# Optochemical Control of DNA Methylation Erasers and Readers enables Kinetic Insights into their Domain-Dependent Interplay

Dissertation

Submitted for the degree of Doctor of Natural Sciences (Dr. rer. nat)

to the

Faculty of Chemistry and Chemical Biology

Technical University Dortmund

By

Lin, Tzu-Chen

林子禎

Dortmund 2022

*"But we must not forget that when radium was discovered no one knew that it would prove useful in hospitals.*
*The work was one of pure science.*
*And this is a proof that scientific work must not be considered from the point of view of the direct usefulness of it.*
*It must be done for itself, for the beauty of science, and then there is always the chance that a scientific discovery may become like the radium a benefit for humanity."*

- The Discovery of Radium. Address by Madame Marie Curie at Vassar College, May 14, 1921 (published as the Ellen S. Richards Monograph No. 2.)

This work was prepared from September 2018 to July 2022 in the group of Prof. Daniel Summerer of the Faculty of Chemistry and Chemical Biology at the Technical University Dortmund.

The research projects presented in this work are funded by the European Research Council (ERC) under the ERC consolidator grant EPICODE No. 723863 to Prof. Dr. Daniel Summerer.

The molecular simulation experiments for the energy minimization models of hMBD1 S45 and T27→**1** were performed by Benedikt Söldner in the group of Prof. Rasmus Linser of the Faculty of Chemistry and Chemical Biology at the Technical University Dortmund.

# Acknowledgments

A Ph.D. journey is like the epigenetic landscape proposed by Conrad Waddington. In this Ph.D. landscape, multiple factors tie the ropes underneath to form an intricate tension network that guided me into the trajectory I have taken and supported me through the journey. I could conquer all the hills and valleys in this journey only because of the supports and guidance I received.

I would like to dedicate my sincere appreciation to my supervisor Prof. Dr. Daniel Summerer, for involving me in the exciting epigenetic and chemical biology research pursued by his team. I'm grateful to receive his excellent supervision and scientific guidance, particularly thankful for his trust that provides me the freedom to explore the unknown.

I also want to express my cordial thanks to Prof. Dr. Heinz Neumann and Dr. Leif Dehmelt for their continuous support and contribution as my thesis advisory committee members.

I want to thank Prof. Dr. Rasmus Linser for agreeing to be the second assessor for this thesis and for his kindness in providing support for molecular simulation from his team.

Thanks also go to Biruntha Ravi (B. Sc.) and Nazmi Uslu (B. Sc.), whom I had the pleasure to advice and work together for the projects. Their commitment eminently advanced the projects, and I believe they will also owe great success in their future careers.

I'm grateful to be supported by lab research assistants Simone Eppmann and Nadine Schmidt through their great experimental techniques and lab management.

I'm more than grateful to be part of the Summerer group and surrounded by many brilliant coworkers who always welcome any scientific or non-scientific conversation and be knowledgeable whenever troubles occur, not to mention the immense mental support I always receive from them. Special thanks go to former and current Summerer group members and friends in CB E02, Dr. Álvaro Muñoz-López, Dr. Anna Witte, Anne Jung, Dr. Benjamin Buchmuller, Brinja Kosel, Damian Schiller, Dominic Kamps, Dr. Jan Wolfgramm, Lena Engelhardt, Dr. Mario Gieß, Dr. Shubhendu Palei, Suchet Nanda, and Sudakshina Banerjee. I would like to express my gratitude to Dr. Shubhendu Palei for contributing his ideas to the project, as well as to Sudakshina Banerjee and Anne Jung for their time and efforts in revising the language of this thesis.

I would like to acknowledge the IMPRS-LM Ph.D. program for providing all the valuable workshops/lectures/trainings, travel fundings, and social events that bring the students together. My heartfelt appreciation goes to the IMPRS-LM coordination office, Dr. Lucia Sironi and Christa Hornemann, for organizing those wonderful events that enriched my Ph.D. journey and for being the strongest backing in the Ph.D. program.

I also want to thank Maria Sergani, Martina Reibner, Petra Alhorn, and Ulrich Schoppe in TU Dortmund for taking care of the administrative affairs and providing a worry-free working environment.

My appreciation also goes to Prof. Dr. Yaowen Wu for providing me the opportunity to join IMPRS-LM (IMPRS-CMB) and work with him in developing novel photoswitchable chemical inducers of dimerization. And I want to extend my thanks to the members of the Wu group in CGC 2, namely Yen-Pu Chang, Dr. Saheli Halder, Yasmin Shaalan, Dr. Laura Klewer, Dr. Xi Chen, Dr. Fu Li, Dr. Ai-Min Yang, Dr. Xiaoyi Xin, Dr. Dale Corkery, Dr. Georgios Konstantinidis, and Dr. Fang-Yi Li.

A special dedication to Prof. Jim-Min Fang, supervisor of my bachelor's and master's thesis and my first academic mentor, to congratulate his retirement in Jun 2022. The valuable knowledge and experiences I obtained while being trained as an organic chemist committed to new drug development, are still the most solid knowledge base that continuously supports my scientific career.

I cannot end the acknowledgment without extending my gratitude to all the people who have played a role in my Ph.D. journey. Especially Dr. Yu-Hsuan Lin, who joined together in the IMPRS-LM, has become one of my best accompanies in life in Germany since we became roommates during the selection round. And I precious all the memories with Dr. Huei-Ru Wu, Dr. Walter Hofer, Dr. Kuan-Hsun and Eva, Dr. Mengqiao Li, Shih-En Chou, and too many to name.

Last but not least, thanks go beyond words to my dearest family, my dad Ran-Long, my brother Zih-Yuan, my husband Jen-Yao and his family, and finally, my mom Fei-Yen. You're the lighthouse that always brings me safely to shore.

# Publications

Part of this work has been published in:

**Lin, T.-C.**, Palei, S., & Summerer, D. (2022). Optochemical Control of TET Dioxygenases Enables Kinetic Insights into the Domain-Dependent Interplay of TET1 and MBD1 while Oxidizing and Reading 5-Methylcytosine. *ACS Chemical Biology*, *17*(7), 1844–1852. https://doi.org/10.1021/acschembio.2c00245

Other publications:

Huang, M. R., Hsu, Y. L., Lin, T. C., Cheng, T. J., Li, L. W., Tseng, Y. W., Chou, Y. shu, Liu, J. H., Pan, S. H., Fang, J. M., & Wong, C. H. (2019). Structure-guided development of purine amide, hydroxamate, and amidoxime for the inhibition of non-small cell lung cancer. *European Journal of Medicinal Chemistry*, *181*, 111551. https://doi.org/10.1016/J.EJMECH.2019.07.054

Chiu, D. C., Lin, T. C., Huang, W. I., Cheng, T. J., Tsai, K. C., & Fang, J. M. (2017). Peramivir analogues bearing hydrophilic side chains exhibit higher activities against H275Y mutant than wild-type influenza virus. *Organic & Biomolecular Chemistry*, *15*(46), 9910–9922. https://doi.org/10.1039/C7OB02374J

Chen, C. L., Lin, T. C., Wang, S. Y., Shie, J. J., Tsai, K. C., Cheng, Y. S. E., Jan, J. T., Lin, C. J., Fang, J. M., & Wong, C. H. (2014). Tamiphosphor monoesters as effective anti-influenza agents. *European Journal of Medicinal Chemistry*, *81*, 106–118. https://doi.org/10.1016/J.EJMECH.2014.04.082

# Table of Contents

## List of Figures

## List of Tables

## Abbreviations

| | |
|---|---|
| °C | Degree Celsius |
| α-KG | α-Ketoglutarate |
| 5caC | 5-Carboxylcytosine |
| 5fC | 5-Formylcytosine |
| 5hmC | 5-Hydroxymethylcytosine |
| 5mC | 5-Methylcytosine |
| A | Adenine |
| aaRS | Aminoacyl-tRNA synthetase |
| AM | Active modification |
| AML | Acute myeloid leukemia |
| AR | Active removal |
| BER | Base excision repair |
| BLUF | Blue light using FAD |
| BODIPY | Borondipyrromethene |
| BSA | Bovine serum albumin |
| C | Cytosine |
| Carb | Carbenicillin |
| CC | Coiled coil |
| CD | Catalytic domain |
| CGIs | CpG islands |
| ChIP | Chromatin immunoprecipitation |
| ChR | Chromatin remodelers |
| CID | Chemical inducers of dimerization |
| Co-IP | Co-immunoprecipitation |
| CRY | Cryptochrome |
| CXXC | Cys-x-x-Cys |
| Cys-rich | Cysteine-rich |
| DDMs | DNA demethylases |
| DMEM | Dulbecco's Modified Eagle Medium |
| DMNB | 4, 5-Dimethoxy-2-nitrobenzyl |
| DMRs | Differential methylated regions |
| DNA | Deoxyribonucleic acid |
| DNMTs | DNA-methyltransferases |
| dNTP | Deoxynucleoside triphosphate |

| | |
|---|---|
| DPBS | Dulbecco's phosphate-buffered saline |
| DSBH | Double-stranded β-helix |
| E. coli | Escherichia coli |
| EDTA | Ethylenediaminetetraacetic acid |
| ESCs | Embryonic stem cells |
| FACS | Fluorescence-activated cell sorting |
| FAD | Flavin adenine dinucleotide |
| FBS | Fetal bovine serum |
| FCM | Flow cytometry |
| FL | Full length |
| FMN | Flavin mononucleotide |
| FRAP | Fluorescence recovery after photobleaching |
| G | Guanidine |
| GI | Gastro-intestinal |
| GR | Glycine-arginine |
| h | Hour(s) |
| HATs | Histone acetyltransferases |
| HDAC | Histone deacetylase |
| HDMs | Histone demethylases |
| HEK293T | Human embryonic kidney 293T |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| HMTs | Histone methyltransferases |
| HP1 | Heterochromatin protein 1 |
| HSAN1E | Hereditary sensory and autonomic neuropathy type 1E |
| ICF syndrome | Immunodeficiency, centromeric region instability, facial anomalies syndrome |
| ICRs | Imprinting control regions |
| KLF4 | Krüppel-like factor 4 |
| LeuRS/LRS | Leucyl tRNA synthetase |
| LINEs | Long interspersed nuclear elements |
| LOV | Light-oxygen-voltage |
| LTR | Long terminal repeat |
| MBD | Methyl-CpG binding domain |
| MCAFs | MBD1-containing chromatin-associated factors |
| MeDIP | Methylated DNA immunoprecipitation |
| mESCs | Mouse embryonic stem cells |
| MFI | Mean fluorescence intensities |
| min | Minute(s) |

| | |
|---|---|
| mRNA | Messenger RNA |
| ncAA | Non-canonical amino acids |
| NMR | Nuclear magnetic resonance |
| NuRD | Nucleosome remodeling and histone deacetylation |
| PCR | Polymerase chain reaction |
| PEI | Polyethylenimine |
| PGCs | Primordial germ cells |
| PHY | Phytochrome |
| P-LISA | Proximity ligation *in situ* assay |
| PPARγ | Peroxisome proliferator-activated receptor-γ |
| PPGs | Photocleavable protecting groups |
| PTL | Photoswitched tethered ligand |
| Pyl | Pyrrolysine |
| PylRS | Pyrrolysyl tRNA synthetase |
| RNA | Ribonucleic acid |
| RT | Room temperature |
| RTT | Rett syndrome |
| SAH | S-adenosyl-L-homocysteine |
| SAM | *S-adenosyl-L-methionine* |
| SeC | Selenocysteine |
| SELEX | Systematic evolution of ligands by exponential enrichment |
| SOC medium | Super Optimal Broth medium |
| T | Thymine |
| TALE | Transcription factor-like effector |
| TDG | Thymine DNA glycosylase |
| TETs | ten-eleven-translocation dioxygenases |
| TFs | Transcription factors |
| TRD | Transcriptional repressor domain |
| Tris | Tris(hydroxymethyl)aminomethane |
| tRNA | Transfer RNA |
| TSSs | Transcription start sites |
| TyrRS | Tyrosyl tRNA synthetase |
| U | Uracil |
| wt | Wild type |

# 1. Abstract

5-Methylcytosine (5mC) is a central epigenetic mark of mammalian DNA. It mainly occurs in cytosine-guanine (CpG) dinucleotides and is recognized competitively by methyl-CpG binding domain (MBD) proteins and ten-eleven-translocation (TET) dioxygenases, which act as methylation readers and erasers to mediate regulatory chromatin crosstalk and epigenome editing, respectively. The dynamic reader-eraser interplay at their common substrate is therefore highly regulated for a coherent transcriptional program. However, mechanistic insights of their interplay are hampered by a lack of suitable methodology to control their activities in living cells.

This work employs light-activatable human TET1 and MBD1 to enable precise temporal control of enzymatic oxidation activity or substrate recognition. Light activation is achieved by genetic encoding of a photocaged serine that can be co-translationally incorporated at critical protein sited in mammalian cells. On the one hand, monitoring the TET1-catalyzed 5mC oxidation kinetics in vivo reveals a multifaceted domain-dependent modulation by MBD1. While the MBD domain of MBD1 negatively regulates TET1 oxidation kinetics and dominates the interplay by competing for the 5mC substrates, the third Cys-x-x-Cys (CXXC3) domain of MBD1 contrarily modulates TET1 activity by binding to nonmethylated CpGs. Intriguingly, the transcriptional repressor domain (TRD) does not influence 5mC oxidation kinetics by TET1. On the other hand, studies with light-activatable MBD1 indicate a domain-dependency of cellular mCpG binding kinetics. Depriving the nonmethylated CpG affinity of the CXXC3 domain enhances binding kinetics, whereas the absence of the TRD domain results in decreased binding kinetics. Moreover, the light-activatable MBD1 can further unveil the mechanism of MBD1-TET1 interplay by uncoupling the process from prior binding events of MBD1. Collectively, this work enables first kinetic insights into the domain-dependent interplay of methylation readers and erasers in the natural chromatin context and provides novel tools to unravel the dynamic chromatin regulation program.

# 1. Zusammenfassung

5-Methylcytosin (5mC) ist eine zentrale epigenetische Markierung der Säugetier-DNA. Es tritt hauptsächlich in Cytosin-Guanin (CpG) Dinucleotiden auf und wird kompetitiv von MBD Proteinen und TET-Dioxygenasen erkannt, die als Methylierungsleser und -löscher fungieren, um regulatorische Chromatin Interaktionen bzw. Epigenomveränderungen zu vermitteln. Das dynamische Leser-Löscher-Wechselspiel an ihrem gemeinsamen Substrat ist daher für ein kohärentes Transkriptionsprogramm stark reguliert. Mechanistische Einblicke in ihr Zusammenspiel sind jedoch durch den Umstand erschwert, daß geeignete Methoden zur Kontrolle ihrer Aktivitäten in lebenden Zellen fehlen.

Diese Arbeit verwendet lichtaktivierbares menschliches TET1 und MBD1, um eine präzise zeitliche Kontrolle der enzymatischen Oxidationsaktivität oder der Substraterkennung zu ermöglichen. Lichtaktivierung wird durch die genetische Kodierung eines Serins mit lichtlabiler Schutzgruppe erreicht, das in Säugerzellen ko-translational an kritischen Positionen der beiden Proteine inkorporiert wird. Einerseits zeigt die Beobachtung der TET1-katalysierten 5mC-Oxidationskinetik in vivo eine facettenreiche domänenabhängige Modulation durch MBD1. Während die MBD-Domäne von MBD1 die TET1-Oxidationskinetik herunterreguliert und die Wechselwirkung dominiert, indem es um die 5mC-Substrate konkurriert, moduliert die dritte Cys-x-x-Cys (CXXC3)-Domäne von MBD1 die TET1-Aktivität durch Bindung an nichtmethylierte CpGs. Interessanterweise beeinflusste die transkriptionelle Repressordomäne (TRD) die 5mC-Oxidationskinetik von TET1 nicht. Andererseits deuten Studien mit lichtaktivierbarem MBD1 auf eine Domänenabhängigkeit der zellulären mCpG-Bindungskinetik hin. Das Entziehen der nichtmethylierten CpG-Affinität der CXXC3-Domäne verstärkt die Bindungskinetik, wohingegen das Fehlen der TRD-Domäne zu einer verringerten Bindungskinetik führt. Darüber hinaus kann das lichtaktivierbare MBD1 den Interaktionsmechanismus von MBD1 und TET1 weiter aufklären, indem es den Prozess von einer vorherigen 5mC-Bindung von MBD1 entkoppelt. Insgesamt ermöglicht diese Arbeit erste kinetische Einblicke in das domänenabhängige Zusammenspiel von Methylierungs-Lesern und -Löschern im natürlichen Chromatin-Kontext und stellt neue Werkzeuge bereit, um das dynamische Chromatin-Regulationsprogramm zu enträtseln.

# 2. Epigenetics and DNA Cytosine Modification

## 2.1. DNA cytosine methylation is an epigenetic mark

Decoding the information embedded in the nucleotide sequence of DNA (genotype) to build up multi-cellular organisms is a process under multilayer control. The central dogma of molecular biology explains the decoding process as a sequential flow of genetic information that gives rise to cellular phenotypes. The four DNA monomers, A, T, C, and G are transcribed residue-by-residue into four RNA monomers, A, U, C, and G (**Figure 1**), where nucleic acid triplets further constitute the basic information units that are translated into proteins (F. Crick, 1970). However, the phenotypic diversity on the cellular level is beyond what can be explained by the central dogma. For instance, the same DNA sequence of a human being results in over 200 different cell types bearing distinct functions and identities. Therefore, studies aiming to unveil the mechanisms underlying the genotype to phenotype transformation has become a special topic in biology.

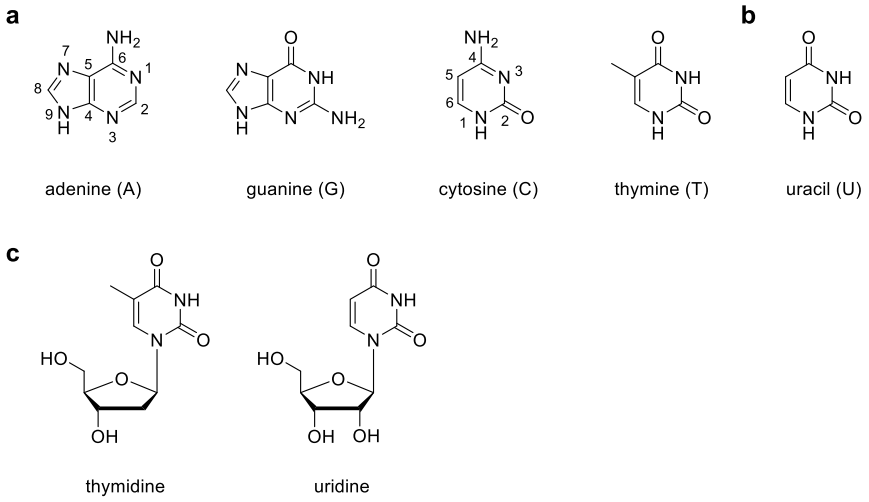


**Figure 1. Chemical structures of DNA and RNA nucleobases. a**) Chemical structures of the DNA purine and pyrimidine bases. **b**) Chemical structure of the uracil base in RNA. **c**) Thymidine nucleoside occurs in DNA, whereas uridine nucleoside occurs in RNA.

The term “epigenetics” was first coined by Waddington in the early 1940s to define “the biological processes by which genotype bring about the phenotype” (Waddington, 1942, 2012). Waddington further formulated the concept of “epigenetic landscape” to metaphorize a system that canalizes developmental decisions toward the phenotype (**Figure 2**) (Tronick & Hunter, 2016; Waddington, 1957). Scientists later discovered that the epigenetic decision-making process involves heritable changes in gene expression and function without altering the DNA sequence. Numerous elements have been found to contribute to altered chromatin activity and epigenetic changes, including the post-synthetic modifications on DNA (DNA methylation) and histone (acetylation, methylation,

phosphorylation, etc.), as well as non-coding RNAs and long-range chromatin interactions (**Figure 2**). In particular, DNA methylation on the fifth carbon of cytosine base, 5-methylcytosine (5mC), is a hallmark in epigenetic regulation (**Figure 4**).

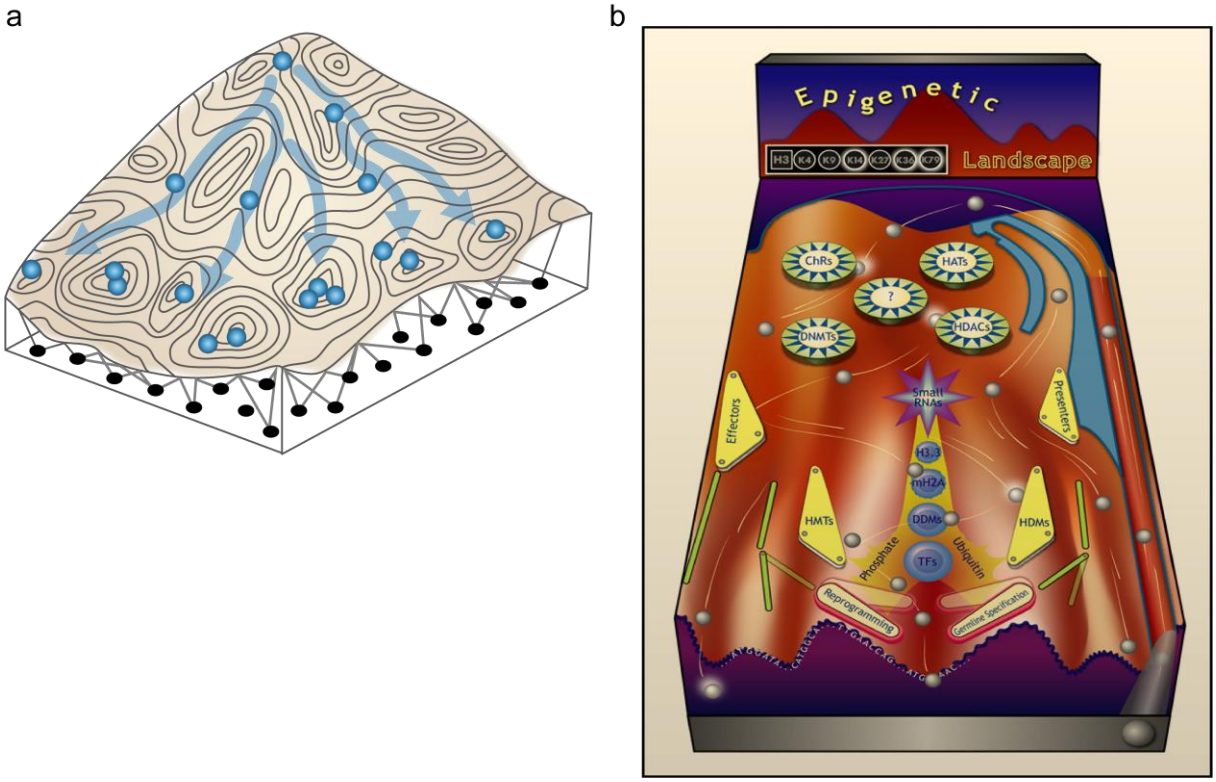


**Figure 2. Classical and current view of the epigenetic landscape. a**) An illustration of the classical epigenetic landscape proposed by Conrad Waddington. It metaphorized the cellular decision-making processes as balls (representing the cells) rolling down the epigenetic landscape, while each permitted trajectory leads to a specific biological outcome or cell-fate. The behind-the-scenes scenario is shown as the tie ropes and pegs (genes) underneath the mountain valleys that form tension networks to shape the landscape. Figure created based on (S. Huang, 2012) with permission (A.4). **b**) Illustration of a current view of epigenetic machinery with more key factors and pathways being uncovered. No specific order of events is implied as the underlying mechanism remains not fully understood (ChR: chromatin remodelers; DNMTs: DNA methyltransferases; HATs: histone acetyltransferases; HDACs: histone deacetylases; HMTs: histone methyltransferases; HDMs: histone demethylases; DDMs: DNA demethylases; TFs: transcription factors). Figure adapted from (Goldberg et al., 2007) with permission (A.4).

5mC plays important roles in transcriptional regulation, differentiation, and development. It is found in a variety of organisms including plants, vertebrates, and fungi. In mammals, about 2-7% of the cytosines are methylated. Nearly all methylation in mammalian somatic cells occurs in CpG dinucleotides and accounts for 60-80% of the total CpGs (Razin & Riggs, 1980), whereas as much as a quarter of methylation happens in a non-CpG context in embryonic stem cells (B. Jin et al., 2011). Indeed, mammalian genomes are globally depleted of CpGs despite the existence of short, interspersed sequences bearing high relative density of non-methylated CpGs termed CpG islands (CGIs) (Deaton & Bird, 2011; Smith & Meissner, 2013). Most CpG islands are protected from

methylation and associated with transcription initiation, they either span the promoter regions of housekeeping genes and developmental regulator genes or sit more remote from the annotated transcription start sites (TSSs) but exhibiting regulatory (enhancer) functions (reviewed in Deaton & Bird, 2011). Hypermethylation of CGIs often leads to transcriptional silencing of associated genes and is an important process involved in genomic imprinting and X-chromosome inactivation (Bird, 2002; Deaton & Bird, 2011; Kohli & Zhang, 2013).

Apart from CGIs, methylation also silences the expression of endogenous repetitive elements including the pericentromeric satellites and transposable elements. The pericentromeric satellites contain tens of thousands of non-coding tandem repeats that are constitutively repressed and provide structural support to eukaryotic chromosomal organization. Moreover, methylation protects genome integrity by preventing the expression and accumulation of repetitive transcripts. Studies have shown that aberrant expression of transposable elements, such as long interspersed nuclear elements (LINEs) and the long terminal repeat (LTR)-containing endogenous retroviruses, leads to disrupted gene function and oncogenicity (reviewed in Zeller & Gasser, 2017).

However, methylation does not solely contribute to gene silencing. Recent genome-wide analysis has revealed a positive correlation between DNA methylation and gene expression in regions within 2 kb of CpG islands, designated as “CpG island shores” (Irizarry et al., 2009). These contrary functions of 5mC suggest its diverse roles in various biological processes and the importance to uncover its dynamic regulation and consequences in the epigenome.

## 2.2. Shaping the DNA methylation landscape

The distribution of 5mC in the genome, namely methylation patterns, is dynamic and closely related to developmental decisions in mammals. Despite being chemically stable, distribution of 5mC vary in time and space to bring biological responses essential for normal development at different stages (reviewed in Smith & Meissner, 2013).

### 2.2.1. Dynamic methylation in mammalian development

The development of a mammalian embryo begins with a zygote from the fertilization of maternal and paternal gametes. A zygote undergoes a series of cell division and differentiation to become a mature multi-cellular organism before birth. On the other hand, the primordial germ cells (PGCs) emerge in the early embryo after gastrulation and continue to develop into mature male or female gametes (**Figure 3**).

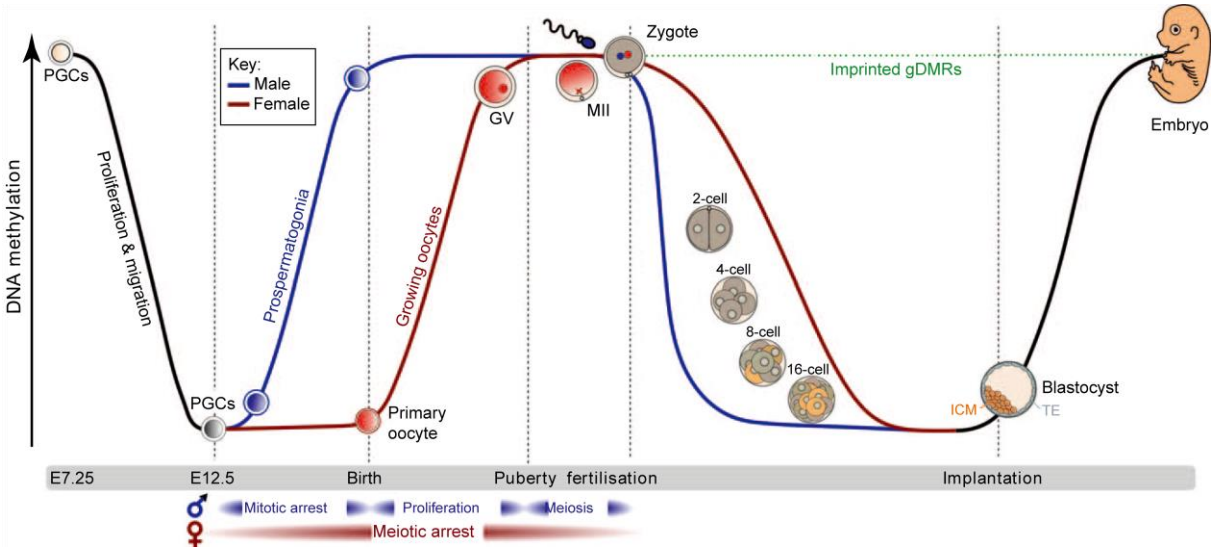


**Figure 3. Dynamic DNA methylation during mammalian embryo development**. DNA methylation is globally erased in the proliferating and migrating primordial germ cells (PGCs) (black line), and subsequently re-established de novo following sex-determination. While methylation in male embryo (blue line) occurs in mitotically arrested cells and is completed before birth, methylation in female embryo (red line) takes place after birth in the meiotically arrested oocytes. Following fertilization, another wave of demethylation occurs in a sex-dependent manner on the parental genomes. Methylation is rapidly removed in the paternal genome (blue line) via an active demethylation pathway, whereas the demethylation in the maternal genome (red line) is much slower and happens in a replication dependent fashion (passive demethylation). Noteworthily, the post-fertilization demethylation does not take place at the imprinted germline DMRs (green dotted line), leading to parental-allele-specific imprinted gene expressions. Figure adapted from (Smallwood & Kelsey, 2012) with permission (A.4).

Two waves of genome-wide demethylation and re-methylation have been shown to occur during embryo development (**Figure 3**). The first wave takes place in the primordial germ cells (PGCs) to reprogram the pre-existing methylation landscape originated from the maternal and paternal gametes. The migrating and proliferating PGCs undergo comprehensive methylation erasure which enables new methylation patterns to be established *de novo* during germ cell development in sex-specific contexts. Methylation in male germ cells begins in the mitotically arrested cells and is completed before birth, hence the establishment of methylation is separated into multiple cell division steps. In contrary, female germ cells only initiate methylation after birth in the meiotic arrested oocytes and will not be further modified (reviewed in Smallwood & Kelsey, 2012).

The second wave of demethylation takes place after fertilization in a sex-dependent manner in parentally contributed genomes (**Figure 3**). The paternal genome which adopts the methylation pattern from male germ cells experiences rapid methylation erasure in the zygote, potentially via the TET-mediated active demethylation mechanism (Mayer et al., 2000; Oswald et al., 2000). On the other hand, demethylation of the maternal genome happens in a replication-dependent fashion (passively diluted through cell division) which is much slower than the demethylation of the paternal genome. Notably, the second wave of demethylation does not affect all the genomic

regions. The allele-specific methylation at the differential methylated regions (DMRs) of the imprinting control regions (ICRs) is escaped from the demethylation and faithfully maintained as a lifelong memory from the parental genome (reviewed in J. R. Edwards et al., 2017; Smallwood & Kelsey, 2012).

### 2.2.2. Turnover of 5mC in the genome

Dynamic regulation of the methylation landscape is enabled by the reversibility of 5mC. The life cycle of 5mC describes its generation and removal in the genome, which starts with methylation writer-catalyzed active modification (AM), then followed by the replication-dependent passive dilution (PD) or methylation eraser-mediated active removal (AR) to restore the unmodified cytosine (**Figure 4**). Active modification is catalyzed by the methylation writers, DNA-methyltransferases (DNMTs), by transferring a methyl group from the *S*-adenosyl-L-methionine (SAM) cofactor onto the fifth carbon of the cytosine pyrimidine ring (**Figure 4**) (Du et al., 2016; Zangi et al., 2010). It is an important process involved in both the *de novo* creation of methylation and the maintenance of methylation patterns between cell generations to ensure faithful epigenetic information transfer. If methylation patterns cannot be efficiently maintained during DNA replication, 5mCs will be gradually replaced by unmodified cytosines and eventually leads to the passive dilution of methylation patterns after progressive rounds of replication. On the other hand, 5mC can be actively removed from the genome by the methylation erasers, ten-eleven-translocation dioxygenases (TETs). TET dioxygenases catalyze the iterative oxidation of 5mC to 5-hydroxymethylcytosine (5hmC) (Kriaucionis & Heintz, 2009; Tahiliani et al., 2009), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (**Figure 4**) (He et al., 2011; Ito et al., 2011; Pfaffeneder et al., 2011), whereby 5fC and 5caC are recognized and removed by the thymine DNA glycosylase (TDG) and leaves abasic sites for the base excision repair (BER) machinery to restore unmodified cytosines (He et al., 2011; Maiti & Drohat, 2011; Weber et al., 2016).

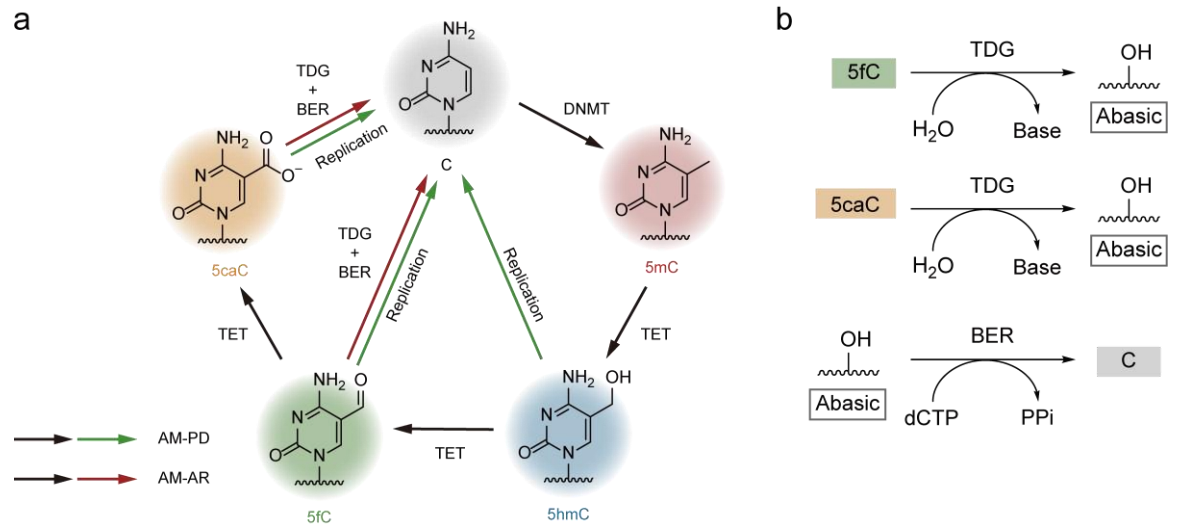


**Figure 4. The complete cycle of cytosine modification. a**) Diagram illustrating the methylation and demethylation pathways. A methyl group is introduced by the DNA methyltransferases (DNMTs) on the fifth carbon of cytosine, giving the 5mC base. The 5mC base is oxidized by the ten-eleven-translocation (TET) enzymes iteratively to 5hmC, 5fC, and 5caC to undergo demethylation. All the oxidation products can be passively removed by replication-dependent dilution (PD), while **b**) the 5fC and 5caC bases can be cleaved from the ribose moiety by the thymine DNA glycosylase (TDG), leaving an abasic site that is subsequently repaired by the base excision repair (BER) machinery with an unmodified cytosine. Figures were created based on (Kohli & Zhang, 2013) with permission (A.4).

Members of the mammalian DNMT family and TET family each owns distinct characteristics and takes part in the turnover of 5mC in different biological contexts. In the following subsections, a detailed mechanistic picture of the 5mC life cycle will be elucidated by revisiting the functional roles of mammalian DNMT family and TET family.

## 2.2.3. Methylation writers

The DNA-methyltransferase (DNMTs) family is consisted of three conserved enzymes: DNMT1, DNMT3A, and DNMT3B (**Figure 5**). DNMT3A and DNMT3B are responsible for establishing methylation *de novo*, a process mainly occuring in early mammalian development to reprogram the epigenome. For example, *de novo* methylation takes place during germ-cell specification after a complete erasure of methylation to reset the epigenetic information carried by male and female gametes (**Figure 3**) (Smallwood & Kelsey, 2012). The origin of specific *de novo* methylation patterns is not well understood, although methylation is considered the "default" state of most of the genome (J. R. Edwards et al., 2010, 2017; Lister et al., 2009). The *de novo* methylation patterns are presumably correlated with different accessibilities of DNMT3 enzymes to distinct genomic regions, and the additional support received from their interacting partners to access specific regions (Bird, 2002). For instance, evidence has shown that DNMT3B specifically targets the pericentromeric repetitive DNA sequences and CGIs on the inactive X-chromosome, and deficiency

of DNMT3B leads to a lack of methylation in these regions (Bird, 2002). Unlike DNMT3s which can work with various contexts such as unsymmetrical 5hmC:C or 5fC:C, DNMT1 strongly prefers hemi-methylated CpG sites with 5mC:C symmetry generated by the incorporation of unmodified cytosine during DNA replication. As a result, DNMT1 is mainly found at the replication foci during S-phase where it functions as the maintenance methyltransferase to methylate CpG on the newly synthesized strand. The specificity of DNMT1 ensures faithful epigenetic information transfer to daughter cells and promises the heritability of methylation pattern between cell generations (Bird, 2002; Holliday & Pugh, 1975; X. Wu & Zhang, 2017; Xu et al., 2010). It is noteworthy that the context preference of DNMT1 also contributes to the passive dilution of 5mC after TET-mediated active methylation due to inefficient maintenance at asymmetric CpG dyads modified with 5hmC, 5fC, and 5caC.

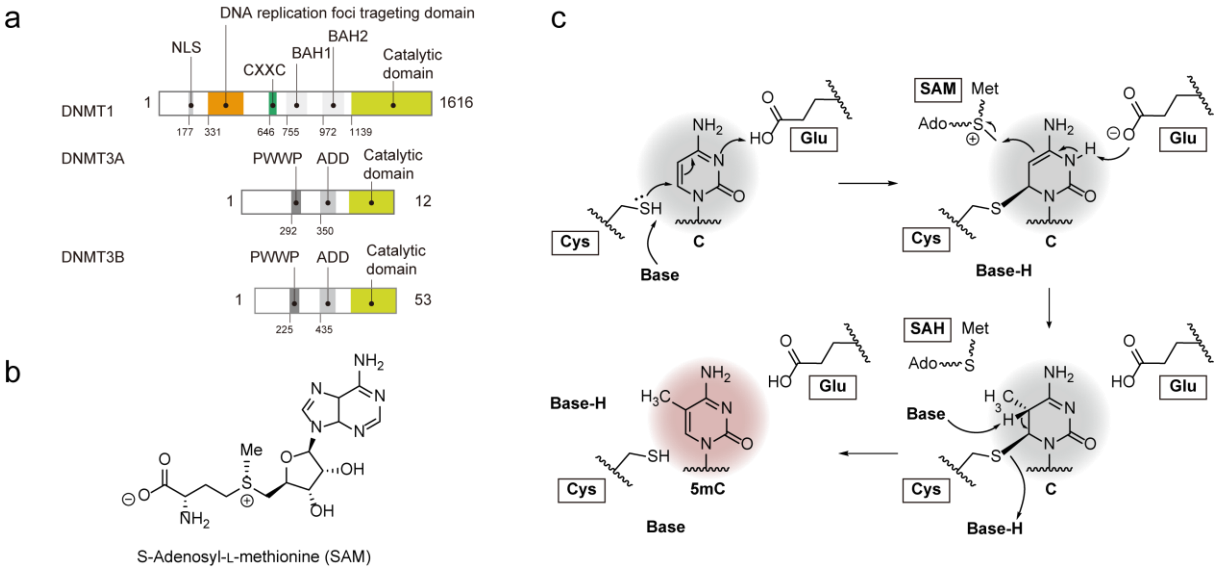


**Figure 5. DNMTs are the DNA cytosine methylation writers. a**) The domain structures of DNMT1, DNMT3A, and DNMT3B. Figure modified from (Denis et al., 2011) with permission (A.4). **b**) Chemical structure of the cofactor SAM. **c**) Mechanism of DNMT-catalyzed cytosine methylation (Du et al., 2016; Zangi et al., 2010). SAM: S-adenosyl-L-methionine; SAH: S-adenosyl-L-homocysteine.

## 2.2.4. Methylation erasers

The TET dioxygenase family has three members: TET1, TET2, and TET3. TET enzymes catalyze the oxidation of 5mC in an iron(II)/α-ketoglutarate (Fe(II)/α-KG)-dependent manner with their conserved core catalytic domain at the C-terminus, which is comprised of a cysteine-rich (Cys-rich) and a double-stranded β-helix (DSBH) domains (**Figure 6a**). In the complex with DNA, the catalytic domain exhibits a compact globular conformation with a central DSBH core buttressed by flanking segments of both the DSBH domain and the Cys-rich domains. The double-stranded DNA sits above the DSBH core with a methylated cytosine flipping out from the base pairing and

inserting into the catalytic pocket (**Figure 6c**). Zooming in the catalytic cavity of mouse TET2 (mTET2), the methyl group is orientated by the residues in the DSBH core in coordination with Fe(II) and α-KG for oxidation (**Figure 6b**). The Fe(II) is chelated in an octahedral coordination by conserved residues H1382, D1384, H1881, oxygen atoms from the 1-carboxylate and 2-keto group of α-KG, and a water molecule which is replaced by $O_2$ in the catalysis (**Figure 6b**). In addition, residue R1261 stabilizes the 1-carboxylate of α-KG, whereas the 5-carboxylate at the other end of α-KG is stabilized by residues H1416, R1896, and S1898 (**Figure 6b**). Most importantly, all residues that contribute to the interactions with Fe(II) and α-KG are highly conserved across TET dioxygenases. Zooming out from the catalytic cavity again, the interaction between DNA and the DSBH core is further supported by two loop regions (L1 and L2) of the Cys-rich subdomains (Cys-N and Cys-C) that wrap around the DSBH core and stabilize its conformation by zinc coordination (**Figure 6c**) (L. Hu et al., 2013, 2015a; Kohli & Zhang, 2013; X. Lu et al., 2015).

TET dioxygenases exhibit a substrate preference for 5mC over 5hmC and 5fC. The 5mC to 5hmC conversion has been shown to be 3-5 times faster than that of 5hmC to 5fC and 5fC to 5caC although all of them are possible substrates in the sequential oxidation (L. Hu et al., 2015a; Ito et al., 2011). In fact, hints for this substrate preference can be found in the molecular mechanism of TET-mediated 5mC oxidation which has been simulated with other Fe(II)/α-KG-dependent dioxygenases. The catalytic cycle is initiated with the oxidation of Fe(II) to reduce a dioxygen molecule and forms a Fe(IV)-peroxo intermediate. The Fe(IV)-peroxo complex is further converted into a Fe(IV)-oxo (Fe(IV)=O) intermediate following the decarboxylation of α-KG co-factor and the release of succinate and $CO_2$. The Fe(IV)-oxo intermediate then abstracts a hydrogen from the methyl group of 5mC to activate the C-H bond for subsequent hydroxylation (**Figure 6d**) (L. Hu et al., 2015a; X. Lu et al., 2015; Tarhonskaya et al., 2014). Structural analysis has elucidated the different availabilities of hydrogen atom in the C-H activation step between 5mC, 5hmC, and 5fC. In the case of 5fC, the hydrogen atom is relatively far from the Fe(IV) complex due to the intramolecular hydrogen bonding, and the planar conformation of 5fC also restricts the rotation of C-H bond to adopt a preferable orientation for reaction (L. Hu et al., 2015a; Münzel et al., 2011). Similar with 5fC, calculations also indicate the tendency of 5hmC to form hydrogen bonds which could prevent the hydrogen abstraction (L. Hu et al., 2015a).

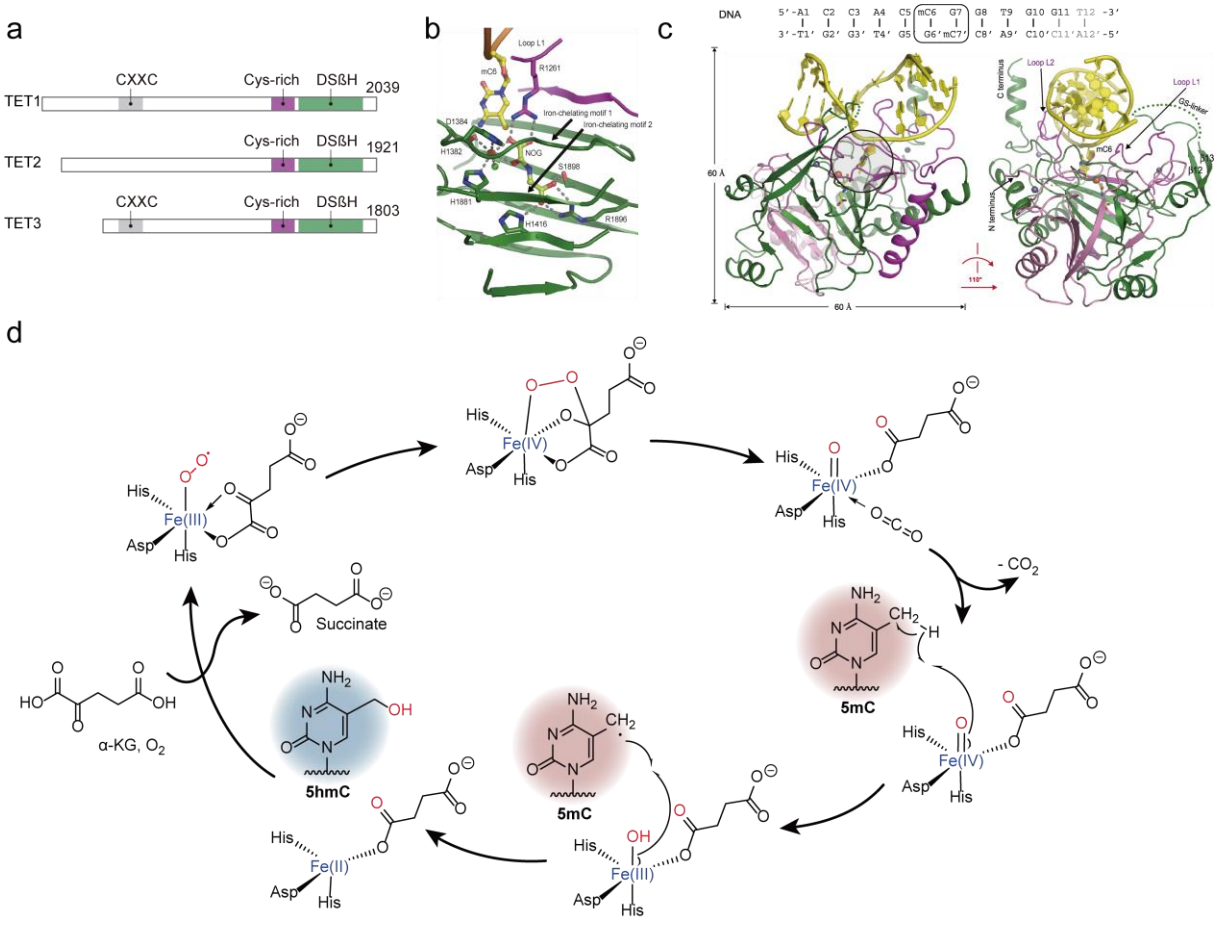


**Figure 6. TET dioxygenases are the cytosine methylation erasers. a**) The domain structures of TET1, TET2, and TET3. **b**) Critical interactions at the DSβH core with NOG (N-oxalylglycine) substituting α-KG. Dashed lines: hydrogen bonds and Fe(II) coordination; green ball: crystallographic water. Figure adapted from (L. Hu et al., 2013) with permission (A.4). **c**) Crystal structure of the TET2-DNA complex view from 2 different angles. Ribbons are color coded as the domain features shown in **a**). DNA is colored in yellow; an iron and three zinc cations are shown in red and gray, respectively. The flipped 5mC base and the catalytic center are marked with gray shadow. Figure adapted from (L. Hu et al., 2013) with permission (A.4). **d**) The proposed catalytic cycle of TET dioxygenase-mediated 5mC oxidation (L. Hu et al., 2015b; X. Lu et al., 2015; Tarhonskaya et al., 2014).

## 2.3. Interpreting DNA methylation landscape

After the epigenetic "script" of DNA methylation has been prepared by the methylation writers and erasers, it is read and interpreted by the methylation readers and associated transcription factors to bring the desired biological outcome. Hence, understanding the mechanism by which the readers and associated factors decode the methylation context can provide more insights into the causality between DNA methylation and biological consequences.

## 2.3.1. Methylation readers

The methyl-CpG binding domain (MBD) protein family is the most well-known DNA methylation reader. The MBD family comprises 11 members including the "core" members MeCP2 and MBD1–4 (**Figure 7a**), they are characterized by a conserved 70-85 amino acids long methyl-CpG binding domain which selectively recognizes single, symmetrically methylated CpG dinucleotides regardless of local sequence contexts (except the MBD of MBD3 which lacks affinity for mCpG) (Cross et al., 1997; Hendrich & Bird, 1998; Lewis et al., 1992; Meehan et al., 1989; Yildirim et al., 2011b). MBD proteins mainly function as transcriptional repressors and play key roles in coordinating the crosstalk between DNA methylation and chromatin regulatory events such as histone modifications and chromatin organization (**Figure 12**). Evidence has also shown that MBD proteins can influence alternative splicing via protein-DNA interaction. For example, MBD1 is involved in mediating CD44 alternative splicing and either the decreased methylation level (by DNMT1/3B double knockout) or MBD1 depletion leads to the skipping of exons (Batsché et al., 2021).

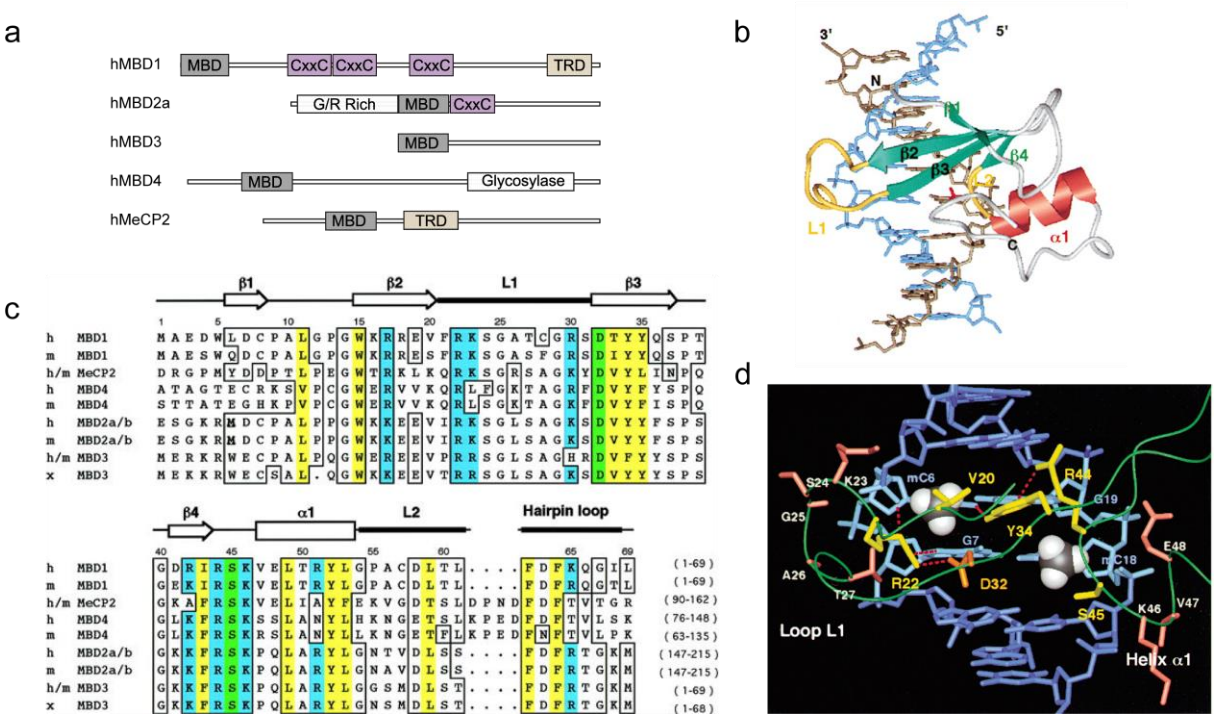


**Figure 7. MBD proteins are the cytosine methylation readers.** **a**) The domain structures of human MBD proteins (MBD1, MBD2a, MBD3, MBD4, and MeCP2). **b**) The energy-minimized averaged structure of the MBD1 MBD-DNA complex showing that MBD binds to the DNA major groove. Figure adapted form (Ohki et al., 2001) with permission (A.4). **c**) MBD domain sequence alignment comparing core MBD family proteins (MBD1–4 and MeCP2) from human (h), mouse (m), and *x.laevis* (x) (numbering shown for human MBD1). The secondary structure of human MBD1 is indicated on top as reference and conserved residues are boxed. Color code marks important residues involved in 5mC recognition (blue: basic; yellow: hydrophobic; green: acidic or polar). Figure adapted from (Ohki et al., 1999) with permission (A.4). **d**) Important contacts between MBD and DNA shown in the energy-minimized averaged structure. Red dashed lines:

proposed intermolecular hydrogen bonds; green lines: protein backbone; blue sticks: DNA. Figure adapted form (Ohki et al., 2001) with permission (A.4).

Moreover, each MBD protein has distinct domain features (**Figure 7a**) and is involved in various regulatory processes. For instance, MeCP2, the first identified and most studied MBD protein, is known to recruit the Sin3-histone deacetylase complex via its transcriptional repressor domain (TRD) (**Figure 7a**) to mediate the histone deacetylation and transcriptional silencing (Jones et al., 1998; Nan et al., 1998). Whereas the largest MBD protein MBD1, which contains three Cys-x-x-Cys (CXXC) zinc finger domains and a TRD domain (**Figure 7a**), is capable of repressing both methylated and non-methylated promoters owing to the affinity of CXXC domain to the non-methylated CpGs (Fujita et al., 1999a).

NMR structural study of the MBD1 MBD-DNA complex reveals that the MBD domain folds into an α/β sandwich structure and binds exclusively on the DNA major groove (**Figure 7b**), indicating that MBD can access mCpG sites in the major groove on nucleosomes without being sterically interfered by core histones (Ohki et al., 2001). Five residues: V20, R22, Y34, R44, and S45, form a hydrophobic patch and mediate the recognition of the symmetrically methylated CpG dinucleotide (**Figure 7d**). These residues are conserved across all functional MBD family proteins with the exception of V20. Side chains of V20 and Y34 as well as the aliphatic part of R22 side chain forms a hydrophobic pocket to accommodate the methyl group of one mC, while the other methyl group is in contact with the aliphatic portion of R44 and S45 (**Figure 7d**). R22 and R44 also facilitate the recognition of the two guanidine bases of mCpG by donating hydrogen bond interactions to form a H-bond/cation-π stair motif (**Figure 7d**) (Rooman et al., 2002). Furthermore, the hydrogen bonding between the hydroxyl group of Y34 and the 4-amino group of mC provides the specificity for the mC:G base pair (Ohki, 1999; Ohki et al., 2001). Apart from the above-mentioned essential interactions, MBD proteins indeed demonstrate distinct affinity profiles for different cytosine modifications presumably due to slight differences of binding pocket residues. Evidence has shown that MBD3 is able to bind 5hmC but lacks the affinity to 5mC which is potentially due to the Y to F replacement at position 34 (**Figure 7c**) (Yildirim et al., 2011b), while MeCP2 and MBD4 exhibit dual binding affinities to both 5mC and 5hmC despite a higher preference to 5mC (Hendrich & Bird, 1998; Lewis et al., 1992; Mellén et al., 2012; Spruijt et al., 2013).

In addition to the MBD protein family, emerging evidence has suggested other methylation readers that lack the typical MBD domain but also interact with methylated CpGs like the Kaiso protein family (Daniel & Reynolds, 1999; Filion et al., 2006) and the SET and RING finger associated (SRA) domain protein family (Spruijt et al., 2013; Unoki et al., 2004). Interestingly, many readers belonging to this class show sequence-dependent mCpG recognition. For example, the Kaiso protein recognizes sequences bearing at least two methylated CpGs within the 5'-CGCG contexts via its three-zinc-finger domain (Prokhortchouk et al., 2001), and the C/EBPα protein recognition

site is created by the methylated CRE sequence (TGACGTCA).

### 2.3.2. Protein readers of oxidized 5mC derivatives

The discovery of proteins capable of recognizing oxidized 5mC derivatives (oxi-5mCs, including 5hmC, 5fC, and 5caC) hints at their potential biological functions that may go beyond their roles as intermediates in the demethylation pathway. Protein readers specifically recruited by oxi-5mCs partially overlap with previously identified 5mC readers (**Figure 8**), which distinguish oxi-5mCs by weakened binding affinities or are even repelled by oxi-5mCs. Two of the known 5mC readers that also show affinity to 5hmC are MeCP2 and UHRF1. UHRF1 binds to both 5mC and 5hmC with similar affinities, while MeCP2 binds to 5hmC with a much lower affinity compared to 5mC (Frauer et al., 2011).

A comprehensive study by Spruijt *et al.* revealed a handful of oxi-5mC readers via mass-spectrometry-based proteomics in mESCs, NPCs (neuronal progenitor cells), and adult mouse brain. (Spruijt et al., 2013). Most of the oxi-5mC binders are highly cell-type and tissue-specific whereby their footprints only partially overlap (**Figure 8**). For instance, UHRF2 is a highly 5hmC-specific binder identified in NPCs but is not expressed in mESCs (Spruijt et al., 2013). Generally, 5fC and 5caC binders are more pronounced than 5hmC binders in mouse embryonic stem cells (mESCs) and are primarily associated with active demethylation and DNA damage response (Spruijt et al., 2013). Moreover, many of the oxi-5mC readers including p53 (Pitolli et al., 2019) and UHRF2 (H. Lu et al., 2016) are implicated in cancer.

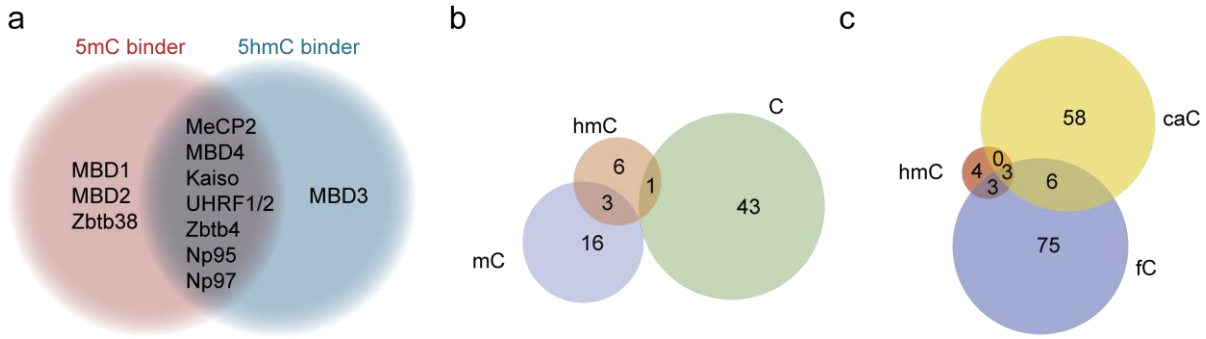


**Figure 8. Protein readers of 5mC and oxi-5mCs**. **a**) Schematic illustration showing methyl-CpG binding proteins that have been reported to selectively bind to 5mC, 5hmC, or both. Figure created based on (Ludwig et al., 2016)(A.4). **b**)**c**) Venn diagram showing numbers of specific protein binders identified using different baits in the proteomic study by Spruijt et al., 2013. Figures adapted from (Spruijt et al., 2013) with permission (A.4)

### 2.3.3. Transcription factors respond to 5mC

A variety of transcription factors (TFs) are associated with DNA methylation-related chromatin regulation and mediate the heritable epigenetic memory (reviewed in Blattler & Farnham, 2013; Zhu et al., 2016). These TFs exhibit differential binding affinities in either positive or negative relationships to 5mC in the CpG context (mCpG) which have been comprehensively profiled by a latest SELEX (systematic evolution of ligands by exponential enrichment) study. While most of the analyzed TFs are not influenced by ("little effect", **Figure 9**) or do not prefer mCpG ("methyl-minus", **Figure 9**), there are 5% of the TFs show divergent preferences for different sequence motifs or CpGs at different positions in the same motif ("multiple effects", **Figure 9**) and 34% of the TFs prefer mCpG over nonmethylated CpG in the corresponding sequences ("methyl-plus", **Figure 9**). Noteworthily, the TFs categorized as "methyl-plus" primarily belong to essential developmental transcription factors such as homeodomain proteins (Yin et al., 2017).

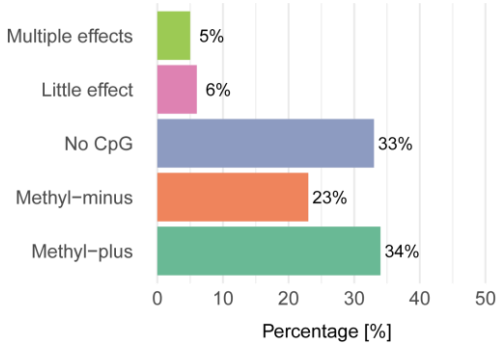


**Figure 9. Classification of transcription factors based on their binding preference to CpG context in a SELEX study** (Yin et al., 2017). Bar chart showing the fraction of TFs that preferentially bind methylated CpGs ("Methyl-plus"), unmethylated CpGs ("Methyl-minus"), or not affected by CpG methylation ("No CpG" or "Little effect"). Notably, 25 TFs (5%) showed differential preferences depend on the CpG positions of the binding sequences or at different motifs, classified as "Multiple effects". Figure created based on the data from (Yin et al., 2017).

Recent findings suggest two mutually inclusive scenarios for the TFs-DNA methylation interaction: on the one hand, methylation status affects the binding pattern of TFs; on the other hand, TFs' binding might contribute to establishing or maintaining the methylation landscape as well. For instance, the nuclear receptor PPARγ (peroxisome proliferator-activated receptor-γ) has been found to recruit TET1 and results in a decreased methylation level surrounding its recognition sites (Fujiki et al., 2013), while an orphan nuclear receptor GCNF induces the methylation of the OCT3/4 promoter via interacting with DNMT3A and DNMT3B (Sato et al., 2006).

Furthermore, some unique TFs function as "pioneer" factors to open the condensed heterochromatin (**Figure 10**) (Buganim et al., 2013), allowing the binding of other transcription factors to activate critical gene expressions during embryonic development and somatic cell

reprogramming. As heterochromatin is often methylated, it's conceivable that the ability to access methylated CpG sites in the compact chromatin regions could support the methylation-dependent pioneer factors to anchor on heterochromatin. In fact, pioneer factors like the Krüppel-like factor 4 (KLF4) shows preference to specific methylated sequences over the corresponding unmethylated sequences (S. Hu et al., 2013; Spruijt et al., 2013; Wan et al., 2017).

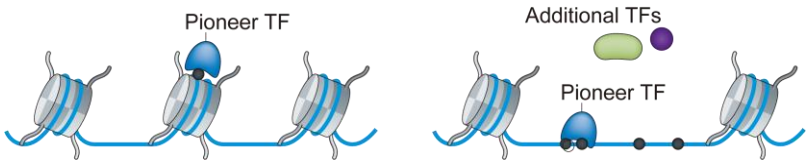


**Figure 10. Pioneer transcription factors open the condensed chromatin for other TFs.** Figure adapted from (Zhu et al., 2016) with permission (A.4).

## 2.4. Disease linked defects in the DNA methylation machinery

Aberrant activities of the components in the DNA methylation machinery, including the writers, readers, and erasers, are major causes of developmental disorders, neurodegenerative diseases, and cancers (reviewed in Ludwig et al., 2016). In this section, examples of the abnormal behaviors of the DNMTs, TETs, MBDs, as well as the associated disease are discussed.

***Disorders associated with DNMT defects***

Aberrant DNMT activities are detrimental to normal cell development. Mutations in the DNMT3B catalytic domain and the accompanying decrease of satellite DNA methylation are the hallmarks of the immunodeficiency, centromeric region instability, facial anomalies (ICF) syndrome (Hansen et al., 1999; B. Jin et al., 2008), while the DNMT3A R882 mutations in acute myeloid leukemia (AML) have been shown to reduce methyltransferase activity (Ley et al., 2010). On the other hand, DNMT1 mutations are associated with adult-onset neurodegenerative disorders such as the hereditary sensory and autonomic neuropathy type 1E (HSAN1E) which causes dementia and hearing loss (Baets et al., 2015).

***Disorder associated with TET defects***

Mis-regulation of TET dioxygenase activity has been shown to correlate with oncogenic aberrant methylation. TET2 mutations disrupting its enzymatic activity are frequently identified in AML patients featuring low 5hmC levels, indicating that impaired active demethylation pathway contributes to myeloid tumorigenesis (Ko et al., 2010). In addition to hematopoietic malignancies, a decrease of 5hmC and consistent TET downregulation was observed in various solid tumors

including human breast, liver, lung, pancreatic, and prostate cancers (Yang et al., 2013).

***Disorder associated with MBD defects***

Hampering the precision of methylation reader ability also leads to various neurological disorders and cancers. X-linked MeCP2 mutations within its MBD and TRD domains impede normal brain development in Rett syndrome (RTT) patients, resulting in progressive loss of motion and mental retardation from 6–18 months after birth (Amir et al., 1999; Hagberg et al., 1983). It has also been revealed that mRNA transcripts of MeCP2 and MBD2 are strongly elevated in breast cancer tissues. The induced MeCP2 and MBD2 protein expression are speculated to facilitate tumor cell proliferation (Billard et al., 2002; Müller et al., 2003).

# 3. Interplay Between DNA Methylation Readers and Erasers

## 3.1. Methylation readers and erasers share a common substrate

The methylation reader, MBD proteins, and the methylation eraser, TET dioxygenases, recognize 5mC as their common substrate but coordinate divergent downstream events. MBD proteins read and interpret the methylation landscape via interacting with other chromatin regulatory factors and modulate the transcriptional activity and organization of chromatin. On the other hand, TET dioxygenases mediate the removal of 5mC from epigenome and are involved in shaping the methylome at different developmental stages. Therefore, the dynamic interplay between MBD proteins and TET dioxygenases at the 5mC substrate must be highly regulated to achieve a coherent transcription program.

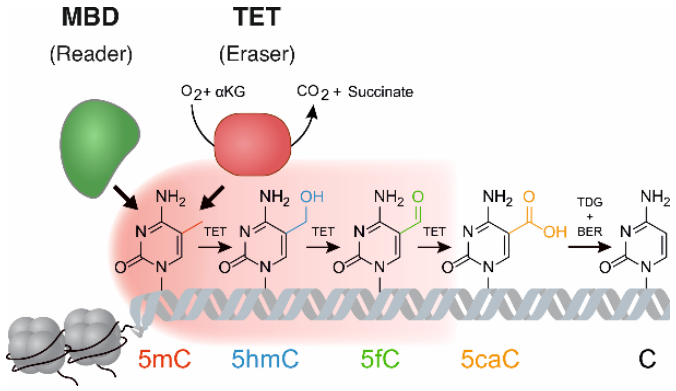


**Figure 11. Cartoon illustration of the interplay between MBDs and TETs at the 5mC substrate**. MBD proteins read and interpret 5mCs to mediate chromatin regulation, whereas TET dioxygenases oxidize 5mC and direct the active demethylation. Figure adapted from (Lin et al., 2022)(A.4).

To understand the importance of the MBD-TET interplay, it's necessary to look at the relevant biological processes that each MBD protein and TET enzyme are involved in to be informed about the potential consequences of disrupted MBD-TET interplay.

## 3.2. Core MBD family proteins: MeCP2 and MBD1-4

Generally being considered as transcription repressors, each core MBD family protein indeed bears distinct domain structures (**Figure 7a**), differential target preferences, and expression patterns (reviewed in Du et al., 2015). In the following sections, the diverse functional roles of MBD proteins are discussed.

***MBD1***

MBD1 is solely expressed in somatic tissues and is deficient in embryonic stem cells (Hendrich & Bird, 1998). It is the largest member among the MBD proteins, containing a C-terminal TRD domain and three internal zinc-finger CXXC domains in the canonical protein sequence. The third CXXC domain of MBD1 has been reported to selectively bind to non-methylated CpGs, the use of both methylated and non-methylated CpG-binding motifs thus provides an additional layer of MBD1 regulatory function (Baubec et al., 2013; Fujita et al., 1999a; Jørgensen et al., 2004). MBD1 mainly functions in mediating transcriptional silencing and heterochromatin formation (**Figure 12a** and **Figure 12h**). Unlike MeCP2, which mediates transcriptional silencing via interacting with the histone deacetylation complex, MBD1 is majorly associated with H3K9 histone methylation machinery by directly interacting with histone methyltransferase Suv39h1 or recruiting SETDB1 via its interacting partner MBD1-containing chromatin-associated factors (MCAFs) (**Figure 12a**) (Fujita, Watanabe, Ichimura, Ohkuma, et al., 2003; Fujita, Watanabe, Ichimura, Tsuruzoe, et al., 2003; Ichimura et al., 2005). It is noteworthy that although MBD1 is not embryonically lethal and the MBD1-knockout mice can survive healthily with normal development, its deficiency still leads to reduced neuron stem cell differentiation and impaired neurogenesis (X. Zhao et al., 2003).

***MBD2***

Similar to MBD1, MBD2 primarily expresses in somatic tissues with only trace expression in embryonic cells (Hendrich & Bird, 1998). MBD2 has a TRD integrated C-terminally to its MBD which is positioned in the middle of the MBD2 sequence, a C-terminal coiled coil (CC) domain, and a N-terminal glycine-arginine (GR) repeat domain. MBD2 is known to be part of the Mi-2/nucleosome remodeling and histone deacetylation (Mi-2/NuRD) complex, a unique complex that couples the histone deacetylases HDAC1/2 with the ATP-dependent chromatin remodeling enzymes CHD3/4 (**Figure 12g**). It was proposed that MBD2 recruits the Mi-2/NuRD complex to methylated regions to convert the active chromatin to hypoacetylated and compact inactive chromatin, mediating the transcriptional repression via a different mechanism than MeCP2 (Denslow & Wade, 2007; Ng et al., 1999). While a complete loss of MBD2 does not affect survival, emerging evidence has linked MBD2 to implications for immunity and cancer, although the exact roles of MBD remain to be clarified (reviewed in Wood & Zhou, 2016).

***MBD3***

MBD3 expresses in both somatic and embryonic cells. The amino acid sequence of MBD3 is highly similar (71.1%) with MBD2 (Hendrich & Bird, 1998) despite that MBD3 is the only mammalian MBD protein that does not have specificity for mCpG (Saito & Ishikawa, 2002). MBD3 purified from

ES cells (as MBD3/NuRD complex) has been found to bind unmethylated and 5hmC-containing DNA (Yildirim et al., 2011b), while later studies reported contrary findings that high concentration of recombinant MBD3 showed higher affinity to 5mC compared to 5hmC (Hashimoto et al., 2012; Spruijt et al., 2013). Compared to MBD2, MBD3 lacks the N-terminal GR domain but keeps the C-terminal CC domain. And like MBD2, MBD3 is a known subunit of the Mi-2/NuRD complex (**Figure 12b**). However, evidence has suggested that the existence of MBD2 and MBD3 in the NuRD complex are mutually exclusive with nearly nonoverlapping functions (Hendrich et al., 2001; le Guezennec et al., 2006). It was proposed that MBD3 is the essential component for NuRD formation and function, whereas MBD2 might act as a transient member for recruiting NuRD to target promoters. This is supported by the importance of MBD3 in embryogenesis, given the evidence that deletion of MBD3 leads to embryonic lethality while MBD2 deletion does not hamper survival (Hendrich et al., 2001).

***MBD4***

MBD4 is also expressed in both somatic and ES cells (Hendrich & Bird, 1998). Unique to other MBD proteins, MBD4 contains a C-terminal glycosylase domain which can repair thymine (T) or uracil (U) mismatches at CpG sites through glycosidic bond cleavage (Bellacosa et al., 1999; Hendrich et al., 1999; Petronzelli et al., 2000). Accordingly, the MBD of MBD4 has been shown to preferentially recognize the G:T mismatch resulting from the hydrolytic deamination of mCpG sites, namely 5mCpG:TpG, hence functions to prevent the mutability of mCpG by directing the repair of 5mCpG:TpG with 5mCpG:CpG (**Figure 12d**) (Hendrich et al., 1999). MBD4 knockout mice appear to be viable and fertile despite the increased C:G to A:T mutagenic transitions at CpG sites; however, stimulated gastro-intestinal (GI) tumorigenesis has been reported when MBD4 deficiency is combined with a germline-specific mutation in the *Apc* tumor suppressor gene in mice (Wong et al., 2002). This indicates that although the deficiency of MBD4 does not induce tumor on its own, the resulting mutagenesis may alter the cancer predisposition phenotype.

***MeCP2***

MeCP2 is abundantly expressed in brain tissue and central nervous systems and has been shown to play key roles in neuron maturation (Jung et al., 2003; Shahbazian et al., 2002). It contains a transcriptional repressor domain (TRD) which interacts with histone deacetylation complex to mediate transcription silencing, including histone deacetylase (HDAC) and co-expressors such as Sin3A (Jones et al., 1998; Nan et al., 1998), c-Ski (Kokura et al., 2001), NcoR (Kokura et al., 2001; Lyst et al., 2013), and CoREST (**Figure 12c**, **Figure 12e**, and **Figure 12f**) (Lunyak et al., 2002). Moreover, MeCP2 is a critical regulator of long-range chromatin remodeling, heterochromatin formation, and chromatin organization (**Figure 12f**) (Agarwal et al., 2007; Kernohan et al., 2014).

Loss of MeCP2 function is known to result in deregulated neuron development and neurological diseases such as the Rett syndrome.

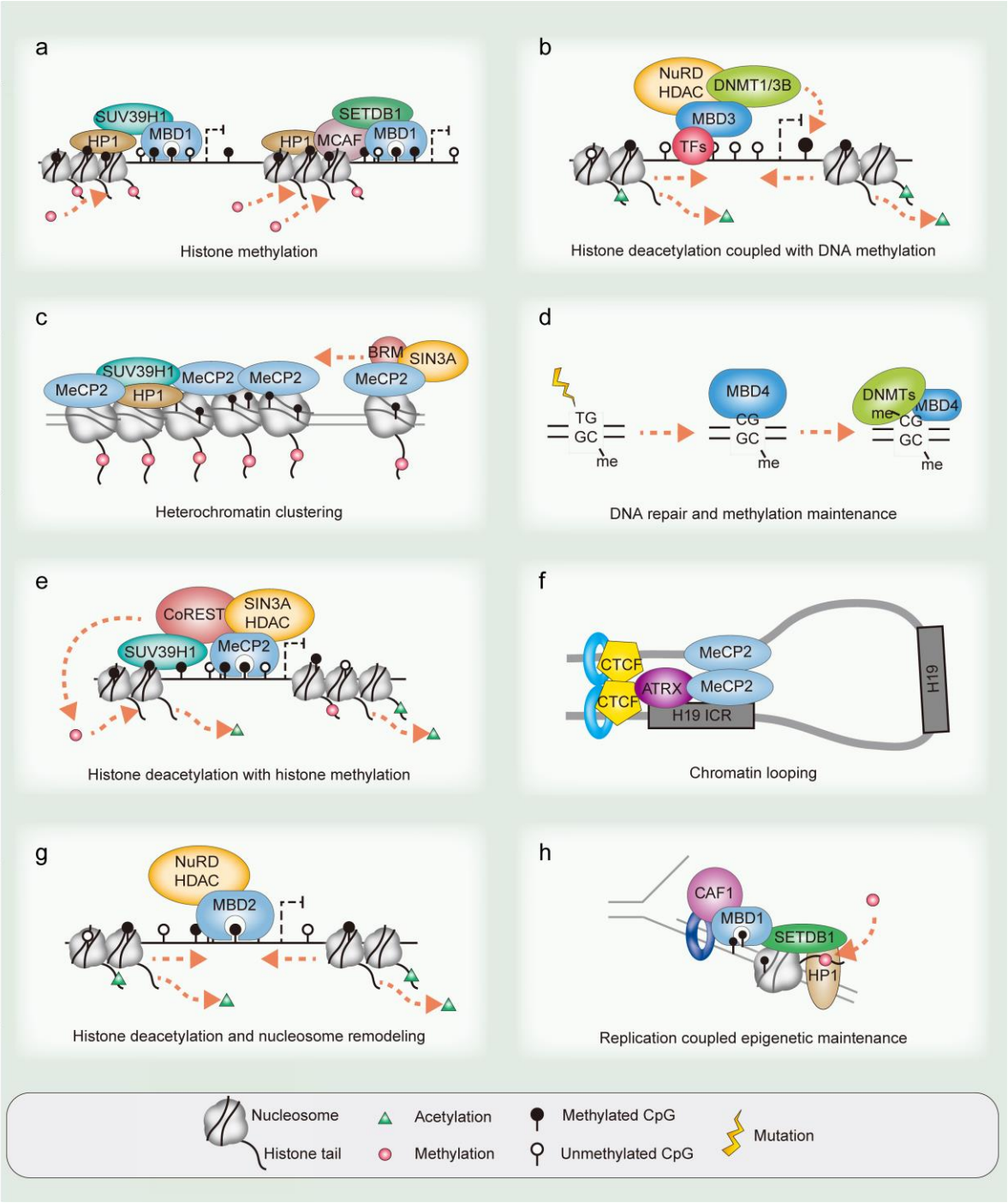


**Figure 12. Summary of biological roles of the MBD proteins**. Figure adapted from (Du et al., 2015)(A.4).

## 3.3. TET family proteins: TET1-3

TET family proteins harbor a highly conserved C-terminal catalytic domain and display the same catalytic activity; however, they are differentially expressed through development and are involved in distinct biological processes (Melamed et al., 2018; Tahiliani et al., 2009). TET3 is the only TET protein that expresses immediately after fertilization, where it mediates the global methylation erasure of the male pronucleus (**Figure 13b**) (the second wave of demethylation discussed in 2.2.1). Whereas TET1 and TET2 are abundantly expressed during the reprogramming of PGCs to finely edit the methylation pattern after the genome-wide *de novo* methylation (**Figure 13b**) (reviewed in Kohli & Zhang, 2013; Melamed et al., 2018; X. Wu & Zhang, 2017). The distinct roles of TET family proteins are therefore discussed in the following sections.

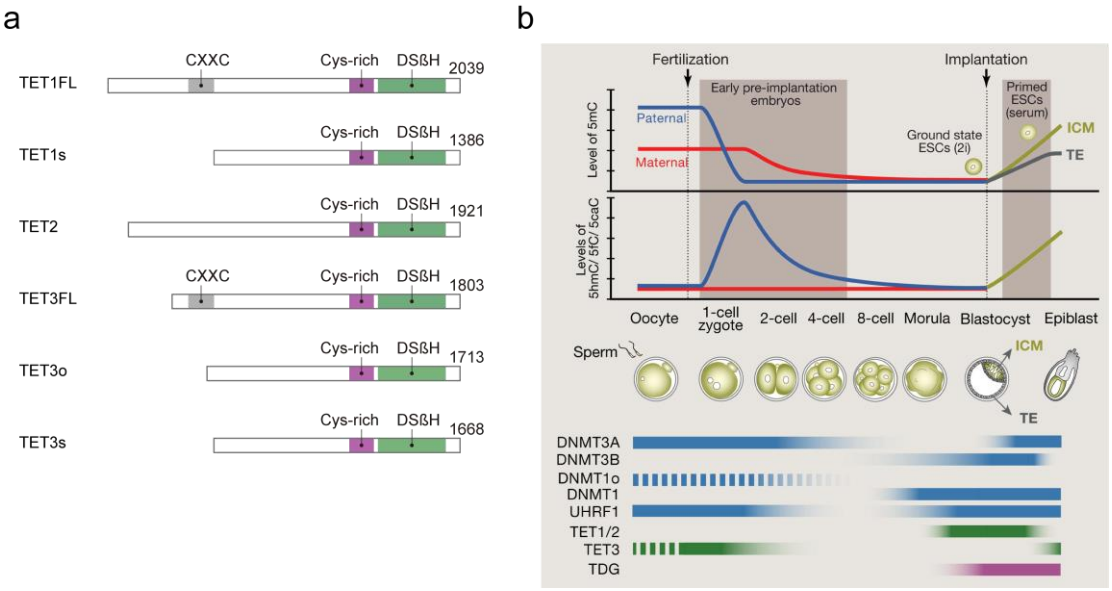


**Figure 13. TET family proteins play roles in different stages of pre-implantation development**. **a**) Domain structures of TET isoforms. **b**) Differential expression level of TETs (and DNA methylation-associated enzymes) during pre-implantation development. TET3 is highly expressed immediately after fertilization to mediate the demethylation in the paternal genome, whereas TET1 and TET2 express in PGCs and ESCs and lead to a globally hypomethylated genome. Figure adapted from (H. Wu & Zhang, 2014) with permission (A.4).

***TET1***

TET1 is highly expressed in ESCs and PGCs but is generally downregulated during differentiation (**Figure 13b**). It has been shown to play crucial roles in mediating the erasure of imprint and activation of meiotic genes during the epigenetic reprogramming of PGCs (Hackett et al., 2013; Yamaguchi et al., 2012), maintaining the self-renewal of ES cells (Ito et al., 2010), as well as directing the ES cell lineage specification (Ito et al., 2010; Koh et al., 2011).

TET1 exists in two distinct isoforms. The canonical full length TET1 has a CXXC zinc-finger domain at its N-terminus and is exclusively expressed in early embryos (ESCs and PGCs), while a shorter isoform lacking the CXXC domain expresses exclusively in somatic cells (**Figure 13a**). Although the N-terminal CXXC of TET1 belongs to the type three CXXC which has little specificity to unmethylated CpGs (Long et al., 2013), it was shown to facilitate the CGIs localization and global chromatin binding of TET1 and is essential for proper TET1 function during embryo development (W. Zhang et al., 2016).

***TET2***

Similar to TET1, TET2 is also expressed in ES cells and functions to maintain the pluripotent state (**Figure 13b**) (Koh et al., 2011). However, deletion of TET1 and TET2 in ESCs has revealed their distinct roles and target preferences. While TET1 preferentially locates to gene promoters, TET2 prefers gene bodies and enhancers and its deficiency leads to delayed gene induction during differentiation (Hon et al., 2014; Y. Huang et al., 2014). The differential binding preference can be attributed to the presence of an N-terminal CXXC domain on TET1 but not on TET2 (**Figure 13a**).

While both the canonical protein sequence of TET1 and TET3 contain a CXXC domain, the ancestral TET2 CXXC domain is now encoded in a separate gene, named *Idax* (or *CXXC4*), due to a chromosome inversion event during evolution (Iyer et al., 2009). The gene product IDAX/CXXC3 is still capable of binding unmethylated CpGs and presents as a regulator of TET2. IDAX preferentially locates at CpG-rich promoters, TSSs, and CGIs. It has been discovered that IDAX directly interacts with the catalytic domain of TET2 and regulates TET2 protein expression; moreover, it might potentially recruit TET2 to genomic targets (Ko et al., 2013).

***TET3***

Unlike TET1 and TET2, TET3 does not express in ES cells but in zygotes and oocytes and is abundantly present in neurons (**Figure 13b**) (Tahiliani et al., 2009).

TET3 exists in three major isoforms, in which the full length TET3 (TET3FL) is characterized by its N-terminal CXXC domain (**Figure 13a**). The TET3 CXXC domain also belongs to the type three CXXC as the one in TET1 but displays surprisingly high affinity to 5caC in addition to unmethylated CpGs (S. G. Jin et al., 2016; Long et al., 2013). TET3FL is the predominant form expressed in neurons, where it specifically localizes at the transcription start sites (TSSs) of a subset of genes involved in lysosome function, mRNA processing, and base excision repair pathway. Cumulating with the 5caC specificity of the TET3 CXXC domain, TET3FL might function in mediating the removal of 5caC via the base excision repair pathway (S. G. Jin et al., 2016).

The other two TET3 isoforms both lack the N-terminal CXXC domain. The shorter isoform, TET3s (**Figure 13a**), is similarly upregulated with TET3FL during the commitment of ES cells toward the neuronal lineage. The longer TET3 isoform, TET3o (**Figure 13a**), has 11 additional amino acids at the N-terminus of TET3FL and is exclusively expressed in oocytes. TET3o is hence the presumed TET oxidase that is responsible for global demethylation in the paternal genome in zygotes (Gu et al., 2011; Iqbal et al., 2011; S. G. Jin et al., 2016).

## 3.4. Recent findings on the MBD-TET interplay

Although investigating the functional roles of MBD proteins and TET dioxygenases is one of the main emphases in the epigenetic research, their interplay did not gain as much attention and remains largely unexplored. State-of-the-art MBD-TET interplay studies can be summarized into two aspects based on the biological information obtained. The static aspect of the interplay highlights the spatial relationship in a particular temporal context and is usually provided by immunoprecipitation and co-localization-based studies, whereas the dynamic aspect enables temporal resolution of the interplay granted by time-lapse imaging or fluorescence recovery after photobleaching (FRAP) assay.

***The static aspect***

The first observation of MBD-TET interplay was suggested by the overlapped genomic loci of MBD3 and TET1. Yildirim et al. mapped the genome-wide localizations of MBD3 by chromatin immunoprecipitation (ChIP) sequencing and found that the MBD3-enriched regions are accompanied with high 5hmC level and largely overlapped with the binding profile of TET1 in mouse embryonic stem cells (mESCs). Knocking down either TET1 or MBD3 expression resulted in impaired MBD3 localization at target genes and decreased global 5hmC level, respectively, which further revealed the mutually dependent relationship between MBD3 and TET1 (Yildirim et al., 2011a). The follow up study by Hainer et al. further unveiled the interdependency of MBD3, TET1, MBD2, and DNMT1, in which the localizations of MBD3 and MBD2 are overlapped and dependent on TET1 catalytic activity, and both MBD3 and MBD2 are required for normal levels of 5hmC and 5mC. Hence, a regulatory loop for modulating gene expression was proposed accordingly, which involves: (1) MBD3-mediated TET1 binding and demethylation (2) MBD3-mediated methylation by DNMT1 (3) MBD2-mediated TET1 binding and 5mC to 5hmC conversion (4) TET1/5hmC-mediated MBD3 binding (Hainer et al., 2016).

MBD3 and its homolog MBD3-like 2 (MBD3L2) were also discovered to modulate the catalytic activity of TET2. Peng et al. found that MBD3 and MBD3L2 co-localize with TET2 at TET2 target sites on chromatin, moreover, the co-expression of MBD3L2 and TET2 indeed promotes the

enrichment of TET2 at target sequences. The co-expression of MBD3 or MBD3L2 and TET2 further resulted in a greatly decreased global 5mC level concurrent with elevated global 5hmC, as well as increased expression of TET2 target genes relevant to cancer development. The collected evidence demonstrates the roles of MBD3 and MBD3L2 in modulating TET2 enzymatic activity by promoting TET2 binding to methylated DNA (Peng et al., 2016).

The spatial proximity and direct interaction between MeCP2 and TET1 were later identified from a series of TET1 interactome analyses including co-immunoprecipitation (Co-IP), proximity ligation *in situ* assay (P-LISA), and pull-down. In addition, MBD1 was also found to co-exist in the Co-IP experiment. But the absence of MBD1 in the TET1 pull down assay implies that MBD1 might not be a direct interaction partner but forms a multiprotein complex with TET1 (Cartron et al., 2013).

***The dynamic aspect***

Since the chromatin state and methylation landscape are dynamically changing, study the regulation of MBD-TET interplay with temporal resolution is of particular interest.

Cardoso and co-workers were the first to dissect the underlying mechanism of MBD2- and MeCP2-regulated TET1 enzymatic activity *in vitro* based on the evidence that co-expression of TET1 and MBD2 or MeCP2 led to reduced 5hmC formation in HEK and C2C12 cells. *In vitro* binding and activity assays further illustrated that prior binding of MBD2 and MeCP2 strongly restricts the access of TET1 to 5mC substrate hence protects the oxidation of 5mC, implying a competitive, DNA coverage- and MBD dwell-time-dependent regulation mechanism. Moreover, TET1 was shown to reactivate major satellite repeats in the neurons of MeCP2 (and MBD2)-deficient mice, hinting at the potential contribution of unconfined TET1 activity in the pathophysiological pathway of Rett syndrome (Ludwig et al., 2017).

In contrary to the negatively regulated TET1 activity in the presence of MBD2 and MeCP2, MBD1 has been shown to interact with TET1 and facilitate the localization and catalytic activity of TET1 by the same group. Zhang et al. demonstrated that MBD1 recruits TET1 but not TET2 and TET3 to accumulate at the pericentromeric loci in murine cells concomitant with elevated local 5hmC level, in which the third CXXC domain of MBD1 was identified to be a crucial component. The decreased protein exchange kinetics of MBD1 in the presence of the TET1 catalytic domain (TET1 CD) in FRAP experiments further reveals the displacement of MBD1 from the oxidized 5mC due to the activity of prior-recruited TET1, providing insights for a detailed mechanistic picture of dynamic MBD1-TET1 interplay (P. Zhang et al., 2017).

Despite the recent studies discussed above, the dynamic MBD-TET interplay remains largely

unexplored to date, particularly the kinetic insights into their dynamic regulation at native chromatins in the cellular environment. The lack of suitable methodologies to control target activities in living cells with high spatial-temporal precision substantially hampers the corresponding research. In this aspect, "light" is an appealing tool for its high spatial-temporal precision, the capability of non-invasively penetrating membrane barriers, the broad spectrum of wavelength selection, and the adjustable intensities.

# 4. Designer Tools enable Light Control of Epigenetic Regulators

## 4.1. Controlling dynamic biological events with light

Implementing photocontrol to study the dynamic regulation of biological targets in all domains of life has become one of the main interests of scientific research in the past decades. The exploration of *in vivo* protein function and regulation is especially aided by photocontrol, as light can noninvasively perform precise spatial-temporal control in living cells or organisms.

Currently, there are two main methods available for applying photocontrol in living systems: optochemical genetics and optogenetics. These methods differ in their receiver modules that bear core chromophores to initiate light-responsive reactions (reviewed in Kneuttinger, 2022). Optogenetic tools fuse protein receivers with natural photoreceptors, which react with chromophore substrates and undergo conformational changes upon light, to translate photo responses into the controls of the hybridized protein receivers (**Figure 16**) (reviewed in Manoilov et al., 2021). Optochemical genetics exploits chemically synthesized small molecules as the core chromophore in accordance with its name, and it can be further categorized into two approaches: the photopharmacology approach (**Figure 14a**) and the codon-specific protein modification approach (**Figure 14c**).

The photopharmacology approach employs protein ligands decorated with light-sensitive functional groups as receivers to influence protein function (**Figure 14a**). Not only the pharmacologically active chemical compounds but also the chemical inducers of dimerization (CID), such as the commonly used rapamycin or the TMP/DHFR conjugation system, can be the protein ligands in this approach (**Figure 14b**) (reviewed in Ankenbruck et al., 2018). But the CID system would require additional genetical engineering to equip target proteins with responsible ligand-binding domains. It is also noteworthy that the photo-pharmacological ligands can either be used alone or tethered onto the target protein to proximately control the protein function. The most well-known example of the tethered ligand approach is the successful application of PTL (photoswitched tethered ligand, **Figure 14a**) in neurons reported by Trauner and Kramer in 2004 (reviewed in Banghart et al., 2004; Fehrentz et al., 2011).

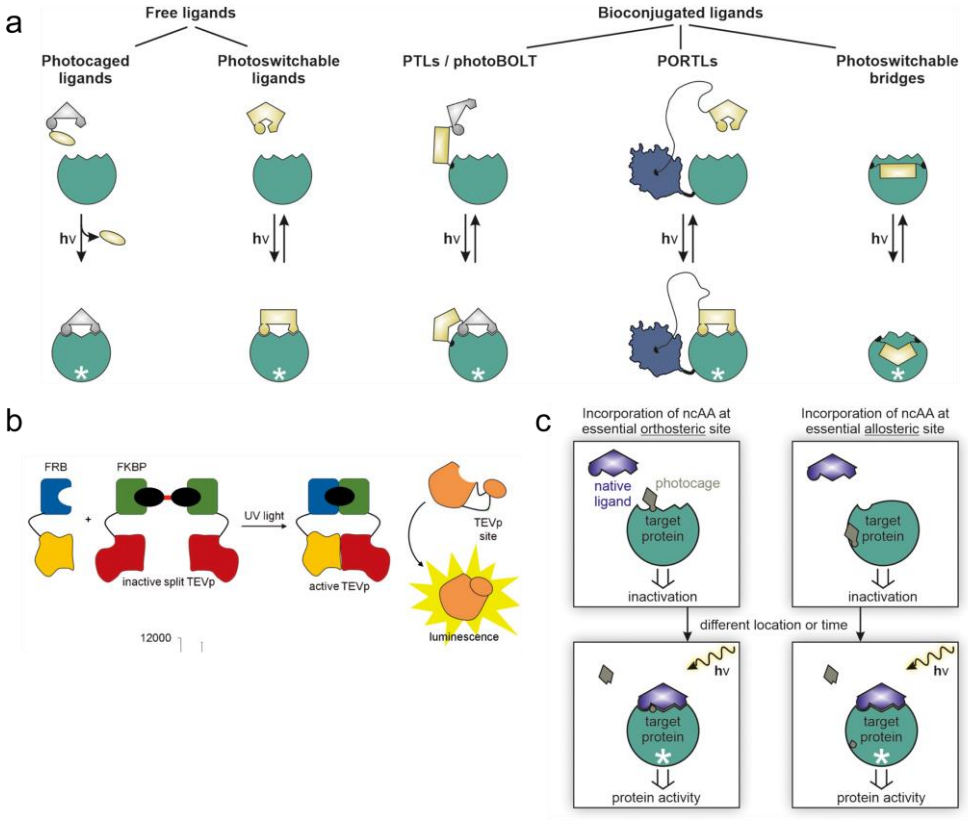


**Figure 14. Principles of photopharmacology approach and codon-specific protein modification approach**. **a**) Main strategies photopharmacology. PTLs: photoswitchable tethered ligands; photoBOLT: photoswitchable biorthogonal ligand tethering; PORTLs: photoswitchable orthogonal remotely tethered ligands. White asterisk: inhibition effect. Figure adapted from (Kneuttinger, 2022)(A.4). **b**) Representative example of chemical inducers of dimerization (CID) via rapamycin-mediated FRB/FKBP interaction. Figure adapted from (Ankenbruck et al., 2018) with permission (A.4). **c**) Concepts of photocontrol via codon-specific protein modification. Figure adapted from (Kneuttinger, 2022)(A.4).

On the other hand, the codon-specific protein modification approach enables chemical modification of photo-responsive functional groups on amino acid residues, and the user-defined site-specific incorporation of the non-canonical amino acid (ncAA) into the target protein receivers to directly control protein function (**Figure 14c**) (reviewed in Ankenbruck et al., 2018; Fehrentz et al., 2011).

Despite the different approaches to impart photocontrol on biological targets, these methods share two pivotal strategies: photoswitch and photocage. Photoswitch employs reversible conformational change of chemical functional groups or photoreceptor modules to enable the switch between on/off state, whereas photocage generally inhibits the function or interaction of target proteins via steric interference in the presence of conformational "cage" which can be "uncaged" upon light irradiation. Photoswitch typically allows multiple rounds of on/off control, whereas photocage, particularly the chemical photocage groups, is often limited to irreversible on/off control.

In order to select a suitable photocontrol method and strategy, the principles to implement the above-mentioned photocontrol methods are discussed in the following sections. Considering that the application of photopharmacology is more restricted to the availability of protein ligands, the next sections will only focus on the optogenetic approach and the codon-specific protein modification approach.

## 4.2. Principles of applying optogenetic tools

Optogenetic tools are built with genetically encoded photosensors fused to an effector module (such as protein binders or regulating peptides) targeting specific cellular events. Common photosensor selections include the blue-light-sensitive LOV (light-oxygen-voltage) domain, CRYs (cryptochrome), BLUF (blue light using FAD), and red/far-red-sensitive PHY (phytochrome), where LOV domain is the smallest among the presently available photosensor modules (**Figure 15c**). Light irradiation triggers the reaction between photosensor modules and chromophores whereby induces a conformational change of photosensor to convey the functional control of effector modules. For instance, the LOV domain employs a cysteine residue to form a covalent adduct at the C4α position of oxidized flavin upon illumination, which provokes hydrogen bond rearrangements in the binding pocket and further propagated to a conformational change of the LOV domain (**Figure 15a** and **Figure 15b**) (Halavaty & Moffat, 2007; Pudasaini et al., 2015). Most blue light photosensors like CRY and BLUF also utilize flavin mononucleotide (FMN, **Figure 15b**) or flavin adenine dinucleotide (FAD) chromophores like the LOV domain. Notably, different photosensor modules undergo different conformational changes resulting in diverse outcomes, which makes optogenetic tools useful for a variety of purposes (**Figure 16**).

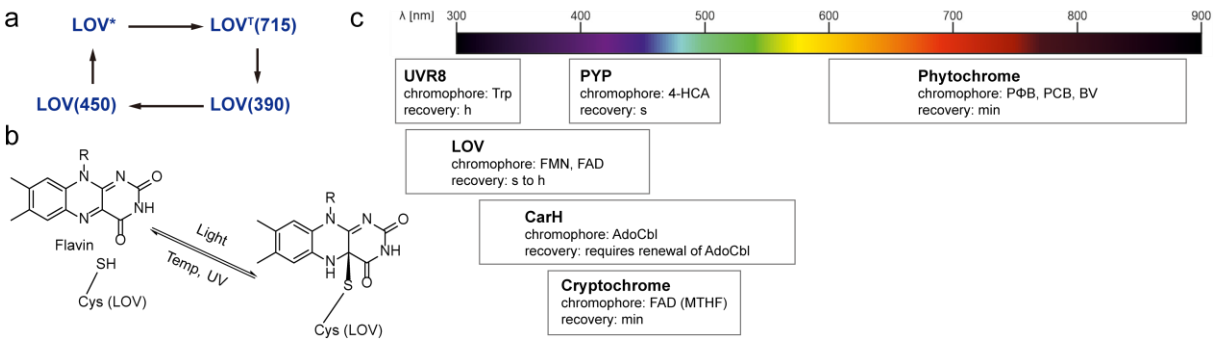


**Figure 15. Optogenetic photosensor modules**. **a**) Scheme of LOV domain photocycle. A LOV domain cysteine residue reacts with flavin to form a flavin-cysteine C4a adduct upon blue light irradiation. The reaction proceeds through an excited singlet state (LOV*) and a triplet species ($LOV^T$), followed by radical recombination to form the C4a adduct LOV(390). The C4a adduct can decay to ground state LOV(450) via either thermal relaxation or UV-sccisson (Pudasaini et al., 2015). **b**) (Pudasaini et al., 2015). **c**) Representative examples of photosensor modules and their respective absorption spectrum. Figure modified from (Kneuttinger, 2022)(A.4).

***Application strategies***

Optogenetic tools enable the control of cellular targets in several means (**Figure 16**) (reviewed in Manoilov et al., 2021). The conformation of effectors, thereby their activity and subcellular localization, can be modulated by photosensors (**Figure 16c**). A recent study demonstrates the allosteric control of Src kinase activity in living cells by an engineered Src-LOV2 fusion where Src is inhibited by LOV2 conformation change upon blue light irradiation (Dagliyan et al., 2016). Moreover, the cellular distribution of LOV2-Src can also be regulated owing to the similarity of inactive Src conformation in LOV2-Src fusion to endogenous inactive Src.

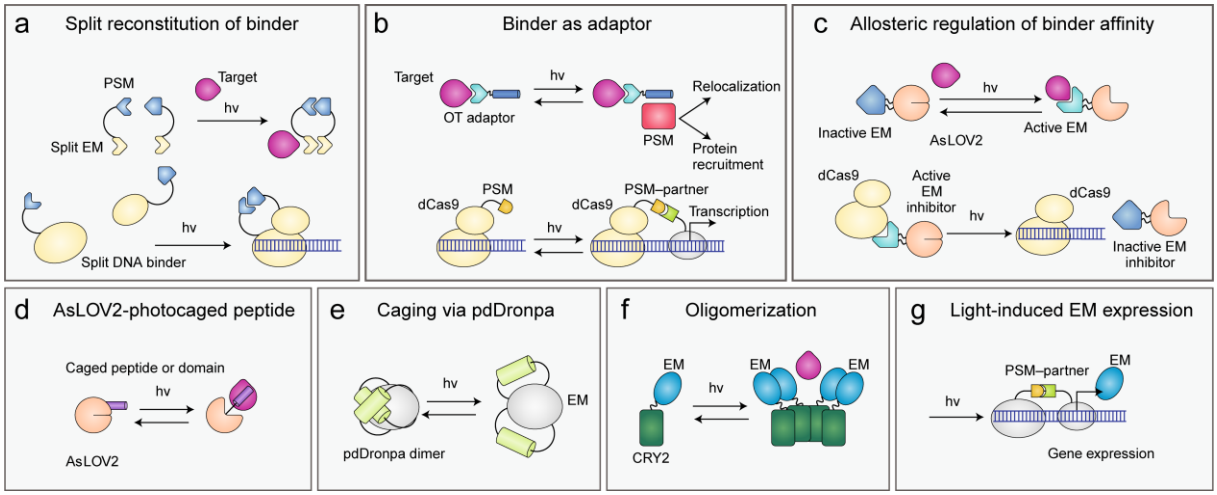


**Figure 16. Principles of optogenetics approaches**. Figure modified from (Manoilov et al., 2021) with permission (A.4).

A different approach to modulating effectors' activity is by caging or uncaging their active sites through the conformation of photosensors. Such an approach typically exploits dimeric photosensors or photosensor-tethered inhibitors to create the open or close conformation (**Figure 16d** and **Figure 16e**). Zhou *et al.* attached two cyan-light-controlled photodissociable dimeric protein domains, pdDronpa, on various kinase domains to cage their activity with dark state dimers and enable activation upon light-induced dissociation (Zhou et al., 2017). An example of caging via tethered inhibitors was illustrated by Schmidt *et al.* using an ion channel peptide ligand tethered on a photoswitch module based on a LOV domain. Once illuminated, the conformational change of the LOV domain uncages the ion channel by bringing away the peptide ligand (Schmidt et al., 2014).

Some photosensors known to form homo-/hetero-dimers or oligomers are employed to control the target binding of effectors by altering their dimerization/oligomerization state (**Figure 16a**, **Figure 16b**, **Figure 16f**, **Figure 16g**). As an example, Yazawa *et al.* utilizes the light-dependent binding of LOV-containing FKF1 and nuclear protein GI to control gene expression. By fusing FKF1

and GI to DNA binding domain Gal4 and transactivation domain of VP16, respectively, the light-induced dimerization successfully activates transcription in mammalian cells (Yazawa et al., 2009). Recent works have further utilized the light-induced CRY2-CIB1 dimerization pair to control the site-specific epigenome editing with de novo DNA methylation writer DNMT3A and methylation eraser TET1. The target specificity was achieved by a CIB1-fused transcription activator-like effector (TALE) targeting the Ascl1 gene promoter, which recruits the CRY2-fused DNMT3A or TET1 upon blue light and alters the local methylation pattern as well as gene expression (Choudhury et al., n.d.; C. L. Lo et al., 2017).

***Limitations***

However, there exist certain limitations of optogenetic tools despite their broad applications. A significant example is their intrinsic reversibility. Photosensors return to their dark state conformation via either relaxation or light-triggered scission (Pudasaini et al., 2015). Reversibility is one of the greater advantages of optogenetic tools but also is a double-edged sword since continuous illumination might be necessary in specific cases to lock the "light" state conformation for the desired effect (Nihongaki et al., 2015). Although efforts have been made on structural optimizations to attenuate the decay, it is more important to carefully evaluate and select suitable photosensor modules with lifetime and decay kinetics matching the timescale of targeted biological processes. Background leakage (residual dark-state function) and off-target effects are also noteworthy drawbacks of optogenetic tools, especially the off-target enzymatic activity when the binding of an enzyme effector is regulated by photosensor dimerization. Some other limitations also include photostability, phototoxicity, and photochemical properties *in vivo*. Therefore, optimization is required for every application in order to achieve a satisfying result.

## 4.3. Principles of codon-specific protein modification

Instead of using light-responsive proteins, chemists take advantage of light-responsive chemical functional groups to expand the optical toolbox for living systems. One of the most important applications of such chemical tools is the codon-specific protein modification to render a target protein light-responsive, which enables precise control of the targeted protein function or activity at single residue resolution by site-selective side chain modification. The available light-responsive side chain moieties can be grouped by their applications in either photocage or photoswitch approach.

### 4.3.1. Light-responsive non-canonical amino acids

***Photocaged ncAAs***

Photocage groups aim to block the function or activity of target proteins either via inhibiting the enzymatic activity conferred by residues like serine and cysteine, or via interrupting local non-covalent interactions that stabilize the protein-substrate, protein-cofactor, or protein-protein interactions (**Figure 14c**). This type of chromophore utilizes photocleavable protecting groups (PPGs) to protect active chemicals/residues, where the active chemical/residue can be released from the PPG as a leaving group upon light irradiation.

A handful of PPGs have been explored to enable precise control of the release of chemicals in biological systems, including derivatives of coumarin, borondipyrromethene (BODIPY), cyanine, and *ortho*-nitrobenzene (reviewed in Bardhan & Deiters, 2019; Klán et al., 2013). Among the PPGs, *o*-nitrobenzyl is the most common moiety in caging biomolecules (**Figure 17a**). It is noteworthy that rather than the generic *o*-nitrobenzyl group which has disadvantages of slow cleavage kinetics and short absorption wavelength (< 350 nm), their dimethoxy or methylenedioxy derivatives (**Figure 17a**) are extremely useful due to their high quantum yield, rapid cleavage, red-shifted absorption spectrum (> 350 nm and strong absorption around 400 nm), and good solubility in biological media (mostly aqueous solution) (Klán et al., 2013; Patchornik et al., 1970; Walker et al., 1993).

*o*-Nitrobenzyl derived PPGs have been used to modify serine (**1**, **Figure 17a**) (Lemke et al., 2007), tyrosine (**7**–**8**, **Figure 17a**) (Deiters et al., 2006), lysine (**9**–**10**, **Figure 17a**) (Chen et al., 2009; Gautier et al., 2010), cysteine (**2**–**4**, **Figure 17a**) (N. Wu et al., 2004), selenocysteine (**5**, **Figure 17a**) (Rakauskaite et al., 2015), and threonine (only been applied in solid phase synthesis) (Mainz et al., 2016) as photocaged ncAAs. The canonical amino acids are attached via ester/thioester or carbamate bond to the benzylic position as leaving groups to be released upon photoreaction of *o*-nitrobenzyl group (**Figure 17b**) (Il'ichev & Wirz, 2000; Klán et al., 2013). However, potential drawbacks of these photocaged ncAAs include the concomitant release of toxic nitrobenzaldehyde by-products and the potential condensation reaction between amine residues and the nitrobenzaldehyde by-product.

Apart from the *o*-nitrobenzyl derived ncAAs, Luo et al. reported a set of coumarin-based photocaged lysine which can be activated via single-photon or two-photon pathway and provide advantageous three-dimensional resolution (**6**, **Figure 17a**) (Luo et al., 2014).

***Photoswitch ncAAs***

Since photocaged ncAAs are limited to irreversible activation, photoswitch ncAAs were developed in order to mimic the reversible biological processes in nature. In this regard, the azobenzene moiety is the most suitable chromophore (**11**–**13**, **Figure 17a**). Azobenzene is known for its unique ability to isomerize between *trans* and *cis* conformations upon different wavelengths of light. It exists in the thermodynamically favored *trans* conformation in dark, isomerized to *cis* conformation upon UV-A or deeply violet light (315–380 nm), and can be reverted to *trans* conformation via thermo-relaxation or visible light-induced thermo-isomerization (Zimmerman et al., 1958). Azobenzenes hold many attractive characteristics, including high extinction coefficients and quantum yields, fast photoswitching rate, and relatively high photostability. These characters allow azobenzenes to isomerize with lower light intensity and be switched over multiple cycles (Fehrentz et al., 2011).

Phenylalanines equipped with various azobenzene derivatives bearing distinct aromatic substituents constitute the currently available photoswitch ncAAs (**11**–**13**, **Figure 17a**). The different geometry and dipole of two azobenzene isomers are useful for reversible modulation of the substrate or ligand binding site to impart activity control of target proteins. For instance, the *trans* conformation is long and extended which often does not disrupt protein conformation or the local interactions in the binding pocket, whereas the *cis* conformation is compact and bulky and may bring large steric interference to destabilize substrate binding (Bose et al., 2006).

Nonetheless, applications of photoswitch ncAAs are not as well-established as the photocaged ncAAs. It is practically difficult for azobenzene to completely switch to *cis* conformation. In addition, it has the potential of being reduced by thiol reagents in cells and organisms, such as glutathione (Boulègue et al., 2007). More applications and improvements of the method remain to be explored.

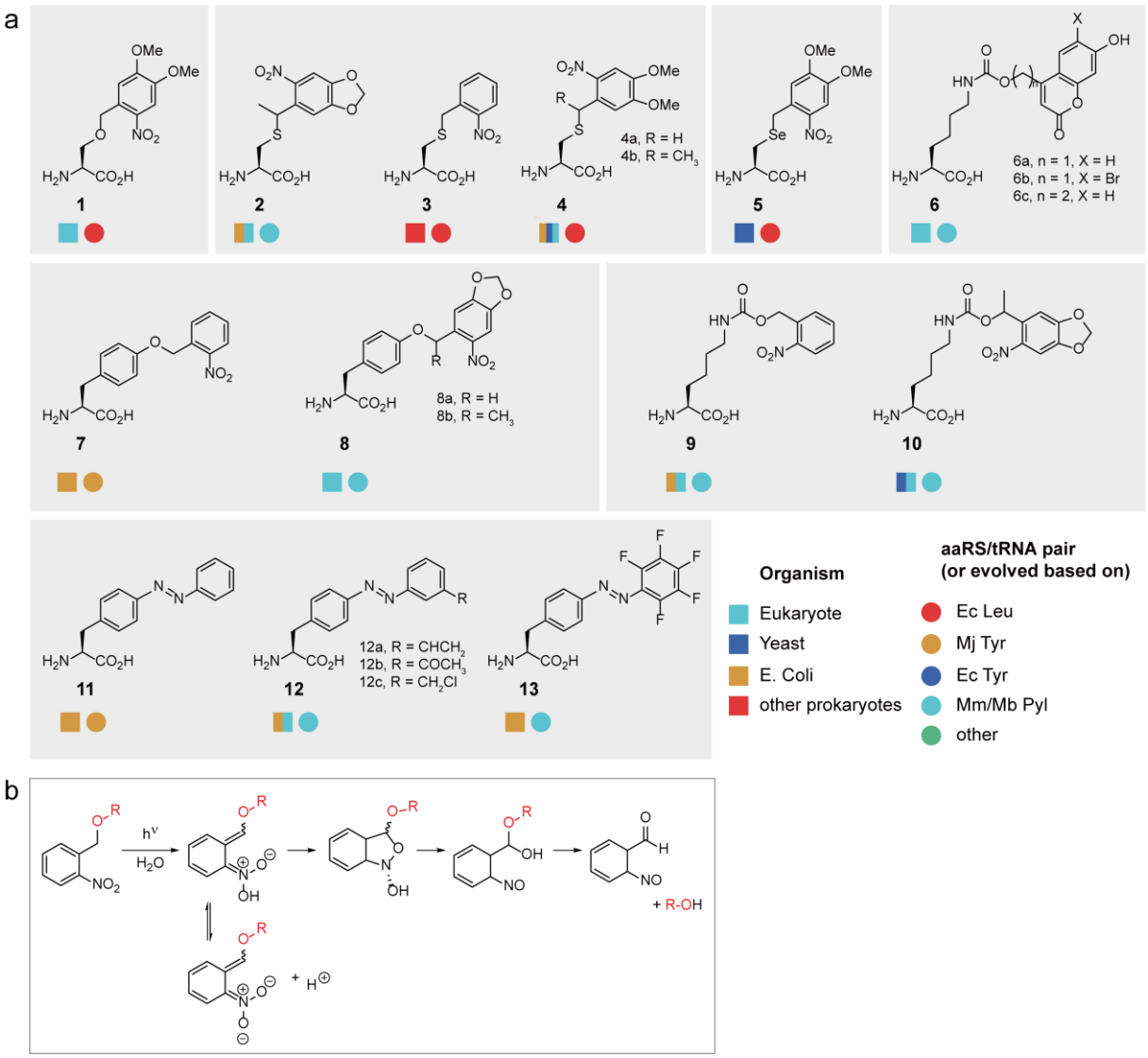


**Figure 17. Examples of photocaged and photoswitch ncAAs**. **a**) Chemical structures of photocaged and photoswitch ncAAs that have been applied in chemical biological studies. Color and shape codes indicate the applicable organism and required amber suppressor pair. Figure created based on (Dumas et al., 2015)(A.4). **b**) Mechanism of *o*-nitrobenzyl photocleavage reaction (Klán et al., 2013).

## 4.3.2. Principles of codon-specific ncAA incorporation via genetic code expansion

With the light-responsive ncAA toolbox in hand, a target protein can be engineered at specific positions to render the desired light response.

Compared to other site-selective protein modification methods, such as solid-phase synthesis and native chemical ligation, codon-specific ncAA incorporation via genetic code expansion is unparalleled in overcoming the challenge of modifying proteins in live cells. Genetic code

expansion allows simultaneous ncAA incorporation into the polypeptide chain during endogenous protein synthesis in response to specifically inserted blank codon(s) on the mRNA, where the most widely used codon is the amber stop codon (UAG).

In the ribosomal polypeptide synthesis, a canonical amino acid is selectively recognized and aminoacylated ("charged") on the cognate tRNA by the aminoacyl-tRNA synthetase (or aminoacyl-tRNA transferase, aaRS) and subsequently transferred onto the growing polypeptide chain if the anti-codon of the cognate tRNA matches the mRNA codon presented in the ribosomal A-site (**Figure 18a**). Therefore, the site-specific ncAA incorporation would require (1) an orthogonal aminoacyl-tRNA synthetase which binds to the ncAA, (2) an orthogonal cognate tRNA which can pair with the orthogonal tRNA synthetase, and (3) a blank codon which can be recognized by the orthogonal tRNA (**Figure 18b**) (reviewed in Lang & Chin, 2014).

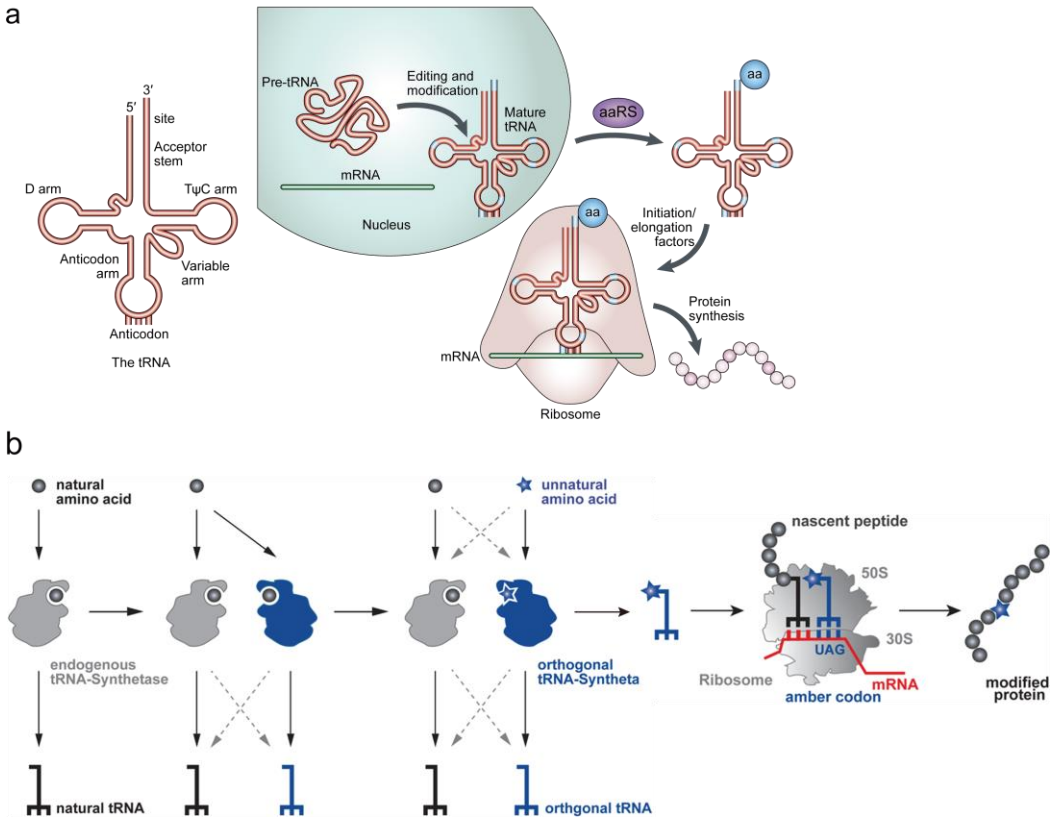


**Figure 18. Overview of the translation machinery and the orthogonal translation of modified protein**. **a**) Schematic illustration of the tRNA secondary structure (major features are marked) and the ribosomal translation process. Figure adapted from (Knight et al., 2001) with permission (A.4). **b**) Schematic illustration of the two-step selection process to evolve orthogonal synthesis/tRNA pairs, and the incorporation of a ncAA into the engineered protein in response to an amber codon. Figure adapted from (Lang & Chin, 2014) with permission (A.4).

***The orthogonality of the ncAA–aminoacyl-tRNA synthetase***

To establish the orthogonality, the chemical structure of the ncAA must be distinguishable from the canonical amino acids and not recognized by the endogenous aminoacyl-tRNA synthetases, while the orthogonal aminoacyl-tRNA synthetase must be capable of binding the ncAA but not canonical amino acids (**Figure 18b**). Most of the orthogonal aminoacyl-tRNA synthetases originally recognize canonical amino acids and require active site evolution to build up the orthogonality, except for the aminoacyl-tRNA synthetase for pyrrolysine (PylRS) (Ibba & Söll, 2002; Polycarpo et al., 2004) which does not recognize any canonical amino acids but can promiscuously bind to a variety of ncAA analogues.

The specificity of ncAA recognition can be further enhanced by engineering the editing domain of the aminoacyl-tRNA synthetase. An editing domain is present in certain aminoacyl-tRNA synthetases to facilitate the correct tRNA aminoacylation by hydrolyzing mischarged aminoacylated tRNAs, whereby mutating critical residues in the editing domain can attenuate the hydrolysis of ncAAs. One example of such strategy is the T252Y mutation in the editing domain of the *E. coli* leucyl tRNA synthetase (LeuRS), which weakens the activity of the editing domain and allows the incorporation of multiple leucine analogues (Tang & Tirrell, 2002).

***The orthogonality of the aminoacyl-tRNA synthetase–tRNA pair***

The orthogonal aminoacyl-tRNA synthetase–tRNA pair must not cross-react with the host aminoacyl-tRNA synthetase–tRNA pair. Meaning, the orthogonal aminoacyl-tRNA synthetase does not charge host tRNAs, and the orthogonal tRNA is not recognized by host aminoacyl-tRNA synthetases (**Figure 18b**). Since the orthogonality of the aminoacyl-tRNA synthetase–tRNA pair reflects the interactions of tRNA or aminoacyl-tRNA synthetase with the host translational components, it is defined with respect to a specific host organism. This orthogonality is typically provided by importing an aminoacyl-tRNA synthetase–tRNA pair from one domain of life into a heterologous host but can also be created *de novo* by directed evolution.

The common aminoacyl-tRNA synthetase–tRNA pairs include: the Pyrrolysyl tRNA synthetase (PylRS)/$^{Pyl}$tRNA from *Methanosarcinae* (*barkeri*, *mazei*, …), the Tyrosyl tRNA synthetase (TyrRS)/$^{Tyr}$tRNA from *Methanococcus jannaschii* and *E. coli*, and the Leucyl tRNA synthetase (LeuRS)/$^{Leu}$tRNA from *E. coli*. Some of these orthogonal pairs have been evolved to incorporate light-responsive ncAAs in mammalian cells. The evolved PylRS/$^{Pyl}$tRNA pairs have been used for encoding photocaged lysines (Chen et al., 2009; Gautier et al., 2010), photocaged cysteine (Nguyen et al., 2014), photocaged tyrosine (Arbely et al., 2012), and photoswitch phenylalanies (Hoppmann et al., 2014). While the evolved LeuRS/$^{Leu}$tRNA pairs have been used for incorporating photocaged serine (Lemke et al., 2007) and cysteine (Kang et al., 2013).

***The orthogonality of the tRNA anticodon–codon pairing***

In nature, all the 64 combinations of mRNA triplet codons are used to encode 20 canonical amino acids and 3 stop signals for translation termination (**Figure 19a**). It is therefore challenging to expand the usage of genetic codons as recoding sense codons (encoding canonical amino acids) will compete with normal protein synthesis, and recoding nonsense codons (stop signals) might disturb the termination machinery. Nonetheless, discovery of deviated codon usage in vertebrate mitochondria and codon reassignments in prokaryote and eukaryote (**Figure 19a**) (reviewed in Ambrogelly et al., 2007; Knight et al., 2001) challenges the concept of nonevolving genetic code from the "frozen accident" hypothesis (F. H. C. Crick, 1968). Stop codon reassignments found in all domains of life shed light on the expanded use of genetic codons. The opal codon UGA is ambiguously used to encode selenocysteine (Sec, **Figure 19b**) for synthesizing selenoproteins in all domains of life despite the limited organisms (Johansson et al., 2005; Leinfelder et al., 1988), while the amber codon UAG is reassigned in methanogenic archaea and bacteria to encode pyrrolysine (Pyl, **Figure 19b**) (Hao et al., 2002; Srinivasan et al., 2002). In the light of the natural expansion of genetic codes, the highest degree of orthogonality can therefore be provided by the least frequent amber codon, which is also the most used codon to date to incorporate ncAAs.

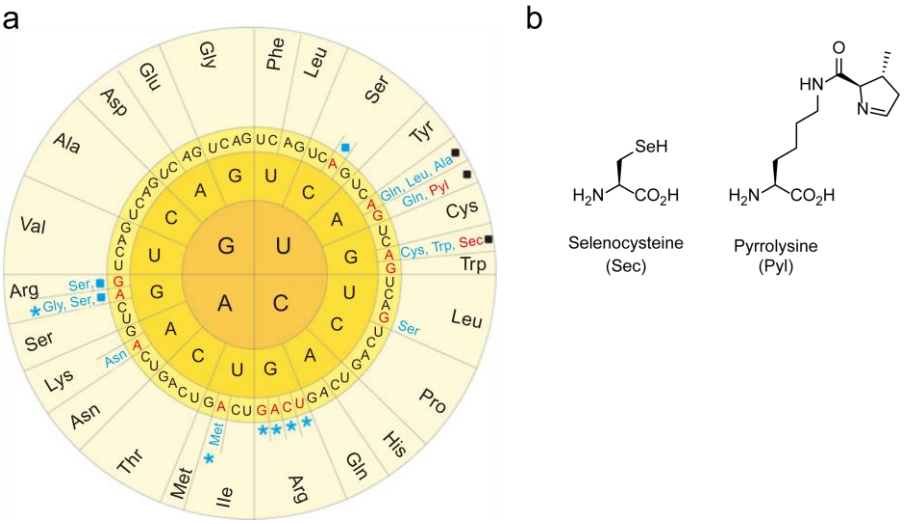


**Figure 19. Summary of codon usage and naturally occurring genetic code expansion**. **a**) The universal genetic codes and codon reassignments in different domains of life. Codons in black are static and not affected by reassignment, while codons in red were found to be reassigned to amino acid residues shown in blue (canonical) or red (noncanonical). Stop codons are indicated with blue squares. Codons marked with blue asterisks are potentially unassigned in some organisms or organelles. Figure adapted from (Ambrogelly et al., 2007) with permission (A.4). **b**) Chemical structures of the naturally occurring noncanonical amino acids, selenocysteine and pyrrolysine.

On the other hand, strategies to enhance the orthogonality has been demonstrated using evolved orthogonal ribosomes. The orthogonal ribosomes are used for the translation of orthogonal mRNAs in parallel to natural ribosomal translation (Rackham & Chin, 2005; Wang et al., 2007).

Since it is not involved in the synthesis of host proteome, the orthogonal ribosomes are flexible to be re-engineered for decoding new blank codons such as the quadruplet codons which are poorly processed by nature ribosomes (Neumann et al., 2010).

### 4.3.3. Application strategies

Photocontrol of protein function at single residue, especially via photocaging, has vast applications in cell biology. It has been used to control the activity of caspase 3 (N. Wu et al., 2004) and β-galactosidase (Deiters et al., 2006), signal transduction (Arbely et al., 2012; Gautier et al., 2011; Lemke et al., 2007; Liaunardy-Jopeace et al., 2017), protein localization (Gautier et al., 2010), gene expression (Chou & Deiters, 2011; W. F. Edwards et al., 2009; Hemphill et al., 2015; Luo et al., 2016), and activating fluorescent or luminescent reporters (Groff et al., 2010; Wilkins et al., 2010; J. Zhao et al., 2013) (reviewed in Ankenbruck et al., 2018; Baker & Deiters, 2014).

One of the most important recent applications of photocaged ncAAs is in studying the epigenetic regulations. The functional studies of epigenetic regulatory elements at chromatin level are often suffered from the challenging druggability and the poor selectivity of small molecule inhibitors. Especially the dynamic and sophisticated nature of chromatin regulation requires tools with high spatial-temporal resolutions to enable kinetic insights uncoupled from background cellular events. In this regard, selective photoactivation of protein function at a user-defined residue provides unparalleled advantages.

Light activation of TET dioxygenases, the epigenetic DNA methylation erasers, has been achieved by genetically encoding a photocaged 4, 5-dimethoxy-2-nitrobenzyl-L-serine (**1**, DMNB-Ser, **Figure 17a**) to replace an active site serine residue S1812 (murine TET2 numbering) responsible for α-KG cofactor binding. The DMNB protecting group serves as a temporary block in the active site to keep TET dioxygenases inactive in cells, and subsequent DMNB removal with light restores the canonical serine residue as well as enzymatic activity. This allows protein inactivation with minimum structural changes. Moreover, only the native protein is involved in the activated biological processes, thus circumvents functional or conformational concerns brought by fusion of photoreceptor domains. The light-activatable TET2 dioxygenase has been demonstrated to activate target gene expression and induce differential transcriptome modulation in mammalian cells in a light-dependent manner. In addition, time-resolved observation revealed the impact of cancer-associated mutations on the kinetics of *in vivo* TET2 catalysis, where the A1505T and P1617H mutations showed striking rate-enhancement over time and hint at additional layers of regulation (Palei et al., 2020).

The light-activatable DNMT enzymes, the DNA methylation writer, were established by

Wolffgramm et al. following a similar strategy. A DMNB-caged cysteine (**4a**, DMNB-Cys, **Figure 17a**) was genetically encoded into the DNMT catalytic site to control the nucleophilicity of residue C710, which activates the pyrimidine base of cytosine for subsequent methylation. Light activation of photocaged DNMT3 illustrated differential de novo methylation kinetics in mammalian cells influenced by frequent DNMT mutations in cancer. Furthermore, a light-activated methylation programming in a target-specific context was showcased by fusing transcription factor-like effector (TALE) domain with photocaged DNMT3 (pcDNMT3). The TALE-pcDNMT3 fusion targeting pericentromeric satellite 3 DNA displayed locus-specific methylation with high sequence specificity as well as clear dose-response to the irradiation time, offering an additional layer of precision to control chromatin regulation (Wolffgramm et al., 2021).

***Limitations***

Despite the unsurpassed strength and precision to control protein function with light in living systems, codon-specific incorporation of light-responsive ncAAs via genetic code expansion has a few limitations.

Firstly, the ability to incorporate multiple different light-responsive residues is limited by the mutually orthogonal aminoacyl-tRNA synthetase–tRNA pairs and the availability of blank codons. Secondly, the application of novel light-responsive ncAAs with improved photochemical properties is restricted by their compatibility with general translational components (elongation factor, ribosome, etc.) and the biological system (solubility, membrane permeability, metabolic stability, etc.). On the other hand, the truncated protein side-products and the potentially mistranslated endogenous proteome due to stop codon suppression remain major concerns of the method.

# 5. Aim of the Work

The main objective of this work is to unravel currently poorly understood MBD-TET interplays, with a particular interest in how the TET dioxygenase activity and the MBD protein function are dynamically regulated in the mammalian genome to achieve a coherent transcriptional program. In light of the lack of mechanistic understanding of MBD-TET interplays and the causalities between dysregulated MBD-TET crosstalk events and methylation-associated human disorders and diseases, this study aims to employ photocontrol on the activity and function of TET dioxygenases and MBD proteins in a human-relevant intracellular model system to enable kinetic insights into their crosstalk in the human genome.

This work intends to investigate the MBD-TET interplay from two aspects: (1) the modulation of the TET dioxygenase enzymatic kinetics by the MBD protein and (2) the kinetics of domain-dependent MBD protein modulation of TET dioxygenase activity. The first part of the study will examine the differential TET-mediated 5mC oxidation kinetics modulated by MBD protein, especially the contribution of individual MBD protein domains, via the light-activatable TET dioxygenase to uncouple prior TET activities. On the other hand, the second part of the study seeks to further elucidate the kinetics of MBD protein modulation dependent on individual domains by uncoupling the methylation reader function through the photocaged MBD protein.

Most importantly, the established system is envisioned to provide a generic strategy for future intracellular mechanistic studies of highly dynamic chromatin regulatory events.

# 6. Light-activatable TET Enables Kinetic Insights into the MBD-TET Interplay

## 6.1. A coexpression screen reveals the MBD1-TET1 interplay

Since the interplay between MBDs and TETs remains largely unexplored, the first goal is to establish a screening strategy and identify potential MBD-TET interactions in living mammalian cells. Literatures have shown that the modulation of MBD proteins often leads to changes in the TET-mediated 5mC oxidation (Ludwig et al., 2017; P. Zhang et al., 2017). Therefore, evaluation of differential *in vivo* TET activities in the presence of MBD proteins would provide a good starting point of the study.

On the other hand, recent studies were primarily conducted in the murine system, namely, studying murine proteins in murine cells. It is thus desirable to gain insights into human-relevant MBD-TET crosstalk events to uncover the mechanistic linkage between aberrant methylation regulation and methylation-associated human diseases, such as neurodegenerative diseases and cancers.

As a result, the design of the MBD-TET interplay screen involves the transient coexpression of full-length human MBD proteins (MBD1, MBD2, MBD3, MBD4, and MeCP2, **Figure 20b**) and the catalytic domain of human TET dioxygenases (TET1, TET2, and TET3, **Figure 20b**) in human embryonic kidney 293T (HEK293T) cells, followed by the flow cytometry (FCM)-assisted global 5hmC quantification at single cell level to evaluate TET activity (**Figure 20a**). Notably, the TET catalytic domains were used in this study not only because they provide stronger 5hmC signals but also because they behave similarly to full-length proteins in an image-based MBD-TET interplay study (P. Zhang et al., 2017). Moreover, the MBD proteins and TET dioxygenases are tagged respectively with EGFP and mCherry at their C-terminus to facilitate orthogonal detection of their cellular expression.

The *in vivo* 5mC oxidation kinetics of different TET catalytic domains (TET CD) were first evaluated in order to determine a suitable time window for the MBD-TET interplay study. The mCherry-tagged human TET catalytic domains (hTET1CD, hTET2CD, and hTET3CD in the following discussion) were separately expressed in HEK293T cells for 16, 24, and 48 hours, then cells were fixed and the TET-mediated 5mC oxidation were detected via immunofluorescence labeling of 5hmC. The global 5hmC level was quantified in all cell populations expressing C-terminal mCherry. hTET1CD and hTET2CD showed prominent increase of 5hmC after 16 hours and comparable enzymatic kinetics, whereas hTET3CD showed much slower oxidation and the 5hmC product only started to increase after 24 hours (**Figure S 2**). Therefore, hTET1CD and hTET2CD are further

screened for their interplay with MBD proteins after 16 hours of coexpression, while hTET3CD was excluded from the study due to its slow oxidation kinetics.

Following the same workflow, the EGFP-tagged human MBD proteins (MBD1, MBD2, MBD3, MBD4, and MeCP2 in the following discussion) or EGFP-only (negative control) were transiently co-expressed with hTET1CD or hTET2CD in HEK293T cells. Cells were fixed 16 hours after transfection and the 5hmC abundance was measured by FCM. The FCM data were additionally grouped by EGFP and mCherry intensities to assess the modulation of TET activity at different MBD/TET expression ratios. In particular, a group defined with medium TET and high MBD expression provides adequate dynamic range and pronounced 5hmC signal over the background and was therefore selected for analysis in the subsequent experiments.

Both the coexpression of MeCP2 and MBD2 led to a slightly decreased 5hmC formation mediated by TET1CD when compared to the EGFP-only negative control (**Figure 20d**), which is consistent with a previous study conducted in HEK293T and mouse myoblasts cells. Coexpression of MBD3 also resulted in slightly reduced 5hmC, while MBD4 did not affect the TET1CD-mediated 5hmC oxidation. On the contrary, coexpression of MBD1 significantly reduced 5hmC formation (**Figure 20b**). This observation surprisingly disagrees with the previously reported colocalization of mouse TET1CD and MBD1 at the murine pericentromeric heterochromatin accompanying an enhanced TET1-mediated 5hmC formation (P. Zhang et al., 2017). The deviation may be due to the differences between human and murine proteins or cell types. In addition, while the reported image-based study emphasized the 5hmC formation at a specific chromatin region, the FCM-based analysis in the current study evaluates the 5hmC abundance at the global level. On the other hand, no significant difference in TET2CD activity was observed in the presence of MBD proteins (**Figure S 3**).

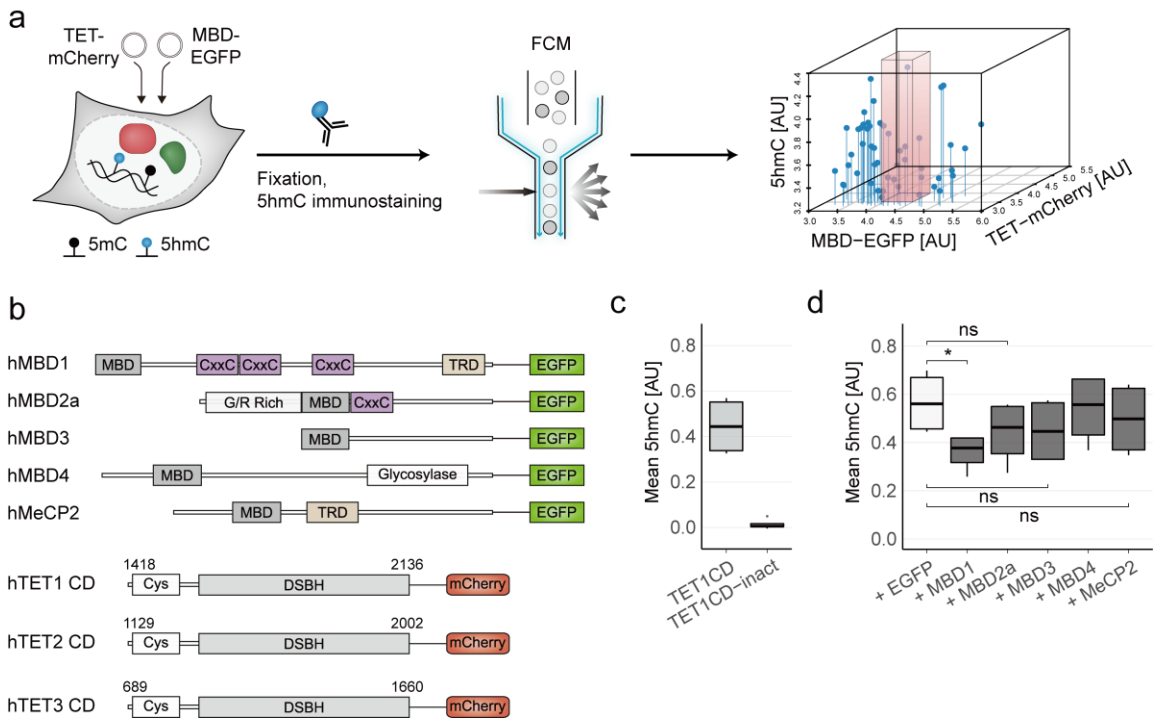


**Figure 20. Study the modulation of TET-catalyzed 5mC oxidation by human MBD proteins**. **a**) Domain structures of the recombinant MBDs and TETs used in this study. The human core MBD family proteins, hMBD1–4 and hMeCP2, are C-terminally tagged with EGFP (CXXC, Cys-x-x-Cys domain; TRD, transcriptional repressor domain). The catalytic domain of human TET dioxygenases, hTET1–3, are C-terminally tagged with m-Cherry (Cys: cysteine-rich domain; DSBH: double-stranded β-helix domain). **b**) Schematic illustration of the FCM-assisted MBD-TET interplay study. Global 5hmC (5mC oxidation product) in single cells is immunostained and measured in FCM as represented in the 3D dot plot, while the red cube indicates the selected cell population for further analysis. **c**) FCM measurement faithfully reflected the differential 5mC oxidation in cells expressing active or inactive TET1CD-mCherry. Measurements were conducted 16 h after transfection. Median intensity of 5hmC immunofluorescence from >100 cells was normalized to the median 5hmC immunofluorescence intensity of the untransfected cell population (**Figure S 1**). Data are from four independent biological replicates. **d**) FCM analysis of cells co-expressing TET1CD-mCherry and different MBD-EGFP constructs. P-values from unpaired student's t-test of four independent biological replicates (*: p <= 0.05; ns: p > 0.05). Figure adapted from (Lin et al., 2022) (A.4).

## 6.2. Light-activatable TET1 enables *in vivo* kinetic studies

Provided the evidence that TET1 activity was significantly downregulated by MBD1, subsequent kinetic studies were focused on the interplay between MBD1 and TET1. A recent work by Palei et al. reported a novel strategy to study *in vivo* TET enzymatic kinetics by direct light activation of TET through genetically encoding a 4,5-dimethoxy-2-nitrobenzyl-L-serine (**1**, **Figure 21a**) (Palei et al., 2020). The photocaged serine derivative **1** replaces an active site residue S2045 in TET1, positioning the 4,5-dimethoxy-2-nitrobenzyl group to introduce steric conflict with α-KG and Fe(II) and thus caging the catalytic activity of TET1 (Palei et al., 2020). In this way, TET1 is expressed in the catalytically inactive state and can be selectively activated with high spatial-temporal precision.

This strategy would uncouple TET enzymatic kinetics from the upstream processes and allow accurate measurement of the modulation by MBD1.

Adopting this approach, a vector encoding hTET1CD with a single in-frame TAG amber codon at S2045 and a mCherry tagged at the C-terminus was constructed to faithfully monitor the expression of full-length hTET1CD by fluorescence (**Figure 21b**). In HEK293T cells co-expressing an evolved *E. coli* amber suppressor leucyl-tRNA-synthetase (LRS)/tRNA$^{Leu}$ pair, mCherry expression was significantly higher in the presence of **1** compared to its absence, indicating high fidelity of incorporation (**Figure 21b** and **Figure S 4**).

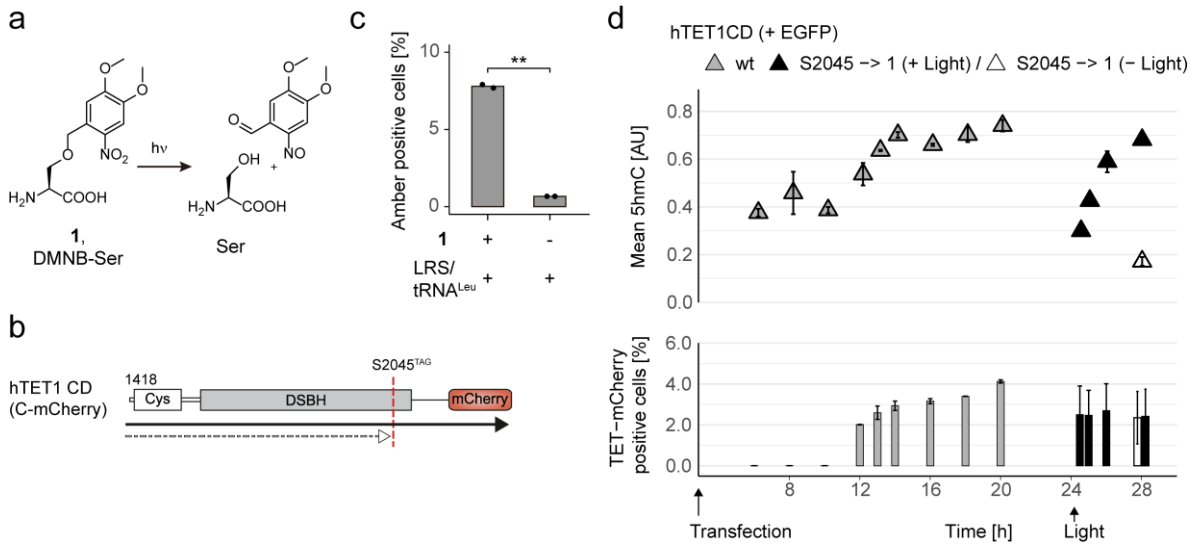


**Figure 21. Light activation of TET1 for studying 5hmC formation kinetics in HEK293T cells**. **a**) Scheme of the photocleavage reaction of **1**. **b**) Domain structure of C-terminally mCherry-tagged hTET1CD-S2045→TAG. **c**) Fidelity of incorporating **1** at hTET1CD S2045→TAG codon analyzed by mCherry signal of HEK293T cells co-expressing the LRS/tRNA$^{Leu}$ pair in the presence or absence of **1**. Data from two independent biological replicates (unpaired t-test, **: p <= 0.01). **d**) Monitoring the expression and activity of hTET1CD in HEK293T cells that grey: express amber-free wt hTET1CD; black: co-express vectors encoding hTET1CD-S2045→TAG and the LRS/tRNA$^{Leu}$ pair in presence of **1** with light irradiation after 24h; white: same as black but without light irradiation (in all cases co-transfected with EGFP-only control). Upper panel shows mean 5hmC intensities of a selected medium TET1 expression group (cell numbers are 30, 42, and 48 for t = 6, 8, and 10 h, respectively; cell numbers are > 1000 for all other time points). Lower panel shows hTET1CD expression by mCherry quantified as % mCherry-positive cells. Error bars from standard error of the mean (SEM) of at least three independent biological replicates. Figure adapted from (Lin et al., 2022) (A.4).

The successful caging and activation of *in vivo* TET1CD activity was further validated in HEK293T cells similarly cotransfected with hTET1CD-S2045 → TAG and LRS/tRNA$^{Leu}$ pair. Cells were incubated with transfection reagents for 3 h and further grew in growth medium containing 0.05 mM **1** for 21 h. The expression of hTET1CD → **1** was terminated by replacing medium with prewarmed DPBS, then cells were irradiated with light (365 nm, 15 W) for 3 min to activate TET1. The cells expressing hTET1CD → **1** showed a sharp 5hmC increase upon light activation, where

an approximate linear response can be observed in the initial 4 h time window. In contrast, 5hmC stays at the basal level even after 4 h in the absence of light irradiation (**Figure 21d**, upper panel, black and white triangles). Moreover, the mCherry fluorescence signal indicated stable expression of hTET1CD → **1** over the 4 h after irradiation, preventing incorrect conclusions on the kinetic measurements due to increasing TET1 expression (**Figure 21d**, lower panel, black and white bars). In comparison to a reference experiment, cells only expressed non-caged wt hTET1CD showed a slow and nonlinear 5hmC increase over 20 h after transfection together with a drastic rise of TET expression at 12 h. This demonstrates the difficulty of measuring accurate kinetics without temporal control of the catalytic activity, primarily due to many rate-limiting steps in the upstream processes including translation, protein folding, and post-translational modification.

To further confirm the fidelity of genomic 5hmC observation, a basic control experiment with or without RNase A treatment was conducted to verify the 5hmC signal obtained from the FCM assay was originated from DNA but not RNA (**Figure S 5**). Moreover, the selected samples in this and later experiments at 2 h after activation showed positive correlation between FCM-based measurement and DNA dot blot assay using the same 5hmC antibody (**Figure S 6**).

Taken together, the ability to uncouple TET catalytic activity enables time-resolved observation of the previously difficult-to-observe regulatory events associated with TET activity.

## 6.3. Kinetic studies reveal the competition between MBD1 and TET1 at mCpGs

With the new tool in hand, the modulation of *in vivo* TET1 catalytic activity by full-length wild type (wt) hMBD1 (**Figure 22a**) can now be studied on the kinetic level. The mCpGs binding ability of the EGFP-tagged recombinant wt hMBD1 was first examined by imaging its nuclear localization in mouse fibroblast NIH/3T3 cells. In these murine cells, the pericentromeric heterochromatin is highly enriched in 5mC and A:T base pairs, resulting in characteristic chromocenter speckles where functional MBD proteins heavily accumulate and are brightly stained by DNA minor groove dyes such as DAPI (Du et al., 2015; X. Wu & Zhang, 2017). The colocalization of wt hMBD1 and DAPI was observed, confirming the mCpG affinity of recombinant hMBD1 (**Figure 22b** and **Figure S 14**).

The kinetic study was conducted in HEK293T cells cotransfected with vectors encoding hTET1CD-S2045 → TAG and the LRS/tRNA$^{Leu}$ pair, as well as a third vector encoding either EGFP-tagged wt hMBD1 or EGFP-only. The TET1CD-mediated 5hmC formation in different MBD/TET expression ratios (**Figure 22c**) was monitored in a time window of 8 hours after light activation. Consistent with the observation in the wild-type MBD-TET coexpression screen, 5hmC was constantly

downregulated over the 8 h time window in the presence of wt hMBD1 compared to the EGFP-only control (**Figure 22d**). In addition, the downregulation dose-dependently correlates with the expression level of wt hMBD1 (**Figure 22e** and **Figure S 7**). Given that both MBD1 and TET1 bind to mCpGs, the dose-dependent downregulation can be explained by reduced accessibility of TET1 to mCpGs due to competing MBD1.

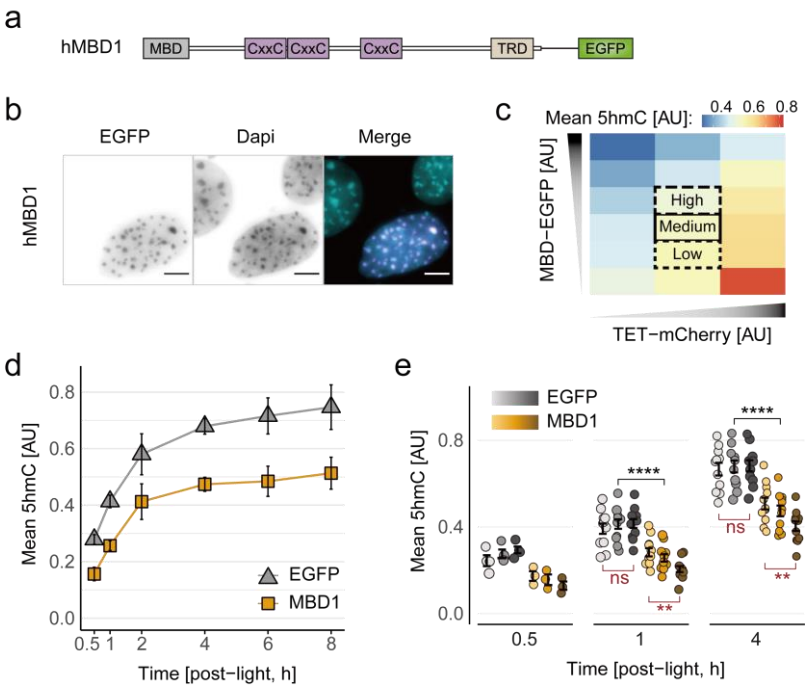


**Figure 22. Kinetic analysis of TET1-mediated 5hmC formation modulated by hMBD1**. **a**) Domain structure of C-terminally EGFP-tagged hMBD1. **b**) Microscopy imaging of wt hMBD1 localization in NIH/3T3 cells. Foci in Dapi staining image indicate the mCpG-rich chromocenters, and the merged image illustrates the co-localization of hMBD1 (magenta) and DAPI foci (cyan). Scale bar: 5 µm. **c**) Exemplary protein expression heat map showing the MBD-TET expression groups selected for 5hmC kinetic analyses. **d**) Kinetic measurements of 5hmC formation mediated by hTET1CD-S2045 → **1** in HEK293T cells co-expressing wt hMBD1 or EGFP-only. Mean global 5hmC intensities from >100 cells in medium MBD/TET expression group are plotted, error bars are from SEM of at least three independent biological replicates. **e**) Analyzing the dose-dependent effect of MBD1 modulated TET1 activity from three different MBD/TET expression groups (gradient bar, from left to right: low, medium, high) at selected time points. Mean 5hmC intensities from at least three independent biological replicates. The P-values from Mann-Whitney test are indicated (*: p <= 0.05; **: p <= 0.01; ****: p <= 0.0001; ns: p > 0.05). Figure adapted from (Lin et al., 2022) (A.4).

To test this hypothesis, two MBD1 mutants lacking mCpG affinity were constructed: one R22C+R44C mutant and one ΔMBD mutant missing the whole MBD domain (hMBD1-R22C+R44C and hMBD1-ΔMBD, **Figure 23a**). Both mutations have been reported to deprive the mCpG affinity of MBD1 (Baubec et al., 2013), and the loss of colocalization with DAPI-bright chromocenters in NIH/3T3 cells was also confirmed (**Figure 23b** and **Figure S 14**). As expected, the hMBD1-R22C+R44C mutant did not downregulate 5hmC formation. Indeed, the 5hmC formation in the initial 2 hours after activation was slightly enhanced and rapidly reached a saturation after 4 h

(**Figure 23c**). This surprising finding was further supported by analyzing three different MBD/TET expression ratios as above (**Figure 22c**). The upregulation of 5hmC formation at the first 1 h showed a dose-dependent correlation that was significant at 0.5 h with the expression level of hMBD1-R22C+R44C, whereas the EGFP-only control did not show any dose-dependency (**Figure 23d** and **Figure S 7**).

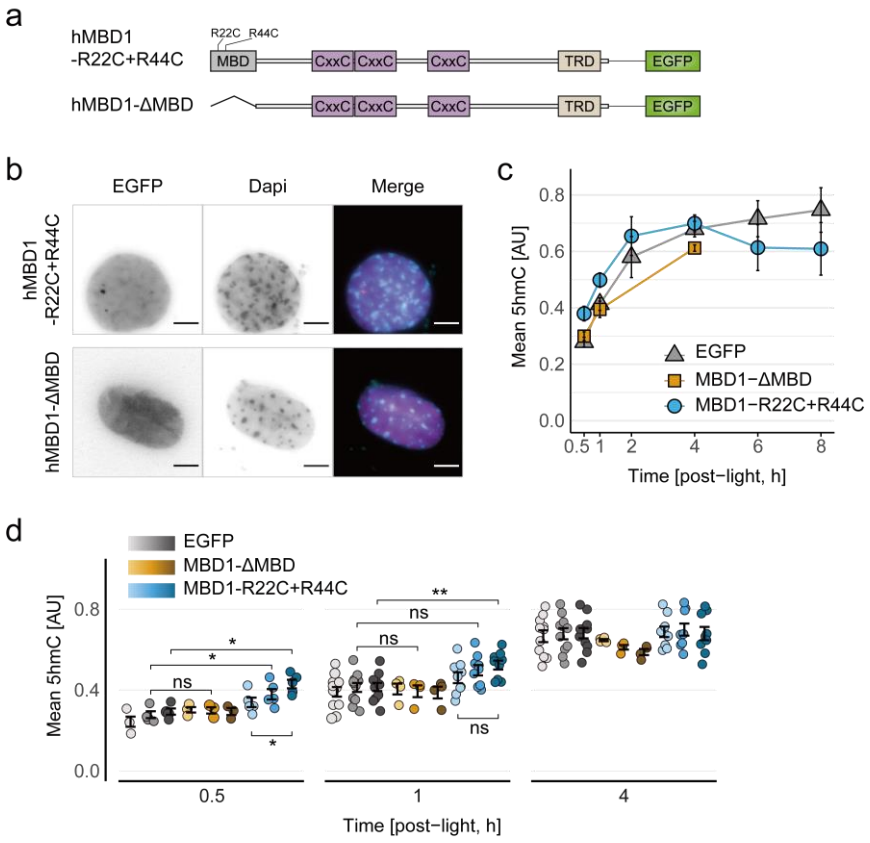


**Figure 23. Kinetic analysis of TET1-mediated 5hmC formation modulated by hMBD1 lacking 5mC binding affinity**. **a**) Domain structure of C-terminally EGFP-tagged hMBD1 R22C+R44C and ΔMBD mutants. **b**) Microscopy imaging of hMBD1 R22C+R44C and ΔMBD mutants in NIH/3T3 cells stained with DAPI as in Figure 22**b**. Scale bar: 5 µm. **c**) Kinetic measurements of 5hmC formation mediated by hTET1CD-S2045 → **1** in HEK293T cells co-expressing either hMBD1-R22C+R44C, hMBD1-ΔMBD, or EGFP-only. Mean global 5hmC intensities from cell populations in the medium MBD/TET expression group (>100 cells) are plotted, error bars indicate SEM of at least three independent biological replicates. **d**) The dose-dependent modulation effects from three different MBD/TET expression ratios (gradient bar, from left to right: low, medium, high) at selected time points. Mean 5hmC intensities from >3 independent biological replicates. The P-values from Mann-Whitney test are indicated (*: p <= 0.05; **: p <= 0.01; ns: p > 0.05). Figure adapted from (Lin et al., 2022) (A.4).

Interestingly, the hMBD1-ΔMBD mutant did not lead to dose-dependently promoted 5hmC formation kinetics as the hMBD1-R22C+R44C mutant but behaved similarly to the EGFP-only control (**Figure 23c** and **Figure 23d**). This result not only confirms the MBD domain-mediated mCpG competition with TET1 but also implies a role of the intact MBD domain in the enhanced TET1 activity by hMBD1-R22C+R44C. Interestingly, coimmunoprecipitation (co-IP) experiments

in a previous study have shown that mouse MBD1 (mMBD1) directly interacts with mouse TET1 (mTET1) both *in vitro* (excluding the DNA-bridging effect) and *in vivo* (expression in HEK293T cells). The interaction involves multi-subdomains of MBD1 and is the strongest in the presence of an intact MBD domain, consistent with the kinetic observations in the current study where hMBD1-ΔMBD did not upregulate TET1 activity (P. Zhang et al., 2017).

It is noteworthy that without light irradiation the cellular 5hmC abundance was not affected by the coexpression of hMBD1 or its mutants, indicating that the 5hmC formation is stringently controlled by the activation of hTET1CD-S2045 → **1** (**Figure S 8**). Additionally, no differences in 5hmC formation kinetics were observed between hMBD1-expressing cells sorted for the N-terminal Flag tag or the C-terminal mCherry of hTET1CD- S2045 → **1**. This provides evidence that the expression of C-terminal truncated hTET1CD (due to termination at amber codon) has no impact on the apparent enzymatic kinetics of the correctly amber-suppressed full-length hTET1CD, for instance, via interactions with hMBD1 (**Figure S 9**).

Overall, these observations suggest that the mCpG competition by a functional MBD domain might not be the only factor in the MBD1 modulated TET1 enzymatic kinetics.

## 6.4. Kinetic studies hint at a role of CXXC3 domain in promoting TET1 activity

The regulatory roles of other hMBD1 domains were further investigated, in which the third CXXC (CXXC3) zinc finger domain that is known for its selective binding to unmethylated CpGs (Jørgensen et al., 2004) is of particular interest and was first interrogated. Two hMBD1 variants which their CXXC3 domains are unable to bind unmethylated CpGs were constructed, one is a natural hMBD1 isoform lacking the complete CXXC3 domain (isoform 7, hMBD1v7, **Figure 24a**) (Fujita et al., 1999b), the other one is mutated at two cysteine residues that are responsible for Zn(II) coordination and CpGs recognition (C338A+C341A, **Figure 24a**) (Clouaire et al., 2010; P. Zhang et al., 2017). Imaging both variants in NIH/3T3 cells confirmed their chromocenter localization owing to functional MBD domains (**Figure 24b** and **Figure S 14**). Moreover, both hMBD1 CXXC3 variants downregulated TET1 oxidation kinetics to the same extent as wt hMBD1 (**Figure 24c** and **Figure S 10**), indicating that the absence of CXXC3 did not influence the competition for mCpGs.

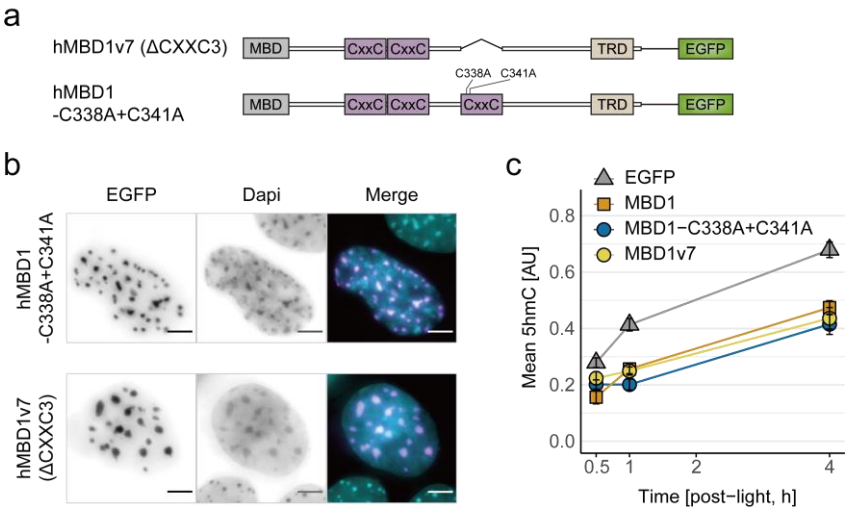


**Figure 24. Kinetic analysis of TET1 activity modulated by hMBD1 deprived of CXXC3 function**. **a**). Domain structures of C-terminally EGFP-tagged hMBD1 CXXC3 mutants (isoform 7 and C338A+C341). **b**) Microscopy imaging of hMBD1 CXXC3 mutants (isoform 7 and C338A+C341) in DAPI-stained NIH/3T3 cells as in Figure 22**b**. Scale bar: 5 µm. **c**) Kinetics of 5hmC formation in HEK293T cells co-expressing hTET1CD-S2045 → **1** and either an hMBD1 CXXC3 mutant (isoform 7 and C338A+C341), wt hMBD1, or EGFP only. Mean global 5hmC intensities from cell populations in the medium MBD/TET expression group (>100 cells) are plotted, error bars indicate SEM of at least three independent biological replicates. Figure adapted from (Lin et al., 2022) (A.4).

To study the role of the CXXC3 domain independently of the strong mCpGs competition of the MBD domain, the MBD domain mutations were again introduced to remove the mCpG affinity. Two new CXXC3 variants bearing the R22C+R44C mutations were constructed: the R22C+R44C+C338A+C341A variant and the isoform7-R22C+R44C variant (**Figure 25a**). Compared to the R22C+R44C mutant, neither of the two new variants bound chromocenters but primarily localized in the nucleoli, implying the role of the CXXC3 domain to anchor hMBD1 to nucleus regions other than chromocenters (**Figure 25b**, **Figure S 14**, and **Figure S 11**). In addition, both variants intriguingly behaved like the EGFP-only control in regulating TET1 activity and did not show dose-dependently enhanced 5hmC formation as the R22C+R44C mutant, which has a functional CXXC3 domain (**Figure 25c** and **Figure S 12**). This observation suggests that the CXXC3 domain and its affinity to the non-methylated CpG regions promote TET1-mediated 5hmC formation.

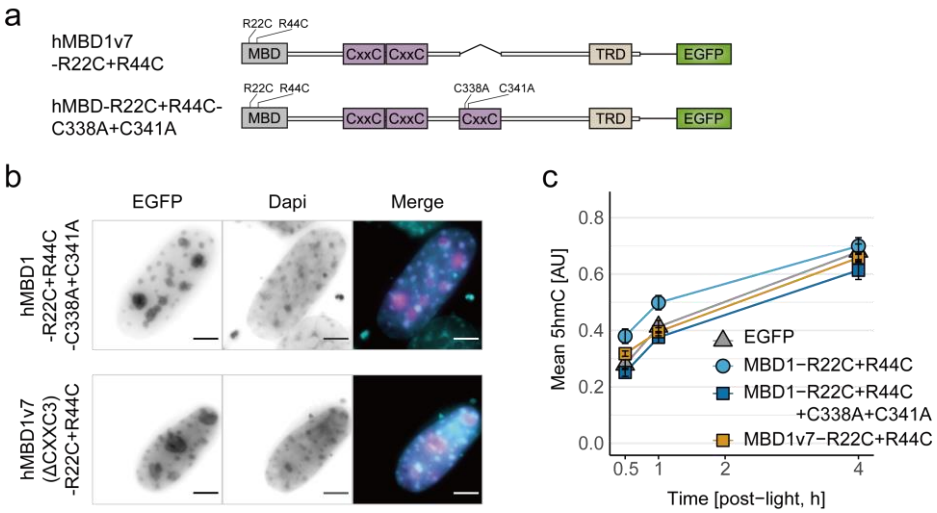


**Figure 25. Role of the CXXC3 domain of hMBD1 in modulating TET1 activity independent of 5mC recognition**. **a**) Domain structures of C-terminally EGFP-tagged hMBD1 mutants (isoform7-R22C+R44C and R22C+R44C+C338A+C341). **b**) Microscopy imaging of hMBD1 mutants (isoform 7-R22C+R44C and R22C+R44C+C338A+C341) in NIH/3T3 cells stained with DAPI as in Figure 22**b**. Scale bar: 5 µm. **c**) Kinetics of 5hmC formation in HEK293T cells co-expressing hTET1CD-S2045 → **1** and hMBD1 mutants (R22C+R44C, R22C+R44C+C338A+C341A, isoform7-R22C+R44C) or EGFP only. Mean global 5hmC intensities from cell populations in the medium MBD/TET expression group (>100 cells) are plotted, error bars are plotted with SEM from at least three independent biological replicates. Figure adapted from (Lin et al., 2022) (A.4).

Interestingly, a previous imaging study has reported a CXXC3-dependent recruitment of mTET1 to pericentromeric heterochromatin (chromocenters) by mMBD1 in mouse fibroblast cells which resulted in locally enhanced 5mC oxidation (P. Zhang et al., 2017). Altogether, the presented data agree with the previous study in a mouse model and further elucidated the CXXC3-dependent MBD1-TET1 interplay at the kinetic level.

## 6.5. The role of TRD domain in regulating TET1 activity

Finally, the role of the transcriptional repressor domain (TRD) of hMBD1 was investigated. The TRD of MBD1 has been shown to interact and recruit multiple chromatin regulatory proteins to mediate condensation and transcriptional silencing (Hameed et al., 2014; Ng et al., 2000). For instance, it interacts with MBD1 Chromatin Associated Factor 1 (MCAF1) and recruits Histone-lysine N-methyltransferase SETDB1 to form heterochromatin protein 1 (HP1) condensed heterochromatin (Fujita, Watanabe, Ichimura, Ohkuma, et al., 2003; Ichimura et al., 2005). Therefore, TRD domain-mediated chromatin condensation may reduce DNA accessibility and downregulate TET1-mediated 5mC oxidation in addition to the competition by the MBD domain.

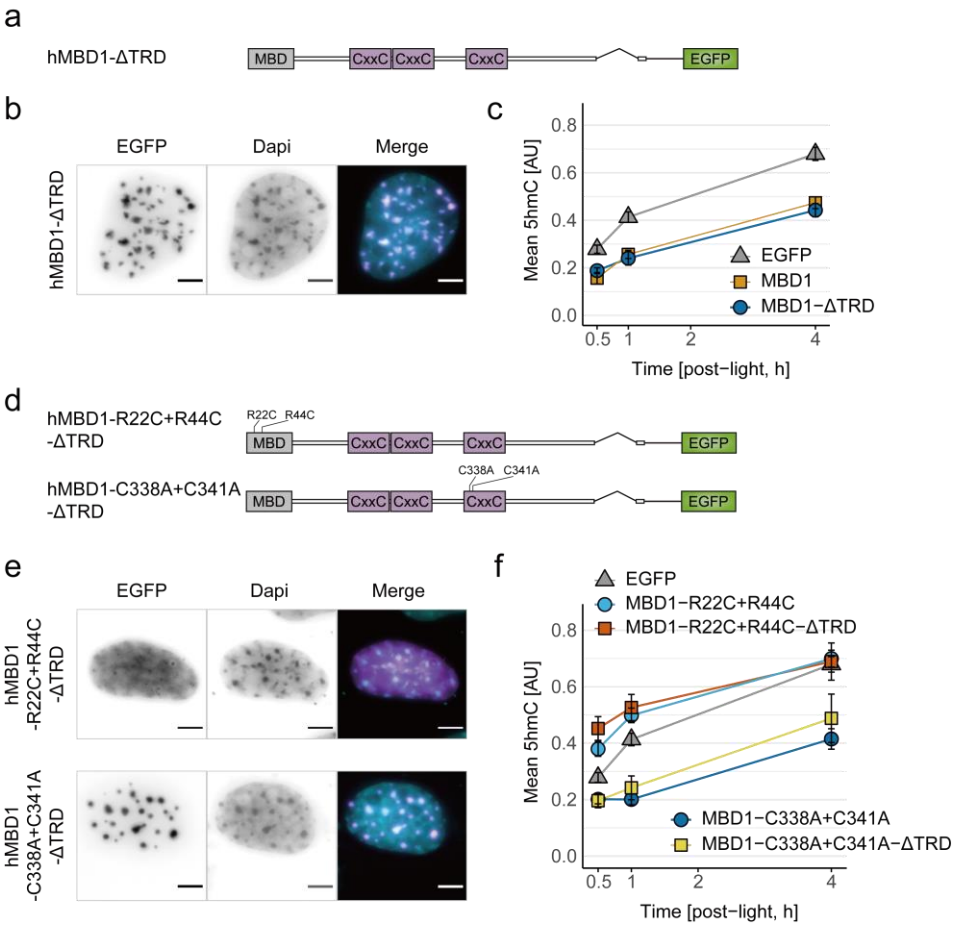


**Figure 26. Role of the TRD domain of hMBD1 in modulating TET1 activity**. **a**) Domain structures of C-terminally EGFP-tagged hMBD1-ΔTRD mutant. **b**) Microscopy imaging of hMBD1-ΔTRD in DAPI-stained NIH/3T3 cells as in Figure 22**b**. Scale bar: 5 µm. **c**) Kinetics of 5hmC formation in HEK293T cells co-expressing hTET1CD-S2045 → **1** and either hMBD1-ΔTRD, wt hMBD1, or EGFP only. Mean global 5hmC intensities from cell populations in the medium MBD/TET expression group (>100 cells) are plotted, error bars show SEM from at least three independent biological replicates. **d**) Domain structures of C-terminally EGFP-tagged hMBD1 ΔTRD mutants (R22C+R44C-ΔTRD and C338A+C341A-ΔTRD). **e**) Microscopy imaging of hMBD1ΔTRD mutants (R22C+R44C-ΔTRD and C338A+C341A-ΔTRD) in DAPI-stained NIH/3T3 cells as in Figure 22**b**. Scale bar: 5 µm. **f**) Kinetics of 5hmC formation mediated by hTET1CD-S2045 → **1** in HEK293T cells co-expressing either hMBD1 ΔTRD mutants (R22C+R44C-ΔTRD and C338A+C341A-ΔTRD), R22C+R44C mutant, C338A+C341A mutant, or EGFP only. Mean global 5hmC intensities from cell populations in the medium MBD/TET expression group (>100 cells) are plotted, error bars indicate SEM from more than three independent biological replicates. Figure adapted from (Lin et al., 2022) (A.4).

To interrogate the impact of the TRD domain, the complete domain (hMBD1-ΔTRD) was truncated since most of the TRD domain is unstructured and multiple residues interact with different chromatin factors (**Figure 26a**) (Hameed et al., 2014). The hMBD1-ΔTRD variant exhibited chromocenter localization in NIH/3T3 cells as expected for its functional MBD domain (**Figure 26b** and **Figure S 14**). Moreover, the ΔTRD variant downregulated the TET1 oxidation kinetics to the same extent as wt hMBD1 in a dose-dependent manner (**Figure 26c** and **Figure S 13a**), implying that TRD did not contribute to additional reduction of TET activity. This observation can

be explained by the masking effect of substrate competition by the MBD domain or because TRD is not yet involved in the interplay within the observation time window in the study. However, literature reported co-IP experiments have shown that a short fragment of C-terminal mouse MBD1 containing TRD domain can still interact with mouse TET1 (P. Zhang et al., 2017). It is thus intriguing whether the direct TRD domain-TET1 interaction is required for the TET1 upregulation involving the CXXC3 domain. New ΔTRD variants with additional R22C+R44C or C338A+C341A mutations were constructed and compared with the same MBD (R22C+R44C) and CXXC3 (C338A+C341A) mutants bearing complete TRD domains for their cellular localization as well as TET1 kinetics (**Figure 26d**). The chromocenter localization (**Figure 26e** and **Figure S 14**) and the downregulated TET1 oxidation kinetics (**Figure 26f** and **Figure S 13c** ) of the C338A+C341A mutants did not exhibit differences in the absence or presence of TRD domain. Similarly, both the R22C+R44C mutants showed enhanced 5hmC formation kinetics but without a significant difference between the absence or presence of the TRD domain (**Figure 26e**, **Figure 26f**, **Figure S 13b**, and **Figure S 14**).

These results suggest two scenarios: (1) the direct TRD domain-TET1 interaction does not contribute to the CXXC3-dependent TET1 upregulation within the observation time window but is potentially a longer process, or (2) the reported interaction only occurs in the specific context of the study, which used murine proteins and a short TRD fragment instead of the full-length MBD1.

## 6.6. Conclusion

The presented study aims to explore the interplay between human MBD proteins and human TET dioxygenases and obtain kinetic insights on their dynamic regulation in mammalian cells. The content of this chapter has been published in Lin et al., 2022.

In the coexpression screening of hTET1 activity with each of the five core MBD family proteins, hMBD1 was found to negatively regulate hTET1 activity. The light-activatable TET1 that was genetically encoded with a photocaged serine further enabled the precise temporal control of the hTET1 activity, allowing the upstream processes (such as translation, post-translational modification, and localization) to be uncoupled from the kinetic observation.

A general trend of reduced 5mC oxidation rate mediated by hTET1 in the presence of the functional MBD domain was observed. This downregulation exhibited dose-dependency and is independent of the function and occurrence of the CXXC3 and TRD domains, which can be explained by the competitive occupancy of the MBD domain at mCpG sites that blocks the access of hTET1. On the other hand, hMBD1 variants bearing a nonfunctional MBD domain but able to bind nonmethylated CpGs (R22C+R44C mutant) surprisingly promoted the TET1 oxidation

kinetics in a dose-dependent manner. This provides a hint of a secondary modulation role of hMBD1 in the interplay which is veiled by the dominant competitive behavior of the MBD domain. Nonetheless, the enhanced TET1 activity indeed depends on the presence of an intact MBD domain, implying that MBD domain per se is involved in this secondary upregulation.

Intriguingly, a correlative previous study that investigated the interplay of murine MBD1 and TET1 proteins without light control in mouse fibroblast cells did not report the dominant downregulation observed in the current study. Instead, a mMBD1 CXXC3-dependent mTET1 localization to pericentromeric heterochromatin accompanied by enhanced local 5hmC formation was observed. Further evidence provided by co-IP experiments suggested that MBD domain is involved in the direct interaction between mMBD1 and mTET1 (P. Zhang et al., 2017). Despite the contradictory downregulation effect observed in the presence of the functional MBD domain, the CXXC3-dependent TET1 activity enhancement concluded in this study is nevertheless consistent with the previous report. Given that unmethylated CpGs are the final product of TET-mediated active demethylation process, CXXC domain might play a role in the product-dependent feedback loop to reinforce TET1 activity (Du et al., 2015). In fact, the canonical full length TET1 sequence also has a N-terminal CXXC domain capable of binding unmethylated CpGs, whereas a shorter TET1 isoform lacking the N-CXXC domain is exclusively expressed in somatic cells. This study only examined the modulation and activity of the TET1 catalytic domain; however, the full-length mTET1 behaved similarly to the catalytic domain in the previous imaging-based study (P. Zhang et al., 2017). It is possible that the CXXC domain of MBD1 functions complementary to the N-CXXC domain of TET1 in regulating the product feedback loop as part of the reader-editor crosstalk.

Interestingly, TET1 activity was not modulated by TRD in this study when comparing hMBD1 variants with or without the TRD domain, whereas a previous study has shown that the TRD domain directly interacts with TET1 (P. Zhang et al., 2017). This observation implies that the TRD-mediated chromatin crosstalk and condensation are not involved in regulating TET1 activity in the model and observation time window in this study.

# 7. Light-activatable MBD1 Reveals the Domain-dependent Cellular Binding Kinetics

## 7.1. Design of a light-activatable MBD1

The last chapter (Chapter 6) discussed the kinetic insights of MBD1-TET1 interplay with respect to TET1 catalytic activity; however, the kinetics of MBD1 domain-dependent modulation remains a missing puzzle piece for completing the picture of MBD1-TET1 interplay at the 5mC substrate. In this regard, a light-activatable MBD1 would enable the 5mC reader activity to be uncoupled from other functions and provide insights into the rate and order of domain-dependent regulatory events.

The study presented in this chapter adopts the approach in the design of light-activatable TET1 to temporally block the 5mC binding affinity of MBD1. In this approach, the 5mC recognition is disrupted by caging an essential serine, cysteine, or threonine residue in the binding pocket with a bulky photolabile 4,5-dimethoxy-2-nitrobenzyl group and can be restored upon light-induced uncaging reaction. Structure analysis showed that two MBD domain amino acids in hMBD1, S45 and T27, provide critical interactions for the 5mC recognition or DNA binding and thus are potent "cageable" residues (**Figure 27a**) (Ohki et al., 1999, 2001). The S45 residue belongs to a hydrophobic pocket of the MBD domain that mediates the interactions with one 5mC, in which its aliphatic side chain provides hydrophobic contact with the methyl group (**Figure 27b**), and serine to alanine mutation at this residue leads to a reduction in DNA binding (Ohki et al., 2001). On the other hand, the T27 residue in loop L1 interacts with the DNA backbone by donating hydrogen bonds and mediating hydrophobic contacts (**Figure 27b**) (Ohki et al., 2001). Therefore, displacing either of the two residues with photocaged serine **1** could potentially introduce steric clash and interrupt the contacts between MBD and 5mC-containing DNA (**Figure 27c** and **Figure 27d**). Moreover, S45 residue is conserved across the human core MBD family proteins (hMBD1–4 and hMeCP2), whereas T27 occurs as either threonine (hMBD1 and hMBD4) or serine (hMeCP2, hMBD2, and hMBD3, notably also mMBD1) (**Figure 7c**). The functional equivalence of threonine and serine at position 27 was verified using the hMBD1-T27S mutant. This mutant showed typical chromocenter localization in NIH/3T3 cells and downregulated TET1 catalytic activity to the same extent as wt hMBD1 (**Figure S 15** and **Figure S 16**), indicating that substituting T27 with photocaged serine **1** will not affect the 5mC recognition and function of MBD1 after uncaging.

As a result, vectors encoding full-length hMBD1 with a single in-frame TAG amber codon at either S45 or T27 and C-terminally tagged with EGFP were constructed (**Figure 27e**). Thereof the EGFP expression enables the monitoring of faithful incorporation of **1** in the full-length hMBD1. In HEK293T cells co-expressing the LRS/$tRNA^{Leu}$ amber suppressor pair, both hMBD1-S45 → TAG

and hMBD1-T27 → TAG constructs demonstrated high fidelity of incorporation by increased EGFP expression in the presence of **1** (**Figure 27f** and **Figure S 17**). Light activation of MBD reader function was further tested in live NIH/3T3 cells by time-lapse imaging. In addition, a hMBD1-R22C+R44C mutant with S45 → **1** that will maintain non-chromocenter and distributed nuclear localization after light irradiation was used to control the fidelity of mCpG binding after uncaging in situ. To enable orthogonal observation in cells, additional vectors with the same amber mutations but replaced the C-terminal EGFP-tags with mCherry-tags were constructed, including hMBD1-S45 → TAG, hMBD1-T27 → TAG, and hMBD1-R22C+R44C-S45 → TAG (**Figure 27g**). The NIH/3T3 cells were co-transfected with either of the mCherry-tagged hMBD1 with S45 or T27 encoded with TAG, the EGFP-tagged hMBD1-R22C+R44C-S45 → TAG, and the LRS/$tRNA^{Leu}$. Cells were incubated with transfection reagents for 3 h and further grew in growth medium containing 0.05 mM **1** for 21 h, then the cell population expressing full-length hMBD1s incorporated with **1** was selected by FACS (fluorescence-activated cell sorting) according to the mCherry or EGFP fluorescence (**Figure S 17**). The sorted cells were recultivated on coverslips for 4 to 6 hours until they adhered to the surface and showed flattened and spread cell bodies. Then DNA was stained with membrane-permeable SiR-DNA (SiR-Hoechst) dye for time-lapse live imaging. Notably, SiR-DNA can be excited by red light and thus is a favorable alternative to prevent simultaneous uncaging of **1** during UV/blue light excitation of traditional DNA stains such as Hoechst 33342 (Lukinavičius et al., 2015).

Interestingly, while hMBD1-S45 → **1** completely lost the chromocenter localization (**Figure 27h**), hMBD1-T27 → **1** showed colocalization with DAPI-bright regions like wt hMBD1 (**Figure S 18**). The result indicates that introducing **1** at S45 can successfully block the MBD reader function but replacing T27 with **1** did not cause steric conflict for 5mC binding. Moreover, hMBD1-S45 → **1** rapidly translocated to chromocenter upon irradiation with violet light (395/25 nm, LED light source of the microscope) for 50 ms (**Figure 27h** and **Figure S 19**), whereas the chromocenter localization of hMBD1-T27 → **1** remained unchanged (**Figure S 18**). On the other hand, hMBD1-S45 → **1** maintained dispersed in the nucleus at dark for at least 20 min before irradiating with light (**Figure 27h**), and the cotransfected hMBD1-R22C+R44C with S45 → **1** maintained the same dispersed distribution both before and after light irradiation (**Figure 27h**). These internal controls validated the faithful light activation of hMBD1-S45 → **1** and proved that violet light alone does not modulate the localization of hMBD1 variants in cells. To further verify the light-activated translocation does not depend on the C-terminal fluorophores, a fluorophore switch experiment was conducted by coexpressing the EGFP-tagged hMBD1-S45 → **1** and mCherry-tagged hMBD1-R22C+R44C with S45 → **1** in NIH/3T3 cells. The alternatively tagged hMBD1 caged variants showed similar behavior as the previous light activation experiments, suggesting that the light-controlled binding behavior is independent of the C-terminal tags (**Figure S 20**).

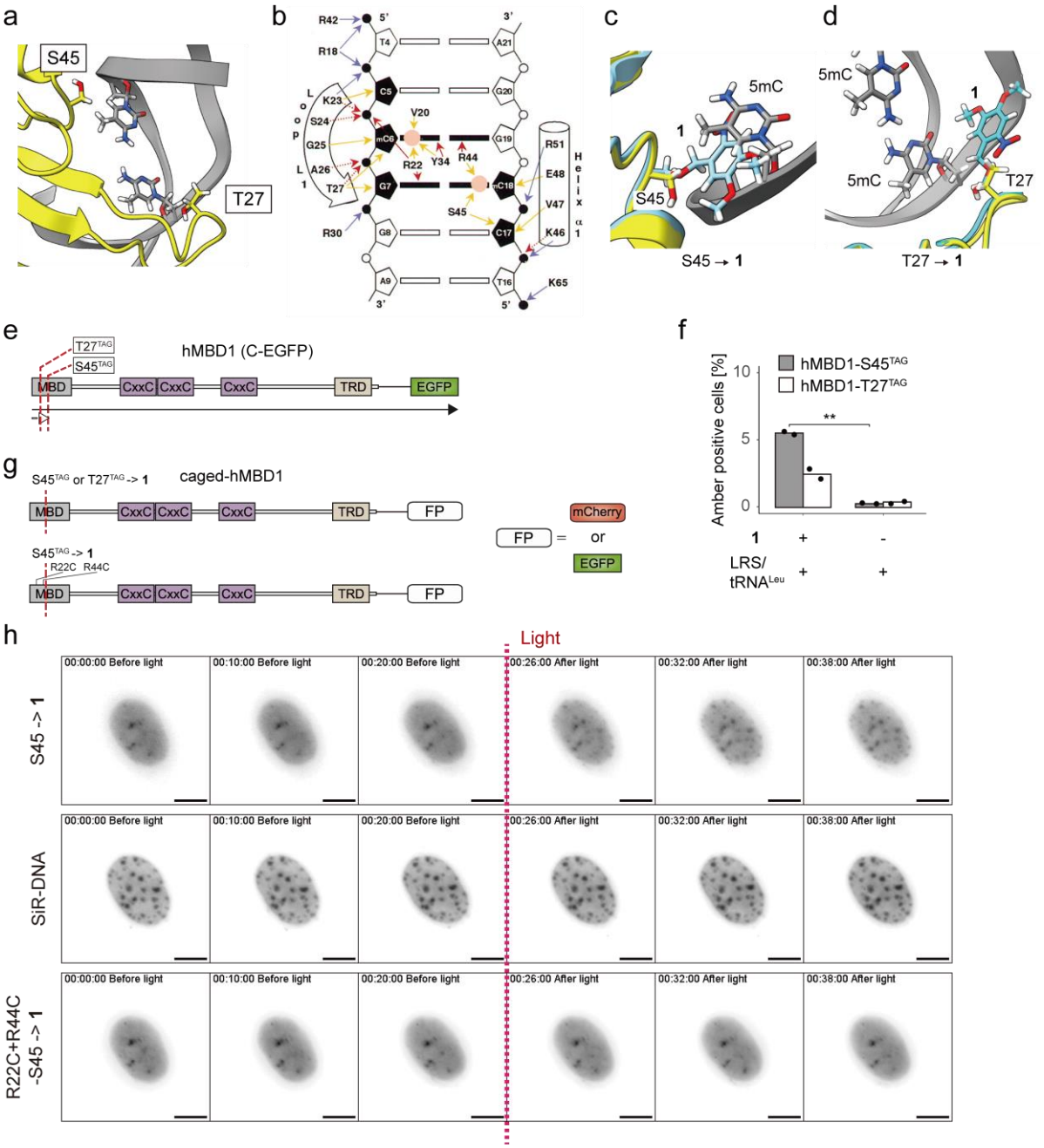


**Figure 27. Controlling the mCpG binding of hMBD1 by light. a**) Positions and orientations of S45 and T27 indicated in the solution structure of wt hMBD1 in complex with 5mC-containing DNA (PDB 1IG4, (Ohki et al., 2001)). Yellow: MBD domain of wt hMBD1; gray: mCpG-containing oligonucleotide. Molecular graphic produced with UCSF ChimeraX (Goddard et al., 2018; Pettersen et al., 2021). **b**) Schematic summary of proposed hMBD1 protein-DNA interactions. DNA bases are shown as rectangular boxes with two methyl groups in mCpG sites marked by pink circles. The deoxyribose rings of DNA are indicated as pentagons and the phosphates as circles. Arrows representing protein-DNA hydrogen bond (red), hydrophobic contact (yellow), and electrostatic interaction (blue), in which hydrogen bonds involving protein amide backbone are shown in red dashed arrows. Figure adapted from (Ohki et al., 2001) with permission (A.4). **c**)**d**) Energy minimized structure of **c**) hMBD1 S45→**1** (cyan) and **d**) hMBD1 T27→**1** (cyan) superimpose with the solution structure of hMBD1-DNA complex (yellow, PDB 1IG4, (Ohki et al., 2001)). Molecular graphics produced by UCSF ChimeraX (Goddard et al., 2018; Pettersen et al., 2021). **e**) Domain structure of vectors

encoding hMBD1 with amber mutation at either S45 or T27 and tagged with C-terminal EGFP. **f**) Fidelity of incorporating **1** at hMBD1 S45 → TAG (gray) or T27 → TAG (white) codon by analyzing EGFP signals in HEK293T cells co-expressing the LRS/tRNA$^{Leu}$ pair in the presence or absence of **1**. Data from two independent biological replicates (unpaired t-test, **: p <= 0.01). **g**) Domain features of vectors encoding S45 → TAG or T27 → TAG mutated hMBD1 wt and R22C+R44C mutant C-terminally tagged with either mCherry or EGFP. **h**) Imaging the simultaneous light activation of mCherry-tagged hMBD1-S45 → **1** and EGFP-tagged hMBD1-R22C+R44C-S45 → **1** in live NIH/3T3 cells. Red dashed line indicating the time point of light treatment. Nucleus is stained with SiR-DNA, intense SiR-DNA foci indicating the mCpG-rich chromocenters. Selected images at different time points (minutes) before and after light activation are shown. Scale bar: 5 μm.

Altogether, the present study developed a novel light-activatable DNA methylation reader by genetically encoding the S45 residue with a photocaged serine **1**. Experiments showed that light irradiation faithfully controls the 5mC reading ability of hMBD1-S45 → **1**, which will further enable cellular kinetic studies on their domain-dependent regulatory events. It is noteworthy that the MBD domain is highly conserved within the core MBD family proteins, giving the potential to photocage other MBD proteins at the same residue.

## 7.2. Light-activatable MBD1 reveals domain-dependent binding kinetics

The light-activatable hMBD1 provides a new tool to temporally uncouple the 5mC binding affinity from the MBD1 domain-dependent modulation of TET1 activity. Particularly interesting are the photocaged hMBD1 domain variants that would allow dissecting the impacts of the CXXC3 and the TRD domain on the mCpG binding kinetics, which could further imply their dynamic modulation in the MBD1-TET1 interplay.

I created vectors encoding an in-frame TAG codon at S45 on the C-terminal EGFP-tagged hMBD1-C338A+C341 and hMBD1-ΔTRD variants (**Figure 28a** and **Figure S 17**), as well as additional vectors encoding the same photocaged hMBD1 domain variants but tagged with mCherry at the C-terminus for fluorophore switch control experiments (**Figure 28a**). Following the same workflow described before (section 7.1), either the EGFP-tagged wt hMBD1, hMBD1-C338A+C341A, or hMBD1-ΔTRD variant with S45 → **1** was coexpressed with the mCherry-tagged internal control hMBD1-R22C+R44C with S45 → **1** and the LRS/tRNA$^{Leu}$ amber suppressor pair in NIH/3T3 cells. The selected cells were first imaged in a 10 min interval for a total of 20 min before activation to monitor the non-specific cellular diffusion, then activated with violet light (395/25 nm, LED) for 50 ms, followed by image recording every 2 min for in total 40 min. All the EGFP-tagged hMBD1 S45 → **1** domain variants demonstrated increasing fluorescence accumulation at chromocenter regions marked by SiR-DNA staining, whereas the non-mCpG binding hMBD1-R22C+R44C-S45 → **1** remained dispersed in the nucleus after activation (**Figure 28b** and **Figure 28c**). Strikingly, the photocaged hMBD1-C338A+C341A exhibited the fastest fluorescence saturation at chromocenters upon light activation compared to the wt and the ΔTRD variants (**Figure 28b** and **Figure 28c**), and the same phenomenon was also observed in the fluorophore switch experiments (**Figure S 20** and **Figure S 21**).

To determine the differential binding kinetics between the constructs, half-lives of fluorescence saturation at chromocenters were calculated by the non-linear least square function based on a pseudo-first-order kinetic equation which is discussed in the following section.

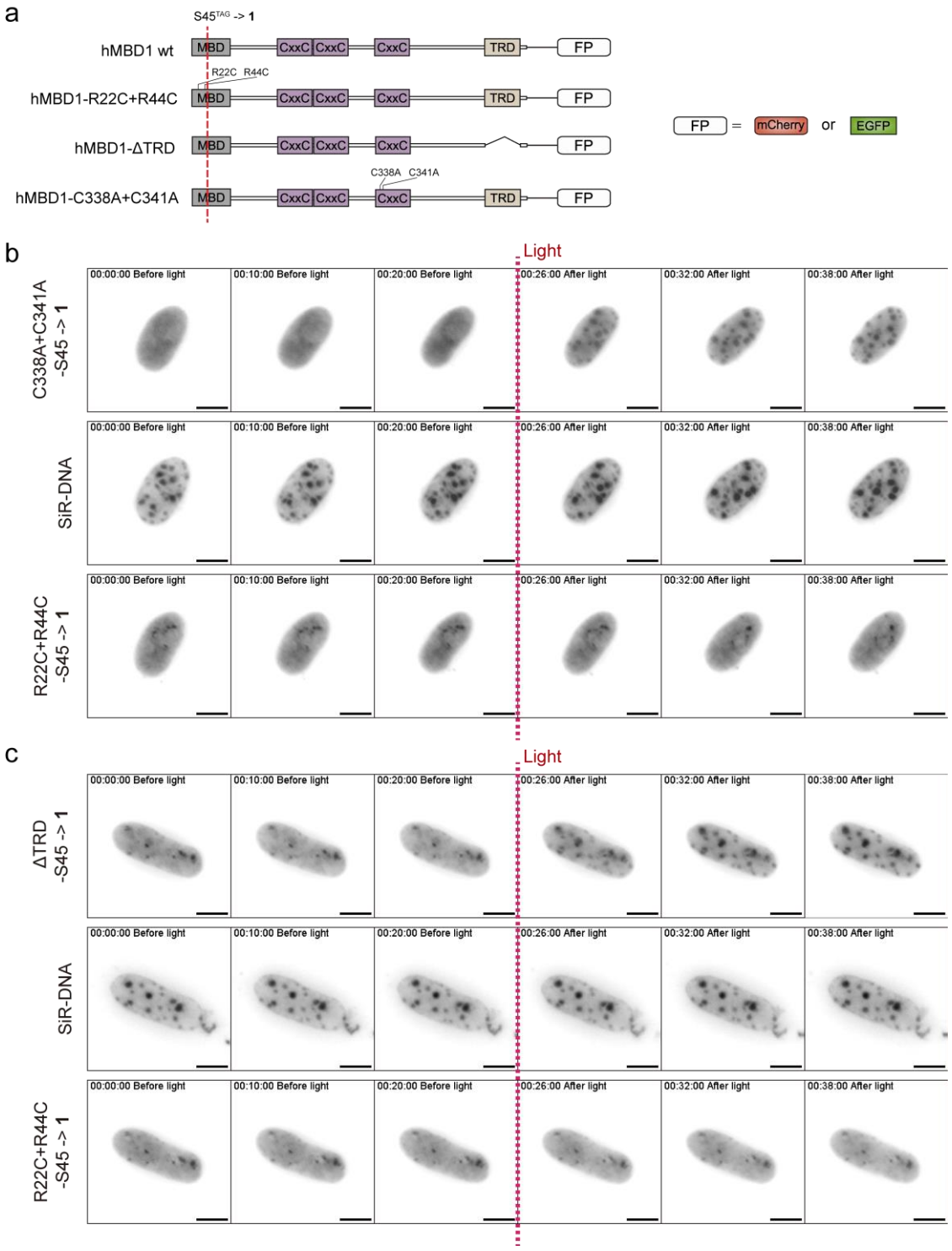


**Figure 28. Study the binding kinetics of hMBD1 CXXC3 and TRD mutant in live NIH/3T3 cells**. **a**) Domain features of vectors encoding S45 → TAG mutated hMBD1 wt, R22C+R44C mutant, C338A+C341A mutant, and ΔTRD variant. All constructs are C-terminally tagged with either mCherry or EGFP. **b**) Imaging the simultaneous light activation of mCherry-tagged hMBD1-C338A+C341A-S45 → **1** and EGFP-tagged hMBD1-R22C+R44C-S45 → **1** in live NIH/3T3 cells as in Figure 27**h**. Scale bar: 5 µm. **c**) Imaging the simultaneous light activation of EGFP-tagged hMBD1-ΔTRD-S45→**1** and mCherry-tagged hMBD1-R22C+R44C-S45 → **1** in live NIH/3T3 cells as in Figure 27**h**. Scale bar: 5 µm.

### 7.2.1. Kinetic analysis and curve fitting method

Since the present study focuses on the mCpG binding kinetics of methylation readers and mCpG sites can be designated by the bright DNA stain (here SiR-DNA) speckles in NIH/3T3 cells, the kinetic analysis measures the increasing fluorescence intensities in the chromocenter regions. In brief, the chromocenter regions were marked by the bright spots in the SiR-DNA staining images, then the mean EGFP and mCherry fluorescence intensities of each chromocenter spot were measured (**Figure S 23**). Additionally, mean fluorescence intensities (MFI) of every chromocenter were normalized to the MFI of the "nucleoplasma" which is the area of the nucleus subtracted by chromocenters, then further normalized to the value at time zero. The resulting normalized MFI values represent the fold increase of mean fluorescence in the chromocenters compared to the initial state. Plotting the normalized chromocenter MFIs yields an exponential curve starting at 20 min after time zero, which is the time when light activation was executed.

Light activation of MBD1 depicts a molecule exchange system similar to the widely used FRAP method in determining cellular dynamic binding interactions. In FRAP experiments, a small region of interest is permanently bleached and the fluorescence recovery rate in this region is measured. FRAP observations are imaging-based complex diffusion-binding models. The fluorescence recovery requires molecules to diffuse into the bleached region and bind to the targets in this "dark" volume, while the diffusion rate further depends on the molecular interactions with other cellular components (B. L. Sprague et al., 2004; B. Sprague & McNally, 2005). In contrast, the light-activated MBD proteins globally displace the endogenous proteins from the mCpG-rich chromocenters, where the fluorescence accumulation at chromocenter regions represents the diffusion and target binding of activated MBD proteins as in FRAP models. Therefore, the simplest diffusion-uncoupled FRAP model of a single binding state was applied to analyze the binding kinetics of light-activated MBD1. This model fits the recovery curve with an inverse exponential decay equation, namely $1 - Ae^{-kt}$ (B. L. Sprague et al., 2004; B. Sprague & McNally, 2005), assuming rapid diffusion and pseudo-first-order (ligand-receptor or substrate-enzyme) association kinetics where the concentration of free molecules does not change noticeably during the time course (Pollard & de La Cruz, 2013; Stroberg & Schnell, 2017).

Consequently, kinetic curves were fitted by the non-linear least square (*nls*) function in R using the following pseudo-first-order association equation:

$$y = y_0 + (y_{max} - y_0)\left(1 - e^{-kt}\right)$$

**Equation 1. Pseudo-first-order association kinetics**. $y =$ Normalized chromocenter MFI at time $t$; $y_0 =$ Normalized chromocenter MFI at the initiation time point (time of activation, $t =$ 20 min); $y_{max} =$ Plateau MFI; $k =$ Rate constant; $t_{1/2} =$ Half-life of saturation $(= (\ln 2)/k)$

The *nls()* function requires estimation of $y_0$, $y_{max}$, and $k$. Since most binding events reached plateau before $t = 60$ min, $y_0$ and $y_{max}$ can be approximated by the MFI at $t = 20$ and 60 min, respectively. Whereas $k$ is estimated by a fixed value $1/20$ (randomly tested value).

### 7.2.2. Comparing domain-dependent binding kinetics

The equation (**Equation 1**) discussed above was used to estimate the binding kinetics of all EGFP-tagged hMBD1-S45 $\rightarrow$ **1** domain variants in the presence of the mCherry-tagged hMBD1-R22C+R44C with S45 $\rightarrow$ **1**. The calculated fluorescence saturation half-lives confirmed that the hMBD1-C338A+C341A mutant has the most rapid mCpG binding rate ($t_{1/2} = 8.45\ min$), while the wt hMBD1 is slower ($t_{1/2} = 10.79\ min$) than the hMBD1-C338A+C341A mutant but faster than the hMBD1-ΔTRD variant ($t_{1/2} = 13.16\ min$) (**Figure 29a**). A same trend was observed in the fluorophore switch control experiments (except for the ΔTRD variant which was not included in the control experiment) with similar half-life values (hMBD1-C338A+C341A: $t_{1/2} = 7.77\ min$, wt hMBD1: $t_{1/2} = 9.13\ min$) (**Figure S 24a**). These data imply that the mCpG binding is stronger in the absence of a functional CXXC3 domain, presumably due to the loss of alternative binding to unmethylated CpGs and the CXXC3-mediated interactions with TET1 as concluded in our published study (Lin et al., 2022) and the previous study (P. Zhang et al., 2017). Conversely, the slow binding kinetics of the hMBD1-ΔTRD variant could be explained by the potentially loose chromatin interactions due to the loss of contacts with other chromatin factors (Hameed et al., 2014; Ng et al., 2000). It is noteworthy that the binding kinetics of EGFP-tagged hMBD1-S45 $\rightarrow$ **1** domain variants alone (in the absence of photocaged hMBD1-R22C+R44C) also demonstrated the same trend and similar half-life values (hMBD1-C338A+C341A: $t_{1/2} = 6.83\ min$, wt hMBD1: $t_{1/2} = 9.79\ min$, hMBD1-ΔTRD: $t_{1/2} = 13.5\ min$) (**Figure 29b**). This result additionally excludes the potential of overexpressed non-mCpG binding MBD proteins forming molecular traps or "sieving" effects which could impede kinetic measurements.

Furthermore, direct binding competition between the hMBD1 wt and the C338A+C341A mutant in cells corroborated their differential binding kinetics. When the photocaged hMBD1 wt and C338A+C341 mutant (orthogonally tagged with EGFP or mCherry) were coexpressed in NIH/3T3 cells and activated simultaneously, the C338A+C341A mutant constantly displayed faster chromocenter binding rate independent of C-terminally tagged fluorophores (**Figure 29c**, **Figure S 22**, and **Figure S 24**). In addition, the CXXC-dependent differential binding kinetics also agrees with a previous FRAP study (P. Zhang et al., 2017), hence substantiating the novel light-activatable methylation reader as a robust alternative approach for studying cellular binding kinetics.

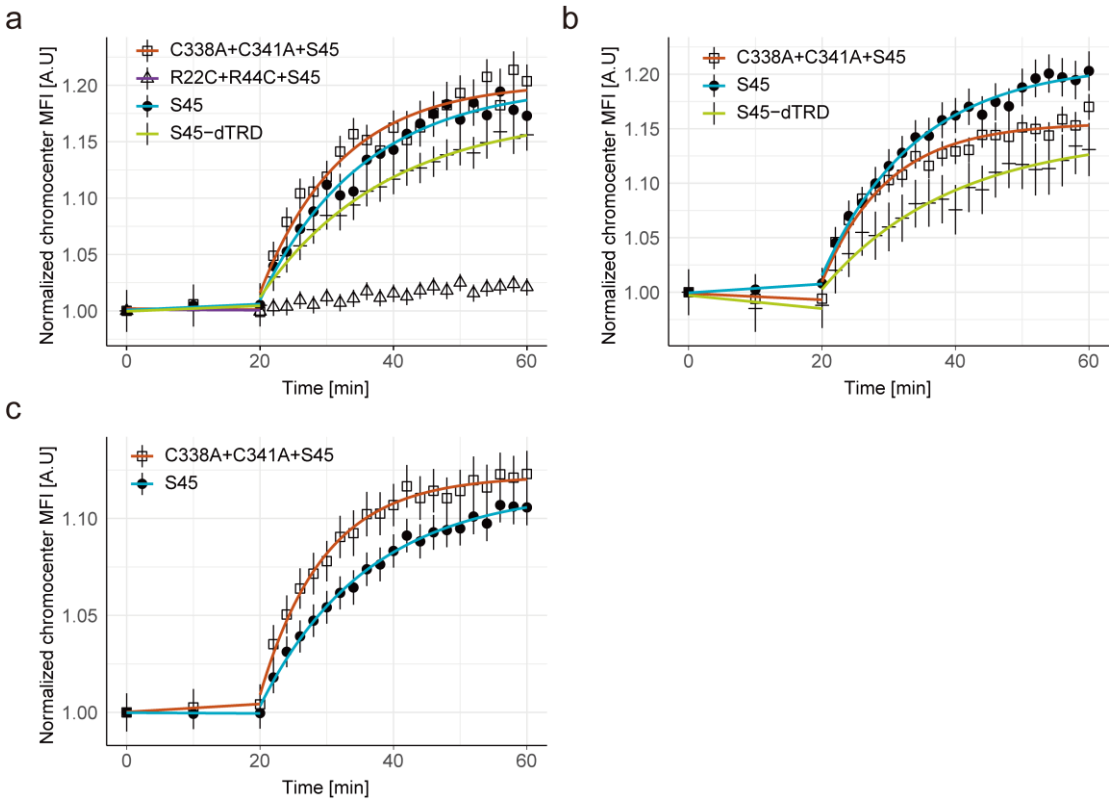


**Figure 29. Cellular mCpG binding kinetics of hMBD1 domain variants**. **a**) Time-resolved measurements of chromocenter fluorescence in NIH/3T3 cells coexpressing EGFP-tagged hMBD1-S45 → **1** (wt, C338A+C341A and ΔTRD variants) and mCherry-tagged hMBD1-R22C+R44C-S45 → **1** upon light activation. The normalized MFI values of recorded chromocenters (N between 100 and 700) from 2 biological independent replicates are plotted by mean and standard error with respect to time. Recorded cell numbers are N = 6, 9, and 12 for wt, C338A+C341A mutant, and ΔTRD variant, respectively. Two additional biological independent fluorophore switch experiments are summarized in Figure S 24**a**. **b**) Measurements of chromocenter fluorescence in cells expressing only EGFP-tagged hMBD1-S45 → **1** (wt, C338A+C341A and ΔTRD variants) upon light activation. The normalized MFI values of recorded chromocenters (N between 100 and 600) are plotted by mean and standard error with respect to time (wt and the C338A+C341A mutant were measured in 2 biological independent replicates, whereas the ΔTRD variant was only measured once). Recorded cell numbers are N = 8, 20, and 4 for wt, C338A+C341A mutant, and ΔTRD variant, respectively. **c**) Measurements of chromocenter fluorescence in cells coexpressing mCherry-tagged wt hMBD1-S45 → **1** and EGFP-tagged hMBD1-C338A+C341A-S45 → **1** upon light activation. The normalized MFI values of recorded chromocenters (N between 800 and 1000) from 2 biological independent replicates are plotted by mean and standard error with respect to time. Recorded cell number is N = 38. Two additional biological independent fluorophore switch experiments are summarized in Figure S 24**b**.

## 7.3. Conclusion

The main goal of the current project is to develop novel light-activatable DNA methylation reader proteins via genetic code expansion to enable kinetic insights into their regulatory events.

Based on the findings of the domain-dependent interplay between hMBD1 and hTET1 described

in Chapter 6, hMBD1 was used as a model for engineering a light-activatable MBD domain. Structural analysis suggested two residues, T27 and S45, where the incorporation of photocaged serine **1** could potentially disrupt the contacts with methylated DNA in the binding pocket, thus temporally blocking the MBD function. Both positions showed faithful incorporation of photocaged serine **1**, although only S45 → **1** demonstrated the expected dispersed localization in the nucleus due to its inability to specifically target mCpG sites. Irradiating hMBD1-S45 → **1** with blue light effectively initiated its translocation to the chromocenter regions like hMBD1 wt. The robust light response was further confirmed by prolonged imaging of its dark state localization and a similarly photocaged hMBD1-R22C+R44C which the localization should not respond to light.

The light-activatable hMBD1 enables uncoupling the upstream processes and the reader ability from the study of domain-dependent mCpG binding kinetics. This new tool revealed the differential binding rates of the wt hMBD1, the CXXC3 mutant (C338A+C341A), and the TRD variant (ΔTRD). The CXXC3 mutant displayed the shortest saturation half-life, potentially due to the removal of alternative binding to unmethylated CpGs and the loss of CXXC3-mediated regulation of chromatin factors such as TET1. The ΔTRD variant exhibited delayed saturation compared to wt, presumably because of the missing interactions with chromatin factors and a loose chromatin binding.

The light activation approach provides a complementary tool to the traditional imaging-based FRAP to study dynamic target binding interactions in living cells. The activation requires relatively mild irradiation that is less likely to destroy fluorophores than FRAP bleaching (50 ms of LED irradiation in the current study, compared to 600 ms of 100% laser irradiation in selected literature) (Rajan et al., 2015), therefore allowing kinetic analysis on the global level (for example, the whole nucleus). Moreover, FRAP is limited to studying proteins localized to visible and clearly-defined cellular features, such as chromocenters, to ensure the recovery observation includes actual target binding events and not purely diffusions. Whereas light activation directly controls the protein function and uncouples the specific binding interactions from diffusions or other interactions in the case of light-activatable MBD1. Most importantly, the light activation approach is not restricted to the demonstrated imaging-based study but allows a wide range of downstream analyses, such as monitoring chromatin regulatory events (via chromatin immunoprecipitation, ChIP), profiling dynamic DNA modification landscape (via MeDIP or beyond), as well as transcriptomics and proteomics studies. For instance, a potential application of the light-activatable MBD1 is studying the kinetics of mCpG competition in the domain-dependent modulation of TET1 activity and downstream regulatory effects.

# 8. Summary and Outlook

The present work established methods to employ direct light control of protein functions in studying complex chromatin regulation in mammalian cells.

The first part of the study demonstrated a novel application of light-activatable TET dioxygenase in unveiling the intracellular crosstalk between chromatin regulatory elements. Genetic code expansion aids the light activation of TET1 in intracellular context by blockage of cofactor binding with a photocaged serine residue, enabling observation at the natural chromatin regulatory network but uncoupled from the upstream TET life cycle. Monitoring the kinetics of TET1-mediated 5hmC formation, the 5mC oxidation product, upon TET1 activation revealed a domain-dependent interplay between MBD1 and TET1. An overall picture of the interplay concludes a dominant 5mC competition from the MBD domain that downregulates TET1-mediated 5mC oxidation. While the CXXC3 domain contributes to the second layer of modulation by promoting TET1 activity via its affinity to non-methylated CpGs, the TRD domain-mediated chromatin interactions do not play roles in the interplay within the monitored 5mC oxidation time window. Notably, the finding on the CXXC3-mediated upregulation of TET1 activity agrees with a previous report on CXX3-dependent recruitment of TET1 in the murine model. Altogether, the established approach in the present study enables unprecedented kinetic insights into the interplay between epigenetic DNA methylation readers and erasers. Additionally, coupling with other analytical approaches such as hMeDIP or RNA-sequencing techniques would allow time-resolved insights into the genomic or transcriptomic regulation by the reader-eraser crosstalk. This approach is expected to provide broad applications in unraveling the regulation network of TET activity, for example, by other chromatin regulators, as well as establishing disease-linked causalities of TET regulation.

The second part of the study highlighted the first light-activatable DNA methylation reader, MBD1, and an example application in monitoring domain-dependent cellular binding kinetics. A genetically encoded photocaged serine at S45 position disrupts the molecular interactions between the S45 residue and the 5mC-containing DNA, thereby temporally blocking 5mC recognition to uncouple the prior binding events of MBD1. Imaging the photocaged MBD1 in live NIH/3T3 cells verified a prompt recovery of localization at mCpG-rich chromocenters upon light irradiation. As a proof of concept, time-resolved imaging of light-activatable MBD1 bearing different domain mutations unveiled differential target binding rates of domain variants in living cells. The photocaged MBD1 bearing a non-functional CXXC3 domain (C338A+C341A) most rapidly accumulated at the chromocenters upon light activation, while the TRD truncation variant (ΔTRD) exhibited the slowest binding rate among all. These observations support the previously proposed model involving the CXXC3 domain as the secondary modulation factor that provides

binding to non-methylated CpGs and potential interactions with TET1. Utilizing the light-activatable MBD1 would enable follow-up insights into the rate and order of the MBD1 domain-dependent modulation of TET1 activity. The same approach can further apply to studies of chromatin regulatory events associated with MBD proteins. On the other hand, kinetic measurements on cellular target binding employing light-activatable MBD1 contribute to a complementary tool of the widely used FRAP method by directly controlling the targeted protein function; nonetheless, the light-activatable methylation reader can offer vast downstream information when coupled with different analytical techniques, for instance, profiling the dynamic regulation of methylome and transcriptome (via MeDIP, RNA-sequencing, etc.).

To summarize, this work demonstrated examples employing direct light control of target protein functions to dissect complex chromatin regulatory events. These approaches are envisioned to provide unparalleled insights into the kinetic regulation of hardly druggable chromatin factors and further connect to the onset of diseases.

# 9. Materials and Methods

Part of the materials and methods described here have been published in (Lin et al., 2022)(A.4).

## 9.1. General information

***Synthesis of oligonucleotides***

The oligonucleotides listed in Table S 1, Table S 2 (o2067), and Table S 3 (o1516 and o1601) were synthesized by Merck KGaA (Darmstadt, Germany). The desalted oligonucleotides were stored as 100 µM stock in TE buffer. The oligonucleotides listed in Table S 2 were provided by Microsynth Seqlab GmbH (Göttingen, Germany).

***Sanger sequencing***

The sequence of constructed plasmids was routinely checked by Sanger sequencing by Microsynth Seqlab GmbH (Göttingen, Germany) or Eurofins Genomics Germany GmbH (Munich, Germany).

***Purification of plasmid DNA and double stranded oligonucleotides***

Plasmid DNAs were isolated from bacteria strains via silica column purification using NucleoSpin® Plasmid EasyPure kit (MachereyNagel, Düren, Germany) following the manufacturer's instruction. Large scale purification of endotoxin-free plasmids for mammalian cell transfection was performed with NucleoBond® Xtra Maxi kit (MachereyNagel).

Double stranded oligonucleotides amplified by PCR or digested from plasmids were identified or separated by DNA electrophoresis (discussed below), then the crude reaction mixture or agarose gels containing desired products were purified using NucleoSpin® Gel and PCR Clean-Up kit (MacheryNagel).

For DNA electrophoresis, 1% (w/v) of agarose gel in 0.5× TBE buffer was prepared. The DNA samples were resolved using 8 – 12 V/cm. The agarose gels were stained with 0.5 µg/mL ethidium bromide, destained with water, and visualized with UV fluorescence.

***Enzymes***

The enzymes used in in this study that are not indicated in the methods are listed below. Corresponding experiments were conducted following the supplier's instructions.

**Table 1. List of enzymes.**

| Enzyme | Company |
|---|---|
| AscI | NEB (New England Biolabs) |
| KpnI | NEB (New England Biolabs) |
| DpnI | NEB (New England Biolabs) |
| PacI | NEB (New England Biolabs) |
| XhoI | NEB (New England Biolabs) |
| XbaI | NEB (New England Biolabs) |
| Q5 HiFi DNA polymerase | NEB (New England Biolabs) |
| Phusion DNA polymerase | NEB (New England Biolabs) |
| KOD hot start DNA polymerase | Merck Millipore |
| T5-Exonuclease | NEB (New England Biolabs) |
| Taq-DNA ligase | NEB (New England Biolabs) |
| T4-DNA ligase | NEB (New England Biolabs) |
| NEBuilder® HiFi DNA assembly cloning kit | NEB (New England Biolabs) |

***Bacteria strains***

**Table 2. *E. coli* strains used in this study.**

| Strain | Genotype | Supplier |
|---|---|---|
| DH5α | F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ– | Invitrogen™ (Thermo Fisher Scientific) |
| DH10B (Top10™) | F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ- | Invitrogen™ (Thermo Fisher Scientific) |

***Chemicals***

The chemicals used in in this study that are not indicated in the methods are listed below.

**Table 3. List of chemicals.**

| Name | CAS No. | Supplier |
|---|---|---|
| 2-Amino-2-(hydroxymethyl)propane-1,3-diol (Tris), buffer grade | **77-86-1** | **Carl Roth** |
| 2-Log DNA ladder | | NEB |

| | | |
|---|---|---|
| Acetic acid | 64-19-7 | Carl Roth |
| Agarose LE, molecular biology grade | 9012-36-6 | Biozym Scientific |
| ammonium acetate | 631-61-8 | Carl Roth |
| Bovine serum albumine (BSA) | 9048-46-8 | Cell Signaling Technology |
| Carbenicillin, disodium salt | 4800-94-6 | Carl Roth |
| Dimethyl sulfoxide (>= 99.7%) | 67-68-5 | Merck |
| Ethanol, absolute | 64-17-5 | Merck |
| Ethidium bromide | 1239-45-8 | Carl Roth |
| Ethylenediaminetetraacetate (EDTA) | 6381-92-6 | Alfa Aeser |
| Formaldehyde, 37 wt. % in H2O, contains 10-15% Methanol as stabilizer | 50-00-0 | Merck |
| Hydrochloric acid, 37% | 7647-01-1 | Merck |
| Isopropanol | 67-63-0 | Fisher Scientific |
| Sodium chloride | 7647-14-5 | Merck |
| Sodium hydroxide | 1310-73-2 | Merck |
| Trisodium citrate, dihydrate | 6132-04-3 | Carl Roth |
| Triton® X-100 | 9002-93-1 | Fluka Chemika |
| Tween® 20 | 9005-64-5 | Fisher Bioreagents |

***Software***

The software used in in this study that are not indicated in the methods are listed below.

**Table 4. List of software.**

| **Name** | **Company** |
|---|---|
| SnapGene v4.3 | Dotmatics |
| NanoDrop 2000 v1.6 | Thermo Fisher Scientific |
| BioDoc Analyze v2.1 | Biometra |
| Adobe Illustrator 2022 v26.4 | Adobe |
| PyMol v2.5.2 | Schrödinger LLC |
| ChemDraw Professional v21.0.28 | PerkinElmer |
| Office 365 | Microsoft |

## 9.2. Construction of plasmids for MBD and TET protein expression

All vectors were derived from pShP2384 which is based on pcDNA3.1-GoldenGate-VP64 (Addgene 47389) with removed VP64 and lacZα gene as described previously (Palei et al., 2020). The mCherry transfection control on pShP2384 was deleted using whole plasmid PCR and re-ligation with primers o3246/o3247 resulting in plasmid pTzL1744. Then, the mCherry sequence amplified with primers o3254/o3255 was inserted into pTzL1744 (amplified with primers o3256/o3257) via Gibson assembly, followed by quick change site-directed mutagenesis (SDM) to correct a frameshift using primers o3284/o3285 (yielding plasmid pTzL1745).

To construct plasmids encoding EGFP-tagged hMBD1, a Myc tag was first introduced into pTzL1745 by quick change SDM using primers o3167/o3257, resulting in pTzL1746. The human full length MBD1 coding sequence was amplified from a human prostate cDNA library (BiocCt 10108-A-GVO-EB) using primers o3292/o3293, then MBD1 and EGFP (amplified with primers o3294/o3295) were assembled with pTzL1746 (amplified by primers o3290/o3291) via Gibson assembly, yielding pTzL1747. Finally, remaining unwanted sequences were removed by quick change using primers o3642/o3643, yielding pTzL1836.

The EGFP-tagged hMBD1 mutants were cloned as follows. The R22C mutation was introduced into pTzL1836 by quick change SDM using primers o3730/o3731 to yield pTzL1947. The R22C+R44C mutant was derived from pTzL1947 by introducing an R44C mutation with primers o3732/o3733 via quick change SDM to yield pTzL1964. The C338A+C341A mutations were introduced into pTzL1836 (hMBD1) and pTzL1964 (hMBD1-R22C+R44C) using primers o4479/o4480, resulting in pTzL2645 and pTzL2646, respectively. The hMBD1-dMBD (aa 1-69 deleted) variant was cloned by Gibson assembly of a truncated hMBD1 sequence (amplified with primers o3293/o4300) and the pTzL1836 backbone (amplified with primers o3386/o3291), yielding pBiR2585. The hMBD1-ΔTRD (aa 529-592 deleted) variant was cloned by Gibson assembly of 2 fragments amplified from pTzL1836 using primers o4302/o3291 and o3292/o4305, yielding pBiR2586. The hMBD1 isoform 7 (hMBD1v7, aa 327-382 deleted from isoform 1 sequence) variant was cloned by Gibson assembly of 3 fragments amplified from pTzL1836 using primers o3292/o4189, o4298/o3293, and o3386/o3291, yielding pBiR2593. The hMBD1v7-R22C+R44C variant was cloned by Gibson assembly of 3 fragments amplified from pTzL1964 using primers o3292/o4189, o4298/o3293, and o3386/o3291, yielding pBiR2628. For EGFP-tagged amber mutant hMBD1-S45$^{TAG}$, the amber mutation was introduced into pTzL1836 by quick change SDM using primers o3756/o3757 to yield pTzL1967. The EGFP-tagged hMBD1-C338A+C341A-S45$^{TAG}$ and hMBD1-ΔTRD-S45$^{TAG}$ were cloned similarly by introducing amber mutation into pTzL2645 and pBiR2586 via quick change SDM using primers o3756/o3757 to yield pNaU2737 and pNaU2738, respectively. Another hMBD1-S45$^{TAG}$ plasmid with additional N-terminal Flag tags and GGGGS linker was cloned by

restriction ligation of hMBD1-S45$^{TAG}$ sequence (from pTzL1967) with vector backbone of pTzL1833 using AscI/KpnI, giving plasmid pTzL2511. The EGFP-tagged hMBD1-R22C+R44C-S45$^{TAG}$ was cloned by introducing amber mutation into pTzL1964 via quick change SDM using primers o4128/o4129 giving pTzL2892. For the mCherry-tagged hMBD1-S45$^{TAG}$, hMBD1-C338A+C341A-S45$^{TAG}$, hMBD1-ΔTRD-S45$^{TAG}$, and hMBD1-R22C+R44C-S45$^{TAG}$, the hMBD1-S45$^{TAG}$ coding sequences (from pTzL1967, pNaU2737, pNaU2738, and pTzL2892, respectively) were ligated with the vector backbone of pTzL1960 using AscI/KpnI, giving pTzL2512, pTzL2889, pTzL2890, and pTzL2902, respectively. For EGFP-tagged amber mutant hMBD1-T27$^{TAG}$, the amber mutation was introduced into pTzL1833 by quick change SDM using primers o4592/o4558 to give pTzL2681.

For EGFP-tagged hMBD3, the human MBD3 isoform 2 (MBD3v2) coding sequence was first amplified from human prostate cDNA using primers o3380/o3381 and inserted in vector backbone of pTzL1747 (amplified by primers o3386/o3291) via Gibson assembly, resulting in pTzL1774. Unwanted sequences were subsequently removed by quick change SDM using primers o3642/o3643 giving pTzL1835. Finally, the canonical human MBD3 sequence (isoform 1) was cloned by inserting the coding sequence of MBD3 aa 5-36 into pTzL1835 via quick change SDM using primers o3810/o3811. For EGFP-tagged hMBD2a, the coding sequence for human MBD2a was amplified from a plasmid encoding human full length MBD2a (Addgene 78141) using primers o3510/o3511, then inserted into the backbone of pTzL1835 (amplified by primers o3386/o3291) via Gibson assembly, resulting in pTzL1889. For EGFP-tagged hMBD4, the coding sequence for human MBD4 was amplified from human prostate cDNA using primers o3382/o3383, then inserted into the backbone of pTzL1835 (amplified by primers o3386/o3291) via Gibson assembly, followed by frameshift correction with primers o3758/o3759 to yield pTzL1948. For EGFP-tagged hMeCP2, the coding sequence for human MeCP2 was amplified from human prostate cDNA using primers o3384/o3385 and inserted into the backbone of pTzL1747 via restriction ligation using AscI/KpnI, resulting in pTzL1773. Unwanted sequences were subsequently removed by quick change SDM using primers o3642/o3643 to afford pTzL1834. For the expression vector encoding EGFP only, the hMBD1 sequence in pTzL1836 was replaced with a $(GGGGS)_3$ linker by restriction/ligation of annealed oligos o3825/o3826 and pTzL1836 backbone using AscI/KpnI, resulting in pTzL1990.

For mCherry-tagged hTET3CD (aa 689–1660), the coding sequence of human TET3 catalytic domain was amplified from plasmid pTzL1837 encoding hTET3CD (cloned by assembling full length human TET3 isoform 1 sequence, which is amplified from Addgene 49446 using primers o3374/o3410, with the AscI/KpnI digested vector backbone of pTzL1745, and subsequently removed the unwanted sequences with primers o3642/o3643) using primers o3752/o3377, and assembled with 2 vector backbone fragments of pTzL1837 amplified with primers o2261/o3596

and o3288/o2260, resulting in plasmid pTzL2050. Mutations that remove catalytic activity (H942Y, D944A) were first introduced into pTzL1837 (full length hTET3 isoform 1) using primers o3754/o3755, then the inactive hTET3CD coding sequence was amplified using primers o3752/o3377 and assembled with 2 vector backbone fragments of pTzL1837 amplified with primers o2261/o3596 and o3288/o2260, resulting in plasmid pTzL2079.

For mCherry-tagged hTET1CD (aa 1418–2136), the coding sequence for the human TET1 catalytic domain was amplified from a plasmid encoding human full length TET1 (Addgene 49792) using primers o3751/o3473, then assembled with 2 vector backbone fragments of pTzL1837 amplified with primers o2261/o3596 and o3288/o2260, resulting in plasmid pTzL1960. The mutations (H1672Y, D1674A) that remove catalytic activity were introduced into hTET1CD plasmid (pTzL1960) using o3762/o3763, resulting in pTzL1970. The plasmid encoding amber mutant hTET1CD-S2045$^{TAG}$ was cloned by restriction ligation of hTET1CD-S2045$^{TAG}$ sequence (digested from pShP2444) and vector backbone of pTzL1960 using AscI/KpnI, yielding pTzL2504. Another hTET1CD-S2045$^{TAG}$ plasmid bearing additional N-terminal Flag tags and GGGGS linker was cloned by restriction ligation of hTET1CD-S2045TAG sequence with vector backbone of pTzL1833 using AscI/XbaI, giving plasmid pTzL2513.

The mCherry-tagged hTET2CD (aa 1129–2002) was cloned by restriction ligation of hTET2CD coding sequence (digested from plasmid pShP2413) and vector backbone of pTzL1960 by AscI/KpnI, yielding pTzL2005. The mCherry-tagged hTET2CD inactive mutant (with H1672Y and D1674A mutations) was cloned similarly by ligating the AscI/KpnI digested hTET2CD coding sequence (from plasmid pShP2416) with vector backbone of pTzL1960 to give pTzL2006.

The orthogonal E. coli leucyl synthetase (ecLRS-BH5) bearing the five previously reported mutations M40G, L41Q, Y499L, Y527G, H537F2 and the suppressor tRNACUA were encoded on the previously reported plasmid pStH1147 (Palei et al., 2020).

## 9.3. Cell culture

HEK293T cells were cultivated in DMEM (Dulbecco's Modified Eagle Medium, w/ 4.5 g/L Glucose, w/o: L-Glutamine, w: Sodium pyruvate, w: 3.7 g/L $NaHCO_3$, PAN Biotech, P04-03600) supplemented with 10% FBS (South America origin, premium grade, PAN Biotech, P30-3306), 2 mM L-glutamine (PAN Biotech, P04-80100), 100 U/mL of penicillin and 0.1 mg/mL of streptomycin (PAN Biotech, P06-07100) in a sterile humidified incubator (≥ 95%) at 37°C and a $CO_2$ level of 5%. For transfection, cells were seeded a day before to reach 70-80% confluency at the time of transfection. Transient plasmid transfection was carried out by the use of poly-ethyleneimine (PEI, 1mg/mL in dd H2O, pH 7) (linear MW 25.000, CAS 9002-98-6, Alfa Aesar).

Mouse embryonic fibroblast NIH/3T3 cells (ATCC, CRL-1658) were maintained in the same conditions described above. The plasmid transfection of NIH/3T3 was done either by PEI as described above, or by electroporation using the 10 µL Neon Transfection System (Invitrogen, Thermo Fisher Scientific Inc.). Briefly, 50,000 cells were resuspended in 10 µL resuspension buffer R (Neon-transfection 10 µL kit, Invitrogen, MPK1096) with 0.25 µg of plasmid and electroporated at a pulse voltage of 1400 volts, pulse width of 20 ms, and pulse number of 2. The cells were subsequently seeded in a 96-well plate containing growth media (DMEM supplemented with 10% FBS and 2 mM L-glutamine as described above, w/o penicillin and streptomycin), then left to adhere in a humidified 37°C incubator with 5% $CO_2$.

## 9.4. Light-activation of TET1

HEK293T cells grown in 6-well cell culture plate (Sarstedt) were transfected with plasmids encoding TET1CD-S2045$^{TAG}$, LeuRS/tRNA$^{Leu}$, and desired MBD proteins. At 3 h post-transfection, growth media was exchanged with media supplemented with 0.05 mM **1** (TOCRIS, 780009-55-4) and allowed expression for 24 h. For light treatment, growth media containing **1** was replaced by warm DPBS (Dulbecco's phosphate-buffered saline, Mg/Ca free, PAN Biotech, P04-361000) and subsequently placed on a 365 nm UV-transilluminator (Witeg DH.WUV00010, 6x 15 W) for 3 min. Immediately after irradiation, DPBS was replaced by pre-warmed growth media (without **1**), and cells were maintained in a humidified 37 °C incubator with 5 % $CO_2$ until harvesting.

## 9.5. Fluorescence microscopy and image analysis

NIH/3T3 cells transfected by PEI or electroporation were grown in black 96-well plate with flat polymer coverslip bottom (ibidi, 89626). After the protein of interest was stably expressed (16-24 h), cells were fixed with 4% formaldehyde for 10–15 min at RT followed by three DPBS washes. Fixed cells were subjected to permeabilization using 0.25% Triton® X-100 for 15-20 min at RT. After three DPBS rinses, nuclei were stained using 1:50 VECTASHIELD® mounting medium with DAPI (Vector Laboratories, H-1200-10) in DPBS for 5 min in the dark and maintained in the same solution during imaging. Experiments were performed using an Olympus IX81 microscope equipped with a Hamamatsu model C10600-10B-H camera. Samples were illuminated with a Lumencor SPECTRA X light engine® NIR version which operates 6 independent LED light sources coupled with bandpass filters (V: 395/25 nm; B: 440/20 nm; C: 470/24; GY: 550/15 nm; R: 640/30 nm; TN: 730/40 nm). Images were acquired using either a 60x or a 100x oil immersion objective and z-stack images (0.5 µm/step) for EGFP (excitation with C, emission filter 554/23 nm), mCherry (excitation with GY, emission filter 635/18), DAPI (excitation with V, emission filter 474/27 nm), and SiR-DNA (excitation with R, emission filter nm). The intensity and subcellular

localization of foci was analyzed from Z-projections of image stacks by maximal intensity (1344 x 1024 pixels, 32 bits) using Fiji distributed ImageJ (1.53q).

## 9.6. Immunostaining and flow cytometry

Cells were harvested at desired time point after transfection or light-activation (described above) from culture dishes by incubating with trypsin (Trypsin 0.05 %/EDTA 0.02 % in PBS, w/o: Ca and Mg, w: Phenol red, PAN Biotech, P10-0231SP) for 3–5 min at 37°C. Following cell detachment, trypsinization was quenched by adding growth media. The resulting cell suspensions were placed in 5 mL round-bottom polystyrene tubes (Falcon, 352058) and washed once with DPBS. After collection by centrifugation, cells were fixed with medium A (Fix & Perm cell permeabilization kit, Thermo Scientific, GAS004) for 15 min at RT and subsequently washed with wash buffer (DPBS with 5% FBS). Then the fixed cells were permeabilized with medium B (Fix & Perm kit) for 20 min. In control experiments with or without RNase A treatment, an additional incubation step with RNase A (10 µg/mL in DPBS, Qiagen, 19101) at 37°C for 30 min was added after permeabilization. Thereafter, cells were resuspended in 2 N HCl and incubated 30 min at RT to denature chromosomal DNA, immediately followed by dilution to a final concentration of 0.4 N HCl with DPBS. The HCl solution was removed by centrifugation, and the cell pellet was washed with wash buffer. Before immunostaining, cells were resuspended in blocking buffer (DPBS with 1% BSA and 0.05% Tween20) and incubated for 1 h at RT or overnight at 4°C with gentle shaking. To detect genomic 5hmC, a rabbit anti-5hmC (Active Motif, 39769) primary antibody and Alexa Fluor 405-conjugated goat anti-rabbit (Invitrogen, A-31556) secondary antibody was used. Cells were incubated with anti-5hmC antibody (1:1000) and 1% BSA in 1x intracellular staining buffer (SONY, 2705010) for 1 h followed by three washing steps with PBST buffer (DPBS with 0.05% Tween20). Then, cells were incubated with AF405-conjugated secondary antibody (1:1000) and 1% BSA in intracellular staining buffer for 1 h. After three washing steps with PBST buffer, cells were resuspended in DSPBS and subjected to a cell-strainer (Falcon, 352235) for FCM measurement. FCM measurements were performed with a Sony Cell Sorter model LE-SH800SFP using 405, 488 and 561 nm lasers coupled with 450/50 (FL1), 525/50 (FL2) and 600/60 (FL3) nm filters to detect AF405, EGFP, and mCherry, respectively. FCM results were exported as flow cytometry standard files (FCS 3.0 or 3.1) by the cell sorter software (v. 2.1.3 or v. 2.1.5, Sony Biotechnology) and analyzed using R as described below.

## 9.7. Fluorescence-activated cell sorting and DNA dot blot

The cells were trypsinized from plates at desired time points after transfection or light-activation and pelleted by centrifugation (described above), followed by resuspension in ice-cold DPBS

containing 1% BSA and subjected to a cell strainer before sorting (Falcon, 352235). Cell sorting was performed with a Sony Cell Sorter model LE-SH800SFP using 488 and 561 nm lasers coupled with 525/50 and 617/30 nm filters to detect EGFP and mCherry, respectively. Fluorescence intensity thresholds for desired sorting population were determined using cells similarly transfected but grown in the absence of **1** as negative control. Cells were kept at 4°C in the sample loading chamber before they were entering the flow system (sample pressure 6, flow rate 37 µL/min). The positive cell population selected according to the defined gate was sorted into 2 mL tubes containing 1 mL DPBS supplemented with 5% FBS at 4°C. The collected cells were pelleted and stored at -20°C. Genomic DNA from sorted cells was extracted using the QIAmp DNA purification kit (Qiagen, 51304). For the positive control in the 5hmC dot blot, a BRCA1 promoter sequence (421 bp) was amplified from HeLa cell cDNA using primers o1516/o1601 and a 5hmdCTP-containing dNTP mix (Zymo, D1040).

Genomic DNA was denatured in 0.4 M NaOH and 10 mM EDTA at 99°C for 10 min, then neutralized with an equal volume of ice-cold 2 M ammonium acetate solution (pH 7.0). Denatured DNA samples were diluted to designated concentrations with TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) and spotted on a Zeta-Probe membrane (high-strength nylon membrane positively charged with quaternary amine groups, Bio-Rad, 1620153) by gravity filtration in an assembled Bio-Dot® apparatus (Bio-Rad, 1706545). The wells were subsequently rinsed with 0.4 M NaOH by applying vacuum until the wells were dry, then the blotted membrane was taken out from the apparatus and washed with 2x SSC buffer (0.3 M NaCl, 0.03 M trisodium citrate). The membrane was either air-dried or vacuum baked at 80°C for 30 min. For immunofluorescence detection of genomic 5hmC, a rabbit anti-5hmC (Active Motif, 39769) primary antibody and an Alexa Fluor 750-conjugated goat anti-rabbit (Invitrogen, A-21039) secondary antibody were used. The blotted membrane was blocked with 5% non-fat milk in TBS buffer (20 mM Tris, 0.5 M NaCl) for 2 h at RT, followed by incubation with anti-5hmC antibody (1: 10,000 in TBS buffer with 0.3% Tween 20 and 5% non-fat milk) overnight at 4°C and anti-rabbit AF750 antibody (1: 5000 in TBS buffer with 0.3% Tween 20 and 5% non-fat milk) for 2 h at RT. Immunofluorescence of 5hmC was visualized with Odyssey® DLx imaging system (LI-COR). To control equal spotting of DNA samples on the membrane, the same blot was stained with 0.02% methylene blue (CAS 122965-43-9, Alfa Aesar, A18174.14) in 0.3 M sodium acetate (pH 5.2).

## 9.8. Fluorescence-activated cell Sorting and light-activation of MBD1

NIH/3T3 cells grown in 10 cm cell culture plate (Sarstedt) were transfected with plasmids encoding indicated MBD1-S45$^{TAG}$ constructs (wt, R22C+R44C, C338A+C341A, or ΔTRD) and

LeuRS/tRNA$^{Leu}$. At 3 h post-transfection, growth media was exchanged with media supplemented with 0.05 mM **1** and allowed expression for 24 h. Then cells were trypsinized, washed once with DPBS, pelleted by centrifugation, resuspended in 500 µL warm DPBS containing 1% BSA and subjected to cell strainer before sorting. Cell sorting was performed with Sony Cell Sorter model LE-SH800SFP using 488 and 561 nm laser coupled with 525/50 and 617/30 nm filter to detect EGFP and mCherry, respectively. Fluorescence intensity thresholds of desired sorting population were determined by cells similarly transfected but grown in the absence of **1** as negative control. Cells were kept at 37°C at sample loading chamber before subjected to flow system (sample pressure 4, flow rate 21 µL/min). Desired cell population was sorted at RT into a 15 mL tube containing 6 mL warm (or RT) imaging media (DMEM containing 4.5 g/L glucose, stable glutamine, sodium pyruvate, 0.5g/L NaHCO3, and 25 mM HEPES from PAN Biotech, P04-01163, supplemented with 100 U/mL of Penicillin, 0.1 mg/mL of Streptomycin, and 10% FBS). The cell collection was pelleted and resuspended in warm conditioned media, then seeded around 500 - 2,000 cells per well into the black 96-well plate (ibidi, 89626) pre-treated with 0.01% poly-L-lysine (CAS 25988-63-0, Sigma-Aldrich, P1274). Finally, cells were kept in a humidified 37°C incubator with 5% $CO_2$ for 4-5 hours to allow adherence.

Live cell imaging was performed after cells have adhered and recovered from cell sorting. Cells were incubated with 1 µM SiR-DNA (SiR-DNA kit from Spirochrome AG, SC007) and 5 µM Verapamil (supplemented by the SiR-DNA kit) in growth media for 1 h in the incubator (37°C, 5% $CO_2$) for DNA staining before experiment, then maintained in imaging media of same SiR-DNA and Varapamil concentrations at 37°C during live cell imaging. Same microscope module and settings were used as described above.

Light activation experiments of hMBD1-S45→**1** constructs were performed with an user-defined automatic acquisition sequence in the cellSens imaging software (cellsens Dimension Version 3, Olympus). The automatic sequence started with image acquisition at 0, 10, and 20 min to control the MBD1 localization before light. Right after image acquisition at 20 min, light activation was performed by taking a DAPI snapshot (395/25 nm violet light illumination, 50 ms exposure with 100% intensity), then cell images were acquired every 2 min for a total of 40 min.

## 9.9. FCM Data Analysis by R

Flow cytometry standard files (FCS 3.0 or 3.1) were processed with R 4.1.2 in Rstudio (2022.02.3+492) using following Bioconductor packages: flowCore (2.0.0) (Meur et al., 2007), flowClust (3.26.0) (K. Lo et al., 2008, 2009), flowDensity (1.22.0) (Malek et al., 2015), flowStats (4.0.0) (Hahne F et al., 2022), and ggcyto (1.16.0) (Phu et al., 2018). Fluorescence intensity data extracted from populations of interest were then analyzed using Tidyverse packages (1.3.1). In

brief, cell populations were identified firstly from multivariate t mixture models, then singlet events were selected by a robust linear model with rlm. Populations showing positive or negative fluorescence signals were further separated by applying thresholds identified from respective negative controls (the upper boundary including 99.9% population in respective channels accordingly to the density distribution). The gated positive population of individual sample was further grouped by their MBD(EGFP) and TET(mCherry) intensities, and the median AF405 intensity (5hmC) of each group was normalized to that of the gated negative population from the same sample (**Figure S 1**). Statistics were performed using ggpubr (0.4.0) package, the implemented significance tests were indicated in the corresponding figure captions.

## 9.10. Image quantification and kinetic curve fitting

The microscopy images were prepared, and the intensity and subcellular localization of foci were analyzed from Z-projections of image stacks by maximal intensity (1344 x 1024 pixels, 32 bits) using the Fiji distribution of ImageJ (1.53q) (Schindelin et al., 2012).

The workflow illustrated in **Figure S 23** was used to analyze the kinetic changes of fluorescence in chromocenters upon light activation. The nuclear boundary was indicated by images acquired for the overexpressed nuclear target proteins in green (EGFP) or red (mCherry) channels, while the chromocenters were indicated by images acquired for SiR-DNA staining in Cy5 channel. To define nuclear boundary, images were processed sequentially with background subtraction (convoluted background subtraction from BioVoxxel toolbox (Brocher Jan, 2022)), contrast enhance, smoothing (Gaussian blur), thresholding, then nuclear boundaries were selected from the resulting binary images. The chromocenters were defined with the same process but only foci within the pre-defined nuclear boundary will be selected. An additional nucleoplasma selection was created by the XOR (exclusive or) selection of the nucleus boundary and chromocenters. Finally, mean fluorescence intensity (MFI) measurements were implemented for all selected foci in green (EGFP) and red(mCherry) channels.

The kinetic analysis of obtained MFIs was performed with R 4.1.2 in Rstudio (2022.02.3+492) using Tidyverse packages (1.3.1). The chromocenter MFIs were first normalized to the nucleus MFI respective to the individual nucleus (Norm.mean), then the normalized MFI values of each time point in the time series are further normalized to the average of normalized MFI value of time = 0, giving normalized chromocenter MFIs (Norm.chc). The resulting values were fitted using Equation 1 with the non-linear least square function, *nls*(), and $t_{1/2}$ values were calculated correspondingly.

## 9.11. Molecular simulation and energy minimization of hMBD1 S45 and T27→1

Models of hMBD1 mutated at positions T27 and S45 each with the ncAA **1** (4,5-dimethoxy-2-nitrobenzyl-L-serine) were energy minimized with Gromacs version 2020.1 (Lindahl et al., 2020). Model 1 of the solution NMR structure of PDB 1IG4 (Ohki et al., 2001) was used as template. For generating the mutation at position T27, threonine was at first replaced by serine using the rotamers tool with the Dunbrack 2010 backbone-dependent rotamer library (Shapovalov & Dunbrack, 2011) implemented in UCSF Chimera (Pettersen et al., 2004). The rotamer with the highest predicted probability was chosen. In the second step, for both cases the mutation was introduced using the "Build Structure" tool implemented in UCSF Chimera. For reflecting the conformational flexibility of the mutated residues, for both mutated models, two additional, alternative starting structures for the energy minimization were generated each by varying the chi1 sidechain dihedral angle about +/-120 degree.

The six starting structures were energy minimized with Gromacs version 2020.1 (Lindahl et al., 2020) using the latest Charmm36 (Croitoru et al., 2021; Vanommeslaeghe et al., 2010) all-atom force field from February 2021. The parameters needed to implement the ncAA **1** into the force field were derived from CGenFF (Vanommeslaeghe et al., 2010, 2012; Vanommeslaeghe & MacKerell, 2012; Yu et al., 2012). The protein models were solvated within a periodic dodecahedron simulation box using TIP4P water molecules. Steepest descent energy minimization was performed until the maximum force fall below 1000 kJ mol^-1 nm^-1. Lennard-Jones 6-12 interactions were smoothly shifted to zero by starting to switch at 1.0 nm until a cutoff distance of 1.2 nm was reached. Coulomb interactions were treated with the Particle-Mesh Ewald (SPME) electrostatics using a cut-off for long range electrostatics of 1.2 nm. The molecular graphics of model illustration were produced by UCSF ChimeraX (Goddard et al., 2018; Pettersen et al., 2021).

# A. Appendix

## A.1. Supplementary figures

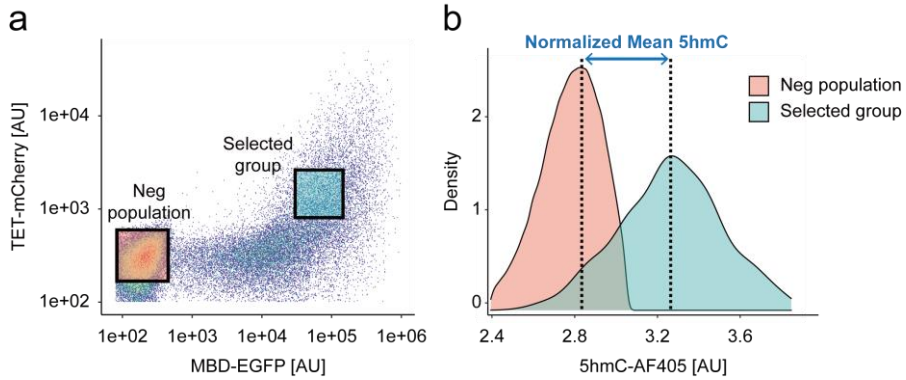


**Figure S 1. FACS data analysis and 5hmC intensity normalization strategy**. **a**) Representative FACS scatter plot showing the binary grouping of MBD(EGFP) and TET expression. Negative population was identified by the gating strategy described in methods. **b**) The median 5hmC intensity of each selected MBD(EGFP)-TET binary group was normalized to the median 5hmC intensity of the negative population in the same sample. The normalized median 5hmC directly correlates to the distance between two peaks in the histogram. Figure adapted from (Lin et al., 2022) (A.4).

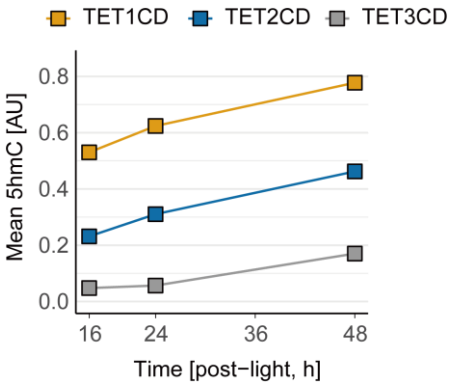


**Figure S 2. Differential oxidation kinetics between the catalytic domain of TETs**. The 5hmC level was measured from HEK293T cells expressing mCherry-tagged wt hTET1CD, hTET2CD, and hTET3CD. Mean global 5hmC intensities in the medium TET expression group (**Figure 22c**) were plotted. This reference experiment was only conducted once without further replication.

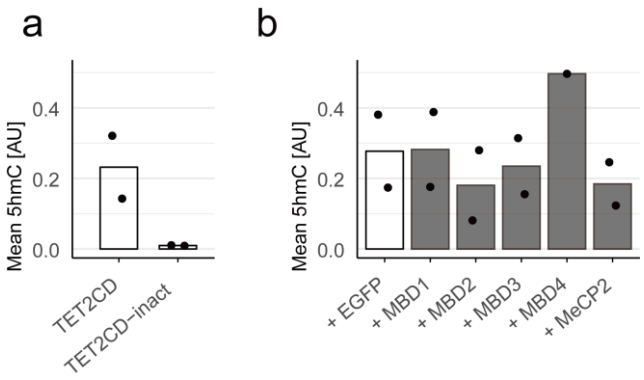


**Figure S 3. Study the modulation of TET2-catalyzed 5mC oxidation by human MBD proteins. a**) FCM measurement showing the catalytic activity of TET2CD-mCherry in comparison to its inactive mutant. Measurements were conducted 16 h after transfection. Mean of normalized 5hmC immunofluorescence from >100 cells were plotted. Data from two independent biological replicates. **d**) FCM analysis as in **a**) from cells co-expressing TET2CD-mCherry and different MBD-EGFP constructs. Data from two biological independent replicates (except for the coexpression of MBD4 which was only measured once).

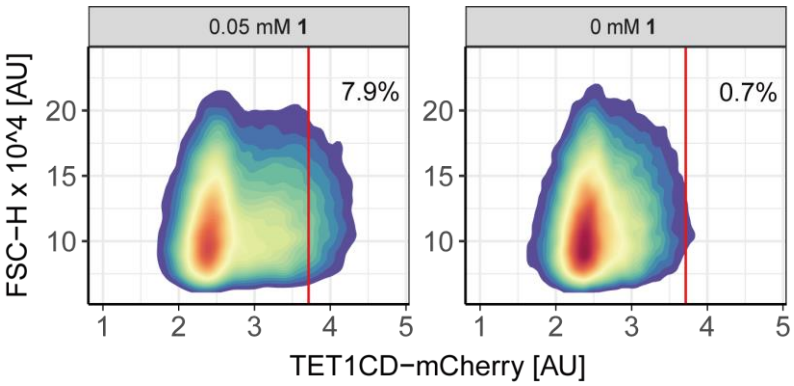


**Figure S 4. Representative FCM density plot showing increased hTET1CD$^{TAG}$-mCherry expression by incorporating 1**. The mCherry fluorescence of HEK293T cells cotransfected with vectors encoding hTET1CD-S2045$^{TAG}$ and (LRS)/tRNA$^{Leu}$ and grown in the presence (left) or absence (right) of 0.05 mM **1** were analyzed by FCM at 24 h after transfection. The intensity threshold for successful incorporation was determined by the fluorescence intensity of cells grown in the absence of **1** (right). Two independent biological replicates were performed. Figure adapted from (Lin et al., 2022) (A.4).

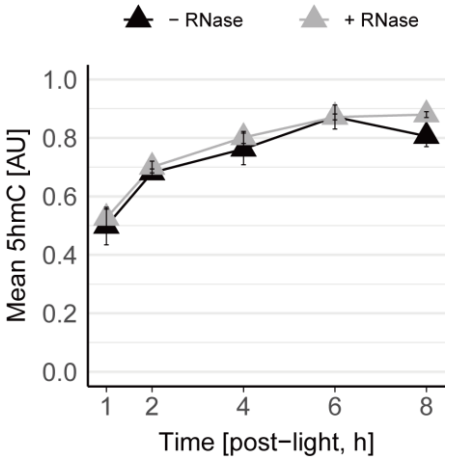


**Figure S 5. Examine the effect of RNA hydroxymethylation on the FCM-assisted global 5hmC measurement in HEK293T cells**. Immunostaining of 5hmC in HEK293T cells co-expressing hTET1CD-S2045→**1** and EGFP control was carried out with (grey) or without (black) RNase treatment and measured the resulting immunofluorescence by FCM. Mean 5hmC intensities from >100 cells in medium MBD/TET expression group are plotted, error bars are from SEM of two independent biological replicates. Figure adapted from (Lin et al., 2022) (A.4).

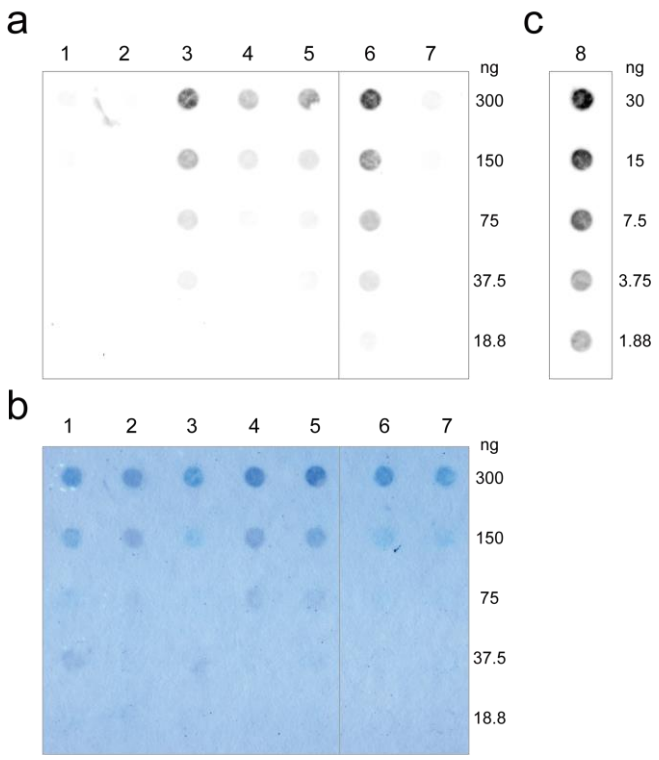


**Figure S 6. Detecting genomic 5hmC by DNA dot blots**. **a**) Genomic 5hmC of sorted (based on EGFP and mCherry fluorescence, described in 9.7) HEK293T cells was detected by DNA dot blot using the same 5hmC antibody as FCM measurements. In brief, HEK293T cells expressing the indicated constructs were irradiated (+) or not irradiated (-) with light (detailed condition listed below) and subjected to cell sorting at 26 h after transfection. For cells expressing hTET1CD-S2045→**1**, 0.05 mM **1** was added 3 h after transfection and cells were grown in **1** for 21 h, then **1** was removed from the media. Light irradiation was carried out after removal of **1** (24 h after transfection). **b**) Loading control of the dot blot in **a**) by methylene blue staining. **c**) The 5hmC content of the BRCA1 promoter sequence (421 bp, containing 21 CpGs) that was PCR-amplified with 5hmdCTP was detected by dot blot as reference. Figure adapted from (Lin et al., 2022) (A.4).

- 1: None (un-transfected) / - light
- 2: hTET1CD S2045 → **1** and EGFP / - light
- 3: hTET1CD S2045 → **1** and hMBD1-R22C+R44C / + light
- 4: hTET1CD S2045 → **1** and wt hMBD1 / + light
- 5: hTET1CD S2045 → **1** and EGFP / + light
- 6: wt hTET1CD / - light
- 7: Inactive hTET1CD / - light

Representative blot from 4 independent experiments is shown.

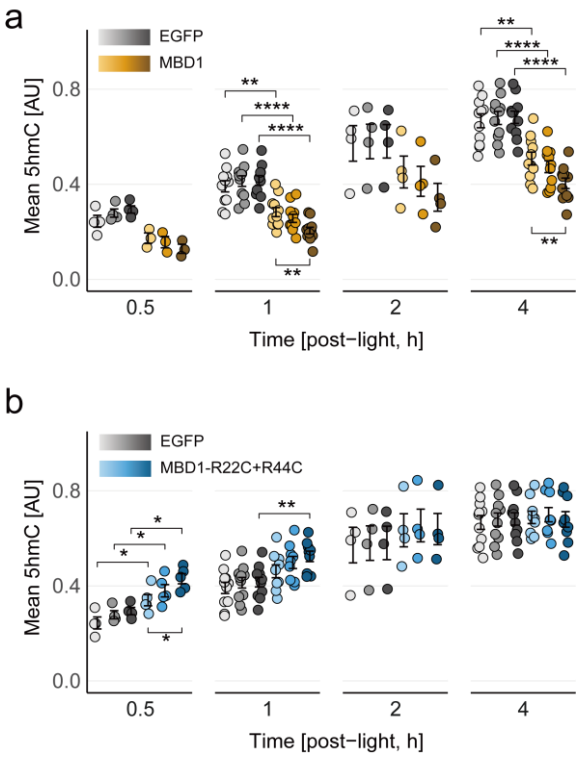


**Figure S 7. Dose-dependent analysis of TET1-mediated 5hmC formation kinetics modulated by hMBD1.** Measurements of 5hmC formation kinetics in HEK293T cells co-expressing hTET1CD-S2045→**1** and **a**) wt hMBD1 or EGFP-only **b**) hMBD1-R22C+R44C or EGFP-only. Global 5hmC levels from three different MBD/TET expression groups (gradient bar, from left to right: low, medium, high) at selected time points are summarized as mean and standard error in plots. Data from at least 3 independent biological replicates (each includes >100 cells) and *P*-values from Mann-Whitney tests are indicated (*: p <= 0.05; **: p <= 0.01; ****: p <= 0.0001). Figure adapted from (Lin et al., 2022) (A.4).

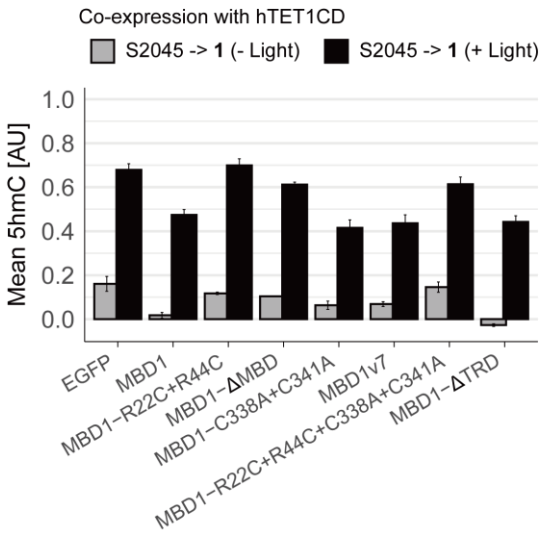


**Figure S 8. The cellular 5hmC is stringently controlled by light activation of hTET1CD-S2045 → 1.** The global 5hmC abundance in HEK293T cells coexpressing hTET1CD-S2045 → **1** and indicated constructs of grey: without light irradiation; black: with light irradiation at 21 h after addition of **1** were measured at 28 h after transfection (0.05 mM **1** was added to growth media at 3h after transfection, cells were further grown for 21h before the removal of **1** from media). Mean 5hmC intensities from the medium MBD/TET expression group of over 100 cells are shown with error bars indicating the SEM of at least 3 independent biological replicates. Figure adapted from (Lin et al., 2022) (A.4).

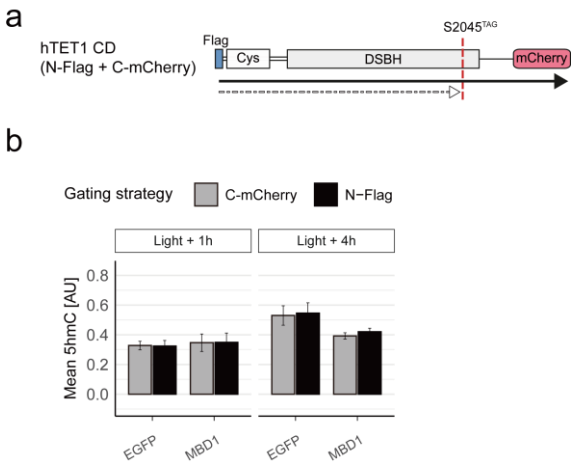


**Figure S 9. Different gating strategies showing that the C-terminus truncated hTET1CD, the amber suppression byproduct, does not influence MBD1 modulation**. **a**) Domain structure hTET1CD-S2045$^{TAG}$ that is N-terminally tagged with Flag and C-terminally tagged with mCherry. The successful amber suppression should result full-length hTET1CD containing both the Flag-tag and mCherry, whereas failure to suppress amber codon will result in termination at S2045 and yield truncation byproducts containing only the Flag-tag. **b**) The potential impact of truncated hTET1CD was investigated using two different gating strategies: population selected only for C-mCherry (gray), or only for N-Flag (black). Both gating methods generated similar 5hmC levels, implying a neglectable effect of truncated hTET1CD (expected to exist in excess) on the observed hTET1CD-mediated 5hmC formation regulated by MBD1. Mean 5hmC intensities from the medium MBD/TET expression group containing over 100 cells are plotted, error bars indicating the SEM of 4 independent biological replicates. Figure adapted from (Lin et al., 2022) (A.4).

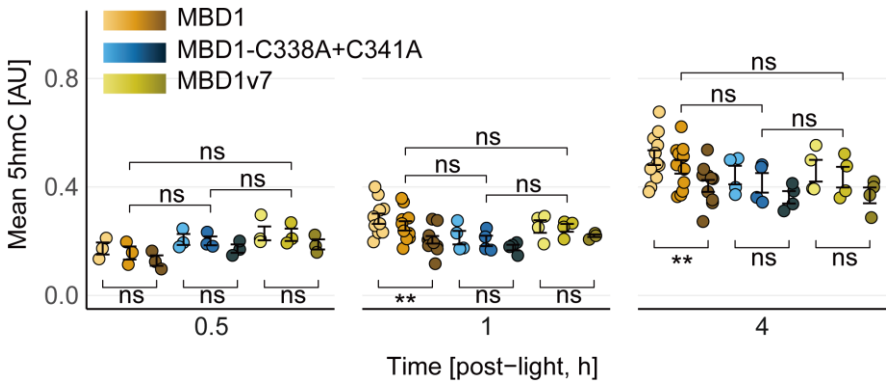


**Figure S 10. Dose-dependent analysis of TET1-mediated 5hmC formation kinetics modulated by hMBD1 CXXC3 mutants (isoform 7 and C338A+C341A)**. Global 5hmC level was measured from HEK293T cells coexpressing hTET1CD-S2045 → **1** with either the wt hMBD1 or the hMBD1 CXXC3 mutants (isoform 7 and C338A+C341A). Mean 5hmC intensities from three different MBD/TET expression groups (gradient bar, from left to right: low, medium, high) (each containing >100 cells) are plotted as mean and standard error. Data from at least 3 independent biological replicates and *P*-values from Mann-Whitney tests are indicated (**: p <= 0.01; ns: p > 0.05). Figure adapted from (Lin et al., 2022) (A.4).

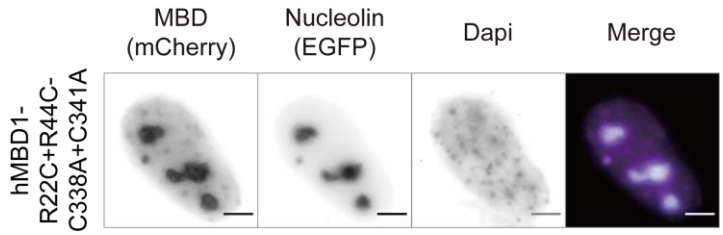


**Figure S 11. The hMBD1-R22C+R44C+C338A+C341A mutant colocalized with nucleolin in NIH/3T3 cells.** Coexpression of mCherry-tagged hMBD1 mutant (magenta) and EGFP-tagged nucleolin (cyan) in NIH/3T3 cells suggested that the R22+R44C+C338A+C341A mutant primarily localized in the nucleoli. Scale bar: 5 μm. Figure adapted from (Lin et al., 2022) (A.4).

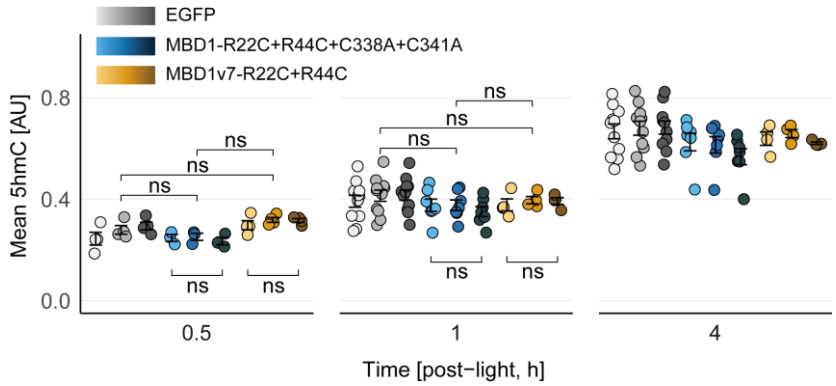


**Figure S 12. Dose-dependent analysis of TET1-mediated 5hmC formation kinetics modulated by the MBD and CXXC3 mutants of hMBD1**. The global 5hmC abundance in HEK293T cells co-expressing hTET1CD-S2045 → **1** with either the hMBD1 mutants (isoform 7-R22C+R44C and R22C+R44C+C338A+C341A) or EGFP-only were measured. Mean 5hmC intensities from three different MBD/TET expression groups (gradient bar, from left to right: low, medium, high), each containing over 100 cells, are summarized in mean and standard error. Data from at least 3 independent biological replicates and *P*-values from Mann-Whitney tests are indicated (ns: $p > 0.05$). Figure adapted from (Lin et al., 2022) (A.4).

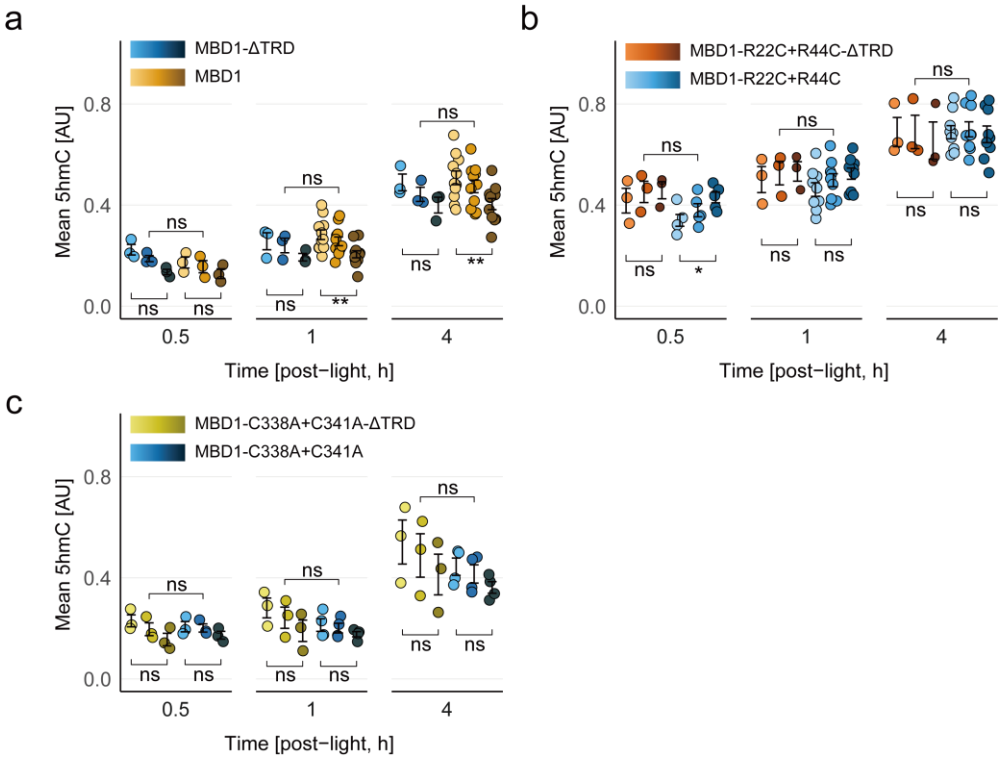


**Figure S 13. Dose-dependent analysis of TET1-mediated 5hmC formation kinetics modulated by the TRD mutants**. Global 5hmC levels were measured from HEK293T cells coexpressing hTET1CD-S2045 → **1** and **a**) hMBD1 wt or ΔTRD mutant **b**) hMBD1 R22C+R44C or R22C+R44C-ΔTRD mutants **c**) hMBD1 C338A+C341A or C338A+C341A-ΔTRD mutants. The mean 5hmC intensities from three different MBD/TET expression groups (each containing >100 cells) (gradient bar, from left to right: low, medium, high) are plotted as mean and standard error. Data from at least 3 independent biological replicates and P-values from Mann-Whitney tests are indicated (*: p <= 0.05; ns: p > 0.05). Figure adapted from (Lin et al., 2022) (A.4).

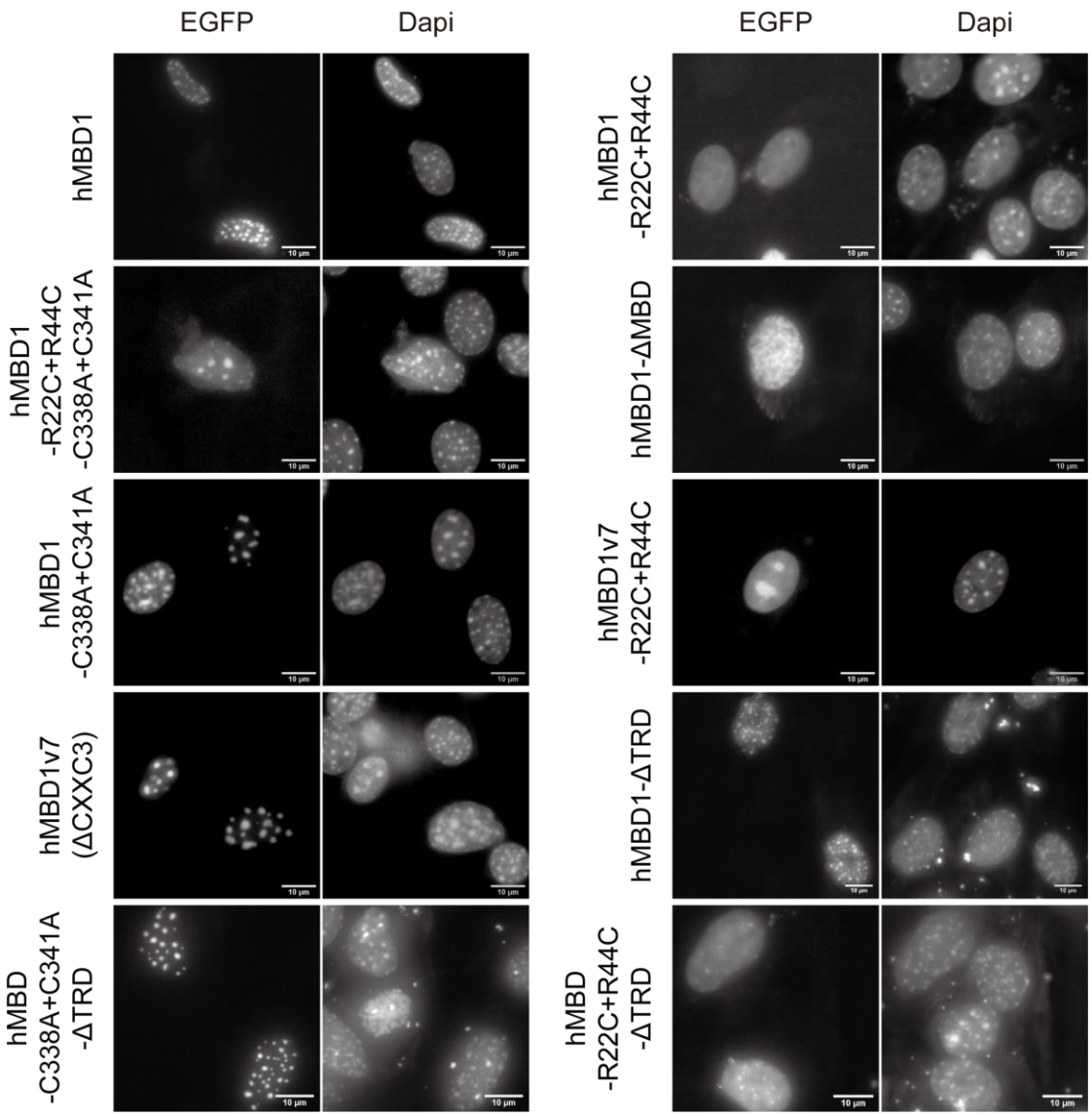


**Figure S 14. Supplementary images of recombinant hMBD1 constructs in NIH/3T3 cells**. Scale bar: 10 µm. Figure adapted from (Lin et al., 2022) (A.4).

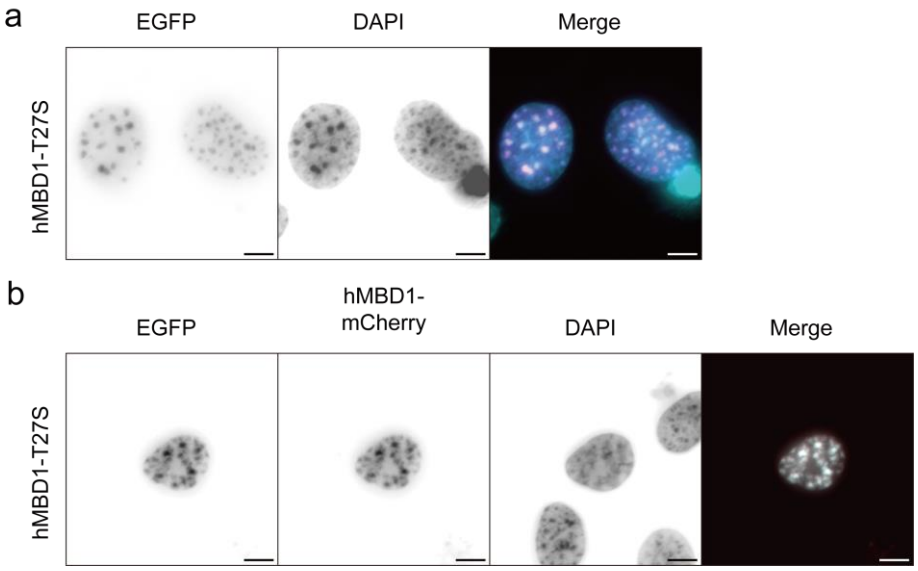


**Figure S 15. MBD1 retains affinity to mCpGs when the Thr 27 residue was replaced by a serine**. **a**) The EGFP-tagged hMBD1-T27S mutant (magenta) colocalized with mCpG-rich chromocenters indicated by DAPI foci (cyan) in NIH/3T3 cells. Scale bar: 5 µm. **b**) The EGFP-tagged hMBD1-T27S mutant (cyan) also co-localized with the mCherry-tagged hMBD1 wt (red) in NIH/3T3 cells. Scale bar: 5 µm.

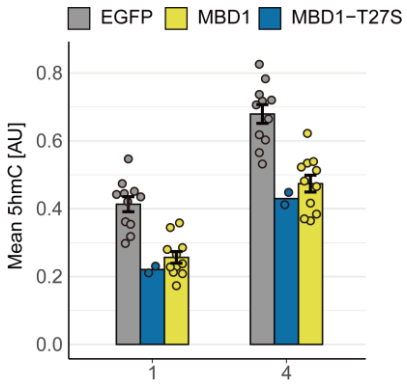


**Figure S 16. The hMBD1 T27S mutant similarly downregulated the hTET1CD-mediated 5hmC formation**. Global 5hmC levels were measured from HEK293T cells coexpressing hTET1CD-S2045 → **1** and either hMBD1 wt or the T27S mutant. The mean 5hmC intensities of the medium MBD/TET expression group (containing >100 cells) from at least two biological replicates were summarized in mean and standard error.

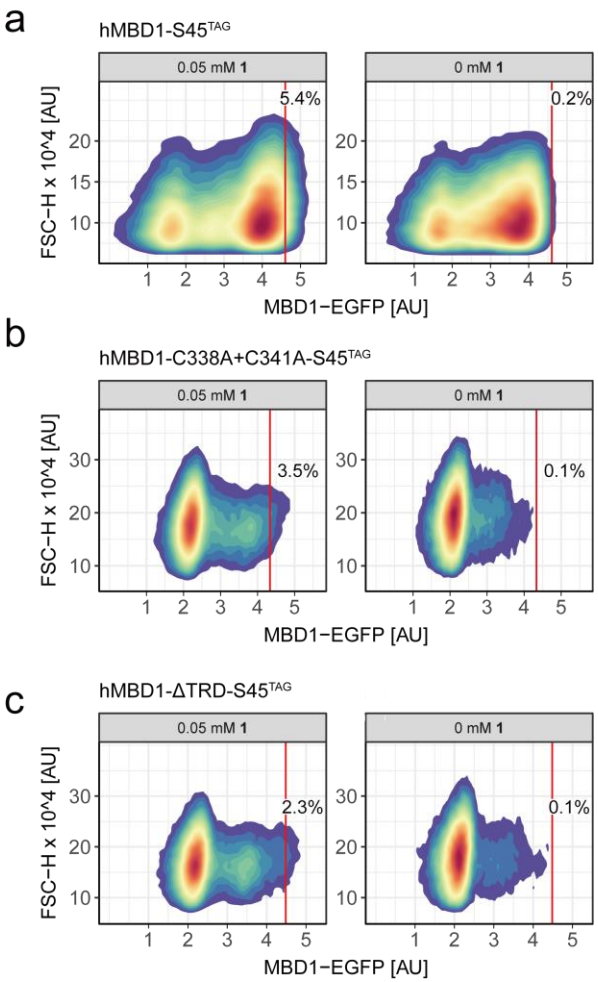


**Figure S 17. Representative FCM density plots showing the incorporation of 1 in indicated hMBD1 constructs and sorting strategies**. The EGFP fluorescence of HEK293T cells cotransfected with vectors encoding LRS/tRNA$^{Leu}$ and **a**) hMBD1-S45$^{TAG}$ **b**) hMBD1-C338A+C341A-S45$^{TAG}$ **c**) hMBD1-ΔTRD-S45$^{TAG}$ and grown in the presence (left) or absence (right) of 0.05 mM **1** were analyzed by FCM at 24 h after transfection. Red lines indicating the intensity thresholds determined by the fluorescence intensity of cells grown in the absence of **1** (right). The determined thresholds were used for sorting cell populations that successfully incorporated **1**. Two independent biological replicates were performed.

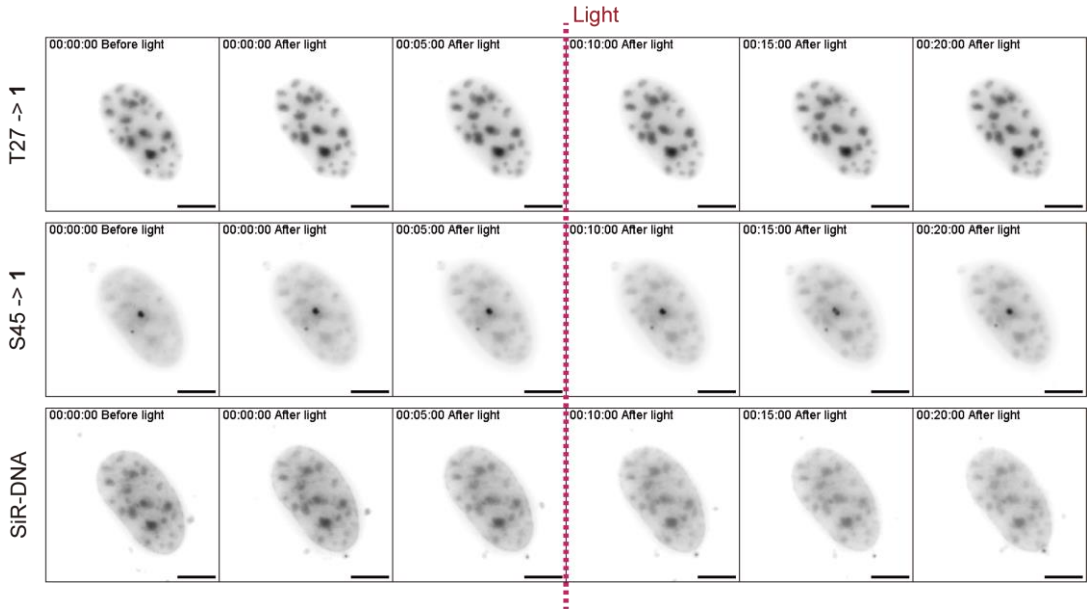


**Figure S 18. Replacing T27 with 1 is not sufficient to suppress the mCpG binding of hMBD1**. NIH/3T3 cells coexpressing the EGFP-tagged hMBD1-T27 → **1** and mCherry-tagged hMBD1-S45 → **1** were sorted by FACS and recultivated until imaging. While hMBD1-S45 → **1** did not locate to the chromocenters, hMBD1-T27 → **1** remained highly accumulated at chromocenters. Moreover, the localization of hMBD1-T27→**1** did not change upon light irradiation as hMBD1-S45 → **1**. The nucleus was stained with SiR-DNA. Selected images at different time points (minutes) before and after light activation are shown. Scale bar: 5 µm.

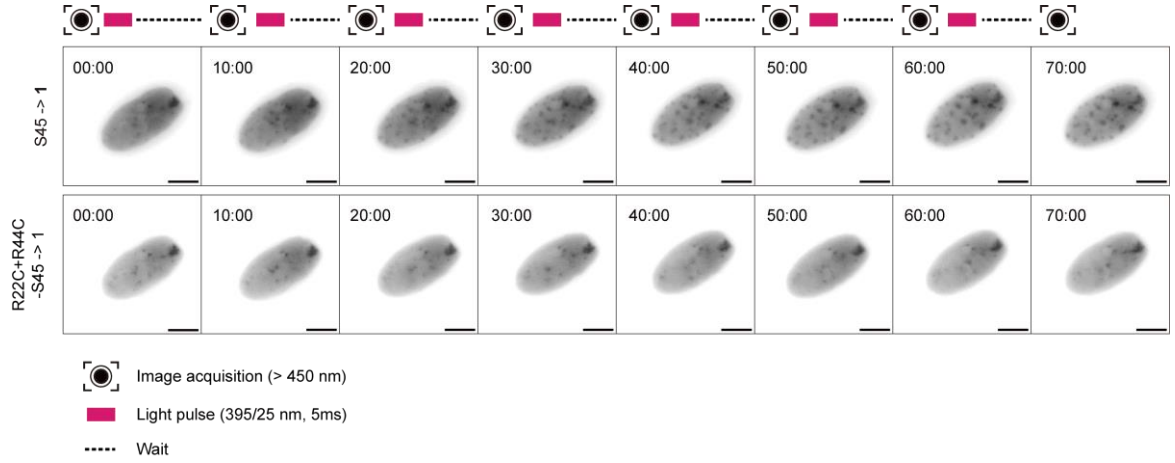


**Figure S 19. Titrating the light dose to uncage 1 and activate hMBD1 in NIH/3T3 cells**. FACS-sorted NIH/3T3 cells coexpressing hMBD1-S45 → **1** and hMBD1-R22C+R44C-S45 → **1** were repeatedly irradiated with 5 ms of violet light (395/25 nm, 100% intensity, LED) every 10 min. After each light pulse, cell images were acquired after 10 min to allow sufficient time for (activated) hMBD1 binding, and directly followed by another activation light pulse. Fluorescence saturation at chromocenters was observed after 50 min, implying a minimum of 30 ms light pulse is required to fully uncage the cellular hMBD1-S45 → **1**. Therefore, a pulse of 50 ms full intensity violet light was determined for the activation experiments, ensuring complete uncaging and MBD1 activation. Scale bar: 5 µm.

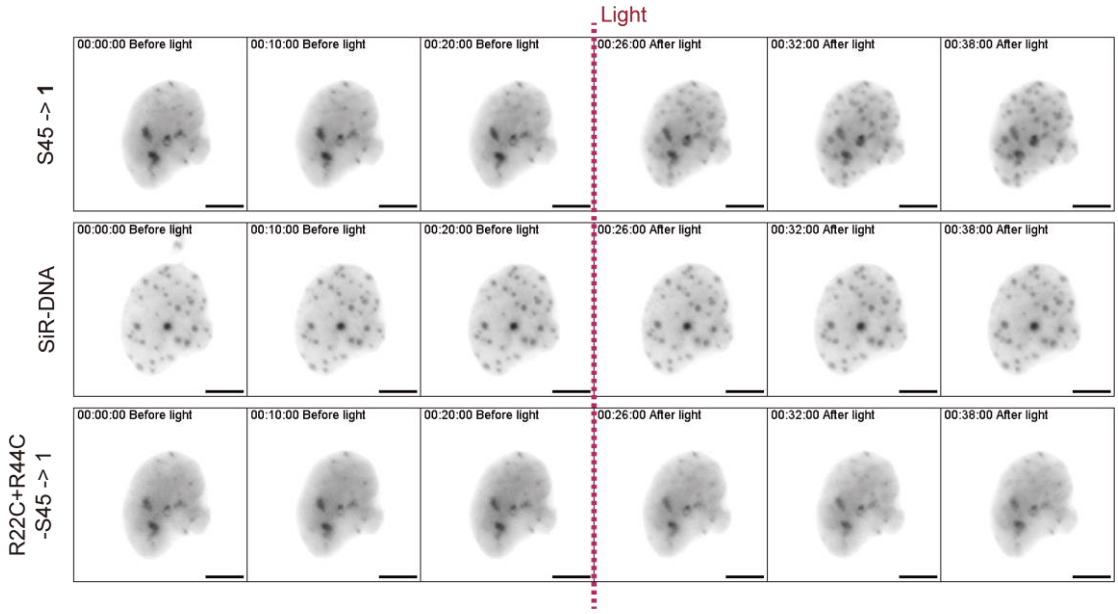


**Figure S 20. Light activation of hMBD1-S45 → 1 imaged with switched C-terminal fluorophore**. NIH/3T3 cells coexpressing EGFP-tagged hMBD1-S45 → **1** and mCherry-tagged hMBD1-R22C+R44C-S45 → **1** were simultaneously irradiated with violet light (395/25 nm, 50 ms). Red dashed line indicating the time point of light treatment. Nucleus is stained with SiR-DNA. Selected images at different time points (minutes) before and after light activation are shown. Scale bar: 5 µm.

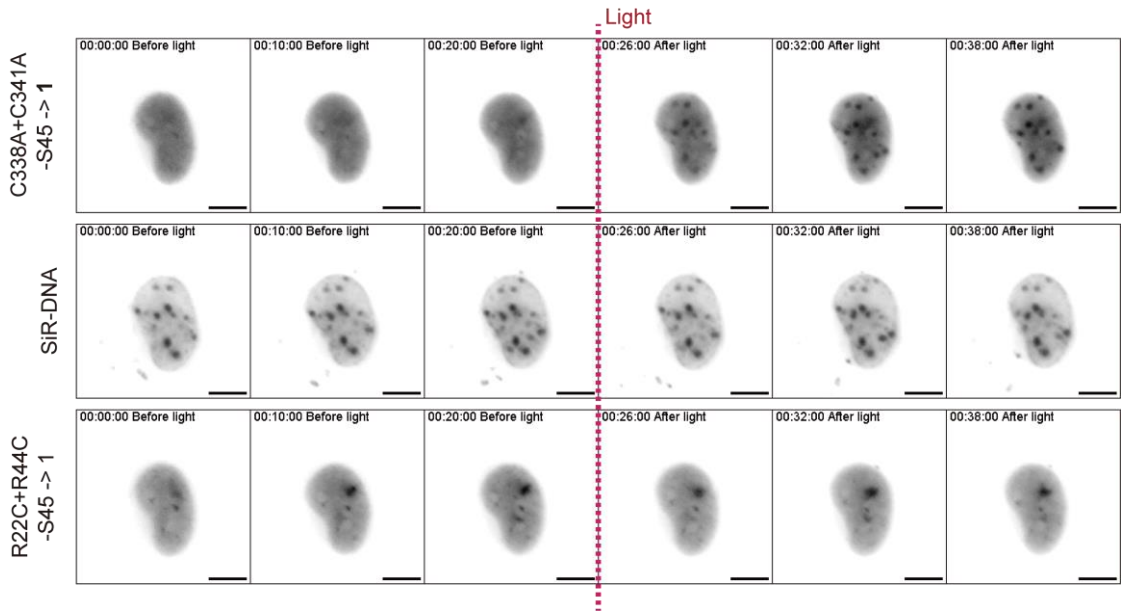


**Figure S 21. Light activation of hMBD1-C338A+C341A-S45 → 1 imaged with switched C-terminal fluorophore**. NIH/3T3 cells coexpressing EGFP-tagged hMBD1-C338A+C341A-S45 → **1** and mCherry-tagged hMBD1-R22C+R44C-S45 → **1** were simultaneously irradiated with violet light (395/25 nm, 50 ms). Red dashed line indicating the time point of light treatment. Nucleus is stained with SiR-DNA. Selected images at different time points (minutes) before and after light activation are shown. Scale bar: 5 µm.

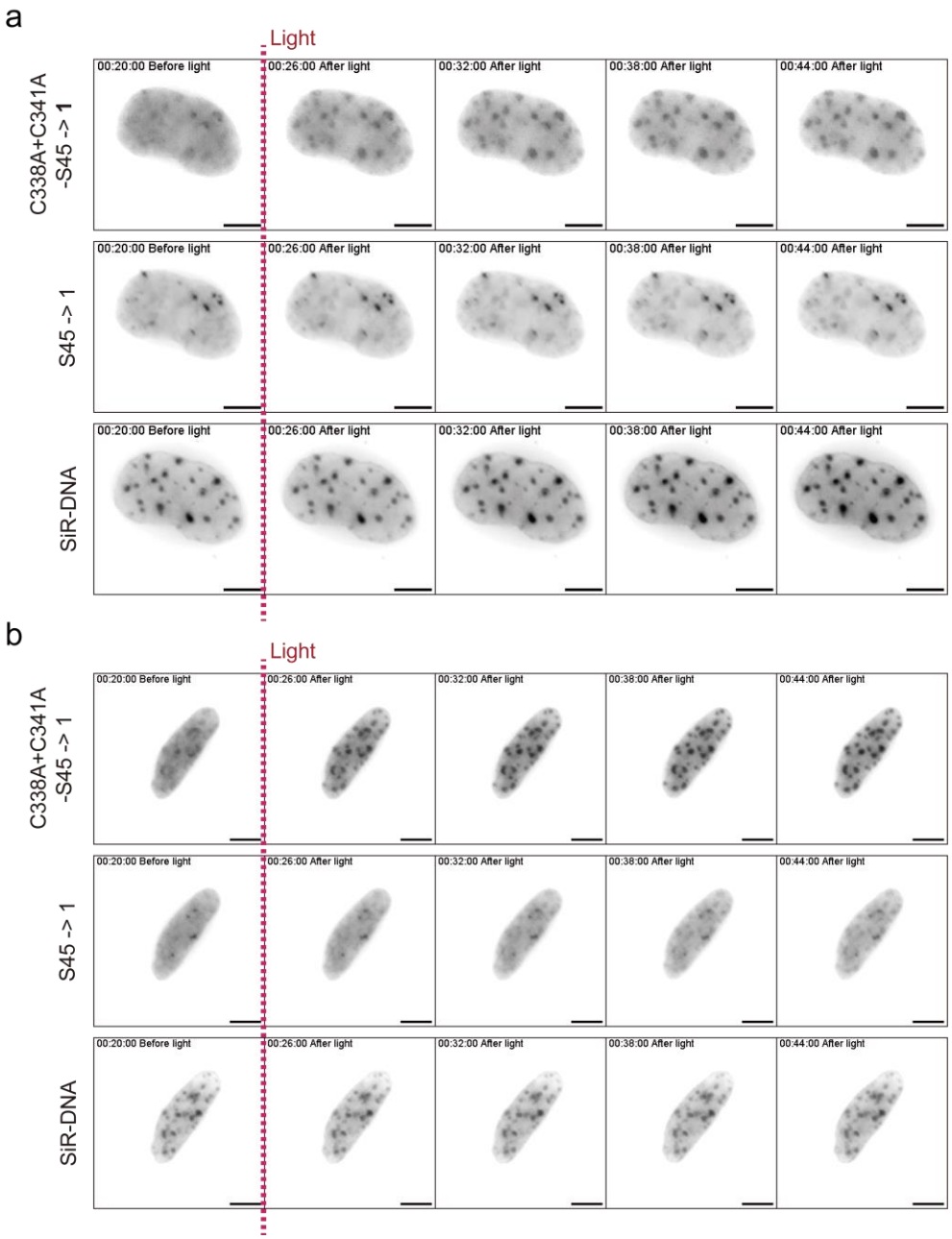


**Figure S 22. Competitive binding of hMBD1- S45 → 1 and hMBD1-C338A+C341A-S45 → 1 upon light activation**. NIH/3T3 cells coexpressing **a**) EGFP-tagged hMBD1-S45 → **1** and mCherry-tagged hMBD1C338A+C341A-S45 → **1 b**) mCherry-tagged hMBD1-S45 → **1** and EGFP-tagged hMBD1-C338A+C341A-S45 → **1** were simultaneously irradiated with violet light (395/25 nm, 50 ms). Red dashed line indicating the time point of light treatment. Nucleus is stained with SiR-DNA. Selected images at different time points (minutes) before and after light activation are shown. Scale bar: 5 µm.

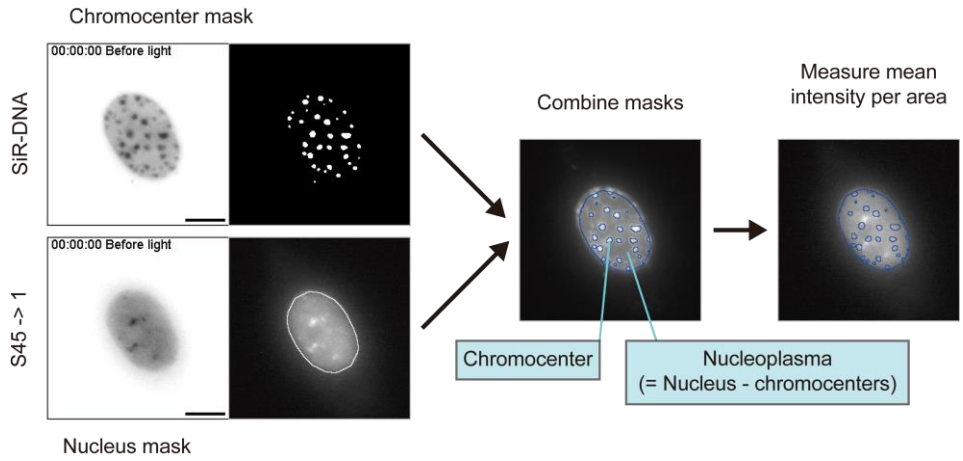


**Figure S 23. Schematic illustration of the image analysis workflow**. Firstly, chromocenter and nucleus masks are defined by images acquired for SiR-DNA staining and EGFP/mCherry fluorescence, respectively. The nucleus mask is further used to confine the selection of chromocenters in order to exclude false positive chromocenter detection outside the nucleus, and an additional "nucleoplasma" mask was created by subtracting chromocenter selection from the nucleus mask. Finally, all selections are redirected to measure the mean EGFP/mCherry fluorescence intensity in designated cell image.

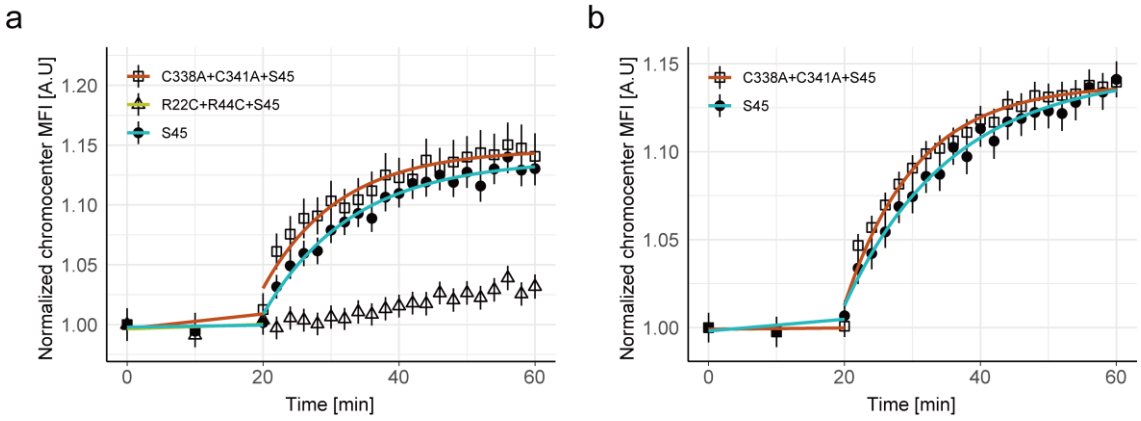


**Figure S 24. Cellular mCpG binding kinetics of hMBD1 domain variants measured from switched fluorophores**. **a**) Kinetic measurements of chromocenter fluorescence in NIH/3T3 cells coexpressing mCherry-tagged hMBD1-S45 → **1** (wt, C338A+C341A and ΔTRD variants) and EGFP-tagged hMBD1-R22C+R44C-S45 → **1** upon light activation. The normalized MFI values of recorded chromocenters (N between 100 and 700) from 2 biological independent replicates are plotted by mean and standard error with respect to time. Recorded cell numbers are N = 13 and 7 for wt and C338A+C341A mutant, respectively. **b**) Measurements of chromocenter fluorescence in cells coexpressing EGFP-tagged wt hMBD1-S45 → **1** and mCherry-tagged hMBD1-C338A+C341A-S45 → **1** upon light activation. The normalized MFI values of recorded chromocenters (N between 600 and 700) from 2 biological independent replicates are plotted by mean and standard error with respect to time. Recorded cell number is N = 26.

## A.2. Supplementary tables

**Table S 1. Oligonucleotides for plasmids construction.**

| Name | Sequence (5' → 3') |
|---|---|
| o2260_JaW | ATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGG |
| o2261_JaW | ACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTG |
| o3167_ShP | CGTTCCGGACTACGCTTCTGAGCAGAAGCTGATCTCAGAGGAGGACCTGTGAATCGGTAGGAATTCGCGGCCG |
| o3246_TzL | CGTCGACGGATCGGGAGCGGCCGCTTCGAGC |
| o3247_TzL | GCTCGAAGCGGCCGCTCCCGATCCGTCGACG |
| o3254_TzL | ggccttacaacacatttgtaggtaccggcggtggagggtcaggtggcggaggttcaGTGAGCAAGGGCGAGGAG |
| o3255_TzL | acgtcgtacgggtagttaatttaattaaCTTGTACAGCTCGTCCATGC |
| o3256_TzL | ttaattaaATTAACTACCCGTACGAC |
| o3257_TzL | tgaacctccgccacctgaccctccaccgccggtaccTACAAATGTGTTGTAAGGC |
| o3284_TzL | GAGCTGTACAAGTTAATTAACTACCCGTACGACG |
| o3285_TzL | CGTCGTACGGGTAGTTAATTAACTTGTACAGCTC |
| o3288_TzL | GGTACCGGCGGTGGAGGG |
| o3290_TzL | tctGAGCAGAAGCTGATCTCAG |
| o3291_TzL | GGCGCGCCCAACTTTGCG |
| o3292_TzL | aacgcaaagttgggcgcgccATGGCTGAGGACTGGCTG |
| o3293_TzL | cgccggtaccgttaatCTGCTTTCTAGCTCCAGGTTTTTTAAG |
| o3294_TzL | tagaaagcagattaacggtaccggcggtggagggtcaggtggcggaggttcaGTGAGCAAGGGCGAGGAG |
| o3295_TzL | tctgagatcagcttctgctcagaCTTGTACAGCTCGTCCATGC |
| o3374_TzL | aacgcaaagttgggcgcgccATGGACTCAGGGCCAGTGTAC |
| o3377_TzL | gaccctccaccgccggtaccGATCCAGCGGCTGTAGGG |
| o3380_TzL | aacgcaaagttgggcgcgccATGGAGCGGAAGAGGTGG |
| o3381_TzL | ccaccgccggtaccgttaatGACGTGCTCCATCTCCGG |
| o3382_TzL | aacgcaaagttgggcgcgccATGGGCACGACTGGGCTG |
| o3383_TzL | ccaccgccggtaccgttaatAGATAGACTTAATTTTTCATGATTTTCCCAAAGCC |
| o3384_TzL | aacgcaaagttgggcgcgccATGGTAGCTGGGATGTTAG |
| o3385_TzL | ccaccgccggtaccgttaatGCTAACTCTCTCGGTCAC |

| | |
|---|---|
| o3386_TzL | ATTAACGGTACCGGCGGTG |
| o3410_TzL | cacctgaccctccaccgccggtaccGATCCAGCGGCTGTAGGG |
| o3473_TzL | cacctgaccctccaccgccggtaccGACCCAATGGTTATAGGGC |
| o3510_TzL | AACGCAAAGTTGGGCGCGCCATGCGCGCGCACCCG |
| o3511_TzL | CCACCGCCGGTACCGTTAATGGCTTCATCTCCACTGTCCATTTC |
| o3596_JaW | GGCGCGCCCAACTTTGCGTTTC |
| o3642_TzL | CTGGctagCGCCATGCCGAAAAAGAAACGC |
| o3643_TzL | GCGTTTCTTTTTCGGCATGGCGctagCCAG |
| o3730_TzL | GAAGCGCCGCGAAGTCTTTTGTAAGTCAGGGGCCACCTGTG |
| o3731_TzL | CACAGGTGGCCCCTGACTTACAAAAGACTTCGCGGCGCTTC |
| o3732_TzL | CACAGGAGACAGGATCTGCAGCAAAGTTGAGCTGACTCG |
| o3733_TzL | CGAGTCAGCTCAACTTTGCTGCAGATCCTGTCTCCTGTG |
| o3751_TzL | aaagaaacgcaaagttgggcgcgccATGGAACTGCCCAC |
| o3752_TzL | GCAAAGTTGGGCGCGCCATGGAGTTCCCCACCTGCGATTG |
| o3754_TzL | GTGCCCACGCCTATAAGGCTCAGCATAACCTCTACAATGGG |
| o3755_TzL | CCCATTGTAGAGGTTATGCTGAGCCTTATAGGCGTGGGCAC |
| o3756_TzL | GACAGGATCCGATAGAAAGTTGAGCTGACTCGATAC |
| o3757_TzL | GAGTCAGCTCAACTTTCTATCGGATCCTGTCTCCTGTG |
| o3758_TzL | GAAGATGTTGCTATGGAATTGGAAAGAGTGGGAGAAGATGAG |
| o3759_TzL | CTCATCTTCTCCCACTCTTTCCAATTCCATAGCAACATCTTC |
| o3762_TzL | CTGTGCTCATCCCTATAGGGCTATTCACAACATGAATAATG |
| o3763_TzL | CATGTTGTGAATAGCCCTATAGGGATGAGCACAGAAGTCC |
| o3810_TzL | GTGGGAGTGCCCGGCGCTCCCGCAGGGCTGGGAGAGGGAAGAAGTGCCCAGAAG GTCGGGGCTGTCGGCCGGCCACAGGGATGTCTTTTACTATAGCCCGAGCGGGAA GAAGTTCC |
| o3811_TzL | CTATAGTAAAAGACATCCCTGTGGCCGGCCGACAGCCCCGACCTTCTGGGCACT TCTTCCCTCTCCCAGCCCTGCGGGAGCGCCGGGCACTCCCACCTCTTCCGCTCC ATGGCG |
| o4128_TzL | AGGAGACAGGATCGCCTAGAAAGTTGAGCTGACTCGATAC |
| o4129_TzL | GAGTCAGCTCAACTTTCTAGGCGATCCTGTCTCCTGTGGG |
| o4189_TzL | CTGGGCAGCAGCCGCTTTAGCTCGTCCTCGTCTACAC |
| o4298_BiR | GTGTAGACGAGGACGAGCTAAAGCGGCTGCTG |
| o4300_BiR | GAAACGCAAAGTTGGGCGCGCCTGCTATCCAGCCCCCAAG |
| o4302_BiR | CCTGGCTGCCCTAGCAAGAGGTCCAAAGACCTTAAAAAACCTGGAGC |
| o4305_BiR | GTTTTTTAAGGTCTTTGGACCTCTTGCTAGGGCAGCCAGG |
| o4479_TzL | AGAACCGCAAGGCCGGGGCCGCAGCAGCCTGCCTACGG |

| | |
|---|---|
| o4480_TzL | TAGGCAGGCTGCTGCGGCCCCGGCCTTGCGGTTCTGCC |
| o4558_TzL | TCTGAGCGTCCACACTAGGCCCCTGACTTGCGAAAG |
| o4592_TzL | AAGTCAGGGGCCTAGTGTGGACGCTCAGACACCTATTAC |

**Table S 2. Oligonucleotides for sequencing.**

| Name | Sequence (5' → 3') |
|---|---|
| o2067_AnW | AACTTGGTCTGACAGTTACCAAT |
| AmpStart | AAAACAGGAAGGCAAAATGC |
| AmpStop | TCAGGCAACTATGGATGAAC |
| BGH-rev | TAGAAGGCACAGTCGAGG |
| EBV-rev | GTGGTTTGTCCAAACTCATC |
| EGFP-C-for | GTCCTGCTGGAGTTCGTG |
| PEN1-737R | TCCAGCTCGACCAGGAT |
| pUCM13-rev-157 | TGCTTCCGGCTCGTATGTTG |
| SV40-for | GCCCCTAACTCCGCCCATCC |

**Table S 3. Oligonucleotides for DNA dot blot.**

| Name | Sequence (5' → 3') |
|---|---|
| o1516_GrK | CTTCCTCTTCCGTCTCTTTCCTTTTACGTCATCCGGGGGCAGACT |
| o1601_GrK | GTCAAAGAATACCCATCTGTCAGCTTCGGAAATCCACTCTCCCAC |
| BRCA1 | CTTCCTCTTCCGTCTCTTTCCTTTTACGTCATCCGGGGGCAGACTGGGTGGCCA<br>ATCCAGAGCCCCGAGAGACGCTTGGCTCTTTCTGTCCCTCCCATCCTCTGATTG<br>TACCTTGATTTCGTATTCTGAGAGGCTGCTGCTTAGCGGTAGCCCCTTGGTTTC<br>CGTGGCAACGGAAAAGCGCGGGAATTACAGATAAATTAAAACTGCGACTGCGCG<br>GCGTGAGCTCGCTGAGACTTCCTGGACGGGGGACAGGCTGTGGGGTTTCTCAGA<br>TAACTGGGCCCCTGCGCTCAGGAGGCCTTCACCCTCTGCTCTGGGTAAAGGTAG<br>TAGAGTCCCGGGAAAGGGACAGGGGGCCCAAGTGATGCTCTGGGGTACTGGCGT<br>GGGAGAGTGGATTTCCGAAGCTGACAGATGGGTATTCTTTGAC |

**Table S 4. Plasmids.**

| Plasmid No | Gene 1 | Gene 2 | Resistance |
|---|---|---|---|
| pStH1147 | LeuRS (T252A) /tRNA | | Amp |
| pTzL1744 | TET2 CD | | Amp |
| pTzL1745 | TET2 CD | mCherry | Amp |
| pTzL1746 | TET2 CD | mCherry | Amp |

| | | | |
|---|---|---|---|
| pTzL1747 | hMBD1 FL | EGFP | Amp |
| pTzL1773 | hMeCP2 | EGFP | Amp |
| pTzL1774 | hMBD3 (isoform2) | EGFP | Amp |
| pTzL1833 | hMBD1 FL | EGFP | Amp |
| pTzL1834 | hMeCP2 FL | EGFP | Amp |
| pTzL1835 | hMBD3 (isoform2) FL | EGFP | Amp |
| pTzL1836 | hMBD1 FL | EGFP | Amp |
| pTzL1837 | hTET3v1 | mCherry | Amp |
| pTzL1889 | hMBD2a | EGFP | Amp |
| pTzL1947 | hMBD1_R22C | EGFP | Amp |
| pTzL1948 | hMBD4 | EGFP | Amp |
| pTzL1960 | hTET1 CD | mCherry | Amp |
| pTzL1964 | hMBD1 R22C+R44C | EGFP | Amp |
| pTzL1967 | hMBD1 S45$^{TAG}$ | EGFP | Amp |
| pTzL1970 | hTET1 CD inactive | mCherry | Amp |
| pTzL1990 | EGFP | | Amp |
| pTzL2005 | hTET2 CD | mCherry | Amp |
| pTzL2006 | hTET2 CD inactive | mCherry | Amp |
| pTzL2050 | hTET3 CD | mCherry | Amp |
| pTzL2079 | hTET3 CD inactive | mCherry | Amp |
| pShP2384 | hTET2CD_S1812$^{TAG}$ | mCherry | Amp |
| pShP2413 | hTET2 CD | | Amp |
| pShP2416 | hTET2 CD inactive | | Amp |
| pShP2444 | hTET1CD_S2045$^{TAG}$ | | Amp |
| pTzL2504 | hTET1CD_S2045$^{TAG}$ | mCherry | Carb |
| pTzL2511 | hMBD1_S45$^{TAG}$ | EGFP | Carb |
| pTzL2512 | hMBD1_S45$^{TAG}$ | mCherry | Carb |
| pTzL2513 | hTET1CD_S2045$^{TAG}$ | mCherry | Carb |
| pBiR2585 | hMBD1_dMBD | EGFP | Carb |
| pBiR2586 | hMBD1_dTRD | EGFP | Carb |
| pBiR2593 | hMBD1_dCXXC3 | EGFP | Carb |
| pBiR2628 | hMBD1_R22C+R44C_dCXXC | EGFP | Carb |
| pTzL2645 | hMBD1_C338A+C441A | EGFP | Carb |
| pTzL2646 | hMBD1_R22C+R44C+C338A+C441A | EGFP | Carb |
| pTzL2681 | hMBD1_T27$^{TAG}$ | EGFP | Carb |
| pNaU2737 | hMBD1_C338A+C341A-S45$^{TAG}$ | EGFP | Amp |

| | | | |
|---|---|---|---|
| pNaU2738 | hMBD1_dTRD-S45$^{TAG}$ | EGFP | Amp |
| pTzL2889 | hMBD1_C338A+C341A-S45$^{TAG}$ | mCherry | Amp |
| pTzL2890 | hMBD1_dTRD-S45$^{TAG}$ | mCherry | Amp |
| pTzL2892 | hMBD1_R22C+R44C-S45$^{TAG}$ | EGFP | Amp |
| pTzL2902 | hMBD1_R22C+R44C-S45$^{TAG}$ | mCherry | Amp |

**Table S 5. Protein coding sequences.**

| **Protein** | **Sequence** |
|---|---|
| hTET1CD | MELPTCSCLDRVIQKDKGPYYTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGKSS<br>HGCPIAKWVLRRSSDEEKVLCLVRQRTGHHCPTAVMVVLIMVWDGIPLPMADRLYTELTEN<br>LKSYNGHPTDRRCTLNENRTCTCQGIDPETCGASFSFGCSWSMYFNGCKFGRSPSPRRFRIDP<br>SSPLHEKNLEDNLQSLATRLAPIYKQYAPVAYQNQVEYENVARECRLGSKEGRPFSGVTACLD<br>FCAHPHRDIHNMNNGSTVVCTLTREDNRSLGVIPQDEQLHVLPLYKLSDTDEFGSKEGMEA<br>KIKSGAIEVLAPRRKKRTCFTQPVPRSGKKRAAMMTEVLAHKIRAVEKKPIPRIKRKNNSTTT<br>NNSKPSSLPTLGSNTETVQPEVKSETEPHFILKSSDNTKTYSLMPSAPHPVKEASPGFSWSPK<br>TASATPAPLKNDATASCGFSERSSTPHCTMPSGRLSGANAAAADGPGISQLGEVAPLPTLSAP<br>VMEPLINSEPSTGVTEPLTPHQPNHQPSFLTSPQDLASSPMEEDEQHSEADEPPSDEPLSDDP<br>LSPAEEKLPHIDEYWSDSEHIFLDANIGGVAIAPAHGSVLIECARRELHATTPVEHPNRNHPT<br>RLSLVFYQHKNLNKPQHGFELNKIKFEAKEAKNKKMKASEQKDQAANEGPEQSSEVNELNQ<br>IPSHKALTLTHDNVVTVSPYALTHVAGPYNHWV |
| hTET2CD | MDFPSCRCVEQIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGKEGKSSQ<br>GCPIAKWVVRRSSSEEKLLCLVRERAGHTCEAAVIVILILVWEGIPLSLADKLYSELTETLRKY<br>GTLTNRRCALNEERTCACQGLDPETCGASFSFGCSWSMYYNGCKFARSKIPRKFKLLGDDPK<br>EEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIEYEHRAPECRLGLKEGRPFSGVTACLDFCA<br>HAHRDLHNMQNGSTLVCTLTREDNREFGGKPEDEQLHVLPLYKVSDVDEFGSVEAQEEKKR<br>SGAIQVLSSFRRKVRMLAEPVKTCRQRKLEAKKAAAEKLSSLENSSNKNEKEKSAPSRTKQT<br>ENASQAKQLAELLRLSGPVMQQSQQPQPLQKQPPQPQQQQRPQQQQPHHPQTESVNSYSAS<br>GSTNPYMRRPNPVSPYPNSSHTSDIYGSTSPMNFYSTSSQAAGSYLNSSNPMNPYPGLLNQN<br>TQYPSYQCNGNLSVDNCSPYLGSYSPQSQPMDLYRYPSQDPLSKLSLPPIHTLYQPRFGNSQSF<br>TSKYLGYGNQNMQGDGFSSCTIRPNVHHVGKLPPYPTHEMDGHFMGATSRLPPNLSNPNM<br>DYKNGEHHSPSHIIHNYSAAPGMFNSSLHALHLQNKENDMLSHTANGLSKMLPALNHDRTA<br>CVQGGLHKLSDANGQEKQPLALVQGVASGAEDNDEVWSDSEQSFLDPDIGGVAVAPTHGSIL<br>IECAKRELHATTPLKNPNRNHPTRISLVFYQHKSMNEPKHGLALWEAKMAEKAREKEEECE<br>KYGPDYVPQKSHGKKVKREPAEPHETSEPTYLRFIKSLAERTMSVTTDSTVTTSPYAFTRVTG<br>PYNRYI |
| hTET3CD | MEFPTCDCVEQIVEKDEGPYYTHLGSGPTVASIRELMEERYGEKGKAIRIEKVIYTGKEGKSSR |

| | |
|---|---|
| | GCPIAKWVIRRHTLEEKLLCLVRHRAGHHCQNAVIVILILAWEGIPRSLGDTLYQELTDTLRK<br>YGNPTSRRCGLNDDRTCACQGKDPNTCGASFSFGCSWSMYFNGCKYARSKTPRKFRLAGDN<br>PKEEEVLRKSFQDLATEVAPLYKRLAPQAYQNQVTNEEIAIDCRLGLKEGRPFAGVTACMDF<br>CAHAHKDQHNLYNGCTVVCTLTKEDNRCVGKIPEDEQLHVLPLYKMANTDEFGSEENQNA<br>KVGSGAIQVLTAFPREVRRLPEPAKSCRQRQLEARKAAAEKKKIQKEKLSTPEKIKQEALELA<br>GITSDPGLSLKGGLSQQGLKPSLKVEPQNHFSSFKYSGNAVVESYSVLGNCRPSDPYSMNSVYS<br>YHSYYAQPSLTSVNGFHSKYALPSFSYYGFPSSNPVFPSQFLGPGAWGHSGSSGSFEKKPDLH<br>ALHNSLSPAYGGAEFAELPSQAVPTDAHHPTPHHQQPAYPGPKEYLLPKAPLLHSVSRDPSP<br>FAQSSNCYNRSIKQEPVDPLTQAEPVPRDAGKMGKTPLSEVSQNGGPSHLWGQYSGGPSMSP<br>KRTNGVGGSWGVFSSGESPAIVPDKLSSFGASCLAPSHFTDGQWGLFPGEGQQAASHSGGRL<br>RGKPWSPCKFGNSTSALAGPSLTEKPWALGAGDFNSALKGSPGFQDKLWNPMKGEEGRIPA<br>AGASQLDRAWQSFGLPLGSSEKLFGALKSEEKLWDPFSLEEGPAEEPPSKGAVKEEKGGGGA<br>EEEEEELWSDSEHNFLDENIGGVAVAPAHGSILIECARRELHATTPLKKPNRCHPTRISLVFY<br>QHKNLNQPNHGLALWEAKMKQLAERARARQEEAARLGLGQQEAKLYGKKRKWGGTVVAE<br>PQQKEKKGVVPTRQALAVPTDSAVTVSSYAYTKVTGPYSRWI |
| hMBD1 | MAEDWLDCPALGPGWKRREVFRKSGATCGRSDTYYQSPTGDRIRSKVELTRYLGPACDLTL<br>FDFKQGILCYPAPKAHPVAVASKKRKKPSRPAKTRKRQVGPQSGEVRKEAPRDETKADTDT<br>APASFPAPGCCENCGISFSGDGTQRQRLKTLCKDCRAQRIAFNREQRMFKRVGCGECAACQV<br>TEDCGACSTCLLQLPHDVASGLFCKCERRRCLRIVERSRGCGVCRGCQTQEDCGHCPICLRPP<br>RPGLRRQWKCVQRRCLRGKHARRKGGCDSKMAARRRPGAQPLPPPPPSQSPEPTEPHPRAL<br>APSPPAEFIYYCVDEDELQPYTNRRQNRKCGACAACLRRMDCGRCDFCCDKPKFGGSNQKR<br>QKCRWRQCLQFAMKRLLPSVWSESEDGAGSPPPYRRRKRPSSARRHHLGPTLKPTLATRTA<br>QPDHTQAPTKQEAGGGFVLPPPGTDLVFLREGASSPVQVPGPVAASTEALLQEAQCSGLSWV<br>VALPQVKQEKADTQDEWTPGTAVLTSPVLVPGCPSKAVDPGLPSVKQEPPDPEEDKEENKD<br>DSASKLAPEEEAGGAGTPVITEIFSLGGTRFRDTAVWLPRSKDLKKPGARKQ |
| hMBD1-<br>ΔMBD | CYPAPKAHPVAVASKKRKKPSRPAKTRKRQVGPQSGEVRKEAPRDETKADTDTAPASFPAP<br>GCCENCGISFSGDGTQRQRLKTLCKDCRAQRIAFNREQRMFKRVGCGECAACQVTEDCGACS<br>TCLLQLPHDVASGLFCKCERRRCLRIVERSRGCGVCRGCQTQEDCGHCPICLRPPRPGLRRQW<br>KCVQRRCLRGKHARRKGGCDSKMAARRRPGAQPLPPPPPSQSPEPTEPHPRALAPSPPAEFI<br>YYCVDEDELQPYTNRRQNRKCGACAACLRRMDCGRCDFCCDKPKFGGSNQKRQKCRWRQC<br>LQFAMKRLLPSVWSESEDGAGSPPPYRRRKRPSSARRHHLGPTLKPTLATRTAQPDHTQAP<br>TKQEAGGGFVLPPPGTDLVFLREGASSPVQVPGPVAASTEALLQEAQCSGLSWVVALPQVKQ<br>EKADTQDEWTPGTAVLTSPVLVPGCPSKAVDPGLPSVKQEPPDPEEDKEENKDDSASKLAPE<br>EEAGGAGTPVITEIFSLGGTRFRDTAVWLPRSKDLKKPGARKQ |
| hMBD1<br>isoform7 | MAEDWLDCPALGPGWKRREVFRKSGATCGRSDTYYQSPTGDRIRSKVELTRYLGPACDLTL<br>FDFKQGILCYPAPKAHPVAVASKKRKKPSRPAKTRKRQVGPQSGEVRKEAPRDETKADTDT<br>APASFPAPGCCENCGISFSGDGTQRQRLKTLCKDCRAQRIAFNREQRMFKRVGCGECAACQV |

| | |
|---|---|
| | TEDCGACSTCLLQLPHDVASGLFCKCERRRCLRIVERSRGCGVCRGCQTQEDCGHCPICLRPP<br>RPGLRRQWKCVQRRCLRGKHARRKGGCDSKMAARRRPGAQPLPPPPPSQSPEPTEPHPRAL<br>APSPPAEFIYYCVDEDELKRLLPSVWSESEDGAGSPPPYRRRKRPSSARRHHLGPTLKPTLAT<br>RTAQPDHTQAPTKQEAGGGFVLPPPGTDLVFLREGASSPVQVPGPVAASTEALLQEAQCSGL<br>SWVVALPQVKQEKADTQDEWTPGTAVLTSPVLVPGCPSKAVDPGLPSVKQEPPDPEEDKEE<br>NKDDSASKLAPEEEAGGAGTPVITEIFSLGGTRFRDTAVWLPRSKDLKKPGARKQ |
| hMBD1-<br>ΔTRD | MAEDWLDCPALGPGWKRREVFRKSGATCGRSDTYYQSPTGDRIRSKVELTRYLGPACDLTL<br>FDFKQGILCYPAPKAHPVAVASKKRKKPSRPAKTRKRQVGPQSGEVRKEAPRDETKADTDT<br>APASFPAPGCCENCGISFSGDGTQRQRLKTLCKDCRAQRIAFNREQRMFKRVGCGECAACQV<br>TEDCGACSTCLLQLPHDVASGLFCKCERRRCLRIVERSRGCGVCRGCQTQEDCGHCPICLRPP<br>RPGLRRQWKCVQRRCLRGKHARRKGGCDSKMAARRRPGAQPLPPPPPSQSPEPTEPHPRAL<br>APSPPAEFIYYCVDEDELQPYTNRRQNRKCGACAACLRRMDCGRCDFCCDKPKFGGSNQKR<br>QKCRWRQCLQFAMKRLLPSVWSESEDGAGSPPPYRRRKRPSSARRHHLGPTLKPTLATRTA<br>QPDHTQAPTKQEAGGGFVLPPPGTDLVFLREGASSPVQVPGPVAASTEALLQEAQCSGLSWV<br>VALPQVKQEKADTQDEWTPGTAVLTSPVLVPGCPSKRSKDLKKPGARKQ |
| hMBD2a | MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRG<br>RWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAP<br>RREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGK<br>KFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIR<br>QTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKG<br>LQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRK<br>QEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA |
| hMBD3 | MERKRWECPALPQGWEREEVPRRSGLSAGHRDVFYYSPSGKKFRSKPQLARYLGGSMDLST<br>FDFRTGKMLMSKMNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNK<br>VKSDPQKAVDQPRQLFWEKKLSGLNAFDIAEELVKTMDLPKGLQGVGPGCTDETLLSAIASA<br>LHTSTMPITGQLSAAVEKNPGVWLNTTQPLCKAFMVTDEDIRKQEELVQQVRKRLEEALMA<br>DMLAHVEELARDGEAPLDKACAEDDDEEDEEEEEEEPDPDPEMEHV |
| hMBD4 | MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSEC<br>NPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSL<br>ANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCK<br>KDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKGIPIKKTKK<br>GCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETLSVTSEENSLVK<br>KKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDI<br>LKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKW<br>TPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVS<br>ELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHPEDHK<br>LNKYHDWLWENHEKLSLS |

| | |
|---|---|
| hMeCP2 | MVAGMLGLREEKSEDQDLQGLKDKPLKFKKVKKDKKEEKEGKHEPVQPSAHHSAEPAEAG<br>KAETSEGSGSAPAVPEASASPKQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDV<br>YLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRGSPSRREQKPPKKPKSPKAPGTG<br>RGRGRPKGSGTTRPKAATSEGVQVKRVLEKSPGKLLVKMPFQTSPGGKAEGGGATTSTQVM<br>VIKRPGRKRKAEADPQAIPKKRGRKPGSVVAAAAAEAKKKAVKESSIRSVQETVLPIKKRKTR<br>ETVSIEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRKSKESSPKGRSSSASSPPKKEHHHHHHH<br>SESPKAPVPLLPPLPPPPPEPESSEDPTSPPEPQDLSSSVCKEEKMPRGGSLESDGCPKEPAKT<br>QPAVATAATAAEKYKHRGEGERKDIVSSSMPRPNREEPVDSRTPVTERVS |

## A.3. Plasmid maps

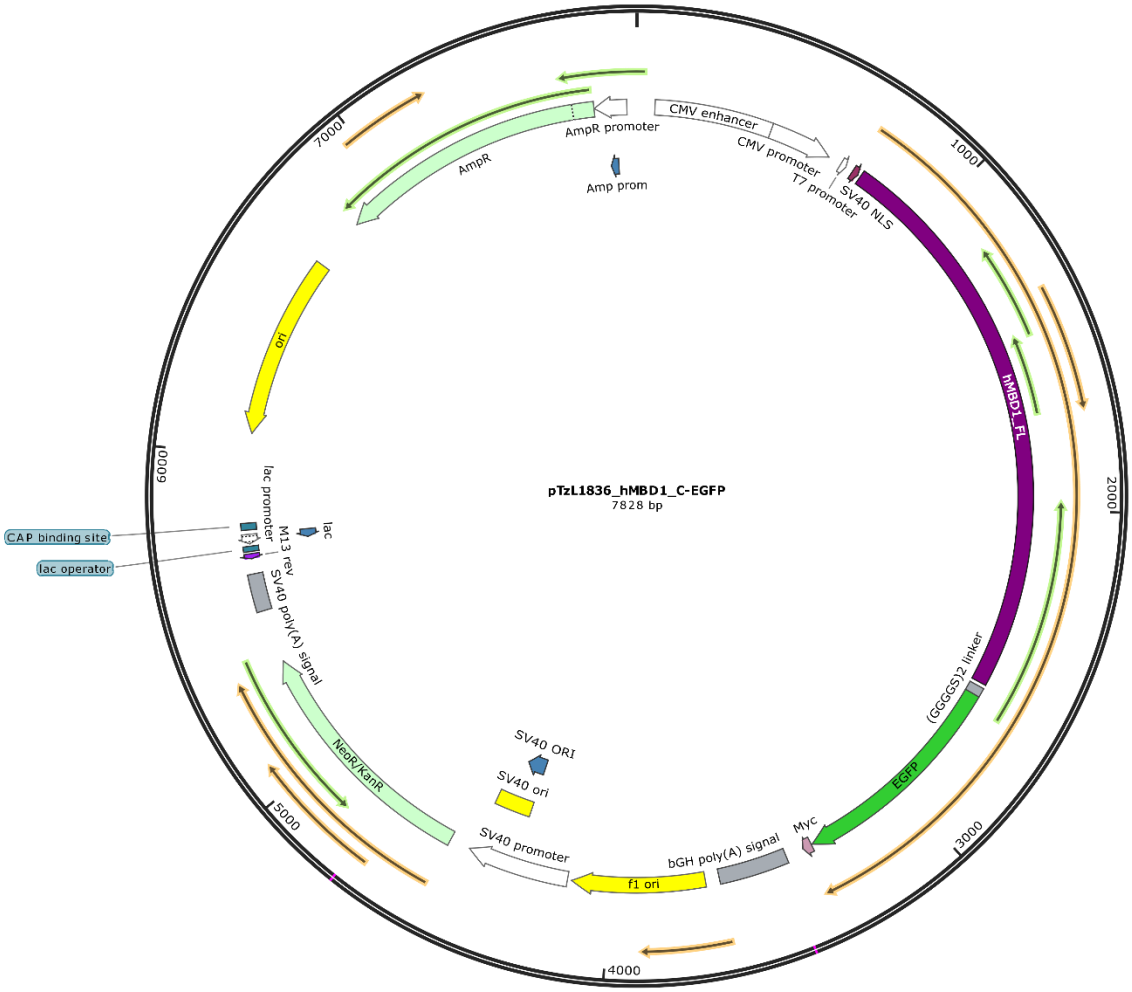


**Figure S 25. Plasmid map of C-terminal EGFP-tagged hMBD1 vector used in the study**. The same vector was used for all the other EGFP-tagged MBD constructs (the encoding MBD protein sequences are listed in Table S 5, mutagenesis is indicated in 0).

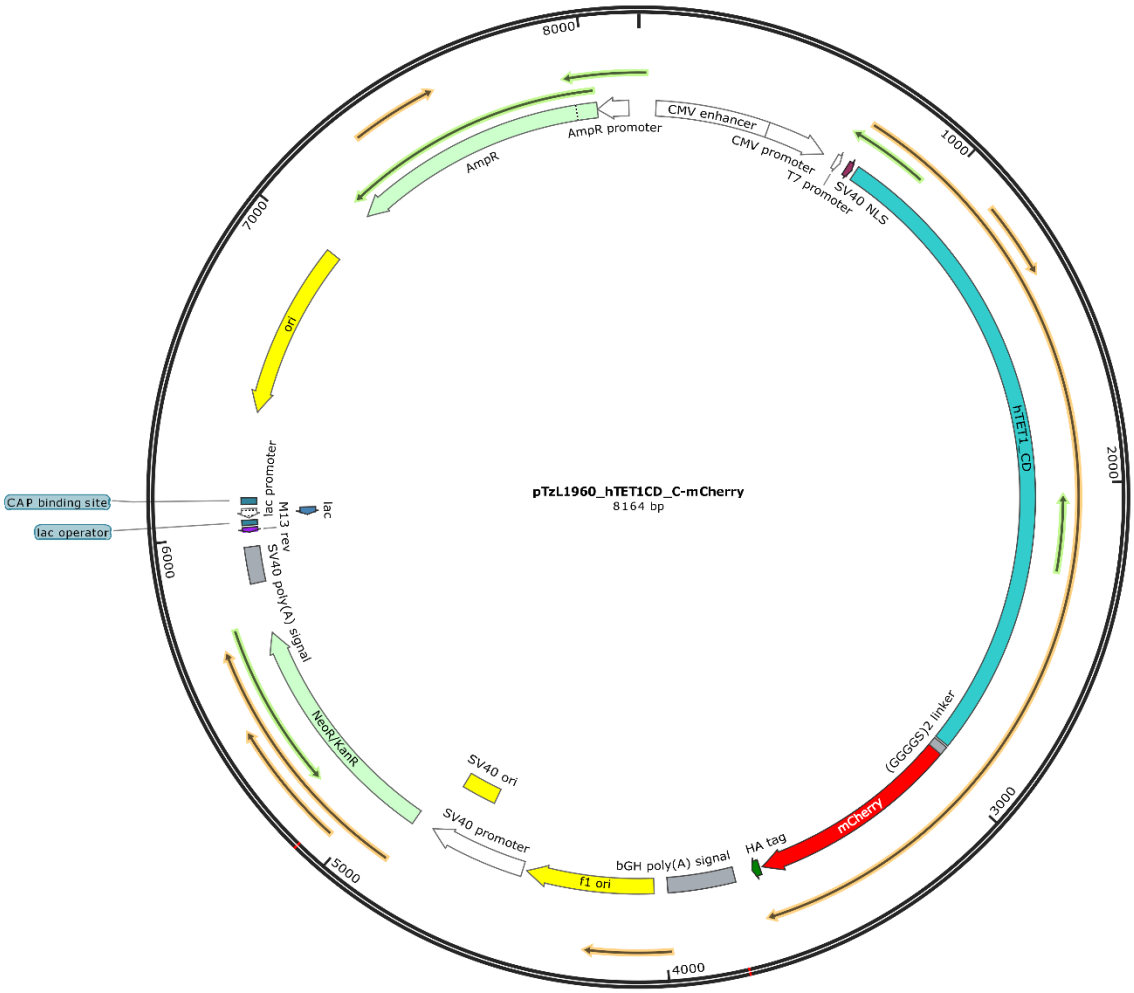


**Figure S 26. Plasmid map of C-terminal mCherry-tagged hTET1CD vector used in the study**. The encoding hTET1CD protein sequences is listed in Table S 5, mutagenesis is indicated in 0.

## A.4. Credits and copyright licenses

The content of **Chapter 6**, some sections in **Chapter 9**, and part of the supplementary figures have been published in (Lin et al., 2022) and licensed under CC BY-NC-ND 4.0, which allows copy and redistribute the material in any medium or format with proper attribution. To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en. The reprinted figures and individual modifications are indicated below:

[1] **Figure 11, Figure 20–26, Figure S 1,** and **Figure S 4–13** are reprinted from (Lin et al., 2022). The following figures are modified from the original publication:

> **Figure 20:** The panels have been rearranged, the domain structures of TET2CD and TET3CD were added to panel b).
>
> **Figure 21**: The panels have been rearranged.
>
> **Figure 22:** Panels f–i were removed from the original figure.
>
> **Figure 23:** Panels a–e were removed from the original figure.
>
> **Figure 24:** Panels d–f were removed from the original figure.
>
> **Figure 25:** Panels a–c were removed from the original figure.

The author is grateful for the permission to use the following copyright materials:

[2] **Figure 2a** is recreated based on the work of (S. Huang, 2012) with permission. Copyright © 2012 WILEY Periodicals, Inc.

[3] **Figure 2b** is reprinted with permission from (Goldberg et al., 2007). Copyright © 2007 Elsevier Inc.

[4] **Figure 3** is reprinted with permission from (Smallwood & Kelsey, 2012). Copyright © 2011 Elsevier Ltd.

[5] **Figure 4b** is created based on the work of (Kohli & Zhang, 2013) with permission. Copyright © 2013, Nature Publishing Group, a division of Macmillan Publishers Limited.

[6] **Figure 5a** is modified from the work of (Denis et al., 2011) with permission. Copyright © 2011 European Molecular Biology Organization.

[7] **Figure 6b** and **Figure 6c** are reprinted with permission from (L. Hu et al., 2013). Copyright

© 2013 Elsevier Inc.

[8] **Figure 7b**, **Figure 7d**, and **Figure 27b** are reprinted with permission from (Ohki et al., 2001). Copyright © 2001 Cell Press.

[9] **Figure 7c** is reprinted with permission from (Ohki et al., 1999). Copyright © 1999 European Molecular Biology Organization.

[10] **Figure 8a** is created based on the work of (Ludwig et al., 2016). Copyright © 2016 Ludwig, Zhang and Cardoso. The original figure is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/.

[11] **Figure 8b** and **Figure 8c** are reprinted with permission from (Spruijt et al., 2013). Copyright © 2013 Elsevier Inc.

[12] **Figure 10** is reprinted with permission from (Zhu et al., 2016). Copyright © 2016, Nature Publishing Group, a division of Macmillan Publishers Limited.

[13] **Figure 12** is created based on the work of (Du et al., 2015) with permission. The original figure is licensed under CC BY-NC-ND 4.0. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

[14] **Figure 13** is reprinted with permission from (H. Wu & Zhang, 2014). Copyright © 2014 Elsevier Inc.

[15] **Figure 14a**, **Figure 14c**, and **Figure 15c** are reprinted from (Kneuttinger, 2022). The original figure is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/.

[16] **Figure 14b** is reprinted with permission from (Ankenbruck et al., 2018). Copyright © 2018 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

[17] **Figure 16** is modified form the work of (Manoilov et al., 2021) with permission. Copyright © 2021, Springer Nature America, Inc.

[18] **Figure 17a** is created based on the work of (Dumas et al., 2015). The original figure is licensed under CC BY 3.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/3.0/.

[19] **Figure 18a** is modified with permission from the work of (Knight et al., 2001). Copyright © 2001, Macmillan Magazines Ltd.

[20] **Figure 18b** is adapted with permission from (Lang & Chin, 2014). Copyright © 2014,

American Chemical Society.

[21] **Figure 19a** is adapted with permission from (Ambrogelly et al., 2007). Copyright © 2006, Nature Publishing Group.

# B. Bibliography

Agarwal, N., Hardt, T., Brero, A., Nowak, D., Rothbauer, U., Becker, A., Leonhardt, H., & Cardoso, M. C. (2007). MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. *Nucleic Acids Research*, *35*(16), 5402–5408. https://doi.org/10.1093/nar/gkm599

Ambrogelly, A., Palioura, S., & Söll, D. (2007). Natural expansion of the genetic code. *Nature Chemical Biology*, *3*(1), 29–35. https://doi.org/10.1038/nchembio847

Amir, R. E., van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl- CpG-binding protein 2. *Nature Genetics*, *23*(2), 185–188. https://doi.org/10.1038/13810

Ankenbruck, N., Courtney, T., Naro, Y., & Deiters, A. (2018). Optochemical Control of Biological Processes in Cells and Animals. *Angewandte Chemie International Edition*, *57*(11), 2768–2798. https://doi.org/10.1002/anie.201700171

Arbely, E., Torres-Kolbus, J., Deiters, A., & Chin, J. W. (2012). Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *Journal of the American Chemical Society*, *134*(29), 11912–11915. https://doi.org/10.1021/ja3046958

Baets, J., Duan, X., Wu, Y., Smith, G., Seeley, W. W., Mademan, I., McGrath, N. M., Beadell, N. C., Khoury, J., Botuyan, M. V., Mer, G., Worrell, G. A., Hojo, K., DeLeon, J., Laura, M., Liu, Y. T., Senderek, J., Weis, J., van den Bergh, P., … Klein, C. J. (2015). Defects of mutant DNMT1 are linked to a spectrum of neurological disorders. *Brain*, *138*(4), 845–861. https://doi.org/10.1093/brain/awv010

Baker, A. S., & Deiters, A. (2014). Optical Control of Protein Function through Unnatural Amino Acid Mutagenesis and Other Optogenetic Approaches. *ACS Chemical Biology*, *9*(7), 1398–1407. https://doi.org/10.1021/cb500176x

Banghart, M., Borges, K., Isacoff, E., Trauner, D., & Kramer, R. H. (2004). Light-activated ion channels for remote control of neuronal firing. *Nature Neuroscience*, *7*(12), 1381–1386. https://doi.org/10.1038/nn1356

Bardhan, A., & Deiters, A. (2019). Development of photolabile protecting groups and their application to the optochemical control of cell signaling. *Current Opinion in Structural Biology*, *57*, 164–175. https://doi.org/10.1016/j.sbi.2019.03.028

Batsché, E., Yi, J., Mauger, O., Kornobis, E., Hopkins, B., Hanmer-Lloyd, C., & Muchardt, C. (2021). CD44 alternative splicing senses intragenic DNA methylation in tumors via direct and indirect mechanisms. *Nucleic Acids Research*, *49*(11), 6213–6237. https://doi.org/10.1093/nar/gkab437

Baubec, T., Ivánek, R., Lienert, F., & Schübeler, D. (2013). Methylation-dependent and -

independent genomic targeting principles of the mbd protein family. *Cell*, *153*(2), 480–492. https://doi.org/10.1016/j.cell.2013.03.011

Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A. T., Matsumoto, Y., Golemis, E. A., Genuardi, M., & Neri, G. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. In *Medical Sciences* (Vol. 96). www.pnas.org.

Billard, L. M., Magdinier, F., Lenoir, G. M., Frappart, L., & Dante, R. (2002). MeCP2 and MBD2 expression during normal and pathological growth of the human mammary gland. *Oncogene*, *21*(17), 2704–2712. https://doi.org/10.1038/sj.onc.1205357

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes and Development*, *16*(1), 6–21. https://doi.org/10.1101/gad.947102

Blattler, A., & Farnham, P. J. (2013). Cross-talk between site-specific transcription factors and DNA methylation states. In *Journal of Biological Chemistry* (Vol. 288, Issue 48, pp. 34287–34294). https://doi.org/10.1074/jbc.R113.512517

Bose, M., Groff, D., Xie, J., Brustad, E., & Schultz, P. G. (2006). The incorporation of a photoisomerizable amino acid into proteins in E. coli. *Journal of the American Chemical Society*, *128*(2), 388–389. https://doi.org/10.1021/ja055467u

Boulègue, C., Löweneck, M., Renner, C., & Moroder, L. (2007). Redox potential of azobenzene as an amino acid residue in peptides. *ChemBioChem*, *8*(6), 591–594. https://doi.org/10.1002/cbic.200600495

Brocher Jan. (2022). *BioVoxxel Toolbox*.

Buganim, Y., Faddah, D. A., & Jaenisch, R. (2013). Mechanisms and models of somatic cell reprogramming. In *Nature Reviews Genetics* (Vol. 14, Issue 6, pp. 427–439). https://doi.org/10.1038/nrg3473

Cartron, P. F., Nadaradjane, A., LePape, F., Lalier, L., Gardie, B., & Vallette, F. M. (2013). Identification of TET1 Partners That Control Its DNA-Demethylating Function. *Genes and Cancer*, *4*(5–6), 235–241. https://doi.org/10.1177/1947601913489020

Chen, P. R., Groff, D., Guo, J., Ou, W., Cellitti, S., Geierstanger, B. H., & Schultz, P. G. (2009). A facile system for encoding unnatural amino acids in mammalian cells. *Angewandte Chemie - International Edition*, *48*(22), 4052–4055. https://doi.org/10.1002/anie.200900683

Chou, C., & Deiters, A. (2011). Light-activated gene editing with a photocaged zinc-finger nuclease. *Angewandte Chemie - International Edition*, *50*(30), 6839–6842. https://doi.org/10.1002/anie.201101157

Choudhury, S. R., Cui, Y., Narayanan, A., Gilley, D. P., Huda, N., Lo, C.-L., Zhou, F. C., Yernool, D., & Irudayaraj, J. (n.d.). *Oncotarget 50380 www.impactjournals.com/oncotarget Optogenetic regulation of site-specific subtelomeric DNA methylation*. www.impactjournals.com/oncotarget/

Clouaire, T., de las Heras, J. I., Merusi, C., & Stancheva, I. (2010). Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA. *Nucleic Acids Research*, *38*(14), 4620–4634. https://doi.org/10.1093/nar/gkq228

Crick, F. (1970). Central Dogma of Molecular Biology. *Nature 1970 227:5258*, *227*(5258), 561–563. https://doi.org/10.1038/227561a0

Crick, F. H. C. (1968). The origin of the genetic code. *Journal of Molecular Biology*, *38*(3), 367–379. https://doi.org/10.1016/0022-2836(68)90392-6

Croitoru, A., Park, S. J., Kumar, A., Lee, J., Im, W., Mackerell, A. D., & Aleksandrov, A. (2021). Additive CHARMM36 Force Field for Nonstandard Amino Acids. *Journal of Chemical Theory and Computation*, *17*(6), 3554–3570. https://doi.org/10.1021/ACS.JCTC.1C00254/SUPPL_FILE/CT1C00254_SI_001.PDF

Cross, S. H., Meehan, R. R., Nan, X., & Bird, A. (1997). *A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins*. http://www.ncbi.nlm.nih.gov/XREFdb,

Dagliyan, O., Tarnawski, M., Chu, P.-H., Shirvanyants, D., Schlichting, I., Dokholyan, N. v, & Hahn, K. M. (2016). Engineering extrinsic disorder to control protein activity in living cells. *Science*, *354*(6318), 1441–1444. https://www.science.org

Daniel, J. M., & Reynolds, A. B. (1999). The Catenin p120 Interacts with Kaiso, a Novel BTB/POZ Domain Zinc Finger Transcription Factor. *Molecular and Cellular Biology*, *19*(5), 3614–3623. https://doi.org/10.1128/MCB.19.5.3614

Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes and Development*, *25*(10), 1010–1022. https://doi.org/10.1101/gad.2037511

Deiters, A., Groff, D., Ryu, Y., Xie, J., & Schultz, P. G. (2006). A genetically encoded photocaged tyrosine. *Angewandte Chemie - International Edition*, *45*(17), 2728–2731. https://doi.org/10.1002/anie.200600264

Denis, H., Ndlovu, 'Matladi N, & Fuks, F. (2011). Regulation of mammalian DNA methyltransferases: a route to new mechanisms. *EMBO Reports*, *12*(7), 647–656. https://doi.org/10.1038/embor.2011.110

Denslow, S. A., & Wade, P. A. (2007). The human Mi-2/NuRD complex and gene regulation. *Oncogene*, *26*(37), 5433–5438. https://doi.org/10.1038/sj.onc.1210611

Du, Q., Luu, P. L., Stirzaker, C., & Clark, S. J. (2015). Methyl-CpG-binding domain proteins: Readers of the epigenome. *Epigenomics*, *7*(6), 1051–1073. https://doi.org/10.2217/epi.15.39

Du, Q., Wang, Z., & Schramm, V. L. (2016). Human DNMT1 transition state structure. *Proceedings of the National Academy of Sciences*, *113*(11), 2916–2921. https://doi.org/10.1073/pnas.1522491113

Dumas, A., Lercher, L., Spicer, C. D., & Davis, B. G. (2015). Designing logical codon reassignment – Expanding the chemistry in biology. *Chemical Science*, *6*(1), 50–69.

https://doi.org/10.1039/C4SC01534G

Edwards, J. R., O'Donnell, A. H., Rollins, R. A., Peckham, H. E., Lee, C., Milekic, M. H., Chanrion, B., Fu, Y., Su, T., Hibshoosh, H., Gingrich, J. A., Haghighi, F., Nutter, R., & Bestor, T. H. (2010). Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. *Genome Research*, *20*(7), 972–980. https://doi.org/10.1101/gr.101535.109

Edwards, J. R., Yarychkivska, O., Boulard, M., & Bestor, T. H. (2017). DNA methylation and DNA methyltransferases. In *Epigenetics and Chromatin* (Vol. 10, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13072-017-0130-8

Edwards, W. F., Young, D. D., & Deiters, A. (2009). Light-activated Cre recombinase as a tool for the spatial and temporal control of gene function in mammalian cells. *ACS Chemical Biology*, *4*(6), 441–445. https://doi.org/10.1021/cb900041s

Fehrentz, T., Schönberger, M., & Trauner, D. (2011). Optochemical Genetics. *Angewandte Chemie International Edition*, *50*(51), 12156–12182. https://doi.org/10.1002/anie.201103236

Filion, G. J. P., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., & Defossez, P.-A. (2006). A Family of Human Zinc Finger Proteins That Bind Methylated DNA and Repress Transcription. *Molecular and Cellular Biology*, *26*(1), 169–181. https://doi.org/10.1128/mcb.26.1.169-181.2006

Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M. C., Antes, I., & Leonhardt, H. (2011). Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE*, *6*(6). https://doi.org/10.1371/journal.pone.0021306

Fujiki, K., Shinoda, A., Kano, F., Sato, R., Shirahige, K., & Murata, M. (2013). PPARγ-induced PARylation promotes local DNA demethylation by production of 5-hydroxymethylcytosine. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms3262

Fujita, N., Takebayashi, S., Okumura, K., Kudo, S., Chiba, T., Saya, H., & Nakao, M. (1999a). Methylation-Mediated Transcriptional Silencing in Euchromatin by Methyl-CpG Binding Protein MBD1 Isoforms. *Molecular and Cellular Biology*, *19*(9), 6415–6426. https://doi.org/10.1128/MCB.19.9.6415

Fujita, N., Takebayashi, S., Okumura, K., Kudo, S., Chiba, T., Saya, H., & Nakao, M. (1999b). Methylation-Mediated Transcriptional Silencing in Euchromatin by Methyl-CpG Binding Protein MBD1 Isoforms. *Molecular and Cellular Biology*, *19*(9), 6415–6426. https://doi.org/10.1128/MCB.19.9.6415/ASSET/7E3C7BAC-1025-4ACF-A14B-8887B6E25470/ASSETS/GRAPHIC/MB0990212007.JPEG

Fujita, N., Watanabe, S., Ichimura, T., Ohkuma, Y., Chiba, T., Saya, H., & Nakao, M. (2003). MCAF Mediates MBD1-Dependent Transcriptional Repression. *Molecular and Cellular Biology*, *23*(8), 2834–2843. https://doi.org/10.1128/MCB.23.8.2834-2843.2003

Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., Chiba, T., & Nakao, M.

(2003). Methyl-CpG Binding Domain 1 (MBD1) Interacts with the Suv39h1-HP1 Heterochromatic Complex for DNA Methylation-based Transcriptional Repression. *Journal of Biological Chemistry*, *278*(26), 24132–24138. https://doi.org/10.1074/jbc.M302283200

Gautier, A., Deiters, A., & Chin, J. W. (2011). Light-activated kinases enable temporal dissection of signaling networks in living cells. *Journal of the American Chemical Society*, *133*(7), 2124–2127. https://doi.org/10.1021/ja1109979

Gautier, A., Nguyen, D. P., Lusic, H., An, W., Deiters, A., & Chin, J. W. (2010). Genetically encoded photocontrol of protein localization in mammalian cells. *Journal of the American Chemical Society*, *132*(12), 4086–4088. https://doi.org/10.1021/ja910688s

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., & Ferrin, T. E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science*, *27*(1), 14–25. https://doi.org/10.1002/pro.3235

Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: A Landscape Takes Shape. *Cell*, *128*(4), 635–638. https://doi.org/10.1016/j.cell.2007.02.006

Groff, D., Wang, F., Jockusch, S., Turro, N. J., & Schultz, P. G. (2010). A new strategy to photoactivate green fluorescent protein. *Angewandte Chemie - International Edition*, *49*(42), 7677–7679. https://doi.org/10.1002/anie.201003797

Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., Xie, Z. G., Shi, L., He, X., Jin, S. G., Iqbal, K., Shi, Y. G., Deng, Z., Szabó, P. E., Pfeifer, G. P., Li, J., & Xu, G. L. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, *477*(7366), 606–612. https://doi.org/10.1038/nature10443

Hackett, J. A., Sengupta, R., Zylicz, J. J., Murakami, K., Lee, C., Down, T. A., & Surani, M. A. (2013). Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine. *Science*, *339*(6118), 448–452. https://doi.org/10.1126/science.1229277

Hagberg, B., Aicardi, J., Dias, K., & Ramos, O. (1983). A Progressive Syndrome of Autism, Dementia, Ataxia, and Loss of Purpos&l Hand Use in Girls: Rett's Syndrome: Report of 35 Cases. *Annals of Neurology*, *14*(4), 471–479.

Hahne F, Gopalakrishnan N, Khodabakhshi AH, Wong C, & Lee K. (2022). *flowStats: Statistical methods for the analysis of flow cytometry data. R package version 4.8.2*. https://www.bioconductor.org/packages/release/bioc/html/flowStats.html

Hainer, S. J., McCannell, K. N., Yu, J., Ee, L.-S., Zhu, L. J., Rando, O. J., & Fazzio, T. G. (2016). DNA methylation directs genomic localization of Mbd2 and Mbd3 in embryonic stem cells. *ELife*, *5*. https://doi.org/10.7554/eLife.21964

Halavaty, A. S., & Moffat, K. (2007). N- and C-terminal flanking regions modulate light-induced signal transduction in the LOV2 domain of the blue light sensor phototropin 1 from Avena sativa. *Biochemistry*, *46*(49), 14001–14009. https://doi.org/10.1021/bi701543e

Hameed, U. F. S., Lim, J., Zhang, Q., Wasik, M. A., Yang, D., & Swaminathan, K. (2014). Transcriptional repressor domain of MBD1 is intrinsically disordered and interacts with its binding partners in a selective manner. *Scientific Reports*, *4*, 1–11. https://doi.org/10.1038/srep04896

Hansen, R. S., Wijmenga, C., Luo, P., Stanek, A. M., Canfield, T. K., Weemaes, C. M. R., & Gartler, S. M. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proceedings of the National Academy of Sciences*, *96*(25), 14412–14417. https://doi.org/10.1073/pnas.96.25.14412

Hao, B., Gong, W., Ferguson, T. K., James, C. M., Krzycki, J. A., & Chan, M. K. (2002). A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase. *Science*, *296*(5572), 1462–1466. https://doi.org/10.1126/science.1069556

Hashimoto, H., Liu, Y., Upadhyay, A. K., Chang, Y., Howerton, S. B., Vertino, P. M., Zhang, X., & Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Research*, *40*(11), 4841–4849. https://doi.org/10.1093/nar/gks155

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C., & Xu, G.-L. (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science*, *333*(6047), 1303–1307. https://doi.org/10.1126/science.1210944

Hemphill, J., Borchardt, E. K., Brown, K., Asokan, A., & Deiters, A. (2015). Optical control of CRISPR/Cas9 gene editing. *Journal of the American Chemical Society*, *137*(17), 5642–5645. https://doi.org/10.1021/ja512664v

Hendrich, B., & Bird, A. (1998). Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Molecular and Cellular Biology*, *18*(11), 6538–6547. https://doi.org/10.1128/mcb.18.11.6538

Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A., & Bird, A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes and Development*, *15*(6), 710–723. https://doi.org/10.1101/gad.194101

Hendrich, B., Hardeland[23], U., Ng, H.-H., Jiricny[3], J., & Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. In *NATURE* (Vol. 401). www.nature.com

Holliday, R., & Pugh, J. E. (1975). DNA Modification Mechanisms and Gene Activity during Development. *Science*, *187*, 226–232. https://www.science.org

Hon, G. C., Song, C. X., Du, T., Jin, F., Selvaraj, S., Lee, A. Y., Yen, C. A., Ye, Z., Mao, S. Q., Wang, B. A., Kuan, S., Edsall, L. E., Zhao, B. S., Xu, G. L., He, C., & Ren, B. (2014). 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Molecular Cell*, *56*(2), 286–297.

https://doi.org/10.1016/j.molcel.2014.08.026

Hoppmann, C., Lacey, V. K., Louie, G. v., Wei, J., Noel, J. P., & Wang, L. (2014). Genetically encoding photoswitchable click amino acids in Escherichia coli and mammalian cells. *Angewandte Chemie - International Edition*, *53*(15), 3932–3936. https://doi.org/10.1002/anie.201400001

Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y. G., Zhu, J., Wang, P., & Xu, Y. (2013). Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell*, *155*(7), 1545–1555. https://doi.org/10.1016/j.cell.2013.11.020

Hu, L., Lu, J., Cheng, J., Rao, Q., Li, Z., Hou, H., Lou, Z., Zhang, L., Li, W., Gong, W., Liu, M., Sun, C., Yin, X., Li, J., Tan, X., Wang, P., Wang, Y., Fang, D., Cui, Q., … Xu, Y. (2015a). Structural insight into substrate preference for TET-mediated oxidation. *Nature*, *527*(7576), 118–122. https://doi.org/10.1038/nature15713

Hu, L., Lu, J., Cheng, J., Rao, Q., Li, Z., Hou, H., Lou, Z., Zhang, L., Li, W., Gong, W., Liu, M., Sun, C., Yin, X., Li, J., Tan, X., Wang, P., Wang, Y., Fang, D., Cui, Q., … Xu, Y. (2015b). Structural insight into substrate preference for TET-mediated oxidation. *Nature*, *527*(7576), 118–122. https://doi.org/10.1038/nature15713

Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H. N., Shin, J., Cox, E., Rho, H. S., Woodard, C., Xia, S., Liu, S., Lyu, H., Ming, G. L., Wade, H., Song, H., Qian, J., & Zhu, H. (2013). DNA methylation presents distinct binding sites for human transcription factors. *ELife*, *2013*(2). https://doi.org/10.7554/eLife.00726

Huang, S. (2012). The molecular and mathematical basis of Waddington's epigenetic landscape: A framework for post-Darwinian biology? *BioEssays*, *34*(2), 149–157. https://doi.org/10.1002/bies.201100031

Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W. A., Kang, J., Zepeda-Martínez, J. A., Pape, U. J., Jacobsen, S. E., Peters, B., & Rao, A. (2014). Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(4), 1361–1366. https://doi.org/10.1073/pnas.1322921111

Ibba, M., & Söll, D. (2002). Genetic Code: Introducing Pyrrolysine. *Current Biology*, *12*(13), R464–R466. https://doi.org/10.1016/S0960-9822(02)00947-8

Ichimura, T., Watanabe, S., Sakamoto, Y., Aoto, T., Pujita, N., & Nakao, M. (2005). Transcriptional repression and heterochromatin formation by MBD1 and MCAF/AM family proteins. *Journal of Biological Chemistry*, *280*(14), 13928–13935. https://doi.org/10.1074/jbc.M413654200

Il'ichev, Y. v., & Wirz, J. (2000). Rearrangements of 2-Nitrobenzyl Compounds. 1. Potential Energy Surface of 2-Nitrotoluene and Its Isomers Explored with ab Initio and Density Functional Theory Methods. *The Journal of Physical Chemistry A*, *104*(33), 7856–7870.

https://doi.org/10.1021/jp000261v

Iqbal, K., Jin, S. G., Pfeifer, G. P., & Szabó, P. E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(9), 3642–3647. https://doi.org/10.1073/pnas.1014033108

Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J. B., Sabunciyan, S., & Feinberg, A. P. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics*, *41*(2), 178–186. https://doi.org/10.1038/ng.298

Ito, S., Dalessio, A. C., Taranova, O. v., Hong, K., Sowers, L. C., & Zhang, Y. (2010). Role of tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, *466*(7310), 1129–1133. https://doi.org/10.1038/nature09303

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., & Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, *333*(6047), 1300–1303. https://doi.org/10.1126/science.1210597

Iyer, L. M., Tahiliani, M., Rao, A., & Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*, *8*(11), 1698–1710. https://doi.org/10.4161/cc.8.11.8580

Jin, B., Li, Y., & Robertson, K. D. (2011). DNA methylation: Superior or subordinate in the epigenetic hierarchy? In *Genes and Cancer* (Vol. 2, Issue 6, pp. 607–617). https://doi.org/10.1177/1947601910393957

Jin, B., Tao, Q., Peng, J., Soo, H. M., Wu, W., Ying, J., Fields, C. R., Delmas, A. I., Liu, X., Qiu, J., & Robertson, K. D. (2008). DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. *Human Molecular Genetics*, *17*(5), 690–709. https://doi.org/10.1093/hmg/ddm341

Jin, S. G., Zhang, Z. M., Dunwell, T. L., Harter, M. R., Wu, X., Johnson, J., Li, Z., Liu, J., Szabó, P. E., Lu, Q., Xu, G. liang, Song, J., & Pfeifer, G. P. (2016). Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration. *Cell Reports*, *14*(3), 493–505. https://doi.org/10.1016/j.celrep.2015.12.044

Johansson, L., Gafvelin, G., & Arnér, E. S. J. (2005). Selenocysteine in proteins—properties and biotechnological use. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1726*(1), 1–13. https://doi.org/10.1016/j.bbagen.2005.05.010

Jones, P. L., Veenstra, G. J. C., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., & Wolffe, A. P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genetics*, *19*(2), 187–191. https://doi.org/10.1038/561

Jørgensen, H. F., Ben-porath, I., Bird, A. P., Jørgensen, H. F., Ben-porath, I., & Bird, A. P. (2004). Mbd1 Is Recruited to both Methylated and Nonmethylated CpGs via Distinct DNA Binding Domains Mbd1 Is Recruited to both Methylated and Nonmethylated CpGs via Distinct DNA Binding Domains. *Molecular and Cellular Biology*, *24*(8). https://doi.org/10.1128/MCB.24.8.3387

Jung, B. P., Jugloff, D. G. M., Zhang, G., Logan, R., Brown, S., & Eubanks, J. H. (2003). The expression of methyl CpG binding factor MeCP2 correlates with cellular differentiation in the developing rat brain and in cultured cells. *Journal of Neurobiology*, *55*(1), 86–96. https://doi.org/10.1002/neu.10201

Kang, J. Y., Kawaguchi, D., Coin, I., Xiang, Z., O'Leary, D. D. M., Slesinger, P. A., & Wang, L. (2013). In vivo expression of a light-activatable potassium channel using unnatural amino acids. *Neuron*, *80*(2), 358–370. https://doi.org/10.1016/j.neuron.2013.08.016

Kernohan, K. D., Vernimmen, D., Gloor, G. B., & Bérubé, N. G. (2014). Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping. *Nucleic Acids Research*, *42*(13), 8356–8368. https://doi.org/10.1093/nar/gku564

Klán, P., Šolomek, T., Bochet, C. G., Blanc, A., Givens, R., Rubina, M., Popik, V., Kostikov, A., & Wirz, J. (2013). Photoremovable Protecting Groups in Chemistry and Biology: Reaction Mechanisms and Efficacy. *Chemical Reviews*, *113*(1), 119–191. https://doi.org/10.1021/cr300177k

Kneuttinger, A. C. (2022). A guide to designing photocontrol in proteins: methods, strategies and applications. *Biological Chemistry*, *403*(5–6), 573–613. https://doi.org/10.1515/hsz-2021-0417

Knight, R. D., Freeland, S. J., & Landweber, L. F. (2001). Rewiring the keyboard: evolvability of the genetic code. *Nature Reviews Genetics*, *2*(1), 49–58. https://doi.org/10.1038/35047500

Ko, M., An, J., Bandukwala, H. S., Chavez, L., Äijö, T., Pastor, W. A., Segal, M. F., Li, H., Koh, K. P., Lähdesmäki, H., Hogan, P. G., Aravind, L., & Rao, A. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature*, *497*(7447), 122–126. https://doi.org/10.1038/nature12052

Ko, M., Huang, Y., Jankowska, A. M., Pape, U. J., Tahiliani, M., Bandukwala, H. S., An, J., Lamperti, E. D., Koh, K. P., Ganetzky, R., Liu, X. S., Aravind, L., Agarwal, S., MacIejewski, J. P., & Rao, A. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*, *468*(7325), 839–843. https://doi.org/10.1038/nature09586

Koh, K. P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C. A., Mostoslavsky, G., Lahesmaa, R., Orkin, S. H., Rodig, S. J., Daley, G. Q., & Rao, A. (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*, *8*(2), 200–213.

https://doi.org/10.1016/j.stem.2011.01.008

Kohli, R. M., & Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. In *Nature* (Vol. 502, Issue 7472, pp. 472–479). https://doi.org/10.1038/nature12750

Kokura, K., Kaul, S. C., Wadhwa, R., Nomura, T., Khan, M. M., Shinagawa, T., Yasukawa, T., Colmenares, C., & Ishii, S. (2001). The Ski Protein Family Is Required for MeCP2-mediated Transcriptional Repression. *Journal of Biological Chemistry*, *276*(36), 34115–34121. https://doi.org/10.1074/jbc.M105747200

Kriaucionis, S., & Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. *Science*, *324*(5929), 929–930. https://doi.org/10.1126/science.1169786

Lang, K., & Chin, J. W. (2014). Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chemical Reviews*, *114*(9), 4764–4806. https://doi.org/10.1021/cr400355w

le Guezennec, X., Vermeulen, M., Brinkman, A. B., Hoeijmakers, W. A. M., Cohen, A., Lasonder, E., & Stunnenberg, H. G. (2006). MBD2/NuRD and MBD3/NuRD, Two Distinct Complexes with Different Biochemical and Functional Properties. *Molecular and Cellular Biology*, *26*(3), 843–851. https://doi.org/10.1128/mcb.26.3.843-851.2006

Leinfelder, W., Zehelein, E., MandrandBerthelot, M., & Bock, A. (1988). Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature*, *331*(6158), 723–725. https://doi.org/10.1038/331723a0

Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M., & Schultz, P. G. (2007). Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nature Chemical Biology*, *3*(12), 769–772. https://doi.org/10.1038/nchembio.2007.44

Lewis, J. D., Ft Meehan, R., Henzel, W. J., Maurer-Fogy, lngrid, Jeppesen, P., Klein, F., & Bird, A. (1992). Purification, Sequence, and Cellular Localization of a Novel Chromosomal Protein That Binds to Methylated DNA. In *Cell* (Vol. 69).

Ley, T. J., Ding, L., Walter, M. J., McLellan, M. D., Lamprecht, T., Larson, D. E., Kandoth, C., Payton, J. E., Baty, J., Welch, J., Harris, C. C., Lichti, C. F., Townsend, R. R., Fulton, R. S., Dooling, D. J., Koboldt, D. C., Schmidt, H., Zhang, Q., Osborne, J. R., … Wilson, R. K. (2010). DNMT3A Mutations in Acute Myeloid Leukemia. *New England Journal of Medicine*, *363*(25), 2424–2433. https://doi.org/10.1056/NEJMoa1005143

Liaunardy-Jopeace, A., Murton, B. L., Mahesh, M., Chin, J. W., & James, J. R. (2017). Encoding optical control in LCK kinase to quantitatively investigate its activity in live cells. *Nature Structural and Molecular Biology*, *24*(12), 1155–1163. https://doi.org/10.1038/nsmb.3492

Lin, T.-C., Palei, S., & Summerer, D. (2022). Optochemical Control of TET Dioxygenases Enables Kinetic Insights into the Domain-Dependent Interplay of TET1 and MBD1 while Oxidizing and Reading 5-Methylcytosine. *ACS Chemical Biology*, *17*(7), 1844–1852.

https://doi.org/10.1021/acschembio.2c00245

Lindahl, Abraham, Hess, & van der Spoel. (2020). *GROMACS 2020.1 Source code* (2020.1). Zenodo. https://doi.org/10.5281/zenodo.3685919

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–322. https://doi.org/10.1038/nature08514

Lo, C. L., Choudhury, S. R., Irudayaraj, J., & Zhou, F. C. (2017). Epigenetic Editing of Ascl1 Gene in Neural Stem Cells by Optogenetics. *Scientific Reports*, *7*. https://doi.org/10.1038/srep42047

Lo, K., Brinkman, R. R., & Gottardo, R. (2008). Automated gating of flow cytometry data via robust model-based clustering. *Cytometry Part A*, *73*(4), 321–332. https://doi.org/10.1002/cyto.a.20531

Lo, K., Hahne, F., Brinkman, R. R., & Gottardo, R. (2009). flowClust: A Bioconductor package for automated gating of flow cytometry data. *BMC Bioinformatics*, *10*(1), 1–8. https://doi.org/10.1186/1471-2105-10-145

Long, H. K., Blackledge, N. P., & Klose, R. J. (2013). ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochemical Society Transactions*, *41*(3), 727–740. https://doi.org/10.1042/BST20130028

Lu, H., Bhoopatiraju, S., Wang, H., Schmitz, N. P., Wang, X., Freeman, M. J., Forster, C. L., Verneris, M. R., Linden, M. A., & Hallstrom, T. C. (2016). *Loss of UHRF2 expression is associated with human neoplasia, promoter hypermethylation, decreased 5-hydroxymethylcytosine, and high proliferative activity* (Vol. 7, Issue 46). www.impactjournals.com/oncotarget

Lu, X., Zhao, B. S., & He, C. (2015). TET family proteins: Oxidation activity, interacting molecules, and functions in diseases. In *Chemical Reviews* (Vol. 115, Issue 6, pp. 2225–2239). American Chemical Society. https://doi.org/10.1021/cr500470n

Ludwig, A. K., Zhang, P., & Cardoso, M. C. (2016). Modifiers and Readers of DNA Modifications and Their Impact on Genome Structure, Expression, and Stability in Disease. *Frontiers in Genetics*, *7*(JUN). https://doi.org/10.3389/fgene.2016.00115

Ludwig, A. K., Zhang, P., Hastert, F. D., Meyer, S., Rausch, C., Herce, H. D., Müller, U., Lehmkuhl, A., Hellmann, I., Trummer, C., Storm, C., Leonhardt, H., & Cardoso, M. C. (2017). Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise. *Nucleic Acids Research*, *45*(5), 2438–2457. https://doi.org/10.1093/nar/gkw1197

Lukinavičius, G., Blaukopf, C., Pershagen, E., Schena, A., Reymond, L., Derivery, E., Gonzalez-Gaitan, M., D'Este, E., Hell, S. W., Gerlich, D. W., & Johnsson, K. (2015). SiR–Hoechst is a far-red DNA stain for live-cell nanoscopy. *Nature Communications 2015 6:1*, *6*(1), 1–7.

https://doi.org/10.1038/ncomms9497

Lunyak, V. v., Burgess, R., Prefontaine, G. G., Nelson, C., Sze, S.-H., Chenoweth, J., Schwartz, P., Pevzner, P. A., Glass, C., Mandel, G., & Rosenfeld, M. G. (2002). Corepressor-Dependent Silencing of Chromosomal Regions Encoding Neuronal Genes. *Science*, *298*(5599), 1747–1752. https://doi.org/10.1126/science.1076469

Luo, J., Arbely, E., Zhang, J., Chou, C., Uprety, R., Chin, J. W., & Deiters, A. (2016). Genetically encoded optical activation of DNA recombination in human cells. *Chemical Communications*, *52*(55), 8529–8532. https://doi.org/10.1039/c6cc03934k

Luo, J., Uprety, R., Naro, Y., Chou, C., Nguyen, D. P., Chin, J. W., & Deiters, A. (2014). Genetically encoded optochemical probes for simultaneous fluorescence reporting and light activation of protein function with two-photon excitation. *Journal of the American Chemical Society*, *136*(44), 15551–15558. https://doi.org/10.1021/ja5055862

Lyst, M. J., Ekiert, R., Ebert, D. H., Merusi, C., Nowak, J., Selfridge, J., Guy, J., Kastan, N. R., Robinson, N. D., de Lima Alves, F., Rappsilber, J., Greenberg, M. E., & Bird, A. (2013). Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nature Neuroscience*, *16*(7), 898–902. https://doi.org/10.1038/nn.3434

Mainz, E. R., Wang, Q., Lawrence, D. S., & Allbritton, N. L. (2016). An Integrated Chemical Cytometry Method: Shining a Light on Akt Activity in Single Cells. *Angewandte Chemie*, *128*(42), 13289–13292. https://doi.org/10.1002/ange.201606914

Maiti, A., & Drohat, A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *Journal of Biological Chemistry*, *286*(41), 35334–35338. https://doi.org/10.1074/jbc.C111.284620

Malek, M., Taghiyar, M. J., Chong, L., Finak, G., Gottardo, R., & Brinkman, R. R. (2015). FlowDensity: Reproducing manual gating of flow cytometry data by automated density-based cell population identification. *Bioinformatics*, *31*(4), 606–607. https://doi.org/10.1093/bioinformatics/btu677

Manoilov, K. Yu., Verkhusha, V. v., & Shcherbakova, D. M. (2021). A guide to the optogenetic regulation of endogenous molecules. *Nature Methods*, *18*(9), 1027–1037. https://doi.org/10.1038/s41592-021-01240-1

Mayer, W., Niveleau, A., Walter, J., Fundele, R., & Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature*, *403*, 501–502. www.nature.com

Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L., & Bird, A. P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell*, *58*(3), 499–507. https://doi.org/10.1016/0092-8674(89)90430-3

Melamed, P., Yosefzon, Y., David, C., Tsukerman, A., & Pnueli, L. (2018). Tet Enzymes, Variants, and Differential Effects on Function. *Frontiers in Cell and Developmental Biology*, *6*(MAR). https://doi.org/10.3389/fcell.2018.00022

Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., & Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell*, *151*(7), 1417–1430. https://doi.org/10.1016/j.cell.2012.11.022

Meur, N. le, Hahne, F., & Ellis, B. (2007). FlowCore: data structures package for flow cytometry data. In *Bioconductor Project*. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:flowCore+:+data+structures+package+for+flow+cytometry+data#2

Müller, H. M., Fiegl, H., Goebel, G., Hubalek, M. M., Widschwendter, A., Müller-Holzner, E., Marth, C., & Widschwendter, M. (2003). MeCP2 and MBD2 expression in human neoplastic and non-neoplastic breast tissue and its association with oestrogen receptor status. *British Journal of Cancer*, *89*(10), 1934–1939. https://doi.org/10.1038/sj.bjc.6601392

Münzel, M., Lischke, U., Stathis, D., Pfaffeneder, T., Gnerlich, F. A., Deiml, C. A., Koch, S. C., Karaghiosoff, K., & Carell, T. (2011). Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine. *Chemistry - A European Journal*, *17*(49), 13782–13788. https://doi.org/10.1002/chem.201102782

Nan, X., Ng, H.-H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., & Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, *393*(6683), 386–389. https://doi.org/10.1038/30764

Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., & Chin, J. W. (2010). Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature*, *464*(7287), 441–444. https://doi.org/10.1038/nature08817

Ng, H.-H., Jeppesen, P., & Bird, A. (2000). Active Repression of Methylated Genes by the Chromosomal Protein MBD1. *Molecular and Cellular Biology*, *20*(4), 1394–1406. https://doi.org/10.1128/MCB.20.4.1394-1406.2000/ASSET/A20D4EC3-0167-4451-AF0F-096CF2241A8C/ASSETS/GRAPHIC/MB0401490007.JPEG

Ng, H.-H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., & Bird, A. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nature Genetics*, *23*(1), 58–61. https://doi.org/10.1038/12659

Nguyen, D. P., Mahesh, M., Elsässer, S. J., Hancock, S. M., Uttamapinant, C., & Chin, J. W. (2014). Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *Journal of the American Chemical Society*, *136*(6), 2240–2243. https://doi.org/10.1021/ja412191m

Nihongaki, Y., Kawano, F., Nakajima, T., & Sato, M. (2015). Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nature Biotechnology*, *33*(7), 755–760. https://doi.org/10.1038/nbt.3245

Ohki, I. (1999). Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. *The EMBO Journal*, *18*(23), 6653–6661. https://doi.org/10.1093/emboj/18.23.6653

Ohki, I., Shimotake, N., Fujita, N., Jee, J.-G., Ikegami, T., Nakao, M., & Shirakawa, M. (2001). Solution Structure of the Methyl-CpG Binding Domain of Human MBD1 in Complex with Methylated DNA. *Cell*, *105*(4), 487–497. https://doi.org/10.1016/S0092-8674(01)00324-5

Ohki, I., Shimotake, N., Fujita, N., Nakao, M., & Shirakawa, M. (1999). Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. *EMBO Journal*, *18*(23), 6653–6661. https://doi.org/10.1093/emboj/18.23.6653

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., & Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. *Current Biology*, *10*(8), 475–478.

Palei, S., Buchmuller, B., Wolffgramm, J., Muñoz-Lopez, Á., Jung, S., Czodrowski, P., & Summerer, D. (2020). Light-Activatable TET-Dioxygenases Reveal Dynamics of 5-Methylcytosine Oxidation and Transcriptome Reorganization. *Journal of the American Chemical Society*, *142*(16), 7289–7294. https://doi.org/10.1021/jacs.0c01193

Patchornik, A., Amit, B., & Woodward, R. B. (1970). Photosensitive protecting groups. *Journal of the American Chemical Society*, *92*(21), 6333–6335. https://doi.org/10.1021/ja00724a041

Peng, L., Li, Y., Xi, Y., Li, W., Li, J., Lv, R., Zhang, L., Zou, Q., Dong, S., Luo, H., Wu, F., & Yu, W. (2016). MBD3L2 promotes Tet2 enzymatic activity for mediating 5-methylcytosine oxidation. *Journal of Cell Science*, *129*(5), 1059–1071. https://doi.org/10.1242/jcs.179044

Petronzelli, F., Riccio, A., Markham, G. D., Seeholzer, S. H., Genuardi, M., Karbowski, M., Yeung, A. T., Matsumoto, Y., & Bellacosa, A. (2000). Investigation of the Substrate Spectrum of the Human Mismatch-Specific DNA N-Glycosylase MED1 (MBD4): Fundamental Role of the Catalytic Domain. In *J. Cell. Physiol* (Vol. 185).

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/JCC.20084

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). <scp>UCSF ChimeraX</scp> : Structure visualization for researchers, educators, and developers. *Protein Science*, *30*(1), 70–82. https://doi.org/10.1002/pro.3943

Pfaffeneder, T., Hackner, B., Truß, M., Münzel, M., Müller, M., Deiml, C. A., Hagemeier, C., & Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angewandte Chemie - International Edition*, *50*(31), 7008–7012. https://doi.org/10.1002/anie.201103899

Phu, V., Jiang, W., Gottardo, R., & Finak, G. (2018). GgCyto: Next generation open-source

visualization software for cytometry. *Bioinformatics*, *34*(22), 3951–3953. https://doi.org/10.1093/bioinformatics/bty441

Pitolli, C., Wang, Y., Candi, E., Shi, Y., Melino, G., & Amelio, I. (2019). P53-mediated tumor suppression: DNA-damage response and alternative mechanisms. In *Cancers* (Vol. 11, Issue 12). MDPI AG. https://doi.org/10.3390/cancers11121983

Pollard, T. D., & de La Cruz, E. M. (2013). Take advantage of time in your experiments: a guide to simple, informative kinetics assays. *Molecular Biology of the Cell*, *24*(8), 1103–1110. https://doi.org/10.1091/mbc.e13-01-0030

Polycarpo, C., Ambrogelly, A., Bérubé, A., Winbush, S. M., McCloskey, J. A., Crain, P. F., Wood, J. L., & Söll, D. (2004). An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proceedings of the National Academy of Sciences*, *101*(34), 12450–12454. https://doi.org/10.1073/pnas.0405362101

Prokhortchouk, A., Hendrich, B., Jørgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., & Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes and Development*, *15*(13), 1613–1618. https://doi.org/10.1101/gad.198501

Pudasaini, A., El-Arab, K. K., & Zoltowski, B. D. (2015). LOV-based optogenetic devices: light-driven modules to impart photoregulated control of cellular signaling. *Frontiers in Molecular Biosciences*, *2*(MAY). https://doi.org/10.3389/fmolb.2015.00018

Rackham, O., & Chin, J. W. (2005). A network of orthogonal ribosome·mrna pairs. *Nature Chemical Biology*, *1*(3), 159–166. https://doi.org/10.1038/nchembio719

Rajan, M., Mortusewicz, O., Rothbauer, U., Hastert, F. D., Schmidthals, K., Rapp, A., Leonhardt, H., & Cardoso, M. C. (2015). Generation of an alpaca-derived nanobody recognizing γ-H2AX. *FEBS Open Bio*, *5*(1), 779–788. https://doi.org/10.1016/j.fob.2015.09.005

Rakauskaite, R., Urbanavičiute, G., Rukšenaite, A., Liutkevičiute, Z., Juškenas, R., Masevičius, V., & Klimašauskas, S. (2015). Biosynthetic selenoproteins with genetically-encoded photocaged selenocysteines. *Chemical Communications*, *51*(39), 8245–8248. https://doi.org/10.1039/c4cc07910h

Razin, A., & Riggs, A. D. (1980). DNA Methylation and Gene Function. *Science*, *210*(4470), 604–610. https://www.science.org

Rooman, M., Liévin, J., Buisine, E., & Wintjens, R. (2002). Cation–π/H-bond Stair Motifs at Protein–DNA Interfaces. *Journal of Molecular Biology*, *319*(1), 67–76. https://doi.org/10.1016/S0022-2836(02)00263-2

Saito, M., & Ishikawa, F. (2002). The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *Journal of Biological Chemistry*, *277*(38), 35434–35439. https://doi.org/10.1074/jbc.M203455200

Sato, N., Kondo, M., & Arai, K. ichi. (2006). The orphan nuclear receptor GCNF recruits DNA

methyltransferase for Oct-3/4 silencing. *Biochemical and Biophysical Research Communications*, *344*(3), 845–851. https://doi.org/10.1016/j.bbrc.2006.04.007

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. https://doi.org/10.1038/nmeth.2019

Schmidt, D., Tillberg, P. W., Chen, F., & Boyden, E. S. (2014). A fully genetically encoded protein architecture for optical control of peptide ligand concentration. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms4019

Shahbazian, M. D., Antalffy, B., Armstrong, D. L., & Zoghbi, H. Y. (2002). Insight into Rett syndrome: MeCP2 levels display tissue-and cell-specific differences and correlate with neuronal maturation. In *Human Molecular Genetics* (Vol. 11, Issue 2).

Shapovalov, M. v., & Dunbrack, R. L. (2011). A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure (London, England : 1993)*, *19*(6), 844. https://doi.org/10.1016/J.STR.2011.03.019

Smallwood, S. A., & Kelsey, G. (2012). De novo DNA methylation: a germ cell perspective. *Trends in Genetics*, *28*(1), 33–42. https://doi.org/10.1016/j.tig.2011.09.004

Smith, Z. D., & Meissner, A. (2013). DNA methylation: Roles in mammalian development. *Nature Reviews Genetics*, *14*(3), 204–220. https://doi.org/10.1038/nrg3354

Sprague, B. L., Pego, R. L., Stavreva, D. A., & McNally, J. G. (2004). Analysis of Binding Reactions by Fluorescence Recovery after Photobleaching. *Biophysical Journal*, *86*(6), 3473–3495. https://doi.org/10.1529/biophysj.103.026765

Sprague, B., & McNally, J. (2005). FRAP analysis of binding: proper and fitting. *Trends in Cell Biology*, *15*(2), 84–91. https://doi.org/10.1016/j.tcb.2004.12.001

Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W. T. C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., Eberl, H. C., Mensinga, A., Brinkman, A. B., Lephikov, K., Müller, U., Walter, J., Boelens, R., van Ingen, H., Leonhardt, H., … Vermeulen, M. (2013). Dynamic readers for 5-(Hydroxy)methylcytosine and its oxidized derivatives. *Cell*, *152*(5), 1146–1159. https://doi.org/10.1016/j.cell.2013.02.004

Srinivasan, G., James, C. M., & Krzycki, J. A. (2002). Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA. *Science*, *296*(5572), 1459–1462. https://doi.org/10.1126/science.1069588

Stroberg, W., & Schnell, S. (2017). On the validity and errors of the pseudo-first-order kinetics in ligand–receptor binding. *Mathematical Biosciences*, *287*, 3–11. https://doi.org/10.1016/j.mbs.2016.09.010

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., & Rao, A. (2009). Conversion of 5-methylcytosine to 5-

hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, *324*(5929), 930–935. https://doi.org/10.1126/science.1170116

Tang, Y., & Tirrell, D. A. (2002). Attenuation of the editing activity of the Escherichia coli Leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. *Biochemistry*, *41*(34), 10635–10645. https://doi.org/10.1021/bi026130x

Tarhonskaya, H., Rydzik, A. M., Leung, I. K. H., Loik, N. D., Chan, M. C., Kawamura, A., McCullagh, J. S. O., Claridge, T. D. W., Flashman, E., & Schofield, C. J. (2014). Non-enzymatic chemistry enables 2-hydroxyglutarate-mediated activation of 2-oxoglutarate oxygenases. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms4423

Tronick, E., & Hunter, R. G. (2016). Waddington, Dynamic systems, and epigenetics. *Frontiers in Behavioral Neuroscience*, *10*(JUN). https://doi.org/10.3389/fnbeh.2016.00107

Unoki, M., Nishidate, T., & Nakamura, Y. (2004). ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. *Oncogene*, *23*(46), 7601–7610. https://doi.org/10.1038/sj.onc.1208053

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., & Mackerell, A. D. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, *31*(4), 671–690. https://doi.org/10.1002/JCC.21367

Vanommeslaeghe, K., & MacKerell, A. D. (2012). Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *Journal of Chemical Information and Modeling*, *52*(12), 3144–3154. https://doi.org/10.1021/CI300363C/SUPPL_FILE/CI300363C_SI_001.ZIP

Vanommeslaeghe, K., Raman, E. P., & MacKerell, A. D. (2012). Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *Journal of Chemical Information and Modeling*, *52*(12), 3155–3168. https://doi.org/10.1021/CI3003649/SUPPL_FILE/CI3003649_SI_001.PDF

Waddington, C. H. (1942). The epigenotype. *Endeavour*, *1*, 18–20.

Waddington, C. H. (1957). *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology.* London: Allen & Unwin.

Waddington, C. H. (2012). The epigenotype. 1942. *International Journal of Epidemiology*, *41*(1), 10–13. https://doi.org/10.1093/ije/dyr184

Walker, J. W., Martin, H., Schmitt, F. R., & Barsotti, R. J. (1993). Rapid release of an .alpha.-adrenergic receptor ligand from photolabile analogs. *Biochemistry*, *32*(5), 1338–1345. https://doi.org/10.1021/bi00056a020

Wan, J., Su, Y., Song, Q., Tung, B., Oyinlade, O., Liu, S., Ying, M., Ming, G., Song, H., Qian, J., Zhu, H., & Xia, S. (2017). Methylated cis-regulatory elements mediate KLF4-dependent gene

transactivation and cell migration. *ELife*, *6*. https://doi.org/10.7554/eLife.20068

Wang, K., Neumann, H., Peak-Chew, S. Y., & Chin, J. W. (2007). Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature Biotechnology*, *25*(7), 770–777. https://doi.org/10.1038/nbt1314

Weber, A. R., Krawczyk, C., Robertson, A. B., Kusnierczyk, A., Vågbø, C. B., Schuermann, D., Klungland, A., & Schar, P. (2016). Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms10806

Wilkins, B. J., Marionni, S., Young, D. D., Liu, J., Wang, Y., di Salvo, M. L., Deiters, A., & Cropp, T. A. (2010). Site-specific incorporation of fluorotyrosines into proteins in escherichia coli by photochemical disguise. *Biochemistry*, *49*(8), 1557–1559. https://doi.org/10.1021/bi100013s

Wolffgramm, J., Buchmuller, B., Palei, S., Muñoz-López, Á., Kanne, J., Janning, P., Schweiger, M. R., & Summerer, D. (2021). Light-Activation of DNA-Methyltransferases. *Angewandte Chemie - International Edition*, *60*(24), 13507–13512. https://doi.org/10.1002/anie.202103945

Wong, E., Yang, K., Kuraguchi, M., Werling, U., Avdievich, E., Fan, K., Fazzari, M., Jin, B., Brown, A. M. C., Lipkin, M., & Edelmann, W. (2002). Mbd4 inactivation increases C→T transition mutations and promotes gastrointestinal tumor formation. *Proceedings of the National Academy of Sciences*, *99*(23), 14937–14942. https://doi.org/10.1073/pnas.232579299

Wood, K. H., & Zhou, Z. (2016). Emerging molecular and biological functions of MBD2, a reader of DNA methylation. In *Frontiers in Genetics* (Vol. 7, Issue MAY). Frontiers Media S.A. https://doi.org/10.3389/fgene.2016.00093

Wu, H., & Zhang, Y. (2014). Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions. *Cell*, *156*(1–2), 45–68. https://doi.org/10.1016/j.cell.2013.12.019

Wu, N., Deiters, A., Cropp, T. A., King, D., & Schultz, P. G. (2004). A genetically encoded photocaged amino acid. *Journal of the American Chemical Society*, *126*(44), 14306–14307. https://doi.org/10.1021/ja040175z

Wu, X., & Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature Reviews Genetics*, *18*(9), 517–534. https://doi.org/10.1038/nrg.2017.33

Xu, F., Mao, C., Ding, Y., Rui, C., Wu, L., Shi, A., Zhang, H., Zhang, L., & Xu, Z. (2010). Molecular and Enzymatic Profiles of Mammalian DNA Methyltransferases: Structures and Targets for Drugs. *Current Medicinal Chemistry*, *17*(33), 4052. https://doi.org/10.2174/092986710793205372

Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., Zhang, K., & Zhang, Y. (2012). Tet1 controls meiosis by regulating meiotic gene expression. *Nature*, *492*(7429), 443–447. https://doi.org/10.1038/nature11709

Yang, H., Liu, Y., Bai, F., Zhang, J. Y., Ma, S. H., Liu, J., Xu, Z. D., Zhu, H. G., Ling, Z. Q., Ye, D., Guan, K. L.,

& Xiong, Y. (2013). Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene*, *32*(5), 663–669. https://doi.org/10.1038/onc.2012.67

Yazawa, M., Sadaghiani, A. M., Hsueh, B., & Dolmetsch, R. E. (2009). Induction of protein-protein interactions in live cells using light. *Nature Biotechnology*, *27*(10), 941–945. https://doi.org/10.1038/nbt.1569

Yildirim, O., Li, R., Hung, J. H., Chen, P. B., Dong, X., Ee, L. S., Weng, Z., Rando, O. J., & Fazzio, T. G. (2011a). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*, *147*(7), 1498–1510. https://doi.org/10.1016/j.cell.2011.11.054

Yildirim, O., Li, R., Hung, J. H., Chen, P. B., Dong, X., Ee, L. S., Weng, Z., Rando, O. J., & Fazzio, T. G. (2011b). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*, *147*(7), 1498–1510. https://doi.org/10.1016/j.cell.2011.11.054

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., Kivioja, T., Dave, K., Zhong, F., Nitta, K. R., Taipale, M., Popov, A., Ginno, P. A., Domcke, S., Yan, J., Schübeler, D., Vinson, C., & Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, *356*(6337). https://doi.org/10.1126/science.aaj2239

Yu, W., He, X., Vanommeslaeghe, K., & MacKerell, A. D. (2012). Extension of the CHARMM General Force Field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of Computational Chemistry*, *33*(31), 2451–2468. https://doi.org/10.1002/JCC.23067

Zangi, R., Arrieta, A., & Cossío, F. P. (2010). Mechanism of DNA Methylation: The Double Role of DNA as a Substrate and as a Cofactor. *Journal of Molecular Biology*, *400*(3), 632–644. https://doi.org/10.1016/J.JMB.2010.05.021

Zeller, P., & Gasser, S. M. (2017). The Importance of Satellite Sequence Repression for Genome Stability. *Cold Spring Harbor Symposia on Quantitative Biology*, *82*, 15–24. https://doi.org/10.1101/sqb.2017.82.033662

Zhang, P., Rausch, C., Hastert, F. D., Boneva, B., Filatova, A., Patil, S. J., Nuber, U. A., Gao, Y., Zhao, X., & Cardoso, M. C. (2017). Methyl-CpG binding domain protein 1 regulates localization and activity of Tet1 in a CXXC3 domain-dependent manner. *Nucleic Acids Research*, *45*(12), 7118–7136. https://doi.org/10.1093/nar/gkx281

Zhang, W., Xia, W., Wang, Q., Towers, A. J., Chen, J., Gao, R., Zhang, Y., Yen, C. an, Lee, A. Y., Li, Y., Zhou, C., Liu, K., Zhang, J., Gu, T. P., Chen, X., Chang, Z., Leung, D., Gao, S., Jiang, Y. hui, & Xie, W. (2016). Isoform Switch of TET1 Regulates DNA Demethylation and Mouse Development. *Molecular Cell*, *64*(6), 1062–1073. https://doi.org/10.1016/j.molcel.2016.10.030

Zhao, J., Zhao, J., Lin, S., Huang, Y., & Chen, P. R. (2013). Mechanism-based design of a photoactivatable firefly luciferase. *Journal of the American Chemical Society*, *135*(20), 7410–7413. https://doi.org/10.1021/ja4013535

Zhao, X., Ueba, T., Christie, B. R., Barkho, B., McConnell, M. J., Nakashima, K., Lein, E. S., Eadie, B. D., Willhoite, A. R., Muotri, A. R., Summers, R. G., Chun, J., Lee, K.-F., & Gage, F. H. (2003). Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. *Proceedings of the National Academy of Sciences*, *100*(11), 6777–6782. https://doi.org/10.1073/pnas.1131928100

Zhou, X. X., Fan, L. Z., Li, P., Shen, K., & Lin, M. Z. (2017). Optical control of cell signaling by single-chain photoswitchable kinases. *Science*, *355*(6327), 836–842. https://doi.org/10.1126/science.aah3605

Zhu, H., Wang, G., & Qian, J. (2016). Transcription factors as readers and effectors of DNA methylation. *Nature Reviews Genetics*, *17*(9), 551–565. https://doi.org/10.1038/nrg.2016.83

Zimmerman, G., Chow, L.-Y., & Paik, U.-J. (1958). The Photochemical Isomerization of Azobenzene [1]. *Journal of the American Chemical Society*, *80*(14), 3528–3531. https://doi.org/10.1021/ja01547a010

# Eidesstattliche Versicherung (Affidavit)

Lin, Tzu-Chen

Name, Vorname
(Surname, first name)

206446

Matrikel-Nr.
(Enrolment number)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden, § 63 Abs. 5 Hochschulgesetz NRW.

Die Abgabe einer falschen Versicherung an Eides statt ist strafbar.

Wer vorsätzlich eine falsche Versicherung an Eides statt abgibt, kann mit einer Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft werden, § 156 StGB. Die fahrlässige Abgabe einer falschen Versicherung an Eides statt kann mit einer Freiheitsstrafe bis zu einem Jahr oder Geldstrafe bestraft werden, § 161 StGB.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offence can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offences of this type is the chancellor of the TU Dortmund University. In the case of multiple or other serious attempts at deception, the candidate can also be unenrolled, Section 63, paragraph 5 of the Universities Act of North Rhine-Westphalia.

The submission of a false affidavit is punishable.

Any person who intentionally submits a false affidavit can be punished with a prison sentence of up to three years or a fine, Section 156 of the Criminal Code. The negligent submission of a false affidavit can be punished with a prison sentence of up to one year or a fine, Section 161 of the Criminal Code.

I have taken note of the above official notification.

Dortmund,

Ort, Datum
(Place, date)

Unterschrift
(Signature)

Titel der Dissertation:
(Title of the thesis):

Optochemical Control of DNA Methylation Erasers and Readers enables Kinetic Insights into their Domain-Dependent Interplay

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel selbstständig und ohne unzulässige fremde Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht.
Die Arbeit hat in gegenwärtiger oder in einer anderen Fassung weder der TU Dortmund noch einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegen.

I hereby swear that I have completed the present dissertation independently and without inadmissible external support. I have not used any sources or tools other than those indicated and have identified literal and analogous quotations.

The thesis in its current version or another version has not been presented to the TU Dortmund University or another university in connection with a state or academic examination.*

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the PhD thesis is the official and legally binding version.

Dortmund,

Ort, Datum
(Place, date)

Unterschrift
(Signature)