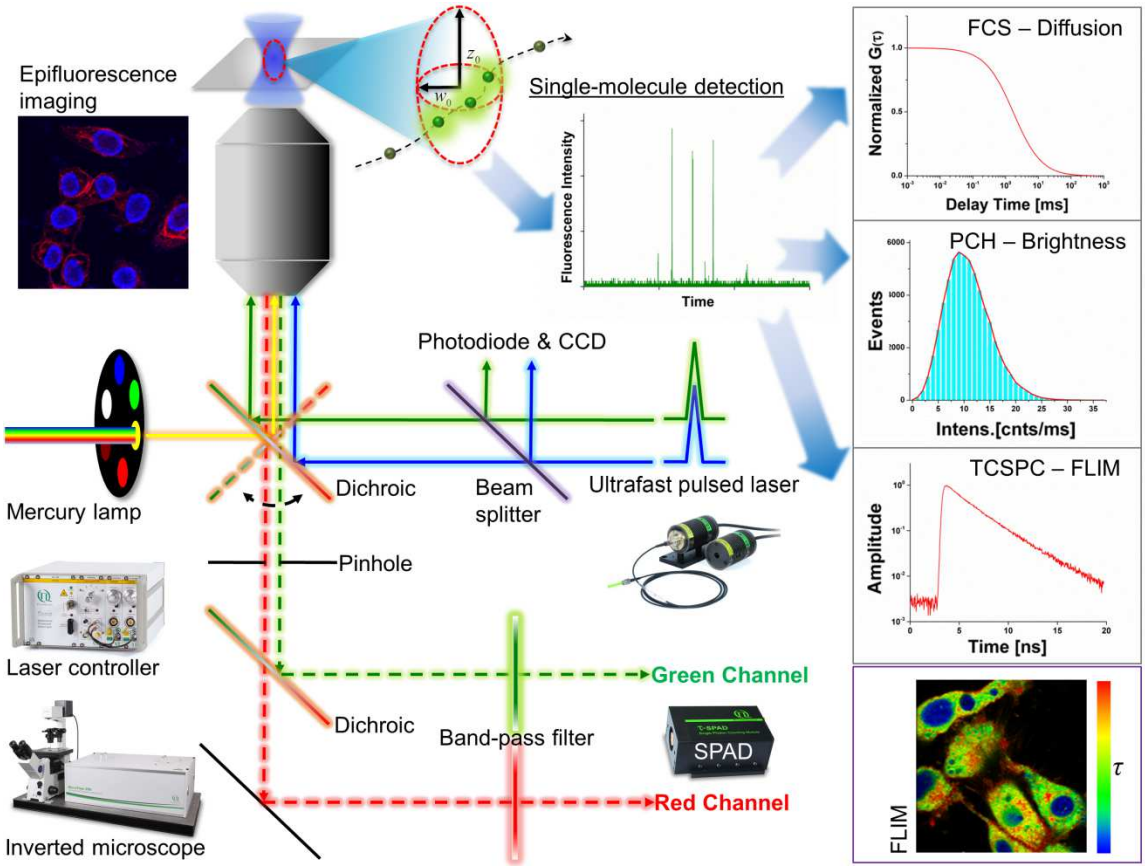


Figure 1.6 Function modules in a single-molecule fluorescence platform. The system is based on an inverted confocal scanning microscope. Ultrafast pulsed lasers and SPAD detectors are implemented to realize time-correlated single-photon counting. In the confocal detection volume, the recorded fluorescence fluctuation can be used for FCS and PCH, and the photon decay pattern can be interpreted in FLIM.

Fluorescence correlation spectroscopy (FCS). First developed by Douglas Madge, Elliot Elson, and W.W. Webb in 1972, FCS takes use of the spontaneous fluorescence fluctuation (at a microsecond resolution) due to the molecular diffusion in-and-out a femtoliter-level volume to biophysically infer the concentration, diffusion rate, hydrodynamic size, and interaction kinetics of the fluorescent targets.[51, 52] In FCS analysis, an autocorrelation function is extracted from the fluorescence fluctuation trace which describes the similarity between two signal events as a function of the time lag (τ):

$$G(\tau) = \frac{\langle \delta F(t) \cdot \delta F(t+\tau) \rangle}{\langle F(t) \rangle^2}$$

where $\langle F \rangle$ is the average fluorescence intensity over time (t), and $\delta F(t) = F(t) - \langle F(t) \rangle$. For 3D free diffusion, the autocorrelation function can be correlated to single-molecule properties using:

$$G(\tau) = \frac{1}{\langle N \rangle} \sum_i \left(f_i \cdot \left(1 + \frac{\tau}{\tau_{Di}} \right)^{-1} \cdot \left(1 + \frac{\tau}{\kappa^2 \tau_{Di}} \right)^{-1/2} \right)$$

where $\langle N \rangle$ is the average number of molecules, f_i is the fraction of species i, κ is the spatial dimension of the detection volume (the ratio of radial w_0 to axial z_0 radii), and τ_{Di} is the characteristic diffusion time (the average time for one molecule dwelling in the detection volume). The concentration of molecules can be calculated as: $C = \langle N \rangle / V_{eff}$, where V_{eff} is the effective detection volume: $V_{eff} = \pi^{3/2} w_0^2 z_0$. From τ_D the diffusion coefficient (D) of molecules can be calculated as: $D = w_0^2 / 4\tau_D$. The hydrodynamic size (R) of molecules can also be determined from the Stokes-Einstein equation: $D = \frac{kT}{6\pi\eta R}$, where k is the Boltzmann's constant and η is the environmental viscosity. For a constrained diffusion condition such as 2D diffusion and anomalous restricted diffusion, different fitting models should be opted accordingly.

Photon counting histogram (PCH). The sensitivity of FCS to distinguish different molecules heavily relies on molecular size, which requires a minimum of 1.6-fold change in hydrodynamic size or 4.1-fold change in molecular weight to separate two species.[53] The

invention of PCH method greatly enriches the single-molecule toolbox with another fluorescent characteristic – molecular brightness (ε), which is especially useful to reflect the stoichiometry of fluorophores having oligomeric status.[54, 55] Collected as the average number of photons emitted from per molecule per sampling time, single-molecule brightness is physically defined by:

$$\varepsilon = I_0^N \beta \eta_w T$$

where T is the integration time increment, η_w is the detection efficiency, I_0 is the maximum excitation at the center of detection volume, N is 1 for single-photon system, β is the excitation probability. The weight averaged $\langle \varepsilon \rangle$ is independent of the ensemble concentration and can be mathematically determined from the fluctuation trace by:

$$\langle \varepsilon \rangle = \frac{\sigma^2 - \langle I \rangle}{\gamma \langle I \rangle} = \sum_i f_i \varepsilon_i$$

where ε_i is the brightness of species i, f_i is its fractional intensity, and σ^2 is the detection variance. The gamma factor (γ) is an excitation-dependent value and is 0.3536 for single-photon confocal system.

Fluorescence lifetime imaging microscopy (FLIM). Fluorescence lifetime (τ, a different parameter from the time lag in autocorrelation function) is defined as the average time one fluorochrome stays at the excited state before returning to the ground state by emitting photons, which can be illustrated as:

$$\tau = \frac{\Gamma}{\Gamma + k_{nr}}$$

where Γ is the emission rate and k_{nr} the non-radiative decay rate, both of which are subjected to physicochemical influences. Fluorescence lifetime is independent of molecule

concentration and excitation intensity so that can be used a robust measure to sense conformational changes, associations, and environmental factors. In the time-domain FLIM, fluorescence lifetime is derived from the time-correlated single-photon counting (TCSPC) histogram that records the arrival time of photons under pulsed excitation (Figure 1.7).

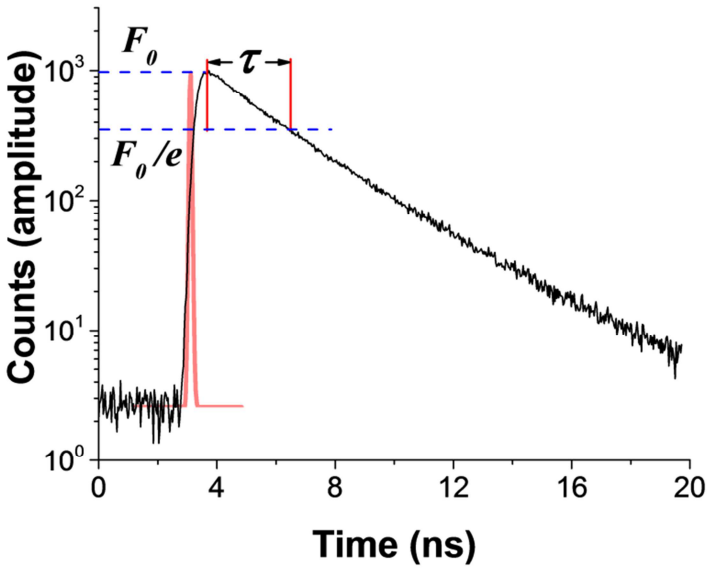


Figure 1.7 Schematic of time-domain fluorescence lifetime measurement. After a pulsed excitation, the arrival time of each collected photon is compiled to a TCSPC histogram. Fluorescence lifetime herein is determined as the decay time for the highest F_0 to become F_0/e.

Förster resonance energy transfer (FRET). Discovered by Theodor Förster in 1946, FRET is a physical process of non-radiative energy transfer between two fluorochromes with proper spectral overlap and dipole-dipole orientation (Figure 1.8). The FRET efficiency (E) greatly depends on the distance (r) between the donor and acceptor with an inverse 6th power law:

$$E = \frac{1}{1+\left(r / R_0\right)^6}$$

FRET mainly manifests within an inter-dipole distance of 1-6 nm and therefore can serve as a biophysical “ruler” to scrutinize molecular interactions. When acceptor molecules are nearby, a reduction in the fluorescence lifetime of donor molecules indicates the occurrence of FRET and is quantitatively related to the FRET efficiency:

$$E = 1 - \frac{\tau_{DA}}{\tau_D}$$

where τ_{DA} and τ_D are the fluorescence lifetimes of donor in the presence and absence of acceptor, respectively.

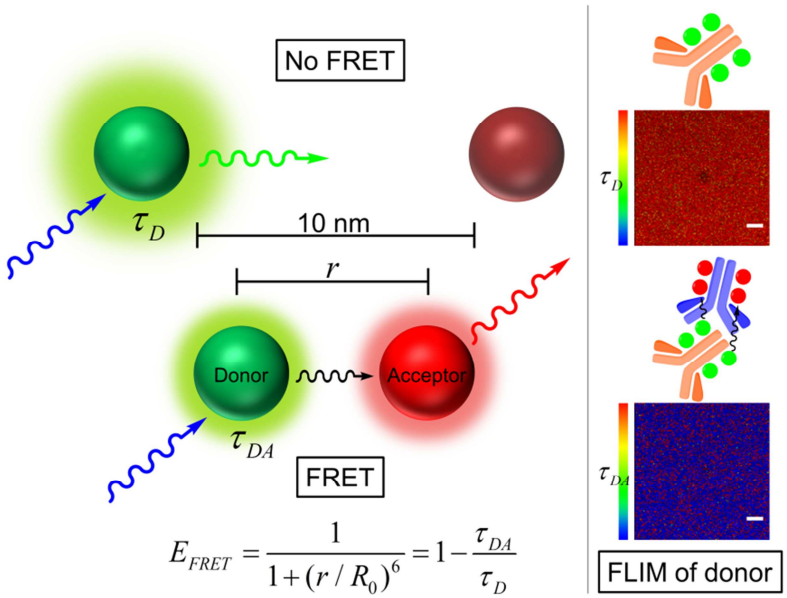


Figure 1.8 FRET and FLIM-FRET. Non-radiative energy transfer would occur when the donor and acceptor are closer than 10 nm. The FRET efficiency depends on the distance between the dipole moments of fluorophores. FRET leads to a reduction in the fluorescence lifetime of donor and can be visualized in FLIM images.

1.2.3 Epigenetic Therapy

Epigenetic pathology has the potential to reshape cellular functions and appreciably foster an oncogenic transformation. As mentioned above, epigenetic silencing of tumor

suppressor genes is one of the driving events in cancer progression and drug resistance. Fortunately, epigenetic errors in cancer, unlike genetic aberrations, can be reversed by proper agents with relative ease and less side effects because epigenetic modifications do not alter the DNA sequence.[56, 57] This property makes epigenetic therapy a promising intervention strategy. DNA methylation and histone deacetylation are the two major mechanisms to compress chromatin and inhibit gene expression. Therefore inhibitors targeting DNMTs and HDACs have been designed and synthesized as a new line of chemotherapy. So far, five epigenetic drugs have been approved by US Food and Drug Administration (FDA): 5-azacytidine (AZA, 2004) and 5-aza-2'-deoxycytidine (DAC, 2006) for treating MDS; vorinostat (2006) and romidepsin (2009) for treating cutaneous T cell lymphoma; and panobinostat (2015) for treating multiple myeloma, while more compounds are under clinical trials or preclinical investigations.[57, 58]

AZA and DAC are cytidine analogues with their 5-carbon replaced by a nitrogen atom so that cannot be methylated by DNMTs anymore. Upon uptake by cells, these cytidine analogues are triphosphorylated and integrated into newly synthesized DNA strands during the S-phase. After several rounds of cell division, the covalent entrapping and subsequent degradation of DNMT1 at the drug incorporation sites would significantly deplete the endogenous methylation reservoir, which achieves a therapeutic effect by reactivating a group of tumor suppressor genes such as p16, RB, PTEN, and hMLH1.[58-61] As an alternative, a handful of non-nucleoside molecules are developed to directly target the catalytic domain of DNMTs or HDACs by which the deregulated genes can be restored.[62, 63]

Besides being applied separately, epigenetic therapies are more often used in combination with other chemo-drugs to achieve a prognostic improvement. To name a few, AZA could potently sensitize prostate cancer to decetaxel and cisplatin;[64] vorinostat and paclitaxel together elicit a much stronger anti-ovarian cancer activity;[65] both DNMT and HDAC inhibitors are effective in sensitizing the tamoxifen-resistant breast cancer to hormone therapies.[66, 67] Another strategy is to look for synergistic effects from combining DNMT and HDAC inhibitors and utilizing their reciprocal modulation on gene activation.[68, 69] One thing worth mentioning, when employing epigenetic treatments, is that they would not only reactivate tumor suppressor genes but also boost the expression of cancer-specific antigens. It thus holds another promise for the incorporation of epigenetic sensitization to the current booming immunotherapy.[70]

1.2.4 Epigenome Editing

In order to fully understand the functional implication of an epigenetic mark, the capability of artificially adding or removing such modification in living organisms is an emerging challenge. From another perspective, searching for tools that can epigenetically regulate genetic activity without altering the coding sequence is becoming a pressing demand for biotechnology. Hence, epigenome editing has been quickly arising as a hot topic in basic research and engineering regime. The first generation of epigenome editing derived from conventional techniques such as small interfering RNA (siRNA) or targeted mutation to systematically disable an epigenetic enzyme, then depleting a relevant modification and its regulatory impact. In this way, the majority of epigenetic enzymes and factors have been characterized regarding their influences on epigenome and phenotypic

traits in a variety of model organisms. However, strategic knockout or overexpression of an epigenetic protein at the global level cannot adequately address the query on epigenetic control over a specific gene of interest and sometimes would even give rise to severe side effects since some epigenetic molecules are indispensable for survival. The past decade has witnessed the soaring popularity of genome engineering enabled by the discovery of novel DNA-targeting modules, including zinc finger nucleases (ZFNs),[71] transcription activator-like effector nuclease (TALENs),[72] and clustered regulatory interspaced short palindromic repeats and its associated bacteria endonuclease 9 (CRISPR/Cas9).[73] With continuing improvements on and implementation of these programmable DNA sequence-recognition proteins, the next generation of loci-specific epigenome editing has been devised.

Rather than by introducing DNA breaks, loci-specific epigenome editing is realized by replacing the nuclease component in ZFNs, TALENs, or CRISPR/Cas9 with an epigenetic effector to confer the desired modification onto the targeted sequence (*e.g.*, connecting a deactivated CRISPR/Cas9 to a favorable epigenetic enzyme[74, 75]) (Figure 1.9). In design, an epigenome editing module should comprise two functional domains: a targeting domain and an effector domain, between which a linker sequence can be inserted according to the covering length.[76] The targeting specificity and effectiveness can be evaluated with affinity-based assays such as MeDIP. The prominent advantages for epigenome editing over genome editing include: first, the outcome from epigenome editing is more predictable whereas the homologues recombination and non-homologous end-joining induced by genome-editing nucleases may lead to entirely different phenotypes; second, the extent of gene expression by epigenome editing can be fine-tuned in contrast to a dichot-

omous on/off effect by genome editing. Substantive attempts have been invested and they successfully laid the foundation on how to harness these powerful tools. For example, by fusing a TET catalytic domain with a zinc finger or TALE proteins, the target gene could be rapidly demethylated for reactivation.[77, 78] Moreover, epigenome editing can be used to investigate the epigenetic code on histone considering that a same type of modification yet at different residuals might cause dissimilar outcomes. To this end, it has been found that demethylating H3K4 or methylating H3K9 could both achieve gene suppression.[79, 80] More recently, upon separating the targeting and effector domains and fusing them with a pair of light-sensitive proteins repectively, loci-specific histone modifications could be manipulated on demand by remote illumination.[81] Thanks to the striving efforts for tool optimization, precision epigenome editing is striding into a new era and being adapted for *in vivo* applications.[82]

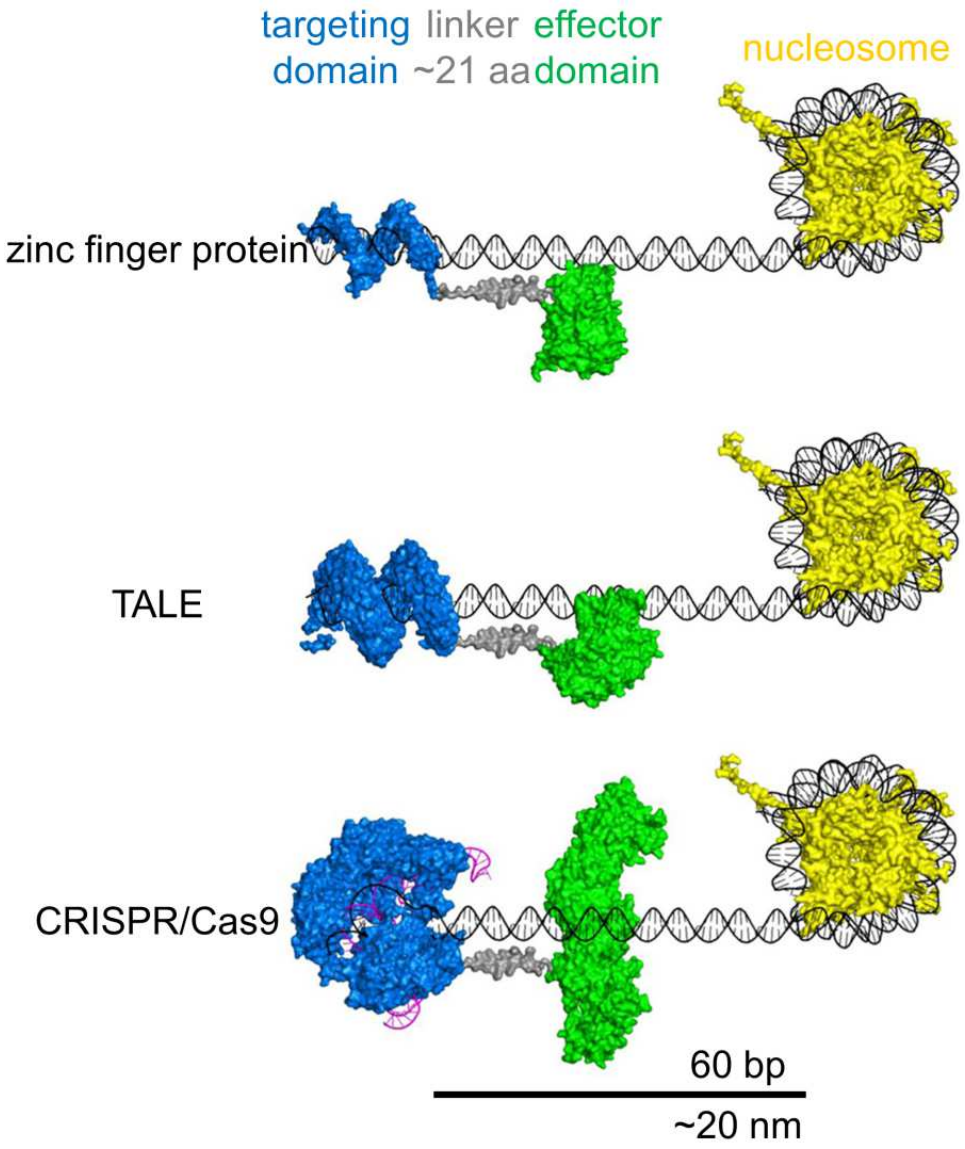


Figure 1.9 Next generation of epigenome-editing tools combining DNA-targeting modules and epigenetic effectors. (figure adapted from reference [75])

CHAPTER 2. UTILIZING REAL-TIME DYNAMICS OF MBD3 TO SENSE ACTIVE DNA DEMETHYLATION

2.1 Introduction

Global DNA hypomethylation, *i.e.*, a genome-wide reduction in 5mC, is one of the most characterized epigenetic anomalies in cancer. However, it is immensely unclear on the triggering mechanisms for such an extensive DNA demethylation, which has become a stumbling block to our in-depth understanding of oncogenesis. Recently, a growing body of evidence has noted that loss of 5hmC also manifested in a number of cancers.[83, 84] In the paradigm of active DNA demethylation, 5mC is depleted through a series of oxidization to 5hmC, 5fC, and 5caC.[85] The concurrent decrease of 5mC and 5hmC in tumor cells indicates that the cancerous DNA hypomethylation might be markedly attributed to an active process involving TET enzymes.[10, 22]

Hypoxia is the typical microenvironment for solid tumors featuring excessive cell growth. As demonstrated before, low oxygen tension is critical for the maintenance of malignant tumorigenic cells (or cancer stem cells) that are responsible for a fatal tumor progression and metastasis.[86] This hypothesis was indirectly supported by a study showing that culturing human fibroblasts under hypoxia could more efficiently reprogram the cells to induced pluripotent stem cells.[87] Complementary to the finding that cancer stem cells possess a much lower DNA methylation compared to physiologically differentiated cells, another line of research has revealed that hypoxic stress could evoke a substantial

DNA demethylation in cancer cells and tissues.[88, 89] These reports collectively suggest that hypoxia may be closely related to active DNA demethylation in tumorigenesis. To fully establish this connection, a cost-effective method for rapid quantification of all cytosine modifications in trace amount of DNA is needed. Although techniques including liquid chromatography-mass spectrometry (LC-MS) and high-throughput next-generation DNA sequencing provide unparalleled power for epigenetic research, they are laborious, time-consuming, and not ideal for multiplex detection.[90] In this regard, an enhanced enzyme-linked immunosorbent assay (ELISA) is first developed to validate the hypoxia-induced active DNA demethylation in HeLa cells.

Furthermore, limited tools exist to capture the ongoing DNA demethylation in single cells. In the following work, with single-molecule FCS the process of active DNA demethylation is successfully captured by monitoring the real-time dynamics of MBD3 protein in living HeLa cells exposed to hypoxia. Among the MBD family members, MBD3 possesses the lowest binding affinity with DNA methylation site due to the residual mutations in its DNA-binding motif (Figure 2.1), which property makes MBD3 a candidate molecule to be detached from chromatin when the occupied CpG site is undergoing an enzymatic conversion. For active DNA demethylation to proceed, the already existing 5mC and 5hmC need to be enzymatically converted to 5fC or 5caC (suitable substrates for TDG-mediated BER). As a binding protein for both 5mC and 5hmC, MBD3 is expected to promptly dissociate from DNA in order to expose the modified cytosine; otherwise its steric hindrance would impede the action of TET and TDG proteins. Hence, timely detection of the subtle changes in the MBD3 diffusion – the transition from bound to unbound state – provides an elegant solution to sense the process of active DNA de-

methylation. FCS is a single-molecule approach that can conveniently quantify protein diffusion with resolution down to microseconds. With this temporal sensitivity, the accelerating MBD3 diffusion in the cell nucleus is proposed to be a precursory event for active DNA demethylation.

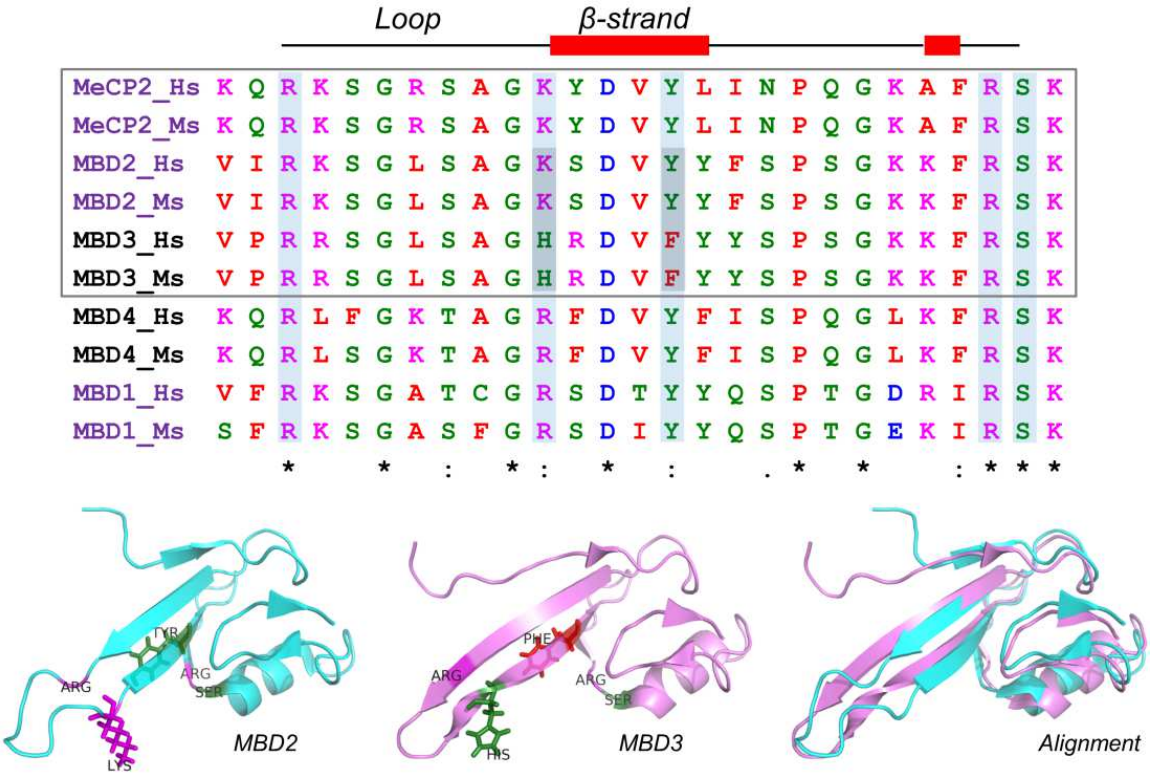


Figure 2.1 Comparison of the core DNA-binding motifs in the major MBD proteins. The five key residues for substrate recognition are highlighted in light blue. The K30H/Y34F substitutions in MBD3 are highlighted in dark grey. The structural alignment for MBD2 (PDB ID: 2KY8) and MBD3 (PDB ID: 2MB7) is generated with PyMOL.

2.2 Methods and Materials

2.2.1 Cell Culture, Plasmid Transfection, and Hypoxia Induction

HeLa cells were routinely cultured in DMEM/F-12 medium supplemented with 10% fetal bovine serum, 100 IU/mL penicillin, and 100 μg/mL streptomycin in a 37°C humid-

ified incubator supplying 5% CO_2. The pGFP-MBD3 plasmid was a kind gift from the Dr. Adrian Bird's group (University of Edinburgh, UK) and transfection was conducted with Lipofectamine LTX & Plus reagents (Life Technology). Before transfection, 5×10^4 cells were seeded onto a sterilized No.1 coverslip (VWR International) placed in 12-well plate with 2 mL fresh culture medium. After the cells reached 70-80% confluence the medium was replaced with antibiotics-free, low-serum medium for another 24 h culture. Then 100 ng plasmid DNA, 2.5 μL LTX reagent, and 0.1 μL Plus reagent were dissolved in 200 μL Opti-MEM medium and incubated for 10 min before adding to the culture well. Transfection lasted for 24 h before hypoxia induction. Since FCS measurement requires an optimal concentration ranging from hundreds of pM to nM, the cells selected for data collection have an average concentration of 80-120 nM MBD3-GFP expressed in their nuclei.

Acute hypoxia was induced with 1 mM sodium dithionite ($Na_2S_2O_4$, Sigma) to fast deplete dissolved oxygen in culture medium, which could efficiently lower the pO_2 to 10 Torr whereas 100% N_2 could only lower the *p*O2 to 30-60 Torr.[91, 92] The pH of medium was re-adjusted to 7.4 by HEPES to buffer the acidification potential of $Na_2S_2O_4$. To maintain the hypoxic condition, medium containing $Na_2S_2O_4$ was changed every 8 h.

2.2.2 Enhanced ELISA for Quantification of DNA Methylation Marks

On the basis of conventional ELISA that takes use of primary and secondary antibodies, a third layer of signal amplification is designed for an enhanced ELISA by conjugating secondary antibody with biotin. When antigen is present, streptavidin coupled to multiple horseradish peroxidase (HRP) molecules would be specifically enriched (Figure 2.2). For a 96-well assay plate, a maximum of 200 ng DNA could be input for analysis,

from which 5fC and 5caC are hard to be detected (less than 0.01% in genomic DNA). With this optimized ELISA, all the four types of cytosine modifications can be simultaneously quantified in the same plate.

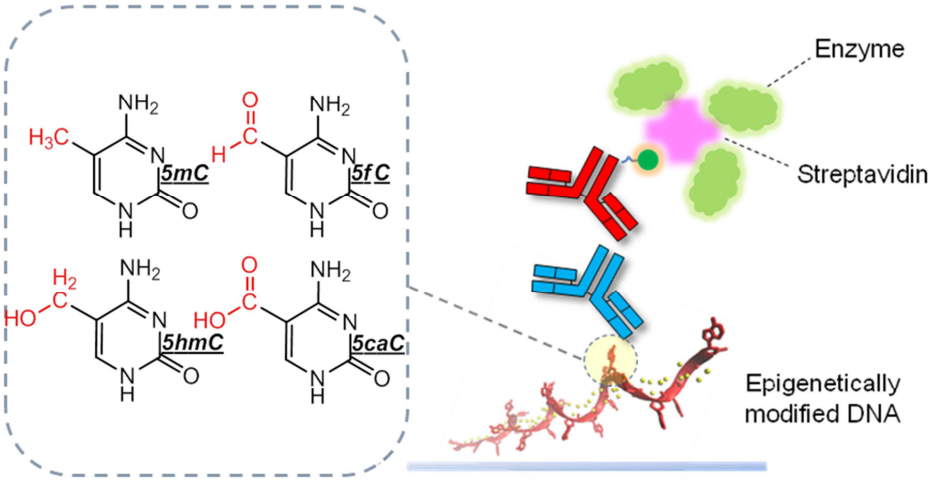


Figure 2.2 Schematic of the enhanced ELISA utilizing the biotin-avidin coupling.

2.2.3 Biotinylation of Secondary Antibody

Secondary antibodies in PBS were labeled with 50-fold molar excess of succinimidyl-6-[biotinamido]-6-hexanamidohexanoate (NHS-LC-LC-biotin, Pierce) dissolved in DMSO. The reaction was performed in 10 mM phosphate buffer with 150 mM NaCl (pH = 7.4) at room temperature for 1 h and the biotinylated antibody was purified by dialysis and size exclusion chromatography with Sephadex G-15 gel column (gel volume: 10 mL). The product was then concentrated and stored at 4°C until use.

2.2.4 Conjugation of Streptavidin with HRP

Streptavidin (Prozyme) and HRP were respectively coupled with succinimidyl 6-(3[2-pyridyldithio] propionamido) hexanoate (LC-SPDP, Pierce) and succinimidyl-4-(N-

maleimidomethyl)cyclohexane-1-carboxy-(6-amidocaproate) (LC-SMCC, Pierce) – both in the molar ratio 1:50 – at room temperature for 1 h. The LC-SPDP was treated with dithiothreitol (DTT, 10 mM) for activation at room temperature for 30 min. The thiolated streptavidin and maleimide-activated HRP (molar ratio 1:5) were purified with Sephadex G-15 gel filtration (pre-equilibrated in 5 mM EDTA) and then immediately mixed. The conjugation was carried out with gentle shaking at room temperature for 2 h. This protein product was stored at 4°C until use.

2.2.5 DNA Extraction and ELISA Quantification

DNA from HeLa cells was extracted with DNeasy Kit (Qiagen). Purified DNA in 100 μL TE buffer (50 ng for 5mC, 200 ng for 5hmC, 5fC, and 5caC) was added to each polystyrene plate microwell followed by mixing with 100 μL DNA coating solution (Pierce) at room temperature for 2 h with gentle agitation. The substrate DNA was then washed three times with 300 μL of DI water in the microwell. To block residual surfaces, the microwell was treated with 0.5% (w/v) casein dissolved in PBS (200 μL) at 37°C for 1 h. Primary antibodies used in this study include: mouse anti-5mC-IgG (Eurogentec), rabbit anti-5hmC/5fC/5caC-IgG (Active Motif). An amount of 0.5 μg/mL of primary antibody diluted in PBS containing 0.5% casein (100 μL) was added and then 1 μg/mL of biotinylated secondary antibody (100 μL) was reacted with the primary antibody. To form avidin-biotin complex, 0.5 μg/mL of streptavidin-HRP conjugate (100 μL) was finally added into the microwell. For signal generation, 200 μL colorimetric substrate mixture [50 mM sodium acetate buffer (pH = 5.1) : 1% (w/v) TMB : 3% (v/v) H_2O_2 = 1,000 : 10 : 1 as volume ratio] was applied for 15 min at room temperature followed by 50 μL of

2 M sulfuric acid to stop the reaction. The optical intensity was measured by a Versa-maxTM absorbance microplate reader (Molecular Device). As shown in Figure 2.3, the limit-of-detection (LOD) of enhanced ELISA is at least ten-fold lower than that of conventional ELISA when using a standard DNA with 16% of the sequence as 5caC. Approximately 2 pg standard DNA (equivalent to 0.0016% 5caC in 200 ng genomic DNA) can be detected with enhanced ELISA whereas 20 pg standard DNA cannot generate a readable signal by conventional ELISA.

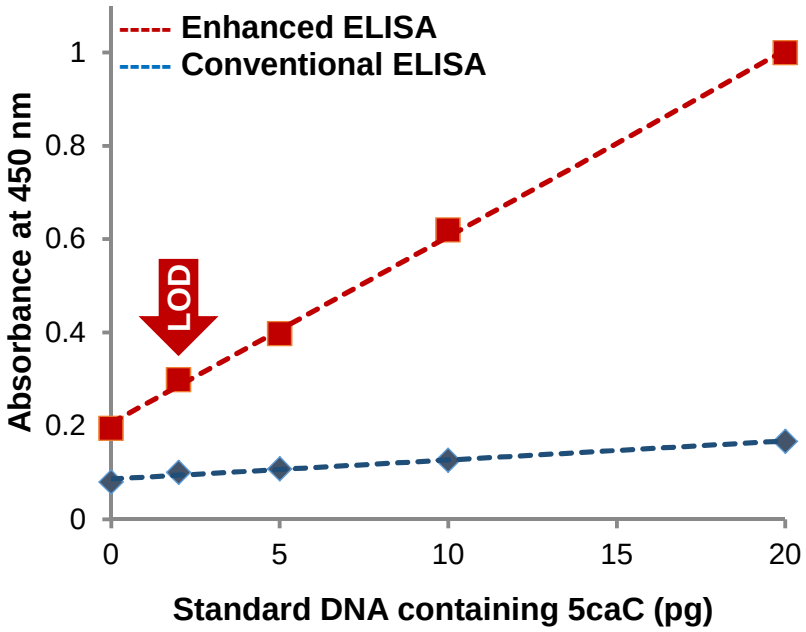


Figure 2.3 Comparison of signal amplification between the conventional and enhanced ELISA for detection of 5caC-containing standard DNA.

2.2.6 Instrumentation for FCS

FCS was performed by using a scanning confocal time-resolved microscope (Micro-time, PicoQuant). A 465 nm picosecond pulsed laser was used to excite the GFP tag. The

power output of excitation was precisely controlled at the μW level by a Sepia PDL 808 driver to prevent phototoxicity. The laser beam was delivered to the sample stage through an apochromatic water immersion objective lens (60×, NA 1.2, Olympus) and the emitted fluorescence was collected by the same objective lens, after which the scattered excitation light was cleaned by a dual-band dichroic (z467/638rpc, Chroma). A 50 um pinhole was employed to block the off-focus photons and the emission signal was finally separated by a band-pass filter (500-540 nm, Chroma) before reaching a SPAD detector (SPCM-AQR-14, PerkinElmer). Information of each collected photon was stored in a time-tagged time-resolved (TTTR) mode (TimeHarp, PicoQuant) in which a time stamp was assigned to record the photon arrival time after the laser pulse.

2.2.7 Data Analysis for FCS

For live-cell FCS measuring nuclear proteins, two biophysical models are often applied to fit the autocorrelation function and characterize the molecular diffusion. The first one is anomalous diffusion model:

$$G(\tau)=\frac{1}{N}\times\left(1+\left(\frac{\tau}{\tau_D}\right)^{\alpha}\right)^{-1}\times\left(1+\frac{1}{K^2}\times\left(\frac{\tau}{\tau_D}\right)^{\alpha}\right)^{-\frac{1}{2}}$$

where α is the correcting factor to offset the diffusion hindrance caused by the crowded environment in cell nucleus. The other one is two-component 3D diffusion model:

$$G(\tau)=\frac{1}{\langle N\rangle}\times\left((1-y)\times\left(1+\frac{\tau}{\tau_D^{free}}\right)^{-1}\times\left(1+\frac{1}{K^2}+\frac{\tau}{\tau_D^{free}}\right)^{-\frac{1}{2}}+y\times\left(1+\frac{\tau}{\tau_D^{bound}}\right)^{-1}\times\left(1+\frac{1}{K^2}+\frac{\tau}{\tau_D^{bound}}\right)^{-\frac{1}{2}}\right)$$

where y is the fraction of bound molecules. In experiments of probing MBD3-GFP, both models were considered and tested as necessary to minimize the fitting residual (χ^2) in

least-square regression algorithm. When fitting with two-component 3D diffusion model, a reference value for τ_D^{free} was predetermined from the cytoplasmic MBD3-GFP.

In parallel, the raw autocorrelation data were also analyzed by the maximum entropy method for FCS (MEMFCS) to validate the robustness of FCS measurement. MEMFCS is a multi-component model to simultaneously minimize χ^2 and maximize entropy S without a fitting priori.[93] The MEMFCS output thus is a continuous distribution of τ_D instead of a presumed number of diffusing components (Figure 2.4). The *G(τ)* function in MEMFCS is expressed as:

$$G(\tau) = \int a_i \left(1+\frac{\tau}{\tau_D}\right)^{-1} \left(1+\frac{1}{\kappa^2}\times\frac{\tau}{\tau_D}\right)^{-\frac{1}{2}} d\tau_D$$

where $S = \sum_{i=1}^{n} p_i \ln p_i$, $p_i = \frac{a_i}{\sum a_i}$. MEMFCS is particularly useful because of its capability to unbiasedly reflect the overall status of molecular diffusion in a complex biological system.

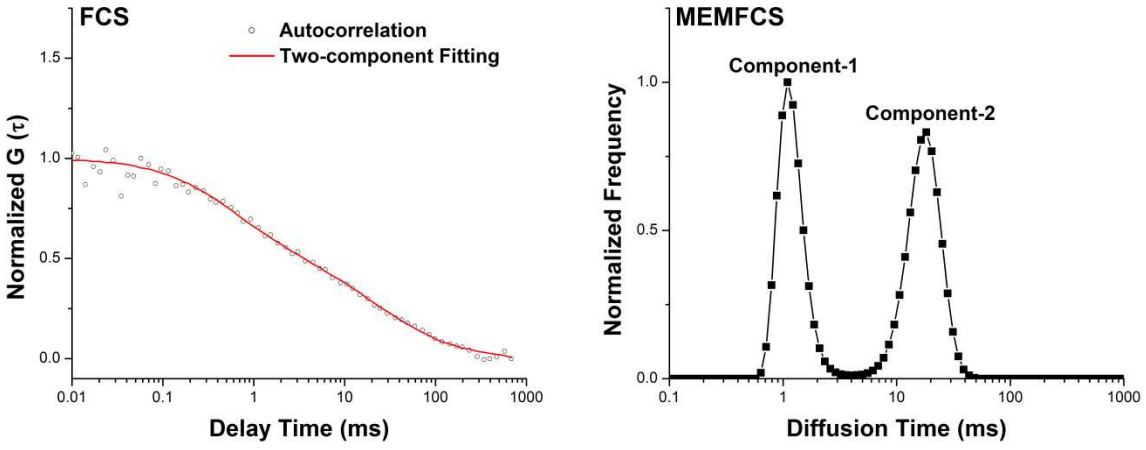


Figure 2.4 Processing FCS data by the standard two-component fitting and MEMFCS.

2.3 Results

2.3.1 Active DNA Demethylation under Hypoxia

To verify the type of DNA demethylation under acute hypoxia, all the four cytosine derivatives (*i.e.*, 5mC, 5hmC, 5fC, and 5caC) were quantified by enhanced ELISA (Figure 2.5A). In 8 hours a global reduction of all modifications was noted, indicating the occurrence of active DNA demethylation. From the decreasing patterns of 5fC and 5caC, two peak periods of active DNA demethylation were suggested: the first and the last 8 hours intermediated by a plateau period. During the peak periods, the active demethylation pathway is in full action because 5fC and 5caC are massively depleted by the TDG enzyme (Figure 2.5B). Considering the heterogeneous cellular responses to the hypoxic stress, these two demethylation periods might be attributed to distinct cell populations, *i.e.*, hypoxia-sensitive and hypoxia-insensitive cells, in which the demethylation rates are not synchronous. The 8-16 h plateau period representing a flat change in the levels of 5hmC/5fC/5caC implies a transition, regarding the contribution to the global hypomethylation, from hypoxia-sensitive cells to hypoxia-insensitive cells. Although precise quantification of 5fC and 5caC is difficult, the enhanced ELISA is capable of precisely determining as low as 1.7% of genomic 5mC and 0.027% of genomic 5hmC in HeLa cells (Figure 2.6), which is in excellent agreement with previous reports and validates the developed ELISA method.

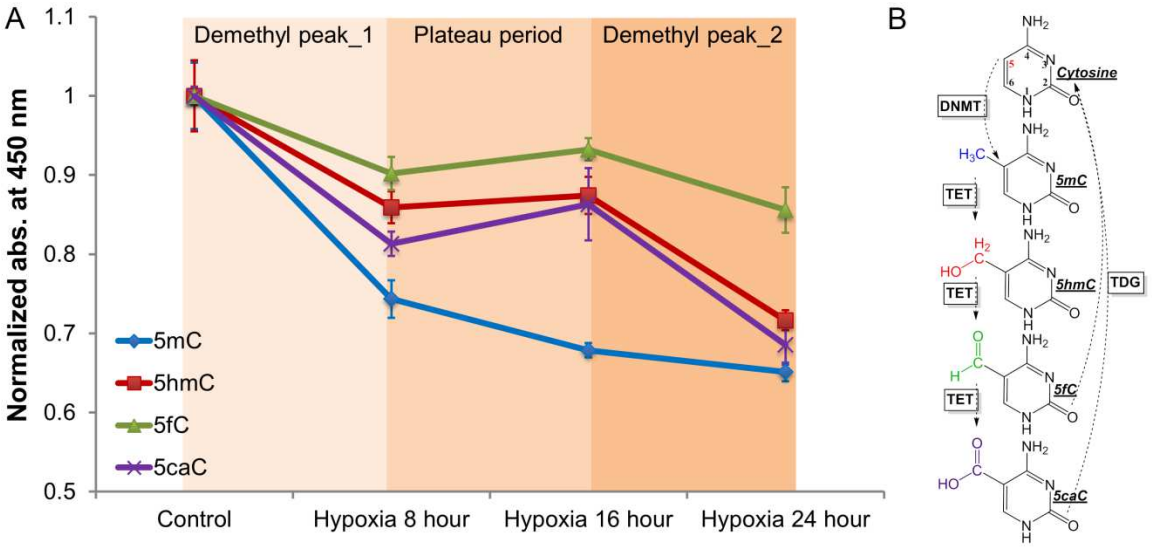


Figure 2.5 Multiplex quantification of 5mC, 5hmC, 5fC, and 5caC indicates a hypoxia-induced active DNA demethylation. (A) Normalized quantities of these four DNA methylation marks from enhanced ELISA. (B) Enzymatic pathway for active DNA demethylation.

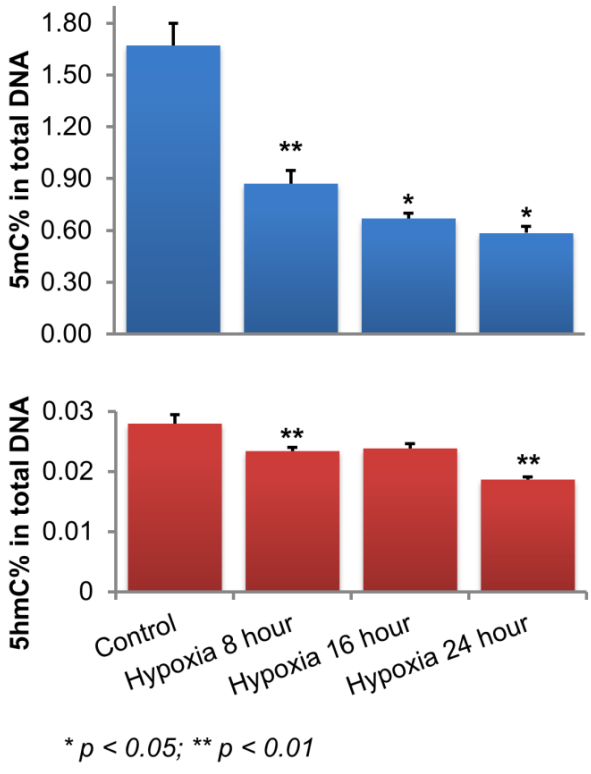


Figure 2.6 Quantitative changes of 5mC and 5hmC under hypoxia.

2.3.2 MBD3 Detachment as Putative Indicator for Active DNA Demethylation

The diffusion characteristics of MBD3-GFP in normoxia cells were determined by FCS: 14.49 ± 0.40 ms in nucleus and 3.10 ± 0.27 ms in cytoplasm. The much slower diffusion in nucleus, as expected, was due to the association with DNA methylation site and other epigenetic co-factors. The diffusion rate of MBD3-GFP was further compared when hypoxia or 3 μM decitabine treatment (agent inducing passive DNA demethylation) (Figure 2.7) was applied. Decitabine achieves a demethylation effect by incorporation into the newly synthesized DNA strand and trapping DNMT1, and so it would not influence the already bound MBD3 in a relatively short time (*e.g.*, within one to two cell cycles). Consistently, after 1 day treatment with decitabine the average diffusion time of MBD3-GFP (13.43 ± 0.57 ms) was nearly the same as in control condition whereas a significantly faster diffusion time (9.89 ± 0.54 ms) was observed in hypoxic cells after 20 hours (Figure 2.8).

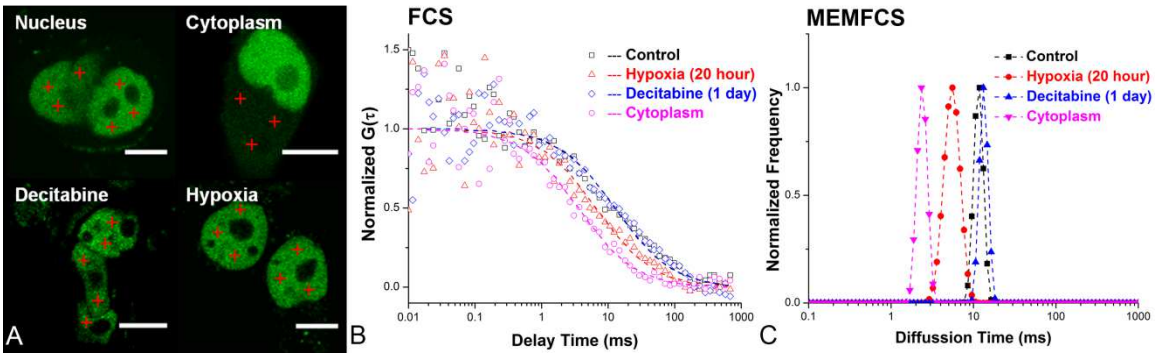


Figure 2.7 FCS measurement for the MBD3-GFP diffusion in single living cells. (A) Data were collected from cell compartments with a proper MBD3-GFP expression, followed by fitting with (B) 3D diffusion models and (C) MEMFCS.

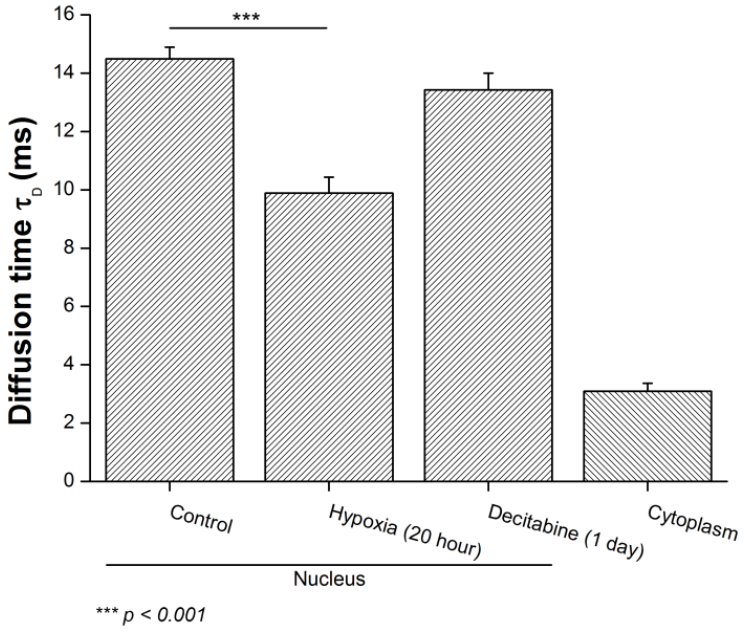


Figure 2.8 Summary of the MBD3-GFP diffusion times (mean + S.E.M., n > 50). Statistical significance was determined by ANOVA ($p = 1.35 \times 10^{-8}$).

2.3.3 Diffusion of MBD3 in Hypoxia-Sensitive and Hypoxia-Insensitive Cells

Following establishing the MBD3 diffusion as an indicator for active DNA demethylation, further experiments were performed to inspect its detailed behavior during the period of hypoxia. Interestingly, based on the MBD3 diffusion, two groups of cells can be distinguished in the first 4 hours of hypoxia. Despite that the majority of cells did not show a radical MBD3 detachment (diffusion time is about 14 ms), in ~30% of the cells the diffusion time of MBD3 displayed a steep reduction (diffusion time = 2.95 ± 0.36 ms). From the microscopic images, the latter group of cells exhibited the morphological properties akin to acutely damaged cells (Figure 2.9A), which was later confirmed by the propidium iodide (PI) uptake experiment (data not shown). Hence, these cells were termed as hypoxia-sensitive cells. Since the viscosity of nucleoplasm is nearly identical to that of cytoplasm,[94] it can be inferred that MBD3 quickly and completely dissociated from the

DNA binding sites in hypoxia-sensitive cells. On the contrary, the diffusion of MBD3 in other hypoxia-insensitive cells experienced a mild and gradual detachment process, as inferred from the slow shift in the distribution of diffusion times (Figure 2.9B). Even after 24 hours, MBD3 in these cells did not seem to be fully detached (diffusion time = 9.80 ± 0.75 ms) (Figure 2.9C). This phenomenon indicates a gentle weakening of the MBD3-DNA association, probably due to a retard activation of the demethylation pathway (*i.e.*, time required for an enhanced expression of TET and TDG proteins by hypoxia-inducible factors or pathways). Taken together, it can be concluded that the hypoxia-induced active DNA demethylation leads to different diffusivities of MBD3 based on the cell state and susceptibility. Hypoxia-sensitive cells predominantly contribute to the initial, massive DNA demethylation in the population while hypoxia-insensitive cells are involved in the long-term, chronic DNA hypomethylation. Notably, the detachment of MBD3 serves as a sensitive measure of active DNA demethylation that could result in a premature cell death, especially to hypoxia-sensitive cells. Starting from the plateau period, the counts representing fully detached MBD3 in FCS measurements substantially decreased from ~30% to ~10% (Figure 2.9C, inset), implicating the elimination of severely damaged cells from the population.

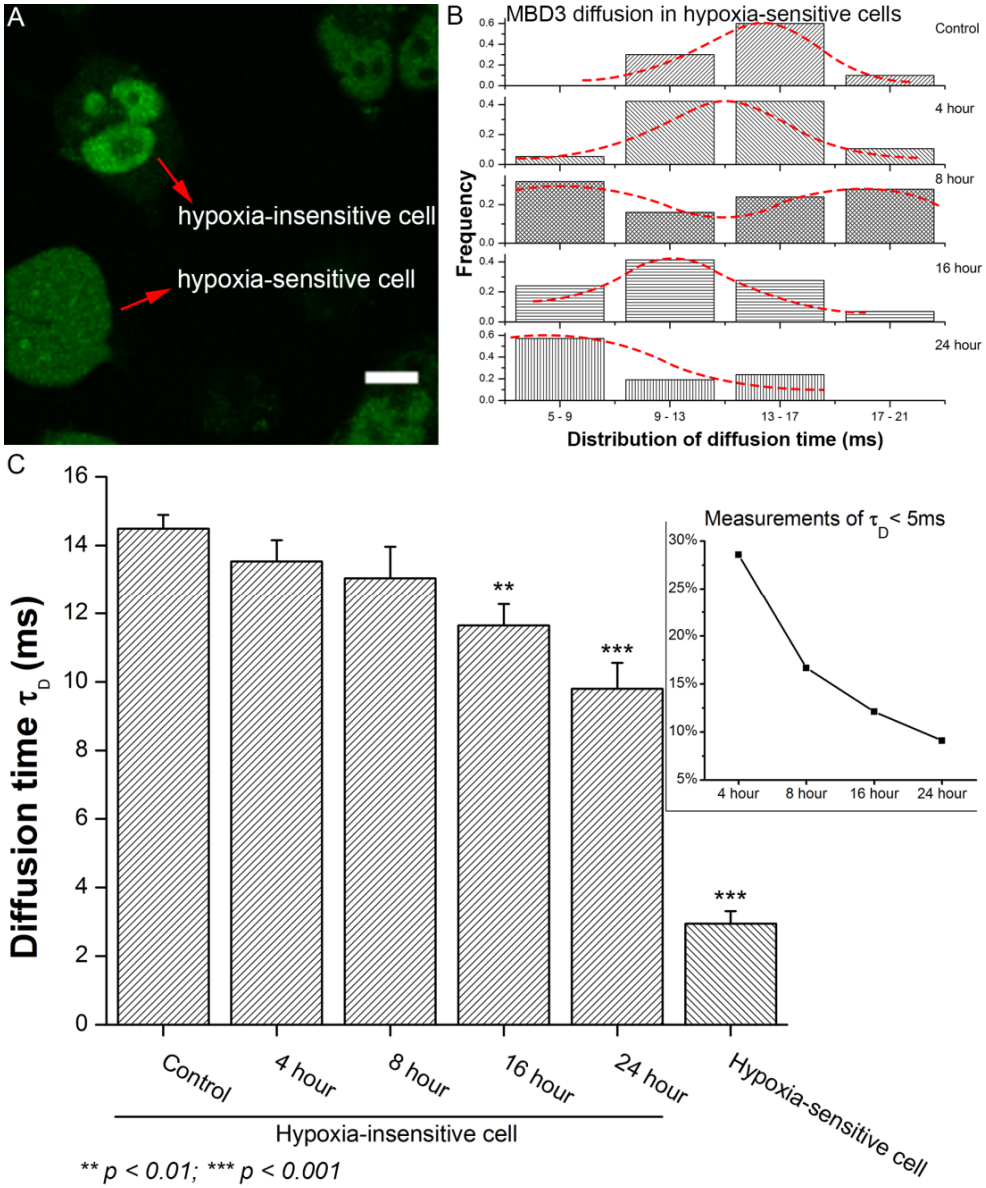


Figure 2.9 Time-course change of the MBD3 diffusion in hypoxia-sensitive and hypoxia-insensitive cells. (A) Morphology of the cells responding variedly to hypoxic stress. (B) The gradual shift of the MBD3 diffusion in hypoxia-insensitive cells. (C) Summary of the diffusion characteristics (by ANOVA tests). Percentage of the measurements representing hypoxia-sensitive cells is provided (inset).

2.4 Discussion

Even though global DNA hypomethylation has been characterized as one of the typical epigenetic events that drive a cancerous transformation, increased research emphasis

has been placed on the mechanistic understanding of promoter hypermethylation in tumor suppressor genes. As the oxidation of 5mC and its role in DNA demethylation is being elucidated, scientists start to recognize that the dynamic turnover of DNA methylation plays a pivotal role in embryogenesis, differentiation, and oncogenesis. This notion raises the necessity of inspecting DNA demethylation in single living cell.

Utilizing single-molecule fluorescence techniques, this study paves the way for studying active DNA demethylation by monitoring the diffusion rate of MBD3 in real time, which is also the pioneering report on the spatiotemporal behavior of MBD proteins in single living cells. In contrast to other MBD proteins that have an nM-level dissociation constant (K_d) with DNA methylation site, the K_d between MBD3 and its binding substrate is at the μM level.[35] Hence, MBD3 is more sensitive to be detached from chromatin in the cells exposed to a triggering stimulation for active DNA demethylation, while the *in vivo* binding substrates preferred by MBD3 needs further investigation. FCS has the quantitative ability to sense minute changes in the diffusion of intracellular biomolecules and thus is a fitted tool to explore the dynamics of DNA demethylation. Under acute hypoxia, active DNA demethylation is initially characterized by an intense detachment of MBD3 in hypoxia-sensitive cells, leading to the first demethylation peak and premature cell death within the first 4-8 hours. In hypoxia-insensitive cells the demethylation process is more relaxed as reflected by the gradual increase in the MBD3 diffusion. To maintain a natural cell state and function, all cytosine derivatives have to be maintained at a tolerable level, partly achieved by the protection by MBD3 at relevant sites. Previous studies have shown that depletion of MBD3 could cause the mice to die in early embryogenesis, signifying the importance of MBD3 in sustaining the overall abundance of DNA methylation.[44] Fu-

ture research may be focused on obtaining a high-resolution distribution landscape of all MBD proteins in the human genome and determining the MBD-dependent susceptibility to active DNA demethylation. In addition, the connections among epigenomic reshaping, hypoxic tolerance, and oncogenic evolution await more in-depth research (Figure 2.10).

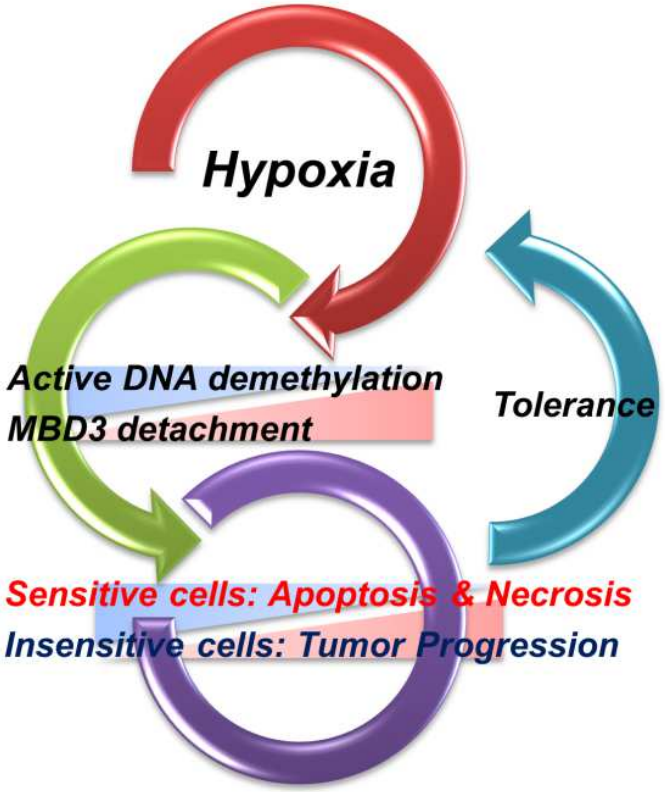


Figure 2.10 Influence of hypoxia on DNA methylation and the fate of cancer cells.

As the sophisticated DNA demethylation network is becoming more understood, it elicits a relevant question: does the hypoxia-induced demethylation reaction terminates at 5hmC? In human cells the removal of 5hmC could be processed by two possible mechanisms: 1) through deamination of 5hmC to 5hmU (5-hydroxymethyluracil) by AID and APOBEC enzymes, followed by BER; 2) or the TET-TDG-mediated oxidation and BER.

Evidently, data from enhanced ELISA supports the latter pathway in that 5hmC, 5fC, and 5caC are concurrently reduced along with 5mC. Moreover, the detachment of MBD3 indicates the TDG-mediated BER that involves a physical removal of cytosine derivatives. However, which specific TET proteins involved in this process remains to be identified. Previously, most of the cancerous hypomethylation were found in the short and long interspersed nuclear elements (SINEs and LINEs) as well as repeat sequences that were long thought as non-coding "junk" DNA.[95, 96] Revealed by the Encyclopedia of DNA Elements (ENCODE) project in 2012, nearly 80% of the human genome is actually functional,[97] necessitating re-interrogation of the epigenetic state at these regions. In recognition of that the patterns of DNA hypomethylation could be largely varied in different cancers, an emerging thought is to study the DNA demethylation in a locus-specific manner so that targeted intervention or therapy can be designed. To fulfil this, methodologies to capture the key moments in a demethylation event and to differentiate heterogeneous cellular responses are of great value.

CHAPTER 3. BEHAVIOR AND FUNCTION OF MBD3 IN DNA METHYLATION HOMEOSTASIS

3.1 Introduction

DNMTs are responsible for adding methyl marks to CpGs; however, the traditional view of classifying the three major human DNMTs to maintenance DNMT and *de novo* DNMT appears to be challenged by cumulative evidence that expounds on a considerable overlap between their functions,[98-100] which in turn inspires a revisiting of the existing knowledge about the DNA methylation homeostasis. Over the past decade it has been observed that the maintenance of DNA methylation rarely follows a rigorous CpG-by-CpG fidelity, but adopts a stochastic model in which only the average density of DNA methylation within a certain region of sequence needs to be conserved.[101, 102] To achieve a balance, the twin forces of methylation and demethylation contend in equilibrium but also work by side. Although the TET family proteins are the only enzyme that catalyzes the active DNA demethylation in mammalian cells, they primarily enrich in germ cells and early embryogenesis instead of in somatic cells.[103] In light of this, it is of great interest to resolve the question on whether additional factors or machineries exist to facilitate the precision of maintenance methylation with limited TETs as well as to prevent a possible over-methylation by DNMTs in the DNA methylation homeostasis.

MBD3 belongs to a family of nuclear proteins in close relation to DNA methylation, but it involves quite different chromatin association and regulatory mechanisms as com-

pared with other MBD proteins.[44, 104] In spite of sharing over 70% of its sequence with MBD2, MBD3 was found not to be a strong binding protein for 5mC.[33, 105, 106] In fact, as a constitutive component in the Mi-2/NuRD complex, MBD3 preferentially localizes at CG-rich promoters and enhancers relating to active genes whose methylation state is poised to frequent turnovers.[107, 108] More interestingly, it has been reported that overexpressing MBD3 could induce a global DNA demethylation,[42, 109] and a co-localization of MBD3, MBD2 and DNMT1 could appear at the DNA replication foci to associate with hemi-methylated DNA.[110] These intriguing findings motivate this study to re-evaluate the underpinning functions of MBD3 in the DNA methylation homeostasis.

In previous chapters, the potency of single-molecule techniques in uncovering epigenetic dynamics was introduced and the diffusion characteristic of MBD3 was successfully correlated with active DNA demethylation by FCS. In this chapter, a comprehensive set of single-molecule fluorescence tools is applied to dissect the behavior of MBD3 from the cell cycle perspective. The research hypothesis regarding the MBD3-MBD2-DNMT1 cooperation lies in the *in vivo* demethylating potential of MBD3 to buffer the activity of DNMT1 during DNA replication as contributing to the epigenetic inheritance of a loci-specific methylation pattern.

In order to obtain an in-depth understanding, several biophysical attributes of MBD3 are characterized – intracellular dynamics, chromatin association, binding stoichiometry, and molecular interactions in each stage of the cell cycle. Fluorescence lifetime correlation spectroscopy (FLCS), an elegant derivative of FCS, is first used to quantify the real-time diffusion of MBD3 in different cell phases and the results suggest that MBD3 has distinct binding partners in the G1-phase and G2-phase. Furthermore, the DNA-binding

stoichiometry of MBD3 is analyzed by PCH that corroborates a salt-dependent “two-site sequential binding” mode in the G1-phase nucleus. In the G2-phase, the co-localization of MBD3, MBD2, and DNMT1 is verified with FLIM-FRET. When depleting MBD3 by siRNAs, the global DNA methylation increases along with promoter hypermethylation in a number of cell cycle-related genes. Altogether, this work constitutes a pioneer effort to revise the model of the DNA methylation homeostasis and paves the way for future research aimed to better interpret the epigenetic inheritance.

3.2 Methods and Materials

3.2.1 Cell Cycle Synchronization

G1-enriched HeLa cells were obtained by treatment with 400 µM L-mimosine for 20-24 h. To achieve a maximum of S-phase population, cells were blocked with 2 mM thymidine for 20 h and released in fresh medium for 10 h, followed by another 20 h thymidine blocking and then culturing with fresh medium for 2-6 h (early S-phase: 2 h; middle S-phase: 4 h; late S-phase: 6 h). Culture of G2-arrested cells was obtained by incubation with fresh medium containing 0.1 µg/mL Colcemid (Life Technologies) for 6 h after two rounds of thymidine blocking. Flow cytometry was used to validate the cell cycle synchronization.[111] In general, cells were fixed with 70% ethanol first and stained with PBS buffer containing 20 µg/mL PI, 0.1% Triton X-100, and 200 µg/mL RNase A. The fluorescence intensity of PI in each single cell was determined by a FC500 MPL system (Beckman Coulter). At least 15,000 cells were collected for one measurement and the generated cell cycle histogram was fitted by the Watson pragmatic model in the Flowjo software (Figure 3.4A).

3.2.2 Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)

Total RNA was extracted and reverse-transcribed with RNeasy Mini Kit (Qiagen) and iScript cDNA Synthesis Kit (Bio-Rad), respectively. Before cDNA synthesis, the genomic DNA was removed by DNase I. PCR reaction was performed in a StepOnePlus system using SYBR Green PCR Master Mix (Life Technologies). *ΔΔCt* method was used to normalize the transcription of target genes to the internal control gene GAPDH. The PCR primers for GAPDH, DNMT1, DNMT3A, and DNMT3B were designed with the Primer-BLAST online tools (http://www.ncbi.nlm.nih.gov/tools/primer-blast/).

Table 3.1 PCR primers used in Chapter 3

GAPDH	(F)CAGCCTCAAGATCATCAGCA (R)TGTGGTCATGAGTCCTTCCA
DNMT3A	(F)TATTGATGAGCGCACAAGAGAGC (R)GGGTGTTCCAGGGTAACATTGAG
DNMT3B	(F)GGCAAGTTCTCCGAGGTCTCTG (R)TGGTACATGGCTTTTGGATAGGA
DNMT1	(F)TACCTGACGACCCTGACCTC (R)CGTTGGCATCAAAGATGGACA
MBD3[112]	(F)GGCCACAGGGATGTCTTTTA (R)TTGACCTGGTTGGAGGAGTCG

3.2.3 FLCS and PCH

In FLCS, separating the “genuine” molecular diffusion from the overall fluctuation is achieved by deconvolution of the TCSPC histogram with a multi-exponential model:

$$I(t) = \sum_{i=1}^{n} a_i \cdot e^{-\frac{t}{\tau_i}}$$

where τ_i is the characteristic decay lifetime of the *i*th component and a_i is its amplitude. A filter function is applied to assign the probability contribution of a single photon to the autocorrelation function for the *i*th component with an optimal weight determined by its

TCSPC channel number *j*. For a given channel *j*: $\sum_{i=1}^{L} f_j^{(i)} = 1$, where *L* is the number of components, and the weight for each component could have an absolute value greater than 1 or be a negative value. These filter functions are numerically calculated by a matrix derived from the measured fluorescence intensity I_j and the decay pattern $p_j^{(i)}$.[113] Hence, the autocorrelation function for a specific species (*i*) can be expressed as:

$$G^{(i)}(\tau) = \frac{\left\langle \sum_j f_j^{(i)} I_i(t) \cdot \sum_j f_j^{(i)} I_i(t+\tau) \right\rangle}{\left\langle \sum_j f_j^{(i)} I_i(t) \right\rangle^2} - 1$$

Taking advantage of the TCSPC filtering, a distorted autocorrelation fluctuation due to the detector noises (*e.g.*, thermal noise, afterpulsing) and other parasitic noises (*e.g.*, elastic and Raman scattering) occurring at the millisecond level can be effectively removed.[114, 115]

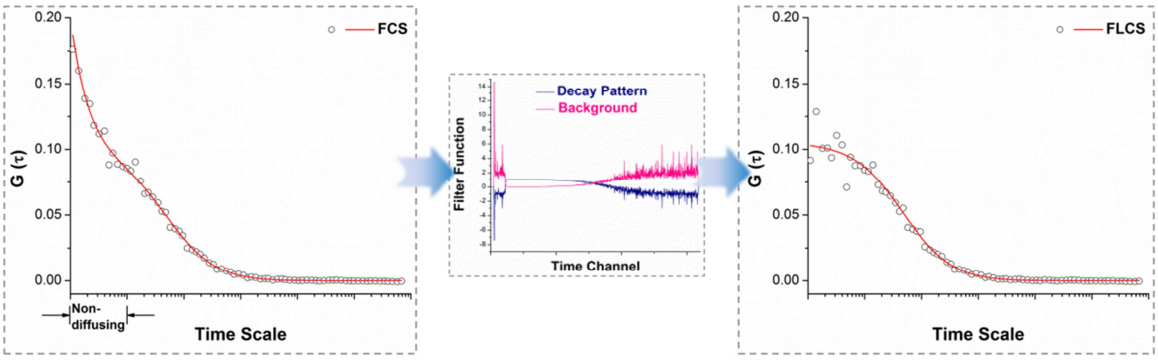


Figure 3.1 TCSPC filter function eliminates the impact of the short-timescale noises to an autocorrelation function.

The basic theory and working mechanism of PCH is elaborated in Chapter 1.2.2. To test the PCH algorithm, Cy3 tetramer was successfully differentiated from Cy3 monomer.

In this experiment, a ten-fold concentrated solution of Cy3 monomers was prepared and it was impossible to distinguish it from the tetramer solution by simply comparing the ensemble fluorescence intensities (otherwise the average brightness could give out an opposite conclusion). However, by mathematically eliminating the influence from the concentration difference, the single-molecule brightness of Cy3 tetramers was determined to be almost four times of that of Cy3 monomers, thus verifying the power of PCH.

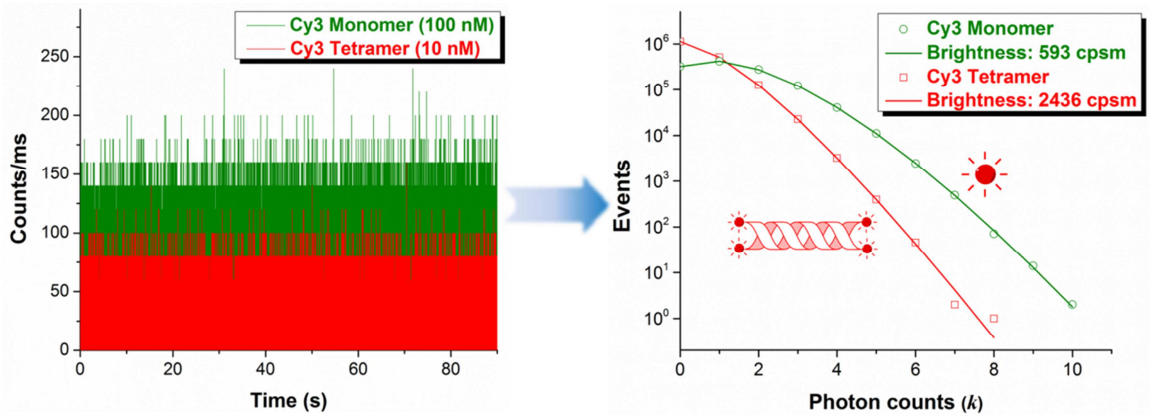


Figure 3.2 *In vitro* demonstration of PCH. Ensemble fluorescence traces (left panel) with an integration window of 50 μs are translated to photon counting histograms followed by fitting with super-Poissonian algorithm (right panel), yielding single-molecule brightness.

3.2.4 Immunofluorescence and FLIM-FRET

Considering the sensitivity of FRET to the inter-molecule distance, to minimize the size effect from antibodies (~10 nm), the primary antibodies (anti-MBD1/MBD2/MBD3-IgG, Abcam; anti-DNMT1-IgG, Novus Biologicals; anti-5hmC/5fC/5caC-IgG, Active Motif) were directly labeled with the FRET-pair dyes including: MBD3-Alexa488 (donor) and MBD1/MBD2/DNMT1/5hmC/5fC/5caC-Alexa555 (acceptor) by using APEX Antibody Labeling Kits (Life Technologies). For immunofluorescence staining, cells were

fixed with 4% ice cold paraformaldehyde (PFA) for 15 min and then rinsed three times with clean PBS buffer. Cell membrane was permeabilized with 0.4% Triton X-100 for 30 min followed by rinse for 15 min. To stain 5hmC, 5fC, and 5caC, DNA was denatured by 1 N hydrogen chloride at 37°C for 30 min. Cells were then blocked with PBS containing 5% goat serum and 0.3% Triton X-100 for 1 h at room temperature. Primary antibodies were diluted 50-100 times in PBS containing 1% BSA and 0.3% Triton X-100 (to a concentration of ~5 nM) and incubated with cells overnight at 4°C. In the next day cells were rinsed 3 times with fresh PBS prior to FLIM imaging. The FLIM-FRET theory and instrumentation are elaborated in Chapter 1.2.2. For a 150 × 150 pixels FLIM image obtained from one scanning, the laser dwelling time at each pixel is about 0.8 ms that is far longer than most fluorescence lifetimes (ns-level). TCSPC can be generated from a single pixel or all the pixels in an image (Figure 3.3B); hence the FLIM measurement can reach single-molecule resolution if few fluorophores exist in one pixel.

3.2.5 Co-Immunoprecipitation (Co-IP) and Western Blot

Co-IP was conducted with the Pierce Co-IP Kit (Thermo Fisher Scientific) referring to an established protocol.[116] In brief, anti-GFP monoclonal antibody (primary amine-free, Pierce) or control IgG was first immobilized using the AminoLink Plus coupling resin. 5 × 10^6 cells transfected with MBD3-GFP in different cell phases were lysed (25 mM Tris, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 5% glycerol; pH = 7.4) and about 1.5 mg precleared lysate was incubated with the antibody-coupled resin at 4°C overnight. After elution, the sample was analyzed by Western immunoblotting. The input proteins were separated by 10% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

followed by membrane transferring and incubation with the primary antibodies. The final signal was developed by Pierce Fast Western Blot Kit and SuperSignal West Pico Substrate.

3.2.6 Transfection of siRNA and Methylation Analysis

MBD3 siRNA, scrambled siRNA, and transfection reagents for knockdown experiments were obtained from Santa Cruz Biotechnology. 50 pmol siRNA was used for a 35 mm Petri dish in which cells reached 70% confluence, according to the manufacturer's instruction. Transfection lasted for 24 h followed by sub-culturing with normal fresh medium for another 72 h. DNA from control and siRNA-treated cells was extracted and purified with DNeasy Blood & Tissue Kit (Qiagen), according to the manufacturer's instruction. Global DNA methylation was quantified with MethylFlash Methylated DNA Quantification Kit (Epigentek). Briefly, 100 ng DNA was deposited to the bottom of the assay well and incubated with capture antibody and detection antibody, sequentially. The final colorimetric signal was normalized to a standard input DNA to calculate the global methylation percentage. 22 genes participating in cell cycle progression were assessed for their promoter CGI methylation by using EpiTect Methyl II Signature PCR Array (Qiagen). The working mechanism for this assay is to digest an equal amount of DNA with methylation-sensitive endonuclease (the unmethylated DNA will be eliminated) or methylation-dependent endonuclease (the methylated DNA will be eliminated). Upon a PCR amplification with specific primers targeting the characterized CGIs of the chosen genes, the ratio of methylated promoters to unmethylated promoters in the sample DNA can be determined.

3.3 Results

3.3.1 Localization Preference and Cell Cycle-Dependent Expression of MBD3

The localization preference of endogenous MBD3 was first probed via FLIM-FRET, to explore the association of MBD3 with the active demethylation products (*i.e.*, the 5mC derivatives). From the results, a relatively weak interaction between MBD3 and 5hmC was noted while MBD3 exhibited a prominent co-localization with 5fC and 5caC (Figure 3.3), which is consistent with the recent findings using conventional methods.[106, 117] However, this cannot fully rule out the possibility of MBD3 as a reader protein for 5hmC because: 1) FRET could be influenced by other factors such as the donor-acceptor orientation; 2) the MBD3-5hmC interaction requires some other co-factors that might block or separate the antibody recognition sites or 3) this connection might be cell type-specific.[105] Nevertheless, this set of experiments provides partial evidence to the assumption that MBD3 associates with the CpGs experiencing active transition in their methylation state. This demethylation potential could be ascribed to the MBD3-mediated further recruitment of TET proteins or a steric hindrance to impede the DNMTs catalysis.

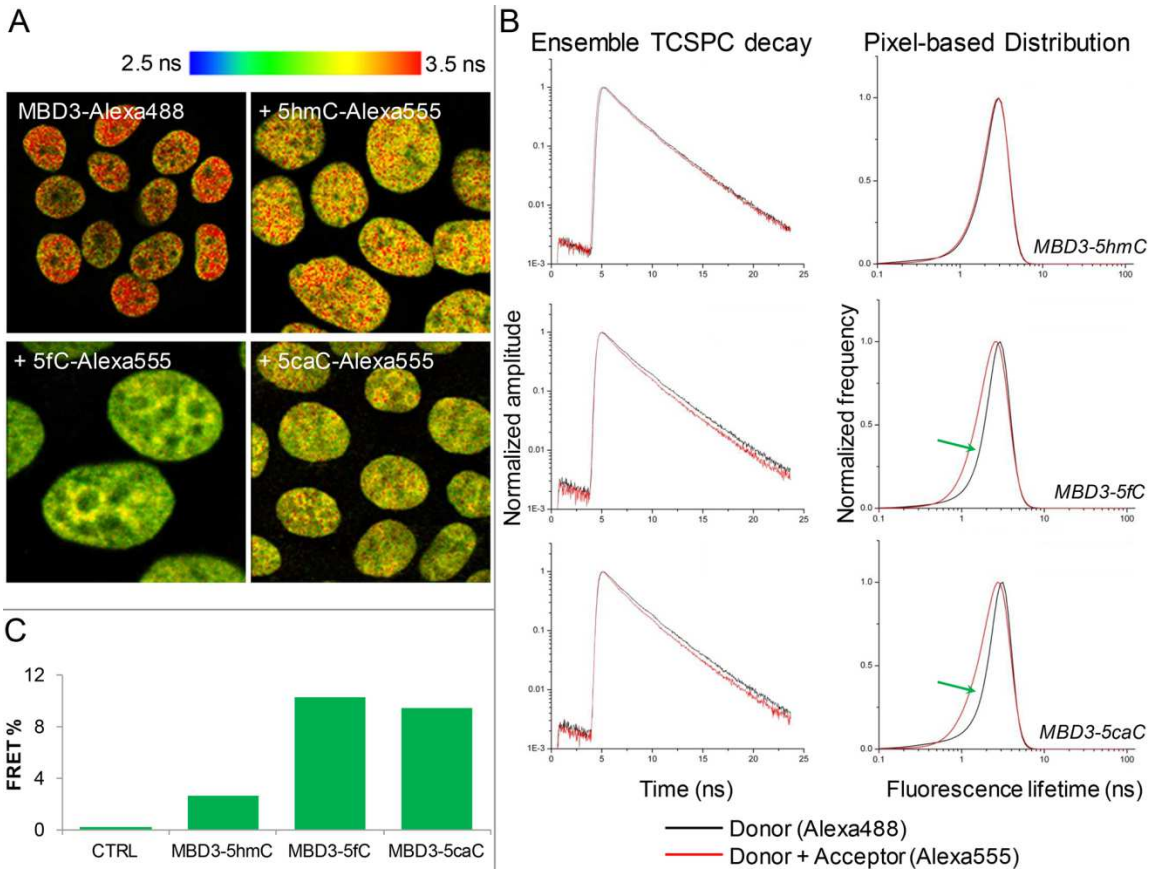


Figure 3.3 Localization preference of MBD3. (A) FLIM images of double-stained cells with directly labeled primary antibodies. (B) Fluorescence lifetime can be determined by fitting the TCSPC from the whole image (left panel), or single pixels to generate a histogram (right panel). (C) FRET efficiencies between MBD3 and different 5mC derivatives.

Next, the transcription of DNMTs and MBD3 in different cell phases was quantified. The qRT-PCR results show that all DNMTs start to increase in the S-phase and maintain their abundance till the G2-phase (Figure 3.4B), which validates the cell cycle-dependent expression of DNMTs.[118] Although it is expected that the transcription of DNMTs should peak during the S-phase rather than in the G2-phase, two critical points to bear in mind are that 1) the DNA maintenance methylation is a delayed event subsequent to DNA replication[119] and 2) in cancerous or transformed cells (*e.g.*, HeLa cells here) the regulation

of DNMTs could become more lagged due to rapid cell proliferation. Interestingly, accompanying with all DNMTs, the transcription of MBD3 also experiences a cell cycle-dependent enrichment. This concurrent transactivation provides preliminary support for a possibly underlying contention or cooperation between MBD3 and DNMTs in the DNA methylation homeostasis.

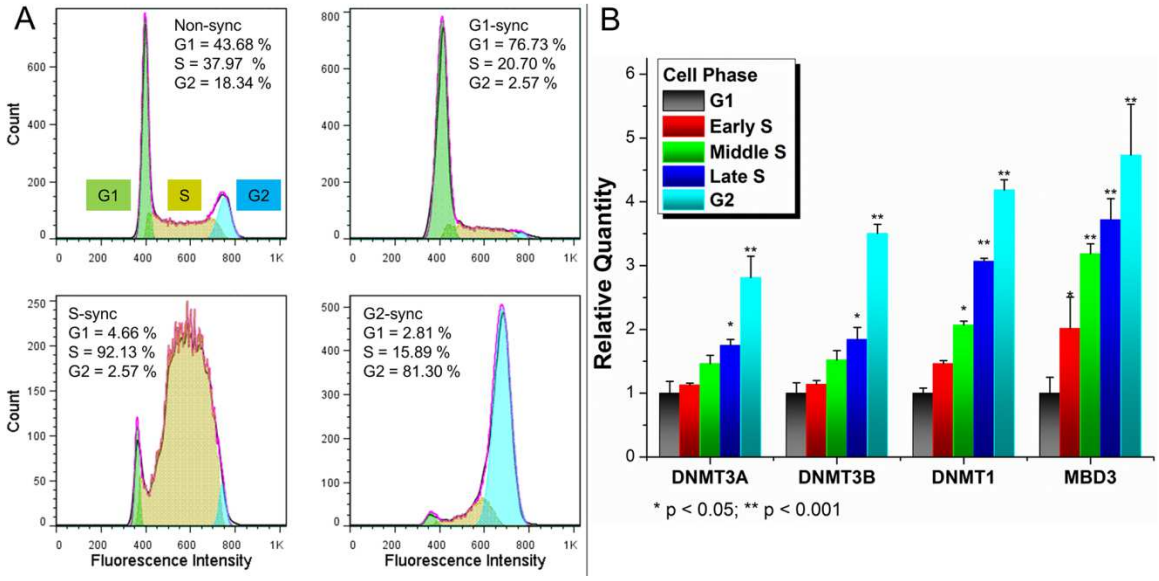


Figure 3.4 Expression of MBD3 is concurrent with DNMTs in cell cycle. (A) Cells enriched in each cell phase were harvested for qRT-PCR. (B) Transcriptional levels are presented as mean with standard deviation (n = 4). One-way ANOVA was used to determine the statistical significance.

3.3.2 The Mobility of MBD3 is Increasingly Constrained during the G1-S-G2 Transition

After transfection, the intracellular diffusion of MBD3-GFP was characterized by FLCS. In non-synchronized cells, the diffusion rate of MBD3-GFP in cytoplasm is found to be much faster (diffusion time = 3.6 ± 0.4 ms) than that in nucleus (diffusion time = 16.2 ± 1.9 ms), which is in agreement with the results in Chapter 2. The restricted diffu-

sion of MBD3-GFP in nucleus could be attributed to two mechanisms: 1) integration into the larger Mi-2/NuRD complex and 2) association with the targeting chromatin site. The cell cycle-dependent diffusion of MBD3-GFP was then determined. Compared to MBD3-GFP in the G1-phase cells, MBD3-GFP in the S-phase and the G2-phase cells exhibits a drastically slower diffusion: diffusion times are 11.5 ± 0.6 ms (G1), 26.6 ± 1.3 ms (S), and 36.9 ± 2.1 ms (G2), respectively (Figure 3.5C). Moreover, an incremental adjustment for anomalous diffusion was noted in fitting the autocorrelation functions from the G1 to S and G2 transition, indicating a strengthened chromatin interaction. Considering that the molecular diffusion is negatively proportional to the hydrodynamic radius, the data from FLCS suggests that, starting from the S-phase, the MBD3-NuRD complex may further associate with or recruit other biomolecules to relevant loci, which results in an enhanced DNA-binding affinity.

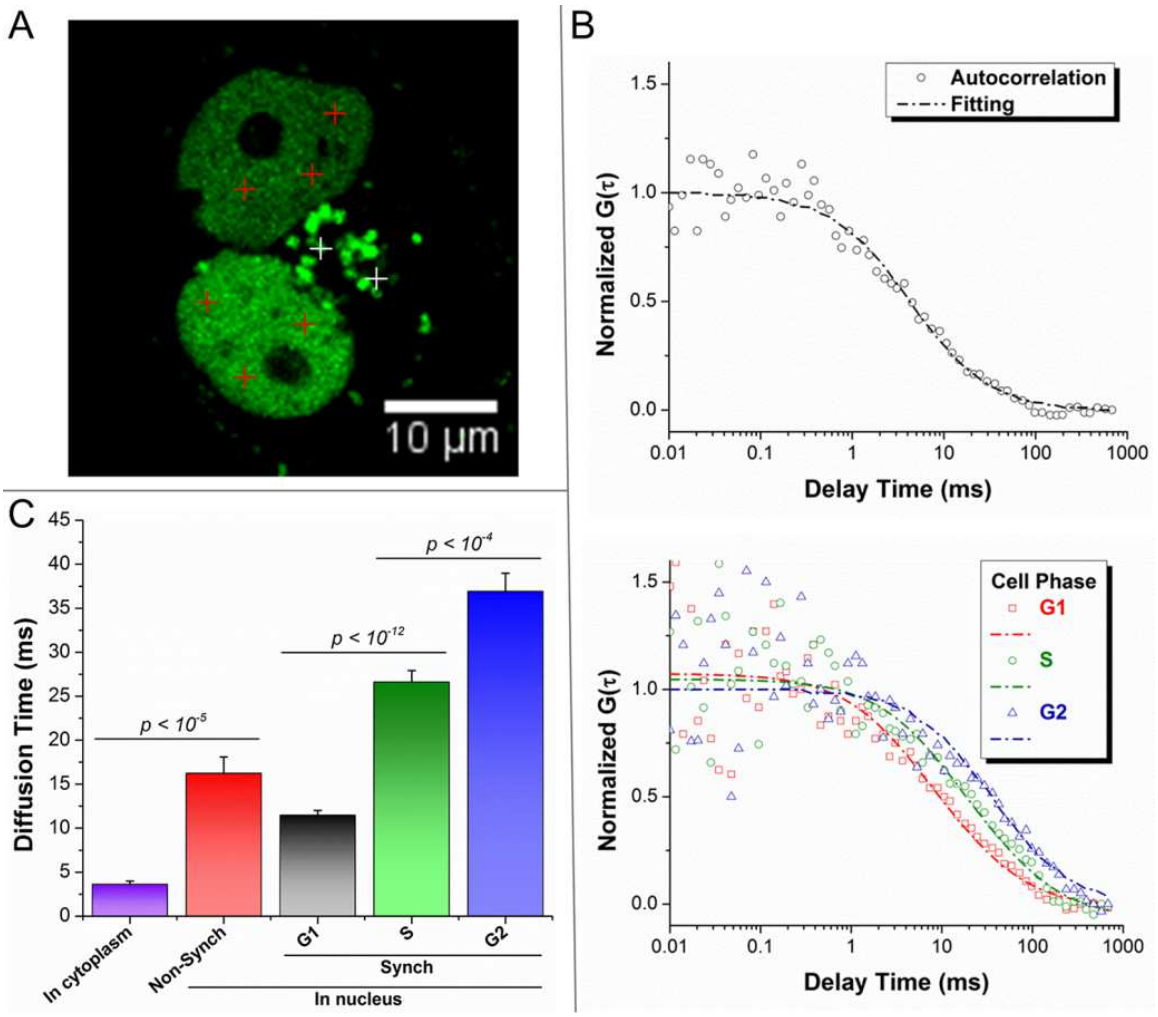


Figure 3.5 The mobility of MBD3 is increasingly restricted during the cell cycle progression. (A) FLCS data collection points from different cell compartments. (B) Autocorrelation functions in different cell phases were obtained from cytoplasm (upper) and nucleus (lower). (C) Summarized diffusion statistics and significance test by ANOVA. Each column represents data from more than 15 cells (mean + S.E.M., n > 30).

3.3.3 The DNA-Binding Stoichiometry of MBD3 Undergoes a Transition during the S-Phase

Following FLCS, the average single-molecule brightness of nuclear MBD-GFP was evaluated by PCH, which uncovered a gradual reduction in the copy number of MBD3-GFP molecules along with the cell cycle progression (Figure 3.6A). To calibrate the GFP

brightness from different stoichiometry, a series of controls were included: monomeric GFP in solution and in cell nucleus, toll-like receptor 9-GFP (TLR9-GFP), and MBD1-GFP in live cells. *In vitro* GFP monomers were measured to establish the baseline brightness that is 1341 ± 48 cpsm (counts per second per molecule) under a 2.5 μW excitation power (Figure 3.6B). Based on this value, about 8% of nonspecific dimerization was determined in living cell nucleus. Then the obtained brightness from dimerizable proteins could be assigned to a proper composition of monomers and dimers. Referring to previous reports, TLR9-GFP and MBD1-GFP were respectively chosen as dimer and monomer controls.[120] From the PCH calculation, the dimer percentages for TLR9-GFP and MBD1-GFP are 78% and 2% in HeLa cells, irrespective of cell phases. Notably, the dimer percentage of MBD3-GFP implicates that its composition stoichiometry experiences a cell cycle-dependent transition: from 84% in the G1-phase to 48% in the S-phase, and eventually to 16% in the G2-phase (Figure 3.6C). According to a recent study, MBD domain was able to dimerize at a single methylated CpG and this stoichiometric pattern was salt/cation-dependent.[121] It was thus proposed that a low-salt environment could undermine the dimerization of MBD3 in living cells. By culturing G1-phase cells in hypotonic medium, the dimer proportion of MBD3-GFP was indeed reduced (Figure 3.6D). In the hypothesis of this study, the stoichiometric transition of MBD3 in cell cycle is driven by the cooperation among MBD3, MBD2, and DNMT1 in the DNA maintenance methylation, which can be combined with the FLCS findings: during the cell cycle progression, MBD3 is gradually incorporated into a larger protein complex (inferred from the slowed diffusion rate) so that gains a stronger binding affinity with chromatin (inferred from the increasing anomalous diffusion behavior). This single-molecule behavior implies that the

difference between the inter-molecule bonding strengths of MBD3-DNA and MBD3-MBD2-DNMT1 mechanistically advance the stoichiometric transition, for which the intracellular microenvironment (*e.g.*, local salt ions) may play a pivotal role.

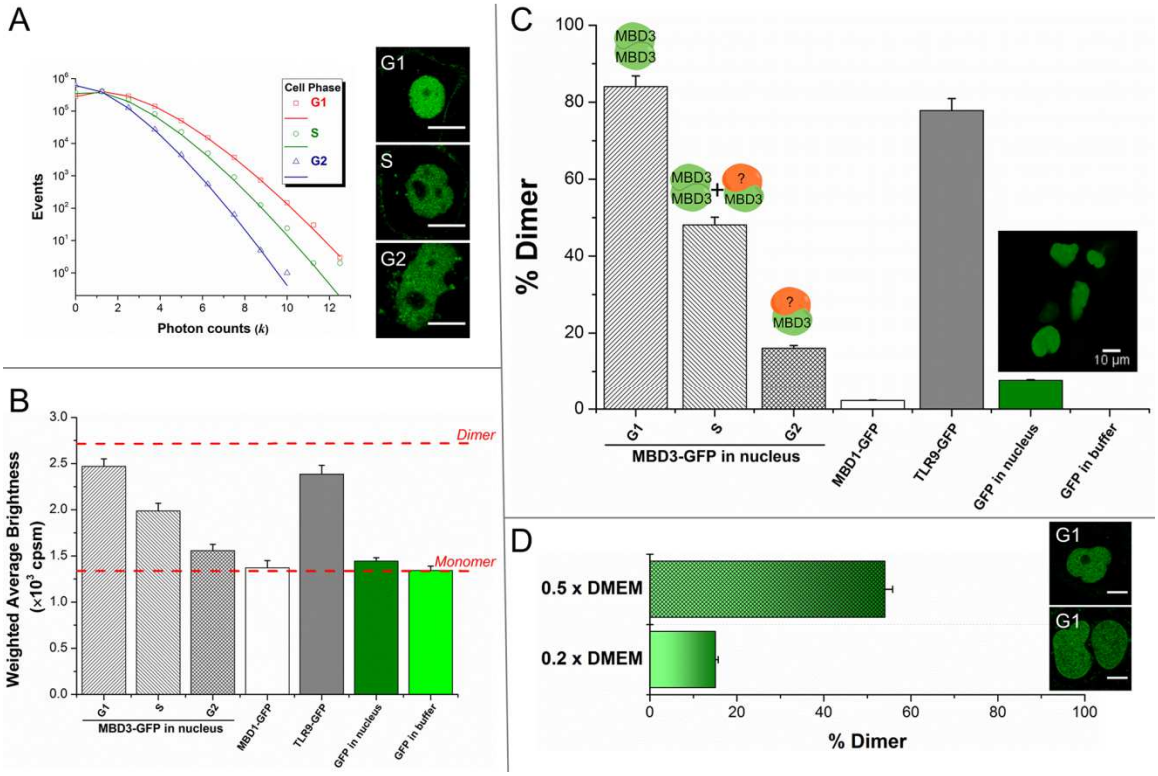


Figure 3.6 Single-molecule brightness analysis implicates a stoichiometric transition of MBD3. (A) PCH from nuclear MBD3-GFP in different cell phases. (B) Brightness of different GFP tagged proteins: pure GFP is the baseline control, TLR9-GFP is the dimer control, and MBD1-GFP is the monomer control. (C) Dimer percentage for each protein is calculated. (D) Hypotonic medium disrupts the dimer formation of MBD3-GFP in the G1-phase nucleus.

3.3.4 MBD3, MBD2, and DNMT1 Act in Concert in DNA Maintenance Methylation

Despite that a number of methods, including Co-IP, affinity electrophoresis, pull-down assay, and two-hybrid screening, are available for evaluation of molecular interactions; a grand challenge is on how to directly visualize such *in situ* information in its original context. Super-resolution optical microscopy has overcome the diffraction limit im-

posed by the physical law to achieve a sub-200 nm resolution for single-cell imaging, but the spatial scale (< 10-20 nm) for most protein-protein and protein-DNA interactions is still beyond that "super-resolution" scope. Hence, in this study FLIM-FRET was employed as a finer measure to probe the MBD3-related co-factors at the 10 nm scale. By directly labeling the primary antibodies, HeLa cells were double-stained and measured by FLIM in different cell phases (Figure 3.7A). As expected, a cell-cycle dependent occurrence of FRET was noted in both the MBD3-MBD2 and the MBD3-DNMT1 combinations, indicating a growing proportion of MBD3, MBD2, and DNMT1 assembled into a larger complex. In the G1-phase, MBD3-Alexa488 in the presence of MBD2-Alexa555 has a slightly shorter lifetime (3.34 ± 0.06 ns) than that in the presence of DNMT1-Alexa555 (3.45 ± 0.06 ns) (Figure 3.7B), confirming that some of the endogenous MBD3-MBD2 co-localization is cell cycle-independent.[5, 122] In the control combination, no FRET signal was detected between MBD3-Alexa488 and MBD1-Alexa555 (Figure 3.8). In parallel, all these interactions were validated by conventional Co-IP (Figure 3.7C).

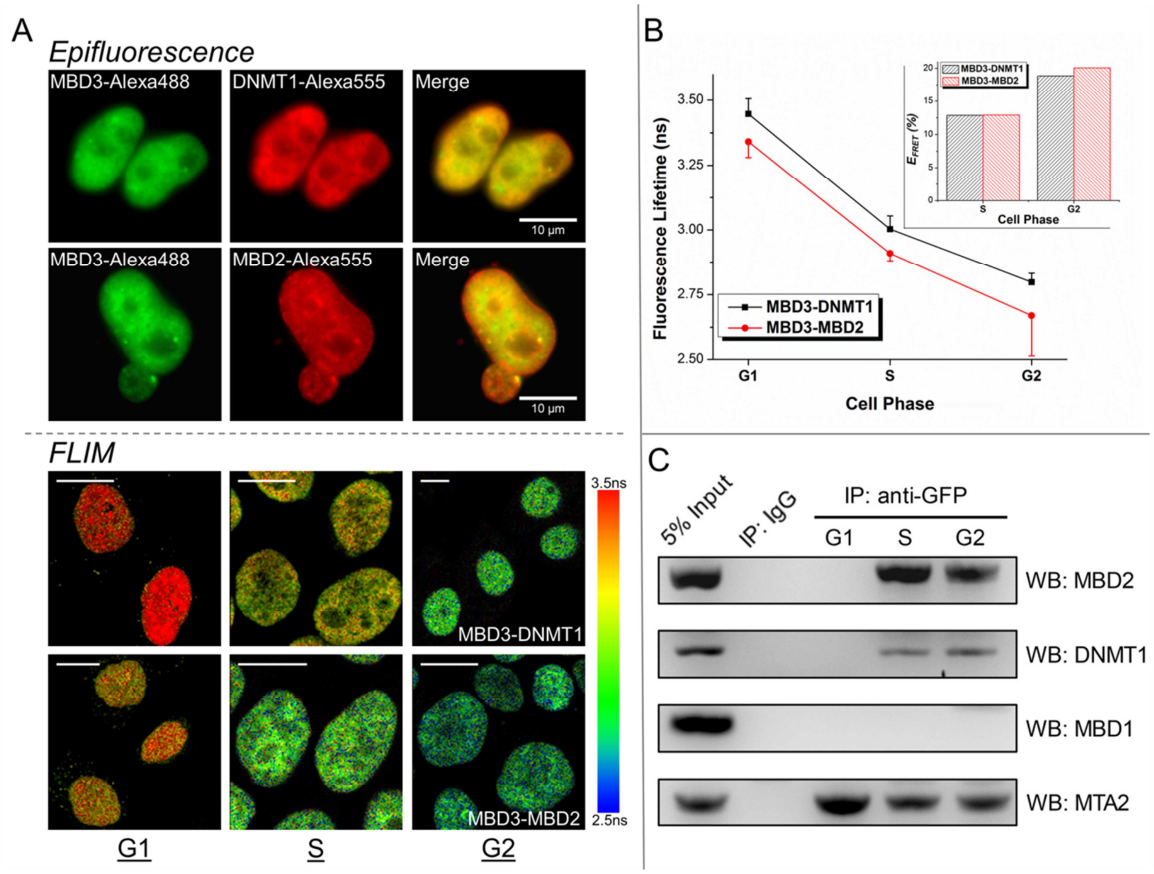


Figure 3.7 Co-localization of MBD3-MBD2-DMNT1 was assessed by FLIM-FRET. (A) Labeled primary antibodies were applied to target the protein pairs (*i.e.*, MBD3-DNMT1, MBD3-MBD2). FLIM images of MBD3-Alexa488 were obtained in different cell phases, with DNMT1 or MBD2 labeled by Alexa555. (B) Fluorescence lifetimes are presented with mean and stand deviation (n = 10 images), based on which FRET efficiencies can be calculated (inset). (C) Results from FLIM-FRET were validated by Co-IP.

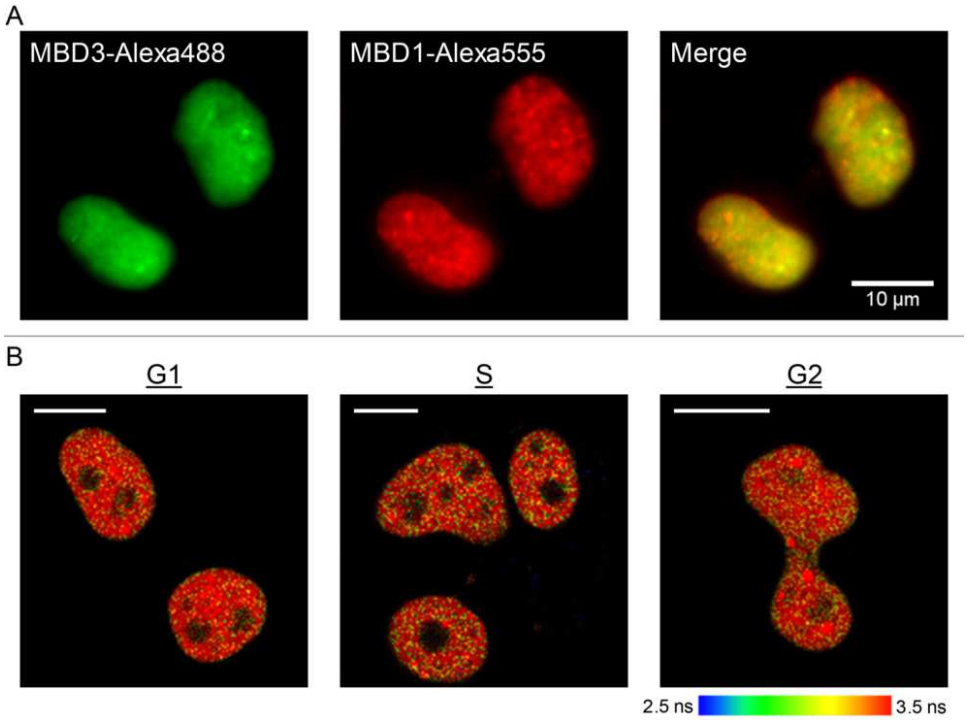


Figure 3.8 MBD3 and MBD1 are not interactive in cell cycle. (A) Double staining was performed with directly labeled antibodies. (B) FLIM of MBD3-Alexa488 was collected in different cell phases.

3.3.5 Insufficiency in MBD3 Disrupts DNA Methylation Homeostasis

To inspect the functional involvement of MBD3 in the DNA methylation homeostasis, a group of siRNAs specifically targeting the MBD3 expression were introduced into HeLa cells. Given the findings that MBD3 facilitated a hypomethylation for its binding sites, it was expected that cells would experience a global hypermethylation when MBD3 was knocked down. Using quantitative immunoassay, the global level of DNA methylation in knockdown cells was found to increase from 1.66% to 2.34% after 72 hours (Figure 3.9B). Afterwards, the methylation levels of promoter CGIs in 22 cell cycle-related genes were evaluated since MBD3 has been discovered to regulate the cell cycle progression.[123] The promoter CGIs of the selected genes, except BRCA2 and RAD9A, were de-

termined to be extensively hypomethylated or unmethylated in HeLa cells, consistent with the ENCODE dataset (retrieved from https://genome.ucsc.edu/) and previous literature.[124] After the MBD3 expression was suppressed, a ubiquitous elevation in the methylation of assessed CGIs was observed, though some changes were relatively subtle perhaps due to the short period of knockdown (Figure 3.9C). In parallel, the global demethylation caused by overexpression of MBD3 was also confirmed (Figure 3.9). Concerning the importance of MBD3 to the Mi-2/NuRD function and DNA maintenance methylation, the cell cycle progression is subjected to dysfunction when MBD3 is deregulated. Therefore, the cell-phase distribution was re-examined and it indicated a difficulty for MBD3-insufficient cells to pass the G1/S checkpoint (Figure 3.9D), resembling a G1-phase arrest upon depletion of Mi-2/NuRD or DNMTs.[125-127] Overall, this line of results supports the hypothesis that the demethylating potential of MBD3 plays a critical role in sustaining the DNA methylation homeostasis and cellular functions.

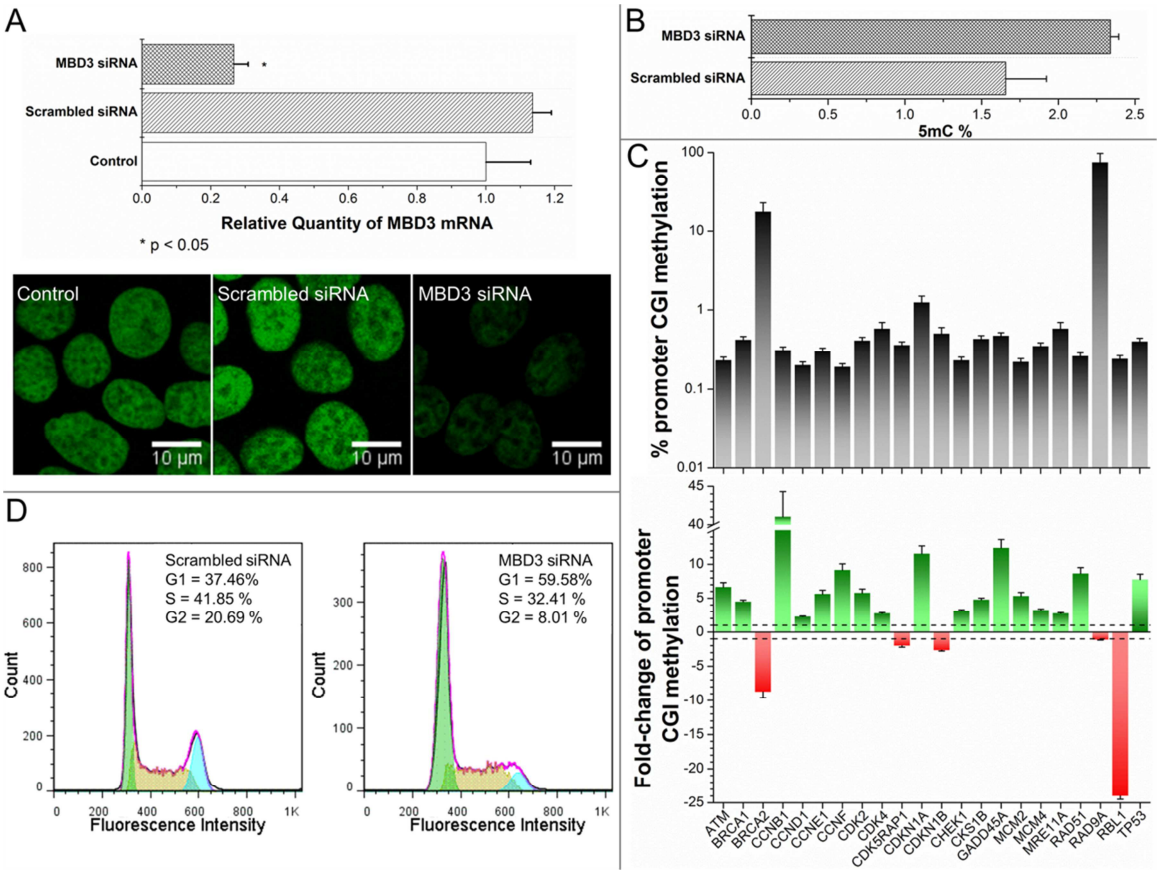


Figure 3.9 Insufficiency in MBD3 disrupts the DNA methylation homeostasis. (A) The MBD3 siRNAs were introduced to HeLa cells for knockdown. (B) 72 hours after knock-down, the global DNA methylation was determined with immunoassay (n = 3), (C) the promoter CGIs methylation of 22 cell cycle-related genes was quantified with PCR arrays (n = 3), and (D) the cell cycle distribution was examined with flow cytometry.

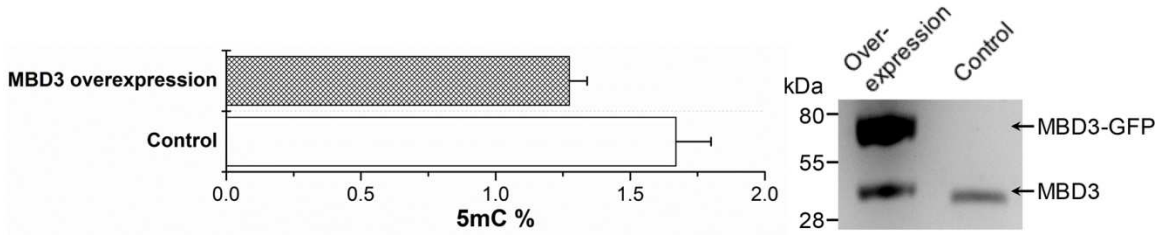


Figure 3.10 Overexpression of MBD3 causes global DNA hypomethylation.

3.4 Discussion

Achieving the DNA methylation homeostasis necessitates the interplay amidst multiple molecules and mechanisms. For a single nucleosome that covers about 200 bp DNA (including the linker), statistically 3-10 CpGs are included and the average extent of methylation marks should be inherited over generations. In order to retain such a steady methylation for a given region, a unified stochastic model was recently proposed[102]:

$$\frac{d\theta_m^i}{dt} = r_{met}^i \times (1-\theta_m^i) - \left(\frac{1}{2} D\left(1- f_{main}^i\right) + r_{demet}^i \right) \times \theta_m^i$$

Perceived from the model, the steady-state methylation level at a specific locus ($\theta_m^{i,ss}$) is synergistically determined by several variables – the rate of cell division (*D*), the efficiency of maintenance methylation (f_{main}^i), the rate of demethylation (r_{demet}^i), the rate of *de novo* methylation (r_{met}^i), and their homeostasis can be expressed as:

$$\theta_m^{i,ss} = \frac{r_{met}^i}{\frac{1}{2} D\left(1- f_{main}^i\right) + r_{demet}^i + r_{met}^i}$$

Derived from the above equation, the forces between demethylation and methylation need to meet the following balance:

$$r_{demet}^i = \frac{1-\theta_m^{i,ss}}{\theta_m^{i,ss}} \times r_{met}^i + \frac{1}{2} D \times f_{main}^i - \frac{1}{2} D$$

Upon incorporating the activity (r^i) of the DNA methylation-related enzymes, it can be roughly defined: $f_{main}^i \propto r_{DNMT1}^i \times r_{MBD2}^i$, $r_{met}^i \propto r_{DNMT3A}^i + r_{DNMT3B}^i$, $r_{demet}^i \propto r_{TETs}^i \times r_{MBD3}^i$. In the G1-phase, MBD3 tends to locate at the active genes and protect their hypomethylated status to avoid DNMT3 invasion. In the S-phase and G2-phase, DNMT1 – together with

DNMT3 – accomplish the DNA maintenance methylation, whereas no experiments have observed the real-time participation of TETs in this process. Yet without a protective or proofreading mechanism, the delicate DNA methylome would be fairly vulnerable to a non-equilibrium whenever DNMT1 overacts (*i.e.*, $f_{main}^{i} > 1$). In line with previous reports, this study reveals the dynamic behavior of MBD3 and its importance to the DNA methylation homeostasis: when cells enter cell cycle, a proportion of MBD3 and MBD2 would co-localize with DNMT1 to offer a regulatory "symbiosis" for the ongoing DNA maintenance methylation (Figure 3.11A). As newly discovered methylation derivatives, 5fC and 5fC emerge to have their independent regulatory functions,[128, 129] and one potential function for MBD3 in the DNA methylation homeostasis is to mark their existence on the parental DNA strand and pass down to the daughter strand (Figure 3.11B). Besides DNA methylation, the MBD2/3-associated Mi-2/NuRD can further facilitate nucleosome remodeling and histone modification in the newly synthesized chromatin. Also in light of the uncovered dynamic behavior of MBD3, it can be inferred that the occupancy sites by MBD3 are varied in different cell phases, which may explain the divergent observations about the genome-wide MBD3 distribution in previous researches.

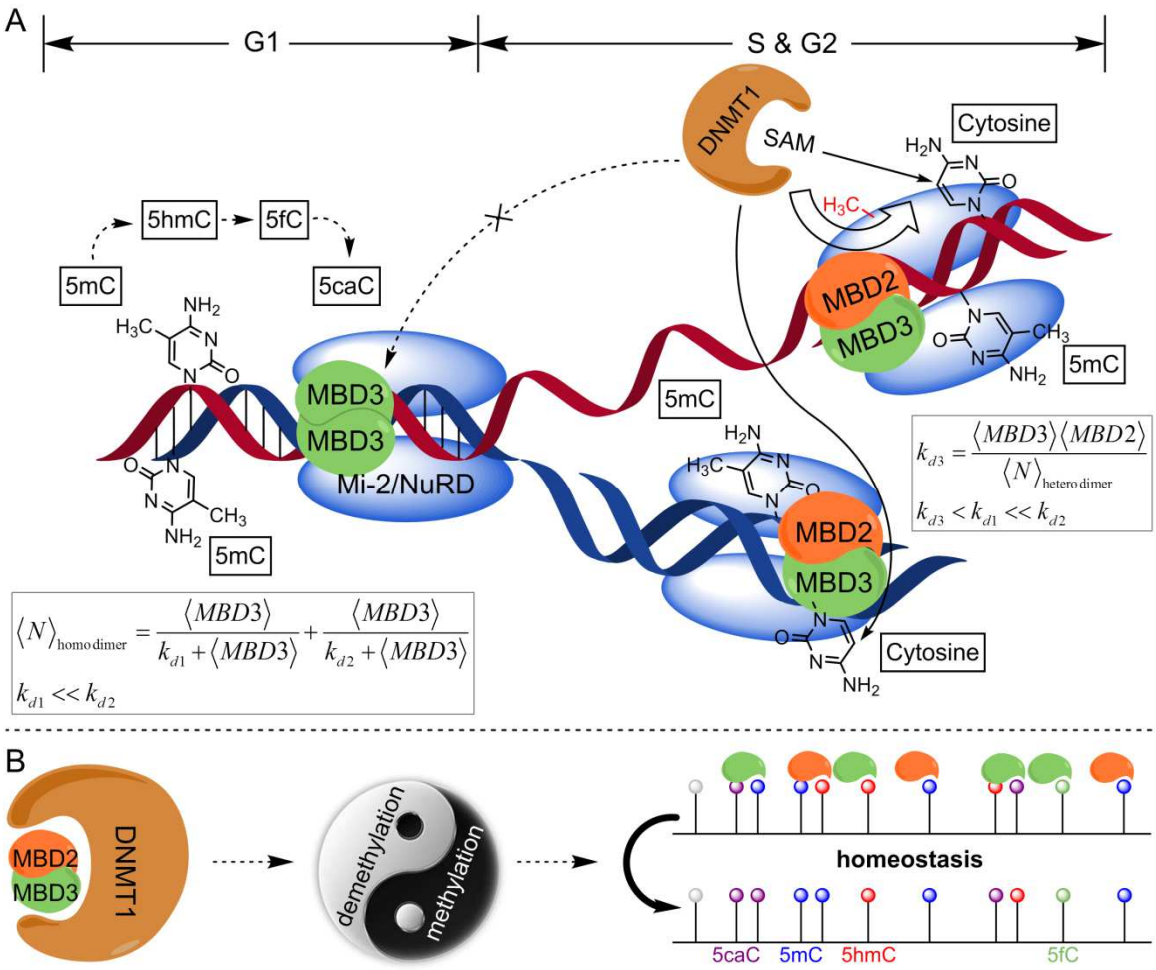


Figure 3.11 Dynamic behavior of MBD3 in the DNA methylation homeostasis. (A) In the G1-phase, MBD3-NuRD adopts a "two-site sequential binding mode" at its binding CGIs and maintains a hypomethylated state. k_{d1} and k_{d2} denote the two dissociation constants for the homodimer. As DNA starts to replicate, the newly synthesized hemi-methylated DNA intends to inherit the methylation density from the parental template, catalyzed by DNMTs (mainly DNMT1). A proportion of MBD3 and MBD2 would cooperate with DNMT1 to complement the maintenance accuracy. k_{d3} denotes the dissociation constant for the heterodimer. (B) When MBD2 and DNMT1 are re-establishing the landscape of 5mC, the demethylating potential of MBD3 can facilitate preserving those 5mC derivatives (*i.e.*, 5hmC, 5fC, and 5caC), providing a balancing force in the DNA methylation homeostasis.

The core region sequence for DNA recognition and binding is highly conserved among human MBD members. To date some already identified residues critical for the

MBD-DNA coupling include VAL20, ARG22, LYS30, TYR34, ARG44 and SER45,[32, 33] in which the arginine residues determine a hydrogen bond and cation-π interaction-based "stair motif" association with CpGs.[34] Although MBD3 is highly identical to MeCP2 and MBD2 in that core region, the two residual mutations (K30H and Y34F) at the binding interface undermine the interaction specificity between MBD3 and 5mC. Corresponding to a recent research showing that MBD3 preferably localizes to the active epigenetic loci, such as enhancers and promoters, herein the FLIM-FRET approach confirms an *in situ* interaction between MBD3 and the 5mC-oxidized derivatives (*i.e.*, 5hmC, 5fC and 5caC that are difficult to be detected with traditional DNA-sequencing methods), to support the participation of MBD3 in the DNA demethylation and gene activation.

The binding stoichiometry between MBD domain and its targeting site was elusive and even controversial as in different studies. Until very recently, a "two-site sequential binding mode" has been biochemically evaluated and it demonstrated that the MBD-mediated DNA binding would tolerate a homodimeric stoichiometry at a single CpG but with dissimilar dissociation constants: approximately 80 nM for the first MBD and 5 μM for the second one.[121] Moreover, a low salt concentration could substantially compromise the MBD to differentiate (methylated)CpG from random DNA sequences as monovalent cations are crucial to suppress the non-specific binding. This explains why the homodimeric state of MBD3 in the G1-phase got significantly depleted when a hypotonic medium was applied. On the contrary, MBD1 was found to exist predominantly in a monomeric state, which might be due to the distribution of MBD1 in the highly condensed heterochromatin where the low-affinity dimerization is physically hindered. One interesting finding requiring future elucidation is about the MBD3 stoichiometric transition from the

G1-phase to the G2-phase, if this transition is not only enforced by a mechanical separation along with the double-helix disentanglement in the S-phase. Theoretically, the cooperation of MBD2-MBD3-DNMT1 aids in achieving the homeostatic DNA maintenance methylation, while the detailed mechanism to form such a multi-component complex is challenging to resolve. MBD2 and DNMT1 have a structural base to co-localize because they could firmly bind with hemi-methylated DNA. For a further incorporation of MBD3, additional actions might be needed. One is through other epigenetic modifications such as histone acetylation. MBD2, in the context of Mi-2/NuRD, can induce a local histone hypoacetylation to subsequently attract MBD3-NuRD. Another way is through the coiled-coil interaction. The key helix domain to form coiled-coil oligomers is highly conversed among MBD proteins, though a stable MBD2-MBD3 heterodimer was hard to detect by gel filtration analysis.[130] However, considering the fact that the coiled-coil interaction is contingent on the protein conformation as well as the surrounding physicochemical factors,[131] it cannot eliminate the possibility of a cell cycle-dependent coiled-coil MBD3-MBD2 interaction. In addition, as a core subunit in the Mi-2/NuRD, p66α can stably oligomerize with either MBD2 or MBD3,[130, 132] an indirect route to link MBD3 and MBD2 could be involved – in the context of a multimer bridged by p66α or other coiled-coil domain proteins. Notably, the dissociation constant for coiled-coil mediated interactions can reach the level of ~10 nM (a much stronger affinity than the MBD-DNA binding), thus conducive to the disintegration of MBD3 homodimers. The primary concern for this assumption is how the coiled-coil interaction gets inhibited in the G1-phase. Last, in the process of recruiting biomolecules to achieve the maintenance methylation, the potential contribution from those classical DNA replication-related proteins, such as PCNA and

UHRF1, should not be overlooked. Further efforts can be placed in understanding the triggering mechanism for this stoichiometric transition and profiling the detailed interactome.

From phenotypic studies, MBD3 has been shown to play a deterministic role in embryogenesis,[44] pluripotency,[133] differentiation,[112] reprograming,[134] and a variety of diseases,[39] highlighting its functional importance. Disrupting the expression of MBD3 unequivocally results in abnormal DNA methylation, but an in-depth exploration is lacking. This study, combining single-molecule probing with biological experiments, opens a new window to research the underlying mechanistic connections. The close relation between MBD3 and DNA demethylation gives rise to a cell cycle-dependent protein behavior and serve as a pivotal balancing strength to merit the DNA methylation homeostasis. By highlighting the location of 5mC derivatives, this unique feature of MBD3 could not only improve the precision and efficiency of TETs but also constitute a protective mechanism to buffer the possible overactivity of DNMTs.

CHAPTER 4. REGULATORY LANDSCAPE AND CLINICAL IMPLICATION OF MBD3 IN HUMAN MALIGNANT GLIOMA

4.1 Introduction

The deadly form of intracranial malignant gliomas, *i.e.*, the World Health Organization (WHO) grade III-IV tumors, constitutes a heavily socioeconomic and healthcare burden. Today, even with significant advances in surgery, radiation therapy, and chemotherapy, yet the median survival for glioma patients, especially for those who are diagnosed with the most aggressive glioblastoma multiforme (GBM, grade IV) is extremely poor. A multitude of factors are involved in gliomagenesis, and so far enormous efforts have been invested in seeking clinical biomarkers for a better patient stratification and therapeutic management. As similar in other cancers, genetic mutations, epigenetic aberrations, environmental stress, signaling malfunction, and immunologic dysfunction synergistically contribute to the etiology of human malignant glioma.[135-137] For GBM patients receiving gross total surgical resection and effective adjuvant treatments, the 5-year survival rate is only 0.05% to 4.7% upon diagnosis,[138] which imperatively necessitates more in-depth and comprehensive research.

Although epigenetic aberrations have been uncovered as a driving force in a large number of human diseases, its mechanistic correlation with malignant glioma is less explored. As a key modulator, DNA methylation, *i.e.*, the addition of a methyl group to the 5-carbon of cytosine, epigenetically dictates the downstream transcription when accumu-

lating at the CpG island of a gene promoter.[4] In the oncogenic process, extensive alterations regarding the quantity and distribution of DNA methylation have been observed.[139] Hence, the DNA methylation homeostasis is essential to prevent a cancerous transformation and require a delicate cooperation among numerous molecules and factors, including MBD proteins.[140] In contrast to other typical MBD proteins (*e.g.*, MBD2 and MeCP2) that specifically associate with 5mC for gene repression, MBD3 evolves to gain a functional diversity. It has been found that MBD3 can either up-regulate or down-regulate its influential genes under particular circumstances.[107] As a core subunit of the Mi-2/NuRD complex, MBD3 also participates in the regulation of nucleosome remodeling and histone acetylation.[5] In Chapters 2 and 3, the significance of MBD3 in the active DNA methylation and DNA methylation homeostasis has been explored. Altogether, its versatile functions render MBD3 a pivotal part in the epigenetic network. Further, considering the importance of MBD3 to brain development and cell differentiation, it will be of translational value to dissect its implication in human malignant glioma.

In this chapter, the genome-wide transcription regulated by MBD3 and the possible correlation with the survival of glioma patients are examined. In human glioma cells, the epigenetic modulation by MBD3 is found to be bidirectional: both active and repressed genes are involved. By profiling the transcriptomic changes in GBM cells upon MBD3 knockdown (MBD3-KD), it is revealed that the differentially expressed genes (DEGs) include a myriad of membrane receptors and transcription-related factors. Moreover, several novel relations with regard to tumor malignancy and progression are discovered. For example, MBD3 is required to facilitate the expression of the major histocompatibility complex (MHC) class II molecules, which are proposed to be the key proteins in anti-

glioma immunity.[141-143] Furthermore, depletion of MBD3 enhances transcription from the oncogenic miR-17-92 cluster at chromosome 13, to promote the tumor progression. The *in vitro* and *in vivo* experiments cogently support an anti-proliferative potential of MBD3. Last, in an attempt to establish MBD3 as a prognostic biomarker, a clinical cohort of glioma patients has been followed up for 60 months. According to the results, a concurrent enrichment of MBD3 and 5hmC in tumor tissue indicates a better progression-free survival (PFS) and overall survival (OS), which confers promising information for future large-scale clinical studies.

4.2 Methods and Materials

4.2.1 Cell Culture and MBD3 Knockdown

For the *in vitro* experiments, human SF767 GBM cells were routinely cultured in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 10% fetal bovine serum, 1% antibiotics, and 1% glutamate. After revival from liquid nitrogen, the cells subcultured for 4 to 6 passages were used for knockdown experiments. Specific siRNAs targeting human MBD3 mRNA and scramble siRNAs were obtained from Santa Cruz Biotechnology (sc-35867). Transfection was conducted using the recommended siRNA Reagent System (sc-45064). The knockdown effects were determined by qRT-PCR and immunofluorescence staining. From the results, a 70-75% of knockdown efficiency can be maintained up to 6 days after transfection (Figure 4.2A), based on which microarray, proliferation assay, and migration assay were performed.

4.2.2 RNA Microarray and Data Analysis

The whole-genome difference in transcriptome between low-grade glioma and GBM, as well as the DEGs in GBM cells before and after MBD3 knockdown were investigated with the SurePrint G3 Human Gene Expression 8 × 60K Microarray Kit (Design ID: 028004, Agilent). This array chip covers 27,958 Entrez Gene RNAs and 7,419 long non-coding RNAs (lncRNAs). Four biological replicates or clinical samples were prepared for each group. With RNA extracted by TRIzol reagent (Life technologies, for biopsy tissues) or RNeasy Mini Kit (Qiagen, for cultured cells), cDNA and Cy3-labeled cRNA were sequentially synthesized with the Agilent LowInput QuickAmp Labeling Kit. The hybridization step lasted for 17 h at 65°C. After mild wash, the array slides were imaged with a SureScan scanner followed by feature extraction. The hybridization quality was assessed by monitoring the spike-in signals of ERCC and E1A control probes. Then the raw fluorescence readouts from the array were normalized to the intensity value of the 75th percentile of all non-control probes. For genes with multiple probes, the expression was numerically averaged. Approximate 2,000 DEGs based on fold-change and *p*-value (fold-change > 3 and *p*-value < 0.05 for clinical samples; fold-change > 1.5 and *p*-value < 0.1 for cultured cells) were determined with moderated *t*-test in the GeneSpring GX software. Protein classification was done with PANTHER (http://www.pantherdb.org).[144, 145] GO analysis was performed with GeneSpring GX. The DEGs-involved cellular functions and diseases were analyzed with the IPA online tool (http://www.ingenuity.com/products/ipa). The information of differentially expressed lncRNAs was manually annotated and cross-verified with referring to:

EMBL-EBI database (http://www.ebi.ac.uk)

lncRNA database (http://lncrnadb.com)

miRBase (http://www.mirbase.org)

Human lincRNA catalog (http://www.broadinstitute.org/genome_bio/human_lincrnas)

The complete microarray raw data have been deposited to NCBI GEO (GSE79878).

4.2.3 Methylation-Specific PCR (MSP) for the CIITA Promoter IV Region

The methylation status in the promoter IV region of the CIITA gene was determined by MSP. After DNA extraction and global methylation analysis (see detailed protocol in Chapter 3.2.6), bisulfite conversion was performed with EZ DNA Methylation-Lightning Kit (Zymo Research). The PCR primers and reaction condition followed the published protocol.[146] The amplicons from unmethylated primers and methylated primers were run through 3% agarose gel pre-stained with the GelRed dye (Biotium). The gel band intensity was quantified with the ImageJ software (http://imagej.nih.gov/ij).

4.2.4 Construction of Lentivirus and Cell Transduction

A MBD3 shRNA was designed for the *in vivo* tumor formation assay. The shRNA (5' – GCAAGATGCTGATGAGCAAGATG – 3') was constructed into the pLent-U6-GFP-Puro plasmid. Then the psPAX2 and pMD2.G packaging plasmids were co-transfected with the pLent-U6-GFP-Puro/pLent-MBD3 shRNAs in HEK293ft cells for 48 h according to manufacturer's instructions. Medium with virus was collected after 48 h of transfection and concentrated by centrifuge. Titer of viruses was tested and then preserved in -80°C. Human U87 GBM cells were infected with the shRNA construct for 48 h and selected with puromycin (0.75 mg/mL). 5×10^6 infected cells (GFP/MBD3 shRNAs) were

injected to the thalamus of mice (BALB/c-nu, body weight 18-22 g, 1-month old). One month after injection, the mice were sacrificed and the tumors were taken out for measurement.

4.2.5 Clinical Samples

A panel of primary glioma biopsy samples with characterized WHO grade and normal brain tissues from non-cancer pathologies were obtained from the Department of Neurosurgery at Xiangya Hospital of Central South University (Changsha, Hunan, China), with informed consent from all involved patients. By matching their clinical properties, MBD3 and 5hmC levels, 14 glioma patients were enrolled to the follow-up cohort, and their basic information was listed in Table 4.1. All patients received surgical resection to the extent feasible, followed by radiation therapy and chemotherapy which were given around 2-4 weeks after the first surgical removal. Radiation therapy was given with a dose of 1.8 Gy/d and a total dose of 50-56 Gy, including 30-36 Gy for brain and spine, plus 20 Gy for the original tumor site as the local boost. Chemotherapy mainly involved oral administration of temozolomide. One treatment cycle of chemotherapy was a consecutive administration of 150 mg/(m^2·d) for 5 days in every 4 weeks. 8 cycles were given in total. All surgical operations were performed in the Department of Neurosurgery at Xiangya Hospital; and the radiation therapy and chemotherapy were administered by the Department of Oncology at Xiangya Hospital. For clinical sample preparation, after a pathologist confirmed the tumor grade according to the 2007 WHO Classification of Tumors of the Central Nervous System, the tumor DNA and RNA were immediately extracted from the surgically resected tissues and the rest of the tumor was paraffin embed-

ded. A 60-month follow-up after surgery was conducted: PFS is defined as the living time before tumor recurrence; OS is defined as the time before clinical death or last follow-up. This study was approved by the Ethic Committee of Xiangya Hospital of Central South University.

Table 4.1 Information of patients enrolled for the clinical follow-up

Patient #	Gender	Age	Symptoms	Tumor Location
G II_1	Female	37	Seizure	L. frontal
G II_2	Male	36	Headache	R. temporal
G II_3	Female	46	Headache; Visual impairment	L. frontal & temporal
G II_4	Male	42	Seizure; L. limb weakness	R. frontal
G III_1	Male	44	R. limb weakness & numbness	L. frontal & parietal
G III_2	Male	48	Nausea & vomiting	L. temporal
G III_3	Male	62	Visual impairment	R. frontal
G III_4	Male	57	Poor memory; Headache	L. insula
GBM_1	Male	37	Seizure; R. limb weakness & numbness	L. frontal & parietal
GBM_2	Male	58	Headache; Poor memory	R. temporal & insula
GBM_3	Male	27	Headache; Dizziness; Seizure	L. temporal
GBM_4	Male	46	Headache; Visual impairment; Poor memory	L. temporal
GBM_5	Female	55	Headache	L. temporal
GBM_6	Female	26	R. limb weakness	R. frontal & parietal

4.2.6 Dot Blot for 5hmC in Clinical Samples

Genomic DNA from normal brain or tumor tissues was isolated by cell lysis buffer (100 mM Tris-HCl, pH = 8.5, 5 mM EDTA, 0.2% SDS, 200 mM NaCl, 0.667 μg/μL proteinase K) at 55°C overnight. In the second day, the samples were added with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 saturated with 10 mM Tris, pH =

8.0, 1 mM EDTA), mixed completely, and centrifuged for 5 min at 14,000 g. The aqueous layer solution was transferred into a new Eppendorf tube and precipitated with equal volume of 100% isopropanol. The genomic DNA was recovered and dissolved with 10 mM Tris-HCl (pH = 8.0). The DNA samples were further sonicated into 500 bp with a Misonix sonicator 3000 system. The concentration of sonicated DNA was determined by Nano-Drop 1000 (Thermo Scientific). The sonicated genomic DNA was denatured in TE buffer with 0.4 N NaOH at 100°C for 10 min, immediately chilled on ice for 10 min, and spotted onto a positive-charged nylon transfer membrane (GE Healthcare) with a Bio-Dot Apparatus (Bio-Rad). The membrane was baked for 30 min at 80°C and then blocked with PBS containing 5% dry milk and 0.1% Triton X-100 for 1 h at room temperature. Rabbit anti-5hmC antibody (1:10,000, Active Motif) was applied to the membrane and incubated at 4°C overnight. In the next day, the membrane was rinsed with fresh PBS and the colorimetric signal was developed after incubation with HRP-conjugated secondary antibody for 30 min. The final signal was visualized with an ECL-Plus system (Amersham Pharmacia Biotech).

4.2.7 Immunohistochemistry for Tissue Slides

5 µm-thick tissue slices were sectioned from surgically resected, formalin-fixed paraffin-embedded tumors and mounted onto polylysine coated slides. Before staining, tissue slides were sequentially deparaffinized and rehydrated using xylene solution and ethanol series. Heat-induced epitope retrieval was carried out in boiling Tris-EDTA buffer (pH = 9.0) for 20 min, followed by thorough wash with 0.025% Triton X-100 buffer. Blocking was conducted in TBS buffer containing 10% goat serum and 1% BSA. The

slides were then incubated with 1:200 diluted rabbit anti-MBD3 antibody (Pierce) or 1:500 diluted rabbit anti-5hmC antibody in a humid 4°C chamber overnight. After suppressing endogenous peroxidase with 0.3% H_2O_2, goat anti-rabbit IgG-HRP conjugates (Life technologies) were applied for 1 h at room temperature. Color signal was developed by using AEC staining Kit (Sigma) and then the slides were sealed in mounting medium prior to optical imaging.

4.3 Results

4.3.1 Loss of MBD3 and Relevant Transcriptomic Changes in Gliomagenesis

A negative correlation between glioma grade and MBD3 expression was noted from immunohistochemistry staining (Figure 4.1A-B). Considering that brain is the organ with the highest level of 5hmC and it has been postulated as one of the major MBD3-binding targets, the amount of 5hmC in the same batch of tumor tissues was thus examined. Evidently, 5hmC displayed a similar reduction pattern in high-grade gliomas (Figure 4.1C). Moreover, in high-grade tumors, the intra-group variation of the MBD3 and 5hmC quantities becomes more significant.

After uncovering the serial loss of MBD3 in the aggravation of glioma grade, the altered gene expression related to MBD3 was profiled by RNA microarray. MBD3 siRNAs were introduced into SF767 cells and the knockdown efficiency was confirmed by qRT-PCR, western blot, and immunofluorescence (Figure 4.2A-B). Next, the genome-wide changes in transcriptome were probed by the Agilent microarray (Figure 4.2C). In parallel, the DEGs between low-grade glioma and GBM tissues were screened for comparison.

As shown in Figure 4.3 and Figure 4.4, a considerable overlap in the impacted protein functions between these two sets of DEGs is observed, which implicates that during the progression from a low-grade to a high-grade glioma MBD3 is involved in a substantial number of gliomagenetic pathways and functions. For the down-regulated DEGs by MBD3-KD, the top-matched gene ontology (GO) is the "MHC II protein complex", which is a surface protein complex presenting antigens to the CD4+ T cells. In analysis of the up-regulated DEGs, the major differences between low-grade gliomas and GBM are "nucleic acid binding proteins" and "transcription factors", while depleting MBD3 further disturbs the expression of many "receptor proteins" (*e.g.,* G-protein coupled receptors) in GBM cells. Other up-regulated DEGs by MBD3-KD include transport proteins (*e.g.*, ion channel proteins and solute carrier family), cytochrome P450 family, growth factors (*e.g.,* FGF and IGF), and dynein proteins, all of which are hallmarks of active cell metabolism and proliferation. From another aspect, the MBD3-KD induced DEG-related cell functions and diseases, as revealed by the Ingenuity Pathway Analysis (IPA), mainly point to "gliomagenesis", "inflammatory responses", "metabolism", "neural signaling" and "neurological disorder" in the central nervous system (CNS) (Figure 4.5). Out of the 7,419 probed lncRNAs, about 600 are determined to be significantly altered upon MBD3-KD, though the majority of their functions require future exploration due to the lack of a well-defined annotation database. One interesting finding is that the lncRNAs regulated by MBD3 is not fully consistent with the natural distribution on chromosomes. Instead, a skewed distribution is observed. Chromosomes 2, 5, 6, 8, 13, and X encompass a higher percentage of the MBD3-influenced lncRNAs than the expected number. Given the importance of lncRNAs to the chromatin integrity, transcription rate, and post-translational

modification, this phenomenon preliminarily suggests that a dysfunction in MBD3 may compromise more functions relating to those specific chromosomes.

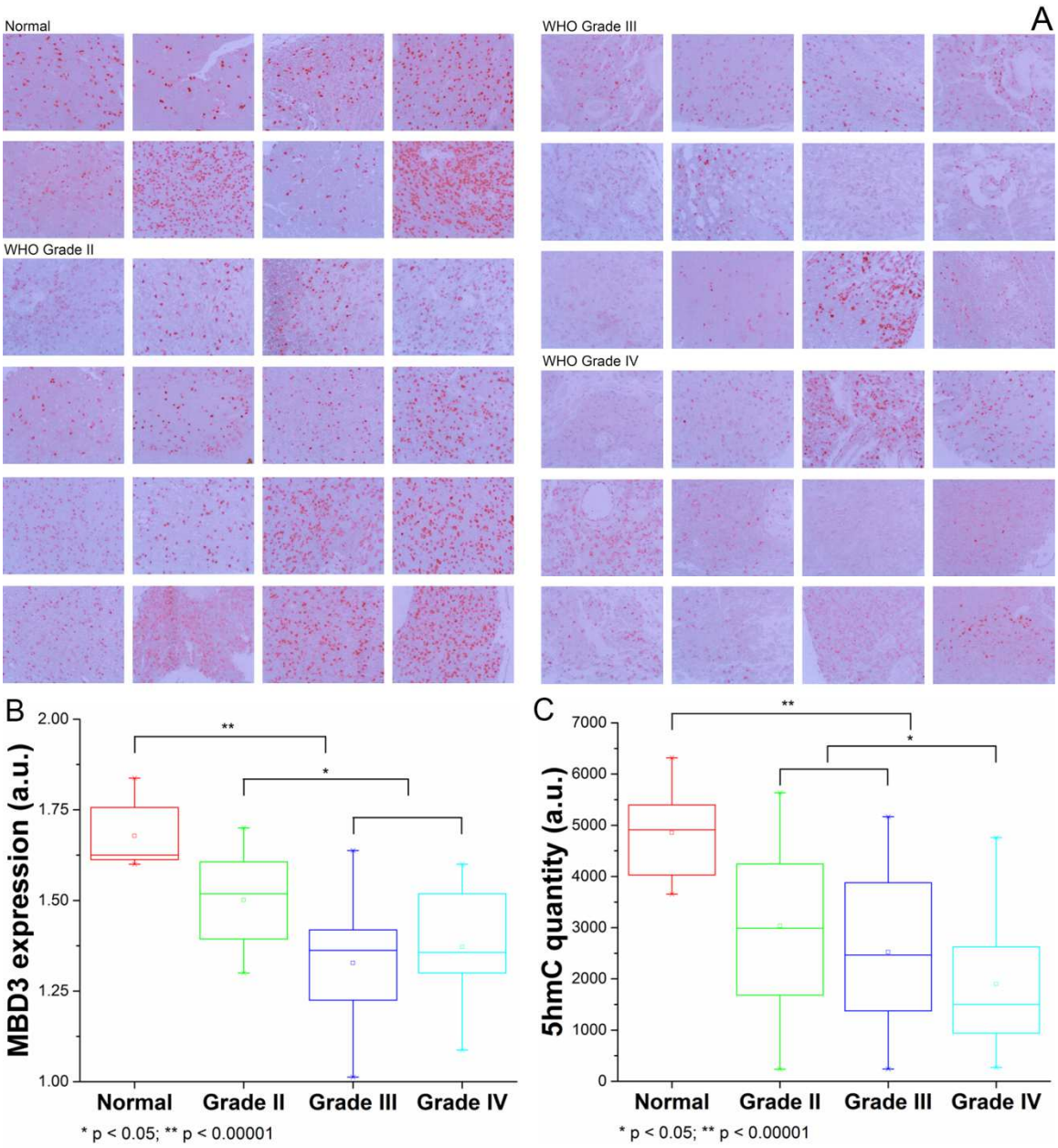


Figure 4.1 Expression pattern of MBD3 in normal brain tissue and malignant glioma. (A) Representative IHC images. Images were taken from an Olympus BX53 microscope fitted with a bright illumination and a 20× objective lens. (B) Semi-quantitative results from the IHC images using the ImageJ software. (C) Quantity of 5hmC from the same groups of patients by dot blot. ANOVA was performed for statistical analysis.

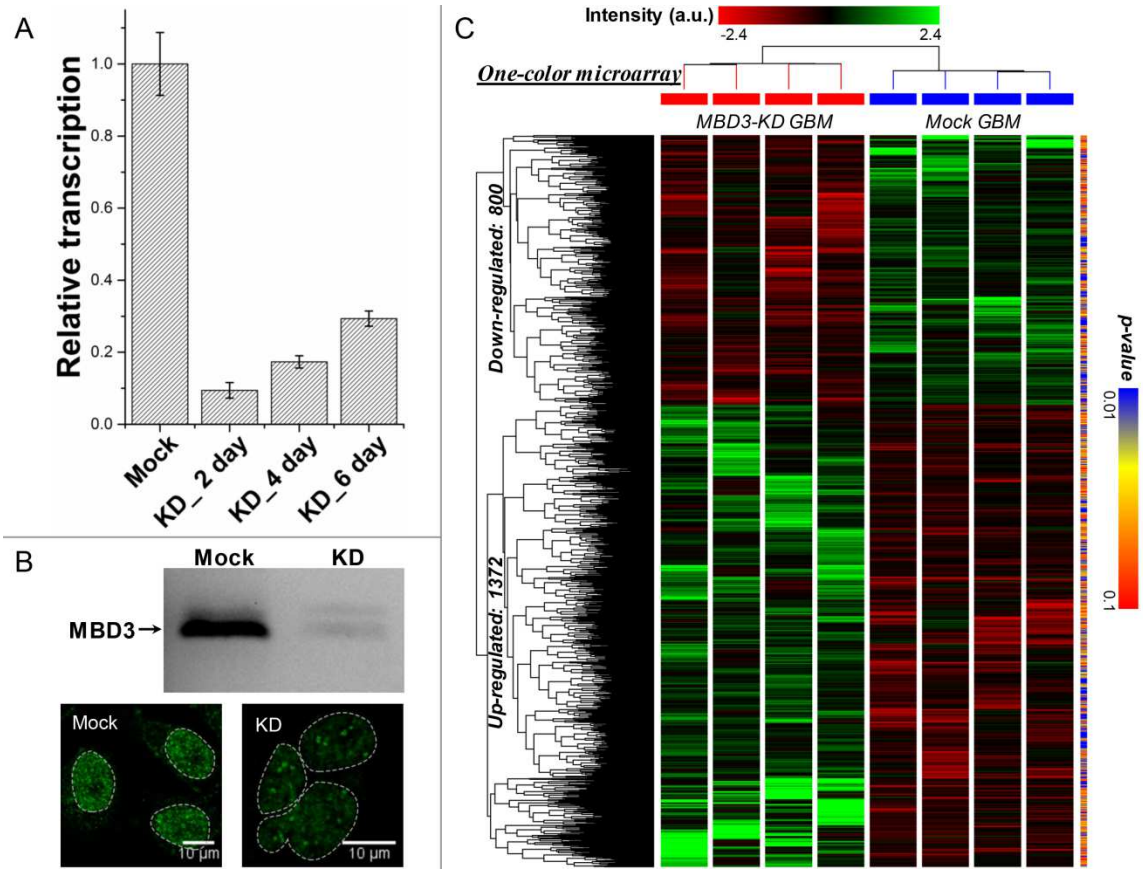
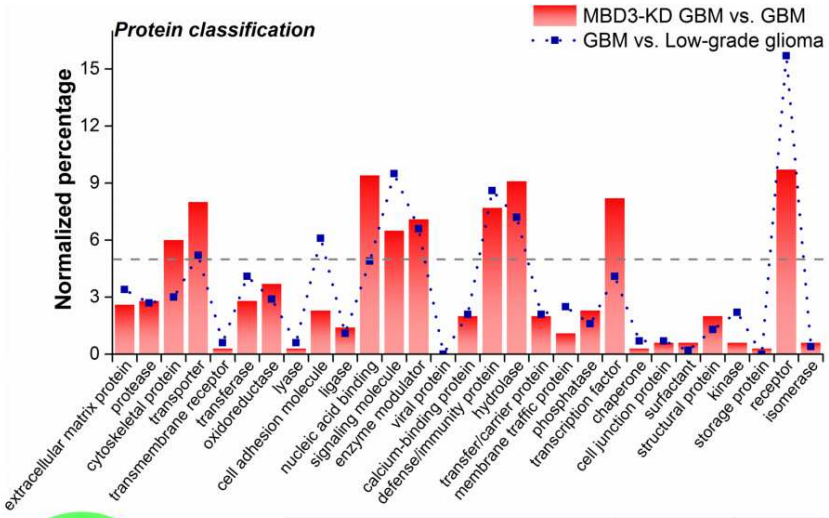
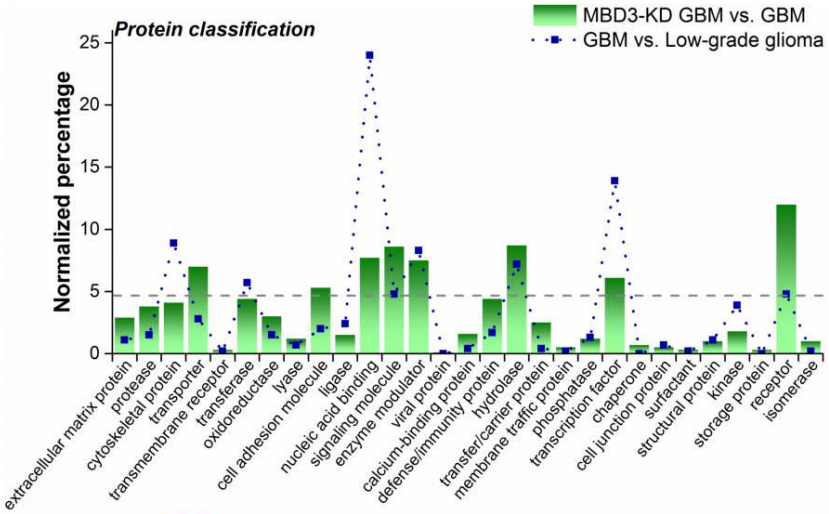


Figure 4.2 Genome-wide transcriptome regulated by MBD3 in GBM cells. The knock-down efficiency by the MBD3-siRNA was validated by (A) qRT-PCR (up to 6 days), (B) western blot and immunofluorescence staining (at day 4). (C) Upon the knockdown of MBD3, the changes in transcriptome were quantified with Agilent one-color microarray.



GO Analysis: MBD3-KD vs. Control	Gene counts	p-value
MHC II protein complex	11	9.77E-17
Vacuolar membrane	17	2.55E-09
Cellular response to cytokine stimulus	21	1.61E-07
Lysosomal membrane	13	2.97E-07
Membrane part	147	1.13E-05

Figure 4.3 Protein classification and GO analysis for the down-regulated genes by MBD3-KD. Differences between GBM and low-grade glioma are presented as the reference (dash line). The number of proteins appearing in both the DEGs is presented in the pie chart. The top 5 matched GO terms mediated by MBD3 is presented in table.



GO Analysis: MBD3-KD vs. Control	Gene counts	p-value
Receptor activity	101	2.00E-09
Transmembrane signaling receptor activity	73	9.42E-08
G protein-coupled receptor signaling pathway	64	6.03E-07
Membrane part	252	2.14E-06
Axoneme	10	3.24E-05

Figure 4.4 Analysis for the up-regulated genes by MBD3-KD.

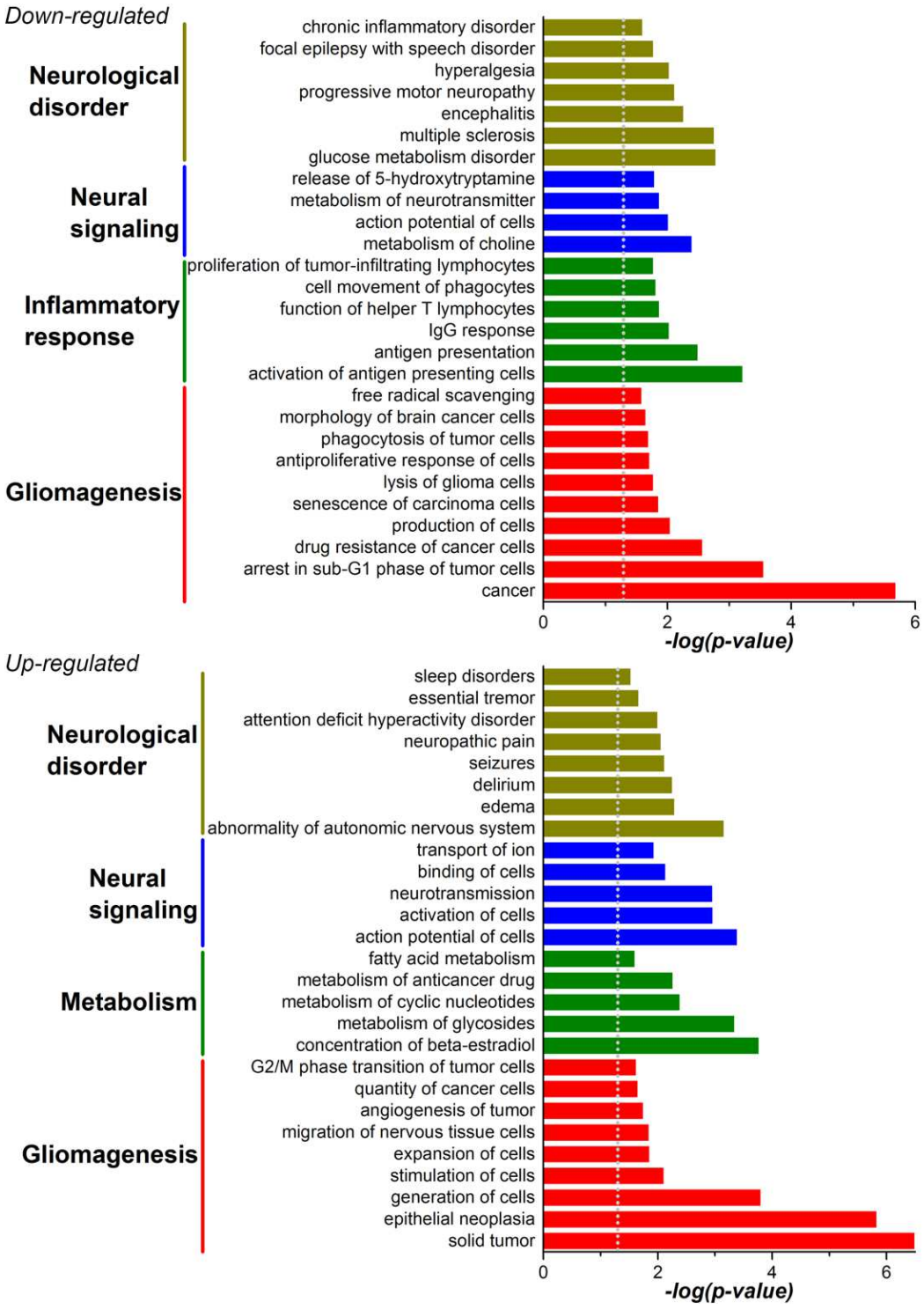


Figure 4.5 Cellular functions and diseases related to the DEGs in MBD3-KD GBM cells were profiled with the IPA software.

4.3.2 Regulatory Function of MBD3 on MHC Class II Molecules

Since the down-regulation of MHC II protein complex was ranked top in GO analysis, the raw array data for each MHC II subunit were extracted (Figure 4.6A). MHC class II protein complex is a heterodimer comprising α chain and β chain that are transcribed from the HLA genes on chromosome 6. Five major HLA-encoding regions exist there: HLA-DM, HLA-DO, HLA-DP, HLA-DQ, and HLA-DR. It has been demonstrated that MHC class II molecules could be constitutively expressed in glioma cells and their levels negatively associate with the tumor invasion and malignancy.[142, 147, 148] As shown in Figure 4.6B, in comparison with low-grade glioma, GBM has a significantly lower transcription of almost every HLA gene. Upon the MBD3 knockdown, ten HLA variants experienced a further decrease. It was speculated that a possible mechanism accounting for this MBD3-induced massive down-regulation was through the MHC class II transactivator (CIITA) gene since it exhibited the same reduction trend from the array results. The transcription of CIITA is subjected to epigenetic inactivation, mainly by DNA methylation in its promoter IV region.[146, 147, 149] In Chapter 3, MBD3 (by extension the Mi-2/NuRD complex) has been demonstrated to maintain a hypomethylated promoter for those genes poised to a dynamic transition in expression. Herein, a global DNA hypermethylation was noted in MBD3-KD GBM cells, again supporting its demethylation potential (Figure 4.7A). Using MSP, the methylation level of the CIITA promoter IV region was found to rapidly increase by about 20% after 72 hours of MBD3 knockdown (Figure 4.7B), which supports an epigenetic control on the CIITA gene by MBD3. Other than DNA methylation, the extensively dysfunctional lncRNAs on chromosome 6 after depletion of MBD3, as aforementioned, might influence the expression of MHC class II molecules as well.

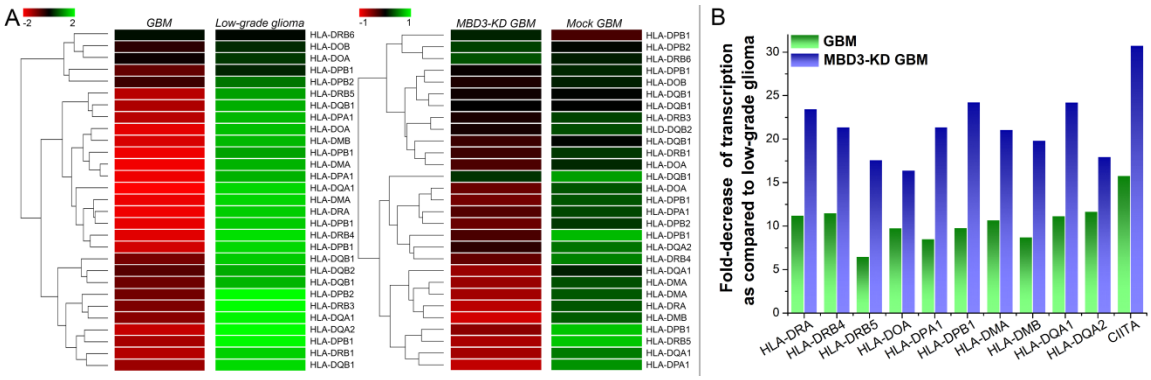


Figure 4.6 MBD3 facilitates the basal expression of MHC class II molecules in glioma cells. The raw signal readings for all the MHC II subunits are presented (left panel). For genes having multiple probes in the array, their transcriptional levels were averaged. The MBD3-regulated HLA genes in GBM and MBD3-KD GBM cells were compared to that in low-grade glioma (right panel).

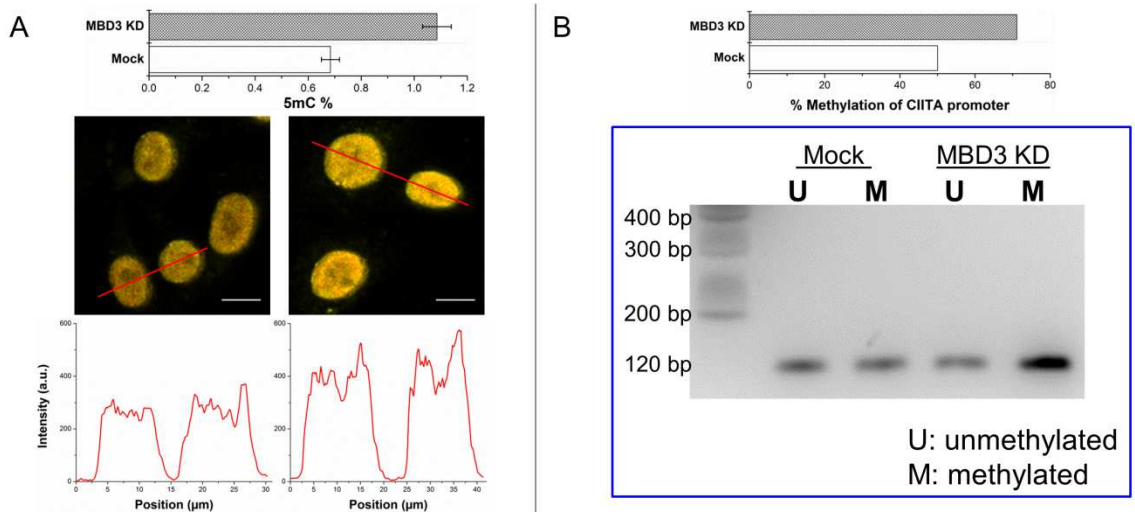


Figure 4.7 MBD3 regulates the expression of the CIITA gene through DNA methylation. (A) Global hypomethylation induced by MBD3-KD was assessed by immunoassay (upper panel, n = 3) and immunofluorescence (lower panel). (B) Increase in the methylation of the CIITA promoter IV region was validated by MSP.

4.3.3 Influence of MBD3 on the Oncogenic miR-17-92 Cluster

The miR-17-92 cluster is among the first characterized miRNAs underpinning a variety of molecular events to promote oncogenesis. From the polycistronic MIR17HG lo-

cus, a precursor transcript (pri-miRNA) is transcribed and can yield six mature miRNAs: miR-17 (further derive the mature miR-17-5p and passenger miR-17-3p), miR-18a, miR-19a, miR-20a, miR-19b-1, and miR-92a-1.[150] These miRNAs separately or collaboratively regulate a broad spectrum of genes including p21, Rb1, Bim, and Pten, thus controlling the cell fate[151, 152] (Figure 4.8A, upper panel). Mainly activated by Myc and E2F proteins, the increased expression of miR-17-92 has been reported in a number of malignancies.[153, 154] From the microarray results, the number of MIR17HG transcripts in GBM is significantly higher than that in low-grade glioma (Figure 4.8A, lower panel). A recent research has revealed that MBD3-NuRD was able to directly compete with c-Myc to repress its transactivation potential.[155] In line with this, the transcriptional level of MIR17HG was expected to be further enhanced by MBD3-KD. Afterwards, the quantity of each mature miRNA from MIR17HG was determined by qRT-PCR, which proved that the processing for the pri-miRNA gave rise to a heterogeneous distribution of the final products (Figure 4.8B, left panel). Specifically, miR-17-3p, miR-18a, miR-19a, and miR-20a experienced a substantial increase upon MBD3 knockdown (Figure 4.8B, right panel). As the miR-17-92 products are closely related to cell proliferation and migration, *in vitro* and *in vivo* assays were performed to establish the connection between MBD3 and the glioma malignancy (Figure 4.9). Collectively, the obtained results support that a loss in MBD3 would release the restraint on Myc and subsequently up-regulate the transcription from the miR-17-92 cluster, leading to a more aggressive tumor growth.

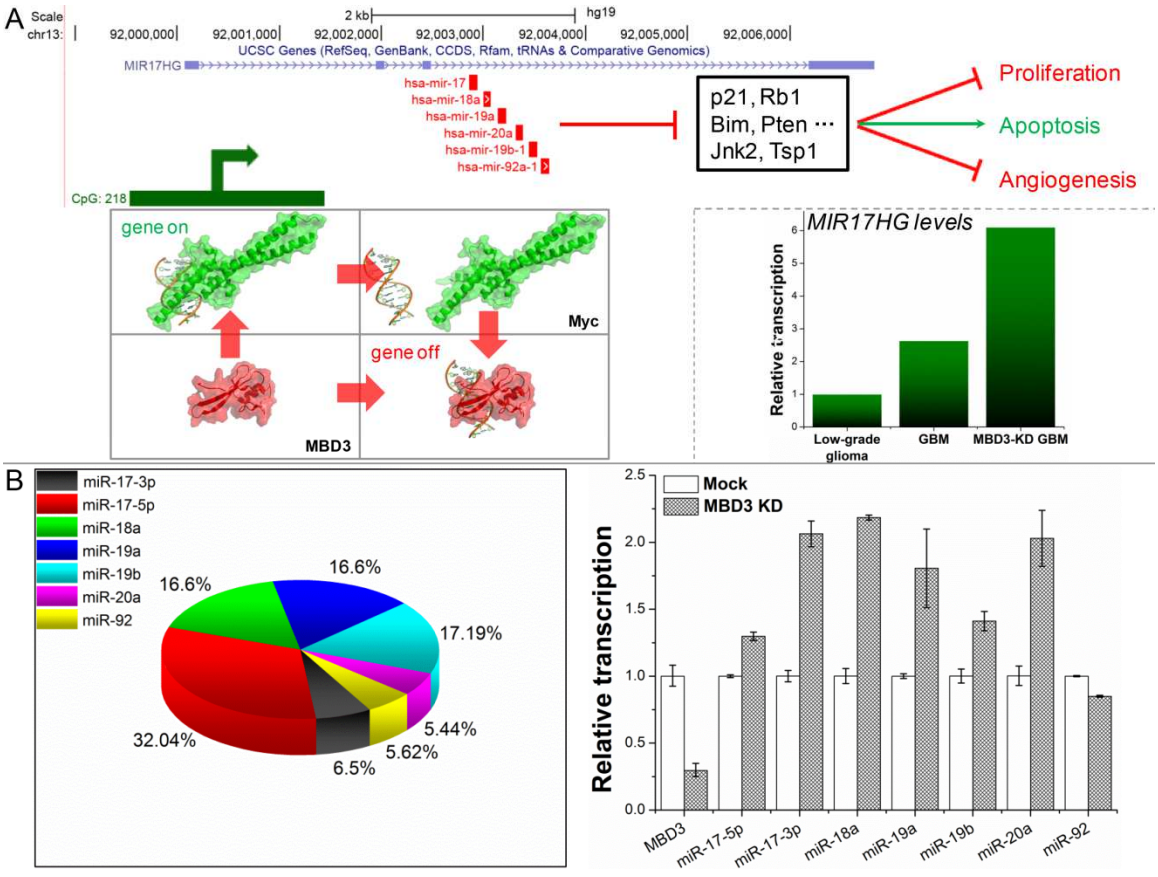


Figure 4.8 MBD3 fine-tunes the Myc-mediated transcription of MIR17HG. (A) The transcriptional regulation and downstream key targets of MIR17HG are illustrated (upper level). The pri-miRNA levels in low-grade glioma, GBM, and MBD3-KD GBM were probed by microarray. (B) Composition of the mature miRNAs from MIR17HG in GBM and their level changes upon MBD3-KD were quantified by qRT-PCR (mean ± standard deviation, n = 3).

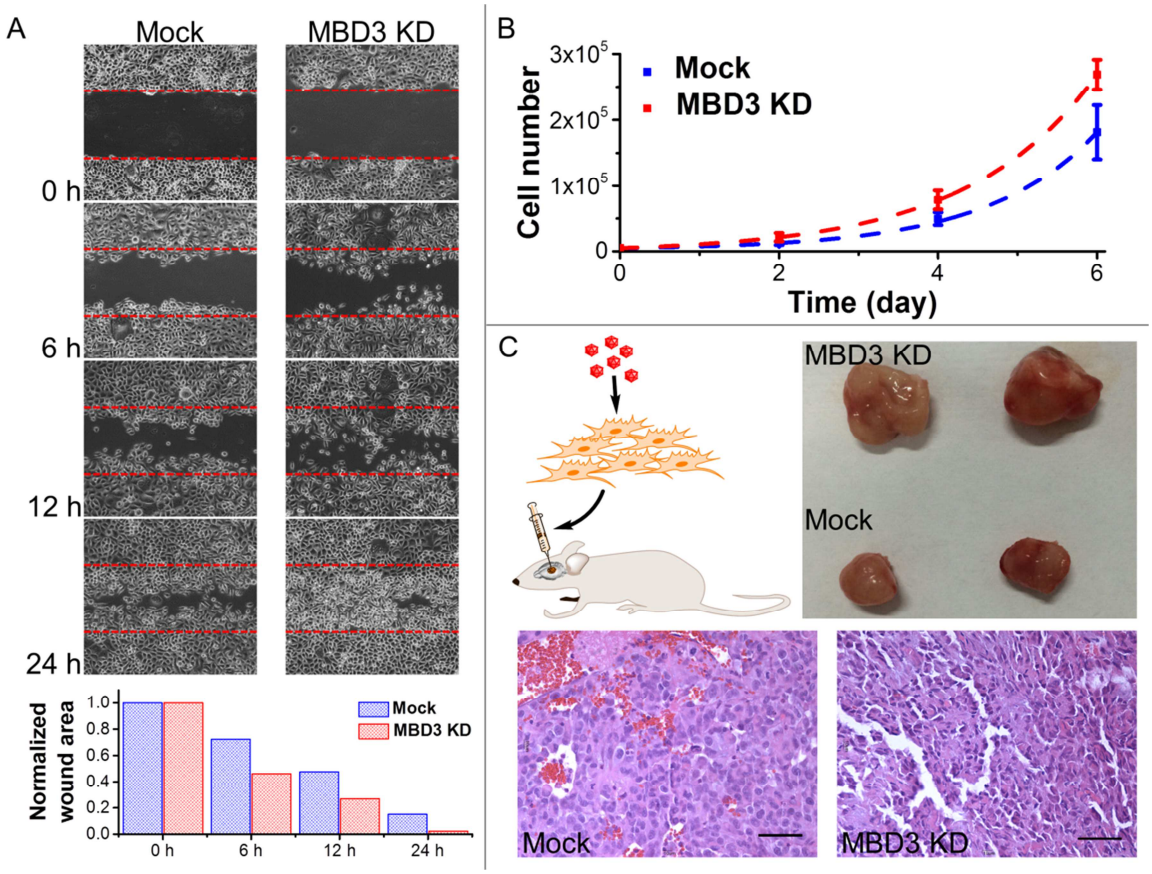


Figure 4.9 Influence of MBD3 on glioma migration and proliferation. (A) Migration assay was performed using GBM cells with and without MBD3-KD. (B) Proliferation rates of GBM cells with and without MBD3 siRNA (n = 4, mean ± standard deviation). (C) *In vivo* tumor formation assay was performed by injecting GBM cells into mouse brain. By H&E staining, the MBD3-KD GBM cells show more pleomorphic and basophilic nuclei (scale bars: 50 µm).

4.3.4 Clinical Implication of MBD3 Expression in Glioma Prognosis

Following the experiments at the molecular and cellular levels, it inspired further efforts to explore the possibility of using MBD3 as a potential prognostic biomarker due to its control on the glioma migration and proliferation. To test this hypothesis, a pilot clinical cohort was designed and the enrolled glioma patients were strictly matched with their demographic attributes, pathological characteristics, surgical outcomes, and adjuvant

therapies. 5hmC has been discovered as the major DNA modification in association with MBD3 and early efforts aimed to exploit the clinical implication of 5hmC in glioma have been attempted.[156-158] The reduction of 5hmC in high-grade glioma was indeed noted in this study. It is extensively perceived that epigenetic modifications and their associated proteins function in a cooperative manner. Since the DNA modification is not convenient to be detected, it is rational to quantitate the relevant binding proteins as an alternative in the clinics. In this regard, a subgroup of 14 patients bearing correlative levels of MBD3 and 5hmC were recruited. After a 60-month follow-up, MBD3 – as well as 5hmC – exhibited a positive correlation with the patients' survival. The PFS of GBM patients who received gross total resection in surgery and adjuvant therapies (*i.e.,* radiation therapy and temozolomide) was considerably proportional to the intra-tumor contents of MBD3 and 5hmC (Figure 4.10). From the MRI imaging, tumors with a relatively high abundance of MBD3 and 5hmC showed a more confined growth whereas a low abundance of MBD3 and 5hmC was connected to a more aggressive, pre-multifocal growth pattern (Figure 4.11). The prognostic connection was also perceivable in the enrolled grade II-III patients (Figure 4.12). The NCI Rembrandt data further support the positive correlation between the expression of MBD3 and the glioma survival, though with a marginal significance of difference (p-value = 0.03, Figure 4.13). Altogether, these clinical data establish the potential of MBD3 as a novel epigenetic marker to complement 5hmC in improvement on the clinical management for malignant glioma.

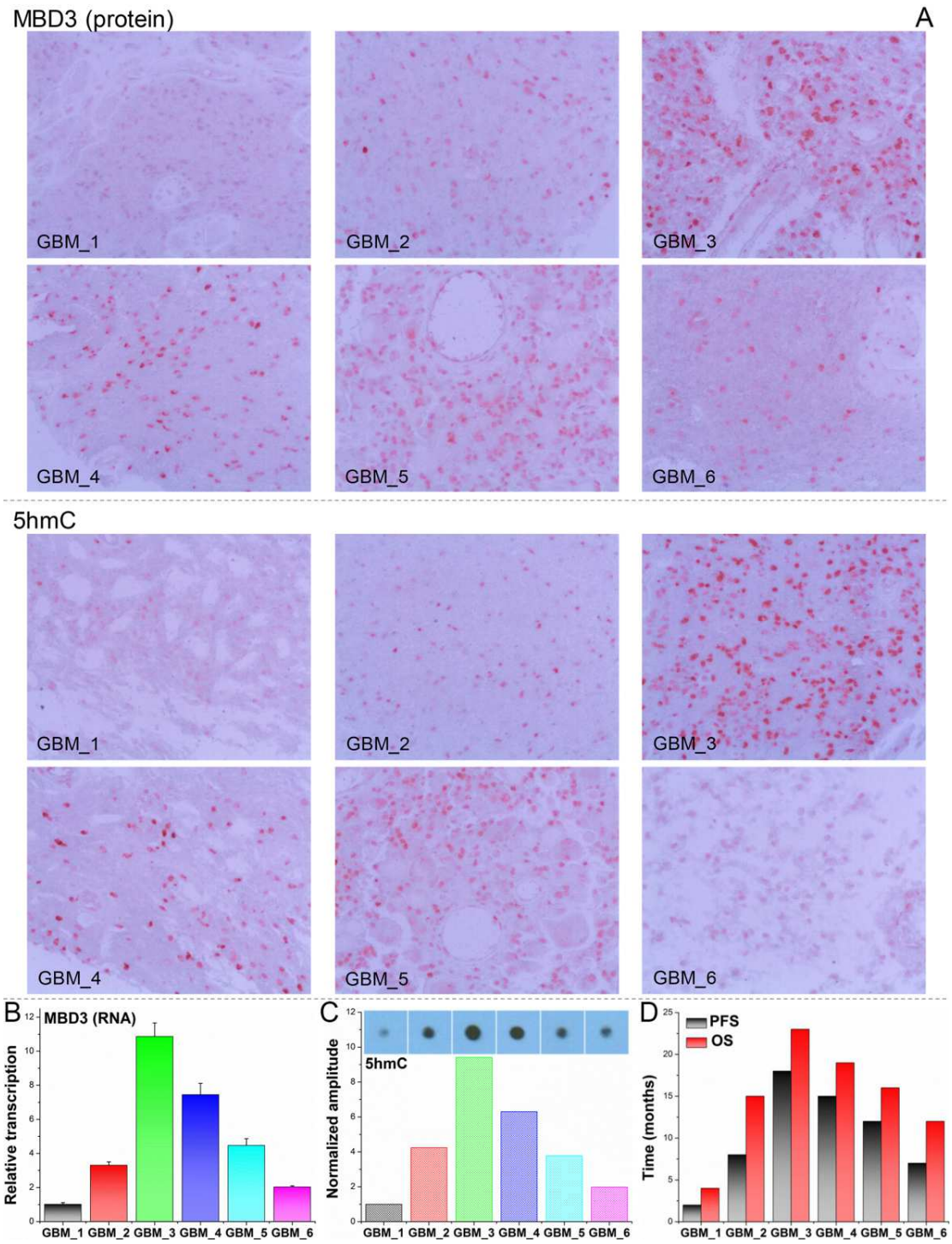


Figure 4.10 Dual-quantification of MBD3 and 5hmC exerts a promising prognostic power for GBM patients. 6 GBM patients with a correlative abundance of MBD3 and 5hmC were followed up after surgery. (A) IHC staining for MBD3 protein and 5hmC. (B) qRT-PCR for MBD3 mRNA. (C) Dot blot for 5hmC. (D) Survival outcomes of the patients. PFS: progression-free survival; OS: overall survival.

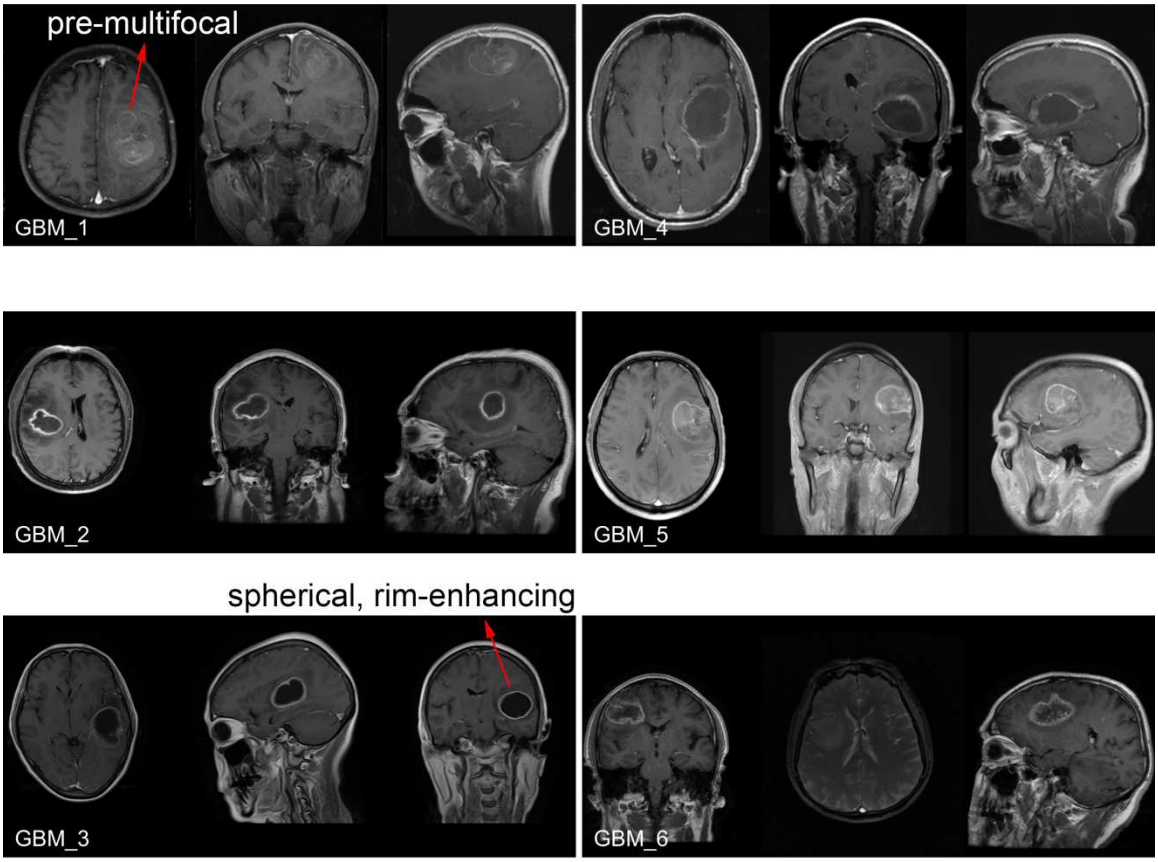


Figure 4.11 Brain MRI images of the GBM patients enrolled in the clinical cohort. Before surgery, the tumor growth was examined by the T1 post-contrast MRI imaging. The pre-multifocal growth of MBD3-low GBM features a rapid tumor expansion, aggressive infiltration to surrounding normal tissue, and poor survival. In comparison, the spherical, rim-enhancing MBD3-high GBM features a relatively mild growth, confined edge, and better survival.

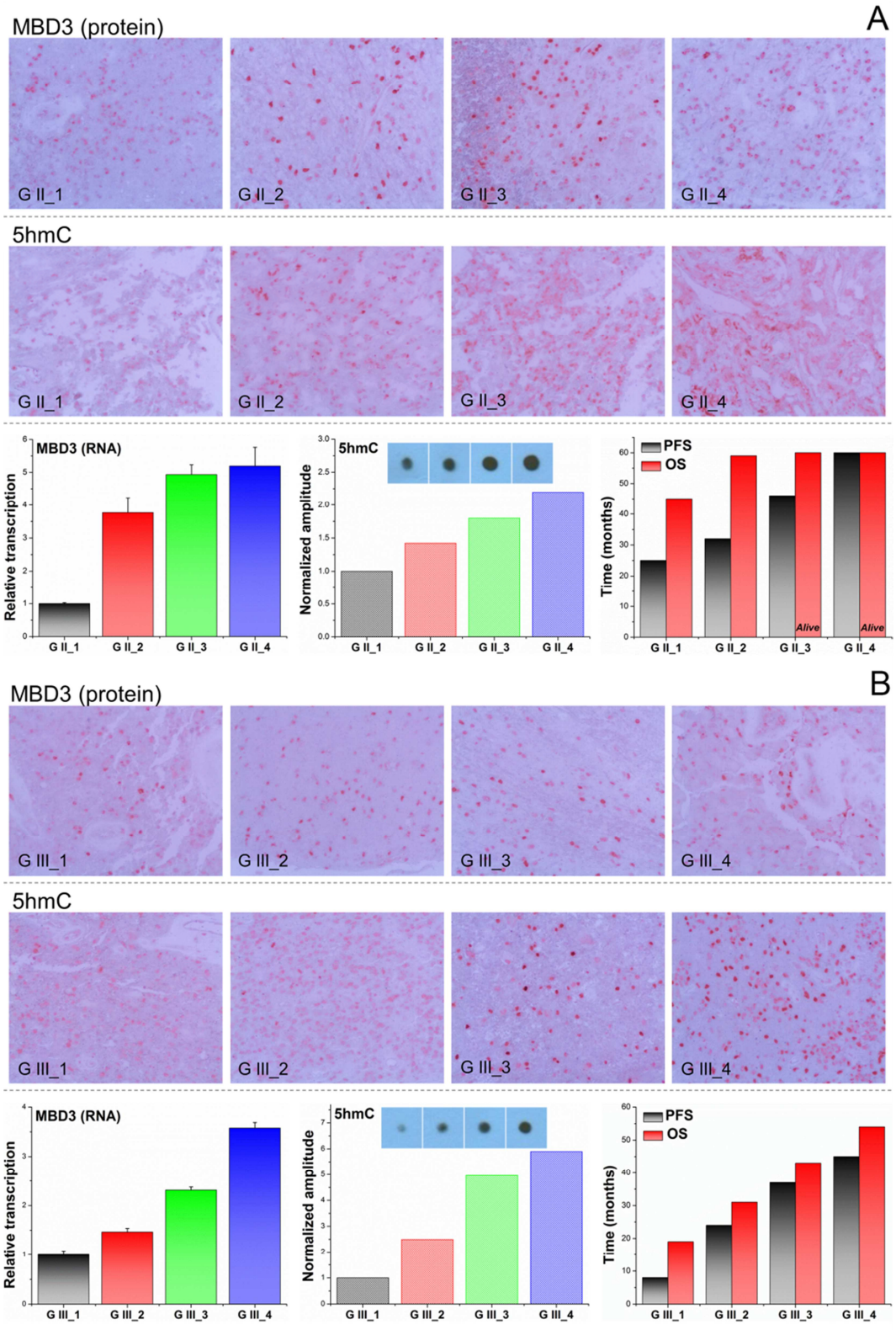


Figure 4.12 Survival outcomes of the enrolled grade II-III patients.

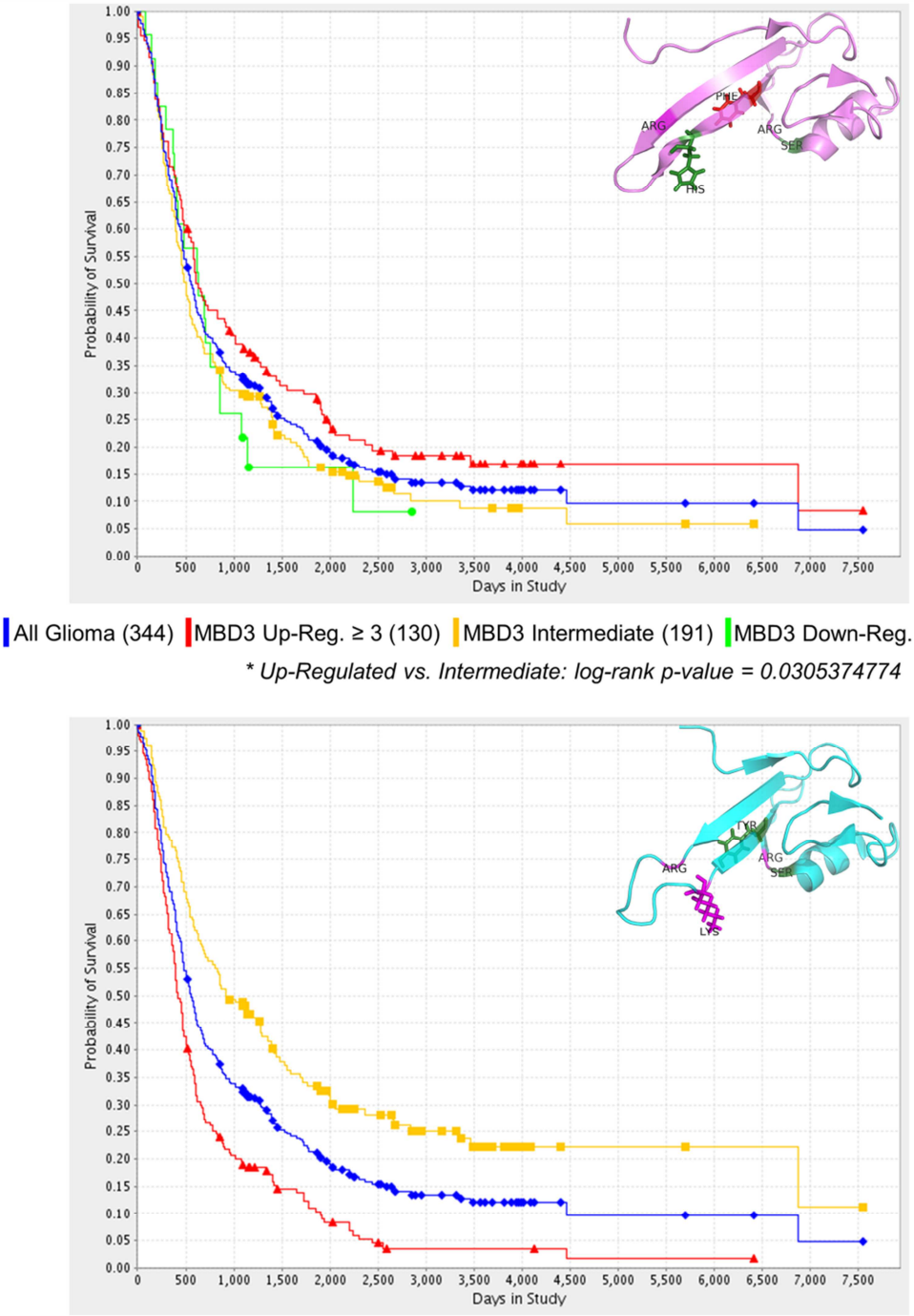


Figure 4.13 Distinct clinical implications of MBD3 and MBD2 in the survival of glioma. (retrieved from the NCI Rembrandt database https://caintegrator.nci.nih.gov/rembrandt)

4.4 Discussion

Although malignant glioma is a prevalent form of primary brain tumor with devastating outcomes, research on the involvement of epigenetic mechanisms in gliomagenesis is relatively sparse. Like other solid tumors, the progression of glioma is fueled by a plethora of epigenetic aberrations, while it is yet uncertain whether the reader proteins for DNA methylation play a role in this pathology even if some preliminary efforts were attempted.[46, 47] In contrast to other MBD proteins (*e.g.,* MBD1, MBD2, and MeCP2), MBD3 in mammalian cells exhibits an elusive and context-dependent interaction with DNA methylation.[159] For example, MBD3 has long been thought to lose its 5mC-binding specificity; however, the *in vivo* co-localization of MBD3 with 5mC and its oxidized derivatives have been unambiguously detected in a series of studies.[5, 105-107, 117] Furthermore, overexpression of MBD3 could induce a global DNA demethylation,[42] again implying the exquisite participation of MBD3 in the DNA methylation homeostasis. Other than being a DNA-binding protein, MBD3 also serves as a constitutional subunit of the Mi-2/NuRD complex that catalyzes histone deacetylation and nucleosome remodeling.[160] MBD3-NuRD has proven to be a key factor in inhibition of tumorigenesis and naïve pluripotency in the recent reports.[155, 161] In spite of some discrepant conclusions, the functional importance of MBD3 to the CNS is evident, from pluripotency to development, from synaptic connectivity to lineage commitment.[134, 162-164] Herein this study further uncovered a MBD3-mediated anti-glioma network. Consistent with the previous research, the microarray results validate that MBD3 exerts an influence on both active and silenced genes in glioma. Bioinformatics analysis and the following *in vitro/in vivo* tests support that a higher level of MBD3 might herald a better survival outcome.

The MBD3-mediated regulation on MHC class II molecules holds a medical significance because it provides us with new avenues to understand the immune escape of gliomagenesis and to compose targeted immunotherapies. In traditional neuroscience, human brain was thought to be an organ with immunological privilege due to the blockade by blood-brain barrier (BBB) and the absence of intracranial lymphatic structures,[165] which can facilitate brain tumors to evade immune surveillance. However, the integrity of BBB in glioma is often sabotaged along with the tumor invasion and deficient vasculature, and the expression of immunological molecules is found in a variety of glial cells, including microglial cells, astrocytes, and even glioma cells. Those discoveries strongly imply an anatomic connection between the CNS and immune system. The recent groundbreaking findings of the glioma infiltrating lymphocytes and meningeal lymphatic vessels consolidate the participation of immune responses in restricting glioma malignancy.[166, 167] Enhancing the recognition of tumor antigens has therefore been proposed to be a promising therapy, and a peptide-based glioma vaccine has been successfully developed to activate the MHC class II molecules-dependent CD4+ T cells.[143] However, an extensive down-regulation of MHC class II expression is a pathological characteristic for invading glioma cells,[142] and thus restoring the expression of MHC class II molecules constitutes the first task to the effectiveness of immunotherapy. In the early attempts of anti-glioma immunotherapy, *in situ* delivery of interferon-gamma (IFN-γ) with viral vectors was applied to reinforce the expression of MHC class II molecules in glioma cells and enhance the anti-tumor activity of T cells.[168, 169] From this work, the expression of MHC class II molecules was confirmed to be utterly low in GBM when compared with that in normal brain tissue and low-grade gliomas. Intriguingly, through epigenetically regulating the CIITA gene

and relevant lncRNAs on chromosome 6, MBD3 could reactivate the expression of MHC class II molecules and may serve as a new therapeutic target to benefit patients from the booming immunotherapies. It is anticipated that tumors bearing a high MBD3 abundance would respond better to therapeutic vaccination and immune-checkpoint-blockade agents (*e.g.*, CTLA-4 and PD-1/PD-1L inhibitors). Furthermore, engineering the MBD3 expression by novel genome-editing tools also holds a great promise for future gene therapies.

Upon recognition of a 6-8 nucleotide "seed" region in target mRNAs to halt translation or trigger the mRNA degradation, miRNAs pose another layer of post-transcriptional regulation in gene expression. A growing body of evidence has supported the involvement of miRNAs in each stage of oncogenesis. The miR-17-92 cluster, also dubbed "oncomiR-1", gives birth to a group of miRNAs that control the cell proliferation, apoptosis, and angiogenesis. Activation of the miR-17-92 cluster (*i.e.*, the MIR17HG on C13orf25) is mainly driven by two classes of transcription factors – the Myc and E2F proteins. In malignant glioma, the copy-number amplification of MIR17HG is rare, indicating its up-regulation primarily through the overactive Myc or E2F. The loss-of-control on the transactivation of miR-17-92 has been identified as a hallmark in hematopoietic malignancies and solid tumors. For instance, overexpression of miR-17-92 was reported in neuroblastoma and correlated with a poor prognosis.[170-172] Other miR-17-92-related cancers include lymphoma,[173] leukemia,[174] retinoblastoma,[175] medulloblastoma,[176] colorectal cancer,[177] breast cancer[178] and so on. In this study, the up-regulation of miR-17-92 in GBM, in comparison to low-grade glioma, can be primarily attributed to the significantly higher expression of n-Myc (up by 4.46-fold) and E2Fs (*e.g.*, E2F1 up by 7.77-fold and E2F2 up by 16.04-fold). In function, MBD3 is able to epigenetically suppress the transactivation

potential of Myc proteins. This explains the secondary increase of the MIR17HG transcripts when MBD3 was depleted in GBM cells.

Encouragingly, MBD3 can complement 5hmC to bring about a remarkable value to predict the glioma survival. This positive correlation inspires a mechanistic exploration. Catalyzed by the TET family enzymes, 5hmC is the first oxidized derivative of 5mC. The content of 5hmC is heterogeneous among different tissues but is most enriched in the CNS (0.3-0.7% of the whole genome).[179, 180] 5hmC locates in the midst of the DNA demethylation pathway, hence critical for the dynamic gene activation. Besides being a demethylation intermediate, 5hmC has further proven to be a predominantly stable DNA modification that regulates a plethora of biological functions through interactions with various partners. MBD3 is one of the few identified factors exhibiting appreciable *in vivo* association with 5hmC, though some recent studies failed to validate the specific *in vitro* MBD3-5hmC interaction. Considering the organ-dependent and dynamic nature of epigenetic modifications as well as their extensive cooperation in the regulatory network, the connection between MBD3 and 5hmC in the CNS ought to be substantive because such an enriched cytosine modification (*i.e.,* 5hmC) has to be interpreted by appropriate readers. Interestingly, the enrichment of MBD3 or the 5mC-binding protein MBD2 indicates opposite prognosis in the glioma survival even though their sequences are highly similar (Figure 4.13). This intra-MBD family divergence highlights the epigenetic intricacy and indirectly signifies a MBD3-MBD2 counteraction in the DNA methylation homeostasis as discovered in Chapter 3. The clinical implication about 5hmC reduction in a couple of cancers, including malignant glioma, has been preliminarily established. This study incorporates one of the major 5hmC-associated factors to bolster the reliability of using ep-

igenetic markers to predict the survival of glioma patients, as well as provides an alternative epigenetic marker easier to be characterized in clinical samples. In long term, it is expected that the obtained information will provide new insights to clinical management of individual patients and epidemiological reference for large-scale studies.

CHAPTER 5. A PLGA-PEG NANO-DELIVERY SYSTEM FOR EPIGENETIC DNA DEMETHYLATING THERAPY

5.1 Introduction

Epigenetic aberration has the potential to reshape cellular functions and is appreciably involved in oncogenic processes. Significantly, epigenetic silencing of tumor suppressor genes is one of the major causes for tumor progression and drug resistance.[181, 182] However, in contrast to genetic mutations, the cancerous epigenetic errors can be reversed through chemotherapy with relative ease and less side effects because epigenetic modifications do not need to change the DNA sequence of a gene. This feature makes epigenetic therapy a promising intervention strategy.[183, 184] The essential goal of epigenetic therapy is to target the diseased chromatin in rapidly dividing tumor cells and return it to a phenotypic "normal state" while only mildly disturbs the epigenome in healthy cells.

The most widely studied epigenetic modification in human cancers is DNA methylation. This modification is one of the prevailing marks for controlling gene expression, genomic stability, chromatin integrity, and cell fate.[185, 186] Regions with dense CpG dinucleotides, namely CpG islands, are predominantly found at gene promoters and usually free of methylation in normal cells. The idea of therapeutically inhibiting DNA methylation emerged from studies showing that a plethora of genes (mainly tumor suppressor genes and apoptosis-related genes) boast a gain-of-methylation (hypermethylation) dur-

ing tumorigenesis.[187, 188] This process is associated with an epigenetic inactivation of the gene expression and thus loss of the protein functions, which leads to the revival of interest in drugs discovered decades ago being potent inhibitors of DNA methyltransferases (DNMTs) to restore relevant anti-tumor activities in malignant cells.[57, 189]

Two commercially available agents targeting DNA methylation – AZA and DAC – are now alternative options for treatment of myelodysplastic syndrome.[190, 191] As cytidine analogues, AZA and DAC exert their effect by incorporating into replicating DNA and inhibiting DNMTs by forming covalent adducts in fast-dividing cells. With more DNMTs entrapped, DNA methylation is incrementally reduced along cell proliferation. Eventually, demethylation at tumor suppressor genes is achieved in daughter cancer cells, but only minimally affects resting or non-dividing cells.[192]

A growing body of evidence (from preclinical tests and clinical trials) suggests that the DNA demethylating drugs might induce an anti-cancer response in solid tumors.[193-196] However, extending the therapeutic effects of DNA demethylating agents to *in vivo* solid tumors still faces substantial hurdles due to the difficulties in drug delivery and cellular uptake.[189] In order to attenuate abnormalities of DNA methylation in cancer cells, several aspects of the current drug formulation need to be carefully improved: aqueous stability, *in vivo* bioavailability, transmembrane efficiency, resistance to efflux, side effects, and long-term efficacy.[197] Among the challenges, effective delivery of the DNA demethylating drugs is perhaps the most significant barrier for expanding the therapeutic indications from blood malignancies to solid tumors.

Efficient accumulation of cytidine analogues in cells relies on the presence of particular membrane nucleoside transport proteins such as the hCNT proteins.[198-200] The ATP-

dependent cellular internalization and short half-life of the DNA demethylating drugs in physiological liquids would compromise the expected therapeutic activities against solid tumors. AZA and DAC are subjected to a series of physicochemical attacks when administered to human body, mainly including degradation by hydrolysis, deformation of the triazine ring, and anomerization of the ribose ring under neutral-to-basic conditions.[201, 202]

One strategy to overcome the pharmacological challenges in drug delivery is to modify the drug structures. However, a considerable number of those modifications could inevitably diminish the therapeutic efficacy and cause unwanted toxicity. A variety of AZA derivatives have been developed, yet with decreased performance compared to the parental drug. Just to name a few here: 5-fluoro-2'-deoxycytidine is highly toxic;[203, 204] 1-(β-D-ribofuranosyl)-1,2-dihydropyrimidin-2-one is less potent;[205] 5,6-dihydro-5-azacytidine has a poor incorporation ability;[206, 207] 2'-deoxy-N4-[2-(4-nitrophenyl)ethoxycarbonyl]-5-azacytidine (NPEOC-DAC) severely impacts liver function;[208] and some drugs require pre-activation by certain enzymes: CP-4200 by plasma membrane lipases,[209] S110 by cellular phosphodiesterases,[210] and NPEOC-DAC by carboxylesterase.[207, 208]

Another strategy is to compose a nanoscale delivery system that has the potential to improve the therapeutic efficacy of demethylating drugs. An ideal delivery vector should protect drugs from premature hydrolysis in blood circulation, increase bioavailability, and enhance cellular internalization. With controlled fabrication, nano-carriers could facilitate the *in situ* retention of demethylating agents, extend the exposure of cancerous tissue to an efficacious drug concentration, and minimize the re-methylation of tumor-suppressor genes until an irreversible elimination of tumor cells. In this regard, the nanomaterial-protected drug molecules become more effective than parental compounds in treating sol-

id tumors.[211, 212] Till today only a handful of nano-vectors have been developed to deliver epigenetic chemo-payloads to solid tumors, but none of the past work has demonstrated the potential of a poly(lactic-co-glycolic acid) (PLGA)-based system for delivery of the DNA demethylating drugs.

Use of the biodegradable PLGA as the drug delivery system has been well defined, while most of the applications were demonstrated to physically encapsulate the drug molecules. However, this formulation is sensitive to various environmental factors that could result in an undesired burst release.[213, 214] Instead, chemical conjugation of a drug with the backbone of PLGA would greatly enhance the drug stability and also facilitate a pH-dependent release. Hence, in this work the aliphatic polyester PLGA is combined with the hydrophilic polyethylene glycol (PEG) to produce an A-B type of di-block copolymer for conjugation with AZA or DAC (Figure 5.1). Under an equimolar drug concentration, the nano-conjugate form of drug experiences a rapid endocytosis and can better reactivate the apoptotic pathway to inhibit cell proliferation. In light of this improved cell uptake, a low concentration of DAC-PLGA-PEG was applied to sensitize the alkylation-resistant glioblastoma cells to temozolomide. These polymers are able to diffuse through loosely packed endothelial cell junctions of blood capillaries in the vicinity of tumor tissue and enrich in tumor cells due to the “enhanced permeation and retention” (EPR) effect. With prolonged circulation and bioavailability, AZA-PLGA-PEG gains a more potent *in vivo* efficacy over free AZA in treating xenografted breast cancers. Notably, the active glycolysis in cancer cells often generates an acidic microenvironment in which the PLGA-PEG-based nano-vector could release the conjugated drug molecules at a faster rate than in normal cells. This unique property substantially avoids an excessive damage to normal

cells. Taken together, this chapter describes a novel delivery strategy for DNA demethylating agents and provides a cost-effective formula for future pharmaceutical applications.

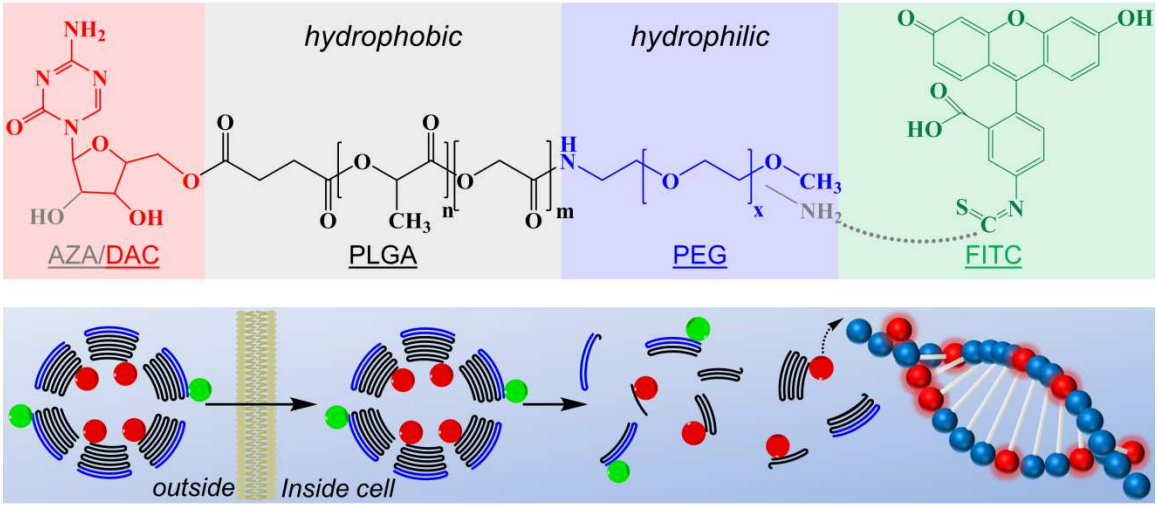


Figure 5.1 Design of the nano-conjugate and principle of delivery.

5.2 Methods and Materials

5.2.1 Synthesis of PLGA-PEG-OCH_3 (Conjugate 1)

The PLGA-PEG di-block was synthesized as described in previous studies with minor modifications.[215] PLGA 50:50 with an average molecular weight 8,586 Da (0.31 mM) was dissolved in dry dichloromethane (DCM, 10 mL). The carboxylic acid group in PLGA was first activated by N,N'-dicyclohexylcarbodiimide (DCC, 0.375 mM) and N-hydroxysuccinimide (NHS, 0.62 mM) at room temperature overnight. The side product dicyclohexylurea (DCU) was filtered out; the obtained PLGA-NHS was precipitated by ice cold diethyl ether (Et_2O) and dried under vacuum. The activated PLGA (2 g) was then added into methoxy polyethylene glycol (H_2N-PEG-OCH_3) (molecular weight = 750 Da, 0.42 mM) solution in 10 mL dry DCM, and the reaction mixture was stirred at room tem-

perature for 24 h. The PLGA-PEG conjugate was precipitated out from the reaction mixture by adding Et_2O and centrifuged at 3,220 g for 10 min. The supernatant was discarded and the polymer was washed with a mixture of Et_2O and hexane 50:50 (3 × 5 mL) to remove excess reagents. Finally, the polymer was dried under vacuum for 24 h to obtain the conjugated product (yield = 92.83%). The coupling efficiency was determined by ^1H NMR spectroscopy: ~99.2% of PLGA was conjugated with PEG.

5.2.2 Activation of PLGA-PEG-OCH_3 by Succinic Anhydride (Conjugate 2)

The PLGA block of Conjugate 1 was modified by reacting with succinic anhydride to produce HOOC-PLGA-PEG. 1.78 g of Conjugate 1 was dissolved in dry DCM (10 mL) containing 4-dimethylaminopyridine (DMAP 25% w/v) as a catalyst, and succinic anhydride (39.3 mg) was added. After stirring at 50°C overnight, the DCM was removed in vacuum. This Conjugate 2 was re-dissolve in dry DCM and precipitated in ice cold Et_2O, dialyzed against water for 4 h using MWCO 3,500 and lyophilized (yield = 81%). The polymer was characterized by ^1H NMR (400 MHz): δ (ppm) = 5.36–5.00 (m, PLGA-CH_2-), 4.94–4.57 (m, PLGA-CH_2-), 3.67–3.58 (m, PEG-CH_2-), 3.38 (s, PEG-OCH_3), 1.52 (s, PLGA-CH_3). The percentage of modification was determined to be 97.2%.

5.2.3 Synthesis of AZA/DAC-PLGA-PEG (Conjugate 3)

Conjugate 2 (0.08 mM) was dissolved in dry DCM containing 4-(dimethylamino) pyridinium-4-toluene sulfonate (DPTS, 0.23 mM) and cooled at 0°C. For the reaction mixture, AZA or DAC (0.09 mM) dissolved in dry N,N-dimethylformamide (DMF) was added with continuous stirring. Then 0.23 mM DCC in pyridine was added slowly to the

solution and the reaction was carried out at 4°C for 72 h under N_2. The progress of reaction was monitored by TLC with $CHCl_3/CH_3OH$ (1:4, v/v) as eluent (product retention factor: Rf_{AZA} = 0.56, Rf_{DAC} = 0.68). The DCM and pyridine were removed under reduced pressure and the product was re-dissolved in DCM and precipitated in ice cold Et_2O. Untreated drug and other impurities were removed followed by dialysis. The dialyzed solution was precipitated in ice cold Et_2O/hexane (50:50) mixture, freeze-dried and kept at -20°C. The drug conjugate was characterized by UV and ^{1}H NMR spectroscopy. ^{1}H NMR of AZA-PLGA-PEG (400 MHz): δ (ppm) = 8.56 (s, 1H, H-6), 7.49-4.53 (bs, 2H, NH_2), 5.76 (s, SA-O-CH_2), 5.65-5.64 (d, 1H, H-1′, J = 4), 5.40-5.41 (bs, 1H, OH-2′), 5.25-5.14 (m, PLGA-CH_2-),5.03-4.99 (m, 2H, H-4′ and OH-3′), 4.91-4.83 (m, PLGA-CH_2-), 3.99 (m, 1H, H-3′), 3.84-3.81 (m, 1H, H-2′), 3.67–3.59 (m, 2H, H-5′; H-5′′), 3.49 (m, PEG-CH_2-), 3.22 (s, PEG-OCH_3), 1.46 (s, PLGA-CH_3). ^{1}H NMR of DAC-PLGA-PEG (400 MHz): δ (ppm) = 8.56 (s, 1H, H-6), 7.52-7.49 (s, 2H, NH_2), 6.15 (t, 1H), 5.26 (s, SA-O-CH2), 5.25 (m, PLGA-CH_2-), 4.91-4.83 (m, PLGA-CH_2-), 4.2 (s, 1H, H), 3.76 (1H, H), 3.6–3.7 (m, 1H, H), 3.7–3.5 (m, 2H, H), 3.49 (m, PEG-CH_2-), 3.22 (s, PEG-OCH_3), 1.46 (s, PLGA-CH_3). The coupling efficiency revealed that over 90% of the carboxyl ends of PLGA were conjugated with AZA.

5.2.4 Synthesis of PLGA-PEG-FITC (Conjugate 4)

The di-block polymer PLGA-PEG-NH_2 (600 mg) was prepared as discussed above and dissolved in 20 mL DCM. FITC (22.3 mg) was dissolved in 250 µL of a mixture of DCM/pyridine and added to the copolymer solution. The reaction was stirred at 4˚C for 24 h in dark. The resultant conjugate was precipitated by ice cold Et_2O (125 mL). The

polymer was centrifuged at 3,220 g for 30 min. The supernatant was discarded and the polymer was washed thoroughly by Et_2O (5 mL × 3) and dried under vacuum overnight. The yield of FITC conjugation was determined using fluorescence spectrometry at an excitation wavelength of 485 and emission at 530 nm.

5.2.5 Preparation and Characterization of Nano-Micelles

Due to the amphipathic property of PLGA-PEG, it would form nano-micelles in solution and the drug molecules are protected in the core. Non-fluorescent (only drug-containing conjugate) and fluorescent (drug-containing conjugate : FITC-containing conjugate = 2 : 1) nano-micelles were prepared by film dispersion method.[216, 217] A proper amount of compounds was dissolved in chloroform at room temperature. The solvent was then removed under reduced pressure to form a dry polymer film. This film was hydrated with 10 mM HEPES buffer (pH = 7.4) at room temperature for 30 min. The nano-micelles were extruded through a 200 nm polycarbonate membrane. The yield of products was determined by UV-Vis spectroscope for non-fluorescent nano-micelles and fluorimeter for fluorescent nano-micelles.

During synthesis, gel permeation chromatography (GPC) was performed to monitor each conjugation using an Agilent 1200 series Wyatt Dawn Heleos-II system equipped with a UV, LSD & RID detector. DMF was used as the mobile phase at a flow rate of 0.3 mL/min through OHpak SB-803 HQ column [8.0 × 300 (mm): inner diameter × length] at 40°C. GPC calibration was constructed using polystyrene standards. The size and morphology of nano-micelles were determined by transmission electron microscope (TEM, Philips, CM1000) with the acceleration voltage of 100 kV. Samples were diluted to 0.5

mg/mL with HEPES buffer and stained with 0.8% uranyl acetate. The surface charge of nano-micelles was measured with a ZetaPlus analyzer (Brookhaven).

5.2.6 Uptake and Release Assays

The uptake and intracellular release of fluorescent nano-micelles were characterized by confocal imaging, FCS, PCH, and FLIM. The pH-dependent drug release kinetics was evaluated with a UV-Vis method. AZA/DAC-PLGA-PEG nano-conjugate was suspended in 500 µL PBS (pH = 5 or 7.4) and dialyzed (MWCO 3,000 Da) against 1.5 mL medium. Then the tubes were placed in an orbital shaker setting at 120 rpm and 37°C. At predetermined time points (1h, 2h, 4h, 8h, 12h, 24h, 48h, 72h), a 10 µL aliquot was withdrawn and replaced with fresh buffer. The released AZA or DAC concentration was quantified at a characteristic wavelength: AZA – λ_{max} 254 nm (pH 5) and 223 nm (pH 7.4); DAC – λ_{max} 244 nm (pH 5) and 220 nm (pH 7.4).[218] The calibration curve for drug quantity was linear over the concentration range from 100 to 4000 ng/mL.

5.2.7 *In vitro* Therapeutic Tests of DAC-PLGA-PEG against Glioblastoma

Human SF767 and U87 GBM cells were cultured in IMDM medium supplemented with 10% fetal bovine serum, 1% antibiotics, and 1% glutamate. MTT assay was used to test the short-term cell viability after DAC-PLGA-PEG and temozolomide treatment. In a 96-well plate, 3,000 cells were seeded in each well to grow for 24 h, followed by incubation with appropriate drugs. After 72 h, the culture medium was quickly removed and the cells were rinsed with clean 1 × PBS. The cells were then incubated with 100 mL fresh medium containing 10 mL thiazolyl blue tetrazolium bromide solution (MTT, 5 mg/mL

in PBS as stock) for 3-4 h and the formed formazan was dissolved with 150 mL acidic isopropanol prior to measurement by a microplate reader at the absorbance of 570 nm. For long-term proliferation assay, 5,000 cells were seeded to a T25 flask at day 0, and the cell number under each specific condition was recorded for 8 days. During the long-term treatment, the global DNA methylation was quantified with colorimetric immunoassay at day 4 and day 8, respectively. In addition, four apoptosis-related genes (p53, p21, Caspase3, and BRCA1 by qRT-PCR) and cell cycle progression (flow cytometry) were evaluated upon different treatments.

Table 5.1 PCR primers used in Chapter 5

GAPDH	(F)CAGCCTCAAGATCATCAGCA (R)TGTGGTCATGAGTCCTTCCA
p53	(F)CCCCTCCTGGCCCCTGTCATCTTC (R)GCAGCGCCTCACAACCTCCGTCAT
p21	(F)GAGGCCGGGATGAGTTGGGAGGAG (R)CAGCCGGCGTTTGGAGTGGTAGAA
Caspase3	(F)ATGATGACATGGCGTGTCATAA (R)AAGGAAAAGGACTCAAATTCTGTTG
BRCA1	(F)CAAGGAACCAGGGATGAAATCAG (R)ATGGCTCCACATGCAAGTTTG

5.2.8 *In vivo* Therapeutic Tests of AZA-PLGA-PEG against Breast Cancer

After testing the *in vitro* cytotoxicity of AZA-PLGA-PEG by MTT assay, the *in vivo* anti-tumor potential was assessed. Female nude mice (18-25 g in body weight) were subcutaneously implanted with 5×10^6 MCF7 cells. Following xenografting, tumors were allowed to grow for another 21 days until reached 190-200 mm^3. Mice were sorted randomly into three groups: control, free AZA treatment, and AZA-PLGA-PEG treatment. A total of 16 doses of PBS saline, free AZA (1 mg/kg), or AZA-PLGA-PEG conjugate (with equivalent AZA content of 1 mg/kg, micelle concentration = 1.53 mg/mL) were

administered via intraperitoneal (IP) injection to each group in four weeks (two cycles of eight consecutive doses). Tumor volume was measured twice a week with an electronic caliper ($v = 0.5 \times L \times W^2$ mm^3) and the body weight of mice was monitored in parallel. All animal handling followed the guidelines of the Purdue Animal Care and Use Committee and approved by the Purdue Center for Cancer Research.

5.2.9 Statistical Analysis

In significance test: Student's *t*-test (two-tail) was used for two sample comparison, and ANOVA was used for multiple sample comparison. All the analyses were performed with the OriginPro software. Without specific note, all data are presented with mean and standard deviation.

5.3 Results and Discussion

5.3.1 Synthesis and Characterization

Cytidine analogues are highly unstable in aqueous condition, which greatly restricts their experimental and clinical applicability. This instability is due to the rapid hydrolysis and opening of the heterocyclic nitrogen ring caused by nucleophilic attacks at the N-5 position. The formed N-(formylamidino)-N-β-d-ribofuranosylurea intermediate would be then slowly converted into 1-β-d-ribofuranosyl-3-guanylurea.[202] This destructive decomposition hinders efficient drug conjugation and storage. New experimental methodologies are thus needed for more effective particle formulation and drug delivery.

Two methods have been developed for incorporation of drug into nano-vectors. The most widely used is physical entrapment, which is quite simple as no chemical modifica-

tion is required. However, burst release is a grand shortcoming due to a premature loss of the physically loaded drug before reaching the target site. The second method is chemical conjugation in which the drug molecules are covalently bonded to nano-vectors and then undergo a controlled release since additional mechanisms are usually needed to trigger the pharmacological action. This strategy has not been widely adopted mainly because of the limited number of reactive sites available on a nano-vector and few conjugation reactions that would not impact the raw compounds. For instance, even a tiny chemical linker could drastically change the structure and therapeutic value of a chemotherapeutic drug.

In this study the drug-containing nano-conjugate was prepared with a three-step synthetic route (Figure 5.2A). First, NHS-activated PLGA reacted with PEG to form the copolymer PLGA-PEG. Then the terminal hydroxyl group of PLGA-PEG was converted to carboxylic group by reaction with succinic acid, and the succinylated PLGA-PEG copolymer was allowed to react with the 5'-OH of AZA or DAC in the presence of DCC and DPTS, to result in the formation of AZA/DAC-PLGA-PEG conjugate. Products were isolated and purified by precipitation and dialysis, followed by characterization with GPC (Figure 5.2B), TEM (Figure 5.2C), and ^{1}H NMR (illustration with AZA-PLGA-PEG in Figures 5.3-5.5). The drug content was determined to be 25 µg/mg for AZA-PLGA-PEG and 22.9 µg/mg for DAC-PLGA-PEG. The formed nano-micelles in solution are spherical in shape with an average size of about 50 nm. The surface charge was found to be nearly neutral by zeta-potential (Table 5.2). The physicochemical characteristics of polymeric nanoparticles including size, shape, and surface charge are critical factors for their stability, *in vivo* pharmacokinetics and bioavailability. In human body, exogenous molecules have to evade the reticuloendothelial system (RES) which can otherwise opsonize

the “invader molecules” and subsequently induce a degradation or phagocytosis by macrophages.[219] PLGA-nanoparticles below 100 nm with a hydrophilic surface modification such as PEG can minimize opsonization and clearance by the RES, which would prolong the circulation time and allow more drug delivered to tumor lesions.[220] Another rationale for PEGylation aims to solve the problem that positive-charged particles easily bind to a variety of cells in the body, such as vascular endothelial cells, before reaching the target loci. Hence, it is favorable to maintain a nano-vector as either neutral or slightly anionic for stabilization and preventing non-specific drug loss.[221, 222]

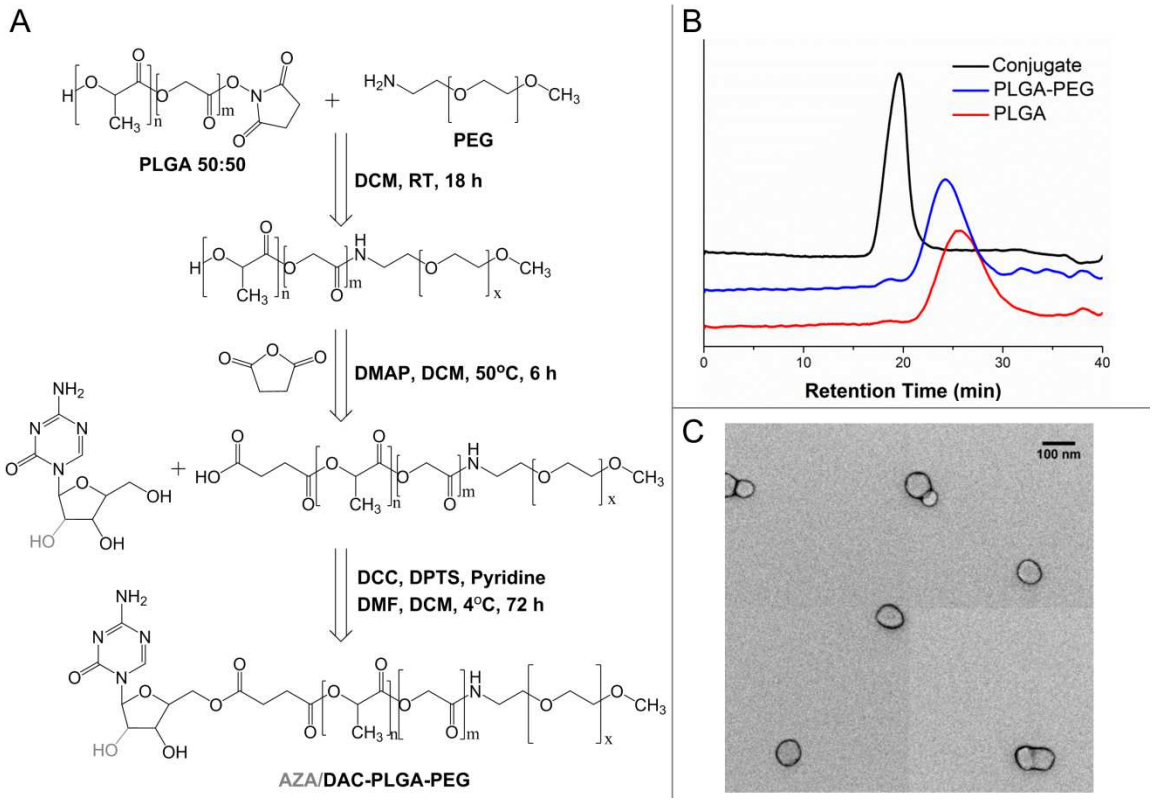


Figure 5.2 Synthesis of AZA/DAC-PLGA-PEG nano-conjugate. (A) Steps for chemical conjugation. (B) Successful conjugation for each component was validated by decreased retention time in GPC. (C) TEM image of the formed nano-micelles.

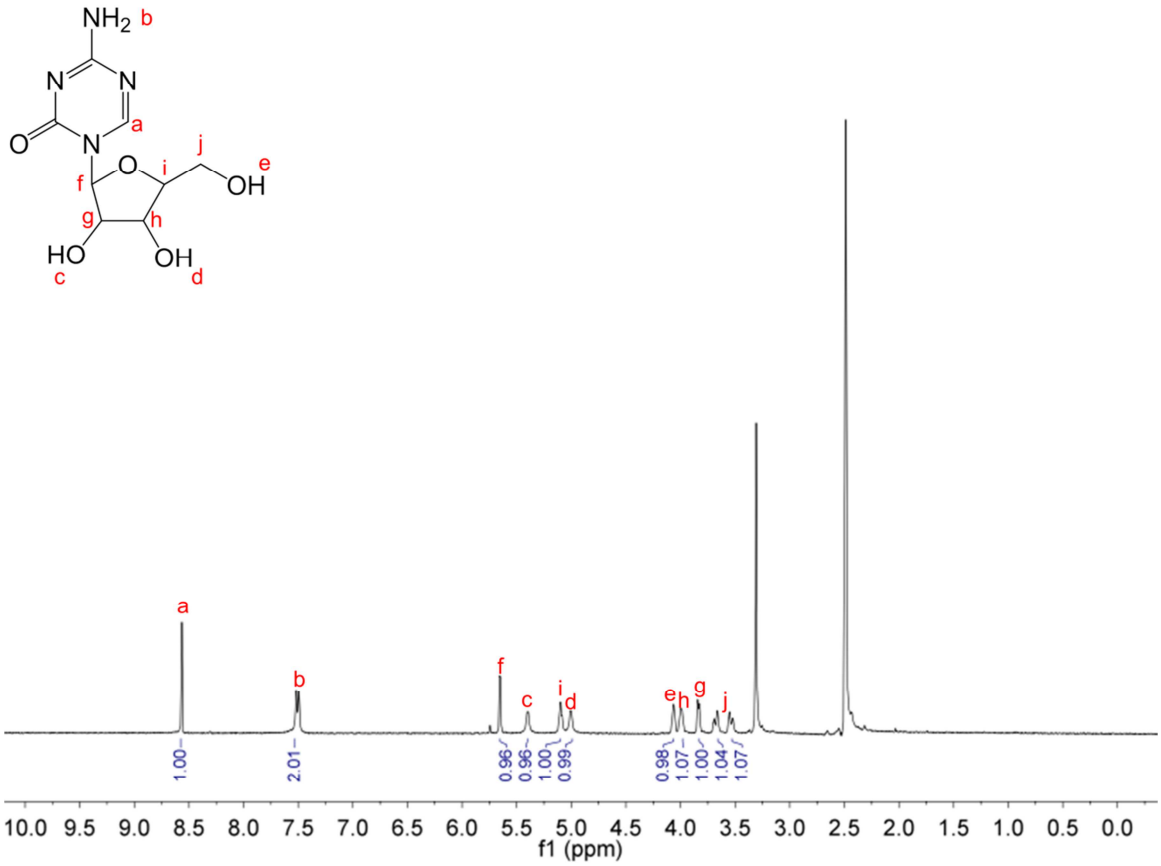


Figure 5.3 ^{1}H NMR spectra and the corresponding peak assignments for AZA.

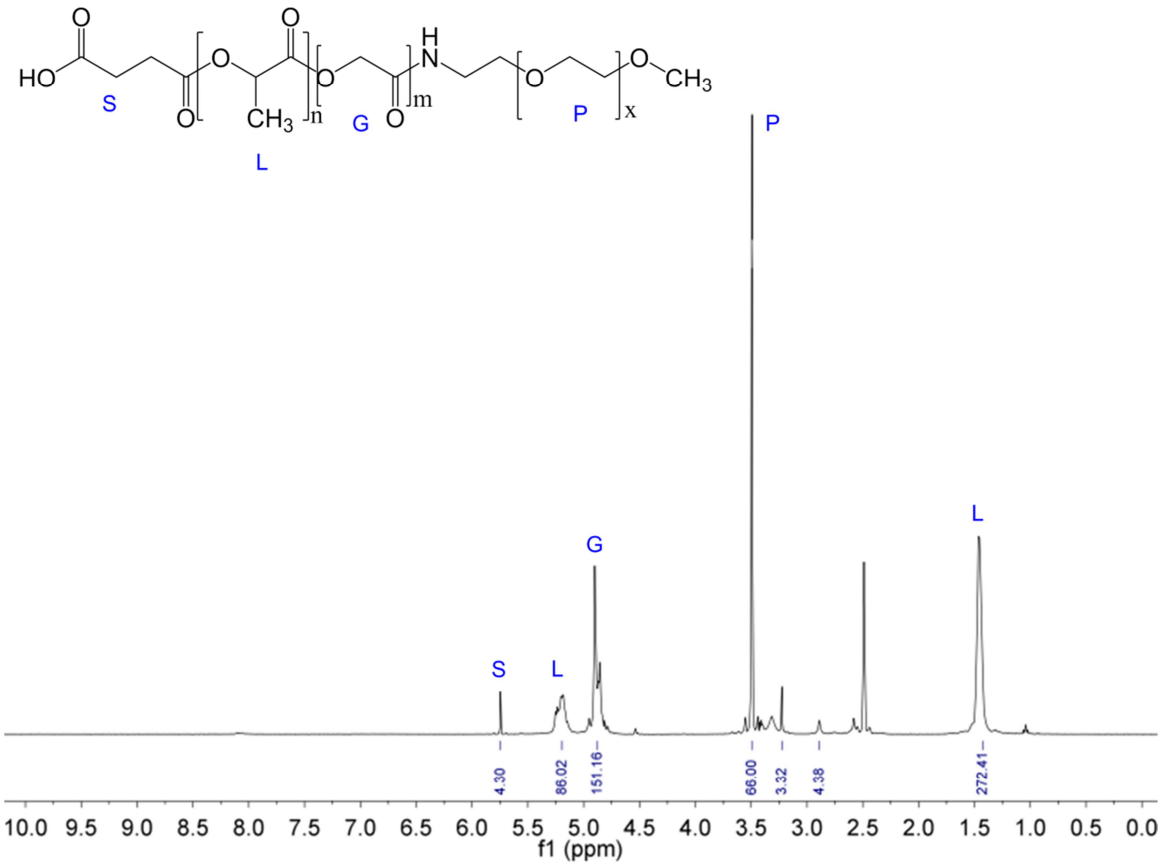


Figure 5.4 ^{1}H NMR spectra and the corresponding peak assignments for activated PLGA-PEG.

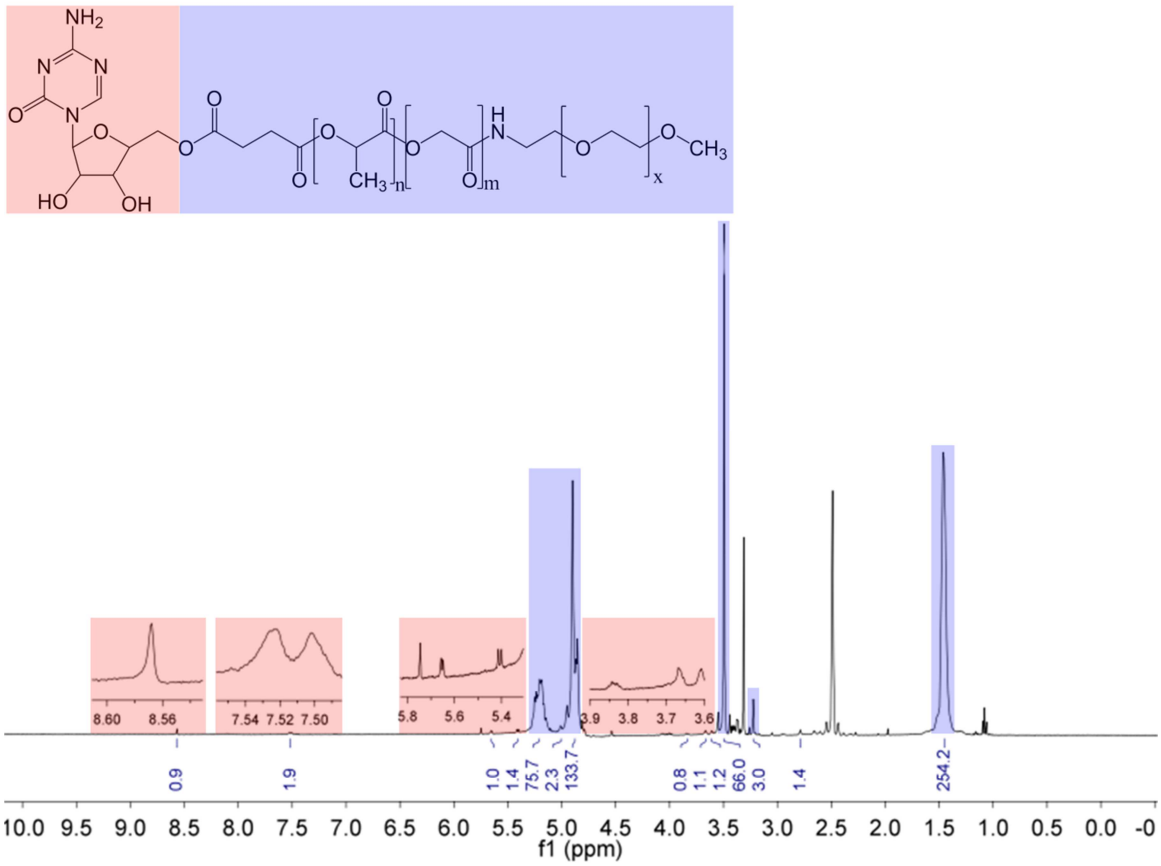


Figure 5.5 ^{1}H NMR spectra and the corresponding peak assignments for AZA-PLGA-PEG.

Table 5.2 Basic characteristics of the nano-micelles with different compositions

Nano-micelle Composition	Particle Size (nm)	Surface Charge (mV)
AZA-PLGA-PEG	45 ± 5.2	-0.87 ± 2.1
DAC-PLGA-PEG	55 ± 6.4	-0.99 ± 1.5
PLGA-PEG-FITC	59 ± 3.4	-1.10 ± 2.3

5.3.2 Uptake and Release Kinetics

To track the drug uptake and release kinetics, fluorescent nano-micelles were incubated with cells for up to 6 hours. From the time-lapse confocal imaging, it can be seen

that the cross-membrane diffusion rapidly completed within 2 hours (Figure 5.6A). Next, a more quantitative exploration with single-molecule fluorescence tools was performed (Figure 5.6B). By mixing 50 nM drug-containing conjugate and 25 nM FITC-containing conjugate, the original concentration of the nano-micelle solution was determined to be 13.8 nM in FCS calculation. The intracellular concentration of nano-micelles could reach 83.5 ± 5.0 nM within the first hour of incubation, validating the high uptake efficiency of this PLGA-PEG-based nano-vector. After 4 hours, the intracellular concentration peaked at 145.8 ± 13.4 nM as shown in Figure 5.6B. Notably, the diffusion coefficient of nano-micelles in living cells did not experience drastic changes within the detection window (Figure 5.6B, inset), which indicates that no significant aggregation or abrupt particle disintegration occurred in 6 hours. From PCH, the average brightness of nano-micelles displayed a gradual increase (Figure 5.6C). It gives rise to a speculation that some PLGA-PEG fragments with FITC tags hydrolyzed from early taken nano-micelles would integrate into newly internalized micelles due to physical adsorption or hydrophobic force. In the fluorescence images, a uniform distribution of nano-micelles over the entire cell was noted without a significant compartmental entrapment, implying an endo-lysosomal elusion (Figure 5.6D). The results from FCS and PCH together suggest that the maximum uptake of the PLGA-PEG nano-micelles could occur within 1 hour (Figure 5.6C, black dash line), followed by an accelerated hydrolysis process (Figure 5.6C, red dash line).

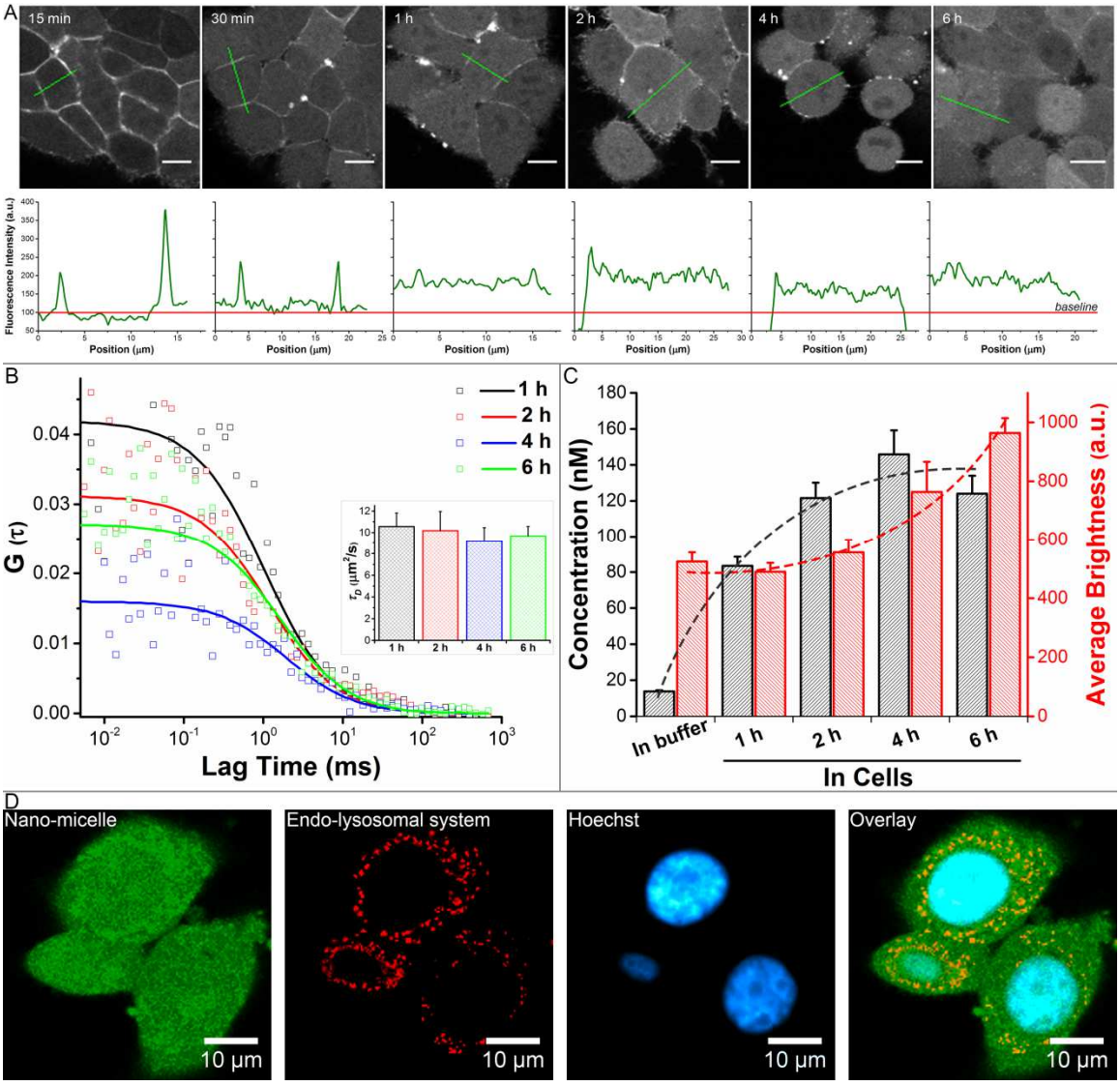


Figure 5.6 Uptake and distribution of PLGA-PEG nano-micelles. (A) Confocal imaging of cells incubated with fluorescent nano-micelles. (B) Autocorrelation functions from live cells were fitted with a 3D diffusion model to obtain the diffusion coefficient (inset) and concentration of internalized nano-micelles. (C) Summary of the average concentration and brightness of nano-micelles in cancer cells (n > 20 cells, mean + S.E.M.). (D) Escape of nano-micelles from the endo-lysosomal system.

To verify the pH-dependent PLGA hydrolysis, *in vitro* drug release assay was performed in acidic and neutral environments. As shown in Figure 5.7B, the hydrolysis rate was significantly faster at an acidic pH than in a neutral environment. Approximately 85%

of the conjugated drug was released at pH 5.0 within the first 12 hours, whereas only ~40% was released at pH 7.4. This supports the susceptibility of the ester bond-based PLGA-PEG conjugate to acid; however, a more comprehensive investigation is required for future *in vivo* applications due to the lack of analytical methods to distinguish the intact AZA/DAC from the degraded intermediates.[192, 223] Moreover, the hydrolysable chemical links within the conjugate render the rate of drug release a near zero-order kinetics under physiological pH, which provides additional benefits for *in vivo* usage.

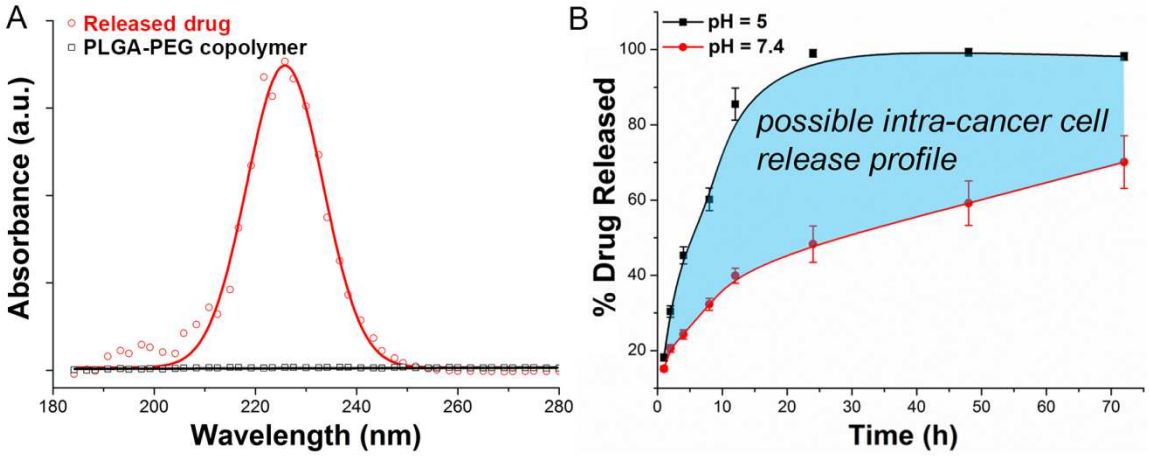


Figure 5.7 pH-dependent drug release from PLGA-PEG nano-micelles. (A) UV-Vis spectra of cytidine analogue and the PLGA-PEG copolymer below 300 nm. (B) *In vitro* drug release kinetics in acidic or neutral environment (n = 3).

In cancer cells with a relative acidity caused by the Warburg effect, upon the hydrolysis of PLGA a fast accumulation of lactic acid and glycolic acid was expected, and these two products would further acidify the intracellular environment to facilitate a drug release cascade. To some extent, the acidifying pH could not only advance the PLGA hydrolysis but also slow down the triazine ring-opening decomposition of the drugs,[218] thus conferring a better efficacy against cancer cells. Since a reduced fluorescence lifetime of

FITC is a sensitive measure of low pH, FLIM was used to assess the PLGA hydrolysis-related acidification in living cells. Figure 5.8A shows the time-course FLIM images at different incubation times. It is evident to see that along the incubation the fluorescence lifetime of FITC was continuously shortening, which can also be perceived from the shift in the lifetime histograms (Figure 5.8B). Taken together, the intracellular behavioral mode for PLGA-PEG nano-micelles was proposed (Figure 5.9). Having a size of 50-60 nm, the formed nano-micelles can rapidly enrich in cancer cells within 1-2 hours and the delivery of AZA/DAC here, to a great extent, is thus independent of nucleoside transporters. Instead, the trend of an initial linear increase and a later saturation stage as revealed by FCS, strongly suggest that the cell internalization of sub-100 nm PLGA-PEG nanoparticles is the outcome of a cooperation between fluid-phase pinocytosis and clathrin-mediated endocytosis,[224] in contrast to the mechanism for large-sized nanoparticles (*e.g.*, via extracellular release of the drug payload to contacting cells).[225] The rapid endo-lysosomal escape is another advantage for the PLGA-PEG-based delivery since a further anionic-to-cationic reversal of the surface charge facilitates the nanoparticle to interact with the endo-lysosomal membrane and achieve the escape in 10 minutes.[226] Afterwards, the internalized nano-micelles begin to be hydrolyzed and are sensitive to acidic pH in cancer cells.

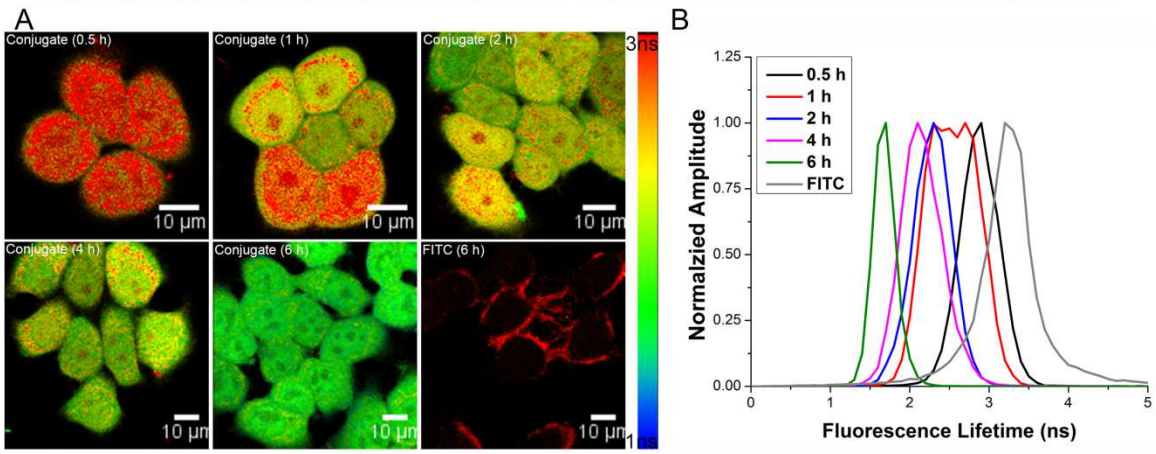


Figure 5.8 PLGA hydrolysis-induced acidification inside cells. (A) FLIM images collected from the FITC channel. (B) Lifetime histograms of the internalized FITC at different time points.

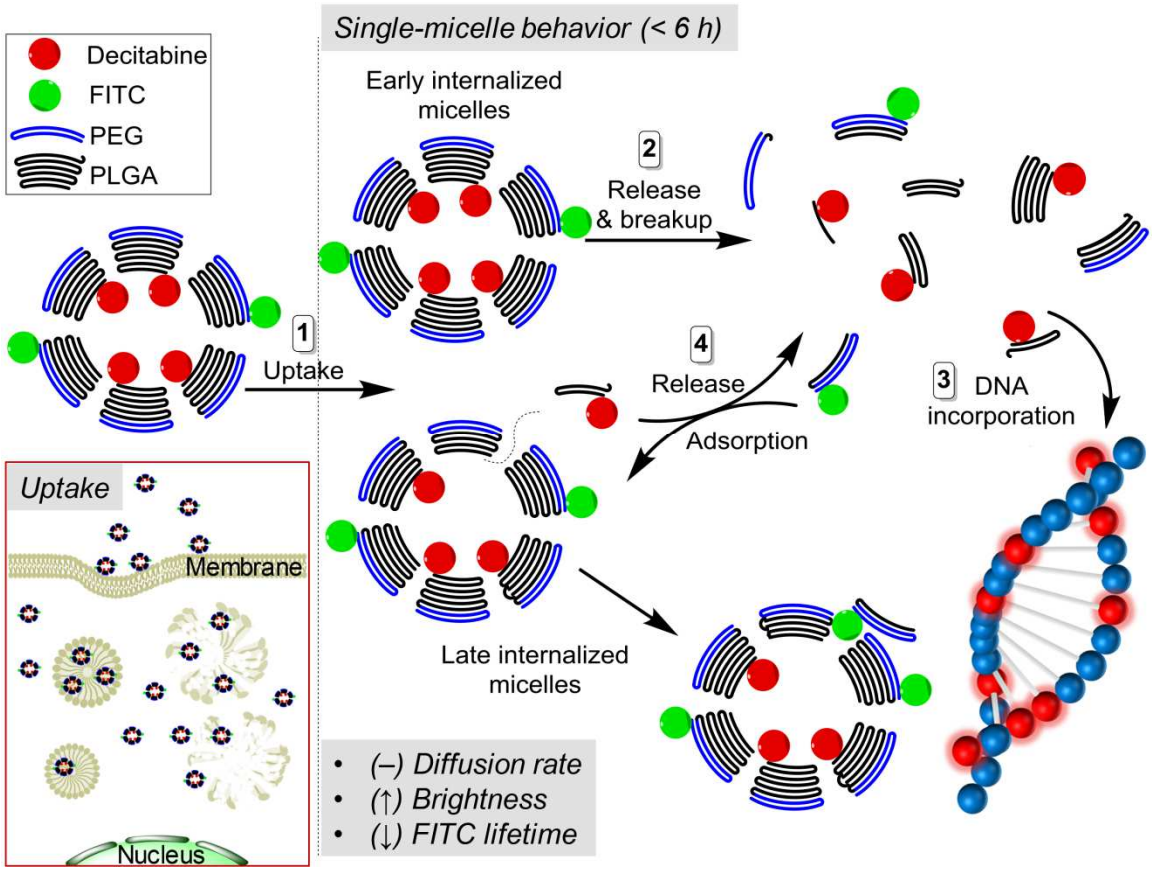


Figure 5.9 Single-molecule behavior of PLGA-PEG-based nano-micelles in cancer cells.

5.3.3 DAC-PLGA-PEG Sensitizes Glioblastoma Cells to Temozolomide

Arising from normal glial cells, glioblastoma (GBM) constitutes the most prevalent primary brain tumor in adults. Even with gross total resection in surgery and adjuvant therapy, the median survival for GBM patients is only 14.6 months and the two-year survival is approximately 30%. As the frontline chemotherapy for GBM, temozolomide is able to elicit a radical anti-tumor effect via inducing alkylation at the guanine residues in DNA. However, after a temporary remission GBM would virulently relapse and become chemo-resistant due to its infiltrating growth and rapid evolution to overcome selection pressure. It has been recently demonstrated that pretreating the chemo-resistant GBM cells with DAC could considerably sensitize GBM cells to temozolomide.[227] In light of this finding, it will be of tremendous value to improve the pharmacodynamics of DAC to further its application in GBM.

First, two GBM cell lines with distinct responses to temozolomide were confirmed by MTT assay. Cells were treated with temozolomide at a concentration of up to 30 μM for 72 hours. From the results, U87 cells exhibited full response to temozolomide treatment and a dosage of 30 μM could suppress the cell proliferation by 90% (IC50 was estimated to be ~20 μM). In contrast, SF767 cells did not show any significant response to 30 μM temozolomide (Figure 5.10A). In the subsequent experiments testing combination therapy, SF767 cells were first sensitized with DAC or DAC-PLGA-PEG with equivalent drug content (50 nM) for 24 hours, followed by treatment with 30 μM temozolomide for another 48 hours. When given separately, both DAC and DAC-PLGA-PEG showed certain cytotoxicity because of the non-specific effect, but the nano-conjugate form of drug was more effective than the free drug due to the improved delivery efficiency. For cells

pretreated with DAC or DAC-PLGA-PEG, 30 μM temozolomide could successfully suppress the proliferation rate by 47% and 62%, respectively (Figure 5.10B). This set of experiments thereby proves the synergistic effect between DAC/DAC-PLGA-PEG and temozolomide.

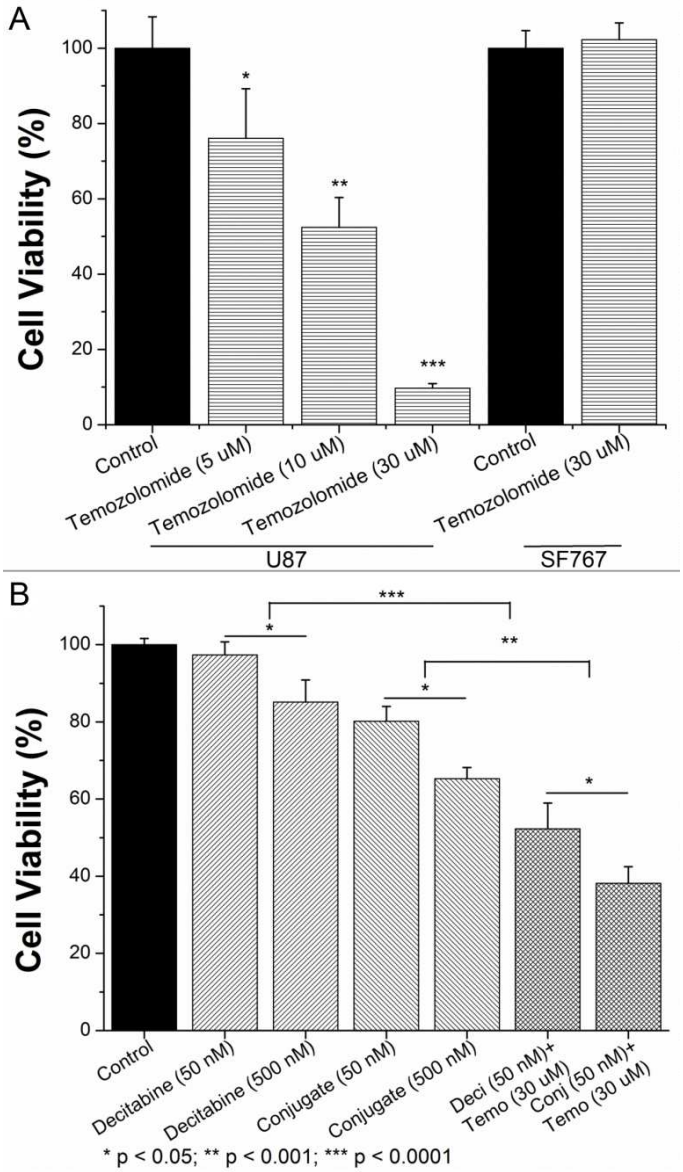


Figure 5.10 Therapeutic potential of combination treatments against GBM cells. (A) Distinct responses to temozolomide in U87 and SF767 cells. (B) Synergistic effects between temozolomide and DAC or DAC-PLGA-PEG were tested in SF767 cells (n = 6).

A critical pitfall for conventional PLGA-based delivery comes from its premature burst release.[228, 229] To solve this problem, the design of this study opted to chemically conjugate DAC to the PLGA backbone instead of physically load the drug. This formulation merits a sustainable and controlled release. Hence, a better long-term efficacy from a combinatorial therapy was expected. As presented in Figure 5.11A, the proliferation rate of SF767 cells under a combination treatment using DAC-PLGA-PEG and temozolomide is significantly lower (by ~70%) than that with DAC and temozolomide in 8 days. It was also noted that a larger extent of DNA demethylation could be achieved by DAC-PLGA-PEG than free DAC (Figure 5.11B), validating a more effective epigenetic remodeling.

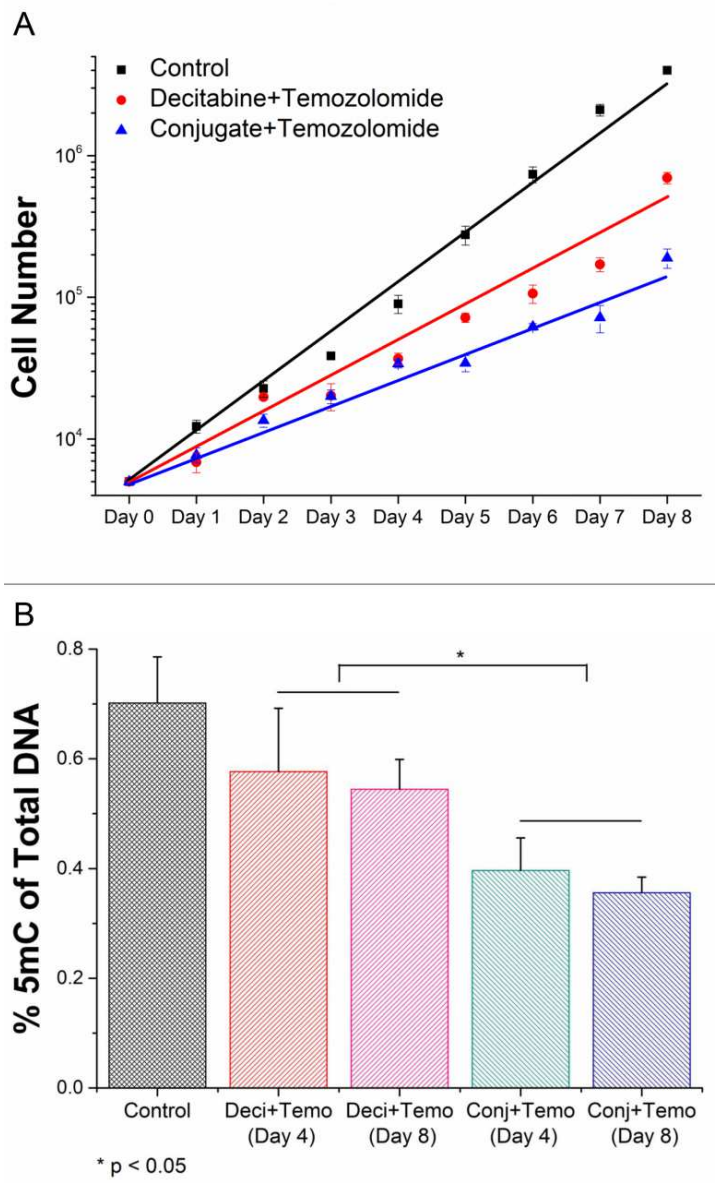


Figure 5.11 Long-term effects of the combination therapy. (A) The proliferation profiles of SF767 cells under combination treatments (n = 3). (B) The global changes in DNA methylation were monitored during the whole process (n = 3).

Since both DAC and temozolomide function by alkylating DNA yet on different nucleotides, the next query is to test whether their synergistic effect stems from triggering a more intense response to DNA damages. Quantitative RT-PCR was thus applied to check four key genes relating to DNA damage and apoptosis: p53, p21, Caspase3, and BRCA1.

Consistent with the cytotoxicity assays, 30 μM temozolomide alone did not evoke any significant changes in the transcription of these genes in SF767 cells, whereas the combination treatment (DAC-PLGA-PEG and temozolomide) radically reactivated p53 and p21 (Figure 5.12A). On the contrary, Caspase3 and BRCA1 were not influenced, verifying a DAC-induced, Caspase-independent apoptosis and the impaired BRCA1 function in glioma cells.[230, 231] The elevation of p53 and p21 could in turn halt the cell cycle progression at the checkpoints. Examined by flow cytometry, a much larger proportion of cells were found to be arrested in the G2/M-boundary when the combination treatment was administered (Figure 5.12B), which is in agreement with the previous research on DAC pharmacology.[232, 233] Collectively, these data suggest that one major mechanism for DAC to potentiate temozolomide in GBM cells hinges on better eliciting the DNA damage-induced proliferation retardation, though it cannot rule out that the observed increase in p53 and p21 might also be attributed to the induced demethylation at their promoters.

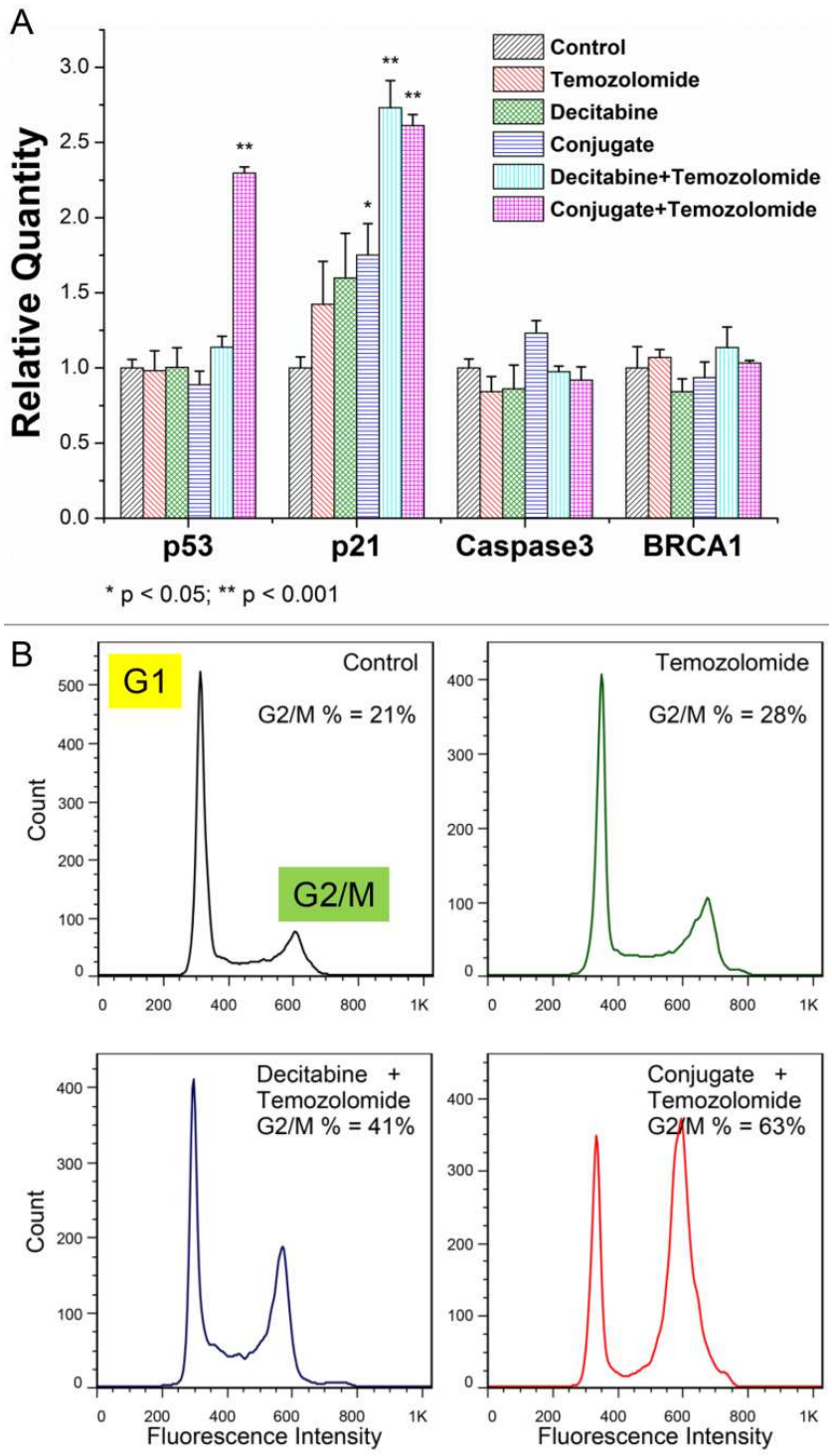


Figure 5.12 DAC-PLGA-PEG nano-conjugate better sensitizes the chemo-resistant GBM cells to temozolomide. (A) Transcriptional levels of four DNA damage-responsive genes were quantified by qRT-PCR after 72 hours of treatment (n = 4). (B) Cell cycle progression was evaluated by flow cytometry.

5.3.4 AZA-PLGA-PEG Enables Better Targeting Efficiency against Breast Cancer in Mouse Model

Similar to DAC-PLGA-PEG, AZA-PLGA-PEG showed a significantly improved cytotoxicity over free AZA in treating MCF7 breast cancer cells (Figure 5.13). As mentioned above, the primary obstacles in utilizing DNA demethylating agents to treat solid tumors are the *in vivo* instability and rapid body clearance. To assess the bioavailability of AZA-PLGA-PEG after circulation through the RES, peritoneal membrane, hepatic portal system, and renal exclusion system, subcutaneously xenografted tumors with MCF7 cells were established in mice. Eight doses of 1 mg/kg AZA-containing solution were IP injected every two weeks when the body weight of mice was closely monitored (Figure 5.14A). After four weeks' treatment, free AZA showed minimal anti-tumor effect whereas AZA-PLGA-PEG could constrain the tumor growth in a time-dependent manner (Figure 5.14B). This *in vivo* experiment clearly demonstrates the superior bioavailability of AZA-PLGA-PEG. Future study could be undertaken to optimize dosage regimens and evaluate combination therapies for other solid tumors.

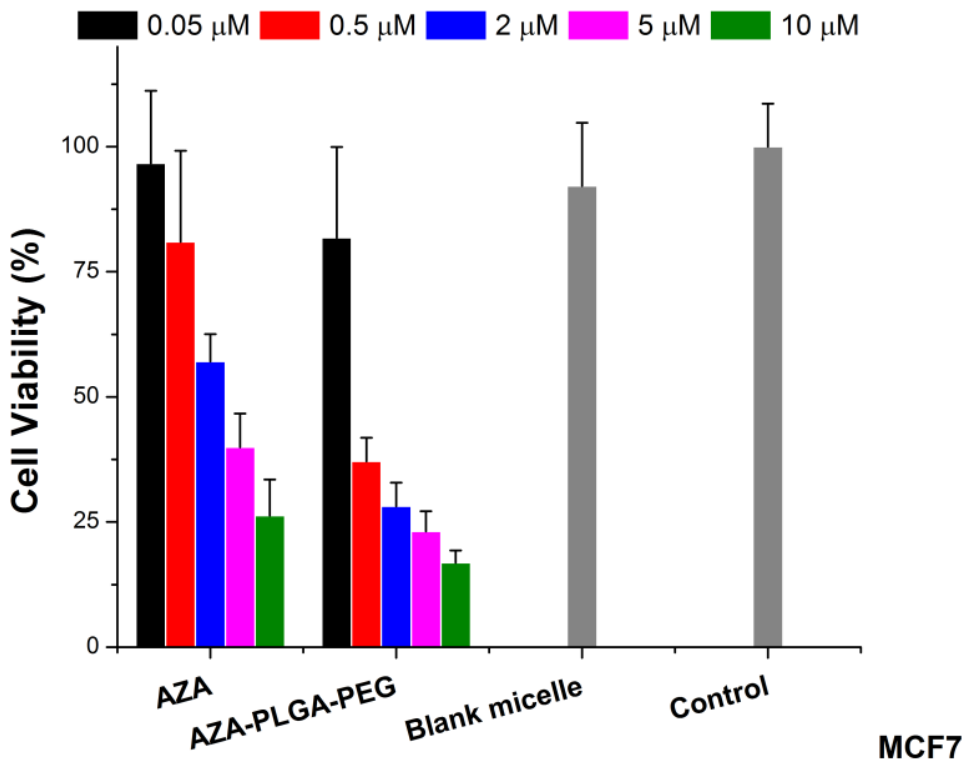


Figure 5.13 Therapeutic effects using free AZA and AZA-PLGA-PEG nano-conjugate against MCF7 cancer cells.

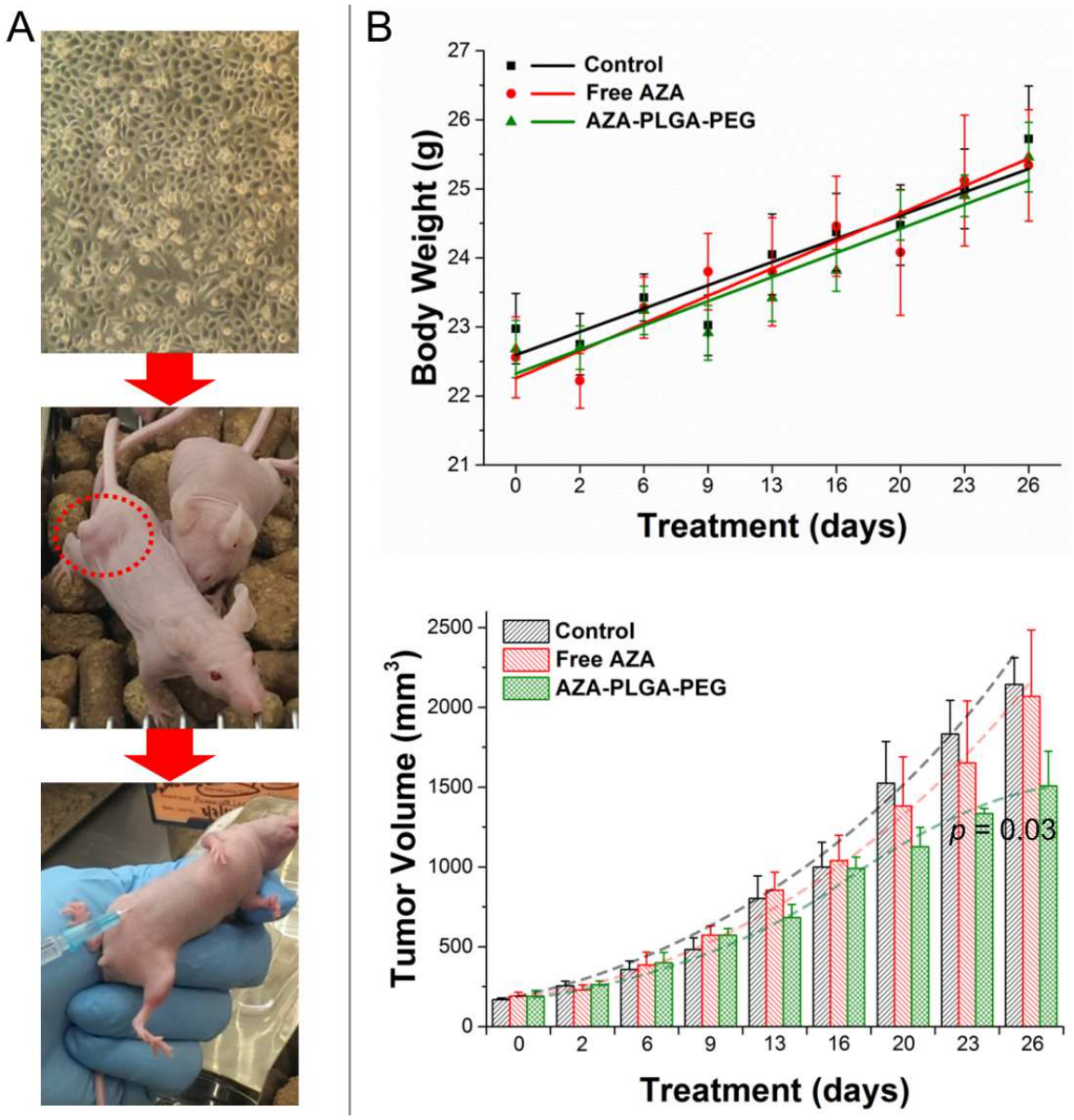


Figure 5.14 *In vivo* anti-tumor activity of AZA-PLGA-PEG with improved drug stability and bioavailability. (A) The body weight of mice was monitored to control severe side effects. (B) The volume of xenografted tumors was recorded during the four-week treatment (n = 5).

5.4 Conclusion

This study aims to address a major problem with regard to drug delivery that has a promising significance to popularize epigenetic therapies. To reach this goal, a nano-

delivery system utilizing PLGA-PEG copolymer was developed and characterized. Upon stabilizing the cytidine analogue drugs by conjugation, the formed nano-micelles feature superior drug solubility, uptake efficiency, pH-dependent release, *in vivo* bioavailability, and anti-proliferative potential. This work constitutes one of the very first to chemically link the drug molecules to the PLGA backbone instead of physical encapsulation, therefore preventing unwanted burst release. By single-molecule fluorescence experiments the rapid internalization and hydrolysis processes in an acidic cancer cell were quantitatively monitored. In the proof-of-concept, DAC-PLGA-PEG was shown to potently sensitize the chemo-resistant GBM cells to temozolomide by reactivation of the p53-p21 pathway. AZA-PLGA-PEG achieved better *in vivo* efficacy targeting breast cancer xenografts in comparison to free drug. Collectively, these results demonstrate the effectiveness of DNA demethylating drugs in treating solid tumors when an appropriate delivery strategy is designed. Hence, this work is expected to be of translational potential to enrich the toolbox of nanomedicine.

CHAPTER 6. OPTOGENETIC TOOLS FOR EPIGENOME EDITING WITH SPATIOTEMPORAL PRECISION

6.1 Introduction

From the ENCODE and Roadmap Epigenomics projects, a huge number of epigenetic marks are statistically associated with specific phenotypes and diseases.[97, 234] However, in order to decode epigenetic language to mechanistic understanding in biology, tools able to control loci-specific epigenetic states are needed. With respect to technical development, the ability to adjust epigenetic modification is also essential for drug screening, disease modeling, epigenetic therapy, and developmental research. The first wave of technical breakthrough came from the invention of agents that can globally change DNA methylation and histone acetylation. Owing to those useful tools, our knowledge of environment-gene interaction, transcriptional regulation, phenotypic diversity, and cancer initiation has been elevated to a higher level. Moreover, some of the developments such as cytidine analogues and histone deacetylase inhibitors have shown promising applications in clinical medicine. Nevertheless, a grand challenge stagnating the first generation of tools for epigenome editing is the limited spatiotemporal accuracy in that the majority of singleplex causations could hardly be researched. In the last decade, loci-specific epigenome editing has been made possible thanks to the discovery of programmable DNA-binding molecules including ZFPs, TALEs, and CRISPR/Cas9 (detailed description in

Chapter 1.2.4).[235] These advances have sparked an explosion of revolution in biotechnology and single-cell engineering. By customizing a DNA-binding motif with an epigenetic modifier, sequence-specific addition or removal of an epigenetic mark can be achieved in individual cells.[74, 75, 77, 80, 236, 237] From this moment, the methodologies for single-cell manipulation are entering an era of nanoscale precision. On the basis of current techniques, this chapter aims to improve another aspect in loci-specific epigenome editing – the temporal precision. By incorporating light-sensitive proteins into the epigenome editing machinery, the timing and extent of epigenetic modulation could be fine-tuned with a standard light source, noninvasively.

The discoveries of the light-sensitive (optogenetic) proteins in plant and microbial origin offer ample opportunities to manipulate the dynamic nature of gene expression with spatiotemporal precision. CRY2-CIB1 and PHYB-PIF6 are representative optogenetic protein pairs and their heterodimerization can be activated by blue light and red light, respectively.[238, 239] Upon illumination by 390-480 nm light, CRY2 is photoactivated to couple with CIB1 whereas the PHYB-PIF6 coupling is dependent on a 650-670 nm illumination. Further, these light-inducible associations are reversible: CRY2-CIB1 will automatically dissociate in the absence of light whilst PHYB-PIF6 needs an additional infrared illumination (~750 nm) to detach.[240, 241] The importance of optogenetic interaction was first recognized as in relation to flowering and circadian rhythms of plants.[242, 243] Recently, genetically engineered molecular constructs containing CRY2-CIB1 or PHYB-PIF6 have been used as novel means for subcellular localization and (epi)genome-editing in a variety of contexts.[81, 244-249] Yet, limited techniques are available to precisely quantify an optogenetic reaction. Herein this present study demonstrates the feasibility of dynamic

monitoring the interaction kinetics of CRY2-CIB1 and PHYB-PIF6 with single-molecule fluorescence tools. Moreover, by constructing a pair of DNMT3A-CRY2-mCherry and TRF1-CIB1-EGFP fusion proteins, this chapter provides the proof-of-concept for selective *de novo* DNA methylation at subtelomeric CpGs with tunable acquisition and temporal precision.

6.2 Methods and Materials

6.2.1 Preparation of Optogenetic Proteins

Plasmids encoding the optogenetic protein pairs were obtained from the Addgene global plasmid repository (http://www.addgene.org/): pCRY2PHR(498 aa)-mCherryN1 (#26866), pCIB1(335 aa)-pmGFP (#28240), pCIBN(the N-terminal 170 aa from CIB1)-pmGFP (#26867), pAL149 [pCMV-PHYB(1-908)-10aaLinker-mCherry-Kras4BCT (#22275)], pAL175 [pSV40-mYFP-PIF6APB (#22276)]. Plasmids were transformed into the stellar competent cells (PT-5056-2, Clontech) for amplification and selection. Extraction and purification of plasmid DNA was performed with QIAprep Spin Miniprep Kit (Qiagen). The blue and red light-induced optogenetic associations were demonstrated in HeLa cells transfected with proper protein pairs. For *in vitro* kinetics experiments, transfected cells were lysed by M-PER mammalian protein extraction reagent (Pierce), followed by acetone precipitation to remove lipid content. The concentration of fluorescent optogenetic proteins in cell lysate was determined by FCS. The phycocyanobilin (PCB) required for PHYB-PIF6 dimerization was extracted by methanolysis from *Spirulina* lysate. 5 μM PCB was added to the reaction buffer or cell culture medium when testing the PHYB-PIF6 interaction.

6.2.2 FCS for Measuring the CRY2-CIB1/N Binding Kinetics

A single blue laser-based FCS was used for measuring the *in vitro* CRY2-CIB1/N binding kinetics. One prominent advantage of this approach is that the 465 nm picosecond pulsed excitation for GFP is also the peak wavelength to induce the CRY2-CIB1/N dimerization. As the size of CRY2 is significantly larger than CIB1/N, the diffusion rate of a free CIB1/N is much faster than that of a CRY2-CIB1/N complex. Hence, the two-component 3D diffusion model can be used to characterize the free and bound forms of CIB1/N-GFP in a quantitative manner. Figure 6.1A outlines the experiment design: a 20-30 µL of the sample droplet containing equimolar concentrations of CRY2-mCherry and CIB1/N-GFP proteins is mounted to a humid microscopic chamber from which a time-course fluorescence fluctuation trace is recorded. Along the blue laser excitation, mono-dispersed CRY2-mCherry and CIB1/N-GFP are expected to couple with each other. Due to the high numeric aperture of the applied objective lens (focal length = 0.25 mm), the excitation beam waist would rapidly expand below and above the focal plane so that the majority of optogenetic proteins in the droplet can be efficiently photoactivated. Then the obtained fluorescence fluctuation trace is segmented to several time windows and an autocorrelation function is generated from each interval. By fitting with the diffusion model algorithm, the proportion of free and bound CIB1/N-GFP molecules at a specific time point can be determined (Figure 6.1B).

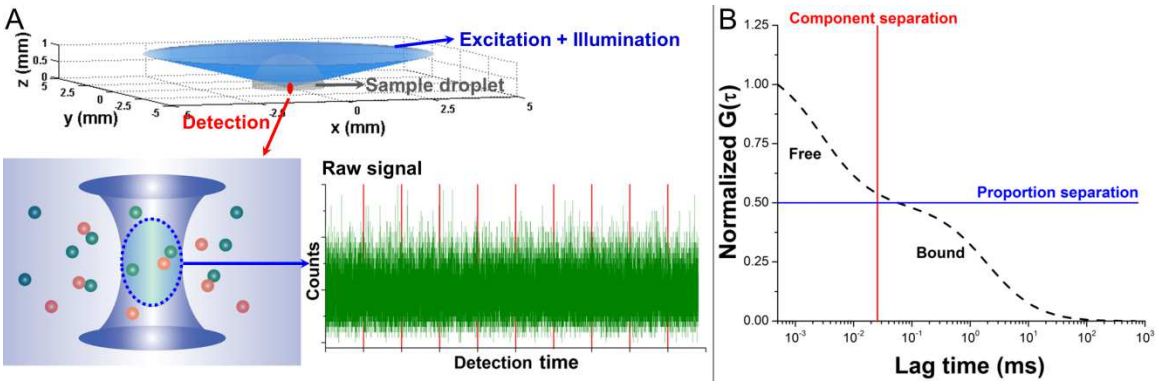


Figure 6.1 Scheme for measuring the CRY2-CIB binding kinetics by FCS. (A) The excitation profile for a sample droplet containing optogenetic proteins. A single blue laser is used to excite GFP and in the meantime to induce the optogenetic association. (B) The two-component 3D diffusion model is used to characterize the free and bound CIB molecules.

6.2.3 PIE-FLCS for Measuring the PHYB-PIF6 Binding Kinetics

In order to monitor the real-time PHYB-PIF6 binding kinetics, two illumination sources were implemented. Here pulsed interleaved excitation-based fluorescence lifetime correlation spectroscopy (PIE-FLCS) was used to reduce photodamage and better separate the single-molecule information of PIF6-YFP (Figure 6.2A). In PIE mode, the red (for dimerization induction) and blue light pulses (for YFP excitation) are intermitted by a 5 ns delay and the laser repetition rate is set to 40 MHz. Therefore, the peak energy imposed on the sample is half of that under a synchronized excitation. Moreover, the red laser-caused autofluorescence and background noises would radically decay in that 5 ns gap period, which benefits a high signal-to-noise ratio (SNR) when collecting the YFP signal (Figure 6.2B).

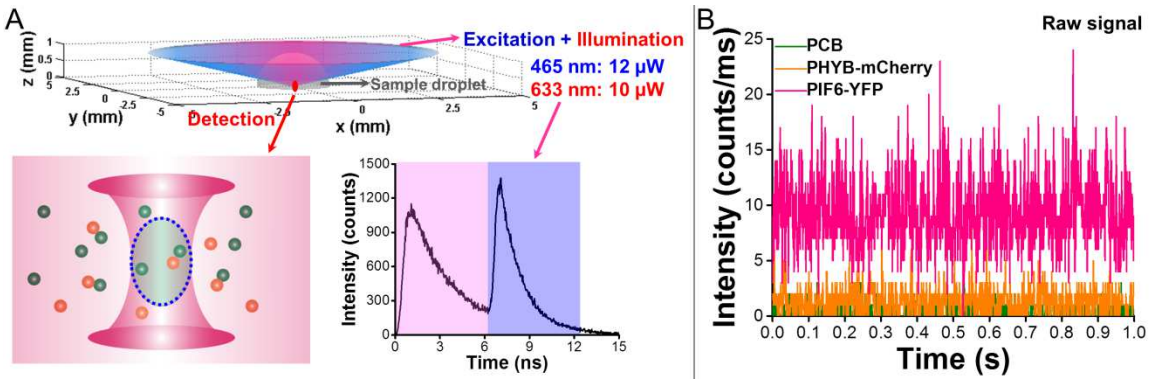


Figure 6.2 Scheme for measuring the PHYB-PIF binding kinetics by PIE-FLCS. (A) The excitation profile for a sample droplet containing optogenetic proteins. The blue laser is for fluorescence detection and the red laser is for optogenetic induction. The PIE mode intermits the laser pulses by 5 ns. (B) The SNR for PIF6-YFP is enhanced from the raw fluorescence signals.

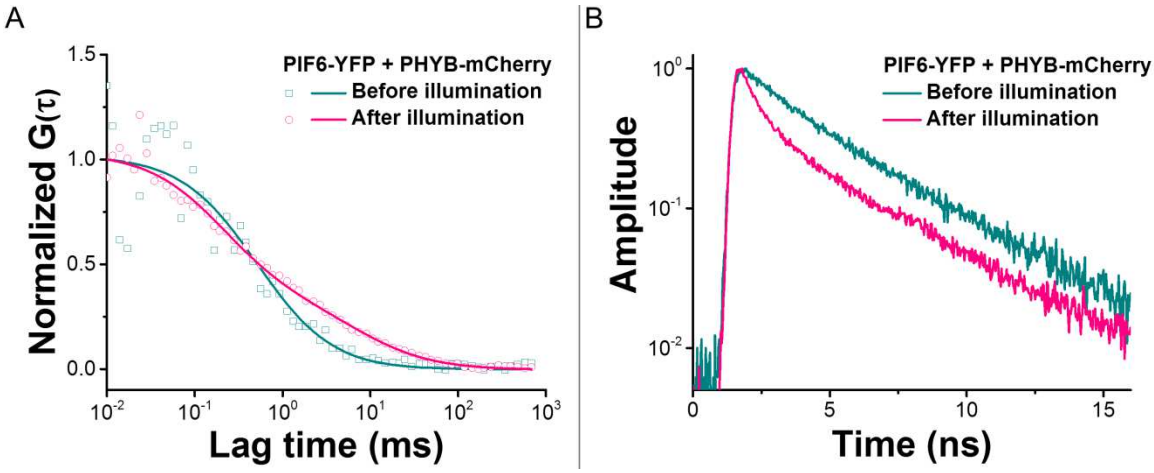


Figure 6.3 Changes in the diffusion rate and fluorescence lifetime of PIF6-YFP due to its association with PHYB-mCherry.

During the PHYB-PIF6 coupling, two events occurring to PIF6-YFP can be used to reflect the dimerization rate. First, the FCS diffusion curve of PIF6-YFP would shift from a one-component mode to a two-component mode (Figure 6.3A). Second, the TCSPC decay of PIF6-YFP would become steeper due to the FRET between YFP and mCherry

(Figure 6.3B). Combining these two sets of information, quantities of the free and bound PIF6-YFP can be precisely calculated.

6.2.4 Molecular Cloning of DNMT3A-CRY2-mCherry and TRF1-CIB1-EGFP

To demonstrate the loci-specific epigenome editing with optogenetic tools, two new fusion proteins were constructed by molecular cloning. CRY2-mCherry and CIB1-EGFP were fused to DNMT3A and TRF1, respectively. The coding sequences of DNMT3A (#35521, Addgene), CRY2-mCherry, EGFP (#23027, Addgene), CIB1, and TRF1 (#16058, Addgene) were PCR amplified from individual templates and assembled to the C-terminus of pHTN HaloTag CMV-neo Vector (Promega).

Sequence information:

>**HaloTag-DNMT3A-linker-CRY2-mCherry** (~223 kDa)
MAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNPTSSYVWRNIIPHVAPTH
RCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVIHDWGSALGFHWAKR
NPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGV
VRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPK
LLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEI
SGEPTTEDLYFQSDNAIASEFGSANTGAMPAMPSSGPGDTSSSAAEREEDRKDGEEQEE
PRGKEERQEPSTTARKVGRPGRKRKHPPVESGDTPKDPAVISKSPSMAQDSGASELLPN
GDLEKRSEPQPEEGSPAGGQKGGAPAEGEGAAETLPEASRAVENGCCTPKEGRGAPAEA
GKEQKETNIESMKMEGSRGRLRGGLGWESSLRQRPMPRLTFQAGDPYYISKRKRDEWLA
RWKREAEKKAKVIAGMNAVEENQGPGESQKVEEASPPAVQQPTDPASPTVATTPEPVGS
DAGDKNATKAGDDEPEYEDGRGFGIGELVWGKLRGFSWWPGRIVSWWMTGRSRAAEGTR
WVMWFGDGKFSVVCVEKLMPLSSFCSAFHQATYNKQPMYRKAIYEVLQVASSRAGKLFP
VCHDSDESDTAKAVEVQNKPMIEWALGGFQPSGPKGLEPPEEEKNPYKEVYTDMWVEPE
AAAYAPPPPAKKPRKSTAEKPKVKEIIDERTRERLVYEVRQKCRNIEDICISCGSLNVT
LEHPLFVGGMCQNCKNCFLECAYQYDDDGYQSYCTICCGGREVLMCGNNNCCRCFCVEC
VDLLVGPGAAQAAIKEDPWNCYMCGHKGTYGLLRRREDWPSRLQMFFANNHDQEFDPPK
VYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGK
IMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLH
DARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNL
PGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKED
ILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVGTSG
SLESGSTPAMKMDKKTIVWFRAMKMDKKTIVWFRRDLRIEDNPALAAAAHEGSVFPVFI

WCPEEEGQFYPGRASRWWMKQSLAHLSQSLKALGSDLTLIKTHNTISAILDCIRVTGAT
KVVFNHLYDPVSLVRDHTVKEKLVERGISVQSYNGDLLYEPWEIYCEKGKPFTSFNSYW
KKCLDMSIESVMLPPPWRLMPITAAAEAIWACSIEELGLENEAEKPSNALLTRAWSPGW
SNADKLLNEFIEKQLIDYAKNSKKVVGNSTSLLSPYLHFGEISVRHVFQCARMKQIIWA
RDKNSEGEESADLFLRGIGLREYSRYICFNFPFTHEQSLLSHLRFFPWDADVDKFKAWR
QGRTGYPLVDAGMRELWATGWMHNRIRVIVSSFAVKFLLLPWKWGMKYFWDTLLDADLE
CDILGWQYISGSIPDGHELDRLDNPALQGAKYDPEGEYIRQWLPELARLPTEWIHHPWD
APLTVLKASGVELGTNYAKPIVDIDTARELLAKAISRTREAQIMIGAAARDPPVATMVS
KGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFA
WDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDG
EFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDA
EVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK

>**HaloTag-EGFP-linker-CIB1-NLS-linker-TRF1** (~149 kDa)
MAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNPTSSYVWRNIIPHVAPTH
RCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVIHDWGSALGFHWAKR
NPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGV
VRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPK
LLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEI
SGEPTTEDLYFQSDNAIASEFMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDA
TYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQER
TIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMAD
KQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEK
RDHMVLLEFVTAAGITLGMDELYKALESGSTPAMNGAIGGDLLLNFPDMSVLERQRAHL
KYLNPTFDSPLAGFFADSSMITGGEMDSYLSTAGLNLPMMYGETTVEGDSRLSISPETT
LGTGNFKKRKFDTETKDCNEKKKKMTMNRDDLVEEGEEEKSKITEQNNGSTKSIKKMKH
KAKKEENNFSNDSSKVTKELEKTDYIHVRARRGQATDSHSIAERVRREKISERMKFLQD
LVPGCDKITGKAGMLDEIINYVQSLQRQIEFLSMKLAIVNPRPDFDMDDIFAKEVASTP
MTVVPSPEMVLSGYSHEMVHSGYSSEMVNSGYLHVNPMQQVNTSSDPLSCFNNGEAPSM
WDSHVQNLYGNLGVGTSVPAGSPKKKRKVEASGRDADPTEEQMAETERNDEEQFECQEL
LECQVQVGAPEEEEEEEEDAGLVAEAEAVAAGWMLDFLCLSLCRAFRDGRSEDFRRTRN
SAEAIIHGLSSLTACQLRTIYICQFLTRIAAGKTLDAQFENDERITPLESALMIWGSIE
KEHDKLHEEIQNLIKIQAIAVCMENGNFKEAEEVFERIFGDPNSHMPFKSKLLMIISQK
DTFHSFFQHFSYNHMMEKIKSYVNYVLSEKSSTFLMKAAAKVVESKRTRTITSQDKPSG
NDVEMETEANLDTRKRSHKNLFLSKLQHGTQQQDLNKKERRVGTPQSTKKKKESRRATE
SRIPVSKSQPVTPEKHRARKRQAWLWEEDKNLRSGVRKYGEGNWSKILLHYKFNNRTSV
MLKDRWRTMKKLKLISSDSED

6.2.5 Illumination for Selective Subtelomeric *de novo* DNA Methylation

In order to illuminate large populations of transfected cells for the downstream bisulfite-pyrosequencing, a mounted high-power LED with a peak wavelength of 455 nm and

a DC2100 LED driver (Thorlabs) were applied. The illumination beam size was adjusted to uniformly cover the surface area of culture vessels. After optimization, the illumination power in all subsequent experiments was determined to be 1-2 mW/cm^2.

The methylation state at the representative subtelomeric CpGs (selected from chromosomes 7q, 8q, 16p, 18p, 21q, and Xp) was quantified with bisulfite-pyrosequencing. Extracted from treated cells, 600 ng genomic DNA was bisulfite converted with EZ DNA Methylation Kit (Zymo Research) and then PCR-amplified against predetermined regions using PyroMark PCR Kit (Qiagen). The PCR products were subjected to pyrosequencing in a PyroMark Q24 system.

Table 6.1 Primers used for bisulfite-PCR. 5'-/5BiosG/ is the 5'-biotinylation tag on the primers for subsequent pyrosequencing

	Forward	Reverse	Prod.
7q	AGTGGATATTTAGGTTTATGATTGTAAT	**5'-/BiosG/**CTCCCCAAAAAACACTTTACTCTTTCACA	257
8q	TGTTTTTTTTTTTTTATTAAAAGGTGAGTT	**5'-/BiosG/**AAAAATCTCCAAAAAACCCTTATAAAC	139
16q	GTTTTAATTGGTTTTTGATTTTGATTAT	**5'-/BiosG/**CTCAACACAAACCTAAAAATCACC	164
18q	**5'-/BiosG/**GTTGTGGTGTAGGTTTATAATTT	ACCCCCAAAAAAAAAACTACAAAAACC	185
21q	TTTTGGAGGATTGTGTTTAGATAGATAT	**5'-/5BiosG/**CCCAACTCCTTCCAAATAAAT	137
Xp	GTGGGTTAGTTTGTATTGAGGTAGGTAT	**5'-/5BiosG/**AAAACCCCCTAACCCCTCC	307

Table 6.2 Information of the sequenced regions and the sequencing primers. Y/R is the potential methylated cytosine positions at the target CpGs

	Sequenced region	Sequencing primer
7q	AAATTTGTTATYGTAGTGGTYGGATTTTTTTGGGTTATTYGAGTGGAGYGG TYGGGTT	TTTATGATTGTAATTTTATAATA
8q	YGGGTAYGTAGTTAGYGTTYGYGATTATTYGTTTATAAGGGTTTTTT	ATTTTTATTAGGAAAGGTGT
16q	AYGGTTGTAATYGGGAAAGATTTTATTTATTGTTAATGYGTTTYGAGTTGT TTTAAAGTTAGGTAGTGTTTTTAAYGTTTGTGTTTAGTAGAATGTTGTTT	ATTGGTTTTTGATTTTGATTATTT
18q	AACCRCCCTCRCAACCCAAATAACCRATTTATCCTACCCRAAAAACATCRC CCTCCTCCRC	CTAAACCCCCCCTTCTACCCCTA
21q	TTYGTAGGTGTGYGYGGYGTTYGTAAGTGGTTAGTATAAYGTYGGGYGAA	GTTTAGATAGATATTGGTTTTTTTA
Xp	TYGAGGTGATTYGGGTGGGAGGTGTATYGTYGTTTATT	GAGGTAGGTATGGGT

6.3 Results and Discussion

6.3.1 Demonstration of Optogenetic Association in Living Cells

First, the CRY2-CIB1 and PHYB-PIF6 mediated subcellular redistribution of tagged proteins was demonstrated in living HeLa cells. The design principle was to localize one part of the optogenetic pair at a specific cell compartment and then attract the complementary part to the same compartment with proper illumination. As shown in Figure 6.4A, TRF1-CIB1-EGFP predominantly locates in cell nucleus due to the presence of TRF1 whereas CRY2-mCherry primarily distributes in cytoplasm. After 30 min illumination with blue light, CRY2-mCherry could be efficiently imported to nucleus because of the optogenetic CRY2-CIB1 dimerization. In another demonstration, PHYB-mCherry is pre-anchored to cell membrane by incorporating a Kras CAAX-motif to its sequence while PIF6-YFP freely diffuses inside cells. After a short red illumination, PIF6-YFP experiences radical redistribution to cell membrane, validating the PHYB-PIF6 coupling (Figure 6.4B). By comparing the YFP fluorescence in cytoplasm and on membrane, a reversion in relative intensity can be clearly seen after illumination. In consistence, the intracellular concentration of PIF6-YFP is reduced from 600 nM to 320 nM in a sampling region. Furthermore, FLIM-FRET confirms the specific PHYB-PIF6 dimerization on cell membrane as only an inter-protein distance below 10 nm can induce the non-radiative energy transfer between YFP and mCherry (Figure 6.4C). Taken together, these experiments prove the effectiveness of optogenetic proteins as powerful tools to translocate biomolecules in human cells.

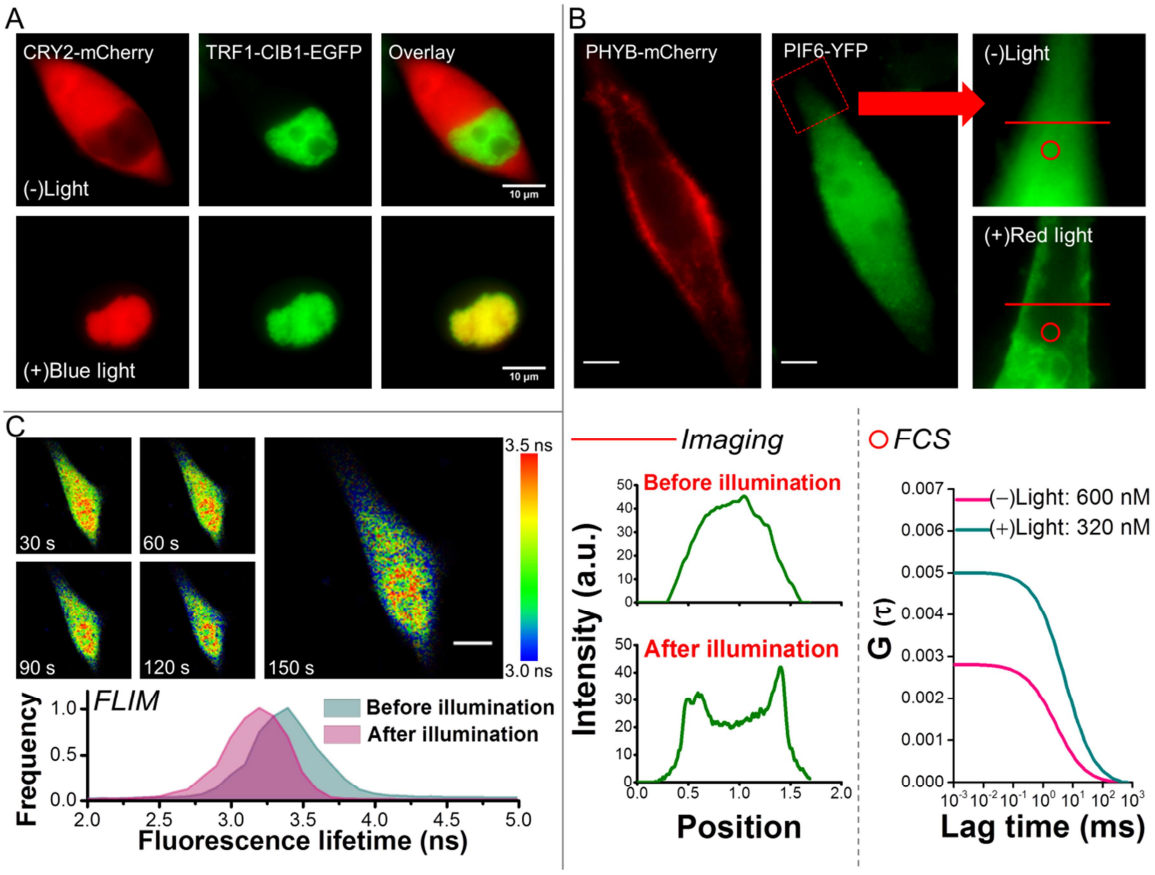


Figure 6.4 Optogenetic associations in living HeLa cells. (A) TRF1-CIB1-EGFP recruits cytoplasmic CRY2-mCherry to nucleus after blue illumination. (B) Membrane-anchored PHYB-mCherry induces redistribution of PIF6-YFP after red illumination. FCS quantitatively confirms the reduction of PIF6-YFP in cytoplasm. (C) FLIM-FRET confirms the specific association between PHYB-mCherry and PIF6-YFP on cell membrane.

6.3.2 *In vitro* CRY2-CIB1/N Binding Kinetics

Initially, fluorescence cross-correlation spectroscopy (FCCS) was proposed to be a fitted technique to probe the CRY2-CIB1/N binding kinetics since they were separately labeled with different fluorescence proteins. However, it was reported that the activity of cryptochrome could be impacted by green irradiation to some extent. A green-to-yellow excitation would suppress the equilibrium of flavin redox critical for CRY2 activation

and thus retard its optogenetic association with CIB1.[250, 251] To avoid this possible negative effect, a single blue laser-based FCS was applied.

In a 20 µL reaction droplet containing 1 nM CRY2-mCherry and CIB1/N-GFP, fluorescence fluctuation was recorded for 300 s and one autocorrelation function was generated from each 50 s time window. In Figure 6.5, a right-shift of the autocorrelation functions was observed for both combinations, indicating more bound proteins being formed along the illumination. By fitting with two-component 3D diffusion model, the free and bound protein components were determined. Compared to CIB1-GFP, CIBN-GFP possesses a faster diffusion rate (*i.e.*, a larger diffusion coefficient) due to its smaller size (Figure 6.6A): 59.77 ± 6.11 $\mu m^2/s$ versus 44.11 ± 4.13 $\mu m^2/s$ in free form, and 2.59 ± 0.62 $\mu m^2/s$ versus 2.19 ± 0.53 $\mu m^2/s$ in bound form. According to a recent study, CYR2 *per se* would be self-clustering under blue illumination in addition to dimerizing with CIB1,[252, 253] which explains why the observed changes in diffusion coefficient from free to bound CIB1/N do not support a 1:1 binding stoichiometry with CRY2. However, the fast mobility of CIBN was found not to increase its coupling efficiency with CRY2 *in vitro* (Figure 6.6B). During the 300 s detection window, CIB1 yielded about 15% more bound component than CIBN. This phenomenon implies that a less active diffusion is conducive to stabilize the CRY2-CIB1 dimerization and the C-terminal 165 amino acids in CIB1 may be required to facilitate the association kinetics. Furthermore, the excitation power-dependent optogenetic association was observed (Figure 6.6B, inset). Overall, this study provides the first quantitative evidence showing that CIB1 is better than CIBN as the optogenetic complement to CRY2.

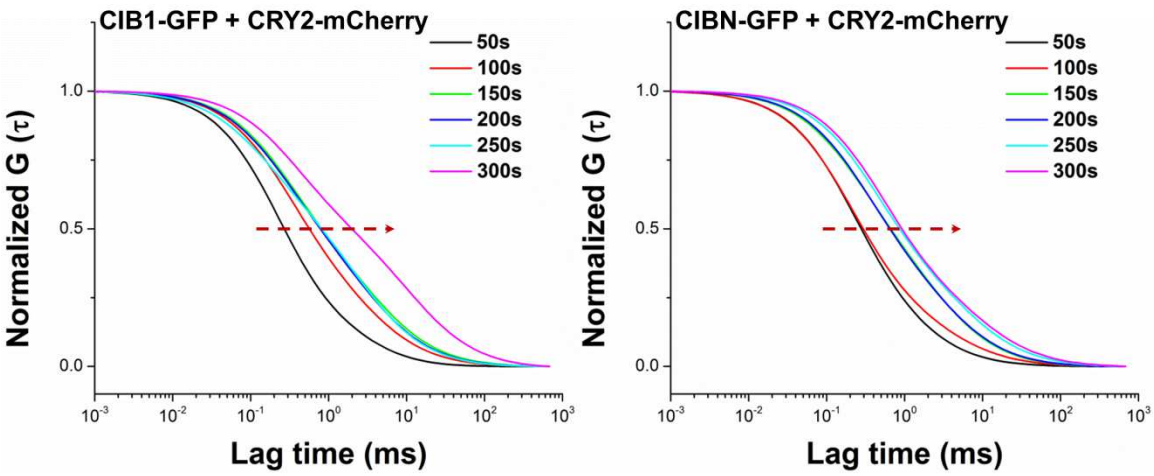


Figure 6.5 *In vitro* association kinetics of CIB1-CRY2 and CIBN-CRY2 probed by FCS.

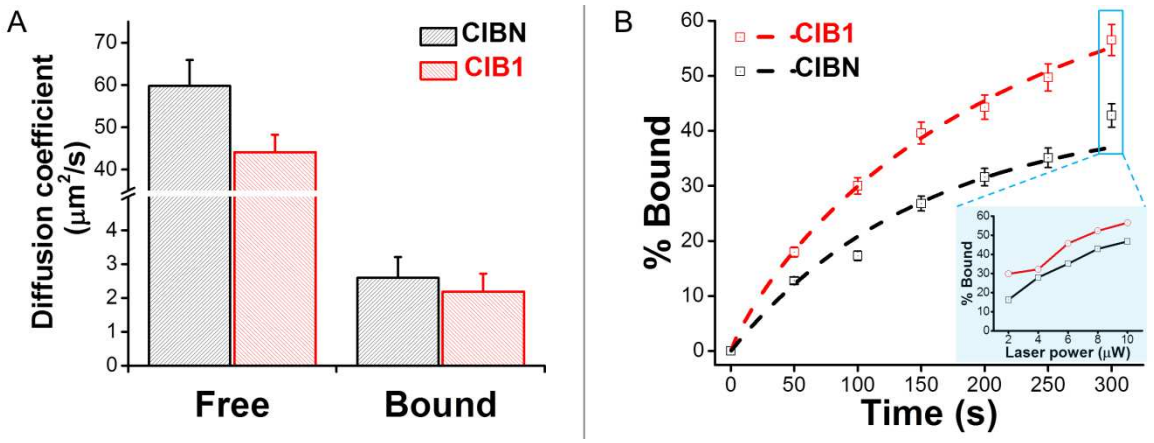


Figure 6.6 Characterization of the CRY2-CIB association rate. (A) Diffusion coefficients of free and bound CIB1/N-GFP (n = 9, mean + S.E.M.). (B) The incremental percentage of bound CIB1/N-GFP in a 300 s detection window (n = 5).

6.3.3 *In vitro* PHYB-PIF6 Binding Kinetics

As perceived from the preliminary cell experiments, the PHYB-PIF6 dimerization occurs much faster than CRY2-CIB1. Therefore, in this part, fluorescence lifetime was included in analysis to better characterize the free and bound PIF6-YFP. A 50 μL reaction droplet containing 100 nM PHYB-mCherry and PIF6-YFP was mounted for PIE il-

lumination. Two methods were used to determine the percentage of free/bound PIF6-YFP at a specific time point. In the TCSPC method, the bound form of PIF6-YFP is distinguished from the free form based on the occurrence of FRET. According to the law of fluorescence decay: $I(t) = I_0 \sum_i a_i \exp\left(-\frac{t}{\tau_i}\right) + c$, the sum TCSPC histogram can be decomposed to three components: the bound (FRET) component, the free (non-FRET) component, and the background (Figure 6.7A). By integrating the photon counts under each component, the percentage of bound proteins can be calculated at a resolution of 1 s (Figure 6.7B). In the second method, more sensitive FLCS is used to precisely determine the concentration of free/bound PIF6-YFP. The differential decay patterns of FRET and non-FRET PIF6-YFP make it possible to generate a TCSPC filter function to weigh the possibility of a detected single photon from which component (Figure 6.8A). With this filter function, a fluorescence lifetime-dependent autocorrelation function can be extracted by FLCS. Hence, featuring a steady fluorescence lifetime, photons from non-FRET PIF6-YFP are separated. Autocorrelation functions are established from each 5 s time window and the concentration information of free PIF6-YFP is obtained (Figure 6.8B). These two methods have their own advantages and in the meantime complement each other: the TCSPC decomposition provides a delicate temporal resolution and FLCS bring about an excellent accuracy, which together uncover the dynamic PHYB-PIF6 interaction.

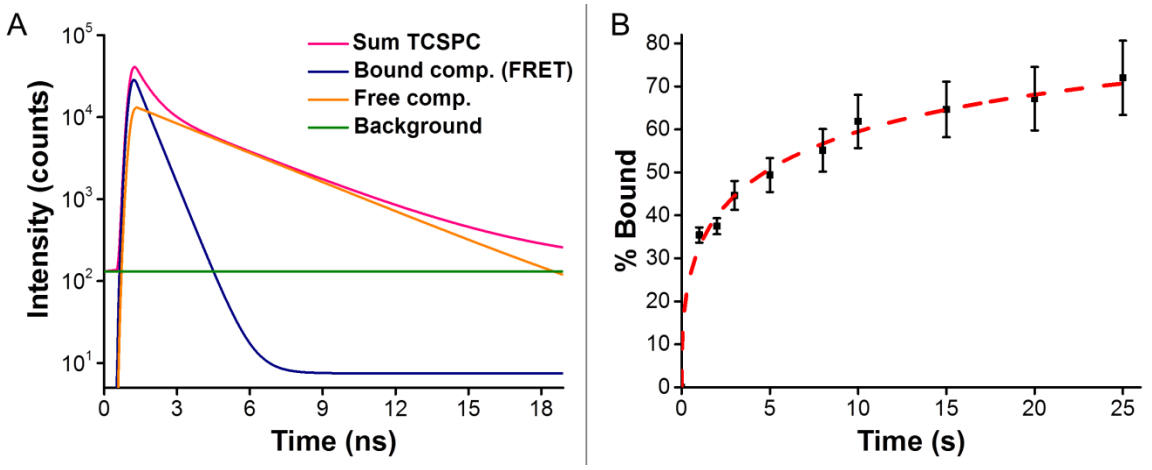


Figure 6.7 TCSPC method for measuring the PHYB-PIF6 binding kinetics. (A) Mathematical decomposition of the sum TCSPC. (B) The association rate of PHYB-PIF6 is calculated by reassigning the photon counts to each lifetime component (n = 5).

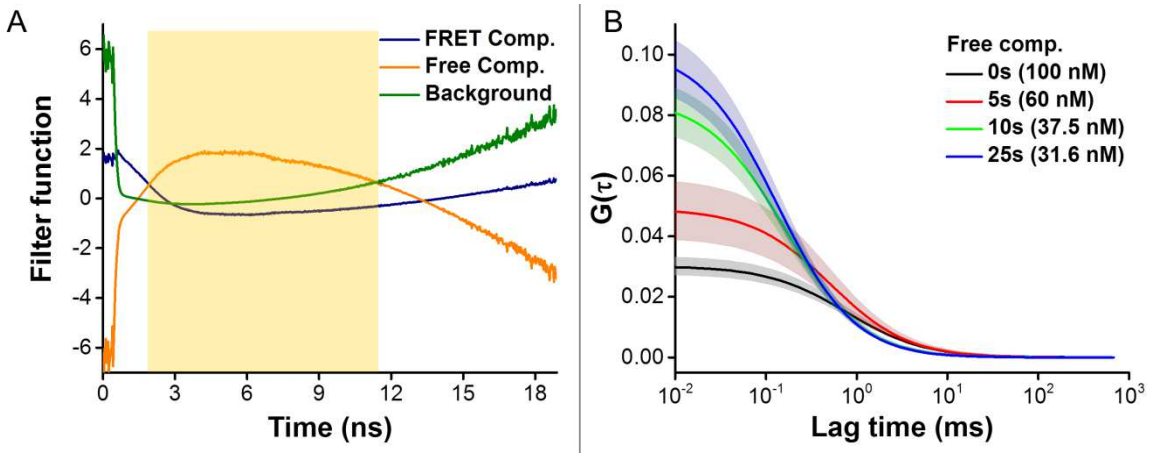


Figure 6.8 FLCS method for determining the concentration of free PIF6-YFP. (A) Based on differential fluorescence lifetimes a filter function is generated to extract photons from the free component. (B) The concentration of free PIF6-YFP is determined by fitting the FLCS autocorrelation functions at each time point.

When comparing the CRY2-CIB1 and PHYB-PIF6 binding kinetics, PHYB-PIF6 exhibits a much more efficient association. For 50% of proteins to get optogenetically coupled, CRY2-CIB1 takes ~250 s whereas PHYB-PIF6 only needs ~5 s, which awaits future research to investigate the underlying mechanisms. One possible explanation could

be related to their functions in plants. Plants heavily rely on these two classes of optogenetic proteins to sense the blue and red wavelengths from sunlight. Considering the fact that blue light has a higher energy than red light, optogenetic proteins responding to the red wavelength might have evolved to gain a better photosensitivity. However, the additional demand for PCB impedes an extensive utilization of the PHYB-PIF6 pair in biological engineering.

6.3.4 Optogenetic Induction of Subtelomeric *de novo* DNA Methylation

Next, two constructs expressing TRF1-CIB1-EGFP and DNMT3A-CRY2-mCherry were introduced to HeLa cells for demonstration of a light-inducible, loci-specific epigenome editing. TRF1-CIB1-EGFP serves as the targeting component to localize the enzymatic reaction of DNMT3A-CRY2-mCherry at subtelomeric regions (Figure 6.9A). Upon transient expression, TRF1-CIB1-EGFP shows a punctate distribution in cell nucleus, highlighting the telomere positions (Figure 6.9B). From FCS measurements, two distinct components were detected: a fast-diffusing free form (D = 5.96 ± 0.58 $\mu m^2/s$) and a slow-diffusing bound form (D = 0.92 ± 0.08 $\mu m^2/s$). A remarkably high percentage of bound form of TRF1-CIB1-EGFP was noted inside those punctate spots (83 ± 17%). In contrast, Nuc-GFP molecules predominantly exist as a free-diffusing form in whole nucleus (D = 11.23 ± 2.06 $\mu m^2/s$). To further confirm the binding specificity of TRF1-CIB1-EGFP to telomeric regions, a co-immunofluorescence staining for endogenous TRF2 proteins was performed and the co-localization between TRF1-CIB1-EGFP and TRF2-Alexa647 was unambiguously observed (Figure 6.9C). These results support that the constructed TRF1-

CIB1-EGFP successfully recognizes the characteristic TTAGGG repeats in telomeric and subtelomeric regions.[254]

Then the blue light-induced association of TRF1-CIB1-EGFP and DNMT3A-CRY2-mCherry was assessed by FLIM-FRET. In control cells, the average fluorescence time of EGFP was determined to be 2.41 ns (Figure 6.10A). On the other hand, the average fluorescence lifetime of EGFP in co-transfected cells was continuously decreased under blue illumination. A 15% reduction in the EGFP fluorescence lifetime was achieved in 5 min at an energy level of 8 mW/cm^2 (Figure 6.10B). This significantly shortened fluorescence lifetime is attributed to the FRET effect between EGFP and mCherry in the optogenetic pair. In the following optimization, the association rate was found to be tunable by the light intensity, for instance, an energy level of 1 mW/cm^2 giving rise to a similar FRET response in about 10 min. These data were validated with Co-IP experiments using EGFP antibody and western blot using DNMT3A antibody. From the illuminated cell lysate, DNMT3A-CRY2-mCherry could be co-immunoprecipitated by EGFP antibody (data not shown). Collectively, this set of experiments proves the formation of TRF1-CIB1-EGFP and DNMT3A-CRY2-mCherry complex at the desired loci after blue light illumination.

After the blue light-mediated re-localization of DNMT3A-CRY2-mCherry to chromosomal ends, the *de novo* methylase activity of DNMT3A is expected to increase the DNA methylation levels at subtelomeric CpGs since no CpGs exist in telomeric regions (Figure 6.11A). Six subtelomeric regions containing a number of CpGs were selected for bisulfite-pyrosequencing after varied illumination durations. The pyrosequencing results corroborate the blue light-induced increase in methylation at most of the CpGs. In general, a longer light exposure results in more *de novo* methylation (Figure 6.11B). For example,

with 4 h illumination the following changes occur: 12% increase at the 2nd CpG site of Chr. 7, 7% increase at the 1st CpG site of Chr. 8q, 9% increase at the 1st CpG site of Chr. 16p, 19% increase at the 2nd CpG site of Chr. 18p, 5% increase at the 8th CpG site of Chr. 21q, and 12% increase at the 3rd CpG site of Chr. Xp. This substantial *de novo* methylation fully demonstrates the efficacy of optogenetic tools even in a short time. The differential increases for each CpG position could be ascribed to their different starting methylation levels and varying distance to the nearest TTAGGG repeat.[255, 256] Admittedly, some slight non-specific *de novo* methylation was noted (*i.e.,* an increase in methylation with 0 h illumination) due to the catalytic action of DNMT3A. However, in a selected centromeric control region, no additional increase in methylation was observed upon cell exposure to blue light. Hence, this line of research demonstrates the feasibility of loci-specific epigenome editing with spatiotemporal precision enabled by optogenetic tools.

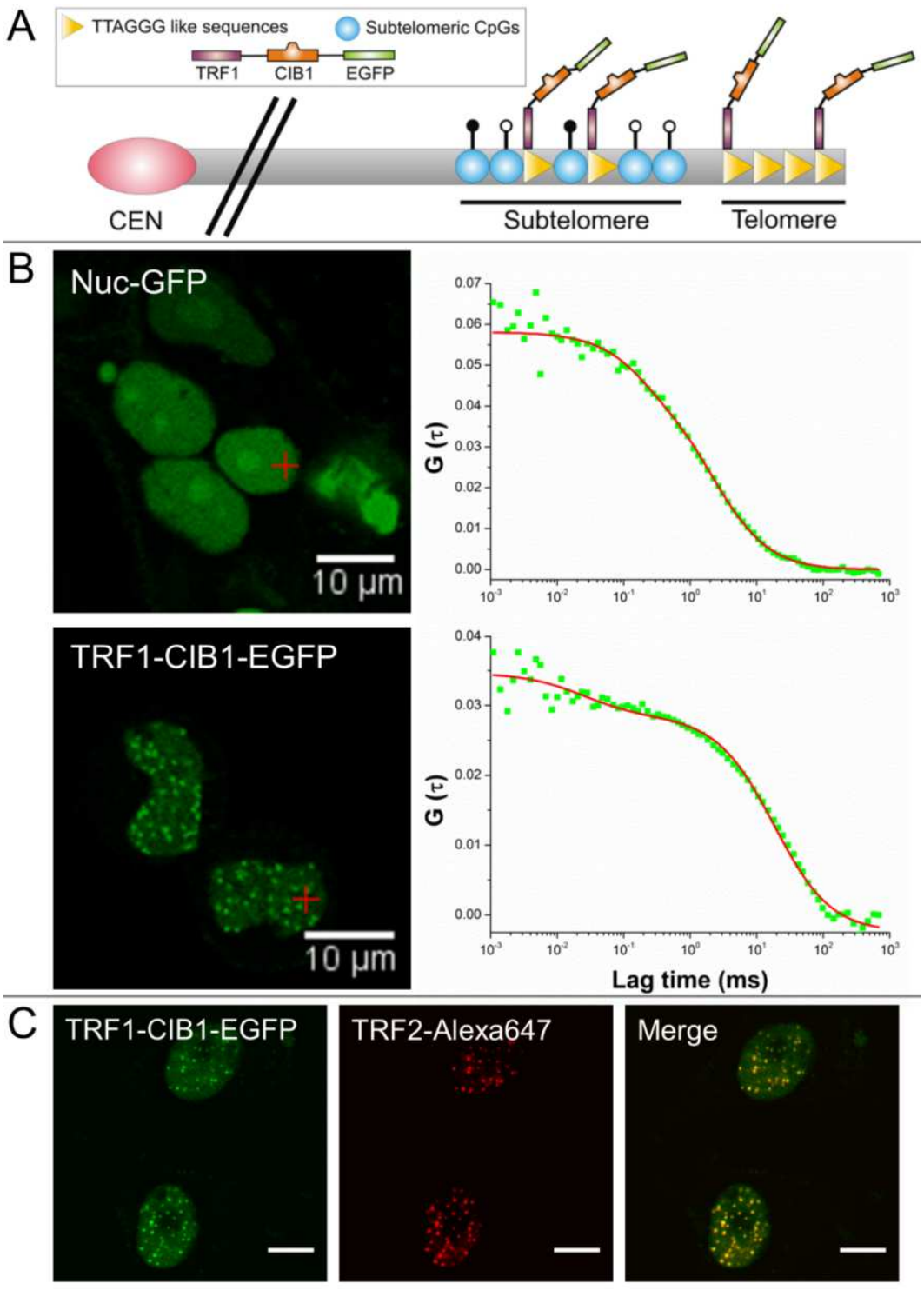


Figure 6.9 Localization of TRF1-CIB1-EGFP to telomeric and subtelomeric regions. (A) Design of the fusion protein. (B) The distribution and diffusion characteristics of TRF1-CIB1-EGFP in nucleus. (C) Co-localization between TRF1-CIB1-EGFP and endogenous TRF2.

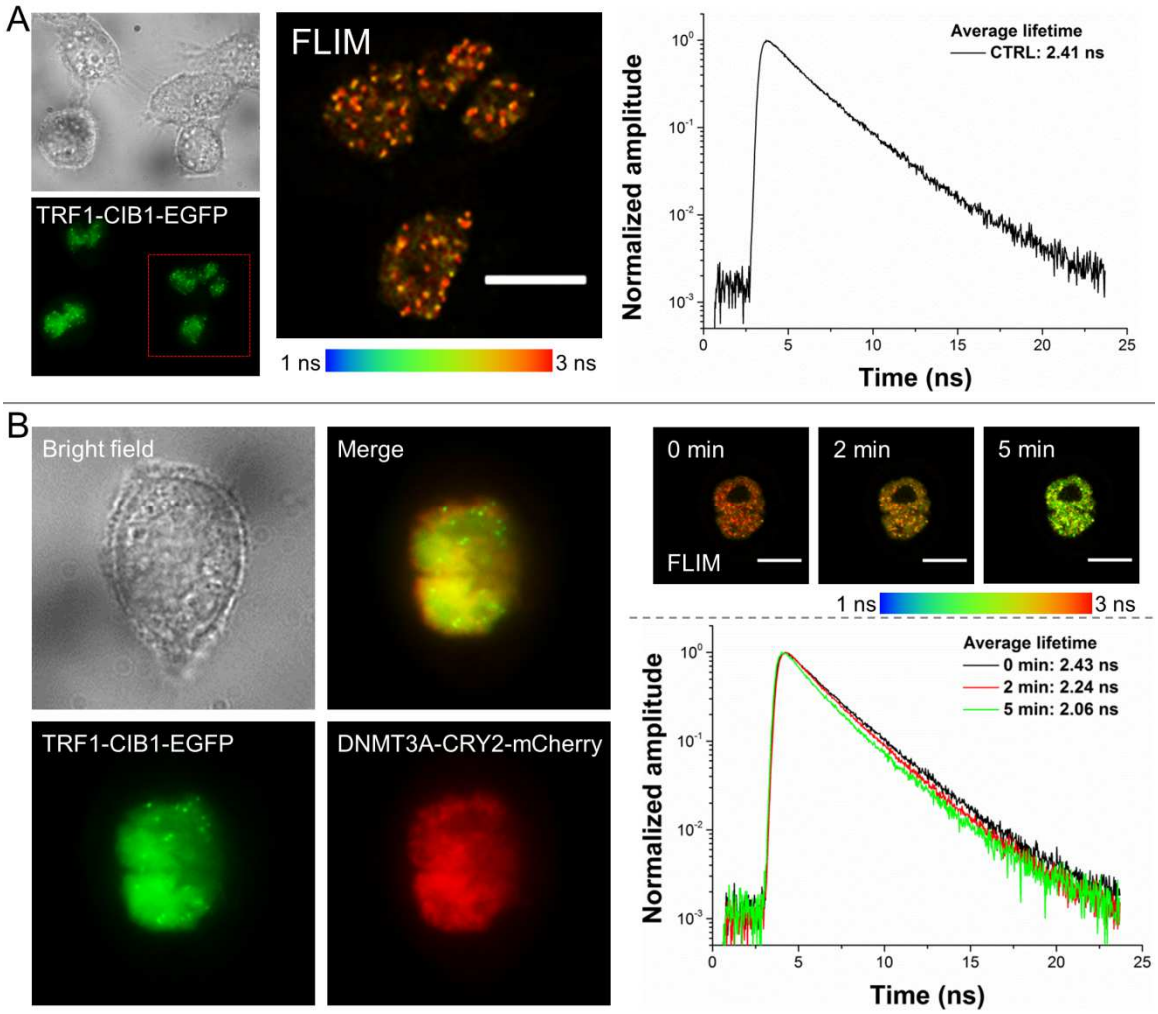


Figure 6.10 Blue light-induced re-localization of DNMT3A-CRY2-mCherry to telomeric regions. (A) FLIM measurement for HeLa cells transfected with TRF1-CIB1-EGFP. (B) FLIM measurement for cells co-transfected with TRF1-CIB1-EGFP and DNMT3A-CRY2-mCherry under blue illumination.

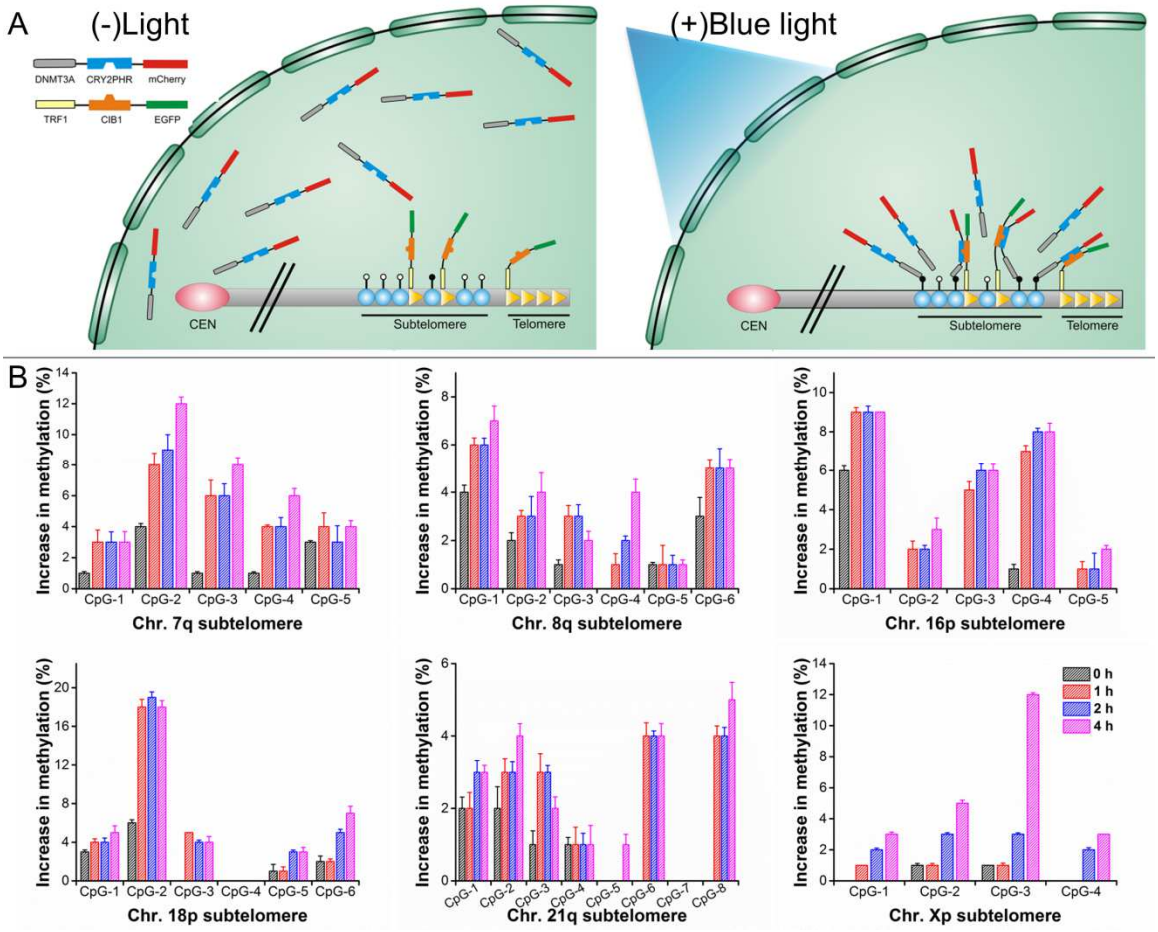


Figure 6.11 Blue light-induced *de novo* DNA methylation at the subtelomeric CpGs. (A) Scheme of the working mechanism for optogenetic induction. (B) Analysis of changes in DNA methylation at the selected CpG positions of six chromosomes.

6.4 Conclusion

Optogenetic protein pairs are emerging as a group of powerful tools in single-cell bioengineering. However, for a better implementation of these photosensitive molecules, quantitative information regarding their associations is needed. In this chapter for the first time the real-time binding kinetics of CRY2-CIB and PHYB-PIF, the major optogenetic pairs responding to blue and red lights, is adequately characterized. By adapting single-molecule fluorescence techniques including FCS, FLIM-FRET, TCSPC, and PIE-FLCS,

optogenetic associations can be quantified with a subsecond resolution. Furthermore, this study contributes to the pioneering efforts aimed to realize an on-demand, loci-specific epigenome editing. By constructing DNA-targeting and epigenetic-effecting proteins that can be photoactivated to function in predetermined sites, *de novo* DNA methylation was successfully achieved at subtelomeric CpGs. Hence this work provides a novel, noninvasive method to manipulate epigenetic state and offers a unique avenue for epigenetic research. In addition, the diversity of the available optogenetic tools makes it possible to multiplexing epigenome editing. Last but not least, *in vivo* application of optogenetic tools is anticipated if advanced illumination facilities such as endoscopic multi-photon excitation are integrated.

CHAPTER 7. CONCLUDING REMARKS AND FUTURE PERSPECTIVES

During the historical transition from the superstitious astrology and alchemy to the modern physics and chemistry, quantitative detection and utilization of mechanical forces, electricity, light, chemical structures, and reaction kinetics have resulted in revolutionary advances. Today, in order to decipher the epigenetic code in a single cell at the single-molecule level, nanoscale measurements and engineering of various biomolecules and activities hold the key. In conventional molecular biology, methods like electrophoresis, PCR, and immunoassays heavily rely on ensemble results estimated from a population of cells, thus overlooking the cellular heterogeneity and dynamic intermediates. In light of advanced optical tools and nanotechnology, scrutinizing the epigenetic network with high spatiotemporal resolutions is made possible and provides us with new dimensions to interpret life and diseases.

Metazoan species, especially human being, contains billions of variedly differentiated cells playing the fundamental roles in every biological process. For cells within the same tissue or organ, at a given time point they might have distinct morphological traits and functional states that are substantially dictated by epigenetic modulations. Single-cell and single-molecule techniques offer us tremendous opportunities to inspect subcellular components with unparalleled sensitivity and specificity. Advanced optics and nanoscale engineering are particularly advantageous in characterizing the epigenetic conservation

and variations implicated in dynamic biological systems. Moreover, the fast evolving and enriching nanobiotechnology will not only benefit the basic research but also hold great promise in healthcare practices.

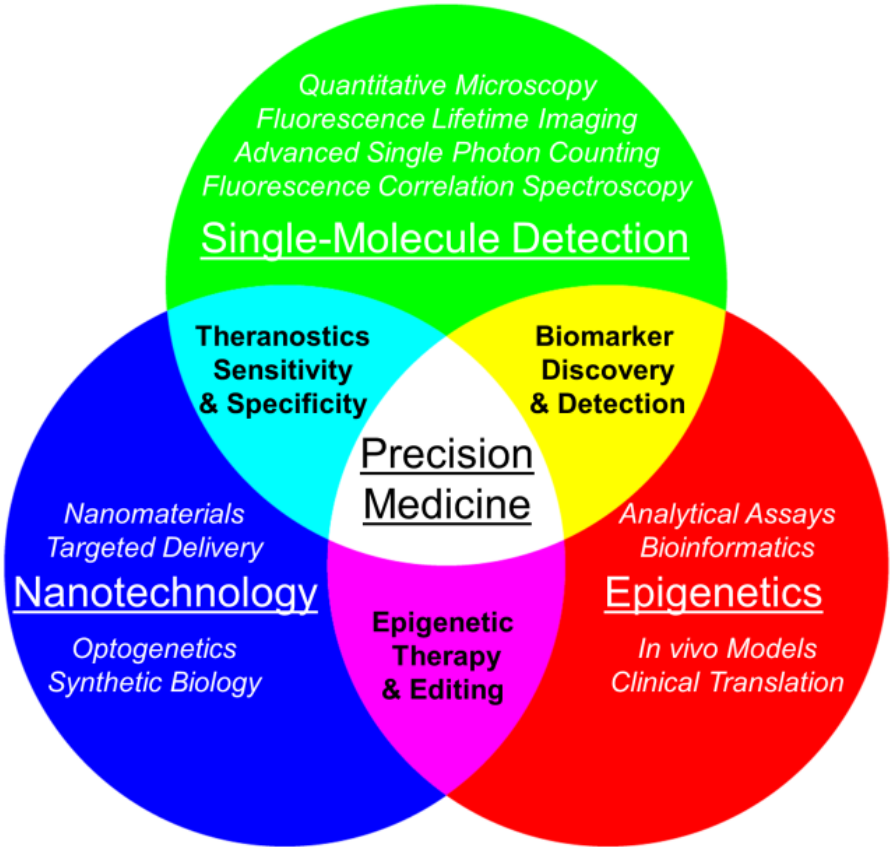


Figure 7.1 Content composition of this dissertation.

My dissertation work aims to dissect the DNA methylation-related molecules and events at the nanoscale using single-molecule detection techniques and nano-engineering approaches (Figure 7.1). My research theme can be summarized to two major aspects: characterization and manipulation. With regard to analytical characterization, my study takes a unique route to quantitatively examine DNA methylation, *i.e.*, via monitoring the

DNA methylation-associated proteins. In Chapter 2, for the first time the dynamic process of active DNA demethylation is successfully correlated with the millisecond-level mobility of MBD3, providing a novel strategy to inspect the global hypomethylation in cancer. In Chapter 3, single-molecule fluorescence tools are further employed to explore the binding preference, real-time association, and stoichiometric transition of MBD3 in a single cell cycle, which constitutes a pioneering effort to analyze the intricacy of DNA methylation homeostasis in single living cells. The results of my study reveal a crucial role of MBD3 in the inheritance of DNA methylation. Inspired by these new findings, in Chapter 4 I have conducted a translational study to delineate the regulatory network by MBD3 in human malignant glioma and established it as a possible prognostic biomarker. With exciting discoveries accumulated, I am fully convinced that single-cell and single-molecule techniques unleash the power for us to better assess epigenetic mechanisms in a micron-sized human cell, therefore complementing the future wave of precision medicine. The other focus of my dissertation lies in controlling DNA methylation with nanoscale accuracy and tunability. In Chapter 5, a novel PLGA-PEG-based di-block polymeric nano-vector is developed to deliver the DNA demethylation drugs, and this formulation features significant improvements on the aqueous stability, *in vivo* bioavailability, cell uptake efficiency, pH-dependent release, and anti-proliferative potential. Other than the global reshaping with DNA demethylation agents, in Chapter 6 the loci-specific epigenome editing using optogenetic proteins is demonstrated. Compared to the first generation of epigenome-editing methodologies, the tools developed herein add another layer of modulation: *i.e.,* the epigenetic status of a specific region can be adjusted on demand with high temporal controllability. Taken together, this dissertation delivers a comprehensive

toolbox for epigenetic research involving DNA methylation and unearths a number of intracellular phenomena inaccessible by traditional means.

Admittedly, a variety of properties with respect to single-cell probing and single-molecule engineering need further optimization. First, to achieve high-throughput readout from most single-molecule techniques is a grand challenge since the trade-off between the detection speed and precision limits the number of assessed targets at a time. Incorporation of the single-molecule analysis to the workflow of flow cytometry could be a possible solution. Second, so far even though numerous single-cell datasets have been obtained, a large proportion of them were from *in vitro* or *ex vivo* conditions. It is more imperative to interpret the behaviors of a single cell within its *in vivo* context. More efforts for whole-body implementation of the newly developed tools such as fluorescence fluctuation spectroscopy, fluorescence lifetime imaging, single-photon counting, optogenetics, and epigenome editing remain to be invested. Third, fully nontoxic and biocompatible nanomaterials are still scarce at present. Therefore, utilization of synthetic biomaterials including peptide nanosensors, xeno-nucleic acids, and DNA origami might be the future options. Last but not least, to correlate the microscale genotype with the macroscale phenotype in epigenetics requires a combinatorial elucidation of the central dogma in biology. With the advent of single-cell sequencing techniques, the hope of obtaining quantitative information of DNA and RNA from a trace amount of cells is becoming a reality. Hence, further research in understanding the epigenetic intricacy and its regulatory impact in a loci-specific manner can be expected.

The biggest challenge begets the best work. Single-cell and single-molecule studies are not commonplace in biomedicine yet but only a fraction of the possibilities, if any

have been explored. Opportunities abound in enhancing our knowledge of embryogenesis, development, aging, metabolic disorders, degenerative diseases, and cancer – the solution to all our unsolved questions exclusively hinges upon the extent we know about and manipulate the function of a single cell, in which millions of actions occur in the timescale ranging from nanoseconds to hours.

LIST OF REFERENCES

LIST OF REFERENCES

1. C. H. Waddington, *Int J Epidemiol*, 2012, **41**, 10-13.
2. T. B. Johnson and R. D. Coghill, *J Am Chem Soc*, 1925, **47**, 2838-2844.
3. A. D. Riggs, *Cytogenet Cell Genet*, 1975, **14**, 9-25.
4. P. A. Jones, *Nat Rev Genet*, 2012, **13**, 484-492.
5. Y. Zhang, H. H. Ng, H. Erdjument-Bromage, et al., *Genes Dev*, 1999, **13**, 1924-1935.
6. N. R. Rose and R. J. Klose, *Biochim Biophys Acta*, 2014, **1839**, 1362-1372.
7. M. W. Kellinger, C. X. Song, J. Chong, et al., *Nat Struct Mol Biol*, 2012, **19**, 831-833.
8. G. Lev Maor, A. Yearim and G. Ast, *Trends Genet*, 2015, **31**, 274-280.
9. M. G. Butler, *J Assist Reprod Genet*, 2009, **26**, 477-486.
10. M. Ehrlich and M. Lacey, *Adv Exp Med Biol*, 2013, **754**, 31-56.
11. Manel Esteller, et al., *Epigenetics in Biology and Medicine*, CRC Press, 2008.
12. M. G. Goll, F. Kirpekar, K. A. Maggert, et al., *Science*, 2006, **311**, 395-398.
13. M. G. Goll and T. H. Bestor, *Annu Rev Biochem*, 2005, **74**, 481-514.
14. T. Bestor, A. Laudano, R. Mattaliano and V. Ingram, *J Mol Biol*, 1988, **203**, 971-983.
15. M. Okano, S. Xie and E. Li, *Nat Genet*, 1998, **19**, 219-220.
16. H. Leonhardt, A. W. Page, H. U. Weier and T. H. Bestor, *Cell*, 1992, **71**, 865-873.
17. Y. Kato, M. Kaneda, K. Hata, et al., *Hum Mol Genet*, 2007, **16**, 2272-2280.
18. P. A. Jones and G. Liang, *Nat Rev Genet*, 2009, **10**, 805-811.
19. K. D. Robertson, *Oncogene*, 2001, **20**, 3139-3155.
20. T. Chen, Y. Ueda, J. E. Dodge, et al., *Mol Cell Biol*, 2003, **23**, 5594-5605.
21. M. Tahiliani, K. P. Koh, Y. Shen, et al., *Science*, 2009, **324**, 930-935.
22. R. M. Kohli and Y. Zhang, *Nature*, 2013, **502**, 472-479.
23. H. Wu and Y. Zhang, *Genes Dev*, 2011, **25**, 2436-2452.
24. S. R. Kinney and S. Pradhan, *Adv Exp Med Biol*, 2013, **754**, 57-79.
25. R. B. Lorsbach, J. Moore, S. Mathew, et al., *Leukemia*, 2003, **17**, 637-641.
26. T. J. Ley, L. Ding, M. J. Walter, et al., *N Engl J Med*, 2010, **363**, 2424-2433.
27. M. J. Walter, L. Ding, D. Shen, et al., *Leukemia*, 2011, **25**, 1153-1158.
28. R. S. Hansen, C. Wijmenga, P. Luo, et al., *Proc Natl Acad Sci U S A*, 1999, **96**, 14412-14417.
29. H. G. Yoon, D. W. Chan, A. B. Reynolds, et al., *Mol Cell*, 2003, **12**, 723-734.
30. G. J. Filion, S. Zhenilo, S. Salozhin, et al., *Mol Cell Biol*, 2006, **26**, 169-181.
31. R. R. Meehan, J. D. Lewis and A. P. Bird, *Nucleic Acids Res*, 1992, **20**, 5085-5092.

32. I. Ohki, N. Shimotake, N. Fujita, et al., *Cell*, 2001, **105**, 487-497.
33. M. Saito and F. Ishikawa, *J Biol Chem*, 2002, **277**, 35434-35439.
34. X. Zou, W. Ma, I. A. Solov'yov, et al., *Nucleic Acids Res*, 2012, **40**, 2747-2758.
35. H. Hashimoto, Y. Liu, A. K. Upadhyay, et al., *Nucleic Acids Res*, 2012, **40**, 4841-4849.
36. T. Uemura, E. Kubo, Y. Kanari, et al., *Cell Struct Funct*, 2000, **25**, 149-159.
37. M. Bostick, J. K. Kim, P. O. Esteve, et al., *Science*, 2007, **317**, 1760-1764.
38. J. Sharif, M. Muto, S. Takebayashi, et al., *Nature*, 2007, **450**, 908-912.
39. Q. Du, P. L. Luu, C. Stirzaker and S. J. Clark, *Epigenomics*, 2015, **7**, 1051-1073.
40. S. K. Bhattacharya, S. Ramchandani, N. Cervoni and M. Szyf, *Nature*, 1999, **397**, 579-583.
41. B. Hendrich, U. Hardeland, H. H. Ng, et al., *Nature*, 1999, **401**, 301-304.
42. S. E. Brown, M. J. Suderman, M. Hallett and M. Szyf, *Gene*, 2008, **420**, 99-106.
43. R. E. Amir, I. B. Van den Veyver, M. Wan, et al., *Nat Genet*, 1999, **23**, 185-188.
44. B. Hendrich, J. Guy, B. Ramsahoye, et al., *Genes Dev*, 2001, **15**, 710-723.
45. H. M. Muller, H. Fiegl, G. Goebel, et al., *Br J Cancer*, 2003, **89**, 1934-1939.
46. J. Schlegel, S. Guneysu and H. D. Mennel, *Oncol Rep*, 2002, **9**, 393-395.
47. D. Zhu, S. B. Hunter, P. M. Vertino and E. G. Van Meir, *Cancer Res*, 2011, **71**, 5859-5870.
48. S. K. Patra, A. Patra, H. Zhao, et al., *Biochem Biophys Res Commun*, 2003, **302**, 759-766.
49. T. B. Pontes, E. S. Chen, C. O. Gigek, et al., *Tumour Biol*, 2014, **35**, 3447-3453.
50. S. Zhao, M. Choi, J. D. Overton, et al., *Proc Natl Acad Sci U S A*, 2013, **110**, 2916-2921.
51. D. Magde, W. W. Webb and E. Elson, *Phys Rev Lett*, 1972, **29**, 705-&.
52. E. Haustein and P. Schwille, *Annu Rev Biophys Biomol Struct*, 2007, **36**, 151-169.
53. U. Meseth, T. Wohland, R. Rigler and H. Vogel, *Biophys J*, 1999, **76**, 1619-1631.
54. Y. Chen, J. D. Muller, P. T. So and E. Gratton, *Biophys J*, 1999, **77**, 553-567.
55. Y. Chen, J. D. Muller, Q. Ruan and E. Gratton, *Biophys J*, 2002, **82**, 133-144.
56. G. Egger, G. Liang, A. Aparicio and P. A. Jones, *Nature*, 2004, **429**, 457-463.
57. C. B. Yoo and P. A. Jones, *Nat Rev Drug Discov*, 2006, **5**, 37-50.
58. C. Balch and K. P. Nephew, *Adv Exp Med Biol*, 2013, **754**, 285-311.
59. J. A. Plumb, G. Strathdee, J. Sludden, et al., *Cancer Res*, 2000, **60**, 6039-6044.
60. C. Batch, T. H. M. Huang, R. Brown and K. P. Nephew, *Am J Obstet Gynecol*, 2004, **191**, 1552-1572.
61. C. B. Yoo, R. Valente, C. Congiatu, et al., *J Med Chem*, 2008, **51**, 7593-7601.
62. J. C. Chuang, C. B. Yoo, J. M. Kwan, et al., *Mol Cancer Ther*, 2005, **4**, 1515-1520.
63. A. C. West and R. W. Johnstone, *J Clin Invest*, 2014, **124**, 30-39.
64. C. Festuccia, G. L. Gravina, A. M. D'Alessandro, et al., *Endocr Relat Cancer*, 2009, **16**, 401-413.
65. C. S. Dietrich, 3rd, V. L. Greenberg, C. P. DeSimone, et al., *Gynecol Oncol*, 2010, **116**, 126-130.
66. M. Sasaki, M. Kaneuchi, S. Fujimoto, et al., *Mol Cell Endocrinol*, 2003, **202**, 201-207.

67. N. T. Phuong, S. K. Kim, S. C. Lim, et al., *Breast Cancer Res Treat*, 2011, **130**, 73-83.
68. S. A. Belinsky, M. J. Grimes, M. A. Picchi, et al., *Cancer Res*, 2011, **71**, 454-462.
69. I. Ecke, F. Petry, A. Rosenberger, et al., *Cancer Res*, 2009, **69**, 887-895.
70. M. Maio, A. Covre, E. Fratta, et al., *Clin Cancer Res*, 2015, **21**, 4040-4047.
71. F. D. Urnov, E. J. Rebar, M. C. Holmes, et al., *Nat Rev Genet*, 2010, **11**, 636-646.
72. N. Sun and H. Zhao, *Biotechnol Bioeng*, 2013, **110**, 1811-1821.
73. P. Mali, K. M. Esvelt and G. M. Church, *Nat Methods*, 2013, **10**, 957-963.
74. I. B. Hilton, A. M. D'Ippolito, C. M. Vockley, et al., *Nat Biotechnol*, 2015, **33**, 510-517.
75. S. Konermann, M. D. Brigham, A. E. Trevino, et al., *Nature*, 2015, **517**, 583-588.
76. T. P. Jurkowski, M. Ravichandran and P. Stepper, *Clin Epigenetics*, 2015, **7**, 18.
77. M. L. Maeder, J. F. Angstman, M. E. Richardson, et al., *Nat Biotechnol*, 2013, **31**, 1137-1142.
78. H. Chen, H. G. Kazemier, M. L. de Groote, et al., *Nucleic Acids Res*, 2014, **42**, 1563-1574.
79. A. W. Snowden, P. D. Gregory, C. C. Case and C. O. Pabo, *Curr Biol*, 2002, **12**, 2159-2166.
80. E. M. Mendenhall, K. E. Williamson, D. Reyon, et al., *Nat Biotechnol*, 2013, **31**, 1133-1136.
81. S. Konermann, M. D. Brigham, A. E. Trevino, et al., *Nature*, 2013, **500**, 472-476.
82. A. W. Cheng, H. Wang, H. Yang, et al., *Cell Res*, 2013, **23**, 1163-1171.
83. M. C. Haffner, A. Chaux, A. K. Meeker, et al., *Oncotarget*, 2011, **2**, 627-637.
84. C. G. Lian, Y. Xu, C. Ceol, et al., *Cell*, 2012, **150**, 1135-1146.
85. Y. F. He, B. Z. Li, Z. Li, et al., *Science*, 2011, **333**, 1303-1307.
86. B. Keith and M. C. Simon, *Cell*, 2007, **129**, 465-472.
87. Y. Yoshida, K. Takahashi, K. Okita, et al., *Cell Stem Cell*, 2009, **5**, 237-241.
88. S. Shahrzad, K. Bertrand, K. Minhas and B. L. Coomber, *Epigenetics*, 2007, **2**, 119-125.
89. A. Pal, T. Srivastava, M. K. Sharma, et al., *J Cell Mol Med*, 2010, **14**, 2646-2654.
90. C. X. Song, C. Yi and C. He, *Nat Biotechnol*, 2012, **30**, 1107-1116.
91. V. Abudara, R. G. Jiang and C. Eyzaguirre, *J Neurophysiol*, 2002, **88**, 639-649.
92. R. G. Jiang and C. Eyzaguirre, *Brain Res*, 2005, **1031**, 56-66.
93. P. Sengupta, K. Garai, J. Balaji, et al., *Biophys J*, 2003, **84**, 1977-1984.
94. L. Liang, X. Wang, D. Xing, et al., *J Biomed Opt*, 2009, **14**, 024013.
95. M. Ehrlich, *Oncogene*, 2002, **21**, 5400-5413.
96. A. S. Wilson, B. E. Power and P. L. Molloy, *Biochim Biophys Acta*, 2007, **1775**, 138-162.
97. The ENCODE Project Consortium, *Nature*, 2012, **489**, 57-74.
98. G. G. Liang, M. F. Chan, Y. Tomigahara, et al., *Mol Cell Biol*, 2002, **22**, 480-491.
99. E. L. Walton, C. Francastel and G. Velasco, *Epigenetics*, 2011, **6**, 1373-1377.
100. J. Arand, D. Spieler, T. Karius, et al., *PLoS Genet*, 2012, **8**, e1002750.
101. A. D. Riggs and Z. Xiong, *Proc Natl Acad Sci U S A*, 2004, **101**, 4-5.
102. A. Jeltsch and R. Z. Jurkowska, *Trends Biochem Sci*, 2014, **39**, 310-318.
103. W. A. Pastor, L. Aravind and A. Rao, *Nat Rev Mol Cell Biol*, 2013, **14**, 341-356.

104. B. Hendrich and A. Bird, *Mol Cell Biol*, 1998, **18**, 6538-6547.
105. O. Yildirim, R. Li, J. H. Hung, et al., *Cell*, 2011, **147**, 1498-1510.
106. C. G. Spruijt, F. Gnerlich, A. H. Smits, et al., *Cell*, 2013, **152**, 1146-1159.
107. T. Shimbo, Y. Du, S. A. Grimm, et al., *PLoS Genet*, 2013, **9**, e1004028.
108. K. Gunther, M. Rust, J. Leers, et al., *Nucleic Acids Res*, 2013, **41**, 3010-3021.
109. S. E. Brown and M. Szyf, *Mol Cell Biol*, 2007, **27**, 4938-4952.
110. K. Tatematsu, T. Yamazaki and F. Ishikawa, *Genes Cells*, 2000, **5**, 677-688.
111. G. Juan, F. Traganos and Z. Darzynkiewicz, *Methods Cell Biol*, 2001, **63**, 343-354.
112. D. Zhu, J. Fang, Y. Li and J. Zhang, *PLoS One*, 2009, **4**, e7684.
113. M. Bohmer, M. Wahl, H. J. Rahn, et al., *Chem Phys Lett*, 2002, **353**, 439-445.
114. J. Enderlein and Gregor I, *Rev Sci Instrum*, 2005, **76**.
115. R. Machan, P. Kapusta and M. Hof, *Anal Bioanal Chem*, 2014, **406**, 4797-4813.
116. A. Adebiyi, D. Narayanan and J. H. Jaggar, *J Biol Chem*, 2011, **286**, 4341-4348.
117. M. Iurlaro, G. Ficz, D. Oxley, et al., *Genome Biol*, 2013, **14**, R119.
118. K. D. Robertson, K. Keyomarsi, F. A. Gonzales, et al., *Nucleic Acids Res*, 2000, **28**, 2108-2113.
119. D. M. Woodcock, D. L. Simmons, P. J. Crowther, et al., *Exp Cell Res*, 1986, **166**, 103-112.
120. J. Chen, S. Nag, P. A. Vidi and J. Irudayaraj, *PLoS One*, 2011, **6**, e17991.
121. S. Khrapunov, C. Warren, H. Cheng, et al., *Biochemistry*, 2014, **53**, 3379-3391.
122. Q. Feng and Y. Zhang, *Genes Dev*, 2001, **15**, 827-832.
123. E. J. Noh, D. S. Lim and J. S. Lee, *Biochem Biophys Res Commun*, 2009, **378**, 332-337.
124. M. Bibikova, J. Le, B. Barnes, et al., *Epigenomics*, 2009, **1**, 177-200.
125. J. D. Knox, F. D. Araujo, P. Bigey, et al., *J Biol Chem*, 2000, **275**, 17986-17990.
126. S. Milutinovic, Q. Zhuang, A. Niveleau and M. Szyf, *J Biol Chem*, 2003, **278**, 14985-14995.
127. T. Yamaguchi, F. Cubizolles, Y. Zhang, et al., *Genes Dev*, 2010, **24**, 455-469.
128. Y. Huang and A. Rao, *Nat Struct Mol Biol*, 2012, **19**, 1061-1064.
129. C. X. Song and C. He, *Trends Biochem Sci*, 2013, **38**, 480-484.
130. M. N. Gnanapragasam, J. N. Scarsdale, M. L. Amaya, et al., *Proc Natl Acad Sci U S A*, 2011, **108**, 7487-7492.
131. M. O. Steinmetz, I. Jelesarov, W. M. Matousek, et al., *Proc Natl Acad Sci U S A*, 2007, **104**, 7062-7067.
132. N. M. Walavalkar, N. Gordon and D. C. Williams, Jr., *J Biol Chem*, 2013, **288**, 3419-3427.
133. K. Kaji, I. M. Caballero, R. MacLeod, et al., *Nat Cell Biol*, 2006, **8**, 285-292.
134. R. L. dos Santos, L. Tosti, A. Radzisheuskaya, et al., *Cell Stem Cell*, 2014, **15**, 102-110.
135. M. L. Goodenberger and R. B. Jenkins, *Cancer Genet*, 2012, **205**, 613-621.
136. B. Kunkle, C. Yoo and D. Roy, *Neurotoxicology*, 2013, **35**, 1-14.
137. S. Kreth, N. Thon and F. W. Kreth, *Cancer Lett*, 2014, **342**, 185-192.
138. Q. T. Ostrom, L. Bauchet, F. G. Davis, et al., *Neuro Oncol*, 2014, **16**, 896-913.
139. H. Heyn and M. Esteller, *Nat Rev Genet*, 2012, **13**, 679-692.

140. O. J. Sansom, K. Maddison and A. R. Clarke, *Nat Clin Pract Oncol*, 2007, **4**, 305-315.
141. R. A. Fenstermaker and M. J. Ciesielski, *Cancer Control*, 2004, **11**, 181-191.
142. D. Zagzag, K. Salnikow, L. Chiriboga, et al., *Lab Invest*, 2005, **85**, 328-341.
143. T. Schumacher, L. Bunse, S. Pusch, et al., *Nature*, 2014, **512**, 324-327.
144. P. D. Thomas, M. J. Campbell, A. Kejariwal, et al., *Genome Res*, 2003, **13**, 2129-2141.
145. H. Mi, B. Lazareva-Ulitsky, R. Loo, et al., *Nucleic Acids Res*, 2005, **33**, D284-288.
146. M. M. Garrity-Park, E. V. Loftus, Jr., W. J. Sandborn, et al., *Gut*, 2009, **58**, 1226-1233.
147. Y. Takamura, H. Ikeda, T. Kanaseki, et al., *Glia*, 2004, **45**, 392-405.
148. Y. Liang, M. Diehn, N. Watson, et al., *Proc Natl Acad Sci U S A*, 2005, **102**, 5814-5819.
149. J. M. Soos, J. I. Krieger, O. Stuve, et al., *Glia*, 2001, **36**, 391-405.
150. A. Tanzer and P. F. Stadler, *J Mol Biol*, 2004, **339**, 327-335.
151. V. Olive, I. Jiang and L. He, *Int J Biochem Cell Biol*, 2010, **42**, 1348-1354.
152. E. Mogilyansky and I. Rigoutsos, *Cell Death Differ*, 2013, **20**, 1603-1614.
153. L. He, J. M. Thomson, M. T. Hemann, et al., *Nature*, 2005, **435**, 828-833.
154. S. Volinia, G. A. Calin, C. G. Liu, et al., *Proc Natl Acad Sci U S A*, 2006, **103**, 2257-2261.
155. Y. Rais, A. Zviran, S. Geula, et al., *Nature*, 2013, **502**, 65-70.
156. S. G. Jin, Y. Jiang, R. Qiu, et al., *Cancer Res*, 2011, **71**, 7360-7365.
157. T. F. Kraus, D. Globisch, M. Wagner, et al., *Int J Cancer*, 2012, **131**, 1577-1590.
158. B. A. Orr, M. C. Haffner, W. G. Nelson, et al., *PLoS One*, 2012, **7**, e41036.
159. R. Menafra and H. G. Stunnenberg, *Front Genet*, 2014, **5**, 428.
160. P. A. Wade, A. Gegonne, P. L. Jones, et al., *Nat Genet*, 1999, **23**, 62-66.
161. C. Aguilera, K. Nakagawa, R. Sancho, et al., *Nature*, 2011, **469**, 231-235.
162. N. Reynolds, P. Latos, A. Hynes-Allen, et al., *Cell Stem Cell*, 2012, **10**, 583-594.
163. Y. Chen, N. P. Damayanti, J. Irudayaraj, et al., *Front Genet*, 2014, **5**, 46.
164. T. Yamada, Y. Yang, M. Hemberg, et al., *Neuron*, 2014, **83**, 122-134.
165. L. L. Muldoon, J. I. Alvarez, D. J. Begley, et al., *J Cereb Blood Flow Metab*, 2013, **33**, 13-21.
166. G. P. Dunn, I. F. Dunn and W. T. Curry, *Cancer Immun*, 2007, **7**, 12.
167. A. Louveau, I. Smirnov, T. J. Keyes, et al., *Nature*, 2015, **523**, 337-341.
168. M. Saleh, N. K. Jonas, A. Wiegmans and S. S. Stylli, *Gene Ther*, 2000, **7**, 1715-1724.
169. M. Ehtesham, K. Samoto, P. Kabos, et al., *Cancer Gene Ther*, 2002, **9**, 925-934.
170. J. H. Schulte, S. Horn, T. Otto, et al., *Int J Cancer*, 2008, **122**, 699-704.
171. L. Fontana, M. E. Fiori, S. Albini, et al., *PLoS One*, 2008, **3**, e2236.
172. J. H. Schulte, T. Marschall, M. Martin, et al., *Nucleic Acids Res*, 2010, **38**, 5919-5928.
173. P. Mu, Y. C. Han, D. Betel, et al., *Genes Dev*, 2009, **23**, 2806-2811.
174. S. Mi, Z. Li, P. Chen, et al., *Proc Natl Acad Sci U S A*, 2010, **107**, 3710-3715.
175. K. Conkrite, M. Sundby, S. Mukai, et al., *Genes Dev*, 2011, **25**, 1734-1745.

176. T. Uziel, F. V. Karginov, S. Xie, et al., *Proc Natl Acad Sci U S A*, 2009, **106**, 2812-2817.
177. M. Dews, A. Homayouni, D. Yu, et al., *Nat Genet*, 2006, **38**, 1060-1065.
178. T. A. Farazi, H. M. Horlings, J. J. Ten Hoeve, et al., *Cancer Res*, 2011, **71**, 4443-4453.
179. S. Kriaucionis and N. Heintz, *Science*, 2009, **324**, 929-930.
180. D. Globisch, M. Munzel, M. Muller, et al., *PLoS One*, 2010, **5**, e15367.
181. P. A. Jones and P. W. Laird, *Nat Genet*, 1999, **21**, 163-167.
182. M. Esteller, P. G. Corn, S. B. Baylin and J. G. Herman, *Cancer Res*, 2001, **61**, 3225-3229.
183. J. P. Issa and H. M. Kantarjian, *Clin Cancer Res*, 2009, **15**, 3938-3946.
184. X. Yang, F. Lay, H. Han and P. A. Jones, *Trends Pharmacol Sci*, 2010, **31**, 536-546.
185. A. Bird, *Genes Dev*, 2002, **16**, 6-21.
186. P. A. Jones and S. B. Baylin, *Nat Rev Genet*, 2002, **3**, 415-428.
187. M. Esteller, *Lancet Oncol*, 2003, **4**, 351-358.
188. J. G. Herman and S. B. Baylin, *N Engl J Med*, 2003, **349**, 2042-2054.
189. J. P. Issa, *Clin Cancer Res*, 2007, **13**, 1634-1637.
190. H. Kantarjian, J. P. Issa, C. S. Rosenfeld, et al., *Cancer*, 2006, **106**, 1794-1803.
191. R. Gurion, L. Vidal, A. Gafter-Gvili, et al., *Haematologica*, 2010, **95**, 303-310.
192. C. Stresemann and F. Lyko, *Int J Cancer*, 2008, **123**, 8-13.
193. K. Appleton, H. J. Mackay, I. Judson, et al., *J Clin Oncol*, 2007, **25**, 4603-4609.
194. D. J. Stewart, J. P. Issa, R. Kurzrock, et al., *Clin Cancer Res*, 2009, **15**, 3881-3888.
195. L. A. Cowan, S. Talwar and A. S. Yang, *Epigenomics*, 2010, **2**, 71-86.
196. S. Mirza, G. Sharma, P. Pandya and R. Ralhan, *Mol Cell Biochem*, 2010, **342**, 101-109.
197. S. Karberg, *Cell*, 2009, **139**, 1029-1031.
198. M. Rius, C. Stresemann, D. Keller, et al., *Mol Cancer Ther*, 2009, **8**, 225-231.
199. V. L. Damaraju, D. Mowles, S. Yao, et al., *Nucleosides Nucleotides Nucleic Acids*, 2012, **31**, 236-255.
200. J. Hummel-Eisenbeiss, A. Hascher, P. A. Hals, et al., *Mol Pharmacol*, 2013, **84**, 438-450.
201. E. Kaminskas, A. Farrell, S. Abraham, et al., *Clin Cancer Res*, 2005, **11**, 3604-3608.
202. A. Argemi and J. Saurina, *Talanta*, 2007, **74**, 176-182.
203. J. H. Beumer, J. L. Eiseman, R. A. Parise, et al., *Clin Cancer Res*, 2006, **12**, 7483-7491.
204. J. Ren, B. N. Singh, Q. Huang, et al., *Cell Signal*, 2011, **23**, 1082-1093.
205. T. Ben-Kasus, Z. Ben-Zvi, V. E. Marquez, et al., *Biochem Pharmacol*, 2005, **70**, 121-133.
206. G. Sheikhnejad, A. Brank, J. K. Christman, et al., *J Mol Biol*, 1999, **285**, 2021-2034.
207. S. P. Lim, P. Neilsen, R. Kumar, et al., *Biodrugs*, 2011, **25**, 227-242.
208. H. M. Byun, S. H. Choi, P. W. Laird, et al., *Cancer Lett*, 2008, **266**, 238-248.

209. B. Brueckner, M. Rius, M. R. Markelova, et al., *Mol Cancer Ther*, 2010, **9**, 1256-1264.
210. C. B. Yoo, S. Jeong, G. Egger, et al., *Cancer Res*, 2007, **67**, 6400-6408.
211. M. N. Tiwari, S. Agarwal, P. Bhatnagar, et al., *Free Radic Biol Med*, 2013, **65**, 704-718.
212. J. B. Hoppe, K. Coradini, R. L. Frozza, et al., *Neurobiol Learn Mem*, 2013, **106**, 134-144.
213. X. Huang and C. S. Brazel, *J Control Release*, 2001, **73**, 121-136.
214. Y. Yeo and K. Park, *Arch Pharm Res*, 2004, **27**, 1-12.
215. J. Cheng, B. A. Teply, I. Sherifi, et al., *Biomaterials*, 2007, **28**, 869-876.
216. B. Yu, R. J. Lee and L. J. Lee, *Methods Enzymol*, 2009, **465**, 129-141.
217. L. Zhang, D. Pornpattananangku, C. M. Hu and C. M. Huang, *Curr Med Chem*, 2010, **17**, 585-594.
218. D. K. Rogstad, J. L. Herring, J. A. Theruvathu, et al., *Chem Res Toxicol*, 2009, **22**, 1194-1204.
219. D. E. Owens, 3rd and N. A. Peppas, *Int J Pharm*, 2006, **307**, 93-102.
220. S. Acharya and S. K. Sahoo, *Adv Drug Deliv Rev*, 2011, **63**, 170-183.
221. S. Sourabhan, K. Kaladhar and C. Sharma, *Trends Biomater Artif Organs*, 2009, **22**, 211-215.
222. M. M. Yallapu, B. K. Gupta, M. Jaggi and S. C. Chauhan, *J Colloid Interface Sci*, 2010, **351**, 19-29.
223. Y. R. Neupane, M. Srivastava, N. Ahmad, et al., *Int J Pharm Sci Res*, 2014, **7**, 294.
224. J. Panyam and V. Labhasetwar, *Pharm Res*, 2003, **20**, 212-220.
225. P. Xu, E. Gullotti, L. Tong, et al., *Mol Pharm*, 2009, **6**, 190-201.
226. J. Panyam, W. Z. Zhou, S. Prabha, et al., *FASEB J*, 2002, **16**, 1217-1226.
227. E. L. Moen, A. L. Stark, W. Zhang, et al., *Mol Cancer Ther*, 2014, **13**, 1334-1344.
228. H. K. Makadia and S. J. Siegel, *Polymers (Basel)*, 2011, **3**, 1377-1397.
229. F. Danhier, E. Ansorena, J. M. Silva, et al., *J Control Release*, 2012, **161**, 505-522.
230. T. Deng and Y. Zhang, *Life Sci*, 2009, **84**, 311-320.
231. A. Granzotto, Z. Bencokova, G. Vogin, et al., *Brain Tumors - Current and Emerging Therapeutic Strategies*, InTech, 2011, 89-104.
232. S. Amatori, F. Papalini, R. Lazzarini, et al., *Lung Cancer*, 2009, **66**, 184-190.
233. D. Y. Shin, H. Sung Kang, G. Y. Kim, et al., *Biomed Pharmacother*, 2013, **67**, 305-311.
234. B. E. Bernstein, J. A. Stamatoyannopoulos, J. F. Costello, et al., *Nat Biotechnol*, 2010, **28**, 1045-1048.
235. D. B. Cox, R. J. Platt and F. Zhang, *Nat Med*, 2015, **21**, 121-131.
236. A. G. Rivenbark, S. Stolzenburg, A. S. Beltran, et al., *Epigenetics*, 2012, **7**, 350-360.
237. A. J. Keung, C. J. Bashor, S. Kiriakov, et al., *Cell*, 2014, **158**, 110-120.
238. M. Ahmad and A. R. Cashmore, *Nature*, 1993, **366**, 162-166.
239. S. Mathews and R. A. Sharrock, *Plant Cell Environ*, 1997, **20**, 666-671.
240. D. Tischer and O. D. Weiner, *Nat Rev Mol Cell Biol*, 2014, **15**, 551-558.

241. R. A. Hallett, S. P. Zimmerman, H. Yumerefendi, et al., *ACS Synth Biol*, 2016, **5**, 53-64.
242. G. C. Whitelam and P. F. Devlin, *Plant Physiol Biochem*, 1998, **36**, 125-133.
243. H. Liu, X. Yu, K. Li, et al., *Science*, 2008, **322**, 1535-1539.
244. A. B. Tyszkiewicz and T. W. Muir, *Nat Methods*, 2008, **5**, 303-305.
245. A. Levskaya, O. D. Weiner, W. A. Lim and C. A. Voigt, *Nature*, 2009, **461**, 997-1001.
246. M. J. Kennedy, R. M. Hughes, L. A. Peteya, et al., *Nat Methods*, 2010, **7**, 973-975.
247. R. M. Hughes, S. Bolger, H. Tapadia and C. L. Tucker, *Methods*, 2012, **58**, 385-391.
248. J. E. Toettcher, O. D. Weiner and W. A. Lim, *Cell*, 2013, **155**, 1422-1434.
249. Y. Nihongaki, F. Kawano, T. Nakajima and M. Sato, *Nat Biotechnol*, 2015, **33**, 755-760.
250. M. Ahmad, N. Grancher, M. Heil, et al., *Plant Physiol*, 2002, **129**, 774-785.
251. R. Sellaro, M. Crepy, S. A. Trupkin, et al., *Plant Physiol*, 2010, **154**, 401-409.
252. L. J. Bugaj, A. T. Choksi, C. K. Mesuda, et al., *Nat Methods*, 2013, **10**, 249-252.
253. D. L. Che, L. Duan, K. Zhang and B. Cui, *ACS Synth Biol*, 2015, **4**, 1124-1135.
254. T. Simonet, L. E. Zaragosi, C. Philippe, et al., *Cell Res*, 2011, **21**, 1028-1038.
255. B. K. Oh, T. H. Um, G. H. Choi and Y. N. Park, *Int J Cancer*, 2011, **128**, 857-868.
256. S. Yehezkel, R. Shaked, S. Sagie, et al., *Front Oncol*, 2013, **3**, 35.

VITA

VITA

Yi Cui was born in Yiyang, Hunan of China, on December 8, 1985. He has studied clinical medicine in Central South University XiangYa School of Medicine and received his M.D. degree in 2009. After that he has completed two years of residency training in the Department of Surgery at XiangYa Hospital. He has obtained the practicing certificate of medical practitioners in 2011. Yi Cui joined the Dr. Joseph Irudayaraj's group at Purdue University from 2011 and started his research in single-cell analysis and single-molecule detection. His work mainly involved the development and implementation of advanced optical techniques and nanotechnology for epigenetic characterization and manipulation. In spring of 2016 he graduated from Purdue University with a Ph.D. degree.

PUBLICATIONS

PUBLICATIONS

Published Journal Articles

Cui Y[*], Wang X[*], Ren W, Liu J, and Irudayaraj J, Optical Clearing Delivers Ultrasensitive Hyperspectral Dark-Field Imaging for Single-Cell Evaluation, *ACS Nano*, **2016**. 10 (3), 3132-3143. DOI: 10.1021/acsnano.6b00142. ([*]equal contribution)

Cui Y, Choudhury SR, and Irudayaraj J, Epigenetic Toxicity of Trichloroethylene: A Single-Molecule Perspective, *Toxicology Research*, **2016**. 5 (2), 641-650. DOI: 10.1039/c5tx00454c

Wang X[*], **Cui Y**[*], and Irudayaraj J, Single-Cell Quantification of Cytosine Modifications by Hyperspectral Dark-Field Imaging, *ACS Nano*, **2015**. 9 (12), 11924-11932. DOI: 10.1021/acsnano.5b04451 ([*]equal contribution)

Choudhury SR, **Cui Y**, Milton JR, and Irudayaraj J, Selective Increase in Subtelomeric DNA Methylation: An Epigenetic Biomarker for Malignant Glioma, *Clinical Epigenetics*, **2015**. 7 (1), 1-11. DOI: 10.1186/s13148-015-0140-y

Cui Y[*], Naz A[*], Thompson DH, and Irudayaraj J, Decitabine Nanoconjugate Sensitizes Human Glioblastoma Cells to Temozolomide, *Molecular Pharmaceutics*, **2015**. 12 (4), 1279-1288. DOI: 10.1021/mp500815b ([*]equal contribution)

Chowdhury B, McGovern A, **Cui Y**, Choudhury SR, Cho IH, Cooper B, Chevassut T, Lossie AC, and Irudayaraj J, The Hypomethylating Agent Decitabine Causes a Paradoxical Increase in 5-Hydroxymethylcytosine in Human Leukemia Cells, *Scientific Reports*, **2015**. 5, 9281. DOI: 10.1038/srep09281

Cui Y and Irudayaraj J, Dissecting the Behavior and Function of MBD3 in DNA Methylation Homeostasis by Single-Molecule Spectroscopy and Microscopy, *Nucleic Acids Research*, **2015**. 43 (6), 3046-3055. DOI: 10.1093/nar/gkv098

Liu J, Cho IH, **Cui Y**, and Irudayaraj J, Single mRNA Detection via Second Harmonic Super-Resolution Microscopy, *ACS Nano*, **2014**. 8 (12), 12418-12427. DOI: 10.1021/nn505096t

Cui Y and Irudayaraj J, Inside Single Cells: Quantitative Analysis with Advanced Optics and Nanomaterials, *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, **2014**. 7 (3), 387-407. DOI: 10.1002/wnan.1321

Cui Y, Choudhury SR, and Irudayaraj J, Quantitative Real-Time Kinetics of Optogenetic Proteins CRY2 and CIB1/N Using Single-Molecule Tools. *Analytical Biochemistry*, **2014**. 458C, 58-60. DOI: 10.1016/j.ab.2014.04.023

Lee K, **Cui Y**, Lee LP, and Irudayaraj J, Quantitative Imaging of Single mRNA Splice Variants in Living Cells. *Nature Nanotechnology*, **2014**. 9 (6), 474-480. DOI: 10.1038/nnano.2014.73

Chen H, Weng TW, Riccitelli M, **Cui Y**, Irudayaraj J, and Choi JH, Understanding the Mechanical Properties of DNA Origami Tiles and Controlling the Kinetics of Their Folding and Unfolding Reconfiguration. *Journal of the American Chemical Society*, **2014**. 136 (19), 6995-7005. DOI: 10.1021/ja500612d

Cui Y, Cho IH, Chowdhury B, and Irudayaraj J, Real-Time Dynamics of Methyl-CpG-Binding Domain Protein 3 and Its Role in DNA Demethylation by Fluorescence Correlation Spectroscopy. *Epigenetics*, **2013**. 8 (10), 1089-1100. DOI: 10.4161/epi.25958

Papers Submitted & in Preparation

Cui Y, Liu J, Wang X, and Irudayaraj J, Beyond Quantification: *In situ* Analysis of Pre-mRNA Alternative Splicing at the Nanoscale, **2016** (invited review, in preparation)

Cui Y, Li J, Wirbisky SE, Freeman JL, Weng L, Liu J, and Irudayaraj J, Regulatory Landscape and Clinical Implication of MBD3 in Human Malignant Glioma, **2016** (submitted)

Naz A[*], **Cui Y**[*], Collins CJ, Thompson DH, and Irudayaraj J, PLGA-PEG Nano-Delivery System for Epigenetic Therapy, **2016** (in revision; [*]equal contribution).

Choudhury SR, **Cui Y**, Lubecka K, Stefanska B, and Irudayaraj J, CRISPR-dCas9 Mediated TET1 Targeting for Selective DNA Demethylation at BRCA1 Promoter, **2016** (in revision)

Choudhury SR, **Cui Y**, Narayanan A, Gilley D, Huda N, Lo C, Zhou F, Yernool D, and Irudayaraj J, Optogenetic Manipulation of Site-Specific Subtelomeric DNA Methylation, **2015** (in revision)