Advances in Experimental Medicine and Biology 1389

Albert Jeltsch
Renata Z. Jurkowska *Editors*

DNA Methyltransferases - Role and Function

Second Edition

Advances in Experimental Medicine and Biology

Volume 1389

Series Editors

Wim E. Crusio, Institut de Neurosciences Cognitives et Intégratives d'Aquitaine, CNRS and University of Bordeaux, Pessac Cedex, France

Haidong Dong, Departments of Urology and Immunology, Mayo Clinic, Rochester, MN, USA

Heinfried H. Radeke, Institute of Pharmacology & Toxicology, Clinic of the Goethe University Frankfurt Main, Frankfurt am Main, Hessen, Germany

Nima Rezaei, Research Center for Immunodeficiencies, Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran

Ortrud Steinlein, Institute of Human Genetics, LMU University Hospital, Munich, Germany

Junjie Xiao, Cardiac Regeneration and Ageing Lab, Institute of Cardiovascular Sciences, School of Life Science, Shanghai University, Shanghai, China

Advances in Experimental Medicine and Biology provides a platform for scientific contributions in the main disciplines of the biomedicine and the life sciences. This series publishes thematic volumes on contemporary research in the areas of microbiology, immunology, neurosciences, biochemistry, biomedical engineering, genetics, physiology, and cancer research. Covering emerging topics and techniques in basic and clinical science, it brings together clinicians and researchers from various fields.

Advances in Experimental Medicine and Biology has been publishing exceptional works in the field for over 40 years, and is indexed in SCOPUS, Medline (PubMed), EMBASE, BIOSIS, Reaxys, EMBiology, the Chemical Abstracts Service (CAS), and Pathway Studio.

2021 Impact Factor: 3.650 (no longer indexed in SCIE as of 2022)

Albert Jeltsch • Renata Z. Jurkowska
Editors

DNA Methyltransferases - Role and Function

Second Edition

Editors
Albert Jeltsch
Department of Biochemistry,
Institute of Biochemistry and
Technical Biochemistry
University of Stuttgart
Stuttgart, Germany

Renata Z. Jurkowska
School of Biosciences, Division of
Biomedicine
Cardiff University
Cardiff, UK

ISSN 0065-2598 ISSN 2214-8019 (electronic)
Advances in Experimental Medicine and Biology
ISBN 978-3-031-11453-3 ISBN 978-3-031-11454-0 (eBook)
https://doi.org/10.1007/978-3-031-11454-0

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

DNA is the key to the inheritance of genetic information in living organisms, and the mechanism of duplication of double-stranded DNA is arguably the molecular process with the widest appreciation in the general public. However, almost since their discovery in DNA, it was known that the nucleobases could be further modified by the addition of methyl groups. DNA methylation patterns resulting from the site-specific presence and absence of methyl marks are often heritable, leading to the classification of DNA methylation as an epigenetic mark. We now know that DNA methylation plays key roles in almost all species, ranging from bacteria to lower and higher eukaryotes and plants. Moreover, changes in DNA methylation are associated with the development of human diseases and the field of epigenetics is currently exploding with connections to nutrition, behavior, and transgenerational inheritance of traits. The comprehension of the relevance of DNA methylation in various fields of biology and medicine has also brought considerable attention to the enzymes responsible for the transfer of methyl groups to DNA, the so-called DNA methyltransferases.

We both have studied DNA methyltransferases for many years, attracted by their complicated mechanisms, beautiful structures, and medical relevance; actually, these enzymes faithfully accompanied us almost our entire career. Therefore, we felt very honored and excited by the offer from SPRINGER to edit a book on these fascinating enzymes, and happily took up this challenge. The second edition of the book now comes 6 years after the first edition, which has been received by the field with great interest and positive feedback. Remarkable progress has been made in DNA methyltransferase research in these few years, with new mechanistic insights and connections to biological processes and human diseases. This revised book provides a compilation of chapters that recapitulate and update many of the developments made in the field, including past achievements and future challenges. All the chapters were written by renowned experts, who themselves made central contributions to this vibrant field.

The introduction of the book (Chap. 1) by Jurkowska and Jeltsch recaptures the development of the field over the last 60 years, highlighting and conceptualizing many key discoveries. Chapter 2 written by Casadesús and Sánchez-Romero describes bacterial DNA methyltransferases and the important role of DNA methylation in bacteria, highlighting many fascinating new findings. The next three chapters cover DNA methylation and DNA methyltransferases in mammals. Tajima et al. focus in Chap. 3 on the

enormous progress made in the structural investigation of the mammalian DNA methyltransferases mainly based on the work from the authors' lab. Then, Jurkowska and Jeltsch describe novel insights into the enzymatic properties of DNMTs and their regulation in cells (Chap. 4). In Chap. 5, Dan and Chen review the important contributions of genetic studies to our current understanding of DNA methylation and DNA methyltransferases. Next, new developments regarding structures and mechanisms of plant DNA methyltransferases are described in Chap. 6 written by Leichter et al., and in Chap. 7, Wedd et al. present the newest insights into the role of DNA methylation in honeybees as an example for DNA methylation in lower eukaryotes. The ongoing progress of science in the field is illustrated by the recent discovery of adenine-N6 methylation in several higher eukaryotes and the discussion of whether this mark, still controversial in some cases, functions as an epigenetic signal, as described in Chap. 8 by O'Brown and Greer. The next chapters focus on the pathways of DNA demethylation (Chap. 9 written by Dean) and the structure and mechanism of TET enzymes, which are involved in this reaction (Chap. 10 contributed by Yin et al.). In Chap. 11, the topic of DNA methylation is approached from another angle, and Liu et al. summarize the mechanisms involved in DNA methylation readout. Next, in Chap. 12, Ren et al. review base flipping as a basic mechanism involved in the setting, reading, and erasing of DNA methylation, illustrating fundamental structural and mechanistic processes. The next part of the book connects DNA MTases to important biological processes and covers their role in disease, with Chap. 13 by Weisenberger et al. recapitulating the now well-established role of DNA methylation in cancer, Chap. 14 by Sarkies describing the intricate involvement of DNA methyltransferases in DNA damage, and Chap. 15 by Yildiz and Zimmer-Bensch discussing the functions of DNA methylation in the brain. The last part of the book is devoted to the ongoing technological development. In Chap. 16, Tost provides a comprehensive review of methods to study DNA methylation. Following on this, Lopez et al. provide an overview of the development and potential application of DNMT inhibitors in cancer and other diseases in Chap. 17, and Cortés-Mancera et al. describe emerging approaches to editing DNA methylation patterns in a targeted manner (Chap. 18). Finally, in Chap. 19, Tomkuvienė et al. describe developing applications of DNA methyltransferases as molecular biology tools to label DNA.

We anticipate many more years of exciting research focusing on DNA methylation and DNA methyltransferases, with many new and groundbreaking discoveries to come. It is our aim that this book serves as a rich, up-to-date, and reliable source of information for specialist scientists, but also students and researchers entering the field, providing them with a solid fundament of knowledge, concepts, and methods for future work. At the same time, it should help researchers to get into this fascinating subject, allowing them to catch up with the current level of knowledge and learn about recent trends in this active field of research.

Stuttgart, Germany — Albert Jeltsch
Cardiff, UK — Renata Z. Jurkowska
June 2022

Contents

1 Mechanisms and Biological Roles of DNA Methyltransferases and DNA Methylation: From Past Achievements to Future Challenges 1
Renata Z. Jurkowska and Albert Jeltsch

2 DNA Methylation in Prokaryotes 21
Josep Casadesús and María A. Sánchez-Romero

3 Domain Structure of the Dnmt1, Dnmt3a, and Dnmt3b DNA Methyltransferases 45
Shoji Tajima, Isao Suetake, Kohei Takeshita, Atsushi Nakagawa, Hironobu Kimura, and Jikui Song

4 Enzymology of Mammalian DNA Methyltransferases 69
Renata Z. Jurkowska and Albert Jeltsch

5 Genetic Studies on Mammalian DNA Methyltransferases . . . 111
Jiameng Dan and Taiping Chen

6 Structure and Mechanism of Plant DNA Methyltransferases 137
Sarah M. Leichter, Jiamu Du, and Xuehua Zhong

7 DNA Methylation in Honey Bees and the Unresolved Questions in Insect Methylomics 159
Laura Wedd, Robert Kucharski, and Ryszard Maleszka

8 N6-methyladenine: A Rare and Dynamic DNA Mark 177
Zach Klapholz O'Brown and Eric Lieberman Greer

9 Pathways of DNA Demethylation 211
Wendy Dean

10 Structure and Function of TET Enzymes 239
Xiaotong Yin, Lulu Hu, and Yanhui Xu

11 Proteins That Read DNA Methylation 269
Ke Liu, Takashi Shimbo, Xiaosheng Song, Paul A. Wade, and Jinrong Min

12 Recent Advances on DNA Base Flipping: A General Mechanism for Writing, Reading, and Erasing DNA Modifications 295
Ren Ren, John R. Horton, Samuel Hong, and Xiaodong Cheng

13 The Role of DNA Methylation and DNA Methyltransferases in Cancer 317
Daniel J. Weisenberger, Ranjani Lakshminarasimhan, and Gangning Liang

14 DNA Methyltransferases and DNA Damage 349
Peter Sarkies

15 Role of DNMTs in the Brain 363
Can Bora Yildiz and Geraldine Zimmer-Bensch

16 Current and Emerging Technologies for the Analysis of the Genome-Wide and Locus-Specific DNA Methylation Patterns 395
Jörg Tost

17 Inhibitors of DNA Methylation 471
Marie Lopez, Julie Gilbert, Jean Contreras, Ludovic Halby, and Paola B. Arimondo

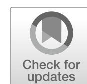
18 Gene-Targeted DNA Methylation: Towards Long-Lasting Reprogramming of Gene Expression? 515
Fabian M. Cortés-Mancera, Federica Sarno, Désirée Goubert, and Marianne G. Rots

19 DNA Labeling Using DNA Methyltransferases 535
Miglė Tomkuvienė, Edita Kriukienė, and Saulius Klimašauskas

About the Editors

Albert Jeltsch is a biochemist who received his PhD in Biochemistry in 1994 at the Medical University Hannover (Germany). After Postdoc and Assistant Professorship (2003) at Justus-Liebig University in Giessen, he was appointed as Associate Professor of Biochemistry at Jacobs University Bremen in 2003, where he became Full Professor of Biochemistry in 2006. Prof. Jeltsch moved to the University of Stuttgart in 2011, where he is heading the Department of Biochemistry at the Institute of Biochemistry and Technical Biochemistry. For more than 25 years, the group of Prof. Jeltsch has investigated the structure, mechanism, and function of bacterial and mammalian DNA methyltransferases. They have long-standing expertise in the field of rational and evolutionary protein design of DNA interacting enzymes and in the design of chimeric methylation enzymes for epigenome editing and gene regulation in eukaryotic cells. In addition, they study the specificity and activity of protein lysine methyltransferases and methyllysine reading domains and the biological role of protein methylation in cells.

Renata Z. Jurkowska is a molecular biologist with general interests in epigenetics, stem cells, and lung biology. She obtained her M.Sc. degree in Biotechnology at Warsaw University (Poland) and went on to complete her PhD in Biochemistry and two postdocs in Germany, where she investigated the specificity and molecular mechanism of mammalian DNA methyltransferases and other epigenetic enzymes for 10 years. In 2015, she joined BioMed X Innovation Center (Germany) as Principal Investigator, where she worked at the interface between academia and industry, leading a biomedical preclinical project in the field of respiratory medicine and epigenetics. In September 2019, she was appointed Senior Lecturer in the School of Biosciences at Cardiff University, UK. Her current research focuses on understanding how epigenetic regulation drives cellular identity in the healthy lung and how dysregulation of epigenetic processes due to environmental insults contributes to the development of lung diseases. Her group uses epigenomics technologies in combination with molecular biology tools to advance the mechanistic understanding of lung diseases, identify epigenetic biomarkers for diagnosis, and devise novel epigenetic therapeutic strategies.



1 Mechanisms and Biological Roles of DNA Methyltransferases and DNA Methylation: From Past Achievements to Future Challenges

Renata Z. Jurkowska and Albert Jeltsch

Abstract

DNA methylation and DNA methyltransferases (MTases)—the enzymes that introduce the methylation mark into the DNA—have been studied for almost 70 years. In this chapter, we review the key developments in the DNA methylation field that have led to our current understanding of the structures and mechanisms of DNA MTases. We discuss the essential biological roles of DNA methylation, including the discovery of DNA methylation, cloning and sequence analysis of the bacterial and eukaryotic MTases, and the elucidation of their structure, mechanism, regulation, and molecular evolution. We describe genetic studies that contributed greatly to the evolving views on the role of DNA methylation in development and diseases, the invention of methods for the genome-wide analysis of DNA methylation, and the biochemical identification of DNA MTases and the TET enzyme family, which is involved in DNA demethylation. We summarize the roles of MTases in bacterial epigenetics and the application of MTases in synthetic biology to generate artificial signaling systems. We finish by highlighting some open questions for the next years of research in the field.

Keywords

DNA methyltransferase · Enzyme mechanism · Protein structure · DNA methylation · MTase

Abbreviations

AdoMet	S-adenosyl-L-methionine
AML	acute myeloid leukemia
HPLC	High-pressure liquid chromatography
MTase	DNA methyltransferase
RM system	Restriction/modification system
SMRT sequencing	Single-molecule real-time sequencing
TET enzyme	Ten-eleven translocation enzyme
TLC	Thin layer chromatography

R. Z. Jurkowska (✉)
School of Biosciences, Cardiff University, Cardiff, UK
e-mail: jurkowskar@cardiff.ac.uk

A. Jeltsch (✉)
Department of Biochemistry, Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Stuttgart, Germany
e-mail: albert.jeltsch@ibtb.uni-stuttgart.de

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_1

1.1 Discovery of DNA Methylation

DNA from various biological sources contains the methylated bases C5-methylcytosine, N4-methylcytosine and N6-methyladenine in addition to the four standard nucleobases (Fig. 1.1a). Methylation of cytosine at the C5-position had been discovered in calf thymus DNA already in 1948 using paper chromatography experiments (Hotchkiss 1948), and 6-methyladenine was found in bacterial DNA in 1955 (Dunn and Smith 1955). N4-methylcytosine, the third and least common methylated base in bacterial DNA, was described for the first time in 1983 (Janulaitis et al. 1983). The methylation of nucleobases at all these positions places the methyl groups in the major groove of the double-stranded B-DNA, where they do not interfere with the Watson/Crick base pairing, but can easily be detected by proteins interacting with the DNA (Fig. 1.1b). DNA methylation can, for example, directly prevent the readout of an AT base pair by glutamine residues in the major groove (Fig. 1.1c). By this and related processes, DNA methylation can control the binding of proteins to DNA and thereby regulate biochemical processes taking place on DNA, for example the expression of the genetic information. Hence, the methylation adds extra information to the DNA that is not encoded in the DNA sequence, and the methylated bases can be considered the fifth, sixth, and seventh letters of the genetic alphabet (Jeltsch 2002). The epigenetic toolbox has been further diversified with the discovery of the oxidized forms of 5-methylcytosine, viz. 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine, in the DNA of many species, including mammals; followed by the discovery of Ten-eleven translocation (TET) enzymes, responsible for generating them (Tahiliani et al. 2009; Kriaucionis and Heintz 2009; Munzel et al. 2010; Pfaffeneder et al. 2011; Ito et al. 2011; He et al. 2011). Despite the interesting properties of the identified methylated bases and their importance in living organisms, many years had passed after the initial discovery of DNA methylation until work with DNA methyltransferases, the enzymes that introduce this modification had been systematically started.

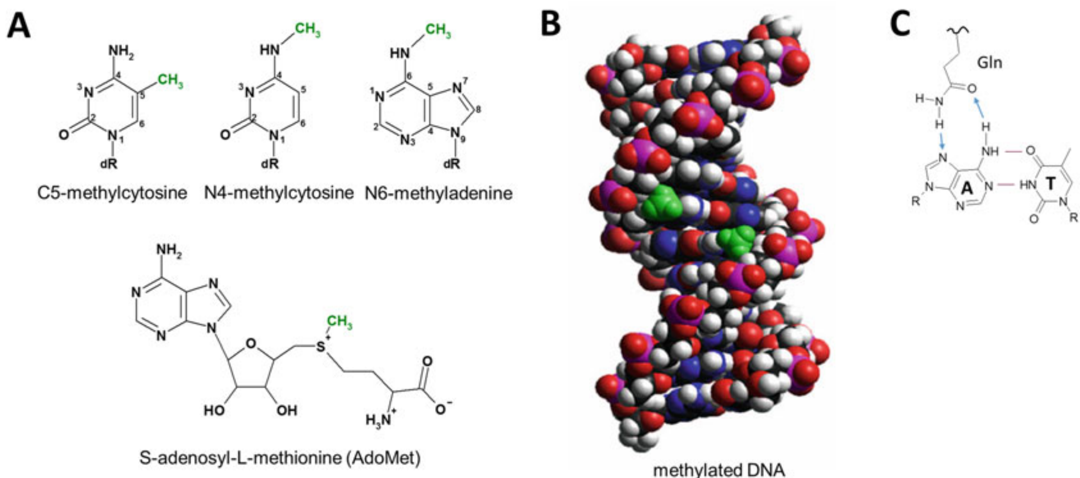


Fig. 1.1 Molecules related to DNA methylation. (**a**) Structures of the methylated bases that occur in DNA and of the AdoMet cofactor, the universal donor for all DNA methylation reactions. (**b**) Space fill model of the structure of B-DNA with a methylated CpG site. The methyl groups are shown in green in the major groove of the DNA. (**c**) Example of the major groove readout of an AT base pair by Gln as proposed by Seeman and colleagues in 1976 (Seeman et al. 1976). This contact is disrupted by methylation of the A at the N6 position

1.2 Discovery and Early Work on DNA MTases

DNA methyltransferases were initially discovered as parts of the restriction/modification (RM) systems, which consist of a DNA methyltransferase and an associated restriction endonuclease (Arber and Dussoix 1962). S-adenosyl-L-methionine (AdoMet)-dependent DNA and RNA methylation activity was first described by Gold in 1963 (Gold et al. 1963) and a series of papers published by Gold in 1964 (Gold et al. 1964; Gold and Hurwitz 1964a, 1964b; Hurwitz et al. 1964a, 1964b). The *E. coli* EcoDam (a solitary bacterial MTase that is not part of an RM system) was initially described in 1973 (Marinus and Morris 1973) and purified in 1982 (Herman and Modrich 1982). The first enzymatic studies with human and murine DNA MTases were reported in the late 1970s and early 1980s (Browne et al. 1977; Gruenbaum et al. 1982). However, in the 1970s and 1980s, DNA MTases remained a kind of passengers in the ongoing molecular biology revolution, due to their functional and genetic association to restriction endonucleases, which were absolutely essential as analytical and cloning tools at this time (Arber and Linn 1969; Boyer 1971; Meselson et al. 1972). In addition, restriction endonucleases and DNA MTases constituted the first model systems to study the sequence-specific DNA recognition, a process essential to the control of gene expression in all forms of life (Modrich 1982).

With the increasing commercial importance of restriction endonucleases, biotech companies were interested to shift the production procedures away from the purification of enzymes from the original bacterial strains toward recombinant expression of cloned enzymes. Therefore, cloning of restriction enzymes moved into the center of scientific and economic interest. It was known that RM systems often reside on mobile genetic elements, with the genes encoding the methyltransferase and the endonuclease located next to each other. Hence, cloning of a DNA fragment containing the methyltransferase gene often led to the cloning of the restriction enzyme gene on the same DNA insert. In a procedure called the "Hungarian trick," the group of Venetianer realized that the special properties of DNA methyltransferases could be exploited to selectively clone genes encoding these enzymes (Szomolanyi et al. 1980). This approach was based on the fact that after the expression of a DNA MTase in cells, the enzyme modified its own encoding DNA. Hence, after shotgun cloning of bacterial genomes, the plasmids containing DNA inserts were isolated and cleaved with a restriction enzyme of interest. The protected DNA likely coded for a methyltransferase, which methylated DNA within the target region of the endonuclease and thereby prevented plasmid cleavage. After cloning of these protected inserts, it turned out that very often the gene for the restriction enzyme was found on the same piece of DNA next to the methyltransferase gene. Almost 20 years later, a similar coupling of genotype and phenotype after the expression of DNA methyltransferases was applied by Tawfik and colleagues to develop a novel approach for protein engineering, which was based on the expression of libraries of MTase mutants in water/oil emulsions (Tawfik and Griffiths 1998).

1.3 DNA MTases Contain Conserved Amino Acids Sequence Motifs

The wide application of the above-described and related cloning procedures led to the cloning of hundreds of restriction enzymes together with their corresponding DNA methyltransferases. Therefore, the group of bacterial DNA methyltransferases provided a rich source of enzymes known to recognize different DNA sequences for enzymatic, biochemical, and evolutionary studies, which has led to many important insights and breakthrough discoveries (Wilson and Murray 1991; Pingoud and Jeltsch 1997; Pingoud et al. 2014). Comparisons of the amino acid sequences of DNA methyltransferases in the early days of multiple sequence alignments

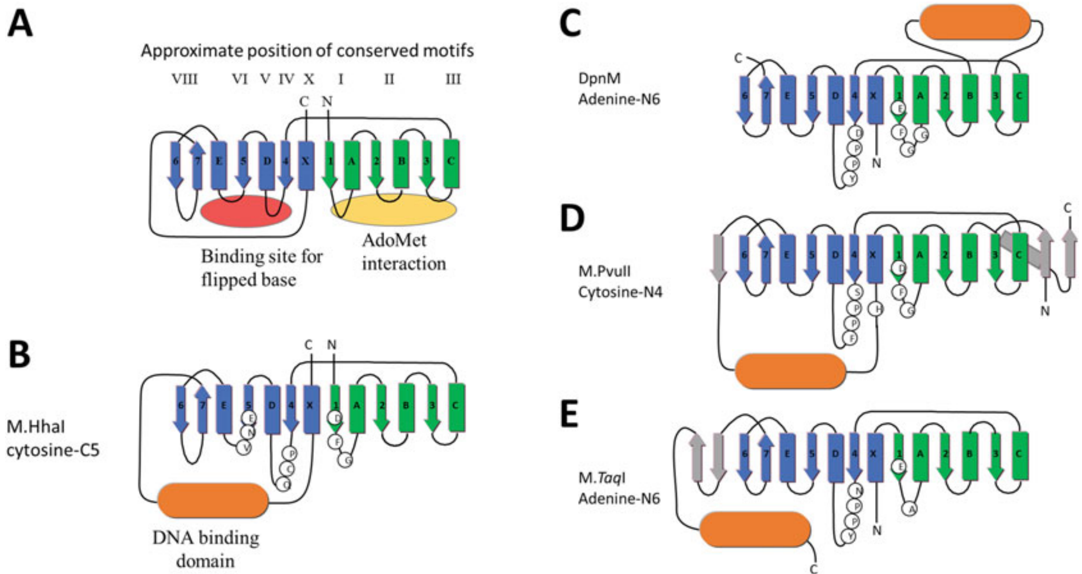


Fig. 1.2 Topological scheme of the universal AdoMet dependent DNA MTase fold. It consists of two Rossman fold half-domains with several conserved amino acid sequences, one subdomain forming the binding site for the AdoMet and the second for the flipped base. DNA recognition is mediated by a DNA binding domain, which is variable in sequence and structure. The linear arrangements of the functional elements vary between different DNA MTases by circular permutation, generating several characteristic subgroups of DNA MTases. (**a**) Schematic representation of the general structure of the DNA MTase fold. (**b**) General structure of Cytosine-C5 MTases. (**c–d**) General structure of three subgroups of Adenine-N6 and Cytosine-N4 MTases

led to the discovery of up to ten amino acid motifs characteristic of cytosine-C5 methyltransferases (Posfai et al. 1989; Klimasauskas et al. 1989; Lauster et al. 1989) (Fig. 1.2). In 1988, Bestor cloned the first mammalian DNA methyltransferase that turned out to share extensive sequence similarity with the bacterial cytosine-C5 methyltransferases in its C-terminal catalytic part (Bestor et al. 1988). It was discovered that bacterial adenine-N6 methyltransferases contained conserved amino acid motifs as well (Fig. 1.2) (Lauster et al. 1987; Guschlbauer 1988) and some of the MTase motifs were shown to be part of the general signature motifs of AdoMet -dependent methyltransferases, including small molecule, protein, and RNA methyltransferases (Kagan and Clarke 1994; Ingrosso et al. 1989). Although statistical methods were insufficient at that time, these studies led to the identification of the key catalytic regions both in adenine-N6 and cytosine-C5 methyltransferases. Many of the most conserved residues in both families of enzymes were shown to be directly involved in the catalytic process (Cheng 1995; Jeltsch 2002) and several amino acid motifs identified in the early alignment studies could later be connected to defined structural elements in the conserved methyltransferase fold (Malone et al. 1995; Roth et al. 1998) (Fig. 1.2).

1.4 Structure and Mechanism of DNA MTases

All DNA methyltransferases use AdoMet as a methyl group donor. Based on their target atom, one can distinguish methyltransferases adding the methyl group to carbon or nitrogen. The former group comprises cytosine-C5, the latter adenine-N6 and cytosine-N4 methyltransferases. All DNA MTases follow a ternary complex mechanism, where the catalytically competent complex consists of the enzyme, the DNA substrate and the AdoMet cofactor. In some enzymes, binding

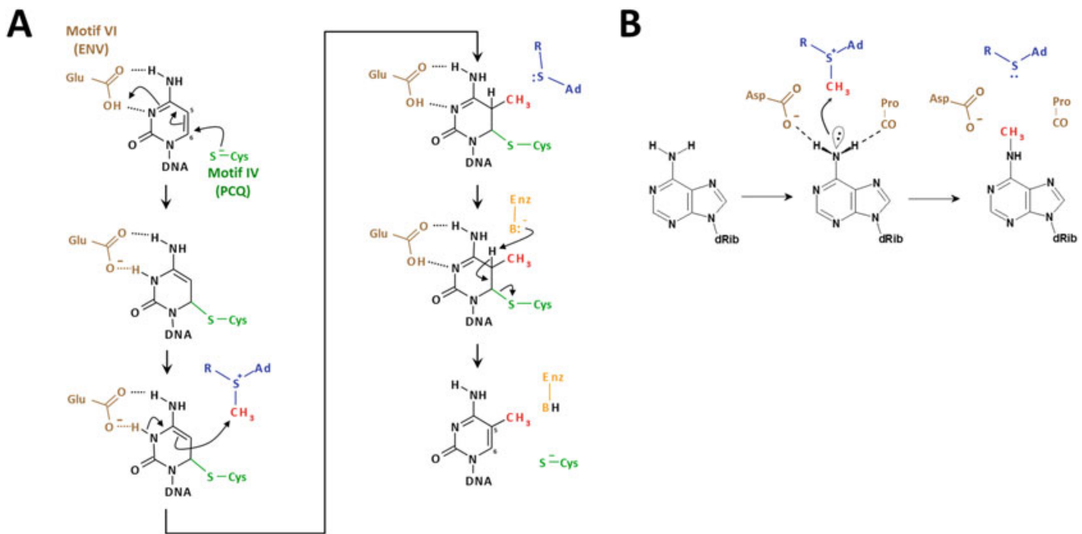


Fig. 1.3 Catalytic mechanism of DNA MTases. (**a**) Mechanism of DNA-(cytosine C5)-MTases. (**b**) Mechanism of DNA-(adenine N6)-MTases

of the DNA substrate and the AdoMet occurs in an ordered reaction, in other cases it is random. Wu and Santi studied the catalytic mechanism of cytosine-C5 methyltransferases and proposed in 1985 that it follows a Michael addition reaction, which is characterized by the formation of a covalent intermediate between the enzyme and the target base (Wu and Santi 1985, 1987) (Fig. 1.3a). Shortly afterwards, Santi and coworkers also showed that adenine methylation proceeds directly at the N6 position, despite the poor nucleophilicity of the N6 atom, and not by a transient transfer of the methyl group to the N1 followed by its shift to the N6 (Pogolotti et al. 1988). Seminal insights into the folding of the methyltransferases and the arrangement of their catalytic centers came with the first structure of a DNA methyltransferase (the bacterial M.HhaI enzyme) that was solved in 1993 (Cheng et al. 1993). In 1994, the publication of the first structure of a DNA methyltransferase (again M.HhaI) with its DNA substrate by Cheng and coworkers led to another conceptual breakthrough regarding the catalytic mechanism of DNA methyltransferases (Klimasauskas et al. 1994) (Fig. 1.4). It was observed that the target base for the methylation reaction was completely rotated out of the DNA helix and inserted into a catalytic pocket of the enzyme in a process called "base flipping", which is necessary to allow for the close access of the catalytic residues to the substrate base. This unexpected and seminal discovery highlighted the flexibility of DNA and the dynamic processes that accompany enzymatic catalysis; research subjects that were intensively studied afterwards. Today, we appreciate the base flipping as a universal process in DNA methylation, but also in other reactions occurring on DNA, including DNA repair (Roberts 1995; Roberts and Cheng 1998). Since then, numerous crystal structures of DNA methyltransferases with bound substrate DNA provided additional confirmation for base flipping among bacterial and eucaryotic enzymes, reinforcing this pivotal discovery.

The first structure of an adenine-N6 MTase (M.TaqI) was published in 1994 as well (but without DNA) (Labahn et al. 1994), unexpectedly showing that both enzyme families contain a large catalytic domain with an identical fold, consisting of a similar six-stranded parallel ß-sheet with a seventh strand inserted in an antiparallel fashion between the fifth and sixth strands (Schluckebier et al. 1995) (Fig. 1.2). This fold is known today as the AdoMet dependent methyltransferase fold (Martin and McMillan 2002). The seven-stranded ß-sheet is flanked by α-helices creating two subdomains with a

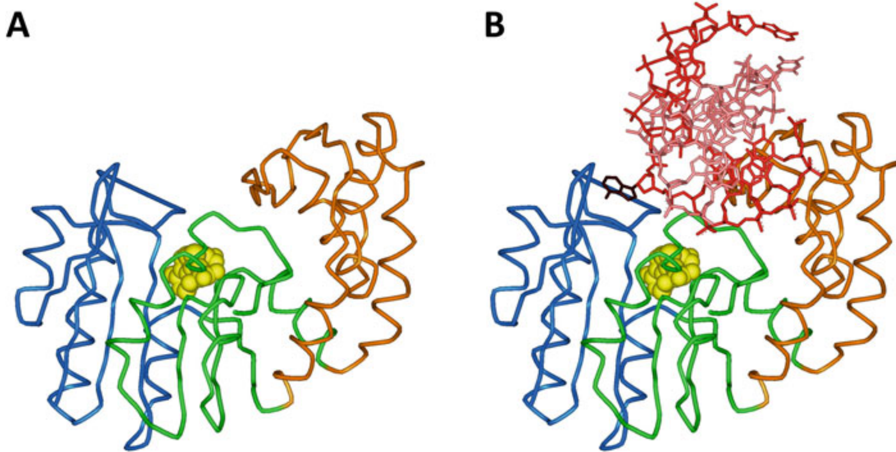


Fig. 1.4 Ribbon model of the structure of the EcoDam DNA MTase (Horton et al. 2006). The AdoMet and the base binding subdomains are shown in green and blue, respectively. AdoMet is displayed in space-filled form in yellow. The DNA binding domain is colored in orange. (**a**) Structure of the EcoDam-AdoMet complex. (**b**) Structure of the complex of EcoDam with bound substrate DNA (red, the flipped adenine base is shown in black)

Rossman fold architecture: one containing the binding site for the AdoMet and the other for the flipped base (Cheng 1995; Jeltsch 2002). In addition, all MTases contain a second less-conserved domain involved in DNA recognition. The first structure of an adenine-N6 MTase with DNA was solved in 2001 (Goedecke et al. 2001). It showed that the N6 of the flipped adenine is positioned in a tetrahedral environment of hydrogen bond donors provided by the conserved residues of the (DNS)PP(YFW) motif, suggesting that its nucleophilicity is increased by a change in hybridization from sp2 to sp3 (Fig. 1.3b). Cytosine-N4 MTases are believed to follow an analogous mechanism, based on the chemical similarity of the methyl-acceptor atom and the observation that the specificity of enzymes form these families overlap, i.e., that adenine-N6 MTases can also methylate cytosine at N4 and Cytosine-N4 MTases also methylate adenine (Jeltsch et al. 1999; Jeltsch 2001).

1.5 Molecular Evolution of MTases

As described above, the conserved structure of the 7-ß strand MTases consists of two half domains with Rossman folds fused to each other. One of them mediates the AdoMet interaction, the second provides the binding sites for the methylation substrates, flipped nucleobases in the case of DNA MTases. The high structural similarity of all DNA MTases and the presence of the conserved motifs with similarities even between different groups of MTases, suggests that these enzymes are monophyletic. Presumably, the two subdomains originated from a duplication of a primordial AdoMet binding Rossman fold domain (Malone et al. 1995). Later, one subdomain continued to bind AdoMet, whereas the second diverged to generate the binding pockets for different methylation substrates including flipped cytosine and adenine bases, but also small molecules like catecholamine or amino acids like arginine, leading to the appearance of various groups of contemporary MTases specific for different methylation substrates. Moreover, the initial MTase ancestor has undergone several modifications during molecular evolution, including circular permutations (Jeltsch 1999; Bujnicki 2002), and in the case of DNA MTases, the insertion of diverse and unrelated DNA binding domain at different places in the consensus structure, leading to the creation of different classes of DNA MTases (Malone et al. 1995) (Fig. 1.2). In Eukaryotes, the catalytic

domain of C5 MTases has taken part in diverse domain shuffling events and became fused with many other chromatin interacting domains found in animals and plants, including PWWP, ADD, BAH, and Chromodomains. The acquisition of the various chromatin and protein/nuclei acid interacting modules by eukaryotic MTases allowed the functional coupling and crosstalk between different layers of epigenetic information, including histone post-translational modifications, chromatin remodeling, and non-coding RNAs.

Unexpectedly, it turned out that the presence of 5-methylcytosine puts an evolutionary burden on organisms that affected the evolution of C5 MTases and shaped cellular DNA methylomes (Jeltsch 2002). One factor contributing to this is the accelerated deamination of 5-methylcytosine (2–4 times) compared to cytosine. The mutational threat of this reaction is further increased by the lower repair efficiency of T/G mismatches (arising from deamination of 5-methylcytosine), as compared to U/G mismatches (arising from deamination of unmethylated cytosine). Therefore, specific repair systems have been developed to handle T/G mismatches, appearing in the sequence context of previously methylated cytosine bases. For example, in mammals, specific repair enzymes exist that act on T/G mismatches in a TpG context, including the 5-methylcytosine binding protein MBD4 (Hendrich et al. 1999). Similarly, in *E. coli* the VSR mismatch endonuclease triggers the repair at T/G mismatches originating from the deamination of methylated CCWGG sites, the target of the endogenous *E. coli* Dcm DNA-(cytosine C5)-MTase (Hennecke et al. 1991). Despite the existence of these and other specialized repair systems, methylation-mediated mutagenesis had a strong influence on the genome evolution of vertebrates, where generally a strong depletion of CpG sites in the DNA sequences is observed (Cross et al. 1994). Recently, it has been shown that C5 MTases also generate 3-methylcytosine in a side reaction by methylating the N3 atom of the cytosine base (Rosic et al. 2018). This modified base is toxic for the cells, because it interferes with RNA synthesis and DNA replication. It can be removed by members of the ALKB2 family of DNA alkylation repair enzymes, in an iron(II) and 2-oxoglutarate-dependent oxidation process. In fact, ALKB2 enzymes are apparently required in all species which express active C5-MTases, because an evolutionary analysis showed that these two enzyme families are very tightly connected, reflecting their essential functional link (Rosic et al. 2018). The disadvantages of the cytosine-C5 methylation system may explain why C5-MTases were repeatedly lost and gained in the evolution of eukaryotes and why they are lacking in several species (Jeltsch 2010; de Mendoza et al. 2021). This effect may also explain why adenine-N6 methylation is far more abundant in the bacterial world than cytosine-C5 methylation.

1.6 Early Views on the Biological Role of DNA Methylation

Methylation of human and mammalian DNA at CpG sites was identified at the beginning of the 1980s (Razin and Riggs 1980; Ehrlich and Wang 1981). In plants, DNA methylation was found also in CNG sites (where N is any nucleotide) (Gruenbaum et al. 1981) and non-symmetric sites. However, the early 1980s was a time when biology mainly focused on the detailed investigation of the so-called "model organisms". While this approach was extremely far-sighted and greatly contributed to the explosion in our understanding of the molecular basis of life, it did not come without risk, as illustrated by the general lack of appreciation for DNA methylation around that time. Due to an unfortunate coincidence, many of the carefully selected model organisms like *S. cerevisiae*, *D. melanogaster*, or *C. elegans* were lacking detectable DNA methylation. Consistently, there was a widespread belief that DNA methylation, although interesting, cannot be very important. On the other hand, it became clear that DNA methylation had an enormous influence on the human genome, when Bird discovered the existence of the CpG islands (Bird 1980; Bird

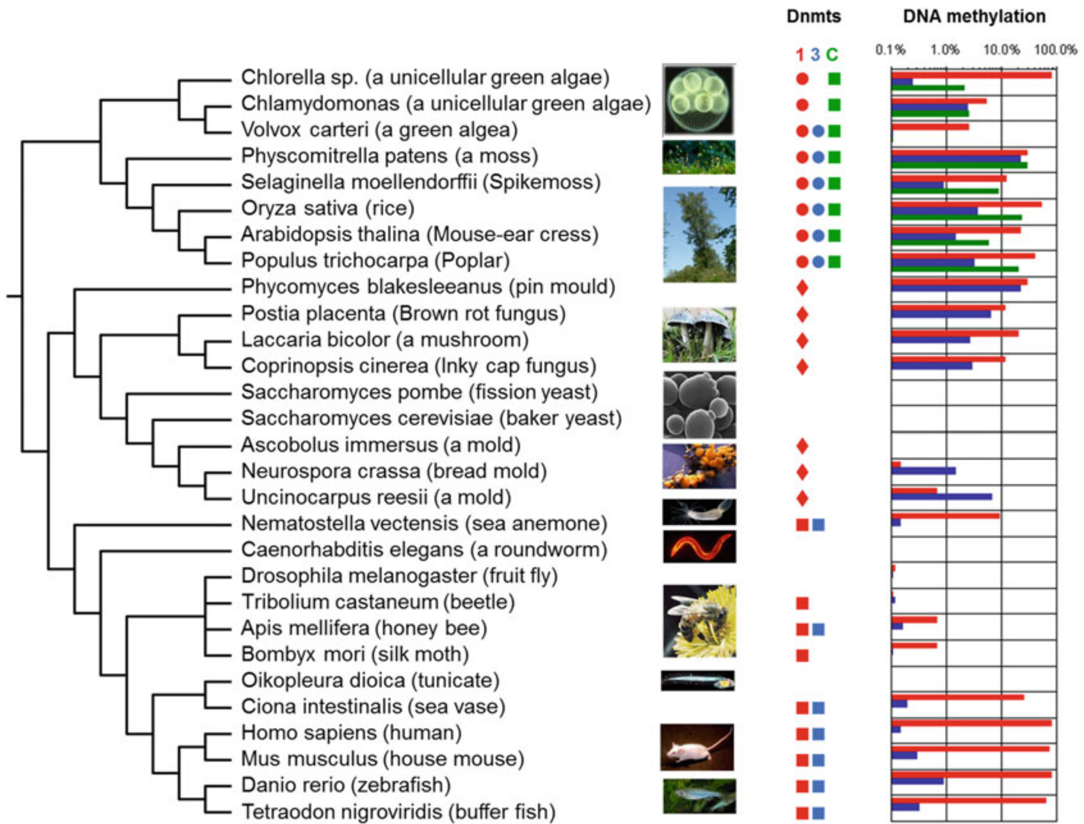


Fig. 1.5 Phylogenic distribution of DNA methylation systems and DNA MTases. The distribution of MTases of the DNMT1 (red), DNMT3 (blue), and chromomethylase families (green) are shown in several characteristic species. Red circles denote plant Met1 homologs, diamond enzymes of the fungal Dim-2 families, and squares DNMT1 homologs. Blue circles denote plant DRM homologs and squares DNMT3 enzymes. DNA methylation data were averaged as described in (Jeltsch 2010) and shown for CpG (red), CHH (blue), and CNG (green). CHG methylation is shown only for plants. The phylogenetic tree was generated with the National Center for Biotechnology Information taxonomy and the Interactive Tree of Life. Reproduced from (Jeltsch 2010) with modifications. Reprinted with permission from AAAS

et al. 1985), which are defined as regions of the high density of CpG sites within the genome that was already known to be globally depleted of this dinucleotide (Swartz et al. 1962). It was realized that the depletion of the CpG sites from the bulk genome was indirectly due to the mutagenic effect of cytosine-C5 methylation (as described above), leading to the preservation of CpG sites only if they were unmethylated, as in CpG islands. Today, we know that DNA methylation systems are found in almost all organisms and the model organisms devoid of them listed above appear to be rather exceptions (Fig. 1.5).

1.7 Genetic Studies on DNMTs in Mammals

While models connecting DNA methylation with known epigenetic phenomena, gene expression, and development (see for example (Riggs 1975; Holliday and Pugh 1975)) were developed, the general skepticism on the essential role of DNA methylation in human biology was only overcome with the discovery of the repressive function of DNA methylation on gene expression (Tazi and Bird 1990) and the finding that mice with a knock-out of DNMT1 (Li et al. 1992), the

only mammalian DNMT known by that time, die during early embryonic development in the uterus. However, as often in science, this discovery led to the next question, because it turned out that DNMT1 knock-out cells were not completely devoid of DNA methylation (Lei et al. 1996), which opened a hunt for additional mammalian DNMTs. Researchers tried to purify additional DNA methyltransferases from human and mouse cells; in parallel, the rising flood of DNA sequences was searched for entries containing the characteristic DNA methyltransferase motifs described above. It was the bioinformatics approach that was successful at the end, leading to the discovery of the DNMT3A and DNMT3B enzymes in 1998 (Okano et al. 1998). Shortly afterwards, both MTases were also shown to be essential in mice (Okano et al. 1999). Soon after, genetic studies showed that DNMT3A together with DNMT3L (a catalytically inactive paralog of DNMT3A and DNMT3B) were needed to set impinging marks in the mouse germline (Bourc'his et al. 2001; Bourc'his and Bestor 2004; Hata et al. 2002; Kaneda et al. 2004). Recently, DNMT3C, a rodent-specific paralog of DNMT3B essential for spermatogenesis and male fertility in mice, has been discovered, further diversifying the repertoire of mammalian MTases (Barau et al. 2016; Jain et al. 2017). It appears to have arisen from an ancient duplication of the DNMT3B gene that occurred ~46 million years ago in the in the Muroidea superfamily during rodent evolution and had been previously wrongly annotated as a nonfunctional pseudogene (Barau et al. 2016; Jain et al. 2017).

DNA methylation provides organisms with an efficient epigenetic regulatory system, which is particularly important in multicellular organisms, because of their need to develop stable cellular differentiation states. It has been proposed that the development of powerful epigenetic systems, comprising DNA methyltransferases, demethylases, and other enzyme systems introducing modifications on histones, had been a critical step in the evolution of multicellular life (Jeltsch 2013). Today, DNA methylation is recognized as an essential epigenetic mark that acts in concert with other chromatin modifications, like histone post-translational modifications, histone variants, or non-coding RNA. In mammals, DNA methylation is involved in the epigenetic processes, like genomic imprinting and X-chromosome inactivation, but it also has global roles in the generation of heterochromatin, silencing of repeats, and gene regulation during development and disease (Jurkowska et al. 2011).

1.8 Structure, Function, and Regulation of Mammalian DNA MTases

While time was progressing, structures of DNMT3A (Jia et al. 2007; Guo et al. 2015; Zhang et al. 2018), DNMT1 (Takeshita et al. 2011; Song et al. 2011, 2012; Adam et al. 2020), and lately also DNMT3B (Gao et al. 2020; Lin et al. 2020) were published showing that complicated regulatory processes, including oligomerization, distinct chromatin contacts, conformational changes and auto-inhibition, all interplay to accurately control the activity of these enzymes. In 1997, targeting of DNMT1 to replication foci via its interaction with PCNA was discovered (Chuang et al. 1997), but later it became clear that the interaction of DNMT1 with UHRF1 is even more essential for the targeting and activity of DNMT1 (Bostick et al. 2007; Sharif et al. 2007). Furthermore, it was found that in addition to the indirect targeting by other complex partners, DNMTs directly interact with chromatin. DNMT3A, DNMT3B, and DNMT3L use their ADD domains for binding to H3 tails unmethylated at K4 (Ooi et al. 2007; Zhang et al. 2010), DNMT3A and DNMT3B use their PWWP domains for binding to H3K36me2/3 (Dhayalan et al. 2010), and DNMT3A1 (an isoform of DNMT3A) binds with its UDR domain H2AK119ub1 (Weinberg et al. 2021). DNMT1 interacts with its replication foci targeting domain (RFTD) with ubiquitinated H3 tails (Nishiyama et al. 2013) and H3K9me3 (Ren et al. 2020) and via its BAH domain it binds

H3K40me3 (Ren et al. 2021). Moreover, the principles of the regulation of the activity and stability of DNMTs via post-translational modifications begin to emerge (Esteve et al. 2011; Deplus et al. 2014), adding another fascinating layer to the study of these enzymes.

1.9 Discovery of TET Enzymes

A similar changeful journey as in the field of DNA methylation was undertaken in the investigation of DNA demethylation, starting from the question of whether an active process of DNA demethylation might exist at all, leading to its discovery and the study of its mechanisms (Ooi and Bestor 2008). It was only in 2009 that the combination of powerful biochemical and bioinformatics approaches led to the discovery of the Ten-eleven Translocation (TET) enzymes (Tahiliani et al. 2009), which oxidize 5-methylcytosine to the hydroxymethyl, formyl or carboxyl state, and the discovery of these modified bases in human DNA (Tahiliani et al. 2009; Kriaucionis and Heintz 2009; Munzel et al. 2010; Pfaffeneder et al. 2011; Ito et al. 2011; He et al. 2011). The exact role of these additional modified bases and the complete pathway of DNA demethylation is not yet fully understood (Wu and Zhang 2017), but it has been well-established that the TDG glycosylase can remove 5-formylcytosine and 5-carboxylcytosine, eventually leading to the generation of an unmethylated cytosine (He et al. 2011; Maiti and Drohat 2011). Moreover, it is becoming increasingly clear that the TET-derived oxidative cytosine bases constitute distinct epigenetic marks by themselves that can be read by specialized reader domains and confer biological functions (Song and Pfeifer 2016; Song et al. 2021). Currently, it is understood that the genome-wide and locus-specific DNA methylation level is determined by a steady state reached through the combined action of MTases, demethylases, and DNA replication (Jeltsch and Jurkowska 2014).

1.10 Methods for Site-Specific Detection of DNA Methylation

The detection of DNA methylation for a long time was based on the initial methods: TLC (followed initially by HPLC, and today by mass spectrometry), allowing for a quantitative overall genome methylation analysis, but without sequence resolution; and restriction digestion using enzymes sensitive to DNA methylation, allowing site-specific genome-wide analysis, but only at defined restriction sites. For cytosine-C5 methylation, this situation dramatically changed with the development of the bisulfite conversion method, which can be combined with a battery of downstream technologies to enable a genome-wide analysis of 5-methylcytosine at single nucleotide resolution (Frommer et al. 1992; Clark et al. 1994). This technology in concert with the breakthroughs in DNA sequencing technologies has enabled researchers starting in 2008 to provide the first genome-wide DNA methylation maps of plant and mouse cells (Cokus et al. 2008; Lister et al. 2008; Meissner et al. 2008). Powerful variants of bisulfite sequencing, including oxidative bisulfite sequencing, have been developed, allowing detection of not only 5-methylcytosine but also its oxidized forms at single base resolution (Booth et al. 2013). Nanopore sequencing is another emerging technique that has been applied for the direct detection of 5-methylcytosine in genomic DNA (Laszlo et al. 2013; Lee et al. 2020; Sakamoto et al. 2020).

For N-methylation, it was only in 2010, almost 20 years after the discovery of the bisulfite technology, when the development of single-molecule real-time (SMRT) sequencing for the first time provided a method for the genome-wide analysis of 6-methyladenine at single nucleotide resolution (Flusberg et al. 2010). This invention was followed by a flurry of bacterial N6-adenine methylomes, including those of *E. coli* and *C. crescentus* (Sanchez-Romero et al. 2015), which provided novel insights into the role of DNA methylation in bacteria in

defense mechanisms, cell division, gene expression, and DNA repair. Recently, nanopore sequencing has been applied for the direct detection of 6-methyladenine in bacterial DNA as well (McIntyre et al. 2019). Moreover, nitrite sequencing has been described for the base-specific readout of 6-methyladenine by DNA sequencing (Mahdavi-Amiri et al. 2021). Conceptually similar to bisulfite sequencing used for the detection of cytosine methylation, nitrite sequencing is based on the selective conversion of adenine to hypoxanthine, which is blocked by adenine-N6 methylation.

1.11 DNA MTases and Bacterial Epigenetics

In bacteria, DNA methylation is involved in the control of DNA replication and repair, host defense by restriction/modification (RM) system, and control of gene expression (Jeltsch 2002; Wion and Casadesus 2006; Casadesus and Low 2006; Sanchez-Romero and Casadesus 2020). Most of the known bacterial DNA MTases introduce N6-methyladenine and belong to the RM systems, which serve as defense systems to protect prokaryotes from bacteriophage infections (Matic et al. 1996). RM systems comprise a restriction endonuclease and a DNA MTase, with both enzymes recognizing the same target DNA sequence. In the most common type II RM systems, the DNA target sites are 4–8 base pair long palindromic sequences. The restriction endonuclease cleaves the DNA at the target sequence only in an unmethylated state, as found on an invading phage DNA during the early steps of infection. The bacterial host cell DNA is kept in a methylated state by the corresponding DNA MTase and thereby it is protected from cleavage (Pingoud and Jeltsch 2001). These systems are complemented by the adaptable CRISPR-Cas systems that have been discovered only recently as an additional defense system of bacteria and archaea against bacteriophages, which is independent of DNA methylation (Bhaya et al. 2011).

Aside from RM systems, bacteria also contain so-called solitary or orphaned MTases that are not accompanied by a restriction enzyme. The *Escherichia coli* deoxyadenosine DNA methyltransferase (Dam) and the *Caulobacter crescentus* cell-cycle-regulated methyltransferase (CcrM) are two well-characterized examples of this type (Jeltsch 2002; Wion and Casadesus 2006; Casadesus and Low 2006; Sanchez-Romero and Casadesus 2020). *E. coli* Dam is involved in DNA mismatch repair, initiation of chromosome replication, and regulation of gene expression, including the pap phase variation in uropathogenic *E. coli* (Marinus and Morris 1973; Lobner-Olesen et al. 2005; Low and Casadesus 2008; Marinus and Casadesus 2009; van der Woude 2011), indicating that DNA methylation patterns are involved in bacterial gene regulation. In all these processes, the main underlying principle is that DNA methylation alters protein–DNA interactions, because the binding of repressor proteins to certain operator sites is impaired by DNA methylation. Although MTases are active in these bacteria, certain target sites can be protected against methylation by bound repressors, generating a methylation pattern. Under specific conditions, these patterns can be heritable and transmit epigenetic information used for example to regulate switching of cell states (Sanchez-Romero and Casadesus 2020). Of note, similar principles can also operate in eukaryotes, including mammals, as many transcriptional regulators show sensitivity to DNA methylation of their binding sites (Yin et al. 2017).

1.12 Role of DNA Methylation in Cancer

In 1983, the first groups reported global hypomethylation of DNA in cancer cells (Feinberg and Vogelstein 1983; Gama-Sosa et al. 1983). Shortly afterwards, the first examples of local hypermethylation at gene promoters of tumor suppressors, leading to their inactivation in cancer cells, were discovered (Baylin et al. 1986; Greger et al. 1989). It is now well-established that

these two processes, global DNA hypomethylation and regional hypermethylation, occur in most tumor cells and are directly connected to the progression of the disease (Baylin 2012; Bergman and Cedar 2013). In 2010, it was discovered that somatic mutations in DNMT3A are prevalently observed in Acute myeloid leukemia (AML) patients, among them the R882H exchange was found with particularly high frequency (Yamashita et al. 2010). Later work has confirmed and extended this finding and showed that mutations in DNMT3A are drivers of the disease process (Hamidi et al. 2015).

The frequent observation of the inactivation of tumor suppressor genes in cancers by hypermethylation has prompted the development of DNA methyltransferase inhibitors for clinical applications. This field was pioneered by Jones with the development of 5-azacytidine (Jones and Taylor 1980), which afterwards was confirmed to form an irreversible covalent complex with DNA methyltransferases (Santi et al. 1984). Later, cofactor analogs were also introduced to inhibit DNA methyltransferases (Reich and Mashhoon 1990). Today, many derivatives of these initial compounds have been developed and several are in clinical use for the treatment of cancer and other diseases (Yang et al. 2010; Fahy et al. 2012). Further development led to the idea of combination treatments, where DNMT inhibitors could be combined with other antitumor drugs, leading to higher efficacy (Ren et al. 2021).

1.13 Application of MTases in Artificial Epigenetic Systems

The field of synthetic biology emerged almost 20 years ago, based on several key developments in molecular biology techniques. Two reports in 2000 showing the construction and engineering of synthetic genetic circuits, a “toggle switch” and a “repressilator”, are now often considered hallmarks of the design of synthetic circuits and the field of synthetic biology in general (Gardner et al. 2000; Elowitz and Leibler 2000). Afterwards, engineered circuits became more diverse and versatile, with an increasing number of control elements integrated. In an artificial DNA methylation-based epigenetic memory system in *E.coli,* the CcrM DNA methyltransferase was combined with a designed zinc finger (ZnF), which only binds to unmethylated target sites (Maier et al. 2017). Hence, it regulates an operon expressing the CcrM MTase, which leads to positive feedback and stable switching of the system. With this system, an initial trigger, like the presence of arabinose, tetracyclin, or DNA damage is memorized in life *E. coli* cells (Maier et al. 2017; Ullrich et al. 2020). A switchable system based on *Salmonella enterica* and dam methyltransferase has been developed as well (Olivenza et al. 2019). Moreover, Park et al. described the development of a full artificial epigenetic system, based on adenine-N6 methylation in a human cell line (Park et al. 2019). It employs the dam MTase as a writer and the DpnI binding domain as reader of the GATC methylation. Notably, it was shown to regulate gene expression in a DNA methylation-dependent manner and conferred the inheritance of the methylation states.

In 1997, Xu & Bestor developed the targeted methylation approach, a method in which a DNA MTase is fused to a DNA binding domain that targets the fusion protein to specific genomic loci and results in the introduction of DNA methylation at these sites (Xu and Bestor 1997). Later, different programable DNA targeting domains, relying on C2H2 zinc fingers, the TAL effector arrays (TALE) or catalytically inactive CRISPR-Cas9 nuclease were fused to an epigenetic writer or reader domains have been employed to synthetically reprogram epigenetic and transcriptional changes in mammalian systems (Kungulovski and Jeltsch 2016; Jurkowska and Jurkowski 2019; Rots and Jeltsch 2018).

1.14 Conclusions and Outlook

Although DNA MTases, the enzymes that introduce methylation into DNA, have been intensively studied, the interest in these enzymes has remained high over many years (Fig. 1.6). This is due to an ever-growing importance of DNA

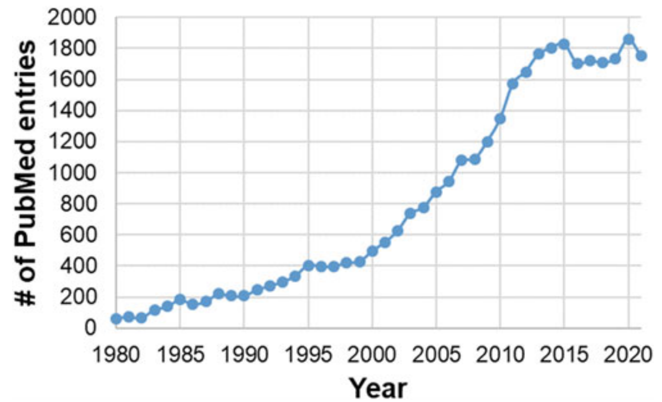


Fig. 1.6 Number of PubMed entries with the term "DNA" and "Methyltransferase" in title or abstract (as of December 2021)

methylation as an epigenetic modification in organismic development and human diseases. Despite decades of active research in the fields of DNA methylation and DNA methyltransferases, and progress in mapping the methylation landscapes at high resolution, many exciting questions still await answers and future challenges extend from our current level of knowledge. How is DNA methylation (and epigenetic information in general) deposited during organismic development and how is it maintained and altered if needed? How are DNA MTases regulated and targeted to achieve these goals? How does DNA methylation interact with other epigenetic systems in mammals, lower eukaryotes, and even bacteria? Will it be soon possible to develop epigenetic antibacterial drugs, addressing processes like phase variation or drug resistance? How can we make use of epigenetic editing, including targeted DNA methylation or demethylation, to repair aberrant, disease-causing epigenetic states and combat diseases like cancer? Which biological function (s) have the oxidized forms of 5-methylcytosine? Is there an active demethylation of N6-methyladenine in the cell? What is the role of this modification in other higher organisms? How can we use DNA methyltransferases in artificial epigenetic circuits more efficiently? We anticipate many more years of exciting research to come in the field of DNA methylation and the study of DNA MTases is an integral objective in this development.

References

Adam S, Anteneh H, Hornisch M, Wagner V, Lu J, Radde NE et al (2020) DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. Nat Commun 11(1): 3723. https://doi.org/10.1038/s41467-020-17531-8

Arber W, Dussoix D (1962) Host specificity of DNA produced by Escherichia coli. I. Host controlled modification of bacteriophage lambda. J Mol Biol 5:18–36

Arber W, Linn S (1969) DNA modification and restriction. Annu Rev Biochem 38:467–500. https://doi.org/10.1146/annurev.bi.38.070169.002343

Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Herault Y et al (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354(6314):909–912. https://doi.org/10.1126/science.aah5143

Baylin SB (2012) The cancer epigenome: its origins, contributions to tumorigenesis, and translational implications. Proc Am Thorac Soc 9(2):64–65. https://doi.org/10.1513/pats.201201-001MS

Baylin SB, Hoppener JW, de Bustros A, Steenbergh PH, Lips CJ, Nelkin BD (1986) DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas. Cancer Res 46(6):2917–2922

Bergman Y, Cedar H (2013) DNA methylation dynamics in health and disease. Nat Struct Mol Biol 20(3): 274–281. https://doi.org/10.1038/nsmb.2518

Bestor T, Laudano A, Mattaliano R, Ingram V (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol 203(4):971–983

Bhaya D, Davison M, Barrangou R (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet 45:273–297. https://doi.org/10.1146/annurev-genet-110410-132430

Bird AP (1980) DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res 8(7): 1499–1504

Bird A, Taggart M, Frommer M, Miller OJ, Macleod D (1985) A fraction of the mouse genome that is derived from islands of nonmethylated. CpG-rich DNA Cell 40(1):91–99

Booth MJ, Ost TW, Beraldi D, Bell NM, Branco MR, Reik W et al (2013) Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. Nat Protoc 8(10):1841–1851. https://doi.org/10.1038/nprot.2013.115

Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317(5845):1760–1764. https://doi.org/10.1126/science.1147939

Bourc'his D, Bestor TH (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature 431(7004):96–99. https://doi.org/10.1038/nature02886

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints. Science 294(5551):2536–2539. https://doi.org/10.1126/science.1065848

Boyer HW (1971) DNA restriction and modification mechanisms in bacteria. Annu Rev Microbiol 25: 153–176. https://doi.org/10.1146/annurev.mi.25.100171.001101

Browne MJ, Turnbull JF, McKay EL, Adams RL, Burdon RH (1977) The sequence specificity of a mammalian DNA methylase. Nucleic Acids Res 4(4):1039–1045

Bujnicki JM (2002) Sequence permutations in the molecular evolution of DNA methyltransferases. BMC Evol Biol 2:3

Casadesus J, Low D (2006) Epigenetic gene regulation in the bacterial world. Microbiol Mol Biol Rev 70(3): 830–856. https://doi.org/10.1128/MMBR.00016-06

Cheng X (1995) Structure and function of DNA methyltransferases. Annu Rev Biophys Biomol Struct 24:293–318. https://doi.org/10.1146/annurev.bb.24.060195.001453

Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ (1993) Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell 74(2):299–307

Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science 277(5334):1996–2000

Clark SJ, Harrison J, Paul CL, Frommer M (1994) High sensitivity mapping of methylated cytosines. Nucleic Acids Res 22(15):2990–2997

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD et al (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452(7184):215–219. https://doi.org/10.1038/nature06745

Cross SH, Charlton JA, Nan X, Bird AP (1994) Purification of CpG islands using a methylated DNA binding column. Nat Genet 6(3):236–244. https://doi.org/10.1038/ng0394-236

de Mendoza A, Poppe D, Buckberry S, Pflueger J, Albertin CB, Daish T et al (2021) The emergence of the brain non-CpG methylation system in vertebrates. Nat Ecol Evol 5(3):369–378. https://doi.org/10.1038/s41559-020-01371-2

Deplus R, Blanchon L, Rajavelu A, Boukaba A, Defrance M, Luciani J et al (2014) Regulation of DNA methylation patterns by CK2-mediated phosphorylation of Dnmt3a. Cell Rep 8(3):743–753. https://doi.org/10.1016/j.celrep.2014.06.048

Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S et al (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem 285(34): 26114–26120. https://doi.org/10.1074/jbc.M109.089433

Dunn DB, Smith JD (1955) Occurrence of a new base in the deoxyribonucleic acid of a strain of bacterium coli. Nature 175(4451):336–337

Ehrlich M, Wang RY (1981) 5-Methylcytosine in eukaryotic DNA. Science 212(4501):1350–1357

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403(6767):335–338. https://doi.org/10.1038/35002125

Esteve PO, Chang Y, Samaranayake M, Upadhyay AK, Horton JR, Feehery GR et al (2011) A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. Nat Struct Mol Biol 18(1):42–48. https://doi.org/10.1038/nsmb.1939

Fahy J, Jeltsch A, Arimondo PB (2012) DNA methyltransferase inhibitors in cancer: a chemical and therapeutic patent overview and selected clinical studies. Expert Opin Ther Pat 22(12):1427–1442. https://doi.org/10.1517/13543776.2012.729579

Feinberg AP, Vogelstein B (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 301(5895):89–92

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA et al (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 7(6):461–465. https://doi.org/10.1038/nmeth.1459

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW et al (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A 89(5):1827–1831

Gama-Sosa MA, Slagel VA, Trewyn RW, Oxenhandler R, Kuo KC, Gehrke CW et al (1983) The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res 11(19):6883–6894

Gao L, Emperle M, Guo Y, Grimm SA, Ren W, Adam S et al (2020) Comprehensive structure-function

characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. Nat Commun 11(1):3355. https://doi.org/10.1038/s41467-020-17109-4

Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403(6767):339–342. https://doi.org/10.1038/35002131

Goedecke K, Pignot M, Goody RS, Scheidig AJ, Weinhold E (2001) Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. Nat Struct Biol 8(2):121–125. https://doi.org/10.1038/84104

Gold M, Hurwitz J (1964a) The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. V. Purification and properties of the deoxyribonucleic acid-Methylating activity of Escherichia Coli. J Biol Chem 239:3858–3865

Gold M, Hurwitz J (1964b) The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. Vi. Further studies on the properties of the deoxyribonucleic acid methylation reaction. J Biol Chem 239:3866–3874

Gold M, Hurwitz J, Anders M (1963) The enzymatic methylation of Rna and DNA, ii. On the species specificity of the methylation enzymes. Proc Natl Acad Sci U S A 50(1):164–169

Gold M, Hausmann R, Maitra U, Hurwitz J (1964) The enzymatic methylation of Rna and DNA. 8. Effects of bacteriophage infection on the activity of the methylating enzymes. Proc Nat Acad Sci U S A 52: 292–297

Greger V, Passarge E, Hopping W, Messmer E, Horsthemke B (1989) Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. Hum Genet 83(2):155–158

Gruenbaum Y, Naveh-Many T, Cedar H, Razin A (1981) Sequence specificity of methylation in higher plant DNA. Nature 292(5826):860–862

Gruenbaum Y, Cedar H, Razin A (1982) Substrate and sequence specificity of a eukaryotic DNA methylase. Nature 295(5850):620–622

Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X et al (2015) Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517(7536):640–644. https://doi.org/10.1038/nature13899

Guschlbauer W (1988) The DNA and S-adenosylmethionine-binding regions of EcoDam and related methyltransferases. Gene 74(1):211–214

Hamidi T, Singh AK, Chen T (2015) Genetic alterations of DNA methylation machinery in human diseases. Epigenomics 7(2):247–265. https://doi.org/10.2217/epi.14.80

Hata K, Okano M, Lei H, Li E (2002) Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development 129(8):1983–1993

He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q et al (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

Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A (1999) The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature 401(6750):301–304. https://doi.org/10.1038/45843

Hennecke F, Kolmar H, Brundl K, Fritz HJ (1991) The vsr gene product of E. coli K-12 is a strand- and sequence-specific DNA mismatch endonuclease. Nature 353(6346):776–778. https://doi.org/10.1038/353776a0

Herman GE, Modrich P (1982) Escherichia coli dam methylase. Physical and catalytic properties of the homogeneous enzyme. J Biol Chem 257(5):2605–2612

Holliday R, Pugh JE (1975) DNA modification mechanisms and gene activity during development. Science 187(4173):226–232

Horton JR, Liebert K, Bekes M, Jeltsch A, Cheng X (2006) Structure and substrate recognition of the Escherichia coli DNA adenine methyltransferase. J Mol Biol 358(2):559–570. https://doi.org/10.1016/j.jmb.2006.02.028

Hotchkiss RD (1948) The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem 175(1):315–332

Hurwitz J, Gold M, Anders M (1964a) The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. 3. Purification of soluble ribonucleic acid-Methylating enzymes. J Biol Chem 239:3462–3473

Hurwitz J, Gold M, Anders M (1964b) The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. Iv. The properties of the soluble ribonucleic acid-Methylating enzymes. J Biol Chem 239:3474–3482

Ingrosso D, Fowler AV, Bleibaum J, Clarke S (1989) Sequence of the D-aspartyl/L-isoaspartyl protein methyltransferase from human erythrocytes. Common sequence motifs for protein, DNA, RNA, and small molecule S-adenosylmethionine-dependent methyltransferases. J Biol Chem 264(33): 20131–20139

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA et al (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

Jain D, Meydan C, Lange J, Claeys Bouuaert C, Lailler N, Mason CE et al (2017) Rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. PLoS Genet 13(8):e1006964. https://doi.org/10.1371/journal.pgen.1006964

Janulaitis A, Klimasauskas S, Petrusyte M, Butkus V (1983) Cytosine modification in DNA by BcnI methylase yields N4-methylcytosine. FEBS Lett 161(1): 131–134

Jeltsch A (1999) Circular permutations in the molecular evolution of DNA methyltransferases. J Mol Evol 49(1):161–164

Jeltsch A (2001) The cytosine N4-methyltransferase M. PvuII also modifies adenine residues. Biol Chem 382(4):707–710. https://doi.org/10.1515/BC.2001.084
Jeltsch A (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chembiochem: a European journal of chemical biology. 3(4):274–293
Jeltsch A (2010) Molecular biology. Phylogeny of methylomes Science 328(5980):837–838. https://doi.org/10.1126/science.1190738
Jeltsch A (2013) Oxygen, epigenetic signaling, and the evolution of early life. Trends Biochem Sci 38(4):172–176. https://doi.org/10.1016/j.tibs.2013.02.001
Jeltsch A, Jurkowska RZ (2014) New concepts in DNA methylation. Trends Biochem Sci 39(7):310–318. https://doi.org/10.1016/j.tibs.2014.05.002
Jeltsch A, Christ F, Fatemi M, Roth M (1999) On the substrate specificity of DNA methyltransferases. Adenine-N6 DNA methyltransferases also modify cytosine residues at position N4. J Biol Chem 274(28):19538–19544
Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449(7159):248–251. https://doi.org/10.1038/nature06146
Jones PA, Taylor SM (1980) Cellular differentiation, cytidine analogs and DNA methylation. Cell 20(1):85–93
Jurkowska RZ, Jurkowski TP (2019) Establishment, erasure and synthetic reprogramming of DNA methylation in mammalian cells. In: Jurga S, Barciszewski J (eds) The DNA, RNA, and histone Methylomes. Springer International Publishing, Cham, pp 1–26
Jurkowska RZ, Jurkowski TP, Jeltsch A (2011) Structure and function of mammalian DNA methyltransferases. Chembiochem: a European journal of chemical biology 12(2):206–222. https://doi.org/10.1002/cbic.201000195
Kagan RM, Clarke S (1994) Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. Arch Biochem Biophys 310(2):417–427. https://doi.org/10.1006/abbi.1994.1187
Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E et al (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature 429(6994):900–903. https://doi.org/10.1038/nature02633
Klimasauskas S, Timinskas A, Menkevicius S, Butkiene D, Butkus V, Janulaitis A (1989) Sequence motifs characteristic of DNA[cytosine-N4] methyltransferases: similarity to adenine and cytosine-C5 DNA-methylases. Nucleic Acids Res 17(23):9823–9832
Klimasauskas S, Kumar S, Roberts RJ, Cheng X (1994) HhaI methyltransferase flips its target base out of the DNA helix. Cell 76(2):357–369
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
Kungulovski G, Jeltsch A (2016) Epigenome editing: state of the art, concepts, and perspectives. Trends Genet: TIG 32(2):101–113. https://doi.org/10.1016/j.tig.2015.12.001
Labahn J, Granzin J, Schluckebier G, Robinson DP, Jack WE, Schildkraut I et al (1994) Three-dimensional structure of the adenine-specific DNA methyltransferase M.Taq I in complex with the cofactor S-adenosylmethionine. Proc Natl Acad Sci U S A 91(23):10957–10961
Laszlo AH, Derrington IM, Brinkerhoff H, Langford KW, Nova IC, Samson JM et al (2013) Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. Proc Natl Acad Sci U S A 110(47):18904–18909. https://doi.org/10.1073/pnas.1310240110
Lauster R, Kriebardis A, Guschlbauer W (1987) The GATATC-modification enzyme EcoRV is closely related to the GATC-recognizing methyltransferases DpnII and dam from E. coli and phage T4. FEBS Lett 220(1):167–176
Lauster R, Trautner TA, Noyer-Weidner M (1989) Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. J Mol Biol 206(2):305–312
Lee I, Razaghi R, Gilpatrick T, Molnar M, Gershman A, Sadowski N et al (2020) Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. Nat Methods 17(12):1191–1199. https://doi.org/10.1038/s41592-020-01000-7
Lei H, Oh SP, Okano M, Juttermann R, Goss KA, Jaenisch R et al (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development 122(10):3195–3205
Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69(6):915–926
Lin CC, Chen YP, Yang WZ, Shen JCK, Yuan HS (2020) Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. Nucleic Acids Res 48(7):3949–3961. https://doi.org/10.1093/nar/gkaa111
Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH et al (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133(3):523–536. https://doi.org/10.1016/j.cell.2008.03.029
Lobner-Olesen A, Skovgaard O, Marinus MG (2005) Dam methylation: coordinating cellular processes. Curr Opin Microbiol 8(2):154–160. https://doi.org/10.1016/j.mib.2005.02.009
Low DA, Casadesus J (2008) Clocks and switches: bacterial gene regulation by DNA adenine methylation. Curr

Opin Microbiol 11(2):106–112. https://doi.org/10.1016/j.mib.2008.02.012
Mahdavi-Amiri Y, Chung KCK, Hili R (2021) Single-nucleotide resolution of N-6-adenine methylation sites in DNA and RNA by nitrite sequencing. Chem Sci 12(2):606–612. https://doi.org/10.1039/d0sc03509b
Maier JAH, Möhrle R, Jeltsch A (2017) Design of synthetic epigenetic circuits featuring memory effects and reversible switching based on DNA methylation. Nat Commun 8:15336. https://doi.org/10.1038/ncomms15336
Maiti A, Drohat AC (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J Biol Chem 286(41):35334–35338. https://doi.org/10.1074/jbc.C111.284620
Malone T, Blumenthal RM, Cheng X (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. J Mol Biol 253(4):618–632. https://doi.org/10.1006/jmbi.1995.0577
Marinus MG, Casadesus J (2009) Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. FEMS Microbiol Rev 33(3):488–503. https://doi.org/10.1111/j.1574-6976.2008.00159.x
Marinus MG, Morris NR (1973) Isolation of deoxyribonucleic acid methylase mutants of Escherichia coli K-12. J Bacteriol 114(3):1143–1150
Martin JL, McMillan FM (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Curr Opin Struct Biol 12(6):783–793
Matic I, Taddei F, Radman M (1996) Genetic barriers among bacteria. Trends Microbiol 4(2):69–72. https://doi.org/10.1016/0966-842X(96)81514-9
McIntyre ABR, Alexander N, Grigorev K, Bezdan D, Sichtig H, Chiu CY et al (2019) Single-molecule sequencing detection of N6-methyladenine in microbial reference materials. Nat Commun 10(1):579. https://doi.org/10.1038/s41467-019-08289-9
Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A et al (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454(7205):766–770. https://doi.org/10.1038/nature07107
Meselson M, Yuan R, Heywood J (1972) Restriction and modification of DNA. Annu Rev Biochem 41:447–466. https://doi.org/10.1146/annurev.bi.41.070172.002311
Modrich P (1982) Studies on sequence recognition by type II restriction and modification enzymes. CRC Crit Rev Biochem 13(3):287–323
Munzel M, Globisch D, Bruckl T, Wagner M, Welzmiller V, Michalakis S et al (2010) Quantification of the sixth DNA base hydroxymethylcytosine in the brain. Angew Chem Int Ed Engl 49(31):5375–5377. https://doi.org/10.1002/anie.201002033
Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K et al (2013) Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature 502(7470):249–253. https://doi.org/10.1038/nature12488
Okano M, Xie S, Li E (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet 19(3):219–220. https://doi.org/10.1038/890
Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99(3):247–257
Olivenza DR, Nicoloff H, Sanchez-Romero MA, Cota I, Andersson DI, Casadesus J (2019) A portable epigenetic switch for bistable gene expression in bacteria. Sci Rep 9(1):11261. https://doi.org/10.1038/s41598-019-47650-2
Ooi SK, Bestor TH (2008) The colorful history of active DNA demethylation. Cell 133(7):1145–1148. https://doi.org/10.1016/j.cell.2008.06.009
Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z et al (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448(7154):714–717. https://doi.org/10.1038/nature05987
Park M, Patel N, Keung AJ, Khalil AS (2019) Engineering epigenetic regulation using synthetic read-write modules. Cell 176(1–2):227–38 e20. https://doi.org/10.1016/j.cell.2018.11.002
Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA et al (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew Chem Int Ed Engl 50(31):7008–7012. https://doi.org/10.1002/anie.201103899
Pingoud A, Jeltsch A (1997) Recognition and cleavage of DNA by type-II restriction endonucleases. Euro J Biochem/FEBS 246(1):1–22
Pingoud A, Jeltsch A (2001) Structure and function of type II restriction endonucleases. Nucleic Acids Res 29(18):3705–3727. https://doi.org/10.1093/nar/29.18.3705
Pingoud A, Wilson GG, Wende W (2014) Type II restriction endonucleases--a historical perspective and more. Nucleic Acids Res 42(12):7489–7527. https://doi.org/10.1093/nar/gku447
Pogolotti AL Jr, Ono A, Subramaniam R, Santi DV (1988) On the mechanism of DNA-adenine methylase. J Biol Chem 263(16):7461–7464
Posfai J, Bhagwat AS, Posfai G, Roberts RJ (1989) Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Res 17(7):2421–2435
Razin A, Riggs AD (1980) DNA methylation and gene function. Science 210(4470):604–610
Reich NO, Mashhoon N (1990) Inhibition of EcoRI DNA methylase with cofactor analogs. J Biol Chem 265(15):8966–8970

Ren W, Fan H, Grimm SA, Guo Y, Kim JJ, Yin J et al (2020) Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. Proc Natl Acad Sci U S A 117(31): 18439–18447. https://doi.org/10.1073/pnas.2009316117

Ren W, Fan H, Grimm SA, Kim JJ, Li L, Guo Y et al (2021) DNMT1 reads heterochromatic H4K20me3 to reinforce LINE-1 DNA methylation. Nat Commun 12(1):2490. https://doi.org/10.1038/s41467-021-22665-4

Riggs AD (1975) X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet 14(1):9–25

Roberts RJ (1995) On base flipping. Cell 82(1):9–12

Roberts RJ, Cheng X (1998) Base flipping. Annu Rev Biochem 67:181–198. https://doi.org/10.1146/annurev.biochem.67.1.181

Rosic S, Amouroux R, Requena CE, Gomes A, Emperle M, Beltran T et al (2018) Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. Nat Genet 50(3):452–459. https://doi.org/10.1038/s41588-018-0061-8

Roth M, Helm-Kruse S, Friedrich T, Jeltsch A (1998) Functional roles of conserved amino acid residues in DNA methyltransferases investigated by Site-directed mutagenesis of the Eco RV Adenine-N6-methyltransferase. Journal of Biological Chemistry 273(28):17333–17342. https://doi.org/10.1074/jbc.273.28.17333. S0021925818804926

Rots MG, Jeltsch A (2018) Editing the epigenome: overview, open questions, and directions of future development. Methods Mol Biol 1767:3–18. https://doi.org/10.1007/978-1-4939-7774-1_1

Sakamoto Y, Sereewattanawoot S, Suzuki A (2020) A new era of long-read sequencing for cancer genomics. J Hum Genet 65(1):3–10. https://doi.org/10.1038/s10038-019-0658-5

Sanchez-Romero MA, Casadesus J (2020) The bacterial epigenome. Nat Rev Microbiol 18(1):7–20. https://doi.org/10.1038/s41579-019-0286-2

Sanchez-Romero MA, Cota I, Casadesus J (2015) DNA methylation in bacteria: from the methyl group to the methylome. Curr Opin Microbiol 25:9–16. https://doi.org/10.1016/j.mib.2015.03.004

Santi DV, Norment A, Garrett CE (1984) Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. Proc Natl Acad Sci U S A 81(22):6993–6997

Schluckebier G, O'Gara M, Saenger W, Cheng X (1995) Universal catalytic domain structure of AdoMet-dependent methyltransferases. J Mol Biol 247(1): 16–20. https://doi.org/10.1006/jmbi.1994.0117

Seeman NC, Rosenberg JM, Rich A (1976) Sequence-specific recognition of double helical nucleic acids by proteins. Proc Natl Acad Sci U S A 73(3):804–808

Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA et al (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 450(7171):908–912. https://doi.org/10.1038/nature06397

Song J, Pfeifer GP (2016) Are there specific readers of oxidized 5-methylcytosine bases? BioEssays 38(10): 1038–1047. https://doi.org/10.1002/bies.201600126

Song J, Rechkoblit O, Bestor TH, Patel DJ (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science 331(6020):1036–1040. https://doi.org/10.1126/science.1195380

Song J, Teplova M, Ishibe-Murakami S, Patel DJ (2012) Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science 335(6069):709–712. https://doi.org/10.1126/science.1214453

Song G, Wang G, Luo X, Cheng Y, Song Q, Wan J et al (2021) An all-to-all approach to the identification of sequence-specific readers for epigenetic DNA modifications on cytosine. Nat Commun 12(1):795. https://doi.org/10.1038/s41467-021-20950-w

Swartz MN, Trautner TA, Kornberg A (1962) Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. J Biol Chem 237:1961–1967

Szomolanyi E, Kiss A, Venetianer P (1980) Cloning the modification methylase gene of Bacillus sphaericus R in Escherichia coli. Gene 10(3):219–225

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y et al (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

Takeshita K, Suetake I, Yamashita E, Suga M, Narita H, Nakagawa A et al (2011) Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proc Natl Acad Sci U S A 108(22): 9055–9059. https://doi.org/10.1073/pnas.1019629108

Tawfik DS, Griffiths AD (1998) Man-made cell-like compartments for molecular evolution. Nat Biotechnol 16(7):652–656. https://doi.org/10.1038/nbt0798-652

Tazi J, Bird A (1990) Alternative chromatin structure at CpG islands. Cell 60(6):909–920

Ullrich T, Weirich S, Jeltsch A (2020) Development of an epigenetic tetracycline sensor system based on DNA methylation. PLoS One 15(5):e0232701. https://doi.org/10.1371/journal.pone.0232701

van der Woude MW (2011) Phase variation: how to create and coordinate population diversity. Curr Opin Microbiol 14(2):205–211. https://doi.org/10.1016/j.mib.2011.01.002

Weinberg DN, Rosenbaum P, Chen X, Barrows D, Horth C, Marunde MR et al (2021) Two competing mechanisms of DNMT3A recruitment regulate the dynamics of de novo DNA methylation at PRC1-targeted CpG islands. Nat Genet 53(6):794–800. https://doi.org/10.1038/s41588-021-00856-5

Wilson GG, Murray NE (1991) Restriction and modification systems. Annu Rev Genet 25:585–627. https://doi.org/10.1146/annurev.ge.25.120191.003101

Wion D, Casadesus J (2006) N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. Nat Rev Microbiol 4(3):183–192. https://doi.org/10.1038/nrmicro1350

Wu JC, Santi DV (1985) On the mechanism and inhibition of DNA cytosine methyltransferases. Prog Clin Biol Res 198:119–129

Wu JC, Santi DV (1987) Kinetic and catalytic mechanism of HhaI methyltransferase. J Biol Chem 262(10)4778–4786

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

Xu GL, Bestor TH (1997) Cytosine methylation targetted to pre-determined sequences. Nat Genet 17(4): 376–378. https://doi.org/10.1038/ng1297-376

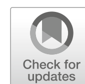
Yamashita Y, Yuan J, Suetake I, Suzuki H, Ishikawa Y, Choi YL et al (2010) Array-based genomic resequencing of human leukemia. Oncogene 29(25): 3723–3731. https://doi.org/10.1038/onc.2010.117

Yang X, Lay F, Han H, Jones PA (2010) Targeting DNA methylation for epigenetic therapy. Trends Pharmacol Sci 31(11):536–546. https://doi.org/10.1016/j.tips.2010.08.001

Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S et al (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 356(6337):eaaj2239. https://doi.org/10.1126/science.aaj2239

Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I et al (2010) Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res 38(13):4246–4253. https://doi.org/10.1093/nar/gkq147

Zhang ZM, Lu R, Wang P, Yu Y, Chen D, Gao L et al (2018) Structural basis for DNMT3A-mediated de novo DNA methylation. Nature 554(7692):387–391. https://doi.org/10.1038/nature25477



2 DNA Methylation in Prokaryotes

Josep Casadesús and María A. Sánchez-Romero

Abstract

The genomes of bacteria, archaea, and phage contain small amounts of C^5-methylcytosine, N^4-methylcytosine, and N^6-methyladenine. Base methylation takes place after DNA replication and is catalyzed by DNA methyltransferases that recognize specific target sequences. Prokaryotic DNA methyltransferases can be classified into two main types: (1) belonging to restriction-modification systems and (2) solitary (or "orphan") enzymes that lack a restriction enzyme partner. All known roles of DNA methylation involve control of interactions between DNA-binding proteins and their cognate sites. Such roles include protection from DNA restriction, strand discrimination during mismatch repair, cell cycle control, and regulation of transcription. DNA methylation often affects the interaction of bacterial pathogens with their hosts, raising the possibility of epigenetic therapies for infectious diseases.

Keywords

DNA methylation · methylome · Restriction-modification systems · methyltransferases

Abbreviations

CcrM	Cell-cycle regulated DNA methyltransferase
Mod	Modification protein in restriction-modification systems
oriC	Origin of replication in the *E. coli* chromosome
R	Purine moiety (A or G) in a nucleotide sequence
SAM	S-adenosyl-L-methionine
SMRT	Single molecule real time
SPI-1	Salmonella pathogenicity island 1
UAS	Upstream activating sequence
Y	Pyrimidine moiety (C or T) in a nucleotide sequence

2.1 Introduction

The presence of C^5-methylcytosine in bacterial DNA was discovered in 1948 (Hotchkiss 1948). In the following decades, N^6-methyladenine (Dunn and Smith 1955) and N^4-methylcytosine (Janulaitis et al. 1983) were found in prokaryotic genomes. The introduction of methyl groups into nucleobases is performed by DNA

J. Casadesús
Departamento de Genética, Facultad de Biología, Universidad de Sevilla, Seville, Spain
e-mail: casadesus@us.es

M. A. Sánchez-Romero (✉)
Departamento de Microbiología y Parasitología, Facultad de Farmacia, Universidad de Sevilla, Sevilla, Spain
e-mail: mtsanchez@us.es

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_2

methyltransferases that recognize specific nucleotide targets and use S-adenosyl-methionine (SAM) as a methyl group donor. Base methylation is postreplicative, and the chemistry of the methylation reaction is known in exquisite detail (Cheng 1995; Horton et al. 2006; Malone et al. 1995; Jeltsch et al. 1999; Jeltsch 2002; Horton et al. 2019). Prokaryotic DNA methyltransferases are active on nonmethylated DNA substrates (de novo methylation) and on hemimethylated substrates produced by DNA replication (maintenance methylation) (M.G. Marinus 1996; Jeltsch 2002; Wion and Casadesus 2006). Maintenance methylation transmits to daughter DNA molecules (and therefore to the offspring) the DNA methylation pattern of the mother cell.

In the 1960s, base methylation was found to be associated with the possession of restriction-modification systems that protect bacteria against phages and other invading DNAs (Arber and Linn 1969). Classical restriction-modification systems are currently classified into four main types on the basis of structural features, pattern of DNA cleavage, and cofactor requirements (Tock and Dryden 2005). In types I and III, the adenine or cytosine methyltransferase is part of a multisubunit enzyme involved in both restriction and modification. In contrast, type II systems consist of two separate enzymes, a restriction endonuclease and an adenine or cytosine methyltransferase. In addition to these three types of systems in which DNA methylation protects against endonucleolytic cleavage, restriction systems that cleave methylated DNA (type IV) have been described (Tock and Dryden 2005).

Under the classical paradigm of DNA methylation as a genome defense mechanism, solitary DNA methyltransferases (e.g., Dam in γ-proteobacteria and CcrM in α-proteobacteria) were considered exceptions. However, current evidence indicates that solitary DNA methyltransferases are common in the genomes of bacteria (Blow et al. 2016), archaea (Harris and Goldman 2020), and bacteriophages (Murphy et al. 2013). This abundance is mainly a consequence of horizontal gene transfer at high rates (Oliveira et al. 2014; Anton and Roberts 2021). However, bacterial genomes also contain solitary DNA methyltransferases evolved from ancestral restriction-modification systems that have lost their restriction enzyme component (Fox et al. 2007a; Blow et al. 2016).

As in eukaryotes, a major role of DNA methylation in bacterial physiology is transcriptional control. Under this paradigm, epigenetic regulation of gene expression was traditionally viewed as a task of solitary DNA methyltransferases and a consequence of long co-evolution that had adapted the epigenome to physiological needs. This view has been overturned in the last decade, and current evidence supports a radically different scenario: that any DNA methyltransferase, either solitary or belonging to a restriction-modification system, can regulate transcription if it happens to methylate a DNA target at a promoter or at a region involved in transcriptional control (Casadesus and Low 2006; Vasu and Nagaraja 2013; Phillips et al. 2019; Sanchez-Romero and Casadesus 2020; Anton and Roberts 2021). In fact, a horizontally acquired DNA methyltransferase can immediately overtake the control of gene expression, triggering a saltational evolutionary event (Chao et al. 2015). Another example of saltational evolution triggered by DNA methylation has been described in *Mycobacterium abscessus*, where acquisition of DNA methyltransferases through horizontal gene transfer increases the pathogenic potential of environmental clones (Bryant et al. 2021).

The study of bacterial DNA methylation was slowed down for decades because of the difficulty to detect methylated bases. One cause of this hurdle was that the bisulfite treatment widely used in eukaryotes to detect C^5-methylcytosine had little utility in prokaryotes as most studies dealt with N^6-methyladenine (Casadesus and Low 2006; Wion and Casadesus 2006). This obstacle has been overcome by advances in nucleic acid technology that permit the detection of N^6-methyladenine and other base modifications as an integral part of the sequencing method. Two such methods, SMRT sequencing (Flusberg et al. 2010) and nanopore sequencing (Rand et al. 2017) can decipher the complete methylation pattern of a bacterial genome (the "methylome"). By identifying the

target(s) of a DNA methyltransferase and the spatial and temporal patterns of genome methylation, these technologies can provide crucial information to unravel the physiological roles of DNA methylation in a given species (Casadesus and Low 2006; Sanchez-Romero et al. 2015; Anton and Roberts 2021; Sanchez-Romero and Casadesus 2020). Furthermore, methylome analysis has extended the study of DNA methylation well beyond its traditional boundaries. For instance, some archaea have been found to harbor solitary DNA methyltransferases (Blow et al. 2016), raising the possibility that DNA methylation may play a role in gene expression or have other unknown functions in a biological kingdom where DNA methylation had remained virtually unknown (Harris and Goldman 2020).

Even though SMRT and nanopore sequencing have made bacterial DNA methylation a burgeoning field, two DNA adenine methyltransferases, CcrM in α-proteobacteria and Dam in γ-proteobacteria, remain paradigms with known roles in bacterial physiology (Casadesus and Torreblanca 1996; Casadesus and Low 2006; Adhikari and Curtis 2016; Mouammine and Collier 2018; Mohapatra et al. 2014b). The overall bias of this chapter toward N^6-methyladenine thus reflects the preponderant involvement of this DNA modification in bacterial epigenetics, and CcrM and Dam receive special attention.

2.2 CcrM Methylation

CcrM (acronym for "cell-cycle regulated methyltransferase") was initially identified in *Caulobacter crescentus,* a bacterial species that undergoes asymmetric cell division producing swarmer and stalked cells. CcrM catalyzes the transfer of a methyl group from SAM to the adenine moiety of 5′GANTC3′ targets, where N is any nucleotide (Stephens et al. 1996; Robertson et al. 2000; Kahng and Shapiro 2001; Mohapatra et al. 2014a; Mouammine and Collier 2018). CcrM is active both as a monomer and as a dimer and shows a slight preference for hemimethylated DNA substrates (Stephens et al. 1996; Robertson et al. 2000; Kahng and Shapiro 2001; Mohapatra et al. 2014a).

CcrM homologs are common in α-proteobacteria. Examples are *Agrobacterium tumefaciens*, the causative agent of crown gall disease in dicotyledonous plants (Kahng and Shapiro 2001), the nitrogen-fixing symbionts of legumes once known as *Rhizobium* and nowadays given other names (Wright et al. 1997; diCenzo et al. 2022) and the animal pathogen *Brucella abortus* (Robertson et al. 2000). A gene that encodes a CcrM homolog known as YhdJ is also found in the genomes of *E. coli* and *Salmonella* (Broadbent et al. 2007; Blow et al. 2016). YhdJ can methylate the 3′ adenosine moiety of 5′AGTCAT3′ targets in vitro. However, YhdJ does not seem to be produced in vivo, at least under laboratory conditions (Broadbent et al. 2007).

In α-proteobacteria, chromosome hemimethylation after the passage of the replication fork is long-lived, especially at loci located near the origin of replication, because CcrM synthesis is restricted to a late stage of chromosome replication, immediately before cell division (Stephens et al. 1996; Collier 2009). The burst of CcrM synthesis is brief because the enzyme is quickly degraded by the Lon protease (Wright et al. 1996). Different temporal patterns of CcrM degradation by the Lon protease are detected in swarmer and stalked cells (Zhou et al. 2019), suggesting that CcrM methylation may play a role in *Caulobacter* cellular dimorphism. CcrM methylation does not govern the initiation of chromosome replication and controls the *Caulobacter* cell cycle by mechanisms that remain incompletely understood (Kozdon et al. 2013; Fioravanti et al. 2013; Mohapatra et al. 2014a). It is conceivable that CcrM methylation may regulate the cell cycle, at least in part, by controlling the transcription of genes that encode cell cycle regulators (Fioravanti et al. 2013).

Hemimethylation of GANTC sites has been shown to activate and to repress transcription in *Caulobacter* (Gonzalez and Collier 2013; Gonzalez et al. 2014). Examples of activation by CcrM hemimethylation are the cell cycle regulatory gene *ctrA* (Mohapatra et al. 2014a;

Gonzalez et al. 2014) and the *creS* gene which encodes crescentin, a structural protein responsible for the curved shape of *Caulobacter* cells (Mohapatra et al. 2020). A noteworthy feature of methylation-dependent *creS* transcription is DNA strand discrimination, which permits transcription in only one of the hemimethylated chromosomes after replication (Mohapatra et al. 2020). This bias may cause differences in CreS synthesis among daughter cells, perhaps contributing to the distinct shapes of swarmer and stalked cells. In turn, transcriptional repression by hemimethylation has been described in the cell division genes *ftsZ* and *mipZ* (Gonzalez et al. 2014).

2.3 Dam Methylation

The Dam methyltransferase, initially characterized in *E. coli*, is present in many γ-proteobacteria (Marinus 1996). Dam is a monomer in solution, and introduces a methyl group into the N^6 position of the adenine residue in 5′GATC3′ sequences (Marinus 1996). Dam is both a de novo methyltransferase and a maintenance methyltransferase, and there is little difference in the rate of in vitro methylation between nonmethylated and hemimethylated DNA substrates (Herman and Modrich 1982). Dam contains two SAM binding sites; one is the catalytic site and the other increases specific binding to DNA, probably through allosteric regulation (Bergerat et al. 1991). Dam is a highly processive enzyme, and methylates about 55 GATC sites per binding event (Urig et al. 2002).

Synthesis of Dam methyltransferase is under transcriptional control. The *dam* gene contains five promoters (Løbner-Olesen et al. 1992), and one of them is regulated by the growth rate. This control adjusts the cellular level of Dam, which consists of about 130 molecules during exponential growth in rich medium. The amount of Dam methyltransferase may be additionally controlled by proteolysis (Calmann and Marinus 2003).

The functions of Dam methylation in bacterial physiology are especially well known in *E. coli* and *Salmonella* owing to decades of research (Marinus 1996; Løbner-Olesen et al. 2005; Boye et al. 1992; Wion and Casadesus 2006). At the turn of the century, the involvement of Dam methylation in *Salmonella* virulence (Garcia-Del Portillo et al. 1999; Heithoff et al. 1999) fostered the study of Dam methylation in other bacterial pathogens, and novel examples of Dam-dependent virulence were found (Heusipp et al. 2007; Giacomodonato et al. 2009; Marinus and Casadesus 2009; Adhikari and Curtis 2016; Sanchez-Romero and Casadesus 2020). Unknown roles of Dam methylation in bacterial physiology may remain to be found as Dam homologs appear to be widespread in the genomes of γ-proteobacteria (Balbontin et al. 2006; Oliveira et al. 2020; Oliveira and Fang 2021).

2.3.1 Role of Dam Methylation in DNA Mismatch Repair

Replication-generated DNA mismatches involve normal (non-damaged) nucleotides that cannot be replaced by base excision repair or by nucleotide excision repair. Furthermore, mismatch repair needs to discriminate between the error-free template strand and the error-prone daughter strand. *E. coli* and other γ-proteobacteria use Dam hemimethylation for this discrimination (Iyer et al. 2006) (Fig. 2.1). Mismatched base pairs are recognized by a protein known as MutS, which recruits two additional proteins, MutL and MutH. When the MutS-MutL-MutH complex is assembled at a DNA mismatch, MutH acquires endonuclease activity and cleaves the phosphodiester bond located at the 5′ side of the G in the closest GATC on the nonmethylated DNA strand (Iyer et al. 2006). After cleavage, the UvrD helicase dislodges MutH from the ternary complex and single strand degradation by exonucleases occurs. The resulting gap is filled by DNA polymerase III, and the nick is sealed by the formation of a phosphodiester bond by DNA ligase (Iyer et al. 2006). Finally, Dam methyltransferase converts the hemimethylated GATC to a fully methylated site (Fig. 2.1). Because MutH cannot cleave methylated DNA,

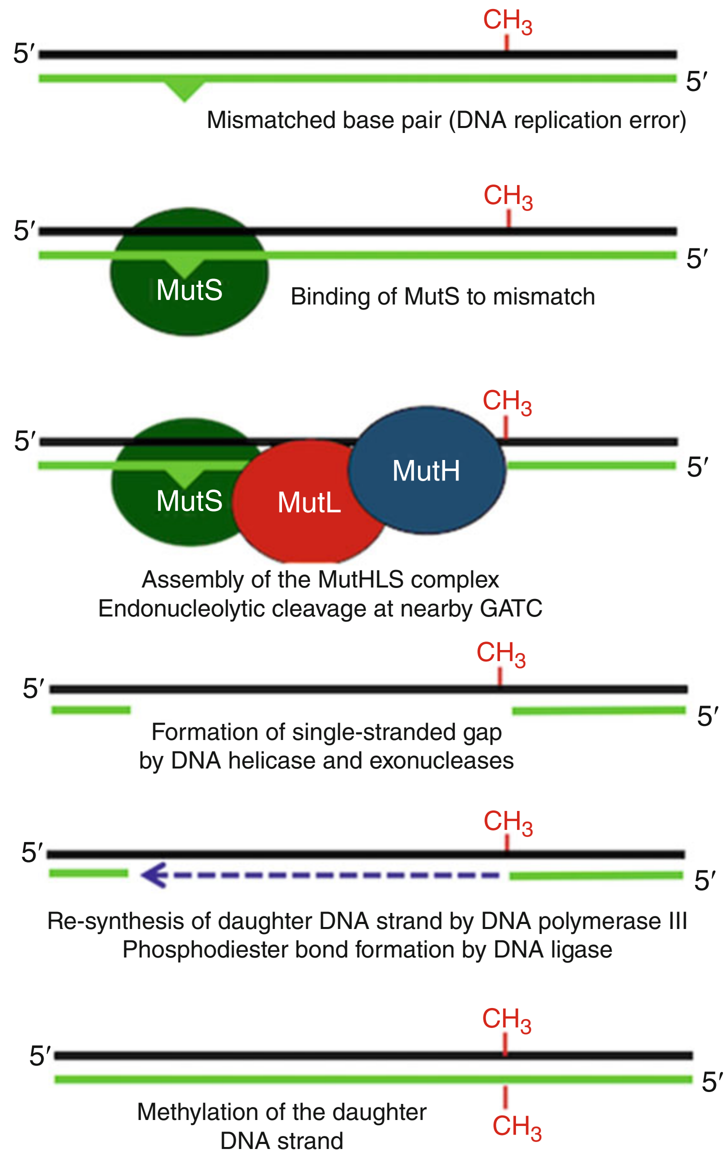


Fig. 2.1 Dam-directed repair of a DNA mismatch. Detection of the mismatch by MutS recruits MutH and MutL, and the MutHLS complex is assembled. MutH-mediated endonucleolytic cleavage of the newly synthesized DNA strand occurs at the nearest GATC. The nicked single strand is then degraded by exonucleases. The gap is filled by DNA polymerase III, and the nick is sealed by DNA ligase. Finally, the hemimethylated GATC is methylated by the Dam methyltransferase

mismatch repair is confined to a short hemimethylated region, estimated to be around 10 kb long, behind the replication fork. Transient lack of GATC methylation in the newly synthesized strand thus provides the signal for DNA strand discrimination (Pukkila et al. 1983).

Upon assembly of the MutH-MutL-MutS complex, MutH endonuclease can cleave nonmethylated GATC duplexes. As a consequence, *dam* mutants can suffer MutH-mediated cleavage in both DNA strands (Iyer et al. 2006). Double-strand DNA breakage explains several traits of enterobacterial *dam* mutants: (1) sensitivity to agents that induce DNA injuries recognized by MutS (Karran and Marinus 1982; Prieto et al. 2004); (2) dependence on homologous recombination and other DNA repair functions to cope with DNA damage (Wang and Smith 1986; Marinus 2000; Torreblanca and Casadesus 1996); and (3) permanent induction of the SOS response (Torreblanca and Casadesus 1996). Another consequence of the involvement of

DNA adenine methylation in mismatch repair is that *dam* mutants of *E. coli* and *S. enterica* show a hypermutable phenotype, indicative of the inability of *dam* mutants to repair DNA replication errors that introduce mismatched base pairs (Glickman et al. 1978). *E. coli* and *Salmonella* strains that overproduce Dam methyltransferase also show elevated mutation rates, higher than in *dam* mutants (Pukkila et al. 1983; Marinus et al. 1984; Torreblanca and Casadesus 1996). Albeit paradoxical at first sight, this observation underscores the relevance of transient hemimethylation as a signal for mismatch repair: excess Dam methyltransferase shortens the hemimethylation period in newly replicated DNA molecules, thus preventing strand discrimination. The need of precise amounts of Dam methyltransferase may explain the tight and complex control of *dam* gene expression (Løbner-Olesen et al. 1992).

2.3.2 Control of Chromosome Replication by Dam Methylation

Initiation of chromosome replication in *E. coli* requires binding of an ATP-bound form of the initiator protein DnaA to the replication origin (*oriC*), followed by separation of the two strands of the double helix and loading of the DnaB DNA helicase (Mott and Berger 2007). The density of GATC sites in the *oriC* region is roughly tenfold higher than the average in the *E. coli* chromosome (Marinus 1996), and DnaA binding is only possible if the GATCs located in the region are methylated; a hemimethylated origin is inactive (Messer et al. 1985).

DNA replication is not followed by immediate methylation of the *oriC*. Actually, the GATC sites within *oriC* remain hemimethylated for a substantial fraction of the cell cycle (Messer et al. 1985; Boye et al. 2000). Extension of the hemimethylation period is a consequence of *oriC* sequestration by a protein called SeqA, which excludes Dam methyltransferase from the *oriC* in the daughter chromosomes (Lu et al. 1994). As long as SeqA-mediated sequestration and concomitant hemimethylation persist, the *oriC* remains inactive and the start of a new replication cycle is delayed (Boye et al. 2000). Binding of SeqA to hemimethylated GATC sites behind the DNA replication fork also plays roles in the spatial organization of the nucleoid (Waldminghaus and Skarstad 2009) and in sister chromosome cohesion (Joshi et al. 2013).

2.3.3 Transcriptional Control by Dam Methylation

If a DNA methyltransferase target is embedded in a promoter or a regulatory region, its methylation state can modulate the binding of RNA polymerase or transcription factors, making transcription responsive to DNA methylation (Wion and Casadesus 2006; Casadesus and Low 2006; Low and Casadesus 2008). Dam methylation-dependent transcriptional controls can be classified into two main types: (i) clock-like controls that use the methylation state of DNA (methylation or hemimethylation) as a signal to couple gene expression to a specific stage of the cell cycle (Low and Casadesus 2008); (ii) switch-like controls that turn off and on gene expression, sometimes in a reversible manner, by the formation of specific DNA methylation patterns. The latter are combinations of methylated and nonmethylated sites reminiscent of the methylation patterns found in eukaryotic chromosomes (Low and Casadesus 2008).

A common procedure for the identification of genes under DNA methylation control is to search for genes that show an altered expression pattern in DNA methyltransferase mutant strains. Such gene expression changes, however, do not necessarily indicate that transcription is DNA methylation-sensitive. An example is found in the DNA-damage responsive SOS regulon of *E. coli* and *Salmonella*, which shows elevated expression in *dam* mutants (M.G. Marinus 1996; Marinus and Casadesus 2009). However, transcription of SOS genes is not under Dam methylation control, and elevated expression in *dam* mutants is a consequence of DNA double-strand breakage by the MutHLS mismatch repair system

(M.G. Marinus 1996; Marinus and Casadesus 2009). To confirm the direct dependence of transcription on DNA methylation, genetic or transcriptomic evidence must be followed by mutational analysis of the putative methyltransferase target(s). For instance, if the elimination of one or more GATC sites abolishes Dam-dependent control, one may tentatively conclude that transcription is Dam-dependent indeed (Camacho and Casadesus 2002). A potential problem of such tests is that site-directed mutagenesis can impair the binding site of a transcriptional regulator and/or disrupt promoter function. The most direct methods to identify a methylation-sensitive DNA–protein interaction are electrophoretic mobility shift analysis and/or DNAse I footprinting using methylated and nonmethylated DNA substrates (Camacho and Casadesus 2002). In certain cases, hemimethylated DNA substrates must be also used (Camacho and Casadesus 2005).

Bioinformatic searches to identify genes whose transcription is controlled by Dam methylation are more difficult than one might expect a priori. It is not always obvious where the search for relevant GATC sites must be performed, because Dam methylation can regulate a promoter from sites more than 100 bp away from the transcription start (A. Hernday et al. 2002; Casadesus and Low 2006; Jakomin et al. 2008; Broadbent et al. 2010; Cota et al. 2012; Sanchez-Romero and Casadesus 2020). In *E. coli*, for instance, the average distance between GATC neighbor sites is 214 base pairs (Henaut et al. 1996), with the obvious consequence that GATC sites at distances potentially relevant for transcriptional control are found at many promoters. To make bioinformatic prediction more uncertain, the presence of a GATC site at a seemingly critical position can be misleading, as exemplified by the *cre* gene of bacteriophage P1 (Sternberg et al. 1986) and the *carAB* operon of *E. coli* (Charlier et al. 1995) whose promoters contain GATC sites not involved in transcriptional control. An additional complication arises from the fact that GATC-less genes can be indirectly controlled by Dam methylation if their transcription is regulated by a factor under direct Dam methylation control. An example is found in *Salmonella* pathogenicity island 1 (SPI-1), a cluster of virulence genes whose expression is reduced in *dam* mutants (Balbontin et al. 2006; Lopez-Garrido and Casadesus 2010). Regulation of SPI-1 by Dam methylation is indirect and requires StdE and StdF, two proteins encoded by the Dam-dependent *std* operon (Lopez-Garrido and Casadesus 2012; Garcia-Pastor et al. 2018b).

2.3.3.1 Temporal Regulation of Gene Expression by Dam Methylation

RNA polymerase and certain transcription factors can discriminate DNA hemimethylation from full (double-stranded) methylation (Wion and Casadesus 2006; Casadesus and Low 2006; Low and Casadesus 2008; Marinus and Casadesus 2009). This discrimination can have physiological significance as DNA hemimethylation is indicative of active growth while two-strand DNA methylation indicates the absence of growth. Hemimethylation can either activate or repress gene expression but activation seems to be more common than repression (Casadesus and Low 2006; Wion and Casadesus 2006).

The insertion element IS10 provides a classical example of transcriptional activation by DNA hemimethylation. The promoter of the IS10 transposase gene contains a GATC site that overlaps the −10 module, and methylation of this site prevents transcriptional initiation, presumably by hindrance of RNA polymerase binding (Roberts et al. 1985). When the replication fork passes by the IS10 transposase promoter, the GATC site becomes hemimethylated, and hemimethylation permits transient transcription (Roberts et al. 1985). An additional feature of Dam-dependent control of the IS10 transposase promoter is strand-specificity: DNA replication generates two daughter IS10 elements that are identical except for their GATC hemimethylation pattern. However, transcription of the transposase gene is permitted in one of the hemimethylated IS10 species only (Roberts et al. 1985).

Another case of transcriptional activation by strand-specific DNA hemimethylation is found in the *traJ* gene of the *Salmonella* virulence

plasmid, a relative of the *E. coli* F sex factor (Camacho and Casadesus 2005). The *traJ* gene encodes a transcriptional activator, and its expression is controlled by multiple factors including Lrp, a global regulator of the bacterial cell (Camacho and Casadesus 2002). Lrp activates *traJ* transcription by binding two cognate sites upstream of the *traJ* promoter, and one site contains a GATC (Camacho and Casadesus 2002). Methylation of this GATC impairs Lrp binding and prevents *traJ* transcription (Fig. 2.2). When replication occurs and the GATC site becomes hemimethylated, Lrp binding activates the transcription of *traJ* in one of the daughter plasmid molecules (Camacho and Casadesus 2005). As in IS10, it is noteworthy that identical DNA molecules can acquire distinct epigenetic properties upon addition of a single methyl group (Low and Casadesus 2008).

In both IS10 and *traJ*, activation of transcription by DNA adenine hemimethylation may permit the production of potentially dangerous cell products during active growth only. Strand-specific DNA hemimethylation may further restrain the synthesis of such products. Indeed, a low amount of IS10 transposase may prevent multiple transposition events and/or other

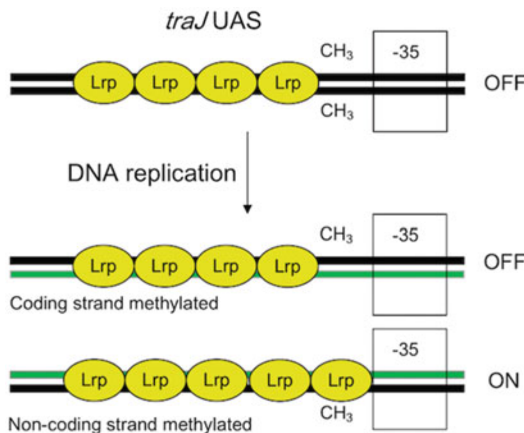


Fig. 2.2 Activation of *traJ* transcription by strand-specific Dam hemimethylation. Passage of the replication fork leaves the *traJ* UAS hemimethylated. The transcriptional activator Lrp binds the *traJ* UAS with different patterns depending on the DNA strand that contains a methyl group. Only one Lrp binding pattern permits transcription

transposase-mediated DNA rearrangements (Casadesus and Low 2006; Low and Casadesus 2008). Furthermore, coupling of transposition to DNA replication, a stage of the cell cycle in which two daughter chromosomes exist, may decrease the danger of lethal transposition (Casadesus and Low 2006). In the case of TraJ, plasmid replication may provide a signal of cellular welfare, and Lrp dependence may ensure that the physiological conditions are appropriate to build the conjugation apparatus and to undertake mating. TraJ synthesis in only one of the daughter plasmids may further relieve the burden of such an energy-consuming process.

An example of transcriptional repression by DNA adenine hemimethylation is found in the *dnaA* gene of *E. coli,* which maps near the origin of chromosome replication, *oriC*. One of the three *dnaA* promoters *(dnaA2)* contains three GATC sites and is only active if the GATCs are methylated (Braun and Wright 1986; Kucherer et al. 1986). After DNA replication, the GATC-rich *oriC-dnaA* region becomes hemimethylated and is sequestered by SeqA (Lu et al. 1994). Because sequestration prevents Dam methyltransferase activity, the *oriC-dnaA* region remains hemimethylated. Initiation of the following chromosome replication round will thus require SeqA release from the *dnaA* promoter and subsequent GATC methylation by the Dam methyltransferase to permit transient transcription of the *dnaA* gene (Waldminghaus and Skarstad 2009).

2.3.3.2 Regulation of Bistability by Dam Methylation

As a rule, hemimethylation of GATC sites in γ-protobacterial genomes is transient because the Dam methyltransferase trails the DNA replication fork at a relatively short distance, and methylation of the daughter DNA strand restores two-strand methylation (Marinus 1996; Wion and Casadesus 2006). However, the activity of the Dam methyltransferase at specific GATC sites can be hindered by the binding of proteins, in a manner reminiscent of sequestration of *oriC* by SeqA (Blyn et al. 1990; Wang and Church 1992; Ringquist and Smith 1992). As a consequence, a

fraction of GATC sites in the genomes of *E. coli* and *Salmonella* are stably undermethylated (hemimethylated or nonmethylated) (Blyn et al. 1990; Wang and Church 1992; Ringquist and Smith 1992; Sanchez-Romero et al. 2020). Because active demethylation is not known to occur in bacteria except during DNA repair (Li et al. 2012), competition between specific DNA-binding proteins and Dam methyltransferase may be a common mechanism to generate stable hemimethylation (Casadesus and Low 2006). Nonmethylation occurs when DNA methyltransferase activity is blocked during two consecutive DNA replication rounds. Some undermethylated GATC sites show distinct methylation states depending on growth conditions, suggesting that undermethylation might be the consequence of protein binding in response to physiological or environmental stimuli (Ringquist and Smith 1992; Tavazoie and Church 1998; Hale et al. 1994). Undermethylation of GATC sites may not be an exception: orphan DNA methyltransferases other than Dam seem to perform incomplete methylation of their targets (Blow et al. 2016). Hindrance of Dam methylation by competing proteins requires that the processivity of Dam methyltransferase is reduced. This reduction typically occurs at a GATC sites that are part of GATC clusters (two or more GATC sites separated by short distances) and contain AT-rich sequences at their boundaries (Peterson and Reich 2006; Coffin and Reich 2008).

The formation of undermethylated GATC sites at promoters or regulatory regions is often an indication of transcriptional control by DNA methylation and can be a source of bistability, reversible or not (Sanchez-Romero and Casadesus 2020). Bistable gene expression generates phenotypic lineages, which can have adaptive value in hostile and changing environments and are often involved in host–pathogen interactions (Sanchez-Romero and Casadesus 2021). A classical example is the *pap* operon of uropathogenic *E. coli* strains which encodes fimbrial adhesins for adherence to the urinary tract epithelium (Blyn et al. 1990; van der Woude et al. 1996; Braaten et al. 1994). Populations of uropathogenic *E. coli* contain a mixture of Pap^{ON} and Pap^{OFF} cells, and the Pap^{OFF} subpopulation is always larger because switching is skewed toward the OFF state (van der Woude et al. 1996; A. Hernday et al. 2002). Pap^{ON} and Pap^{OFF} subpopulations harbor distinct DNA methylation patterns in the *pap* regulatory region, which contains two GATC sites of the reduced processivity type. In the OFF state, $GATC_{prox}$ is nonmethylated and $GATC_{dist}$ is methylated. In the ON state, $GATC_{prox}$ is methylated and $GATC_{dist}$ is nonmethylated (Fig. 2.3).

The methylation-blocking protein that creates Dam methylation patterns at the *pap* operon is the global regulator Lrp. The upstream regulatory region (UAS) of *pap* contains six sites for Lrp binding, and two such sites contain GATCs. When the *pap* operon is not transcribed (OFF state), Lrp is bound to the three downstream sites and represses transcription. Binding to the downstream sites reduces Lrp affinity for the upstream sites, generating a feedback loop that propagates the OFF state (Hernday et al. 2003). Occupancy of the downstream sites prevents methylation of $GATC_{prox}$ while $GATC_{dist}$ is methylated. This DNA methylation pattern undergoes endless propagation unless a protein called PapI is present (Kaltenbach et al. 1995; Hernday et al. 2003).

Synthesis of the switching factor PapI is low and probably noisy (Hernday et al. 2002; Hernday et al. 2003). Above a critical threshold, PapI stimulates the translocation of Lrp to the three upstream binding sites in the *pap* UAS (van der Woude et al. 1996; Casadesus and Low 2006). Binding of Lrp and PapI prevents methylation of $GATC_{dist}$, which consequently becomes nonmethylated. In turn, $GATC_{prox}$ is no longer bound by Lrp and is methylated by the Dam methyltransferase. This configuration ($GATC_{dist}$ nonmethylated, $GATC_{prox}$ methylated) permits *pap* transcription (Casadesus and Low 2006; van der Woude et al. 1996). A positive feedback loop sustains the Pap^{ON} state: one of the proteins encoded by the *pap* operon, PapB, enhances transcription of the *papI* gene (van der Woude et al. 1996). Under laboratory conditions, the ON state

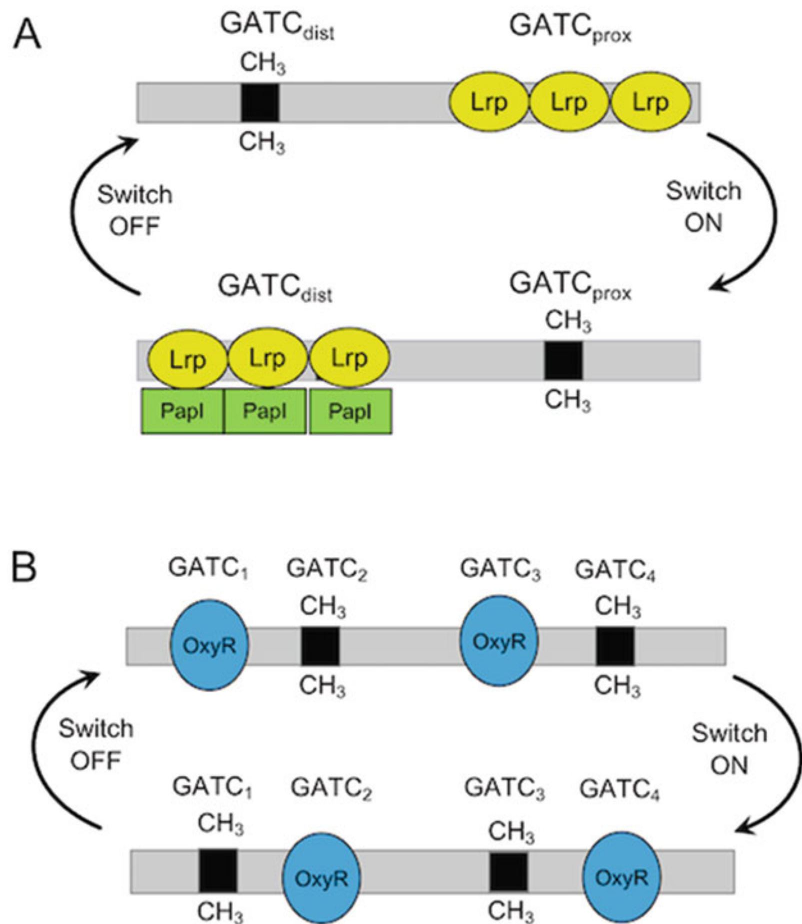


Fig. 2.3 Phase variation in the *pap* and *opvAB* operons. DNA hemimethylation states during switching are not shown

is perpetuated during 10–12 generations on average, probably with large fluctuations (Casadesus and Low 2006).

Switching from ON on OFF requires a decrease in the concentration of PapI below a critical threshold, perhaps by proteolytic degradation (van der Woude et al. 1996). In the absence of PapI, Lrp is unable to bind the upstream regulatory sites and translocates to the downstream sites. Release of the upstream sites by Lrp permits methylation of GATC$_{dist}$, and Lrp translocation to the downstream sites hinders methylation of GATC$_{prox}$ (van der Woude et al. 1996; Casadesus and Low 2006). The Pap^{OFF} pattern (GATC$_{dist}$ methylated, GATC$_{prox}$ nonmethylated) is thus restored (Casadesus and Low 2006, 2013; Low and Casadesus 2008; van der Woude et al. 1996).

Other phase variation systems regulated by Dam methylation and Lrp are the *foo, clp,* and *pef* fimbrial operons, and the architecture of their regulatory regions reminds of *pap* (Casadesus and Low 2006). Certain phase variation loci controlled by Dam methylation use DNA-binding regulators other than Lrp, including OxyR and HdfR (Table 2.1). A paradigm of this class is the *E. coli agn43* locus, which encodes an outer membrane protein involved in biofilm formation and host–pathogen interaction (Henderson and Owen 1999; Danese et al. 2000; Luthje and Brauner 2010). Binding of OxyR to the *agn43* regulatory region blocks methylation of three GATC sites, and inhibits *agn43* transcription (Waldron et al. 2002; Haagmans and van der Woude 2000) Switching to the Agn43^{ON} state requires full (two-strand) GATC methylation,

Table 2.1 Examples of bacterial genes under transcriptional control by DNA methylation-sensitive proteins or protein complexes

Species	Locus	Methylation-sensitive protein or protein complex	Active state of the promoter or the regulatory region	Function
Caulobacter crescentus	*ctrA*	GcrA	Hemimethylated	Cell cycle control
Caulobacter crescentus	*creS*	GcrA	Hemimethylated	Crescentin (structural protein)
E. coli	*Tnp (IS10)*	RNA polymerase	Hemimethylated	Transposition
E. coli	*dnaA*	SeqA	Methylated	DNA replication
E. coli	*Pap*	Lrp	Methylation pattern	Fimbriae
E. coli	*agn43*	OxyR	Methylated	Adhesion
E. coli	*sciH*	Fur	Nonmethylated	Type VI secretion
Neisseria meningitidis	*Eda*	MisR	Methylated	Enter-Doudorov aldolase
Salmonella enterica	*traJ*	Lrp	Hemimethylated	Plasmid transfer
Salmonella enterica	*Std*	HdfR	Nonmethylated	Fimbriae and pleiotropic control of gene expression
Salmonella enterica	*Gtr*	OxyR	Methylation pattern	Lipopolysaccharide modification
Salmonella enterica	*opvAB*	OxyR	Methylation pattern	Lipopolysaccharide modification

and may be facilitated by the fact that the *agn43* GATC sites do not have flanking sequences able to reduce the processivity of Dam methyltransferase. Hence, if the GATCs are not bound by OxyR, Dam will processively methylate them. Switching to the Agn43^{OFF} state can occur after DNA replication, when the three GATCs are hemimethylated (Wallecha et al. 2002). OxyR has a higher affinity for *agn43* DNA containing hemimethylated GATCs than for fully methylated *agn43* DNA. Thus, switching to OFF will be possible if OxyR binds to the GATC region before Dam methylates the GATC sites (Wallecha et al. 2003; Kaminska and van der Woude 2010).

Phase variation systems reminiscent of *agn43* include the *gtr* (glycotransferase) locus of bacteriophage P22 (Broadbent et al. 2010), certain *gtr* loci of the *Salmonella* chromosome (Davies et al. 2013), and the *Salmonella opvAB* operon (Cota et al. 2012). All these loci encode proteins that modify the bacterial lipopolysaccharide, and their transcription is controlled by Dam methylation and OxyR. In the *opvAB* operon, binding of OxyR generates distinct patterns of DNA methylation in OpvAB^{ON} and OpvAB^{OFF} cells, and Dam-dependent regulation is especially complex as it involves 2 OxyR binding sites (actually, 4 half-sites) and 4 GATC sites (Cota et al. 2016). Expression of *opvAB* produces a subpopulation of *Salmonella* cells resistant to phages that use the O-antigen as receptor (Cota et al. 2015).

Another complex Dam-dependent switch is found in the *std* operon of *Salmonella enterica*, which encodes fimbriae that promote adhesion to fucose residues in the mucus layer of the large intestine (Chessa et al. 2009). *Salmonella* populations contain Std^{OFF} and Std^{ON} cells, and the Std^{ON} subpopulation is small under laboratory conditions (Garcia-Pastor et al. 2018b) and large in the intestine (Suwandi et al. 2019). The environmental signal(s) and mechanisms that control subpopulation sizes remain hypothetical but the molecular basis of transcriptional bistability is well known (Garcia-Pastor et al. 2019). The *std* operon contains two genes that encode transcriptional regulators (Garcia-Pastor et al. 2018b). One such protein, StdF, is an autogenous activator of *std* transcription while StdE is a transcriptional activator of the *hdfR* gene, which encodes the

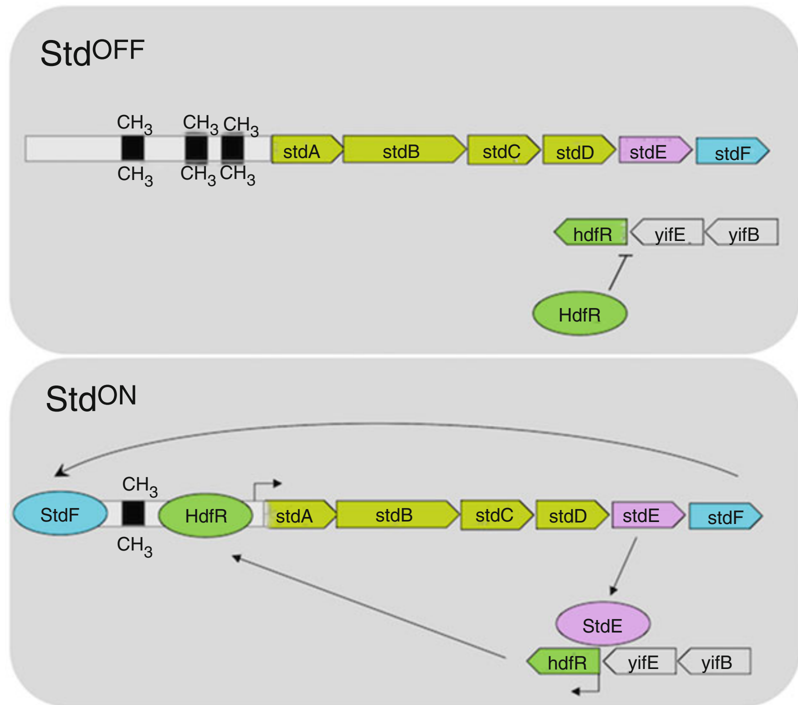


Fig. 2.4 Model for formation of Std^{OFF} and Std^{ON} lineages. In the Std^{ON} lineage, StdE disrupts autogenous repression of *hdfR* expression and increases the HdfR level above a critical threshold required for *std* transcription. StdF is a transcriptional co-activator, which binds an upstream site in the *std* UAS

main activator of *std* transcription (Fig. 2.4). Activation of *std* expression thus involves a double feedback loop that generates an AND logic gate (Garcia-Pastor et al. 2019). Further complexity arises from the fact that StdE and StdF regulate transcription of more than one hundred *Salmonella* genes, in some cases as activators and in others as repressors (Lopez-Garrido and Casadesus 2012; Garcia-Pastor et al. 2018b). Std^{OFF} and Std^{ON} cells thus differ in multiple phenotypic traits besides the presence or absence of Std fimbriae, a phenomenon that may be viewed as a rudimentary example of bacterial differentiation. Formation of cell variants during infection may adapt *Salmonella* subpopulations to distinct animal niches, either as a division of labor or as a bet-hedging strategy (Garcia-Pastor et al. 2018a; Sanchez-Romero and Casadesus 2021). Pleiotropic control of gene expression by StdE and StdF may be also reminiscent of the phase-variable regulons described later in this chapter.

In the above examples, DNA methylation patterns generate cell-to-cell differences in gene expression by promoting transcriptional bistability in a reversible, programmed manner. However, an example of Dam-dependent bistability lacking programmed reversion is found in the *sci1* operon of *E. coli* (Brunet et al. 2011), and additional examples may exist.

Another feature that departs from classical Dam-dependent bistability is found in genes where Dam methylation produces cell-to-cell variation by promoting graded transcription, perhaps through enhancement of noise (Sanchez-Romero et al. 2020). The possibility that the presence of N^6-methyladenine at or near promoters increases noise may be supported by old observations indicating that base methylation decreases the thermodynamic stability of the double helix (Engel and von Hippel 1978) and alters DNA curvature (Diekmann 1987; Polaczek et al. 1998).

2.4 Phase Variable DNA Adenine Methylation

A significant fraction, perhaps one-fifth, of restriction-modification systems of types I and III show phase-variable expression (Atack et al.

2018; Atack et al. 2020). In type III systems, phase variation usually involves contraction or expansion of short sequence repeats, generating cell lineages with an active (ON) or an inactive system (OFF) (Seib et al. 2020). In type I systems, phase variation is often triggered by recombination, which causes switching in target specificity. In either case, switching between OFF and ON states or change in DNA methyltransferase specificity can alter gene expression in a pleiotropic manner, producing a phase-variable regulon or "phasevarion." Only a modest survey of phasevarions is made here as comprehensive reviews have been published (Srikhanta et al. 2005, 2009b, 2010; Phillips et al. 2019; Seib et al. 2020).

Certain type III phasevarions conserve their restriction-modification activity (Fox et al. 2007b); in others, however, the modification gene (*mod*) remains active but the restriction enzyme is inactivated by mutation. Whatever the case, phase-variable synthesis of Mod methyltransferase generates two subpopulations of bacterial cells, one of which contains N^6-methyladenine in its genome while the other subpopulation does not. As a consequence, each lineage shows a distinct pattern of expression of DNA methylation-sensitive loci (Srikhanta et al. 2005, 2009a, 2010).

A paradigm among type III restriction-modification systems is the ModA phasevarion of the respiratory pathogen *Haemophilus influenzae*. More than twenty ModA alleles with different target specificities have been described, each controlling in a strain-specific manner the expression of genes involved in virulence, evasion of the immune system, antibiotic resistance, and other adaptive traits (Atack et al. 2015).

Virulence-related phasevarions have been also described in *Neisseria meningitidis,* an opportunistic pathogen that can cause meningitis and other pathologies, and in *Neisseria gonorrheae,* the causative agent of gonorrhea. As above, Mod alleles exist and each Mod variant controls the expression of a distinct set of genes, whose number differs among variants (e.g., 250 in ModA11, 26 in ModA12, 54 in ModA13, etc.) (Srikhanta et al. 2010; Seib et al. 2020). In *N. gonorrhoeae,* lack of the ModA13 methyltransferase impairs antimicrobial resistance, invasion of epithelial cells, and biofilm formation (Srikhanta et al. 2009a). In turn, the *N. gonorrhoeae* NgoAXP restriction-modification system controls biofilm formation, adhesion to host cells and epithelial cell invasion, and may regulate the transcription of more than 100 genes (Kwiatek et al. 2015). In many cases, it is not known whether phasevarion-mediated control of transcription is direct or indirect. An exception is found in *Neisseria meningitidis*, where methylation of the 3′ adenine at 5′ACGTAGG3′ targets by ModA11 methyltransferase directly controls transcription of the *eda* gene, which encodes Entner–Doudoroff aldolase, an enzyme involved in glucose catabolism to pyruvate (Jen et al. 2020). Type III phasevarions, each with a distinct role in virulence, have been also described in *Moraxella catarrhalis* (Blakeway et al. 2014), *Helicobacter pylori* (Bayliss et al. 2006; Srikhanta et al. 2017b) and *Kingella kingae* (Srikhanta et al. 2017a).

Even though the study of type III phase-variable DNA methylation has been so far centered on pathogenic species, bioinformatic analysis predicts the existence of phasevarions in environmental bacteria including *E. coli* and *Pseudomonas* (Atack et al. 2018). A potential involvement of type III phasevarions in bacteriophage resistance has been proposed (Atack et al. 2020).

Among type I phasevarions, a fascinating example has been described in *Streptococcus pneumoniae*, an opportunistic pathogen that causes several types of acute infection including pneumonia and meningitis (Weiser et al. 1994). Switching occurs by homologous recombination and produces six DNA methyltransferase variants (Manso et al. 2014; De Ste Croix et al. 2017). Each variant generates a distinct pattern of genome methylation that gives rise to cell types with distinct virulence properties. Formation of such lineages may facilitate adaptation during different stages of the infection (Manso et al. 2014; Li et al. 2016; Oliver et al. 2017).

A N^6-methyladenine phasevarion involving a restriction-modification system of the rare type

IIG (presumably derived from a type I system) has been described in *Campylobacter jejunii,* and a potential role in restriction of bacteriophage replication has been proposed (Anjum et al. 2016).

2.5 Additional Examples of DNA Adenine Methylation

SMRT and nanopore sequencing, accompanied by increasing refinement of bioinformatic procedures, have identified active DNA methyltransferase genes in numerous bacterial genera (Blow et al. 2016; Beaulaurier et al. 2018; Atack et al. 2018; Anton and Roberts 2021). These studies remain at a descriptive stage in most cases but exceptions exist. For instance, DNA methylation is required for virulence in the Gram-positive pathogen *Clostridioides difficile,* where the lack of a solitary DNA methyltransferase known as CamA impairs sporulation and biofilm formation (Oliveira et al. 2020). CamA methylates the 3′ adenine of 5′CAAAAA3′ targets (Oliveira et al. 2020) and its catalytic mechanism presents the unusual feature that SAM binds the enzyme with low affinity (Zhou et al. 2021a). This trait has raised the possibility of designing specific inhibitors (Zhou et al. 2021b). Another example of biomedical interest involves two DNA methyltransferases of the opportunistic pathogen *Burkholderia caenocepacia*, one solitary and the other type III, each involved in control of particular virulence-related traits (Vandenbussche et al. 2020). An additional, noteworthy case involves CtsM, an orphan DNA adenine methyltransferase of the food poisoning pathogen *Campylobacter jejuni*. Methylation of the 3′ adenine of 5′RAATTY3′ sites by CtsM controls DNA discrimination in natural transformation (Beauchamp et al. 2017).

An intriguing connection between DNA adenine methylation and virulence is found in *Mycobacterium tuberculosis*. Strains of the Euro-American lineage of *M. tuberculosis* harbor a DNA adenine methyltransferase called MamA, which is absent from strains of the Beijing lineage. MamA methylation appears to control survival in hypoxia, a stress condition found during human infection and may regulate the expression of a number of *M. tuberculosis* genes (Shell et al. 2013). A different DNA adenine methyltransferase is found in the Beijing lineage, suggesting that strain-specific differences in DNA methylation might control lineage-specific features, perhaps explaining differences in virulence and transmissibility (Shell et al. 2013).

2.6 C^5-Methylcytosine

A C^5-methylcytosine methyltransferase known as Dcm was described in *E. coli* several decades ago and is present in other enteric bacteria (Marinus 1996). Dcm methylates the C^5 position of internal cytosine residues in 5′CCAGG3′ and 5′CCTGG3′ sites, and its existence has been a long-lasting paradox (Marinus 1996; Marinus and Casadesus 2009). Hydrolytic deamination of C^5-methylcytosine produces thymine, generating T:G mismatches. Even though enteric bacteria possess a repair system that restores C:G pairs before replication, mutations due to C^5-methylcytosine deamination are frequent, especially in stationary phase (Poole et al. 2001). Mutational hotspots are thus created (Cherry 2018). The paradox is further strengthened by the fact that the potential benefits of C^5-methylcytosine formation remain a mystery as loss of Dcm does not have obvious phenotypic consequences, at least under laboratory conditions (Marinus and Casadesus 2009).

Recent studies, however, suggest that Dcm methylation may play physiological roles in *E. coli*, perhaps of a subtle nature, including regulation of gene expression. *E. coli dcm* mutants show increased expression of the sigma factor RpoS (Kahramanoglou et al. 2012) and overexpress a membrane protein involved in ethidium bromide transport (Militello et al. 2014). However, it remains unclear whether control of transcription is direct or indirect.

In *Helicobacter pylori*, the JHP1050 C^5-methylcytosine methyltransferase provides a paradigm for transcriptional control by C^5-methylcytosine in prokaryotes. This enzyme

derives from an ancestral restriction-modification system and its absence causes loss of adherence to host cells, impaired competence for DNA uptake, altered cell shape, and susceptibility to copper (Estibariz et al. 2019). These phenotypes correlate with altered patterns of gene expression, and transcriptional control by C^5-methylcytosine appears to be direct (Estibariz et al. 2019).

Another *H. pylori* C^5-methylcytosine methyltransferase, HpyAVIBM, controls the expression of genes involved in motility, adhesion, and virulence but it has not been established whether the effect is direct or indirect (Kumar et al. 2012). Because DNA repeats are present in the *hpyAVIBM* coding sequence, one may speculate that repeat expansion and/or contraction might cause phase-variable expression, thus forming a C^5-methylcytosine phasevarion (Kumar et al. 2012).

In *Vibrio cholerae*, an orphan C^5-methylcytosine methyltransferase known as VhcM is necessary for optimal growth both in vitro and during infection and modulates stress responses. Transcriptome changes in *vhcM* deletion mutants may be a consequence of direct transcriptional control but indirect effects are also conceivable (Chao et al. 2015).

2.7 N^4-Methylcytosine

N^4-methylcytosine produced by DNA methyltransferases of restriction-modification systems is scarce in most bacterial genomes (Beaulaurier et al. 2018). Exceptions are thermophilic bacteria where the presence of N^4-methylcytosine may have selective value because it is more resistant to heat-induced deamination than C^5-methylcytosine (Ehrlich et al. 1985). Aside from its role in protection from restriction, the biological functions of N^4-methylcytosine remain largely unknown. An exception is found in a strain of *Helicobacter pylori* where deletion of a solitary DNA methyltransferase that introduces a methyl group at the N^4 position in the 5′ C of 5′TCTTC3′ motifs alters the expression of over one hundred genes and results in pleiotropic virulence defects (S. Kumar et al. 2018). As in other studies discussed above, transcriptional control may be direct or indirect (S. Kumar et al. 2018).

2.8 Biomedical and Biotechnological Applications of Dam Methylation

Because attenuation of virulence in DNA methyltransferase mutants is widespread among human pathogens, one can consider the possibility of using DNA methyltransferase inhibitors as antibacterial drugs (Marinus and Casadesus 2009). Such drugs have been already described for the DNA adenine methyltransferases CcrM and Dam (Benkovic et al. 2005; Mashhoon et al. 2006). In pathogens in which DNA methylation controls virulence but is not essential, inhibitors can be expected to attenuate virulence by transforming wild-type bacteria into phenocopies of DNA methyltransferase mutants. In principle, such drugs should be harmless for the host because N^6-methyladenine is rare, if not absent, in mammalian cells (Ratel et al. 2006; Xiao et al. 2018; Douvlataniotis et al. 2020). However, the effects of DNA methylation inhibitors on the normal intestinal microbiome cannot be predicted, and experimental tests have not yet been reported. If DNA methyltransferase inhibitors pass such tests, they may be advantageous over bactericidal antibiotics because the absence of lethal selection can be expected to reduce the frequency of resistance mutations. In the case of Dam methyltransferase inhibitors, another therapeutic use might be the enhancement of the efficacy of antibiotics (Cohen et al. 2016). Lack of Dam methylation increases the bactericidal activity of β-lactams because exposure to the antibiotic causes oxidative damage, which induces the SOS response including error-prone DNA repair. As a consequence, mismatches are produced. In the absence of DNA strand discrimination, the MutHLS mismatch repair system generates double-strand breaks (Iyer et al. 2006). Such breaks further activate the SOS response, thus generating a toxic feedback loop that

potentiates the lethal action of the antibiotic (Cohen et al. 2016). Another therapeutic strategy based on DNA methylation envisages the use of bacteriophages that encode DNA methyltransferases in phage therapy, in the hope that methylation of the phage genome will delay the killing of the incoming phage by host restriction-modification systems (Murphy et al. 2013).

Application of DNA to synthetic biology is another emerging field. Synthetic memory devices able to store information in the form of CcrM-dependent methylation patterns have been described in *E. coli* (Maier et al. 2017; Klingel et al. 2021). Other studies have used either CcrM or Dam methylation to build biological sensors. Examples include a CcrM methylation-based tetracycline sensor (Ullrich et al. 2020) and Dam methylation-dependent switches aimed to promote the formation of bacterial subpopulations with predetermined sizes (Olivenza et al. 2019) and to detect the presence of bacteriophages in environmental samples (Olivenza et al. 2020). A side benefit of synthetic biology manipulations is that they sometimes reveal subtle, previously overlooked traits of the biological system under study (Olivenza et al. 2019). Following Richard Feynman's dictum, "What I cannot create I do not understand," use of DNA methylation in synthetic biology can thus be expected to reveal hitherto unknown aspects of DNA methylation biochemistry.

Acknowledgments Research in our laboratory is supported by grant PID2020-116995RB-I00 from the Ministerio de Ciencia e Innovación of Spain and the European Regional Fund.

References

Adhikari S, Curtis PD (2016) DNA methyltransferases and epigenetic regulation in bacteria. FEMS Microbiol Rev 40(5):575–591. https://doi.org/10.1093/femsre/fuw023

Anjum A, Brathwaite KJ, Aidley J, Connerton PL, Cummings NJ, Parkhill J et al (2016) Phase variation of a Type IIG restriction-modification enzyme alters site-specific methylation patterns and gene expression in Campylobacter jejuni strain NCTC11168. Nucleic Acids Res 44(10):4581–4594. https://doi.org/10.1093/nar/gkw019

Anton BP, Roberts RJ (2021) Beyond restriction modification: Epigenomic roles of DNA methylation in prokaryotes. Annu Rev Microbiol 75:129–149. https://doi.org/10.1146/annurev-micro-040521-035040

Arber W, Linn S (1969) DNA modification and restriction. Annu Rev Biochem 38:467–500. https://doi.org/10.1146/annurev.bi.38.070169.002343

Atack JM, Srikhanta YN, Fox KL, Jurcisek JA, Brockman KL, Clark TA et al (2015) A biphasic epigenetic switch controls immunoevasion, virulence and niche adaptation in non-typeable Haemophilus influenzae. Nat Commun 6:7828. https://doi.org/10.1038/ncomms8828

Atack JM, Yang Y, Seib KL, Zhou Y, Jennings MP (2018) A survey of type III restriction-modification systems reveals numerous, novel epigenetic regulators controlling phase-variable regulons; phasevarions. Nucleic Acids Res 46(7):3532–3542. https://doi.org/10.1093/nar/gky192

Atack JM, Guo C, Litfin T, Yang L, Blackall PJ, Zhou Y et al (2020) Systematic analysis of REBASE identifies numerous type I restriction-modification systems with duplicated, distinct hsdS specificity genes that can switch system specificity by recombination. mSystems 5(4):e00497-20. https://doi.org/10.1128/mSystems.00497-20

Balbontin R, Rowley G, Pucciarelli MG, Lopez-Garrido J, Wormstone Y, Lucchini S et al (2006) DNA adenine methylation regulates virulence gene expression in salmonella enterica serovar typhimurium. J Bacteriol 188(23):8160–8168. https://doi.org/10.1128/JB.00847-06

Bayliss CD, Callaghan MJ, Moxon ER (2006) High allelic diversity in the methyltransferase gene of a phase variable type III restriction-modification system has implications for the fitness of Haemophilus influenzae. Nucleic Acids Res 34(14):4046–4059. https://doi.org/10.1093/nar/gkl568

Beauchamp JM, Leveque RM, Dawid S, DiRita VJ (2017) Methylation-dependent DNA discrimination in natural transformation of campylobacter jejuni. Proc Natl Acad Sci U S A 114(38):E8053–E8061. https://doi.org/10.1073/pnas.1703331114

Beaulaurier J, Schadt EE, Fang G (2018) Deciphering bacterial epigenomes using modern sequencing technologies. Nat Rev Genet. https://doi.org/10.1038/s41576-018-0081-310.1038/s41576-018-0081-3

Benkovic SJ, Baker SJ, Alley MR, Woo YH, Zhang YK, Akama T et al (2005) Identification of borinic esters as inhibitors of bacterial cell growth and bacterial methyltransferases, CcrM and MenH. J Med Chem 48(23):7468–7476. https://doi.org/10.1021/jm050676a

Bergerat A, Guschlbauer W, Fazakerley GV (1991) Allosteric and catalytic binding of S-adenosylmethionine to Escherichia coli DNA adenine methyltransferase

monitored by 3H NMR. Proc Natl Acad Sci U S A 88(15):6394–6397

Blakeway LV, Power PM, Jen FE, Worboys SR, Boitano M, Clark TA et al (2014) ModM DNA methyltransferase methylome analysis reveals a potential role for Moraxella catarrhalis phasevarions in otitis media. FASEB J 28(12):5197–5207. https://doi.org/10.1096/fj.14-256578

Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R et al (2016) The Epigenomic landscape of prokaryotes. PLoS Genet 12(2): e1005854. https://doi.org/10.1371/journal.pgen.1005854

Blyn LB, Braaten BA, Low DA (1990) Regulation of pap pilin phase variation by a mechanism involving differential dam methylation states. EMBO J 9(12): 4045–4054

Boye E, Marinus MG, Løbner-Olesen A (1992) Quantitation of dam methyltransferase in Escherichia coli. J Bacteriol 174(5):1682–1685. https://doi.org/10.1128/jb.174.5.1682-1685.1992

Boye E, Løbner-Olesen A, Skarstad K (2000) Limiting DNA replication to once and only once. EMBO Rep 1(6):479–483. https://doi.org/10.1093/embo-reports/kvd116

Braaten BA, Nou X, Kaltenbach LS, Low DA (1994) Methylation patterns in pap regulatory DNA control pyelonephritis- associated pili phase variation in E. coli. Cell 76(3):577–588

Braun RE, Wright A (1986) DNA methylation differentially enhances the expression of one of the two E. coli dnaA promoters in vivo and in vitro. Mol Gen Genet 202(2):246–250

Broadbent SE, Balbontin R, Casadesus J, Marinus MG, van der Woude M (2007) YhdJ, a nonessential CcrM-like DNA methyltransferase of Escherichia coli and salmonella enterica. J Bacteriol 189(11):4325–4327. https://doi.org/10.1128/JB.01854-06

Broadbent SE, Davies MR, van der Woude MW (2010) Phase variation controls expression of salmonella lipopolysaccharide modification genes by a DNA methylation-dependent mechanism. Mol Microbiol 77(2):337–353. https://doi.org/10.1111/j.1365-2958.2010.07203.x

Brunet YR, Bernard CS, Gavioli M, Lloubes R, Cascales E (2011) An epigenetic switch involving overlapping fur and DNA methylation optimizes expression of a type VI secretion gene cluster. PLoS Genet 7(7):e1002205. https://doi.org/10.1371/journal.pgen.1002205

Bryant JM, Brown KP, Burbaud S, Everall I, Belardinelli JM, Rodriguez-Rincon D et al (2021) Stepwise pathogenic evolution of mycobacterium abscessus. Science 372(6541):eabb8699. https://doi.org/10.1126/science.abb8699

Calmann MA, Marinus MG (2003) Regulated expression of the Escherichia coli dam gene. J Bacteriol 185(16): 5012–5014

Camacho EM, Casadesus J (2002) Conjugal transfer of the virulence plasmid of salmonella enterica is regulated by the leucine-responsive regulatory protein and DNA adenine methylation. Mol Microbiol 44(6):1589–1598

Camacho EM, Casadesus J (2005) Regulation of *traJ* transcription in the *salmonella* virulence plasmid by strand-specific DNA adenine hemimethylation. Mol Microbiol 57:1700–1718

Casadesus J, Low D (2006) Epigenetic gene regulation in the bacterial world. Microbiol Mol Biol Rev 70(3): 830–856. https://doi.org/10.1128/MMBR.00016-06

Casadesus J, Low DA (2013) Programmed heterogeneity: epigenetic mechanisms in bacteria. J Biol Chem 288(20):13929–13935. https://doi.org/10.1074/jbc.R113.472274

Casadesus J, Torreblanca J (1996) Methylation-related epigenetic signals in bacterial DNA. In: Russo VEA, Martienssen RA, Riggs AD (eds) Epigenetic mechanisms of gene regulation. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp 141–153

Chao MC, Zhu S, Kimura S, Davis BM, Schadt EE, Fang G et al (2015) A cytosine methyltransferase modulates the cell envelope stress response in the cholera pathogen [corrected]. PLoS Genet 11(11):e1005666. https://doi.org/10.1371/journal.pgen.1005666

Charlier D, Gigot D, Huysveld N, Roovers M, Pierard A, Glansdorff N (1995) Pyrimidine regulation of the Escherichia coli and salmonella typhimurium carAB operons: CarP and integration host factor (IHF) modulate the methylation status of a GATC site present in the control region. J Mol Biol 250(4):383–391. https://doi.org/10.1006/jmbi.1995.0384

Cheng X (1995) Structure and function of DNA methyltransferases. Annu Rev Biophys Biomol Struct 24:293–318. https://doi.org/10.1146/annurev.bb.24.060195.001453

Cherry JL (2018) Methylation-induced Hypermutation in natural populations of bacteria. J Bacteriol 200(24): e00371-18. https://doi.org/10.1128/JB.00371-18

Chessa D, Winter MG, Jakomin M, Baumler AJ (2009) Salmonella enterica serotype typhimurium Std fimbriae bind terminal alpha(1,2)fucose residues in the cecal mucosa. Mol Microbiol 71(4):864–875. https://doi.org/10.1111/j.1365-2958.2008.06566.x

Coffin SR, Reich NO (2008) Modulation of Escherichia coli DNA methyltransferase activity by biologically derived GATC-flanking sequences. J Biol Chem 283(29):20106–20116. https://doi.org/10.1074/jbc.M802502200

Cohen NR, Ross CA, Jain S, Shapiro RS, Gutierrez A, Belenky P et al (2016) A role for the bacterial GATC methylome in antibiotic stress survival. Nat Genet 48(5):581–586. https://doi.org/10.1038/ng.3530

Collier J (2009) Epigenetic regulation of the bacterial cell cycle. Curr Opin Microbiol 12(6):722–729. https://doi.org/10.1016/j.mib.2009.08.005

Cota I, Blanc-Potard AB, Casadesus J (2012) STM2209-STM2208 (opvAB): a phase variation locus of salmonella enterica involved in control of O-antigen chain

length. PLoS One 7(5):e36863. https://doi.org/10.1371/journal.pone.0036863

Cota I, Sanchez-Romero MA, Hernandez SB, Pucciarelli MG, Garcia-Del Portillo F, Casadesus J (2015) Epigenetic control of salmonella enterica O-antigen chain length: a tradeoff between virulence and bacteriophage resistance. PLoS Genet 11(11):e1005667. https://doi.org/10.1371/journal.pgen.1005667

Cota I, Bunk B, Sproer C, Overmann J, Konig C, Casadesus J (2016) OxyR-dependent formation of DNA methylation patterns in OpvABOFF and OpvABON cell lineages of salmonella enterica. Nucleic Acids Res 44(8):3595–3609. https://doi.org/10.1093/nar/gkv1483

Danese PN, Pratt LA, Dove SL, Kolter R (2000) The outer membrane protein, antigen 43, mediates cell-to-cell interactions within Escherichia coli biofilms. Mol Microbiol 37(2):424–432. doi:mmi2008

Davies MR, Broadbent SE, Harris SR, Thomson NR, van der Woude MW (2013) Horizontally acquired glycosyltransferase operons drive salmonellae lipopolysaccharide diversity. PLoS Genet 9(6): e1003568. https://doi.org/10.1371/journal.pgen.1003568

De Ste Croix M, Vacca I, Kwun MJ, Ralph JD, Bentley SD, Haigh R et al (2017) Phase-variable methylation and epigenetic regulation by type I restriction-modification systems. FEMS Microbiol Rev 41 (Supp_1):S3–S15. https://doi.org/10.1093/femsre/fux025

diCenzo GC, Cangioli L, Nicoud Q, Cheng JHT, Blow MJ, Shapiro N et al (2022) DNA methylation in ensifer species during free-living growth and during nitrogen-fixing symbiosis with medicago spp. mSystems 7(1): e0109221. https://doi.org/10.1128/mSystems.01092-21

Diekmann S (1987) DNA methylation can enhance or induce DNA curvature. EMBO J 6(13):4213–4217

Douvlataniotis K, Bensberg M, Lentini A, Gylemo B, Nestor CE (2020) No evidence for DNA N (6)-methyladenine in mammals. Sci Adv 6(12):eaay3335. https://doi.org/10.1126/sciadv.aay3335

Dunn DB, Smith JD (1955) The occurrence of 6-methylaminopurine in microbial deoxyribonucleic acids. Biochem J 60(339th Meeting):xvii

Ehrlich M, Gama-Sosa MA, Carreira LH, Ljungdahl LG, Kuo KC, Gehrke CW (1985) DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine. Nucleic Acids Res 13(4):1399–1412. https://doi.org/10.1093/nar/13.4.1399

Engel JD, von Hippel PH (1978) Effects of methylation on the stability of nucleic acid conformations. Studies at the polymer level. J Biol Chem 253(3):927–934

Estibariz I, Overmann A, Ailloud F, Krebes J, Josenhans C, Suerbaum S (2019) The core genome m5C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in helicobacter pylori. Nucleic Acids Res 47(5): 2336–2348. https://doi.org/10.1093/nar/gky1307

Fioravanti A, Fumeaux C, Mohapatra SS, Bompard C, Brilli M, Frandi A et al (2013) DNA binding of the cell cycle transcriptional regulator GcrA depends on N6-adenosine methylation in Caulobacter crescentus and other Alphaproteobacteria. PLoS Genet 9(5): e1003541. https://doi.org/10.1371/journal.pgen.1003541PGENETICS-D-13-00133

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 7:461–465

Fox KL, Dowideit SJ, Erwin AL, Srikhanta YN, Smith AL, Jennings MP (2007a) Haemophilus influenzae phasevarions have evolved from type III DNA restriction systems into epigenetic regulators of gene expression. Nucleic Acids Res 35(15):5242–5252. https://doi.org/10.1093/nar/gkm571

Fox KL, Srikhanta YN, Jennings MP (2007b) Phase variable type III restriction-modification systems of host-adapted bacterial pathogens. Mol Microbiol 65(6): 1375–1379. https://doi.org/10.1111/j.1365-2958.2007.05873.x

Garcia-Del Portillo F, Pucciarelli MG, Casadesus J (1999) DNA adenine methylase mutants of salmonella typhimurium show defects in protein secretion, cell invasion, and M cell cytotoxicity. Proc Natl Acad Sci U S A 96(20):11578–11583

Garcia-Pastor L, Puerta-Fernandez E, Casadesus J (2018a) Bistability and phase variation in salmonella enterica. Biochim Biophys Acta Gene Regul Mech. https://doi.org/10.1016/j.bbagrm.2018.01.003

Garcia-Pastor L, Sanchez-Romero MA, Gutierrez G, Puerta-Fernandez E, Casadesus J (2018b) Formation of phenotypic lineages in salmonella enterica by a pleiotropic fimbrial switch. PLoS Genet 14(9): e1007677. https://doi.org/10.1371/journal.pgen.1007677

Garcia-Pastor L, Sanchez-Romero MA, Jakomin M, Puerta-Fernandez E, Casadesus J (2019) Regulation of bistability in the std fimbrial operon of salmonella enterica by DNA adenine methylation and transcription factors HdfR, StdE and StdF. Nucleic Acids Res. https://doi.org/10.1093/nar/gkz530

Giacomodonato MN, Sarnacki SH, Llana MN, Cerquetti MC (2009) Dam and its role in pathogenicity of salmonella enterica. J Infect Dev Ctries 3(7):484–490. https://doi.org/10.3855/jidc.465

Glickman B, van den Elsen P, Radman M (1978) Induced mutagenesis in dam- mutants of Escherichia coli: a role for 6-methyladenine residues in mutation avoidance. Mol Gen Genet 163(3):307–312

Gonzalez D, Collier J (2013) DNA methylation by CcrM activates the transcription of two genes required for the division of Caulobacter crescentus. Mol Microbiol 88(1):203–218. https://doi.org/10.1111/mmi.12180

Gonzalez D, Kozdon JB, McAdams HH, Shapiro L, Collier J (2014) The functions of DNA methylation by

CcrM in Caulobacter crescentus: a global approach. Nucleic Acids Res 42(6):3720–3735. https://doi.org/10.1093/nar/gkt1352

Haagmans W, van der Woude M (2000) Phase variation of Ag43 in Escherichia coli: dam-dependent methylation abrogates OxyR binding and OxyR-mediated repression of transcription. Mol Microbiol 35(4):877–887. doi:mmi1762

Hale WB, van der Woude MW, Low DA (1994) Analysis of nonmethylated GATC sites in the Escherichia coli chromosome and identification of sites that are differentially methylated in response to environmental stimuli. J Bacteriol 176(11):3438–3441. https://doi.org/10.1128/jb.176.11.3438-3441.1994

Harris AJ, Goldman AD (2020) The complex phylogenetic relationships of a 4mC/6mA DNA methyltransferase in prokaryotes. Mol Phylogenet Evol 149:106837. https://doi.org/10.1016/j.ympev.2020.106837

Heithoff DM, Sinsheimer RL, Low DA, Mahan MJ (1999) An essential role for DNA adenine methylation in bacterial virulence. Science 284(5416):967–970

Henaut A, Rouxel T, Gleizes A, Moszer I, Danchin A (1996) Uneven distribution of GATC motifs in the Escherichia coli chromosome, its plasmids and its phages. J Mol Biol 257(3):574–585. https://doi.org/10.1006/jmbi.1996.0186

Henderson IR, Owen P (1999) The major phase-variable outer membrane protein of Escherichia coli structurally resembles the immunoglobulin A1 protease class of exported protein and is regulated by a novel mechanism involving dam and oxyR. J Bacteriol 181(7):2132–2141

Herman GE, Modrich P (1982) Escherichia coli dam methylase. Physical and catalytic properties of the homogeneous enzyme. J Biol Chem 257(5):2605–2612

Hernday A, Krabbe M, Braaten B, Low D (2002) Self-perpetuating epigenetic pili switches in bacteria. Proc Natl Acad Sci U S A 29:29

Hernday AD, Braaten BA, Low DA (2003) The mechanism by which DNA adenine methylase and PapI activate the pap epigenetic switch. Mol Cell 12(4):947–957

Heusipp G, Falker S, Schmidt MA (2007) DNA adenine methylation and bacterial pathogenesis. Int J Med Microbiol 297(1):1–7. https://doi.org/10.1016/j.ijmm.2006.10.002

Horton JR, Liebert K, Bekes M, Jeltsch A, Cheng X (2006) Structure and substrate recognition of the Escherichia coli DNA adenine methyltransferase. J Mol Biol 358(2):559–570. https://doi.org/10.1016/j.jmb.2006.02.028

Horton JR, Woodcock CB, Opot SB, Reich NO, Zhang X, Cheng X (2019) The cell cycle-regulated DNA adenine methyltransferase CcrM opens a bubble at its DNA recognition site. Nat Commun 10(1):4600. https://doi.org/10.1038/s41467-019-12498-7

Hotchkiss RD (1948) The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem 175(1):315–332

Iyer RR, Pluciennik A, Burdett V, Modrich PL (2006) DNA mismatch repair: functions and mechanisms. Chem Rev 106(2):302–323. https://doi.org/10.1021/cr0404794

Jakomin M, Chessa D, Baumler AJ, Casadesus J (2008) Regulation of the salmonella enterica std fimbrial operon by DNA adenine methylation, SeqA, and HdfR. J Bacteriol 190(22):7406–7413. https://doi.org/10.1128/JB.01136-08

Janulaitis A, Klimasauskas S, Petrusyte M, Butkus V (1983) Cytosine modification in DNA by BcnI methylase yields N4-methyl-cytosine. FEBS Lett 161(1):131–134. https://doi.org/10.1016/0014-5793(83)80745-5

Jeltsch A (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chembiochem 3(4):274–293. https://doi.org/10.1002/1439-7633(20020402)3:4<274::AID-CBIC274>3.0.CO;2-S

Jeltsch A, Christ F, Fatemi M, Roth M (1999) On the substrate specificity of DNA methyltransferases. Adenine-N6 DNA methyltransferases also modify cytosine residues at position N4. J Biol Chem 274(28):19538–19544

Jen FE, Scott AL, Tan A, Seib KL, Jennings MP (2020) Random switching of the ModA11 type III DNA methyltransferase of Neisseria meningitidis regulates Entner-Doudoroff aldolase expression by a methylation change in the eda promoter region. J Mol Biol 432(21):5835–5842. https://doi.org/10.1016/j.jmb.2020.08.024

Joshi MC, Magnan D, Montminy TP, Lies M, Stepankiw N, Bates D (2013) Regulation of sister chromosome cohesion by the replication fork tracking protein SeqA. PLoS Genet 9(8):e1003673. https://doi.org/10.1371/journal.pgen.1003673

Kahng LS, Shapiro L (2001) The CcrM DNA methyltransferase of *agrobacterium tumefaciens* is essential, and its activity is cell-cycle regulated. J Bacteriol 183:3065–3075

Kahramanoglou C, Prieto AI, Khedkar S, Haase B, Gupta A, Benes V et al (2012) Genomics of DNA cytosine methylation in Escherichia coli reveals its role in stationary phase transcription. Nat Commun 3:886. https://doi.org/10.1038/ncomms1878

Kaltenbach LS, Braaten BA, Low DA (1995) Specific binding of PapI to Lrp-pap DNA complexes. J Bacteriol 177(22):6449–6455

Kaminska R, van der Woude MW (2010) Establishing and maintaining sequestration of dam target sites for phase variation of agn43 in Escherichia coli. J Bacteriol 192(7):1937–1945. https://doi.org/10.1128/JB.01629-09

Karran P, Marinus MG (1982) Mismatch correction at O6-methylguanine residues in E. coli DNA. Nature 296(5860):868–869. https://doi.org/10.1038/296868a0

Klingel V, Kirch J, Ullrich T, Weirich S, Jeltsch A, Radde NE (2021) Model-based robustness and bistability analysis for methylation-based, epigenetic memory

systems. FEBS J 288(19):5692–5707. https://doi.org/10.1111/febs.15838

Kozdon JB, Melfi MD, Luong K, Clark TA, Boitano M, Wang S et al (2013) Global methylation state at base-pair resolution of the Caulobacter genome throughout the cell cycle. Proc Natl Acad Sci U S A 110(48): E4658–E4667. https://doi.org/10.1073/pnas.1319315110

Kucherer C, Lother H, Kolling R, Schauzu MA, Messer W (1986) Regulation of transcription of the chromosomal dnaA gene of Escherichia coli. Mol Gen Genet 205(1): 115–121

Kumar R, Mukhopadhyay AK, Ghosh P, Rao DN (2012) Comparative transcriptomics of H. pylori strains AM5, SS1 and their hpyAVIBM deletion mutants: possible roles of cytosine methylation. PLoS One 7(8):e42303. https://doi.org/10.1371/journal.pone.0042303. PONE-D-11-13104

Kumar S, Karmakar BC, Nagarajan D, Mukhopadhyay AK, Morgan RD, Rao DN (2018) N4-cytosine DNA methylation regulates transcription and pathogenesis in helicobacter pylori. Nucleic Acids Res 46(7): 3429–3445. https://doi.org/10.1093/nar/gky126

Kwiatek A, Mrozek A, Bacal P, Piekarowicz A, Adamczyk-Poplawska M (2015) Type III methyltransferase M.NgoAX from Neisseria gonorrhoeae FA1090 regulates biofilm formation and interactions with human cells. Front Microbiol 6:1426. https://doi.org/10.3389/fmicb.2015.01426

Li D, Delaney JC, Page CM, Yang X, Chen AS, Wong C et al (2012) Exocyclic carbons adjacent to the N6 of adenine are targets for oxidation by the Escherichia coli adaptive response protein AlkB. J Am Chem Soc 134(21):8896–8901. https://doi.org/10.1021/ja3010094

Li J, Li JW, Feng Z, Wang J, An H, Liu Y et al (2016) Epigenetic switch driven by DNA inversions dictates phase variation in Streptococcus pneumoniae. PLoS Pathog 12(7):e1005762. https://doi.org/10.1371/journal.ppat.1005762. PPATHOGENS-D-16-01075

Løbner-Olesen A, Boye E, Marinus MG (1992) Expression of the Escherichia coli dam gene. Mol Microbiol 6(13):1841–1851

Løbner-Olesen A, Skovgaard O, Marinus MG (2005) Dam methylation: coordinating cellular processes. Curr Opin Microbiol 8:154–160

Lopez-Garrido J, Casadesus J (2010) Regulation of salmonella enterica pathogenicity island 1 by DNA adenine methylation. Genetics 184(3):637–649. https://doi.org/10.1534/genetics.109.108985

Lopez-Garrido J, Casadesus J (2012) Crosstalk between virulence loci: regulation of salmonella enterica pathogenicity island 1 (SPI-1) by products of the std fimbrial operon. PLoS One 7(1):e30499. https://doi.org/10.1371/journal.pone.0030499. PONE-D-11-22508

Low DA, Casadesus J (2008) Clocks and switches: bacterial gene regulation by DNA adenine methylation. Curr Opin Microbiol 11(2):106–112. https://doi.org/10.1016/j.mib.2008.02.012

Lu M, Campbell JL, Boye E, Kleckner N (1994) SeqA: a negative modulator of replication initiation in E. coli. Cell 77(3):413–426. https://doi.org/10.1016/0092-8674(94)90156-2

Luthje P, Brauner A (2010) Ag43 promotes persistence of uropathogenic Escherichia coli isolates in the urinary tract. J Clin Microbiol 48(6):2316–2317. https://doi.org/10.1128/JCM.00611-10

Maier JAH, Mohrle R, Jeltsch A (2017) Design of synthetic epigenetic circuits featuring memory effects and reversible switching based on DNA methylation. Nat Commun 8:15336. https://doi.org/10.1038/ncomms15336

Malone T, Blumenthal RM, Cheng X (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. J Mol Biol 253(4):618–632. https://doi.org/10.1006/jmbi.1995.0577

Manso AS, Chai MH, Atack JM, Furi L, De Ste Croix M, Haigh R et al (2014) A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. Nat Commun 5:5055. https://doi.org/10.1038/ncomms6055

Marinus MG (1996) Methylation of DNA. In: Neidhardt FC (ed) Escherichia coli and salmonella. Cellular and molecular biology. ASM Press, Washington, D.C., pp 782–791

Marinus MG (2000) Recombination is essential for viability of an Escherichia coli dam (DNA adenine methyltransferase) mutant. J Bacteriol 182(2):463–468

Marinus MG, Casadesus J (2009) Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. FEMS Microbiol Rev 33(3):488–503. https://doi.org/10.1111/j.1574-6976.2008.00159.x

Marinus MG, Poteete A, Arraj JA (1984) Correlation of DNA adenine methylase activity with spontaneous mutability in Escherichia coli K-12. Gene 28(1): 123–125

Mashhoon N, Pruss C, Carroll M, Johnson PH, Reich NO (2006) Selective inhibitors of bacterial DNA adenine methyltransferases. J Biomol Screen 11(5):497–510. https://doi.org/10.1177/1087057106287933

Messer W, Bellekes U, Lother H (1985) Effect of dam methylation on the activity of the E. coli replication origin, oriC. EMBO J 4(5):1327–1332

Militello KT, Mandarano AH, Varechtchouk O, Simon RD (2014) Cytosine DNA methylation influences drug resistance in Escherichia coli through increased sugE expression. FEMS Microbiol Lett 350(1): 100–106. https://doi.org/10.1111/1574-6968.12299

Mohapatra SS, Fioravanti A, Biondi EG (2014a) DNA methylation in Caulobacter and other Alphaproteobacteria during cell cycle progression. Trends Microbiol 22(9):528–535. https://doi.org/10.1016/j.tim.2014.05.003

Mohapatra SS, Fioravanti A, Biondi EG (2014b) DNA methylation in Caulobacter and other

Alphaproteobacteria during cell cycle progression. Trends Microbiol 22:528–535. https://doi.org/10.1016/j.tim.2014.05.003
Mohapatra SS, Fioravanti A, Vandame P, Spriet C, Pini F, Bompard C et al (2020) Methylation-dependent transcriptional regulation of crescentin gene (creS) by GcrA in Caulobacter crescentus. Mol Microbiol 114(1):127–139. https://doi.org/10.1111/mmi.14500
Mott ML, Berger JM (2007) DNA replication initiation: mechanisms and regulation in bacteria. Nat Rev Microbiol 5(5):343–354. https://doi.org/10.1038/nrmicro1640
Mouammine A, Collier J (2018) The impact of DNA methylation in Alphaproteobacteria. Mol Microbiol 110(1):1–10. https://doi.org/10.1111/mmi.14079
Murphy J, Mahony J, Ainsworth S, Nauta A, van Sinderen D (2013) Bacteriophage orphan DNA methyltransferases: insights from their bacterial origin, function, and occurrence. Appl Environ Microbiol 79(24):7547–7555. https://doi.org/10.1128/AEM.02229-13
Oliveira PH, Fang G (2021) Conserved DNA methyltransferases: a window into fundamental mechanisms of epigenetic regulation in bacteria. Trends Microbiol 29(1):28–40. https://doi.org/10.1016/j.tim.2020.04.007
Oliveira PH, Touchon M, Rocha EP (2014) The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res 42(16):10618–10631. https://doi.org/10.1093/nar/gku734
Oliveira PH, Ribis JW, Garrett EM, Trzilova D, Kim A, Sekulovic O et al (2020) Epigenomic characterization of Clostridioides difficile finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis. Nat Microbiol 5(1):166–180. https://doi.org/10.1038/s41564-019-0613-4
Olivenza DR, Nicoloff H, Sanchez-Romero MA, Cota I, Andersson DI, Casadesus J (2019) A portable epigenetic switch for bistable gene expression in bacteria. Sci Rep 9(1):11261. https://doi.org/10.1038/s41598-019-47650-2
Olivenza DR, Casadesus J, Ansaldi M (2020) Epigenetic biosensors for bacteriophage detection and phage receptor discrimination. Environ Microbiol 22(8):3126–3142. https://doi.org/10.1111/1462-2920.15050
Oliver MB, Basu Roy A, Kumar R, Lefkowitz EJ, Swords WE (2017) Streptococcus pneumoniae TIGR4 phase-locked opacity variants differ in virulence phenotypes. mSphere 2(6):e00386-17. https://doi.org/10.1128/mSphere.00386-17. mSphere00386-17
Peterson SN, Reich NO (2006) GATC flanking sequences regulate dam activity: evidence for how dam specificity may influence pap expression. J Mol Biol 355(3):459–472. https://doi.org/10.1016/j.jmb.2005.11.003
Phillips ZN, Husna AU, Jennings MP, Seib KL, Atack JM (2019) Phasevarions of bacterial pathogens - phase-variable epigenetic regulators evolving from restriction-modification systems. Microbiology. https://doi.org/10.1099/mic.0.000805
Polaczek P, Kwan K, Campbell JL (1998) GATC motifs may alter the conformation of DNA depending on sequence context and N6-adenine methylation status: possible implications for DNA-protein recognition. Mol Gen Genet 258(5):488–493
Poole A, Penny D, Sjoberg BM (2001) Confounded cytosine! Tinkering and the evolution of DNA. Nat Rev Mol Cell Biol 2(2):147–151. https://doi.org/10.1038/35052091
Prieto AI, Ramos-Morales F, Casadesus J (2004) Bile-induced DNA damage in Salmonella enterica. Genetics 168(4):1787–1794. https://doi.org/10.1534/genetics.104.031062
Pukkila PJ, Peterson J, Herman G, Modrich P, Meselson M (1983) Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in Escherichia coli. Genetics 104(4):571–582
Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akeson M et al (2017) Mapping DNA methylation with high-throughput nanopore sequencing. Nat Methods 14(4):411–413. https://doi.org/10.1038/nmeth.4189
Ratel D, Ravanat JL, Charles MP, Platet N, Breuillaud L, Lunardi J et al (2006) Undetectable levels of N6-methyl adenine in mouse DNA: cloning and analysis of PRED28, a gene coding for a putative mammalian DNA adenine methyltransferase. FEBS Lett 580(13):3179–3184. https://doi.org/10.1016/j.febslet.2006.04.074
Ringquist S, Smith CL (1992) The Escherichia coli chromosome contains specific, unmethylated dam and dcm sites. Proc Natl Acad Sci U S A 89(10):4539–4543
Roberts D, Hoopes BC, McClure WR, Kleckner N (1985) IS10 transposition IS regulated by DNA adenine methylation. Cell 43(1):117–130. doi:0092-8674(85)90017-0
Robertson GT, Reisenauer A, Wright R, Jensen RB, Jensen A, Shapiro L et al (2000) The *Brucella abortus* CcrM DNA methyltransferase is essential for viability, and its overexpression attenuates intracellular replication in murine macrophages. J Bacteriol 182:3482–3489
Sanchez-Romero MA, Casadesus J (2020) The bacterial epigenome. Nat Rev Microbiol 18(1):7–20. https://doi.org/10.1038/s41579-019-0286-2
Sanchez-Romero MA, Casadesus J (2021) Waddington's landscapes in the bacterial world. Front Microbiol 12:685080. https://doi.org/10.3389/fmicb.2021.685080
Sanchez-Romero MA, Cota I, Casadesus J (2015) DNA methylation in bacteria: from the methyl group to the methylome. Curr Opin Microbiol 25:9–16. https://doi.org/10.1016/j.mib.2015.03.004
Sanchez-Romero MA, Olivenza DR, Gutierrez G, Casadesus J (2020) Contribution of DNA adenine methylation to gene expression heterogeneity in salmonella enterica. Nucleic Acids Res 48(21):11857–11867. https://doi.org/10.1093/nar/gkaa730

Seib KL, Srikhanta YN, Atack JM, Jennings MP (2020) Epigenetic regulation of virulence and Immunoevasion by phase-variable restriction-modification Systems in Bacterial Pathogens. Annu Rev Microbiol 74:655–671. https://doi.org/10.1146/annurev-micro-090817-062346

Shell SS, Prestwich EG, Baek SH, Shah RR, Sassetti CM, Dedon PC et al (2013) DNA methylation impacts gene expression and ensures hypoxic survival of mycobacterium tuberculosis. PLoS Pathog 9(7):e1003419. https://doi.org/10.1371/journal.ppat.1003419

Srikhanta YN, Maguire TL, Stacey KJ, Grimmond SM, Jennings MP (2005) The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. Proc Natl Acad Sci U S A 102(15):5547–5551. https://doi.org/10.1073/pnas.0501169102

Srikhanta YN, Dowideit SJ, Edwards JL, Falsetta ML, Wu HJ, Harrison OB et al (2009a) Phasevarions mediate random switching of gene expression in pathogenic Neisseria. PLoS Pathog 5(4):e1000400. https://doi.org/10.1371/journal.ppat.1000400

Srikhanta YN, Fox KL, Jennings MP (2009b) The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. Nat Rev Microbiol 8(3):196–206. https://doi.org/10.1038/nrmicro2283

Srikhanta YN, Fox KL, Jennings MP (2010) The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. Nat Rev Microbiol 8(3):196–206. https://doi.org/10.1038/nrmicro2283

Srikhanta YN, Fung KY, Pollock GL, Bennett-Wood V, Howden BP, Hartland EL (2017a) Phasevarion-regulated virulence in the emerging pediatric pathogen Kingella kingae. Infect Immun 85(12):e00319-17. https://doi.org/10.1128/IAI.00319-17

Srikhanta YN, Gorrell RJ, Power PM, Tsyganov K, Boitano M, Clark TA et al (2017b) Methylomic and phenotypic analysis of the ModH5 phasevarion of helicobacter pylori. Sci Rep 7(1):16140. https://doi.org/10.1038/s41598-017-15721-x

Stephens C, Reisenauer A, Wright R, Shapiro L (1996) A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. Proc Natl Acad Sci U S A 93(3):1210–1214

Sternberg N, Sauer B, Hoess R, Abremski K (1986) Bacteriophage P1 cre gene and its regulatory region. Evidence for multiple promoters and for regulation by DNA methylation. J Mol Biol 187(2):197–212. doi:0022-2836(86)90228-7

Suwandi A, Galeev A, Riedel R, Sharma S, Seeger K, Sterzenbach T et al (2019) Std fimbriae-fucose interaction increases salmonella-induced intestinal inflammation and prolongs colonization. PLoS Pathog 15(7):e1007915

Tavazoie S, Church GM (1998) Quantitative whole-genome analysis of DNA-protein interactions by in vivo methylase protection in E. coli. Nat Biotechnol 16(6):566–571

Tock MR, Dryden DT (2005) The biology of restriction and anti-restriction. Curr Opin Microbiol 8(4):466–472. https://doi.org/10.1016/j.mib.2005.06.003

Torreblanca J, Casadesus J (1996) DNA adenine methylase mutants of salmonella typhimurium and a novel dam-regulated locus. Genetics 144(1):15–26

Ullrich T, Weirich S, Jeltsch A (2020) Development of an epigenetic tetracycline sensor system based on DNA methylation. PLoS One 15(5):e0232701. https://doi.org/10.1371/journal.pone.0232701

Urig S, Gowher H, Hermann A, Beck C, Fatemi M, Humeny A et al (2002) The Escherichia coli dam DNA methyltransferase modifies DNA in a highly processive reaction. J Mol Biol 319(5):1085–1096. https://doi.org/10.1016/S0022-2836(02)00371-6. S0022-2836(02)00371-6

van der Woude M, Braaten B, Low D (1996) Epigenetic phase variation of the pap operon in *Escherichia coli*. Trends Microbiol 4(1):5–9. https://doi.org/10.1016/0966-842x(96)81498-3

Vandenbussche I, Sass A, Pinto-Carbo M, Mannweiler O, Eberl L, Coenye T (2020) DNA methylation epigenetically regulates gene expression in Burkholderia cenocepacia and controls biofilm formation, cell aggregation, and motility. mSphere 5(4):e00455-20. https://doi.org/10.1128/mSphere.00455-20

Vasu K, Nagaraja V (2013) Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev 77(1):53–72. https://doi.org/10.1128/MMBR.00044-12

Waldminghaus T, Skarstad K (2009) The Escherichia coli SeqA protein. Plasmid 61(3):141–150. https://doi.org/10.1016/j.plasmid.2009.02.004

Waldron DE, Owen P, Dorman CJ (2002) Competitive interaction of the OxyR DNA-binding protein and the dam methyltransferase at the antigen 43 gene regulatory region in Escherichia coli. Mol Microbiol 44(2):509–520

Wallecha A, Munster V, Correnti J, Chan T, van der Woude M (2002) Dam- and OxyR-dependent phase variation of agn43: essential elements and evidence for a new role of DNA methylation. J Bacteriol 184(12):3338–3347

Wallecha A, Correnti J, Munster V, van der Woude M (2003) Phase variation of Ag43 is independent of the oxidation state of OxyR. J Bacteriol 185(7):2203–2209

Wang MX, Church GM (1992) A whole genome approach to in vivo DNA-protein interactions in E. coli. Nature 360(6404):606–610

Wang TC, Smith KC (1986) Inviability of dam recA and dam recB cells of Escherichia coli is correlated with their inability to repair DNA double-strand breaks produced by mismatch repair. J Bacteriol 165(3):1023–1025. https://doi.org/10.1128/jb.165.3.1023-1025.1986

Weiser JN, Austrian R, Sreenivasan PK, Masure HR (1994) Phase variation in pneumococcal opacity:

relationship between colonial morphology and nasopharyngeal colonization. Infect Immun 62(6): 2582–2589. https://doi.org/10.1128/IAI.62.6.2582-2589.1994

Wion D, Casadesus J (2006) N6-methyladenine: an epigenetic signal for DNA-protein interactions. Nat Rev Microbiol 4(3):183–192. https://doi.org/10.1038/nrmicro1350

Wright R, Stephens C, Zweiger G, Shapiro L, Alley MR (1996) Caulobacter Lon protease has a critical role in cell-cycle control of DNA methylation. Genes Dev 10(12):1532–1542. https://doi.org/10.1101/gad.10.12.1532

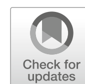
Wright R, Stephens C, Shapiro L (1997) The CcrM DNA methyltransferase is widespread in the alpha subdivision of proteobacteria, and its essential functions are conserved in rhizobium meliloti and Caulobacter crescentus. J Bacteriol 179(18):5869–5877

Xiao CL, Zhu S, He M, Chen D, Zhang Q, Chen Y et al (2018) N(6)-Methyladenine DNA modification in the human genome. Mol Cell 71(2):306–318. e7. https://doi.org/10.1016/j.molcel.2018.06.015

Zhou X, Wang J, Herrmann J, Moerner WE, Shapiro L (2019) Asymmetric division yields progeny cells with distinct modes of regulating cell cycle-dependent chromosome methylation. Proc Natl Acad Sci U S A 116(31):15661–15670. https://doi.org/10.1073/pnas.1906119116

Zhou J, Horton JR, Blumenthal RM, Zhang X, Cheng X (2021a) Clostridioides difficile specific DNA adenine methyltransferase CamA squeezes and flips adenine out of DNA helix. Nat Commun 12(1):3436. https://doi.org/10.1038/s41467-021-23693-w

Zhou J, Horton JR, Yu D, Ren R, Blumenthal RM, Zhang X et al (2021b) Repurposing epigenetic inhibitors to target the Clostridioides difficile-specific DNA adenine methyltransferase and sporulation regulator CamA. Epigenetics:1–12. https://doi.org/10.1080/15592294.2021.1976910



3 Domain Structure of the Dnmt1, Dnmt3a, and Dnmt3b DNA Methyltransferases

Shoji Tajima, Isao Suetake, Kohei Takeshita, Atsushi Nakagawa, Hironobu Kimura, and Jikui Song

Abstract

In mammals, three major DNA methyltransferases, Dnmt1, Dnmt3a, and Dnmt3b, have been identified. Dnmt3a and Dnmt3b are responsible for establishing DNA methylation patterns produced through their de novo-type DNA methylation activity in implantation stage embryos and during germ cell differentiation. Dnmt3-like (Dnmt3l), which is a member of the Dnmt3 family but does not possess DNA methylation activity, was reported to be indispensable for global methylation in germ cells. Once the DNA methylation patterns are established, maintenance-type DNA methyltransferase Dnmt1 faithfully propagates them to the next generation via replication. All Dnmts possess multiple domains. For instance, Dnmt3a and Dnmt3b each contain a Pro-Trp-Trp-Pro (PWWP) domain that recognizes the histone H3K36me2/3 mark, an Atrx-Dnmt3-Dnmt3l (ADD) domain that recognizes unmodified histone H3 tail, and a catalytic domain that methylates CpG sites. Dnmt1 contains an N-terminal independently folded domain (NTD) that interacts with a variety of regulatory factors, a replication foci-targeting sequence (RFTS) domain that recognizes the histone H3K9me3 mark and H3 ubiquitylation, a CXXC domain that recognizes unmodified CpG DNA, two tandem Bromo-Adjacent-homology (BAH1 and BAH2) domains that read the H4K20me3 mark with BAH1, and a catalytic domain that preferentially methylates hemimethylated CpG sites. In this chapter, the structures and functions of these domains are described.

Keywords

Catalytic pocket · PWWP domain · CXXC domain · RFTS domain · BAH domain · Target recognition domain · Target cytosine

S. Tajima (✉) · H. Kimura
Institute for Protein Research, Osaka University, Osaka, Japan
e-mail: tajima@protein.osaka-u.ac.jp

I. Suetake
Department of Nutritional Sciences, Faculty of Nutritional Sciences, Nakamura Gakuen University, Fukuoka, Japan

K. Takeshita
RIKEN Spring-8 Center, Sayo-gun, Hyogo, Japan

A. Nakagawa
Laboratory of Supramolecular Crystallography, Institute for Protein Research, Osaka University, Osaka, Japan

J. Song (✉)
Department of Biochemistry, University of California Riverside, Riverside, CA, USA
e-mail: jikui.song@ucr.edu

Abbreviations

AdoMet	*S*-Adenosyl-L-methionine
BAH domain	Bromo-adjacent-homology domain

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_3

DMR	Differentially methylated region
ES cells	Embryonic stem cells
ICF syndrome	Immunodeficiency, centromeric instability, and facial anomalies syndrome
NTD	The N-terminal independently folded domain
RFTS domain	Replication foci-targeting sequence domain
SAH	*S*-Adenosyl-L-homocysteine
SRA domain	The SET and RING-associated domain
TDG	Thymine DNA glycosylase
Tet enzyme	Ten-eleven translocation enzyme
TRD	Target recognition domain

3.1 DNA Methylation and Methyltransferases in Mammals

The methylation patterns of genomic DNA are established at an early stage of embryogenesis. Once the global methylation patterns are established, they are maintained during replication in a cell lineage-dependent manner (Fig. 3.1a). In mammals, a second methylation reprogramming occurs in gametogenesis. The global DNA methylation patterns are removed during an early stage of germ cell development and reestablished before meiosis in gonocytes in males and growing oocytes in females (Bird 2002). The expression of more than a hundred genes on autosomes is regulated in a sex-dependent manner, these genes being called imprinted genes. These genes are characterized by differentially methylated regions (DMRs), which undergo distinct DNA methylation in the male and female genomes. Generally, the DMR methylation patterns are established in germ cells at an identical stage to that of global DNA methylation (Kaneda et al. 2004). In mammals, three major DNA methyltransferases, Dnmt1, Dnmt3a, and Dnmt3b, have been identified (Bestor et al. 1988; Okano et al. 1998). Dnmt3a and Dnmt3b are responsible for establishing DNA methylation patterns produced through their de novo-type DNA methylation activity in implantation stage embryos and during germ cell differentiation (Okano et al. 1999). In addition, Dnmt3c, an enzyme closely related to Dnmt3b, methylates evolutionarily young transposons in the mouse male germ line (Barau et al. 2016; Jain et al. 2017). Dnmt3-like (Dnmt3l), which is a member of the Dnmt3 family but does not possess DNA methylation activity, was reported to be indispensable for global methylation in germ cells (Bourc'his et al. 2001; Hata et al. 2002). Once the DNA methylation patterns are established, the maintenance-type DNA methyltransferase Dnmt1 faithfully propagates them to the next generation after DNA replication. Dnmt1 preferentially methylates hemimethylated CpG sites, which appear after DNA replication and repair.

3.2 Enzymes Responsible for the Establishment of DNA Methylation Patterns

In mammals, the establishment of DNA methylation patterns is mainly mediated by the de novo DNA-(cytosine C5)-methyltransferases Dnmt3a and Dnmt3b, which are encoded in distinct gene loci (Aoki et al. 2001; Okano et al. 1999). Their domain arrangements are similar, each comprising a Pro-Trp-Trp-Pro (PWWP), Atrx-Dnmt3-Dnmt3l (ADD), and C-terminal catalytic domain (Fig. 3.1b). The PWWP domain is reported to bind to DNA (Qiu et al. 2002) and histone tails (Dhayalan et al. 2010) and the ADD domain to interact with various proteins including histone tails, as described below (Brenner et al. 2005; Fuks et al. 2001; Otani et al. 2009). In addition, Dnmt3a isoform 1 (Dnmt3a1) contains a ubiquitin-dependent recruitment region (UDR) that binds to monoubiquitylated histone H2A lysine 119 (H2AK119ub) (Fig. 3.1b) (Weinberg et al. 2021). Dnmt3l, a homolog of Dnmt3a and Dnmt3b, possesses no conserved domain for DNA methylation but contains an ADD domain (Aapola et al. 2000) and is necessary for global

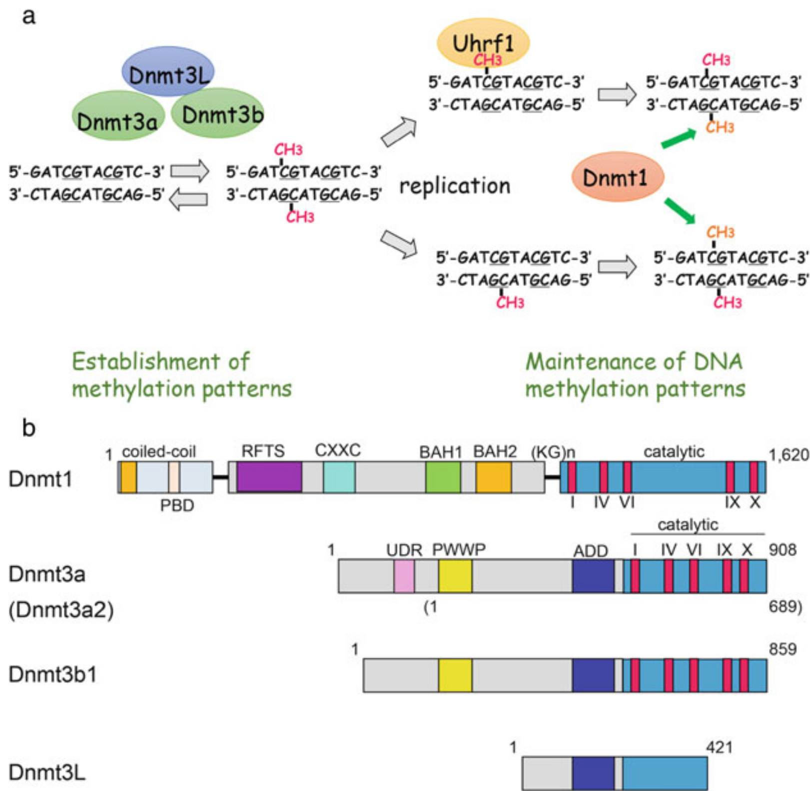


Fig. 3.1 Schematic illustration of establishment and maintenance of DNA methylation patterns. (**a**) The methylation patterns of genomic DNA are established at an early stage of embryogenesis by de novo-type DNA methyltransferases, Dnmt3a and Dnmt3b, with the aid of Dnmt3l. Once the global methylation patterns are established, they are maintained during replication by maintenance DNA methyltransferase Dnmt1 in collaboration with Uhrf1 in a cell lineage-dependent manner. (**b**) Schematic illustration of mammalian DNA methyltransferases, Dnmt1, Dnmt3a, and Dnmt3b. Dnmt3a has a short isoform utilizing different promoter and a transcription start site, Dnmt3a2. Dnmt3l, a member of the Dnmt3 family, lacks the catalytic domain and thus does not exhibit DNA methylation activity

DNA methylation (Bourc'his et al. 2001; Hata et al. 2002).

3.2.1 PWWP Domain

The PWWP domain of Dnmt3 enzymes, comprising 100–150 amino acid residues, is characterized by a central core sequence motif of Pro-Trp-Trp-Pro. It was hypothesized that the domain contributes to protein–protein interactions, especially of proteins involved in cell division, growth, and differentiation, based on a comparison of 39 proteins containing a PWWP domain (Stec et al. 2000). The PWWP domains of Dnmt3a and Dnmt3b tether them to chromatin regions (Ge et al. 2004), especially to pericentric heterochromatin and thus are responsible for their DNA methylation (Chen et al. 2004).

The PWWP domain of Dnmt3b, comprising a beta-barrel structure with 5 beta-strands followed by a five-helix bundle, turned out to be a fold responsible for DNA binding (Qiu et al. 2002). Positively charged Lys and Arg residues on the surface of the domain are expected to be the sites for DNA binding. The beta-barrel part of the PWWP domain is homologous to that of the SAND domain (named after Sp100, AIRE-1, NucP41/75, DEAF-1), which is a DNA-binding motif, and the Tudor domain, which is generally a

histone-binding motif. The PWWP domain of Dnmt3b binds to histone H3 tri-methylated at lysine 36 (H3K36me3) via a hydrophobic cage (Fig. 3.2a) (Rondelet et al. 2016), which is responsible for the recruitment of Dnmt3b, but not that of Dnmt3a, to the H3K36me3-containing gene body for de novo methylation (Baubec et al. 2015). A point mutation in the PWWP domain in Dnmt3b was found to be the cause of the immunodeficiency, centromeric instability, and facial anomalies (ICF) syndrome (Shirohzu et al. 2002), which is the consequence of hypomethylation of the pericentromere (Hansen et al. 1999; Okano et al. 1999).

The PWWP domain of Dnmt3a is highly homologous to that of Dnmt3b, possessing a corresponding hydrophobic cage (Fig. 3.2b). The Dnmt3a PWWP domain binds to both H3K36me2 and H3K36me3, with a subtle preference toward H3K36me2 (Dhayalan et al. 2010; Dukatz et al. 2019; Weinberg et al. 2019). The PWWP-H3K36me2 interaction ensures DNA methylation at the intergenic repeats (Weinberg et al. 2019). The PWWP domain of Dnmt3a also binds to DNA, which is important for chromatin association of Dnmt3a, though the affinity toward DNA is one order of magnitude lower compared to that of the PWWP domain of Dnmt3b (Dukatz et al. 2019; Purdy et al. 2010).

ZHX1, a member of the zinc finger and homeobox protein family, interacts with the PWWP domain of Dnmt3b and contributes to gene silencing (Kim et al. 2007). In addition, thymine DNA glycosylase (TDG), which is a T/G mismatch glycosylase, interacts with the PWWP and/or catalytic domains of Dnmt3a to modulate its DNA methylation activity. TDG was postulated to be responsible for the removal of formylcytosine and carboxylcytosine, which are the oxidation products of methylcytosine via hydroxymethylcytosine for active demethylation initiated by Ten-eleven translocation (Tet) enzymes (He et al. 2011; Maiti and Drohat 2011). The interaction between TDG and Dnmt3a suggests their functional interplay.

3.2.2 ADD Domain

The plant homeodomain (PHD)-like ADD domain is rich in Cys residues and reportedly binds to many factors. The ADD domain of Dnmt3a was reported to bind to corepressor RP58 (Fuks et al. 2001), oncogene c-myc (Brenner et al. 2005), Lys 9 histone H3 (H3K9) methylase Suv39h1 and heterochromatin protein 1 (HP1) beta (Fuks et al. 2003), H3K9 methylase Setdb1 (Li et al. 2006), or histone H3 un-methylated at Lys 4 (H3K4me0) (Otani et al. 2009; Zhang et al. 2010).

The three-dimensional structure of the ADD domain of Dnmt3a is similar to those of Dnmt3l and ATRX (Argentaro et al. 2007; Ooi et al. 2007) (Fig. 3.2c) and possibly Dnmt3b as well (Zhang et al. 2010). The affinity of the ADD domain of Dnmt3a to histone H3 tail is in the sub-micromolar range and is decreased by methylation modification at Lys 4 (Otani et al. 2009). This explains why the H3K4me3, which is a mark associated with active gene promoters, protects DNA from methylation (Okitsu and Hsieh 2007; Weber et al. 2007). X-ray crystallography showed that the histone H3 tail fits into the shallow groove of the PHD finger motif in the ADD domain. The main chain of Arg 2 to Thr 6 of histone H3 forms hydrogen bonds with the ADD, and this induces a conformational change of the ADD (Otani et al. 2009). The mode of recognition of the H3K4me0 by the ADD domain of Dnmt3a is similar to that of Dnmt3l (Fig. 3.2d), although the affinity is tenfold higher. As described below, Dnmt3l interacts directly with Dnmt3a and Dnmt3b (Suetake et al. 2004), and the proteins exist as a complex in embryonic stem (ES) cells (Li et al. 2007). Selective recognition of H3K4me0 by the ADD domains of Dnmt3a (Dnmt3b) and Dnmt3l may recruit de novo methyltransferases to the sites to be methylated. Conversion of the Dnmt3a ADD domain into a H3 Lys 4 methylation or H3 Thr 3 phosphorylation-insensitive module via protein engineering led to altered gene expression and/or

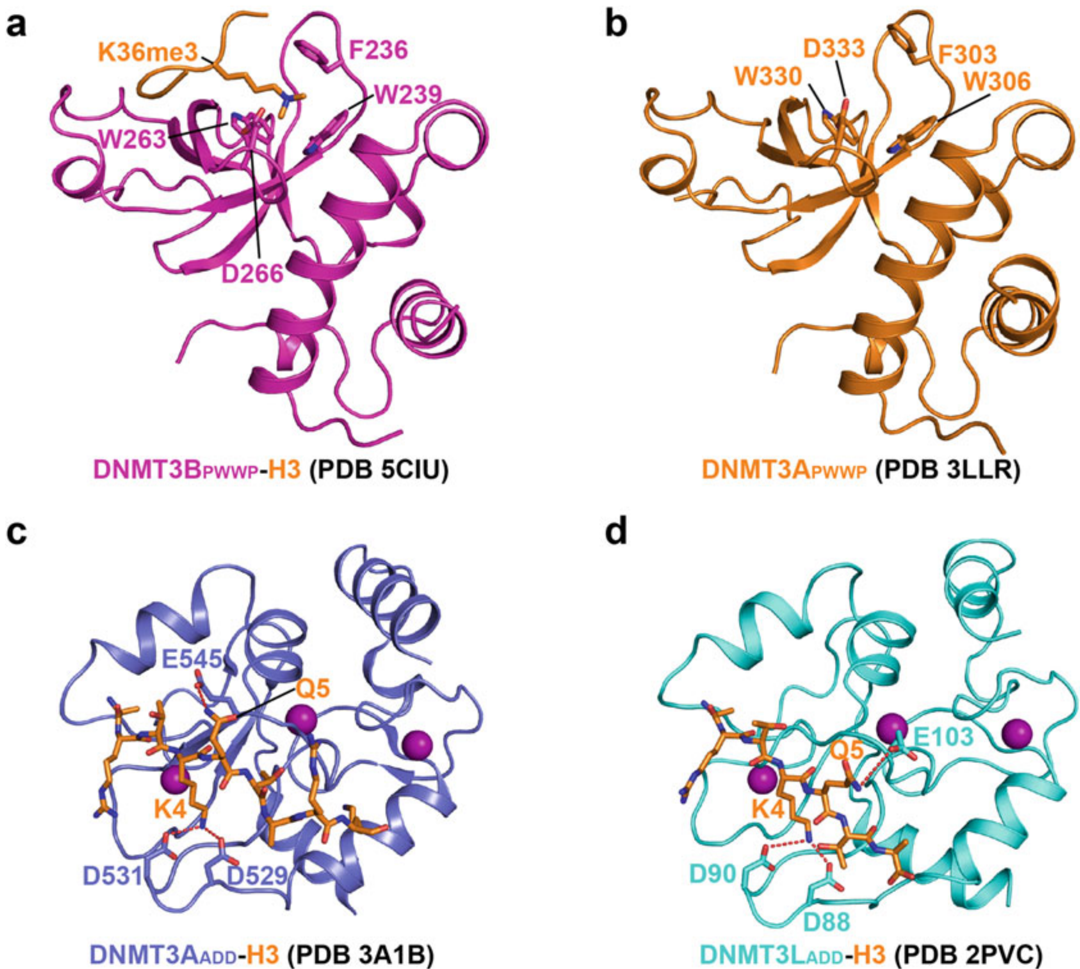


Fig. 3.2 Structures of the PWWP and ADD domains of Dnmt3s. (**a**) Ribbon diagram of the human DNMT3B PWWP domain bound to a H3K36me3 peptide (orange sticks), with the H3K36me3-binding cage shown in stick representation. (**b**) Ribbon diagram of the human DNMT3A PWWP domain, with the potential H3K36me2-binding cage shown in stick representation. (**c**, **d**) Ribbon diagram of the ADD domain of human DNMT3A (PDB accession number 3A1B) (**c**) and human DNMT3L (PDB accession number 2PVC) (**d**) bound to a histone H3 peptide (orange sticks). The H3-interacting residues are conserved in Dnmt3l and Dnmt3a. The zinc ions are shown in sphere representation

chromosome instability in mouse ES cells (Noh et al. 2015), confirming the contribution of the ADD domain to Dnmt3 targeting and function.

Interestingly, the ADD domain of Dnmt3a is located at a position that inhibits the accession of substrate DNA to the catalytic domain (Guo et al. 2015). The binding of the N-terminal tail of histone H3 induces rearrangement of the ADD domain to change its position to the one that DNA can access. Enhancement of de novo methylation at the chromatin region enriched in nucleosomes containing H3K4me0 reported previously (Li et al. 2011; Zhang et al. 2010) may be well explained by the conformational rearrangement of the ADD domain positioning (Guo et al. 2015) (Fig. 3.3). It will be important to determine whether or not other factors that are reported to interact with the ADD domain of Dnmt3a or Dnmt3b induce similar rearrangement of the enzyme to enhance de novo DNA methylation

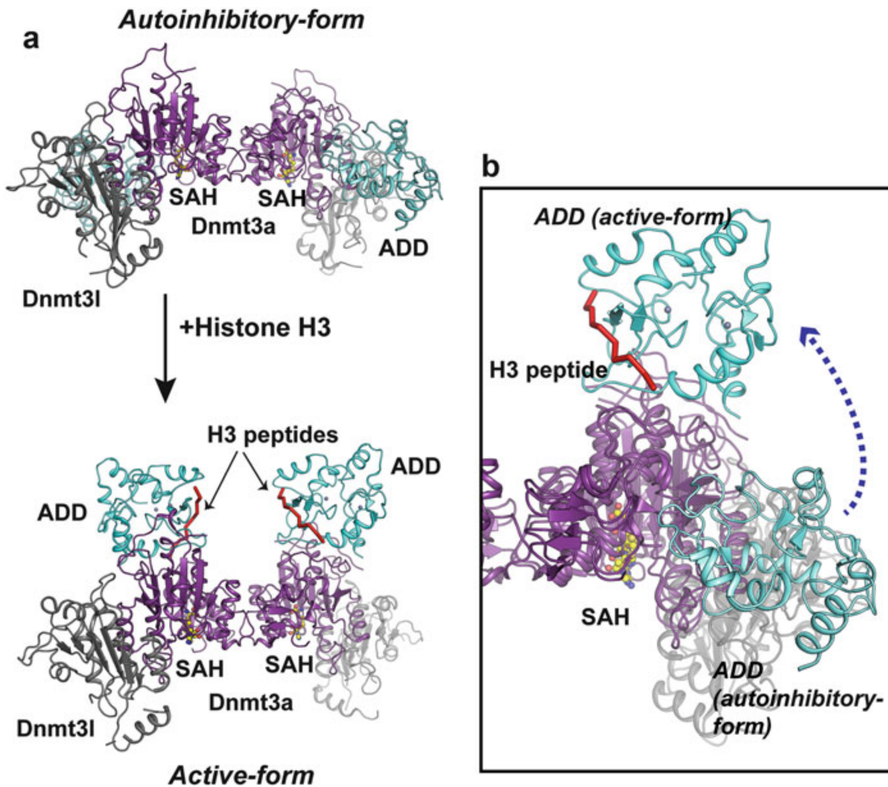


Fig. 3.3 Autoinhibition of DNMT3A by the ADD domain and histone H3 tail-induced activation of DNA methylation activity. (**a**) Ribbon illustrations of the structure of the complex of the catalytic domain with the ADD domain of Dnmt3a and the C-terminal half of Dnmt3l without (*upper*) or with (*lower*) a histone H3 tail. The catalytic domain is shown in magenta, the ADD domain in cyan, and the C-terminal region of Dnmt3l in gray. *S*-Adenosyl-L-homocysteine (SAH) is in yellow and the histone H3 tail in red. In the absence of a histone H3 tail, substrate DNA cannot gain access to the catalytic center as the ADD domain is in a position that inhibits the DNA binding (autoinhibitory form; PDB accession number 4U7P). The addition of a histone H3 tail (*red*) drastically changes the position of the ADD domain to one that allows accession of DNA to the catalytic center (active form; PDB accession number 4U7T). (**b**) Superimposition of the active and autoinhibitory forms. The *dotted arrow* indicates the movement of the ADD domain from the histone H3 tail free to the bound form

activity, or rather reinforce ADD-mediated inhibition, as reported for methyl CpG binding protein 2 (MeCP2) (Rajavelu et al. 2018).

3.2.3 Catalytic Domain

In the catalytic domains of Dnmt3a and Dnmt3b, ten motifs characteristic of DNA-(cytosine C5)-methylation activity are conserved (Kumar et al. 1994). Dnmt3a and Dnmt3b interact through their catalytic domains with the C-terminal domain of Dnmt3l, and this interaction enhances de novo DNA methylation activity (Chen et al. 2005; Suetake et al. 2004). The crystal structure of the catalytic domain of Dnmt3a in complex with the C-terminal domain of Dnmt3l has been determined (Jia et al. 2007). It is a heterotetramer comprising two Dnmt3a molecules in the center and one Dnmt3l molecule at each edge (Fig. 3.3) (Jia et al. 2007; Jurkowska et al. 2011). The association of two Dnmt3a catalytic subunits in the center of the heterotetramer presumably increases the affinity for substrate DNA and is therefore crucial for DNA methylation activity. In the absence of Dnmt3l, however, Dnmt3a

tends to polymerize using the same interaction surface as Dnmt3l. As the two interaction surfaces of Dnmt3a that cause polymerization contribute to its heterochromatin formation, it was proposed that the formation of the complex with Dnmt3l may promote releasing Dnmt3a from heterochromatin and facilitates Dnmt3a access to the substrate DNA (Jurkowska et al. 2011). It was also proposed that this inhibition of polymerization of Dnmt3a by Dnmt3l can be the underlying mechanism for the enhancement of DNA methylation activity of Dnmt3a (Jurkowska et al. 2011), especially in germ line cells to increase Dnmt3a availability and DNA methylation activity for the generation of global DNA methylation (Bourc'his et al. 2001; Hata et al. 2002; Kaneda et al. 2004). The association of Dnmt3a with Dnmt3l also serves to stabilize Dnmt3a in mouse ES cells (Veland et al. 2018).

The structures of the C-terminal domains of human DNMT3A-DNMT3L and DNMT3B-DNMT3L tetramers in complex with CpG DNA have been reported (Fig. 3.4a) (Anteneh et al. 2020; Gao et al. 2020b; Lin et al. 2020; Zhang et al. 2018). For the DNMT3A-DNMT3L-DNA complex, the two DNMT3A molecules bind to a single DNA duplex containing two separate CpG/ZpG ("Z" denotes zebularine) sites with 14bp interval, with the base of each zebularine flipped into the active site of DNMT3A, where it is stabilized via a covalent linkage with catalytic cysteine C710 and hydrogen-bonding interactions. Structural analysis of the DNMT3A-DNMT3L-DNA complexes reveals three major DNA-binding regions of DNMT3A: the catalytic loop (residues G707-K721), a loop (residues R831-F848) from the target recognition domain (TRD), and the DNMT3A/3A interface. The catalytic loop interacts with the DNA minor groove, the TRD loop interacts with the DNA major groove, while the DNMT3A/3A interface interacts with the DNA backbone of the segment bridging the two CpG sites. The substrate binding promotes the structural ordering of the TRD loop. Structural comparison of the CGT- and CGA-bound DNMT3A-DNMT3L complex further reveals that both the CpG recognition and the intramolecular interaction between the TRD loop and the DNMT3A/3A interface occur in a context-dependent fashion: In the DNMT3A-DNMT3L-CGT DNA complex, the side chain of R882 forms hydrogen bonds with both the backbone and the side chain of S837, which stabilizes the TRD loop, wherein R836 forms a direct hydrogen bond with the CpG guanine (G6) (Fig. 3.4b). However, the hydrogen bonding interaction between R882 and the backbone of S837 is disrupted in the DNMT3A-DNMT3L-CGA DNA complex. Meanwhile, TRD loop residue N838 replaces R836 to form a hydrogen bond with G6, while R836 is repositioned toward the +1- and +2-flanking nucleotides for hydrogen-bonding and van der Waals contacts (Fig. 3.4c). No protein-DNA interaction was observed for the C-terminal domain of DNMT3L. The 14bp co-methylation spacing of the DNMT3A-DNMT3L complex was supported by a subsequent biochemical analysis (Gao et al. 2020a). However, its functional implication remains unclear.

The structures of human DNMT3B-DNMT3L-DNA complexes are highly similar to those of the DNMT3A-DNMT3L-DNA complexes (Fig. 3.4a). Nevertheless, the catalytic loop, the TRD loop, and the DNMT3B/3B interface all exhibit subtle differences in DNA contacts. First, the central interface of both catalytically active subunits engages in fewer DNA contacts in the DNMT3B-DNMT3L-DNA complexes than it does in the DNMT3A-DNMT3L-DNA complexes. Second, unlike DNMT3A R882 which is hydrogen bonded to the backbone of the TRD loop in the CGT complex, the corresponding DNMT3B R823 points away from the TRD loop in both CGT and CGA complexes (Fig. 3.4d,e). Third, unlike DNMT3A which interacts with the CpG site via R836 in the CGT complex but N838 in the CGA complex, DNMT3B interacts with the CpG site via an asparagine (N779)-mediated hydrogen bond regardless of the CGT or CGA context. Finally, a side-chain hydrogen bond is formed between DNMT3B N656 and R661 in the DNMT3B-DNMT3L-DNA complexes but not in the corresponding sites in the DNMT3A-DNMT3L-DNA complexes, due to the substitution of

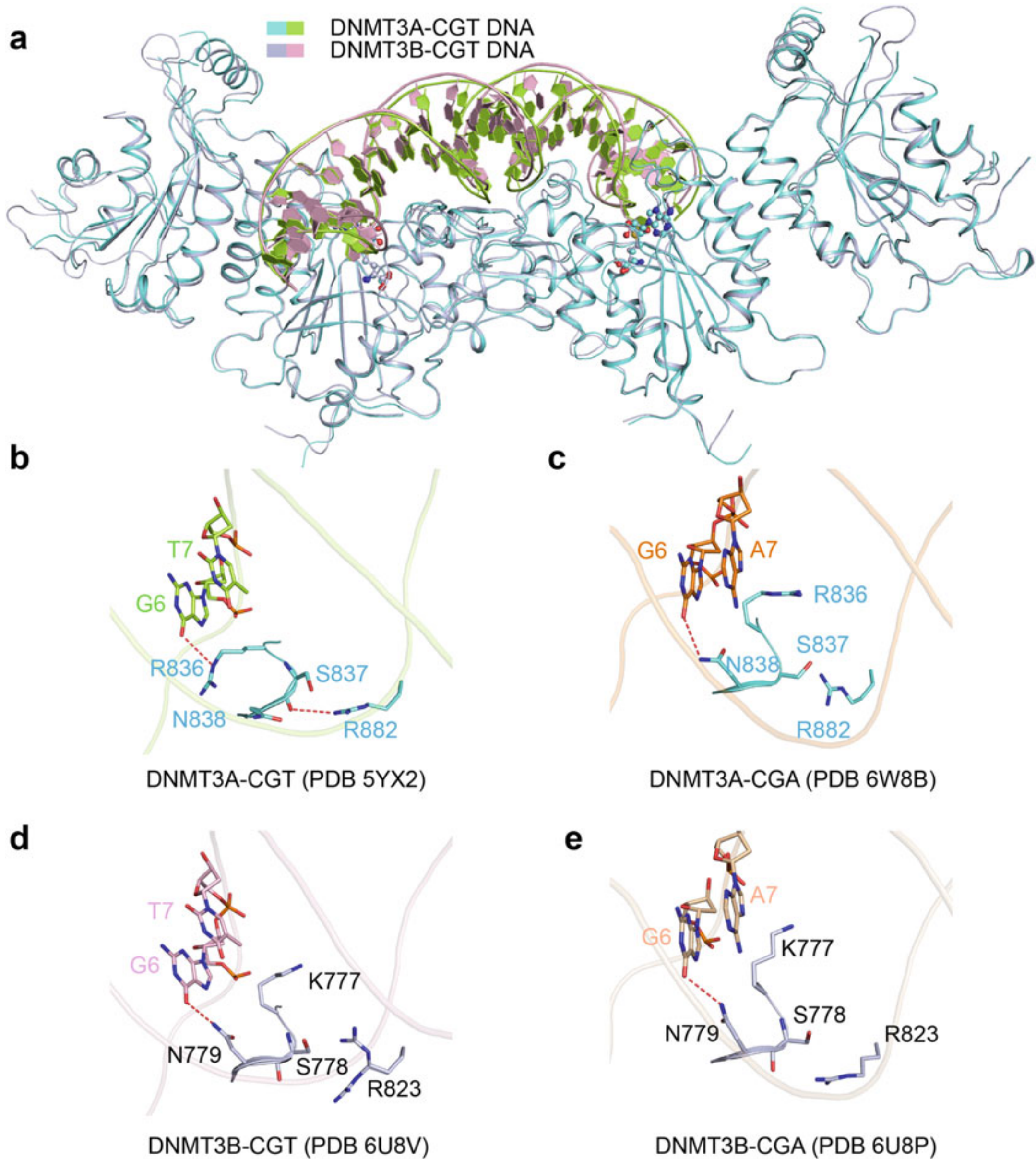


Fig. 3.4 Dnmt3a and Dnmt3b show similar but distinct substrate recognition mechanisms. (**a**) Structural overlay of human DNMT3A and DNMT3B catalytic domains covalently bound to ZpGpT/ApCpG (denoted as CGT) DNA. The SAH molecules are shown in ball-and-stick representation. (**b**, **c**) Close-up view of the CpG-specific interaction by human DNMT3A TRD-loop residues in the context of CGT (**b**) and CGA (**c**) motif. The hydrogen bonds are shown as dashed lines. (**d**, **e**) Close-up view of the CpG-specific interaction by human DNMT3B TRD-loop residues in the context of CGT (**d**) and CGA (**e**) motif

DNMT3B N656 with DNMT3A I715. Consistent with these structural observations, biochemical and cellular analyses revealed distinct CpG specificity and flanking sequence preferences between Dnmt3a and Dnmt3b (Gao et al. 2020b; Lee et al. 2017; Lin et al. 2002; Lister et al. 2009, 2013; Wienholz et al. 2010), providing an explanation to their overlapped but distinct functionality during development (Okano et al. 1999). For instance, the fact that Dnmt3b prefers a purine,

whereas Dnmt3a prefers a pyrimidine, in the +1-flanking site explains why Dnmt3b-associated ICF mutations lead to pronounced hypomethylation of satellite II repeat, which is enriched with the CGA motif (Prosser et al. 1986).

The activity of Dnmt3a is also regulated by isoform 3 of Dnmt3b (Dnmt3b3), an inactive form of Dnmt3b, in somatic cells (Duymich et al. 2016). The cryo-EM structure of Dnmt3a2 in complex with Dnmt3b3 and nucleosome reveals that the C-terminal domain of Dnmt3b3 associates with the catalytic domain of Dnmt3a in a similar manner as the Dnmt3a-Dnmt3l complex (Xu et al. 2020). Furthermore, an interaction was identified between Dnmt3b3 and the acidic patch of the nucleosome, which regulates the Dnmt3a2-mediated DNA methylation in cells (Xu et al. 2020).

3.2.4 Functions of Other Regions

An N-terminal sequence upstream of the PWWP domain, present in Dnmt3a1 but not in the Dnmt3a2 isoform, strongly binds to DNA. This contributes to the DNA methylation activity and localization of the enzyme in nuclei (Suetake et al. 2011). As mentioned earlier, this region also contains a UDR domain that binds to H2AK119ub, which mediates the recruitment of Dnmt3a1 to H2AK119Ub-decorated genomic regions (Weinberg et al. 2021). The N-terminal sequence of Dnmt3b, which exhibits no homology with that of Dnmt3a, binds to centromere protein C (CENP-C). CENP-C is a constitutive centromere component and is necessary for mitosis. It was proposed that CENP-C recruits Dnmt3b to both centromeric and pericentromeric satellite repeats to methylate these regions (Gopalakrishnan et al. 2009). Moreover, it was reported that an Arg residue in the N-terminal region of Dnmt3a undergoes citrullination by peptidylarginine deiminase 4 (PADI4), which stabilizes Dnmt3a and increases the DNA methylation level of the promoter of the *p21* gene (Deplus et al. 2014). Moreover, Dnmt3b binds to NEDD8 (neuronal precursor cell-expressed developmentally down-regulated protein 8), which is a small ubiquitin-like protein, through the region between the ADD and catalytic domains. NEDD8-modified Cullin 4A (CUL4A), which is essential for repressive chromatin formation, binds to Dnmt3b as well (Shamay et al. 2010).

3.2.5 Factors That Guide Dnmt3 to the Regions to Be Methylated

There have been several reports on the factors bringing Dnmt3 enzymes to specific sequences such as gene promoters. This mechanism is supported by the observation that a short DNA sequence (methylation-determining region, MDR) can determine the DNA methylation state (Lienert et al. 2011). Sequence-specific DNA-binding proteins may recognize such a sequence. For example, Dnmt3a binds to the corepressor complex of PR48 (regulatory subunit of protein phosphatase 2A)/HDAC1 (histone deacetylase 1) or proto-oncogene c-Myc through the ADD domain (Brenner et al. 2005; Fuks et al. 2001). Dnmt3b is reported to be tethered to the centromeric and pericentromeric heterochromatin regions through interaction with CENP-C to methylate the regions (Gopalakrishnan et al. 2009). Both Dnmt3a and Dnmt3b cooperate with EVI1 (oncogene product) to bind and methylate the expression-controlling region of miRNA 124–3 (Senyuk et al. 2011). Moreover, it was reported that noncoding RNA is involved in the targeting of Dnmt3b to de novo methylation sites. Promoter-associated RNA (pRNA), which binds the promoter of rRNA coding genes and forms a DNA/RNA triplex, recruits Dnmt3b to its target regions (Schmitz et al. 2010). However, it was also reported that the DNA/RNA heteroduplex rather inhibits the de novo methylation activities of both Dnmt3a and Dnmt3b in vitro (Ross et al. 2010).

In addition to the direct interaction with a DNA-binding protein or RNA, indirect interaction with the factors that bind to sequence-specific DNA-binding proteins has been reported. The Krüppel-associated box (KRAB) zinc finger

protein family, which determines target regions for methylation, comprising more than 300 genes (Liu et al. 2013), is an example. ZFP57, a KRAB zinc finger protein, binds to DNA in a sequence-specific manner and plays crucial roles in the establishment and maintenance of the methylation of imprinted genes through interaction with Tripartite motif containing 28 (Trim28, a.k.a. KAP1 or TIF1β) (Quenneville et al. 2011, 2012). Trim28 interacts with Dnmt3a, Dnmt3b, and Dnmt1 (Zuo et al. 2012) and acts as a scaffold to guide Dnmts to a variety of target sequences utilizing sequence-specific binding of KRAB zinc finger proteins. As a similar example, NEDD8, which is a ubiquitin-like small protein modifier, acts as a tag in guiding Dnmt3b to NEDDylated proteins (Shamay et al. 2010). The main target of NEDDylation is Cullin, which plays a role in heterochromatin formation.

However, the recruitment of Dnmt3a to specific genomic regions does not always introduce DNA methylation. Although Dnmt3a is recruited to a target sequence by Ezh2 (enhancer of zeste homolog 2), a component of polycomb repressive complex 2 (PRC2) (Rush et al. 2009); MBD3 (methyl-CpG binding domain protein 3), an intrinsic component of corepressor complex NuRD (nucleosome remodeling deacetylase); Brg1 (Brahma-related gene-1), an ATPase subunit of Swi/Snf chromatin remodeling factor (Datta et al. 2005); or p53 (Wang et al. 2005), this recruitment does not affect the DNA methylation state of the target regions.

3.2.6 Correlation Between de novo DNA Methylation and Histone Modifications

The histone tail modifications directly recruit de novo-type Dnmt3a or Dnmt3b to the site of DNA methylation. As described above, the PWWP domain of Dnmt3a recognizes H3K36me2/H3K36me3 to enhance the DNA methylation activity (Dhayalan et al. 2010; Weinberg et al. 2019), and the ADD domain binds H3K4me0 (Li et al. 2011; Otani et al. 2009) to enhance the DNA methylation activity (Li et al. 2011). The histone H3 tail with K4me3 inhibits DNA methylation by Dnmt3a (Li et al. 2011; Zhang et al. 2010), protecting H3K4me3-rich regions from DNA methylation. Dnmt3l, a member of the Dnmt3 family with no methylation activity, also contains an ADD domain and recognizes H3K4me0 (Ooi et al. 2007), as described above. H3K4me0 recruits and activates the Dnmt3a and Dnmt3l de novo methyltransferase complex to methylate the genome. In addition, the PWWP domains of Dnmt3a and Dnmt3b are reported to be a motif for DNA binding (Purdy et al. 2010; Qiu et al. 2002) and bringing Dnmt3a or Dnmt3b to heterochromatin (Chen et al. 2004; Ge et al. 2004). Thus, the PWWP in the amino-terminal half of Dnmt3a or Dnmt3b is one of the determinants of methylation-site targeting. Trim28, which is reported to interact directly with Dnmt3a (Zuo et al. 2012), also interacts with Setdb1, a histone H3K9 methyltransferase, and HP1 (Matsui et al. 2010), which recognizes H3K9me2/3.

3.3 Enzymes Responsible for the Maintenance of DNA Methylation Patterns

Dnmt1 is mainly responsible for maintaining DNA methylation patterns during replication or after DNA damage repair. Dnmt1 is a large molecule, comprising ~1,600 amino acid residues. Dnmt1 is composed of several domains: the N-terminal independently folded domain (NTD), replication foci-targeting sequence (RFTS) domain, CXXC domain, two bromo-adjacent-homology (BAH1 and BAH2) domains, and the catalytic domain (Fig. 3.1b). The domains are folded almost independently and interact with each other to form a functional DNA methyltransferase. The three-dimensional structures of mouse and human Dnmt1 with all the domains except for the NTD have been reported (Takeshita et al. 2011; Zhang et al. 2015).

3.3.1 NTD

The NTD of mouse Dnmt1 comprising amino acids (aa) 1–248 folds independently (Suetake et al. 2006). This domain functions as a binding platform for the factors that regulate the Dnmt1 function. The 1–118 aa sequence in the NTD, which is a typical coiled-coil structure and is lacking in the oocyte-specific Dnmt1 isoform (Gaudet et al. 1998; Mertineit et al. 1998), binds Dnmt1 associated protein 1 (DMAP1), which is a factor that represses transcription by cooperating with histone deacetylase HDAC2. DMAP1 binds to Dnmt1 at replication foci to assist the maintenance of the heterochromatin state as well (Rountree et al. 2000).

Proliferating cell nuclear antigen (PCNA), which binds DNA polymerase δ and other factors related to replication, is a prerequisite factor for replication. PCNA binds to the 160–178 aa sequence of mouse Dnmt1 (Chuang et al. 1997; Jimenji et al. 2019). The binding helps Dnmt1 maintain the methylation profile of the daughter DNA (Chuang et al. 1997) and recruits Dnmt1 to replication foci at the early and middle stages of the S-phase (Egger et al. 2006; Schermelleh et al. 2007). Therefore, it is thought to be involved in the replication-dependent DNA methylation process. However, the NTD domain containing the PCNA-binding motif is dispensable for the maintenance of the differentially methylated regions (DMRs) of imprinted genes, at least in ES cells (Garvilles et al. 2015). The cell-cycle regulating Rb protein is also reported to bind to the NTD (Robertson et al. 2000).

Interestingly, many epigenetic factors that may contribute to the formation and maintenance of heterochromatin are reported to bind to the NTD. De novo-type DNA methyltransferases Dnmt3a and Dnmt3b (Kim et al. 2002), heterochromatin-binding protein beta (HP1 beta) that selectively recognizes H3K9me2/3 (Fuks et al. 2003), and G9a that specifically methylates H3K9 (Esteve et al. 2006) bind to the NTD. All these interacting factors are related to the formation of heterochromatin, indicating that maintenance-type DNA methyltransferase Dnmt1 is tightly linked to histone methylation modification.

Although its function is not known, the NTD binds to cyclin-dependent kinase-like 5 (CDKL5) (Kameshita et al. 2008) and casein kinase (Sugiyama et al. 2010) and undergoes phosphorylation. The CDKL5 is reported to be a causative kinase for Rett syndrome. Rett syndrome is known to be caused mainly by a mutation in the *MeCP2* gene, of which the translation product specifically binds to methylated DNA and is a component of the corepressor complex. An impairment of the interaction between Dnmt1 and CDKL5 may contribute to the pathogenic process of Rett syndrome (Kameshita et al. 2008). Casein kinase 1 also interacts with the NTD. Phosphorylation with casein kinase 1 inhibits the DNA-binding activity of the NTD (Sugiyama et al. 2010). The function of the N-terminal region, which is a platform for the regulatory factors of Dnmt1, also seems to be regulated by different types of kinases (Esteve et al. 2011; Lavoie et al. 2011; Lavoie and St-Pierre 2011).

In addition, the NTD contains the DNA-binding 119–197 aa sequence, which overlaps with the PCNA-binding motif. The sequence contains an AT-hook-like motif and binds to the minor groove of AT-rich DNA. The DNA binding competes with the PCNA binding. Arg 133 and 136 in the sequence are crucial for the DNA-binding activity (Suetake et al. 2006). It has been proposed that this DNA-binding activity of the N-terminal domain contributes to the localization of Dnmt1 to AT-rich genome regions such as *Line1*, satellite, and the promoter of tissue-specific silent genes to maintain the fully methylated state of the repaired region that is hemimethylated (Suetake et al. 2006).

After the NTD, a flexible linker follows. Partial digestion with proteases can release the NTD 1–248 aa and the C-terminal part 291–1620 aa sequences (Suetake et al. 2006). According to the crystal structure of mouse Dnmt1 291–1620 aa, the structure of the RFTS domain has been determined after Pro 357 (Takeshita et al. 2011). The sequence starting from 249 to 356 aa seems to be a flexible region lacking an ordered structure. It has been reported that deletion of this region from Dnmt1 decreases maintenance methylation of the

genome (Borowczyk et al. 2009). However, it has recently been reported that even with deletion of the entire NTD including this region, Dnmt1 is fully active as a maintenance methyltransferase, at least in ES cells (Garvilles et al. 2015). The 1–353 sequence, which contains the NTD and the linker, binds to un-methylated DNA with CpG (Fatemi et al. 2001). However, as described above, the NTD also contains a DNA-binding domain, which exhibits a preference not for the CpG sequence but for an AT-rich (Suetake et al. 2006). The function of this linker is ambiguous at this moment.

3.3.2 RFTS Domain

The RFTS domain follows the NTD. This domain is necessary for Dnmt1 localization at the replication region at the late S-phase (Leonhardt et al. 1992). This recruitment depends on the tethering of Uhrf1 (ubiquitin-like with PHD and ring finger domains 1) to the hemimethylated DNA that appears after replication, and it is a prerequisite event for the replication-dependent maintenance of DNA methylation (Bostick et al. 2007; Sharif et al. 2007). Uhrf1 selectively binds to hemimethylated DNA through the SET and RING-associated (SRA) domain (Arita et al. 2008; Avvakumov et al. 2008; Hashimoto et al. 2008), to which the RFTS domain of Dnmt1 directly binds (Bashtrykov et al. 2014a; Berkyurek et al. 2014). Direct interaction of the RFTS domain with the SRA domain accelerates the hemimethylated DNA accession to the catalytic center. The SRA of Uhrf1 and Dnmt1 cannot bind to the same CpG site at the same time due to steric hindrance (Arita et al. 2008; Song et al. 2012). This clearly indicates that there must be a mechanism to hand the hemimethylated CpG from the SRA domain over to the catalytic center of Dnmt1, which may be involved in the direct interaction between the RFTS and SRA domains. How the hemimethylated CPG is transferred from the SRA domain to the catalytic center of Dnmt1 remains unclear.

The structure of the human RFTS domain itself has been elucidated (Syeda et al. 2011), exhibiting a two-lobe fold that is almost identical to that in the catalytically active mouse Dnmt1 (Takeshita et al. 2011) and human DNMT1 (Zhang et al. 2015). The position of the RFTS domain in the catalytically active Dnmt1 is intriguing. Since the RFTS domain is inserted into the catalytic pocket, a substrate DNA cannot gain access to the catalytic center due to steric hindrance. The position of the RFTS domain is stabilized by hydrogen bonds between the RFTS and catalytic domains. When the substrate DNA is short, DNA methylation activity is inhibited due to the positioning of the RFTS domain (Bashtrykov et al. 2014b; Berkyurek et al. 2014; Syeda et al. 2011). Surprisingly, even if the RFTS domain occupies the catalytic pocket, Dnmt1 can methylate DNA when it is longer than 12 bp and a length of about 30 bp is necessary for its full activity (Berkyurek et al. 2014). When the substrate DNA size is 12 bp, which is exactly the size that fits into the catalytic pocket of Dnmt1 (Song et al. 2012), Dnmt1 cannot methylate substrate DNA. The DNA methylation activity of Dnmt1 that lacks the RFTS domain toward short hemimethylated DNA is efficiently inhibited by ectopically added RFTS domain (Berkyurek et al. 2014; Syeda et al. 2011). Since the full DNA methylation activity is acquired when the substrate DNA is longer than 30 bp, the catalytic domain of Dnmt1 may increase the DNA-binding affinity by two DNA-binding sites to trigger the removal of the RFTS domain from the catalytic pocket.

Amino acid residues Lys 23 (Nishiyama et al. 2013), Lys 14 and Lys18 of histone H3 (Ishiyama et al. 2017; Qin et al. 2015) are reported to be ubiquitylated. These modifications cooperate with the H3K9me3 mark to interact with the RFTS domain, thereby modulating maintenance methylation by Dnmt1 (Fig. 3.5a) (Ishiyama et al. 2017; Ren et al. 2020). The ubiquitin molecules bind to the N-terminal subdomain of the RFTS domain, while the H3 residues occupy the cleft between the N- and C-lobe, through eviction of the linker C-terminal to the RFTS domain (Fig. 3.5a). Binding to the H3K9me3 mark is further stabilized by the site corresponding to human DNMT1 W465, along with the residues

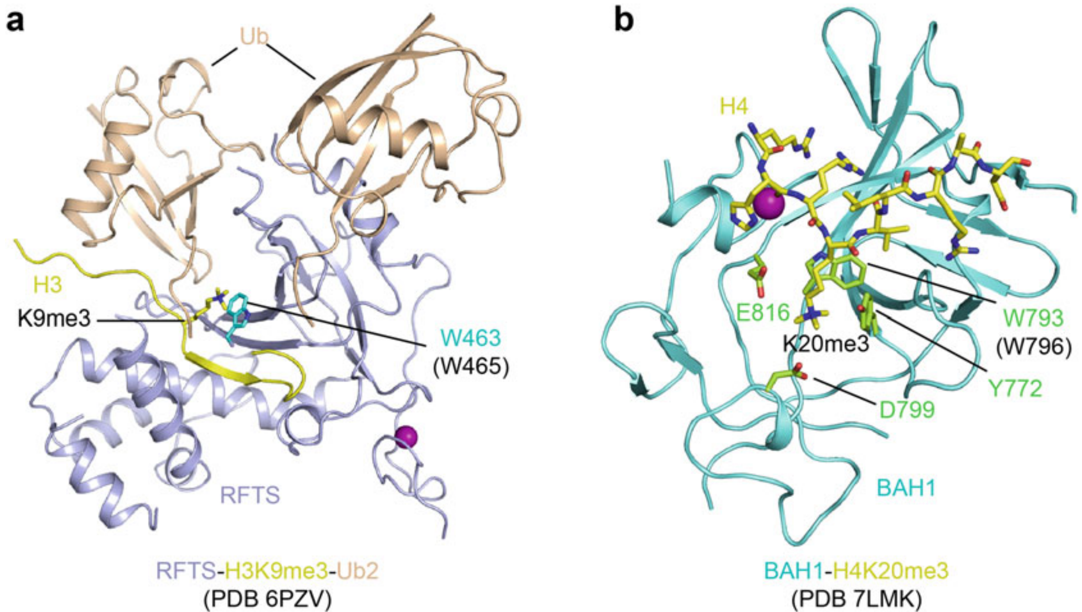


Fig. 3.5 Recognition of histone marks by Dnmt1 RFTS and ADD domains (**a**) Ribbon diagram of bovine Dnmt1 RFTS domain (slate) bound to a H3K9me3 peptide (yellow) and two ubiquitin molecules (wheat), with the H3K9me3-binding W463 shown in stick representation. The W463-corresponding site in human DNMT1 (W465) is shown in parathesis. (**b**) Ribbon diagram of bovine Dnmt1 BAH1 domain (cyan) bound to a H4K20me3 peptide (yellow), with the H4K20me3-binding cage residues shown in stick representation. The W793-corresponding site in human DNMT1 (W796) is shown in parenthesis. The zinc ions are shown as purple spheres in (**a**) and (**b**)

from one of the bound ubiquitin molecules (Fig. 3.5a). Introducing H3K9me3/H3 ubiquitylation binding-defective mutations to human DNMT1 led to a severe loss of genomic methylation and impairment of genome stabilization (Ren et al. 2020). In addition, Uhrf1 ubiquitylates lys 15 and lys 24 of PAF15 (PCNA-associated factor 15) during the early S-phase (Gonzalez-Magana et al. 2019; Nishiyama et al. 2020). The PAF15 protein with dual mono-ubiquitylation in turn recruits Dnmt1 to the replication foci via an interaction with the RFTS domain, an event important for the maintenance DNA methylation at early replicating domains (Gonzalez-Magana et al. 2019; Nishiyama et al. 2020).

Interestingly, the ring finger motif of Uhrf1, which is a prerequisite factor for replication-dependent maintenance methylation, is involved in the ubiquitylation of histones and DNMT1 as an E3 ligase (Du et al. 2010; Nishiyama et al. 2013; Qin et al. 2015). The tandem Tudor domain and the PHD finger of Uhrf1 recognize H3K9me3 and H3R2me0 (Arita et al. 2012), as well as a C-terminal poly-basic region of Uhrf1 (Fang et al. 2016; Gao et al. 2018; Gelato et al. 2014). Mutations within the tandem Tudor domain of UHRF1, which inhibit the recognition of H3K9me3, partly inhibit the maintenance DNA methylation (Rothbart et al. 2012; Zhao et al. 2016b), again indicating the cross-talk between DNA methylation and histone modification.

Following the RFTS domain, there are three residues, Phe 631, 634, and 635 (numbering based on mouse Dnmt1), in an alpha-helix structure interacting with Tyr 1243 and Phe 1246, which are adjacent to the PCQ loop in catalytic domain motif IV, of which the Cys residue covalently binds to the target cytosine at the sixth carbon. The interactions pull the PCQ loop toward the DNA-binding pocket (Takeshita et al. 2011).

3.3.3 CXXC

The CXXC domain contains two zinc atoms forming zinc finger motifs, which are known to bind DNA-containing un-methylated CpG. This motif is conserved among Dnmt1, mammalian trithorax-group protein, myeloid/lymphoid leukemia (MLL) (Cierpicki et al. 2010), CXXC-type zinc finger protein 1 (CXXC1) (Voo et al. 2000), methyl-CpG-binding protein 1 (MBD1) (Cross et al. 1997), and other proteins (Long et al. 2013). The CXXC domain of Dnmt1 contains two C4-type zinc fingers. The backbone structure of the CXXC domain does not change even when the RFTS domain is deleted ((Takeshita et al. 2011); Hashimoto et al., PDB accession number 3SWR), or the CXXC is bound to un-methylated DNA (Song et al. 2011).

The CXXC domain was initially proposed to be essential for the DNA methylation activity of Dnmt1 (Pradhan et al. 2008). However, this notion was later challenged by observations that removal of the CXXC domain does not substantially impair the Dnmt1 activity on DNA (Frauer et al. 2011; Song et al. 2011, 2012). When the RFTS domain is deleted, the autoinhibitory linker between the CXXC and BAH1 domains falls into the catalytic pocket, blocking the DNA from accessing the active site (Song et al. 2011). Song et al. proposed that binding of the CXXC domain to un-methylated DNA is a mechanism to inhibit its accession to the catalytic center of Dnmt1 and thus limits Dnmt1 from de novo methylation. This autoinhibitory mechanism cooperates with the intrinsic substrate specificity of the catalytic domain (Song et al. 2012) to modulate Dnmt1-mediated maintenance DNA methylation. This model predicts that deletion or mutation of the CXXC domain would increase de novo-type methylation activity. Although this hypothesis was supported by the observations for the RFTS-removed fragment (Song et al. 2011), full-length Dnmt1 with mutations in the CXXC domain did not show appreciable change in the specificity toward hemimethylated DNA in vitro ((Bashtrykov et al. 2012); Suetake, unpublished observation). Reconstitution of mouse *Dnmt1*$^{-/-}$ ESCs with Dnmt1 containing CXXC mutations led to a similar genomic DNA methylation level to that with wild-type Dnmt1 (Frauer et al. 2011), likely due to the redundancy of Dnmt1 regulation. At present, the autoinhibition mechanism involving the CXXC domain to prevent de novo methylation (Song et al. 2011) awaits further investigation.

The autoinhibitory linker assumes a helical structure in DNA-free Dnmt1 (Takeshita et al. 2011; Zhang et al. 2015), but becomes extended in the RFTS-deleted Dnmt1 ((Song et al. 2011); Hashimoto et al., PDB accession number 3SWR). A mutation or deletion of this linker changes Dnmt1 into an extended conformation and enhances the DNA methylation activity toward 12 bp DNA (Zhang et al. 2015). Since such a short DNA cannot be methylated by Dnmt1 in the absence of the SRA domain of Uhrf1 (Berkyurek et al. 2014), it is reasonable to assume that this region plays a crucial role in the release of the RFTS domain from the catalytic pocket.

3.3.4 Two BAH Domains

The CXXC domain is followed by two tandem BAH domains. The BAH domains consist of a beta-sheet core and are functionally correlated to chromatin processes. The BAH domains from many other proteins, including the "remodels the structure of complex" component RSC2 (Chambers et al. 2013), Silent information regulator 3 (Sir3) (Armache et al. 2011; Arnaudo et al. 2013; Yang et al. 2013), the origin recognition complex subunit 1 (ORC1) (Kuo et al. 2012), BAH domain and coiled-coil containing 1 (BAHCC1) (Fan et al. 2020), BAHD1 (Fan et al. 2021; Zhao et al. 2016a), SHORT LIFE (SHL) (Qian et al. 2018), EARLY BOLTING IN SHORT DAY (EBS) (Li et al. 2018b; Yang et al. 2018), and anti-silencing 1 (ASI1)-IMMUNOPRECIPITATED PROTEIN 3 (AIPP3) (Zhang et al. 2020), interact with nucleosomes with various histone modifications. The two BAH domains of Dnmt1 are connected through an alpha-helix, which is dumbbell shaped

(Song et al. 2011; Takeshita et al. 2011). At the end of the BAH1 domain, just before the helix linker, there is a zinc finger motif which stabilizes DNMT1.

The BAH1 domain of Dnmt1 binds to histone H4K20me3 via a hydrophobic cage (Fig. 3.5b) (Ren et al. 2021). In the apo form of Dnmt1, this hydrophobic cage is shielded by the autoinhibitory linker. The interaction between the BAH1 domain and the H4K20me3 mark causes the displacement of the autoinhibitory linker, which in turn allosterically activates Dnmt1 (Ren et al. 2021). Single-molecule fluorescence resonance energy transfer (smFRET) analysis further indicated that the histone interactions of the BAH1 and RFTS domains both lead to enhanced conformational dynamics between the RFTS domain and the catalytic domain (Ren et al. 2021). Mutation of the hydrophobic cage residue W796 to alanine in human DNMT1 results in a H4K20me3 binding-defective but hyperactive enzyme, leading to DNA hypomethylation at the H4K20me3-decorated regions (e.g. Line1), but enhanced methylation at regions that lack H4K20me3 (Ren et al. 2021). Interestingly, the BAH1 W796A mutation can partially restore the DNA methylation that was reduced by the RFTS W465A mutation, raising a notion of the functional cooperation between the BAH1 and RFTS domains (Ren et al. 2021). Consistently, cells transfected with human DNMT1 containing W465A/W796A double mutation appear less sensitive to ionization radiation than those with DNMT1 W465A single mutation (Ren et al. 2021).

The BAH2 domain possesses a long loop protruding from its body, of which the distal end interacts with the TRD in the catalytic domain, and adjacent residues interact directly with the substrate DNA (Song et al. 2012). The function of BAH2 remains elusive, although evidence suggests that it may regulate Dnmt1-mediated de novo DNA methylation in vivo (Yarychkivska et al. 2018a).

The KG-repeat between the BAH2 and the catalytic domain is conserved among species (Kimura et al. 1996; Tajima et al. 1995). This repeat was observed to interact with ubiquitin-specific protease 7 (USP7), which is a deubiquitinating enzyme (Qin et al. 2011). This interaction increases DNA methylation activity possibly through stabilizing Dnmt1 (Cheng et al. 2015) or deubiquitylation of H3 (Yamaguchi et al. 2017). Acetylation of the Lys residues in the KG-repeat impairs the Dnmt1-USP7 interaction and promotes degradation of Dnmt1. On the other hand, a separate study showed that the Dnmt1-USP7 interaction unlikely plays a major role in the stabilization of Dnmt1 in somatic cells (Yarychkivska et al. 2018b).

3.3.5 Catalytic Domain

Similar to other Dnmts, the ten motifs characteristic of DNA-(cytosine C5)-methyltransferases are conserved in the catalytic domain of Dnmt1. The DNA methylation mechanism of Dnmt1 is assumed to be identical to that of M.HhaI (Kumar et al. 1994). However, different from in M.HhaI (Cheng et al. 1993), the position of the side chain of Cys in the PCQ loop, which is expected to form a covalent bond with the sixth carbon of the target cytosine base (Song et al. 2012), turns toward target cytosine on the addition of methyl-group donor *S*-adenosyl-L-methionine (SAM) even in the absence of DNA (Takeshita et al. 2011). The side chain of the Cys faces away when SAM is catabolized to *S*-adenosyl-L-homocysteine (SAH) after the transfer of a methyl group in mouse Dnmt1. Interestingly, the side chain of the Cys in the PCQ loop of human DNMT1 does not completely face away even in the SAH-binding form (Zhang et al. 2015). The effect of this difference between the mouse and human enzymes remains to be determined.

The TRD in the catalytic domain of Dnmt1 is exceptionally long compared to those in other DNA methyltransferases. The TRD covers the hemimethylated DNA and holds the methylated cytosine through hydrophobic interactions (Fig. 3.6a) (Song et al. 2012). The target cytosine in the hemimethylated CpG is flipped out and inserts into the active site of Dnmt1 (Fig. 3.6a). According to the three-dimensional structure of

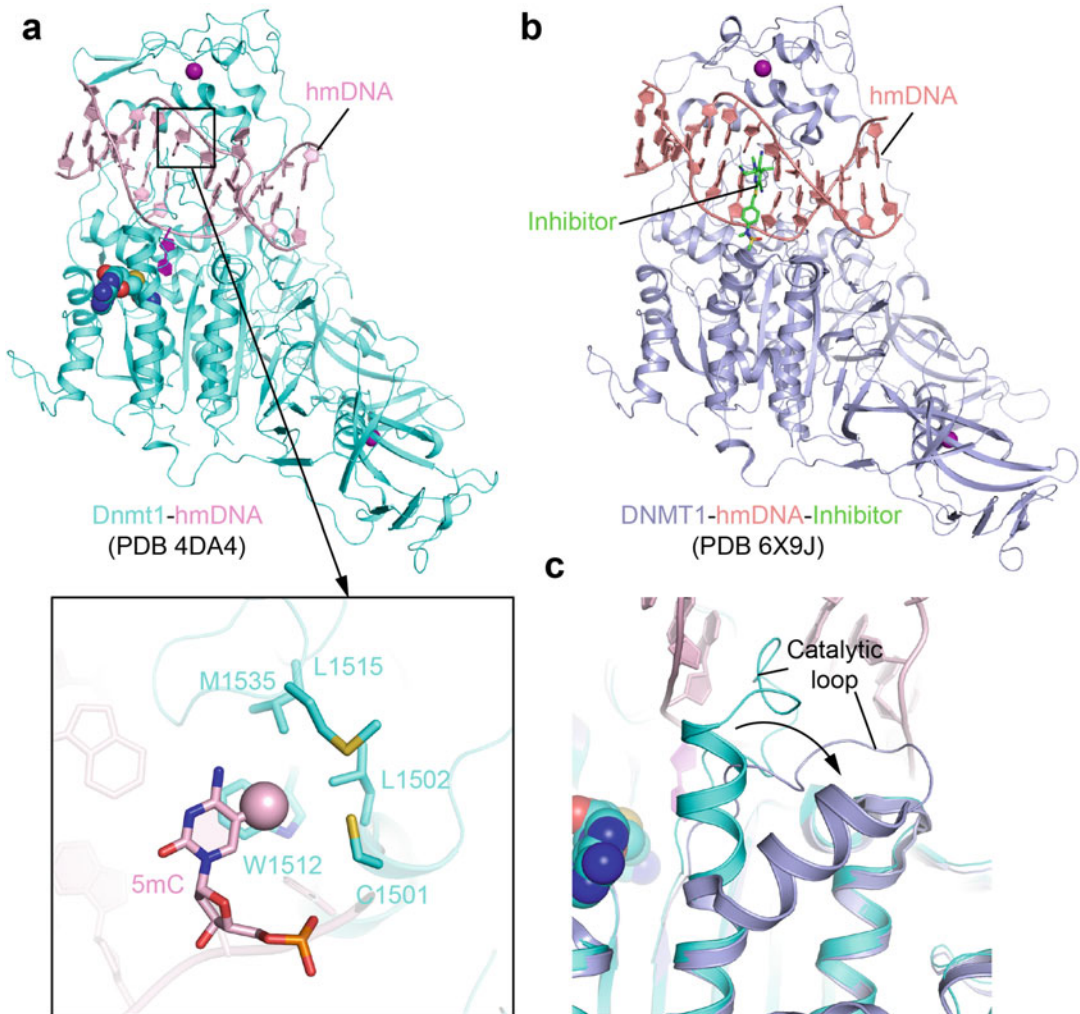


Fig. 3.6 Structure, mechanism and inhibition of Dnmt1-mediated maintenance DNA methylation. (**a**) Structure of mouse Dnmt1 C-terminal fragment (residues 731-1602) bound to hemimethylated CpG DNA (hmDNA), with flipped 5-fluorocytosine colored in purple. The van der Waals contacts between the 5-methyl group (sphere representation) of 5mC in the template strand and Dnmt1 residues are shown in expanded view. (**b**) Structure of human DNMT1 C-terminal fragment (residues 729–1600) bound to hemimethylated CpG DNA and inhibitor GSK3830052. (**c**) Comparison of the active site conformation between the enzymatically active (PDB 4DA4) and inactive (PDB 6X9J) complex. The conformational shift of the helix C-terminal to the catalytic loop is indicated by arrow

the complex with hemimethylated DNA and the DNA methylation activity of the truncated Dnmt1, the recognition and selective methylation of hemimethylated DNA is at least in part underpinned by the catalytic domain (Bashtrykov et al. 2012; Song et al. 2012). A reversible small molecule inhibitor was identified to inhibit Dnmt1 activity through both DNA intercalation and its interaction with Dnmt1 (Fig. 3.6b) (Pappalardi et al. 2021). The transition between catalytically active and inactive states of Dnmt1 is accompanied by a straight-to-kinked switch of the alpha-helix following the catalytic loop (Fig. 3.6c) (Pappalardi et al. 2021; Song et al. 2011, 2012; Ye et al. 2018).

In addition to its regulatory domains, the enzymatic activity of Dnmt1 is fine-tuned by the flanking sequence of the hemimethylated CpG sites (Adam et al. 2020). Structural comparison of Dnmt1 in complex with DNAs containing hemimethylated *G*CG, *A*CG, and *C*CG motifs (underlined are the CpG-flanking nucleotides) reveals distinct base-flanking mechanisms (Adam et al. 2020), which presumably causes the differential methylation activity of Dnmt1 on these substrates and impacts the dynamic landscape of DNA methylation in health and disease.

3.4 Cross-Talk Between De Novo-Type and Maintenance-Type DNA Methyltransferases

Establishment of DNA methylation patterns is mainly performed by de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and their maintenance during replication is carried out by Dnmt1, as described above. However, it has been reported that Dnmt3a and/or Dnmt3b are also necessary for maintaining the methylation of repeat elements (Liang et al. 2002). In *Dnmt3a* and *Dnmt3b* double-knockout ES cells, DNA methylation gradually decreased during culture (Chen et al. 2003). A similar decrease in DNA methylation has been observed in mouse embryonic fibroblasts after *Dnmt3b* deletion (Dodge et al. 2005). These reports indicate that not only Dnmt1 but also de novo-type DNA methyltransferases Dnmt3a and/or Dnmt3b contribute to the maintenance DNA methylation. There has been a report that Dnmt3a and Dnmt3b interact with Dnmt1 at the NTD (Kim et al. 2002). It is unlikely, however, that Dnmt3a and Dnmt3b coexist with Dnmt1 at replication foci, since Dnmt1 is loaded at an early stage of replication, and Dnmt3a and Dnmt3b at a rather late stage of replication (Alabert et al. 2014). Therefore, the molecular mechanism of the cooperation with de novo-type Dnmts in maintenance DNA methylation remains to be determined.

As for the establishment of DNA methylation patterns, it was expected that Dnmt1 exhibits de novo methylation activity in vivo (Christman et al. 1995). Actually, Dnmt1 exhibits a significant level of de novo-type DNA methylation activity in vitro (Fatemi et al. 2001; Vilkaitis et al. 2005) and ex vivo (Biniszkiewicz et al. 2002; Haggerty et al. 2021; Li et al. 2018a; Takagi et al. 1995; Vertino et al. 1996; Wang et al. 2020; Yarychkivska et al. 2018a). In *Dnmt3a* and *Dnmt3b* knockout ES cells, ectopically introduced DNA (Lorincz et al. 2002) as well as endogenous regions (Arand et al. 2012) undergo de novo DNA methylation. Dnmt1 apparently favors de novo methylation near preexisting methylation sites (Arand et al. 2012; Vilkaitis et al. 2005). Therefore, although its physiological meaning is elusive, Dnmt1 also causes de novo DNA methylation in vivo. The cross-talk of de novo and maintenance DNA methylations is discussed in broader context in Jones and Liang (Jones and Liang 2009) and Jeltsch and Jurkowska (Jeltsch and Jurkowska 2014).

3.5 Conclusions and Perspective

Elucidation of the domain structures of Dnmts has provided important information in understanding the molecular mechanisms of DNA methylation. Indeed, the complexes of the ADD domain of Dnmt3a with histone H3, the PWWP domain of Dnmt3b with H3K36me3, the RFTS domain of Dnmt1 with H3K9me3 and H3 ubiquitylation, and the BAH1 domain of Dnmt1 with H4K20me3 illustrated their functions in the target recruitment and/or allosteric activation of the enzymes. Co-crystal structures of Dnmt3a with Dnmt3l and DNA and that of Dnmt1 with hemimethylated DNA have provided a clue to understand the DNA methylation mechanism. The domain rearrangement of Dnmt3a by histone H3 tail and occupation of the catalytic pocket of Dnmt1 by the RFTS domain have lifted the veils of DNA methylation tricks. In the near future, by utilizing the structural information, the biochemical approach with site-directed mutagenesis might provide further information in understanding molecular mechanisms of DNA methylation

regulation. To this end, we need more structural information including complexes with other factors.

In addition to the high-resolution crystal structures, NMR may possibly provide us with more dynamic structural information in solution, and analysis by single-particle cryogenic electron microscopy can be a powerful technology to analyze large complexes that may be involved in DNA methylation regulation in the chromatin environment.

References

Aapola U, Kawasaki K, Scott HS, Ollila J, Vihinen M, Heino M, Shintani A, Kawasaki K, Minoshima S, Krohn K et al (2000) Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics 65:293–298

Adam S, Anteneh H, Hornisch M, Wagner V, Lu J, Radde NE, Bashtrykov P, Song J, Jeltsch A (2020) DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. Nat Commun 11:3723

Alabert C, Bukowski-Wills JC, Lee SB, Kustatscher G, Nakamura K, de Lima Alves F, Menard P, Mejlvang J, Rappsilber J, Groth A (2014) Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components. Nat Cell Biol 16:281–293

Anteneh H, Fang J, Song J (2020) Structural basis for impairment of DNA methylation by the DNMT3A R882H mutation. Nat Commun 11:2294

Aoki A, Suetake I, Miyagawa J, Fujio T, Chijiwa T, Sasaki H, Tajima S (2001) Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases. Nucleic Acids Res 29:3506–3512

Arand J, Spieler D, Karius T, Branco MR, Meilinger D, Meissner A, Jenuwein T, Xu G, Leonhardt H, Wolf V et al (2012) In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. PLoS Genet 8:e1002750

Argentaro A, Yang JC, Chapman L, Kowalczyk MS, Gibbons RJ, Higgs DR, Neuhaus D, Rhodes D (2007) Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. Proc Natl Acad Sci U S A 104:11939–11944

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature 455:818–821

Arita K, Isogai S, Oda T, Unoki M, Sugita K, Sekiyama N, Kuwata K, Hamamoto R, Tochio H, Sato M et al (2012) Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1. Proc Natl Acad Sci U S A 109: 12950–12955

Armache KJ, Garlick JD, Canzio D, Narlikar GJ, Kingston RE (2011) Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 A resolution. Science 334:977–982

Arnaudo N, Fernandez IS, McLaughlin SH, Peak-Chew SY, Rhodes D, Martino F (2013) The N-terminal acetylation of Sir3 stabilizes its binding to the nucleosome core particle. Nat Struct Mol Biol 20:1119–1121

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, Arrowsmith CH, Dhe-Paganon S (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature 455:822–825

Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Herault Y, Guillou F, Bourc'his D (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354:909–912

Bashtrykov P, Jankevicius G, Smarandache A, Jurkowska RZ, Ragozin S, Jeltsch A (2012) Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. Chem Biol 19:572–578

Bashtrykov P, Jankevicius G, Jurkowska RZ, Ragozin S, Jeltsch A (2014a) The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. J Biol Chem 289:4106–4115

Bashtrykov P, Rajavelu A, Hackner B, Ragozin S, Carell T, Jeltsch A (2014b) Targeted mutagenesis results in an activation of DNA methyltransferase 1 and confirms an autoinhibitory role of its RFTS domain. Chembiochem: A Euro J Chem Biol 15:743–748

Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, Akalin A, Schubeler D (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520: 243–247

Berkyurek AC, Suetake I, Arita K, Takeshita K, Nakagawa A, Shirakawa M, Tajima S (2014) The DNA methyltransferase Dnmt1 directly interacts with the SET and RING finger-associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic center to hemi-methylated DNA. J Biol Chem 289:379–386

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

Biniszkiewicz D, Gribnau J, Ramsahoye B, Gaudet F, Eggan K, Humpherys D, Mastrangelo MA, Jun Z, Walter J, Jaenisch R (2002) Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality. Mol Cell Biol 22:2124–2135

Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16:6–21

Borowczyk E, Mohan KN, D'Aiuto L, Cirio MC, Chaillet JR (2009) Identification of a region of the DNMT1 methyltransferase that regulates the maintenance of genomic imprints. Proc Natl Acad Sci U S A 106: 20806–20811

Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317:1760–1764

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints. Science 294:2536–2539

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

Chambers AL, Pearl LH, Oliver AW, Downs JA (2013) The BAH domain of Rsc2 is a histone H3 binding domain. Nucleic Acids Res 41:9168–9182

Chen T, Ueda Y, Dodge JE, Wang Z, Li E (2003) Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol 23:5594–5605

Chen T, Tsujimoto N, Li E (2004) The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Mol Cell Biol 24:9048–9058

Chen ZX, Mann JR, Hsieh CL, Riggs AD, Chedin F (2005) Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. J Cell Biochem 95: 902–917

Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ (1993) Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell 74:299–307

Cheng J, Yang H, Fang J, Ma L, Gong R, Wang P, Li Z, Xu Y (2015) Molecular mechanism for USP7-mediated DNMT1 stabilization by acetylation. Nat Commun 6:7023

Christman JK, Sheikhnejad G, Marasco CJ, Sufrin JR (1995) 5-Methyl-2'-deoxycytidine in single-stranded DNA can act in cis to signal de novo DNA methylation. Proc Natl Acad Sci U S A 92:7347–7351

Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science 277: 1996–2000

Cierpicki T, Risner LE, Grembecka J, Lukasik SM, Popovic R, Omonkowska M, Shultis DD, Zeleznik-Le NJ, Bushweller JH (2010) Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. Nat Struct Mol Biol 17:62–68

Cross SH, Meehan RR, Nan X, Bird A (1997) A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nat Genet 16:256–259

Datta J, Majumder S, Bai S, Ghoshal K, Kutay H, Smith DS, Crabb JW, Jacob ST (2005) Physical and functional interaction of DNA methyltransferase 3A with Mbd3 and Brg1 in mouse lymphosarcoma cells. Cancer Res 65:10891–10900

Deplus R, Blanchon L, Rajavelu A, Boukaba A, Defrance M, Luciani J, Rothe F, Dedeurwaerder S, Denis H, Brinkman AB et al (2014) Regulation of DNA methylation patterns by CK2-mediated phosphorylation of Dnmt3a. Cell Rep 8:743–753

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

Dodge JE, Okano M, Dick F, Tsujimoto N, Chen T, Wang S, Ueda Y, Dyson N, Li E (2005) Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. J Biol Chem 280: 17986–17991

Du Z, Song J, Wang Y, Zhao Y, Guda K, Yang S, Kao HY, Xu Y, Willis J, Markowitz SD et al (2010) DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination. Sci Signal 3:ra80

Dukatz M, Holzer K, Choudalakis M, Emperle M, Lungu C, Bashtrykov P, Jeltsch A (2019) H3K36me2/3 Binding and DNA Binding of the DNA Methyltransferase DNMT3A PWWP Domain Both Contribute to its Chromatin Interaction. J Mol Biol 431:5063–5074

Duymich CE, Charlet J, Yang X, Jones PA, Liang G (2016) DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. Nat Commun 7:11453

Egger G, Jeong S, Escobar SG, Cortez CC, Li TW, Saito Y, Yoo CB, Jones PA, Liang G (2006) Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. Proc Natl Acad Sci U S A 103:14080–14085

Esteve PO, Chin HG, Smallwood A, Feehery GR, Gangisetty O, Karpf AR, Carey MF, Pradhan S (2006) Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. Genes Dev 20:3089–3103

Esteve PO, Chang Y, Samaranayake M, Upadhyay AK, Horton JR, Feehery GR, Cheng X, Pradhan S (2011) A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. Nat Struct Mol Biol 18:42–48

Fan H, Lu J, Guo Y, Li D, Zhang ZM, Tsai YH, Pi WC, Ahn JH, Gong W, Xiang Y et al (2020) BAHCC1 binds H3K27me3 via a conserved BAH module to mediate gene silencing and oncogenesis. Nat Genet 52:1384–1396

Fan H, Guo Y, Tsai YH, Storey AJ, Kim A, Gong W, Edmondson RD, Mackintosh SG, Li H, Byrum SD et al (2021) A conserved BAH module within

mammalian BAHD1 connects H3K27me3 to Polycomb gene silencing. Nucleic Acids Res 49: 4441–4455

Fang J, Cheng J, Wang J, Zhang Q, Liu M, Gong R, Wang P, Zhang X, Feng Y, Lan W et al (2016) Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition. Nat Commun 7:11197

Fatemi M, Hermann A, Pradhan S, Jeltsch A (2001) The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. J Mol Biol 309:1189–1199

Frauer C, Rottach A, Meilinger D, Bultmann S, Fellinger K, Hasenoder S, Wang M, Qin W, Soding J, Spada F et al (2011) Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. PLoS One 6:e16627

Fuks F, Burgers WA, Godin N, Kasai M, Kouzarides T (2001) Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. EMBO J 20:2536–2544

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

Gao L, Tan XF, Zhang S, Wu T, Zhang ZM, Ai HW, Song J (2018) An Intramolecular Interaction of UHRF1 Reveals Dual Control for Its Histone Association. Structure 26(304-311):e303

Gao L, Anteneh H, Song J (2020a) Dissect the DNMT3A- and DNMT3B-mediated DNA Co-methylation through a Covalent Complex Approach. J Mol Biol 432:569–575

Gao L, Emperle M, Guo Y, Grimm SA, Ren W, Adam S, Uryu H, Zhang ZM, Chen D, Yin J et al (2020b) Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. Nat Commun 11:3355

Garvilles RG, Hasegawa T, Kimura H, Sharif J, Muto M, Koseki H, Takahashi S, Suetake I, Tajima S (2015) Dual Functions of the RFTS Domain of Dnmt1 in Replication-Coupled DNA Methylation and in Protection of the Genome from Aberrant Methylation. PLoS One 10:e0137509

Gaudet F, Talbot D, Leonhardt H, Jaenisch R (1998) A short DNA methyltransferase isoform restores methylation in vivo. J Biol Chem 273:32725–32729

Ge YZ, Pu MT, Gowher H, Wu HP, Ding JP, Jeltsch A, Xu GL (2004) Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. J Biol Chem 279:25447–25454

Gelato KA, Tauber M, Ong MS, Winter S, Hiragami-Hamada K, Sindlinger J, Lemak A, Bultsma Y, Houliston S, Schwarzer D et al (2014) Accessibility of different histone H3-binding domains of UHRF1 is allosterically regulated by phosphatidylinositol 5-phosphate. Mol Cell 54:905–919

Gonzalez-Magana A, de Opakua AI, Merino N, Monteiro H, Diercks T, Murciano-Calles J, Luque I, Bernado P, Cordeiro TN, Biasio A et al (2019) Double Monoubiquitination Modifies the Molecular Recognition Properties of p15(PAF) Promoting Binding to the Reader Module of Dnmt1. ACS Chem Biol 14:2315–2326

Gopalakrishnan S, Sullivan BA, Trazzi S, Della Valle G, Robertson KD (2009) DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. Hum Mol Genet 18:3178–3193

Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X, He S, Shi P, Dong L, Li G et al (2015) Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517:640–644

Haggerty C, Kretzmer H, Riemenschneider C, Kumar AS, Mattei AL, Bailly N, Gottfreund J, Giesselmann P, Weigert R, Brändl B et al (2021) Dnmt1 has de novo activity targeted to transposable elements. Nat Struct Mol Biol 28:594–603

Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM, Gartler SM (1999) The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci U S A 96:14412–14417

Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature 455: 826–829

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

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

Ishiyama S, Nishiyama A, Saeki Y, Moritsugu K, Morimoto D, Yamaguchi L, Arai N, Matsumura R, Kawakami T, Mishima Y et al (2017) Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance. Mol Cell 68(350-360):e357

Jain D, Meydan C, Lange J, Claeys Bouuaert C, Lailler N, Mason CE, Anderson KV, Keeney S (2017) rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. PLoS Genet 13:e1006964

Jeltsch A, Jurkowska RZ (2014) New concepts in DNA methylation. Trends Biochem Sci 39:310–318

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449: 248–251

Jimenji T, Matsumura R, Kori S, Arita K (2019) Structure of PCNA in complex with DNMT1 PIP box reveals the

basis for the molecular mechanism of the interaction. Biochem Biophys Res Commun 516:578–583
Jones PA, Liang G (2009) Rethinking how DNA methylation patterns are maintained. Nat Rev Genet 10:805–811
Jurkowska RZ, Rajavelu A, Anspach N, Urbanke C, Jankevicius G, Ragozin S, Nellen W, Jeltsch A (2011) Oligomerization and binding of the Dnmt3a DNA methyltransferase to parallel DNA molecules: heterochromatic localization and role of Dnmt3L. J Biol Chem 286:24200–24207
Kameshita I, Sekiguchi M, Hamasaki D, Sugiyama Y, Hatano N, Suetake I, Tajima S, Sueyoshi N (2008) Cyclin-dependent kinase-like 5 binds and phosphorylates DNA methyltransferase 1. Biochem Biophys Res Commun 377:1162–1167
Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature 429:900–903
Kim GD, Ni J, Kelesoglu N, Roberts RJ, Pradhan S (2002) Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. EMBO J 21:4183–4195
Kim SH, Park J, Choi MC, Kim HP, Park JH, Jung Y, Lee JH, Oh DY, Im SA, Bang YJ et al (2007) Zinc-fingers and homeoboxes 1 (ZHX1) binds DNA methyltransferase (DNMT) 3B to enhance DNMT3B-mediated transcriptional repression. Biochem Biophys Res Commun 355:318–323
Kimura H, Ishihara G, Tajima S (1996) Isolation and expression of a Xenopus laevis DNA methyltransferase cDNA. J Biochem 120:1182–1189
Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, Roberts RJ, Wilson GG (1994) The DNA (cytosine-5) methyltransferases. Nucleic Acids Res 22:1–10
Kuo AJ, Song J, Cheung P, Ishibe-Murakami S, Yamazoe S, Chen JK, Patel DJ, Gozani O (2012) The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. Nature 484:115–119
Lavoie G, St-Pierre Y (2011) Phosphorylation of human DNMT1: implication of cyclin-dependent kinases. Biochem Biophys Res Commun 409:187–192
Lavoie G, Esteve PO, Laulan NB, Pradhan S, St-Pierre Y (2011) PKC isoforms interact with and phosphorylate DNMT1. BMC Biol 9:31
Lee JH, Park SJ, Nakai K (2017) Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s. Sci Rep 7:11295
Leonhardt H, Page AW, Weier HU, Bestor TH (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell 71:865–873
Li H, Rauch T, Chen ZX, Szabo PE, Riggs AD, Pfeifer GP (2006) The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. J Biol Chem 281:19489–19500
Li JY, Pu MT, Hirasawa R, Li BZ, Huang YN, Zeng R, Jing NH, Chen T, Li E, Sasaki H et al (2007) Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. Mol Cell Biol 27:8748–8759
Li BZ, Huang Z, Cui QY, Song XH, Du L, Jeltsch A, Chen P, Li G, Li E, Xu GL (2011) Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. Cell Res 21:1172–1181
Li Y, Zhang Z, Chen J, Liu W, Lai W, Liu B, Li X, Liu L, Xu S, Dong Q et al (2018a) Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. Nature 564:136–140
Li Z, Fu X, Wang Y, Liu R, He Y (2018b) Polycomb-mediated gene silencing by the BAH-EMF1 complex in plants. Nat Genet 50:1254–1261
Liang G, Chan MF, Tomigahara Y, Tsai YC, Gonzales FA, Li E, Laird PW, Jones PA (2002) Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Mol Cell Biol 22:480–491
Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schubeler D (2011) Identification of genetic elements that autonomously determine DNA methylation states. Nat Genet 43:1091–1097
Lin IG, Han L, Taghva A, O'Brien LE, Hsieh CL (2002) Murine de novo methyltransferase Dnmt3a demonstrates strand asymmetry and site preference in the methylation of DNA in vitro. Mol Cell Biol 22:704–723
Lin CC, Chen YP, Yang WZ, Shen JCK, Yuan HS (2020) Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. Nucleic Acids Res 48:3949–3961
Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:315–322
Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD et al (2013) Global epigenomic reconfiguration during mammalian brain development. Science 341:1237905
Liu Y, Zhang X, Blumenthal RM, Cheng X (2013) A common mode of recognition for methylated CpG. Trends Biochem Sci 38:177–183
Long HK, Blackledge NP, Klose RJ (2013) ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochem Soc Trans 41:727–740
Lorincz MC, Schubeler D, Hutchinson SR, Dickerson DR, Groudine M (2002) DNA methylation density influences the stability of an epigenetic imprint and Dnmt3a/b-independent de novo methylation. Mol Cell Biol 22:7572–7580
Maiti A, Drohat AC (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active

demethylation of CpG sites. J Biol Chem 286:35334–35338
Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y (2010) Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature 464:927–931
Mertineit C, Yoder JA, Taketo T, Laird DW, Trasler JM, Bestor TH (1998) Sex-specific exons control DNA methyltransferase in mammalian germ cells. Development 125:889–897
Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K, Shimamura S, Arita K, Kodama T, Ishikawa F et al (2013) Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature 502:249–253
Nishiyama A, Mulholland CB, Bultmann S, Kori S, Endo A, Saeki Y, Qin W, Trummer C, Chiba Y, Yokoyama H et al (2020) Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation. Nat Commun 11:1222
Noh KM, Wang H, Kim HR, Wenderski W, Fang F, Li CH, Dewell S, Hughes SH, Melnick AM, Patel DJ et al (2015) Engineering of a Histone-Recognition Domain in Dnmt3a Alters the Epigenetic Landscape and Phenotypic Features of Mouse ESCs. Mol Cell 59:89–103
Okano M, Xie S, Li E (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet 19:219–220
Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99:247–257
Okitsu CY, Hsieh CL (2007) DNA methylation dictates histone H3K4 methylation. Mol Cell Biol 27:2746–2757
Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD et al (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448:714–717
Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M (2009) Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep 10:1235–1241
Pappalardi MB, Keenan K, Cockerill M, Kellner WA, Stowell A, Sherk C, Wong K, Pathuri S, Briand J, Steidel M et al (2021) Discovery of a first-in-class reversible DNMT1-selective inhibitor with improved tolerability and efficacy in acute myeloid leukemia. Nat Can 2:1002–1017
Pradhan M, Esteve PO, Chin HG, Samaranayke M, Kim GD, Pradhan S (2008) CXXC domain of human DNMT1 is essential for enzymatic activity. Biochemistry 47:10000–10009
Prosser J, Frommer M, Paul C, Vincent PC (1986) Sequence relationships of three human satellite DNAs. J Mol Biol 187:145–155
Purdy MM, Holz-Schietinger C, Reich NO (2010) Identification of a second DNA binding site in human DNA methyltransferase 3A by substrate inhibition and domain deletion. Arch Biochem Biophys 498:13–22
Qian S, Lv X, Scheid RN, Lu L, Yang Z, Chen W, Liu R, Boersma MD, Denu JM, Zhong X et al (2018) Dual recognition of H3K4me3 and H3K27me3 by a plant histone reader SHL. Nat Commun 9:2425
Qin W, Leonhardt H, Spada F (2011) Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. J Cell Biochem 112: 439–444
Qin W, Wolf P, Liu N, Link S, Smets M, La Mastra F, Forne I, Pichler G, Horl D, Fellinger K et al (2015) DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. Cell Res 25:911–929
Qiu C, Sawada K, Zhang X, Cheng X (2002) The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nat Struct Biol 9:217–224
Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A et al (2011) In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell 44:361–372
Quenneville S, Turelli P, Bojkowska K, Raclot C, Offner S, Kapopoulou A, Trono D (2012) The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. Cell Rep 2:766–773
Rajavelu A, Lungu C, Emperle M, Dukatz M, Brohm A, Broche J, Hanelt I, Parsa E, Schiffers S, Karnik R et al (2018) Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2. Nucleic Acids Res 46: 9044–9056
Ren W, Fan H, Grimm SA, Guo Y, Kim JJ, Yin J, Li L, Petell CJ, Tan XF, Zhang ZM et al (2020) Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. Proc Natl Acad Sci U S A 117:18439–18447
Ren W, Fan H, Grimm SA, Guo Y, Kim JJ, Yin J, Li L, Petell CJ, Tan XF, Zhang ZM et al (2021) DNMT1 reads heterochromatic H4K20me3 to reinforce LINE-1 DNA methylation. Nat Commun 12(1):2490
Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL, Wolffe AP (2000) DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat Genet 25:338–342
Rondelet G, Dal Maso T, Willems L, Wouters J (2016) Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. J Struct Biol 194:357–367

Ross JP, Suetake I, Tajima S, Molloy PL (2010) Recombinant mammalian DNA methyltransferase activity on model transcriptional gene silencing short RNA-DNA heteroduplex substrates. Biochem J 432:323–332

Rothbart SB, Krajewski K, Nady N, Tempel W, Xue S, Badeaux AI, Barsyte-Lovejoy D, Martinez JY, Bedford MT, Fuchs SM et al (2012) Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat Struct Mol Biol 19: 1155–1160

Rountree MR, Bachman KE, Baylin SB (2000) DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nat Genet 25: 269–277

Rush M, Appanah R, Lee S, Lam LL, Goyal P, Lorincz MC (2009) Targeting of EZH2 to a defined genomic site is sufficient for recruitment of Dnmt3a but not de novo DNA methylation. Epigenetics 4:404–414

Schermelleh L, Haemmer A, Spada F, Rosing N, Meilinger D, Rothbauer U, Cardoso MC, Leonhardt H (2007) Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. Nucleic Acids Res 35:4301–4312

Schmitz KM, Mayer C, Postepska A, Grummt I (2010) Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev 24:2264–2269

Senyuk V, Premanand K, Xu P, Qian Z, Nucifora G (2011) The oncoprotein EVI1 and the DNA methyltransferase Dnmt3 co-operate in binding and de novo methylation of target DNA. PLoS One 6:e20793

Shamay M, Greenway M, Liao G, Ambinder RF, Hayward SD (2010) De novo DNA methyltransferase DNMT3b interacts with NEDD8-modified proteins. J Biol Chem 285:36377–36386

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

Shirohzu H, Kubota T, Kumazawa A, Sado T, Chijiwa T, Inagaki K, Suetake I, Tajima S, Wakui K, Miki Y et al (2002) Three novel DNMT3B mutations in Japanese patients with ICF syndrome. Am J Med Genet 112:31–37

Song J, Rechkoblit O, Bestor TH, Patel DJ (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science 331:1036–1040

Song J, Teplova M, Ishibe-Murakami S, Patel DJ (2012) Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science 335:709–712

Stec I, Nagl SB, van Ommen GJ, den Dunnen JT (2000) The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? FEBS Lett 473:1–5

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

Suetake I, Hayata D, Tajima S (2006) The amino-terminus of mouse DNA methyltransferase 1 forms an independent domain and binds to DNA with the sequence involving PCNA binding motif. J Biochem 140:763–776

Suetake I, Mishima Y, Kimura H, Lee YH, Goto Y, Takeshima H, Ikegami T, Tajima S (2011) Characterization of DNA-binding activity in the N-terminal domain of the DNA methyltransferase Dnmt3a. Biochem J 437:141–148

Sugiyama Y, Hatano N, Sueyoshi N, Suetake I, Tajima S, Kinoshita E, Kinoshita-Kikuta E, Koike T, Kameshita I (2010) The DNA-binding activity of mouse DNA methyltransferase 1 is regulated by phosphorylation with casein kinase 1delta/epsilon. Biochem J 427: 489–497

Syeda F, Fagan RL, Wean M, Avvakumov GV, Walker JR, Xue S, Dhe-Paganon S, Brenner C (2011) The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. J Biol Chem 286:15344–15351

Tajima S, Tsuda H, Wakabayashi N, Asano A, Mizuno S, Nishimori K (1995) Isolation and expression of a chicken DNA methyltransferase cDNA. J Biochem 117:1050–1057

Takagi H, Tajima S, Asano A (1995) Overexpression of DNA methyltransferase in myoblast cells accelerates myotube formation. Eur J Biochem 231:282–291

Takeshita K, Suetake I, Yamashita E, Suga M, Narita H, Nakagawa A, Tajima S (2011) Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proc Natl Acad Sci U S A 108:9055–9059

Veland N, Lu Y, Hardikar S, Gaddis S, Zeng Y, Liu B, Estecio MR, Takata Y, Lin K, Tomida MW et al (2018) DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells. Nucleic Acids Res 47(1):152–167

Vertino PM, Yen RW, Gao J, Baylin SB (1996) De novo methylation of CpG island sequences in human fibroblasts overexpressing DNA (cytosine-5-)-methyltransferase. Mol Cell Biol 16:4555–4565

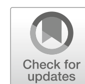
Vilkaitis G, Suetake I, Klimasauskas S, Tajima S (2005) Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. J Biol Chem 280:64–72

Voo KS, Carlone DL, Jacobsen BM, Flodin A, Skalnik DG (2000) Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol Cell Biol 20:2108–2121

Wang YA, Kamarova Y, Shen KC, Jiang Z, Hahn MJ, Wang Y, Brooks SC (2005) DNA methyltransferase-3a interacts with p53 and represses p53-mediated gene expression. Cancer Biol Ther 4:1138–1143

Wang Q, Yu G, Ming X, Xia W, Xu X, Zhang Y, Zhang W, Li Y, Huang C, Xie H et al (2020) Imprecise

DNMT1 activity coupled with neighbor-guided correction enables robust yet flexible epigenetic inheritance. Nat Genet 52:828–839
Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 39:457–466
Weinberg DN, Papillon-Cavanagh S, Chen H, Yue Y, Chen X, Rajagopalan KN, Horth C, McGuire JT, Xu X, Nikbakht H et al (2019) The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. Nature 573: 281–286
Weinberg DN, Rosenbaum P, Chen X, Barrows D, Horth C, Marunde MR, Popova IK, Gillespie ZB, Keogh M-C, Lu C et al (2021) Two competing mechanisms of DNMT3A recruitment regulate the dynamics of de novo DNA methylation at PRC1-targeted CpG islands. Nat Genet 53:794–800
Wienholz BL, Kareta MS, Moarefi AH, Gordon CA, Ginno PA, Chedin F (2010) DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo. PLoS Genet 6:e1001106
Xu T-H, Liu M, Zhou XE, Liang G, Zhao G, Xu HE, Melcher K, Jones PA (2020) Structure of nucleosome-bound DNA methyltransferases DNMT3A and DNMT3B. Nature 586:151–155
Yamaguchi L, Nishiyama A, Misaki T, Johmura Y, Ueda J, Arita K, Nagao K, Obuse C, Nakanishi M (2017) Usp7-dependent histone H3 deubiquitylation regulates maintenance of DNA methylation. Sci Rep 7:55
Yang D, Fang Q, Wang M, Ren R, Wang H, He M, Sun Y, Yang N, Xu RM (2013) Nalpha-acetylated Sir3 stabilizes the conformation of a nucleosome-binding loop in the BAH domain. Nat Struct Mol Biol 20: 1116–1118
Yang Z, Qian S, Scheid RN, Lu L, Chen X, Liu R, Du X, Lv X, Boersma MD, Scalf M et al (2018) EBS is a bivalent histone reader that regulates floral phase transition in Arabidopsis. Nat Genet 50:1247–1253
Yarychkivska O, Shahabuddin Z, Comfort N, Boulard M, Bestor TH (2018a) BAH domains and a histone-like motif in DNA methyltransferase 1 (DNMT1) regulate de novo and maintenance methylation in vivo. J Biol Chem 293:19466–19475
Yarychkivska O, Tavana O, Gu W, Bestor TH (2018b) Independent functions of DNMT1 and USP7 at replication foci. Epigenetics Chromatin 11:9
Ye F, Kong X, Zhang H, Liu Y, Shao Z, Jin J, Cai Y, Zhang R, Li L, Zhang YW et al (2018) Biochemical Studies and Molecular Dynamic Simulations Reveal the Molecular Basis of Conformational Changes in DNA Methyltransferase-1. ACS Chem Biol 13:772–781
Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, Rathert P, Brandt O, Reinhardt R, Fischle W et al (2010) Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res 38(13):4246–4253
Zhang ZM, Liu S, Lin K, Luo Y, Perry JJ, Wang Y, Song J (2015) Crystal Structure of Human DNA Methyltransferase 1. J Mol Biol 427:2520–2531
Zhang ZM, Lu R, Wang P, Yu Y, Chen D, Gao L, Liu S, Ji D, Rothbart SB, Wang Y et al (2018) Structural basis for DNMT3A-mediated de novo DNA methylation. Nature 554:387–391
Zhang YZ, Yuan J, Zhang L, Chen C, Wang Y, Zhang G, Peng L, Xie SS, Jiang J, Zhu JK et al (2020) Coupling of H3K27me3 recognition with transcriptional repression through the BAH-PHD-CPL2 complex in Arabidopsis. Nat Commun 11:6212
Zhao D, Zhang XJ, Guan HP, Xiong XZ, Shi XM, Deng HT, Li HT (2016a) The BAH domain of BAHD1 is a histone H3K27me3 reader. Protein & cell 7:222–226
Zhao Q, Zhang J, Chen R, Wang L, Li B, Cheng H, Duan X, Zhu H, Wei W, Li J et al (2016b) Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals. Nat Commun 7:12464
Zuo X, Sheng J, Lau HT, McDonald CM, Andrade M, Cullen DE, Bell FT, Iacovino M, Kyba M, Xu G et al (2012) Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. J Biol Chem 287: 2107–2118



4 Enzymology of Mammalian DNA Methyltransferases

Renata Z. Jurkowska and Albert Jeltsch

Abstract

DNA methylation is a hot topic in basic and biomedical research. Despite tremendous progress in understanding the structures and biochemical properties of the mammalian DNA methyltransferases (DNMTs), principles of their targeting and regulation in cells have only begun to be uncovered. In mammals, DNA methylation is introduced by the DNMT1, DNMT3A, and DNMT3B enzymes, which are all large multi-domain proteins containing a catalytic C-terminal domain and a complex N-terminal part with diverse targeting and regulatory functions. The sub-nuclear localization of DNMTs plays an important role in their biological function: DNMT1 is localized to replicating DNA and heterochromatin via interactions with PCNA and UHRF1 and direct binding to the heterochromatic histone modifications H3K9me3 and H4K20me3. DNMT3 enzymes bind to heterochromatin via protein multimerization and are targeted to chromatin by their ADD, PWWP, and UDR domains, binding to unmodified H3K4, H3K36me2/3, and H2AK119ub1, respectively. In recent years, a novel regulatory principle has been discovered in DNMTs, as structural and functional data demonstrated that the catalytic activities of DNMT enzymes are under a tight allosteric control by their different N-terminal domains with autoinhibitory functions. This mechanism provides numerous possibilities for the precise regulation of the methyltransferases via controlling the binding and release of the autoinhibitory domains by protein partners, chromatin interactions, non-coding RNAs, or posttranslational modifications of the DNMTs. In this chapter, we summarize key enzymatic properties of DNMTs, viz. their specificity and processivity, and afterwards focus on the regulation of their activity and targeting via allosteric processes, protein interactions, and posttranslational modifications.

Keywords

DNA methyltransferase · DNMT1 · DNMT3A · DNMT3B · Enzyme mechanism · Enzyme regulation · Enzyme specifcity · DNA methylation

Abbreviations

5mC 5-methylcytosine

R. Z. Jurkowska (✉)
School of Biosciences, Cardiff University, Cardiff, UK
e-mail: jurkowskar@cardiff.ac.uk

A. Jeltsch (✉)
Institute of Biochemistry and Technical Biochemistry, Department of Biochemistry, University of Stuttgart, Stuttgart, Germany
e-mail: albert.jeltsch@ibtb.uni-stuttgart.de

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_4

ADD domain	ATRX-Dnmt3-DNMT3L domain
AdoHcy	S-adenosyl-L-homocysteine
AdoMet	S-adenosyl-L-methionine
AML	acute myeloid leukaemia
BAH domain	Bromo-adjacent homology domain
CpG	cytosine-guanine dinucleotide sequence
DMAP1	DNA methyltransferase-associated protein 1
DMR	differentially methylated region
DNMT	(mammalian) DNA nucleotide methyltransferase
ES cells	embryonic stem cells
HDAC	histone deacetylase
ICF	Immunodeficiency-centromeric instability-facial anomalies syndrome
lncRNA	long non-coding RNA
KG repeats	lysine-glycine repeats
KO	knock out
MBD	methyl-binding domain
miRNA	micro-RNA
MTase	methyltransferase
ncRNA	non-coding RNA
PCNA	proliferating cell nuclear antigen
PBD	PCNA binding domain
PHD	plant homeodomain
PTM	posttranslational modification
RING	really interesting new gene
RFTD	replication foci targeting domain
SIRT1	sirtuin 1
SRA domain	SET and RING-associated domain
TET	Ten-eleven translocation
TRD	target recognition domain
TTD	tandem Tudor domain
UBL	ubiquitin-like domain
UDR	ubiquitin-dependent recruitment
UHRF1	ubiquitin-like with PHD and ring finger domains 1
USP7	ubiquitin-specific peptidase 7

4.1 Introduction

The expression of genes in multicellular organisms is coordinated during development and cellular differentiation by epigenetic information comprising DNA methylation, histone tail posttranslational modifications (PTMs), and non-coding RNAs [for general reviews on Molecular Epigenetics cf. (Allis and Jenuwein 2016)]. In mammals, DNA methylation mainly occurs at the C5-position of the cytosine residues, primarily in CpG dinucleotide sequences [for general reviews on DNA methylation cf. (Ambrosi et al. 2017; Schubeler 2015; Jeltsch and Jurkowska 2014)]. However, only certain CpG sites are methylated, resulting in the establishment of a tissue- and cell type-specific pattern of DNA methylation consisting of modified and unmodified sites. In different cell types, approximately 60–80% of all CpGs in the human genome are modified (3–8% of all cytosines). Notably, the correct methylation pattern is essential for development and human health, and several diseases, including cancer, are associated with aberrant DNA methylation [for reviews cf. (Zhao et al. 2021; Weinberg et al. 2019; Bergman and Cedar 2013; Suva et al. 2013; Hamidi et al. 2015)].

In mammals, DNA methylation patterns are introduced during early development and maturation of germ cells by DNA methyltransferases (MTases) DNMT3A and DNMT3B, with the help of the stimulatory factor DNMT3L (Jurkowska et al. 2011a; Jeltsch and Jurkowska 2016). DNMT3A and DNMT3B have been traditionally designated as de novo DNA MTases, as they do not display any significant preference between hemimethylated and unmethylated DNA (Okano et al. 1998; Gowher and Jeltsch 2001). In agreement with this role, they are highly expressed in undifferentiated cells and germ cell precursors, and present at much lower levels in somatic cells. In the cell nucleus, they localize to pericentromeric heterochromatin (Chen et al. 2004; Ge et al. 2004; Barau et al. 2016), where they are tightly bound to nucleosomes containing

methylated DNA (Jeong et al. 2009; Sharma et al. 2011). Mice and other rodents also contain an additional DNMT3B-related DNA methyltransferase called DNMT3C, which is specifically expressed in testis. It is required for methylation and silencing of retrotransposons during spermatogenesis and hence critical for male fertility in mice (Barau et al. 2016; Jain et al. 2017).

After their establishment, DNA methylation patterns are perpetuated through cell divisions, with small tissue-specific changes. The palindromic nature of the CpG sites provides an elegant mechanism for the inheritance of the DNA methylation mark because the methylation information is encoded in both DNA strands. During DNA replication the fully methylated CpG sites are converted into a hemimethylated state, with the parental strand carrying the original methylation marks and the daughter strand devoid of methylation. The methylation pattern is copied after each round of DNA replication by the maintenance methyltransferase DNMT1. This enzyme is present at the replication fork, where it quickly methylates hemimethylated CpG dinucleotides, thereby restoring the original DNA methylation pattern (Petryk et al. 2021). DNMT1 is ubiquitously and highly expressed in proliferating cells, representing the major DNA MTase activity in somatic tissues throughout mammalian development, but it is present only at low levels in non-dividing cells (Robertson et al. 1999).

However, recent data showed that this traditional division of tasks into de novo and maintenance methyltransferases is an oversimplification. DNA methylation is more correctly described as a dynamic process of ongoing methylation and demethylation, and DNMT1, DNMT3A, and DNMT3B all play roles in both de novo and maintenance methylation (Jeltsch and Jurkowska 2014). Hence, the dynamic regulation and targeting of DNMTs and Ten-eleven translocation (TET) DNA demethylating enzymes controls the methylation state of each CpG site, thereby governing all the biological processes associated with DNA methylation. Consequently, the complex role of DNA methylation in human biology cannot be decoded without a thorough mechanistic understanding of the properties of the DNMTs, including their regulation, targeting, and interaction with chromatin and other epigenetic factors.

4.2 General Features of Mammalian DNMTs

4.2.1 Structure and Domain Composition of Mammalian DNMTs

Structural and biochemical data provided compelling evidence that the arrangement of the specific domains in DNMTs plays a central role in the regulation of the biological functions of these enzymes. The general architecture of all mammalian DNMTs is similar. They all are multi-domain proteins, in which two functional parts can be distinguished, a large N-terminal regulatory part and a smaller C-terminal part, required for catalysis (Fig. 4.1) (Jeltsch 2002; Hermann et al. 2004a; Jurkowska et al. 2011a). The N-terminal parts of variable size are different between DNMT1 and DNMT3 proteins. They guide the nuclear localization of the enzymes and mediate their interaction with other proteins, regulatory nucleic acids (like non-coding RNAs), and chromatin. They are also subject to posttranslational modifications (PTMs) and are involved in the allosteric regulation of the enzymes' activity and specificity.

The C-terminal domains harboring the catalytic centers of the enzymes are required for binding of the S-adenosyl-L-methionine (AdoMet) cofactor and the DNA substrate. They contain ten conserved amino acid motifs characteristic of the common structure of all DNA-(cytosine-C5)-MTases, called the "AdoMet-dependent MTase fold", which consists of a mixed seven-stranded β-sheet, formed by six parallel β-strands and a seventh strand inserted in an anti-parallel

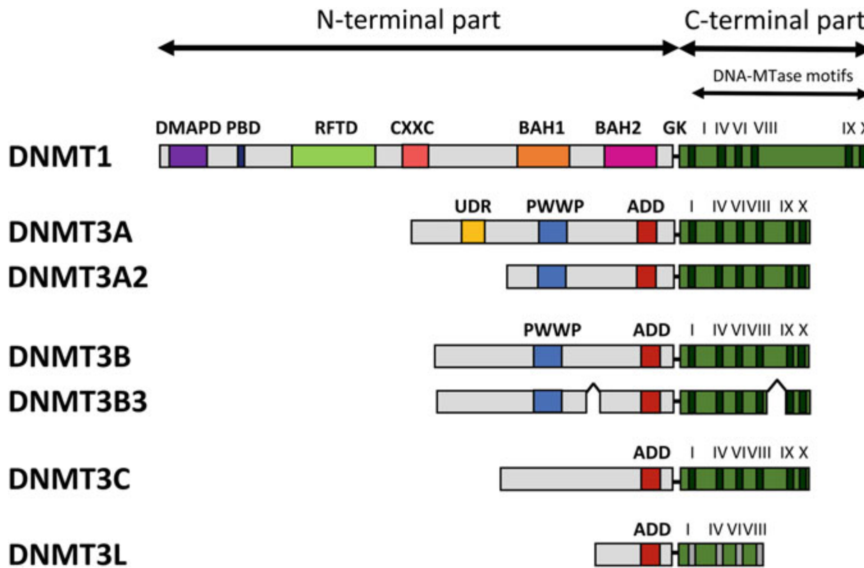


Fig. 4.1 Domain structure of the mammalian DNMT enzymes. Abbreviations used: *DMAPD* DNA methyltransferase-associated protein 1 interacting domain, *PDB* PCNA binding domain, *RFTD* replication foci targeting domain, *CXXC* CXXC domain, *BAH1 and BAH2* Bromo-adjacent homology domains 1 and 2, *GK* glycine lysine repeats, *UBR* ubiquitin-dependent recruitment domain, *PWWP* PWWP domain, *ADD* ATRX-DNMT3-DNMT3L domain. DNMT3C is a rodent-specific enzyme

orientation into the sheet between strands 5 and 6. Six α-helices surround the central β-sheet on both sides (Cheng and Blumenthal 2008; Jeltsch 2002). The C-terminal domain is involved in the cofactor binding (motifs I and X), binding of the flipped substrate cytosine base, and the methyl group transfer (motifs IV, VI, and VIII). The non-conserved region between motifs VIII and IX, the so-called target recognition domain (TRD), is involved in substrate DNA recognition and specificity.

4.2.2 Catalytic Mechanism of C5-MTases

DNA-(cytosine C5)-methyltransferases catalyze the transfer of the methyl group from an AdoMet cofactor molecule to the C5-position of cytosine residues. In this reaction, 5-methylcytosine (5mC) is created and the AdoMet is converted into S-adenosyl-L-homocysteine (AdoHcy), which is then released from the enzyme. The transfer of the activated methyl group from AdoMet to the C5-position of the cytosine requires a close contact between the enzyme's active site and the substrate base. Such proximity is not possible while the base is located in the DNA double helix; therefore, DNA methyltransferases flip their target base out of the DNA during catalysis and bury it into a hydrophobic pocket of their active center. This base flipping mechanism was first discovered in 1994 for the bacterial DNA C5-MTase M.HhaI (Klimasauskas et al. 1994). Later, it became clear that it is common to all DNA methyltransferases, including the mammalian enzymes (Cheng and Roberts 2001; Jeltsch 2002) and flipping of the cytosine base was observed in different crystal structures of DNMT1, DNMT3A, and DNMT3B with bound substrate DNA (Song et al. 2012; Adam et al. 2020; Zhang et al. 2018; Gao et al. 2020b; Lin et al. 2020).

The methylation of the C5-position of cytosine is not an easy chemical task, because the cytosine is an electron-poor aromatic system. Therefore, its C5-atom is not intrinsically reactive and it will

not attack the activated methyl sulfonium group of the AdoMet spontaneously. Hence, a key step in the catalysis of DNA-(cytosine C5)-methyltransferases is the nucleophilic attack of the catalytic cysteine residue located in a PCQ motif (motif IV) on the C6 position of the cytosine ring, leading to the formation of a covalent bond between the enzyme and the substrate base. Thereby, the negative charge density at the C5-atom of the cytosine increases, so that it can attack the methyl group of the cofactor. It has been postulated that the nucleophilic attack of the cysteine might be facilitated by a transient protonation of the cytosine ring at the endocyclic nitrogen atom (N3) by an enzyme-derived acid; the conserved glutamate residue from an ENV motif (motif VI) has been proposed to carry out this reaction. In addition, it stabilises the flipped cytosine by forming an H-bond to the N4-amino group. The second arginine residue from an RXR motif (motif VIII) may be involved in the stabilization of both the glutamate and the cytosine base as well. The addition of the methyl group to the base is followed by a deprotonation of the C5-atom, catalyzed by a so far unknown proton acceptor, which resolves the covalent bond between the enzyme and the base in an elimination reaction and re-establishes aromaticity (Cheng and Roberts 2001; Jeltsch 2002). For DNMT1, kinetic isotope effects confirmed this two-step mechanism (Du et al. 2016). For DNMT3A, mutations of the key catalytic residues reduced the catalytic activity, confirming their critical role in catalysis (Reither et al. 2003; Gowher et al. 2006; Lukashevich et al. 2016).

Unexpectedly, DNA-(cytosine C5)-methyltransferases, including DNMT3A, also introduce low levels of methylation at the N3-atom of the cytosine ring, forming 3-methylcytosine (3mC) (Rosic et al. 2018), which is a toxic DNA alkylation lesion that interferes with RNA synthesis and DNA replication. 3mC is removed by members of the ALKB2 family of DNA alkylation repair enzymes that are evolutionarily strongly connected to DNA methyltransferases, reflecting their close functional link (Rosic et al. 2018). Mechanistically, the 3mC methylation is likely introduced after positioning the flipped cytosine base in an inverted conformation into the active site pocket of the DNMT (Dukatz et al. 2019b). Further mechanistic details of DNMTs, including their sequence specificity, processivity, oligomerization, and the mechanism of DNA and chromatin binding will be discussed below.

4.2.3 Regulation and Targeting of DNMTs

Despite tremendous progress in understanding the biochemical properties of the mammalian DNA methyltransferases, their genomic targeting combined with regulation of their activity is still insufficiently understood. Recent discoveries demonstrated the involvement of the N-terminal parts of the mammalian DNMTs in enzyme targeting and regulation. In this context, different domains of DNMT3A, DNMT3B, and DNMT1 were shown to directly bind modified histone H3 tails. Moreover, various domains (ADD domain in DNMT3A and CXXC and RFT domains in DNMT1) engage in autoinhibitory interactions with the catalytic domain, demonstrating that the activity of the enzymes is under precise allosteric control. Similarly, the interactions of the N-terminal domains of DNMTs with other proteins regulate the enzymes' activities and genome targeting. Thus, allosteric control represents a unifying concept in the regulation of DNMTs, which sets the stage for additional regulatory cues. By influencing the allosteric conformational changes of DNMTs, interacting proteins or RNAs, chromatin modifications or PTMs can affect key enzymatic properties of DNMTs, including their activity and eventually specificity (Jeltsch and Jurkowska 2016).

Several interaction partners of DNMTs have been described so far and their effect on the MTases has been studied mechanistically. This

includes PCNA (Chuang et al. 1997), DNMT3L (Bourc'his et al. 2001; Hata et al. 2002; Chedin et al. 2002; Gowher et al. 2005a), UHRF1 (Sharif et al. 2007; Bostick et al. 2007; Meilinger et al. 2009), MeCP2 (Fuks et al. 2003b; Kimura and Shiota 2003; Rajavelu et al. 2018), p53 (Wang et al. 2005; Sandoval and Reich 2019), or USP7 (Du et al. 2010; Felle et al. 2011). Other important interaction partners like HP1-beta (Fuks et al. 2003a), Mbd3 (Datta et al. 2005), MYC (Brenner et al. 2005), PU.1 and RP58 transcription factors (Suzuki et al. 2006; Fuks et al. 2001), zinc-finger proteins ZHX1 and Trim28 (Kim et al. 2007; Quenneville et al. 2011), protein lysine methyltransferases (PKMTs) G9a, SUV39H1 (Fuks et al. 2003a), EZH2 (Vire et al. 2006), and SETDB1 (Li et al. 2006), histone deacetylase (HDAC1) (Fuks et al. 2000; Fuks et al. 2001), and remodeling factors HELLS (Myant and Stancheva 2008; Zhu et al. 2006), SMARCA4 (Datta et al. 2005), and hSNF2 (Geiman et al. 2004) have been reported, but their interaction with DNMTs has not yet been mechanistically investigated in great details. Finally, various aspects of the biological function of DNMTs, including their targeting and activity in cells, are regulated by posttranslational modifications (PTMs). Until now, several PTMs, including phosphorylation, acetylation, ubiquitination, SUMOylation, and methylation, have been identified on mammalian DNMTs in proteomic studies (http://www.phosphosite.org). PTMs are ideally suited to mediate regulation of DNMTs' function, either by direct effects on catalytic activity or by recruiting modification-specific readers that could influence the enzymes' stability, activity, localization, or interaction with other proteins. Notably, the few modifications that have been functionally characterized revealed the important regulatory potential of the PTMs, opening the field for future research.

Finally, non-coding RNA (ncRNA) is an emerging player in chromatin regulation (Holoch and Moazed 2015; Rinn and Chang 2012) and RNA molecules have been shown to influence DNA methylation. In plants, a process of RNA-dependent DNA methylation exists, in which the RNA sequence directly guides DNA methylation (Matzke and Mosher 2014). Though this pathway is absent in mammals, binding of small and long non-coding RNAs (lncRNA) to mammalian DNMTs has been shown to guide and regulate their activity. In addition, the piRNA-mediated DNA methylation in the germline of many animals, including mammals (Iwasaki et al. 2015), recapitulates many features of an RNA-directed DNA methylation pathway. Recently, SPOCD1 has been identified to bind to the PIWI protein MIWI2 and DNMT3A/DNMT3L and to play an essential role in targeting DNA methylation to piRNA binding sites (Zoch et al. 2020), but many further details of piRNA-directed DNA methylation process are not yet well understood at the molecular level. The direct regulation of DNA methylation by genome-encoded non-coding RNAs adds another fascinating dimension to the complex interplay between the genetic information (encoded in the DNA sequence) and the epigenetic information (encoded in the chromatin modification pattern, including DNA methylation), urging more research in this direction.

4.3 Structure, Function, and Mechanism of DNMT1

4.3.1 Domain Composition of DNMT1

DNMT1 is a large enzyme, comprising 1620 amino acids in mice and 1616 amino acids in humans, but different isoforms of DNMT1, resulting from alternative splicing or use of an alternative promoter have been described (Hermann et al. 2004a; Jurkowska et al. 2011a). DNMT1 contains multiple functional domains located in the N-terminal part that is joined to the C-terminal part by a flexible linker composed of lysine-glycine (KG) repeats (Fig. 4.1). The N-terminal part serves as a platform for the assembly of various proteins involved in the control of chromatin structure and gene regulation.

The very N-terminus of DNMT1 contains the DNA methyltransferase-associated protein 1 (DMAP1) interaction domain that is involved in the interaction of DNMT1 with DMAP1, a transcriptional repressor, mediating the stability of DNMT1 in cells (Rountree et al. 2000). Next to it, the proliferating cell nuclear antigen (PCNA) binding domain (PBD) has been mapped (Chuang et al. 1997). The interaction with PCNA is involved in the targeting and tethering of DNMT1 to the replication fork during S-phase, which supports DNA methylation in the cell (Egger et al. 2006). The same region also contains an AT-hook-like DNA binding motif (Suetake et al. 2006). The replication foci targeting domain (RFTD) following next is involved in the targeting of DNMT1 to replication foci (Leonhardt et al. 1992) and centromeric chromatin (Easwaran et al. 2004). This domain interacts with UHRF1 (ubiquitin-like with PHD and ring finger domains 1), which harbors an SRA (SET and RING-associated) domain that specifically binds to hemimethylated DNA (see below). Moreover, the RFTD binds to ubiquitinated histone H3 tails, a modification introduced by the RING domain of UHRF1 (Nishiyama et al. 2013; Qin et al. 2015), in the context of H3K9me3 (Ren et al. 2020), a major heterochromatic histone PTM in the mammalian epigenome (Jeltsch et al. 2019). Next, the N-terminal part of DNMT1 contains a CXXC domain that binds unmethylated DNA and is implicated in DNMT1 regulation (Pradhan et al. 2008; Song et al. 2011; Bashtrykov et al. 2012a). The CXXC domain is followed by the BAH1 and BAH2 (Bromo-adjacent homology 1 and 2) domains. BAH1 binds H4K20me3 (Ren et al. 2021), another key heterochromatic histone PTM. Hence, through its N-terminal part, DNMT1 interacts with other proteins and specific histone marks, contributing to the crosstalk between DNA methylation and other epigenetic modifications.

The C-terminal domain of DNMT1 contains the catalytic center of the enzyme, but is not active in an isolated form, both in vitro and in vivo, despite the presence of all motifs required for catalysis (Fatemi et al. 2001; Margot et al. 2003). The structural arrangement of the particular domains in DNMT1 has been revealed by crystallographic studies (Song et al. 2011, 2012; Takeshita et al. 2011; Syeda et al. 2011) (Fig. 4.2). They demonstrated that the various domains in the N-terminal part of DNMT1 contact the C-terminal catalytic domain from different sides, explaining why the isolated terminal domain lacks catalytic activity.

4.3.2 Structures of DNMT1 and Allosteric Regulation

In recent years, several structures of truncated DNMT1 proteins (lacking various parts of the N-terminus) have been solved (Song et al. 2011, 2012; Takeshita et al. 2011; Zhang et al. 2015b; Adam et al. 2020). They all confirmed that the catalytic domain of DNMT1 adopts the typical AdoMet-dependent MTase fold described above. These studies also revealed that the enzyme unexpectedly undergoes large domain rearrangements, which allosterically regulate its catalytic activity (Fig. 4.2).

A DNMT1 C-terminal fragment lacking the RFT and CXXC domains adopted an open conformation, in which the enzyme was able to bind the hemimethylated substrate DNA (with a GG*CG*GC sequence) and showed high catalytic activity (Song et al. 2012) (Fig. 4.2a). This complex represented a real breakthrough in the field, as it provided the first example of a mammalian DNMT structure solved with substrate DNA bound in the active site. As expected, it showed the target cytosine flipped out of the DNA helix and bound to DNMT1 in a manner reminiscent of other DNA MTases. Moreover, this structure also revealed additional unforeseen rearrangements in the DNMT1–DNA structure, including the formation of a non-Watson/Crick base pair of the orphan G residue with a G flanking the CpG site. The (then orphaned) C of the flanking G:C base pair was rotated out of the DNA helix in a

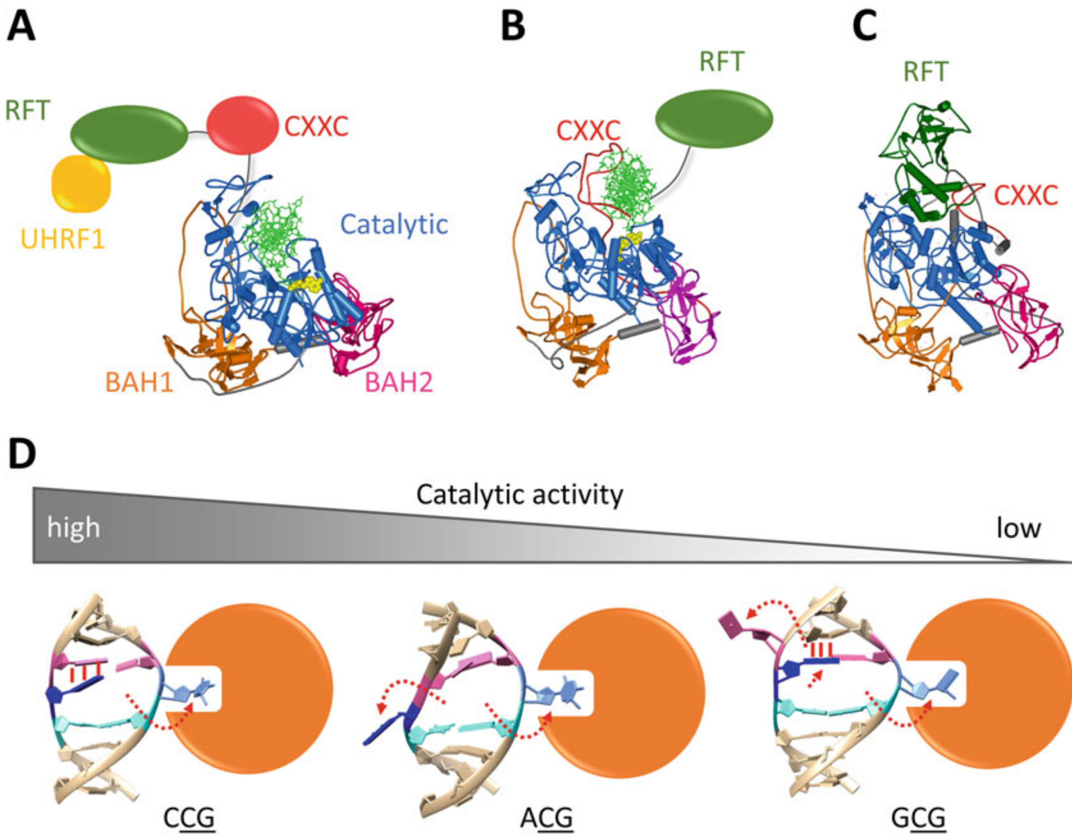


Fig. 4.2 Structures of DNMT1 with different N-terminal domains. (**a**) DNMT1 in an active conformation with DNA (green) bound in the active site (Song et al. 2012) (pdb 4D4A). Removal of the autoinhibitory RFTD can be triggered by UHRF1 interaction (Berkyurek et al. 2014; Bashtrykov et al. 2014a). (**b**) DNMT1 with unmethylated DNA bound to the autoinhibitory CXXC domain (Song et al. 2011) (pdb 3PTA). (**c**) DNMT1 with the RFT domain blocking access to the active site (Takeshita et al. 2011) (pdb 3AV4). (**d**) Flanking sequence-dependent base flipping mechanism observed in different DNMT1 structures. In the C*CG* structure (Adam et al. 2020) (pdb 6W8W) only the target cytosine (light blue) is rotated out of the double helix and bound by the enzyme (symbolized by the orange circle). In the A*CG* structure (Adam et al. 2020) (pdb 6W8V), both the target cytosine and the orphaned G (dark blue) are rotated out of the double helix and in the G*CG* structure (Song et al. 2012) (pdb 4DA4), the orphaned G forms a non-canonical base pair with the G(−1) and the corresponding C(−1′) (pink) is flipped out in the opposite direction. Catalytic activity of the complexes was inversely correlated with the extent of the conformational changes of the DNA upon complex formation. DNMT1 is symbolized by an orange circle. Reprinted in modified version from Jeltsch et al. (2021) with permission from Elsevier

direction roughly opposite to the target C flipping (Fig. 4.2d, GCG complex). Several contacts of the enzyme to the target CpG site observed in the structure were validated in kinetic studies as essential for the enzyme activity and the recognition of the CpG site (described in detail below in Sect. 4.3.3) (Bashtrykov et al. 2012b).

Additional recent structures of this DNMT1 fragment in complex with different DNA sequences strikingly revealed a DNA sequence-dependent base flipping mechanism (Adam et al. 2020) (Fig. 4.2d). The structure of DNMT1 bound to a hemimethylated TA*CG*GA substrate showed flipping of the target C and its Watson/Crick partner G, but no formation of a non-canonical base pair. In turn, the structure of DNMT1 bound to a hemimethylated TC*CG*TA substrate only showed target base flipping (Adam

et al. 2020). A kinetic analysis uncovered strong differences in the methylation rates of DNMT1 depending on the sequences flanking the target CG site. A comparison of the kinetic and structural data showed that the extent of the conformational rearrangements during base flipping was anti-correlated with the methylation rates of the corresponding substrates.

A structure of a larger C-terminal fragment of DNMT1 also containing the CXXC domain showed a CpG site-specific binding of an unmethylated DNA, but this interaction surprisingly occurred not at the C-terminal domain containing the active center, but at the CXXC domain (Song et al. 2011) (Fig. 4.2b). This observation led to the proposal that the CXXC domain has an autoinhibitory function and acts as a specificity filter in DNMT1 by preventing unmethylated DNA from accessing the active site. Kinetic experiments with this DNMT1 version indeed revealed an influence of the CXXC domain on the specificity of DNMT1 (Song et al. 2011). Surprisingly, similar experiments conducted with the full-length DNMT1 did not provide evidence for a role of the CXXC domain in the specificity of DNMT1 (Bashtrykov et al. 2012a), indicating that this point deserves further attention.

Finally, a crystal structure of an almost complete DNMT1 fragment, but without DNA provided additional seminal insight into the mechanism of DNMT1 by showing that the RFT domain inhibits the enzyme through binding to the active site cleft of the catalytic domain (Takeshita et al. 2011) (Fig. 4.2c). The autoinhibition was observed in biochemical studies as well (Takeshita et al. 2011; Syeda et al. 2011) and engineering of this interface altered the conformation of DNMT1, generating a methyltransferase that was hyperactive in vitro and in cells (Bashtrykov et al. 2014b).

Importantly, the arrangement of different domains in DNMT1 is controlled by long linker regions, which form tight interactions with surface clefts of the domains. Both the linkers and the clefts are subject to many reported PTMs in DNMT1, including phosphorylation, acetylation, and ubiquitination (http://www.phosphosite.org), which might directly control the positioning of these domains in DNMT1 and thereby enzymatic activity. Accumulating evidence indicates that the autoinhibitory mechanism of the RFT domain plays a central role as an allosteric trigger in DNMT1 (Fig. 4.2) that can be influenced by protein partners and chromatin binding. Indeed, the interaction of the RFTD with UHRF1 stimulates the activity of DNMT1 by relieving autoinhibition (Berkyurek et al. 2014; Bashtrykov et al. 2014a). Similarly, its interaction with ubiquitinated H3 and H3K9me3 also leads to DNMT1 activation (Nishiyama et al. 2013; Qin et al. 2015; Ren et al. 2020) and the binding of H4K20me3 to the BAH1 domain modulates the conformation of autoinhibitory linker regions connecting the different domains of DNMT1 (Ren et al. 2021).

Structural studies combined with molecular dynamics simulations showed that the helix following the catalytic loop in DNMT1 can adopt either a kinked or straight conformation. Mutational data suggested that the structural transition between these states is necessary for DNMT1 activity (Ye et al. 2018). Later, it was shown that these conformational changes are also dependent on the DNA sequence flanking the target sites and that the most active complex shows the least conformational changes (Adam et al. 2020). Hence, protein partners and chromatin interactions can regulate DNMT1 activity by influencing the allosteric conformation of the enzyme.

4.3.3 Specificity of DNMT1

DNMT1 shows a preference for hemimethylated DNA over unmethylated substrates, supporting its role as a maintenance MTase (Bashtrykov et al. 2012a; Bashtrykov et al. 2012b; Fatemi et al. 2001; Goyal et al. 2006; Song et al. 2012). Its intrinsic preference for hemimethylated DNA has been estimated to be about 30–40 fold (Jeltsch 2006), but it depends on the exact substrate sequence, its length, and the reaction conditions. This preference has been investigated for decades, as it is one of the mechanistic

foundations of the role of DNA methylation in the transfer of epigenetic information. We know now that it is molecularly based on the sequence-specific interaction of hemimethylated CpG sites with the active center of the enzyme that is mediated by the interaction of the methyl group with a hydrophobic pocket formed by the enzyme (Song et al. 2012). More precisely, the methyl group of the 5mC is placed into a pocket formed by C1501, L1502, W1512, L1515, and M1535, which explains the preference of the enzyme for hemimethylated target sites. Further details of this process could be uncovered once a structure of DNMT1 with an unmethylated DNA bound to the active center becomes available. The recognition of the 5mC-G base pair is based on side-chain- and backbone-mediated H-bonds of M1535, K1537, Q1538, and R1237 to the edges of the CpG base pair in the major and minor groove (Song et al. 2012). These interactions explain why the 5mC and the corresponding G in the target DNA strand are very accurately recognized by DNMT1 and cannot be exchanged by other nucleotides (Bashtrykov et al. 2012b). The requirement for a close contact between the catalytic domain of DNMT1 and its substrate DNA also explains the finding that the activity of DNMT1 on nucleosomal DNA is restricted to the linker DNA regions (Mishima et al. 2017).

Two recent studies investigated DNA replication-coupled maintenance of DNA methylation by DNMT1, providing novel evidence for de novo methylation activity of DNMT1 post-replication (Wang et al. 2020b; Ming et al. 2020). Genetic studies showed that the de novo activity of DNMT1 is particularly strong at intracisternal A particles (IAP) retrotransposons, possibly contributing to their stable silencing (Haggerty et al. 2021). This activity was dependent on UHRF1 acting as a universal cofactor of DNMT1, as well as H3K9me3 and TRIM28, suggesting that it crosstalks with the KRAB/TRIM28/SETDB1 silencing complex, which delivers H3K9me3 at retrotransposons (Haggerty et al. 2021; Markouli et al. 2021). In this context, H3K9me3 interaction could be mediated by the tandem Tudor domain (TTD) of UHRF1 (Nady et al. 2011) or by its direct interaction with the RFT domain of DNMT1 (Ren et al. 2020) (see below).

Recent evidence suggests that besides specificity for hemimethylated sites, Dnmt1 also has a preference for certain sequence contexts flanking the target CpGs. Biochemical experiments investigating the methylation of CpG sites in a randomized sequence context uncovered about 100-fold differences in the methylation rates of hemimethylated CpG sites placed in a variable NNCGNN sequence context (Adam et al. 2020). A comparison of the DNMT1–DNA structures on preferred and disfavored substrates with the kinetic data revealed the mechanistic basis for some of the observed flanking sequence preferences (Adam et al. 2020; Jeltsch et al. 2021). The disfavor for a G in the target strand in the −1 flanking base pair can be explained, because it allows the formation of the non-canonical G-G base pair with the orphaned G, which is accompanied by base flipping of the C in the non-target strand seen in the low-activity GG*CG*GC complex. The observed disfavor for a G in the non-target strand at the −2 flanking base pair could be explained because it could stack to the non-canonical G:G base pair and further stabilize this low-activity conformation. In turn, the preference for a G in the non-target strand at the −1 site could be related to its ability to stack with the orphaned G, keeping it inside of the DNA helix and thereby stabilizing the highly active conformation seen in the TC*CG*TA complex. In addition to these direct effects, minor groove width at the +1 to +3 flank correlated with DNMT1 activity as well. Notably, the comparison with genomic methylation data from various sources showed that the flanking sequence preferences of DNMT1 highly correlate with the flanking site-dependent modulation of genomic DNA methylation levels in human and mouse cells, indicating that the preferences determined in vitro affect genomic DNA methylation patterns in cells (Adam et al. 2020).

4.3.4 Processivity of DNMT1

DNMT1 is a highly processive enzyme, able to methylate long stretches of hemimethylated DNA without dissociation from the substrate, a property that fits perfectly to its function as a molecular copy machine at the replication fork (Goyal et al. 2006; Hermann et al. 2004b; Vilkaitis et al. 2005). A recent study revealed that DNMT1 undergoes a conformational change after DNA binding from an open into a closed conformation capable of processive methylation. Once the enzyme has adopted the closed conformation, it has a 97% chance of staying on the DNA and continuing processive DNA methylation after each methylation event (Adam et al. 2020). Interestingly, processive methylation is possible only in one strand of the DNA, which indicates that DNMT1 does not exchange DNA strands while moving along its substrate (Hermann et al. 2004b). These biochemical findings are in perfect agreement with the structure of DNMT1 with bound substrate DNA (Song et al. 2012), showing that the enzyme enwraps the DNA, which enables it to slide along the substrate and catalyze several successive methylation reactions without dissociation from the DNA. Due to its high processivity, DNMT1 is a very effective enzyme, ideally suited to follow DNA replication and methylate the newly synthetized DNA strand before the chromatin is reassembled.

4.3.5 Allosteric Regulation and Targeting of DNMT1

The sub-nuclear localization of DNMT1 changes dynamically during the cell cycle (Hermann et al. 2004a; Jurkowska et al. 2011a). The enzyme is diffusely distributed in the nucleus during interphase (when cells are not replicating) but localizes to replication foci in the early and mid-S-phase (in cells actively synthesizing DNA). During progression of the S-phase, the sub-nuclear pattern of DNMT1 changes from small, punctuate, and abundant structures in early S-phase to fewer, large, toroidal structures in late S-phase, which co-localized with late replicating heterochromatic satellite DNA (Leonhardt et al. 1992; O'Keefe et al. 1992; Easwaran et al. 2004; Liu et al. 2013; Schneider et al. 2013). In addition, some DNMT1 remains associated with the centromeric heterochromatin in G2 phase even after heterochromatin replication. In murine ESC cells, DNMT1 shows a heterochromatic distribution (Ren et al. 2020, 2021). Three regions of DNMT1 have been implicated in the targeting of the enzyme to the replication foci during S-phase, namely the PCNA binding domain (PBD) (Chuang et al. 1997), the replication foci targeting domain (RFTD) (Leonhardt et al. 1992), and the BAH domains (Liu et al. 1998), which will be described in the following chapters in more detail (Figs. 4.1 and 4.3).

4.3.5.1 The DNMT1–PCNA Interaction

Deletion of RFTD or BAH domains did not affect the delivery of DNMT1 to the replication fork (Easwaran et al. 2004), suggesting that the PBD domain has a central role in this process. Through this domain, DNMT1 directly interacts with PCNA, the so-called processivity factor of the replication machinery that forms a ring around the DNA helix (Chuang et al. 1997). In addition, both proteins co-localize in vivo, indicating that PCNA might recruit DNMT1 to the replication fork and load it onto DNA. Indeed, the expression of a truncated DNMT1, which lacked parts of the PBD domain, led to a delay in the re-methylation of DNA after replication (Egger et al. 2006). However, it did not cause massive defects in DNA methylation, indicating that the interaction of PCNA with DNMT1 contributes to the efficiency of DNA re-methylation, but it is not essential for this process. In addition, in vitro experiments provided evidence that the interaction with PCNA increases the DNA binding and catalytic activity of DNMT1 (Iida et al. 2002).

The interaction of DNMT1 with heterochromatin occurs in a replication-independent manner (Easwaran et al. 2004) and is mediated in part by the PBD domain of DNMT1 and by UHRF1, as described in the next paragraph. Direct interactions with heterochromatic histone marks

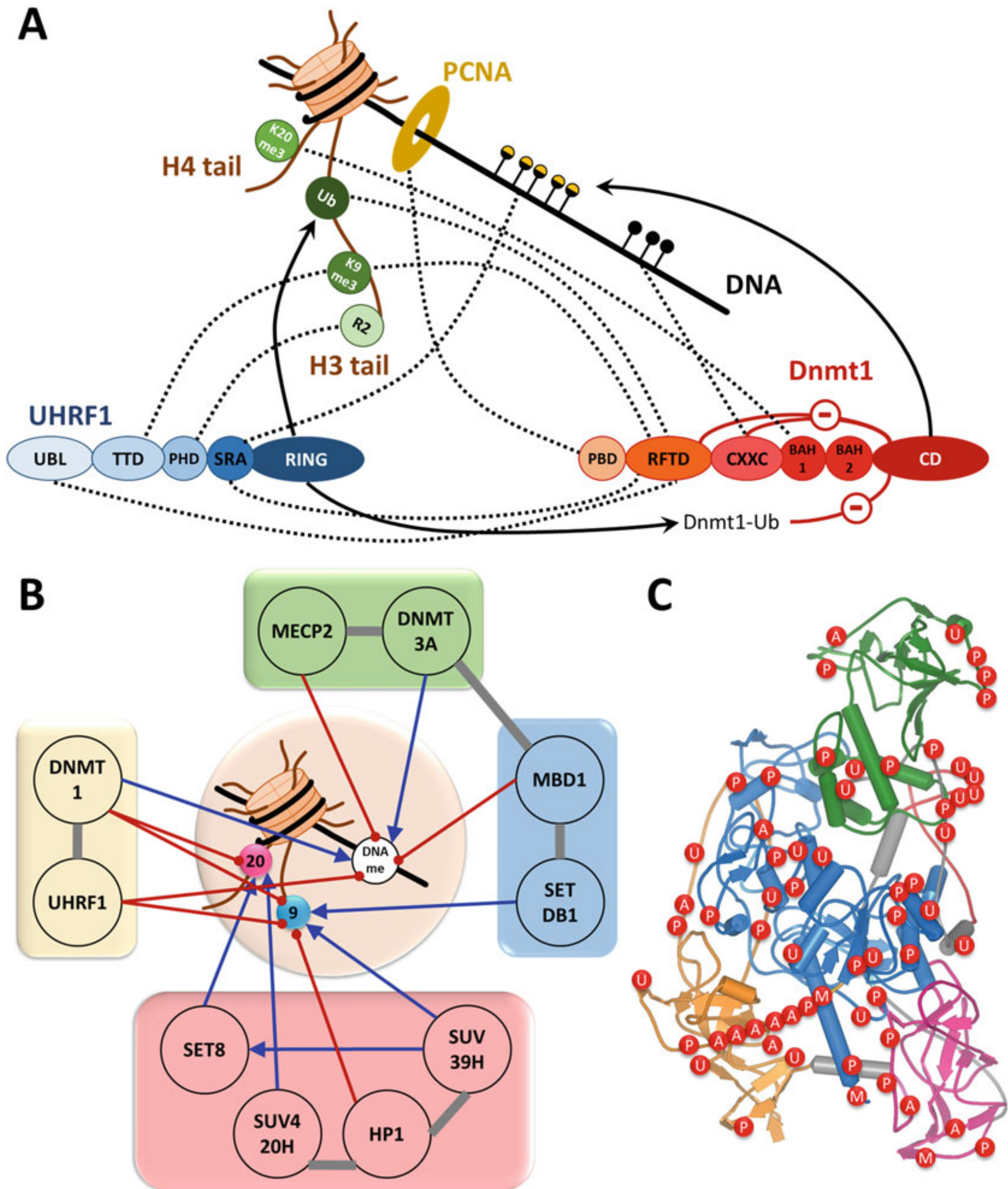


Fig. 4.3 Regulatory networks controlling the activity and stability of DNMT1 and heterochromatic DNA methylation. (**a**) Schematic illustration of the complex interplay between DNMT1, UHRF1, replication forks, and chromatin. Enzymatic activities are indicated by solid lines with arrows. Binding ("reading") interactions are symbolized by dotted lines. For details cf. the text. Abbreviations used: *CD* catalytic domain, *PCNA* Proliferating cell nuclear antigen, *UBL* Ubiquitin-like domain, *TTD* Tandem Tudor domain, *PHD* Plant homeodomain, *SRA* SET and RING-associated domain, *RING* Really interesting new gene, *Ub* ubiquitinated H3 tail. For DNMT1 domain abbreviations, refer to the legend of Fig. 4.1. (**b**) Schematic illustration of the four different chromatin modification sub-networks involved in the establishment and maintenance of the heterochromatic DNA methylation, H3K9me3 (light blue circle) and H4K20me3 (pink circle). Catalytic activities are shown as blue arrows, chromatin reading interactions as dark red lines and protein/protein interaction as gray lines. (**c**) Selection of known PTMs on human DNMT1.

mediated by the RFT and BAH1 domains of DNMT1 further contribute to DNMT1 genomic targeting and methylation of heterochromatic regions (Ren et al. 2020, 2021).

4.3.5.2 The DNMT1–UHRF1 Interaction

Another key pathway of DNMT1 targeting was discovered with the finding that UHFR1 is essential for maintaining DNA methylation in mammals (Bostick et al. 2007; Sharif et al. 2007). UHRF1 specifically binds to hemimethylated DNA via its SET and RING-associated (SRA) domain (Bostick et al. 2007; Avvakumov et al. 2008; Hashimoto et al. 2008; Arita et al. 2008) and its localization to replicating heterochromatin is dependent on the presence of hemimethylated DNA (Sharif et al. 2007) and specific histone PTMs (Nady et al. 2011; Rothbart et al. 2012). UHFR1 co-localizes with DNMT1 and PCNA at replicating heterochromatic regions during mid to late S-phase and DNMT1 association with chromatin is lost in UHFR1 knock out (KO) cells (Sharif et al. 2007; Bostick et al. 2007). It interacts with DNMT1 through the RFTD domain, partially explaining the central role of this domain in the localization of DNMT1 to the replication foci (Leonhardt et al. 1992). Remarkably, the phenotype of the UHFR1 KO in mice mimics that of DNMT1 KO, as UHRF1-deficient embryos die shortly after gastrulation and show significantly reduced levels of DNA methylation (Sharif et al. 2007), indicating that UHRF1 has a central role in the maintenance of DNA methylation. These data led to a model that UHFR1 recruits DNMT1 to the replicated hemimethylated DNA to facilitate its efficient re-methylation (Jeltsch 2008) (Figs. 4.2 and 4.3).

Later, it was found that two domains of UHRF1 recognize histone marks: the tandem Tudor domain (TTD) of UHRF1 binds methylated lysine 9 and unmethylated lysine 4 on histone 3 tail (Nady et al. 2011; Rothbart et al. 2012) and the plant homeodomain (PHD) of UHRF1 binds to unmodified arginine 2 of the H3 tail (Hu et al. 2011; Rajakumara et al. 2011; Wang et al. 2011). The interaction with H3K9me3 is required for the proper localization of UHRF1 to heterochromatin and maintenance of DNA methylation, since a mutation in TTD, which prevents binding to H3K9me3, abolished both functions (Nady et al. 2011; Rothbart et al. 2012). Similarly, disruption of H3R2 binding in UHRF1 abolished DNA methylation by DNMT1 in cells (Qin et al. 2015). These data indicate that the coordinated recognition of two histone marks, H3K9me3 and H3R2, as well as the interaction with hemimethylated DNA by UHRF1, are all necessary for the guidance of DNMT1 and faithful maintenance of DNA methylation (Rothbart et al. 2013; Liu et al. 2013) (Fig. 4.3).

In addition to its role in the targeting of DNMT1, UHRF1 also directly stimulates the catalytic activity of DNMT1, by interacting with the RFT domain of DNMT1 and preventing the autoinhibitory conformation (Berkyurek et al. 2014; Bashtrykov et al. 2014a). Moreover, the RING domain of UHRF1 ubiquitinates H3 at K18 and K23 (Nishiyama et al. 2013; Qin et al. 2015). Ubiquitinated H3 is bound by DNMT1 as described in the next paragraph, increasing its methyltransferase activity. Furthermore, UHRF1 is involved in the ubiquitination of DNMT1, which reduces DNMT1's stability (see below).

However, UHRF1 also forms a stable interaction with DNA Ligase 1 methylated at K126 (which is bound to the TTD instead of H3K9me3) (Ferry et al. 2017). This interaction occurs at replication forks, where the ligase is needed to seal the Okazaki fragments.

Fig. 4.3 (continued) Phosphorylations, acetylations, methylations, and ubiquitinations are represented by red circles labeled with P, A, M, or U, respectively. RFTD is shown in green, CXXC as a red loop, BAH1 in orange, BAH2 in violet, and the catalytic domain in blue. Reprinted from Jeltsch and Jurkowska (2016) with permission from Oxford University Press

Interestingly, the replication-coupled maintenance activity of DNMT1 is determined by the UHRF1–Ligase 1 and PCNA–DNMT1 interactions, while its replication-independent activity depends on nucleosome occupancy and the interaction between UHRF1 and methylated H3K9 (Ming et al. 2020). This finding is in agreement with the observation that the activity of DNMT1 is inhibited by nucleosome formation as mentioned above (Mishima et al. 2017). All these observations demonstrate that UHRF1 is a key multifaceted regulator of DNMT1 and the entire maintenance DNA methylation machinery (Fig. 4.3).

4.3.5.3 Binding of the DNMT1-RFTD to Ubiquitinated H3 Tails

The UHRF1-dependent ubiquitination of histone H3 has an essential role in DNMT1 function, as the catalytically inactive UHRF1 RING mutant failed to recruit DNMT1 to the replication sites (Nishiyama et al. 2013). The molecular mechanism of this finding has begun to be uncovered with the observation that DNMT1 preferentially associates with monoubiquitinated H3 through its RFT domain and that this interaction leads to the activation of the methyltransferase (Nishiyama et al. 2013; Qin et al. 2015). The binding to the monoubiquitinated H3 peptide increased the methylation activity of DNMT1 on a substrate with multiple hemimethylated CpG sites (Mishima et al. 2020), indicating that it may contribute to the efficiency of DNA methylation maintenance. This stimulatory effect was reduced by mutations in the RFTD of DNMT1 that are linked to human autosomal dominant cerebellar ataxia, deafness, and narcolepsy (ADCA-CN) (Mishima et al. 2020).

The ubiquitination of the H3 tail is introduced by the RING domain of UHRF1, which is an E3 ligase (Nishiyama et al. 2013; Qin et al. 2015). Monoubiquitination of H3 has been detected at K14, K18, and K23. Structural and biochemical studies showed that the dual monoubiquitinated H3 (at K18 and K23) peptide is bound preferentially by DNMT1 in a binding cleft located in the RFT domain (Ishiyama et al. 2017). The binding of H3-K18Ub/23Ub results in a conformational change of the RFTD, leading to an increase in DNMT1 activity. In addition, the RFTD also binds the ubiquitin-like (UBL) domain of UHRF1, further strengthening the DNMT1–UHRF1 interaction. Notably, both the ubiquitin ligase and the ubiquitin-like domain of UHRF1 are required for the heterochromatic localization of DNMT1 and DNA methylation of repeat elements (Li et al. 2018). Consistently, the USP7 deubiquitinase, which removes histone ubiquitination, has been shown to suppress DNMT1 recruitment and DNA methylation (Li et al. 2020). In addition, PCNA-associated factor 15 (PAF15) undergoes dual monoubiquitination by UHRF1 in a DNA replication-coupled manner and it thereby recruits DNMT1 to replicating chromatin (Nishiyama et al. 2020). Strikingly, during early S-phase, UHRF1 preferentially ubiquitinates PAF15, whereas H3Ub2 predominates during late S-phase, suggesting that the mechanism of DNMT1 recruitment changes between the early and late replicating DNA regions. Taken together, these data indicate an important additional connection between the chromatin interactions of DNMT1 and UHRF1, which is essential for efficient maintenance methylation to occur (Fig. 4.3).

4.3.5.4 Binding of DNMT1 to Heterochromatic Chromatin Marks

DNA methylation, H3K9me3 and H4K20me3 together constitute a characteristic modification state called constitutive heterochromatin (Jeltsch et al. 2019). Recent work has demonstrated that DNMT1 directly binds to both, H3K9me3 and H4K20me3, which explained the heterochromatic localization of DNMT1 and provided novel connections between these chromatin modifications and DNA methylation. A structural and biochemical study demonstrated that the RFT domain of DNMT1 preferentially interacts with ubiquitinated H3 peptides if they also contain H3K9me3, leading to the stimulation of DNMT1 activity (Ren et al. 2020). Structural analysis revealed the H3 tail bound in a surface cleft of the RFT domain with the interaction sites for ubiquitin on the surface of RFTD and

H3K9me3 binding mediated by a non-conventional binding site formed by RFTD and one ubiquitin moiety. The mutation of a tryptophan residue critical for the H3K9me3 interaction led to a global reduction of DNA methylation in cells, underscoring the functional relevance of this interaction. Another recent study discovered that the first BAH domain of DNMT1 (BAH1) binds H4K20me3, contributing to heterochromatin targeting of DNMT1 and DNA methylation (Ren et al. 2021). Structural analysis revealed binding of the H4 tail to the BAH1 domain and recognition of H4K20me3 by an aromatic half-cage. The binding of the H4 tail led to the displacement of the autoinhibitory linker between the CXXC and BAH1 domains, causing an allosteric activation of DNMT1. Disruption of the H3K9me3 or H4K20me3 binding led to a loss of the heterochromatic localization of DNMT1 in murine ES cells (Ren et al. 2020, 2021).

DNA methylation, H3K9me3 and H4K20me3 form an interconnected network of chromatin modifications that defines the constitutive heterochromatin state (Fig. 4.3b). Previous work has already identified several molecular connections between readers (HP1ß and UHRF1 for H3K9me3 and MBD1 for DNA methylation) and writers of these modifications (SUV39H1/H2 and SETDB1 for H3K9 methylation, SET8 and SUV420H1/H2 for H4K20 methylation). For example, H3K9 methylation stimulates H4K20 methylation, because HP1 recruits SUV420H enzymes (Schotta et al. 2004) and SUV39H1 stimulates the activity of SET8 (Kudithipudi et al. 2017). Similarly, HP1 stimulates further spreading of H3K9 methylation by interaction with SUV39H enzymes (Raurell-Vila et al. 2017). UHRF1 functions as a critical cofactor of DNMT1 (Liu et al. 2013) and DNA methylation recruits SETDB1 via MDB1 binding (Markouli et al. 2021). The data showing that DNMT1 also directly binds to H3K9me3 and H4K20me3 connect this network even more, ensuring efficient methylation and silencing of heterochromatin and repetitive sequences. These complex interactions elegantly illustrate the cooperation between various layers of epigenetic modifications that all establish and reinforce specific epigenetic states and biological outcomes.

4.3.5.5 Regulation of Activity and Specificity of DNMT1 by Nucleic Acid Binding

DNMT1 possesses multiple DNA binding sites, which contribute to the allosteric regulation of its activity and specificity. Many groups reported that the enzyme shows reduced specificity in the presence of methylated DNA (Fatemi et al. 2001, 2002; Christman et al. 1995; Bacolla et al. 1999). This effect was due to an increase in the rate of de novo methylation of unmodified DNA, while the methylation of hemimethylated DNA was weakly inhibited (Fatemi et al. 2001; Goyal et al. 2006). The increase in the methylation efficiency of unmethylated DNA indicates that the binding of the methylated DNA to the N-terminal domain of the enzyme induces an allosteric activation for the methylation of unmethylated substrates. The molecular mechanism of the allosteric activation of DNMT1 is not well understood, the CXXC domain (Fatemi et al. 2001) and the residues 284–287 of the murine DNMT1 (Pradhan and Esteve 2003) have been implicated in this process. Therefore, it is likely that DNA binding to the CXXC domain is involved in these effects. In addition, an inhibitory effect of unmethylated DNA was demonstrated in several studies (Svedruzic and Reich 2005; Flynn et al. 2003; Bacolla et al. 1999), suggesting that binding of an unmethylated DNA to the N-terminal part of DNMT1 leads to a repression of the enzymatic activity on hemimethylated DNA. The binding site for this substrate inhibition effect was localized in the first 501 amino acids of DNMT1 (Bacolla et al. 2001). Additional evidence suggests that binding of the methylated DNA to the N-terminal inhibition site also caused de-repression of the enzyme (Bacolla et al. 2001). Whether the inhibition and stimulation effects observed in these various studies are due to binding to the same or different sites and to what extent different DNAs compete for the different sites is not clear.

Interestingly, all studies agree that binding to unmethylated DNA at a secondary site reduces

the activity of DNMT1, while binding to methylated DNA increases its activity. This observation could be related to the fact that DNA methylation patterns in the human genome are highly bimodal (Eckhardt et al. 2006; Meissner et al. 2008; Zhang et al. 2009), meaning that the genomic regions tend to be either highly methylated or almost unmethylated. The occurrence of the bimodal methylation patterns could be explained by the allosteric binding of the substrate DNA to a secondary site because DNMT1 would be activated on methylated regions and inactivated on unmethylated DNA. Consequently, highly methylated regions will tend to gain methylation, whereas lowly methylated regions will tend to lose even their residual methylation.

In addition to DNA, DNMT1 binds various RNA molecules. Initial studies showed that DNMT1 purified from insect cells contains inhibitory RNA (Glickman et al. 1997a). Later, it was discovered that RNA binding regulates the activity of DNMT1 in a locus-specific manner. A long non-coding RNA (lncRNA) originating from the CEBPA locus was observed to bind and inhibit DNMT1 and prevent the methylation of this locus. Similar effects were observed for several other loci on a genomic scale (Di Ruscio et al. 2013). Based on these findings, the authors proposed a model, in which the ncRNAs transcribed at one locus function as a shield for this locus preventing its methylation. Thereby, the expression of the locus would be perpetuated. Later, it was also reported that DNMT1 binds to miRNAs like miR-155-5p (Zhang et al. 2015a). Other studies showed regulation of DNMT1 by DNMT1-associated lncRNAs, leading to global changes in DNA methylation and gene regulation in cancer cells (Merry et al. 2015; Somasundaram et al. 2018). A specific example of this mechanism is the DACOR1 lncRNA, which is a positive regulator of DNA methylation (Somasundaram et al. 2018). Similar to lncRNAs, miRNAs function as inhibitors of DNMT1 and the transfection of miRNAs to cells caused changes in cellular methylation (Zhang et al. 2015a). RNA binding was mapped to the catalytic domain of DNMT1 (Di Ruscio et al. 2013; Zhang et al. 2015a), and it was reported that miRNAs can act as DNA competitive inhibitors (Zhang et al. 2015a). These findings suggest that the inhibition of DNMT1 by miRNAs is based on a direct competition of the RNA and DNA for access to the catalytic center. However, it is well conceivable that the additional DNA binding sites described above bind regulatory RNAs as well. These important features of the interaction of DNMT1 with regulatory DNA and RNA are not well understood at a molecular level and deserve additional experimental work.

4.3.6 PTMs of DNMT1

4.3.6.1 Phosphorylation of DNMT1

DNMT1 is subject to several posttranslational modifications like phosphorylation, methylation, ubiquitination, acetylation, and SUMOylation, (Fig. 4.3c). Following the initial identification of S515 as a major phosphorylation site in DNMT1 purified from insect cells (Glickman et al. 1997b), several more phosphorylated serine and threonine residues have been identified in targeted and high-throughput proteomics approaches with DNMT1 purified from human or mouse cells. Currently, >60 phosphorylation sites have been mapped on human and mouse DNMT1 (http://www.phosphosite.org), but only a few of them have been functionally studied. The phosphorylated S515 is involved in the interaction between the N-terminal and catalytic domains of DNMT1 which is necessary for the activity of the enzyme (Goyal et al. 2007). Phosphorylation of S146 introduced by casein kinase 1 delta/epsilon decreases the DNA binding affinity of DNMT1 (Sugiyama et al. 2010), and phosphorylation of S127 and S143 regulates the interaction of DNMT1 with PCNA and UHRF1 (Hervouet et al. 2010). Moreover, phosphorylation of DNMT1 by PKC has been reported, but the target sites have not yet been identified (Lavoie et al. 2011). S143 of DNMT1 is phosphorylated by AKT1, which leads to the stabilization of the methyltransferase (Esteve et al. 2011). A specific 14–3-3 family reader protein for this modification has been identified

(Esteve et al. 2016). It binds phosphorylated DNMT1, leading to the inhibition of DNMT1 activity, aberrant DNA methylation, and cell invasion (Esteve et al. 2016). The functional significance of many of the other phosphorylations in DNMT1 still awaits elucidation. In particular, the influence of the PTMs on the allosteric regulation of DNMT1 activity and specificity needs to be studied.

4.3.6.2 Acetylation and Ubiquitination of DNMT1

Multiple acetylation sites have been identified on DNMT1 up to date in proteomics analyses (Kim et al. 2006; Choudhary et al. 2009; Peng et al. 2011) (http://www.phosphosite.org); however, their functional significance has only begun to be revealed. Initial experiments with deacetylase inhibitors demonstrated the involvement of acetylation in the control of DNMT1 stability (Zhou et al. 2008; Peng et al. 2011). Based on this, an elegant mechanism regulating the abundance of DNMT1 during cell cycle was identified. It starts with the acetylation of DNMT1 in the KG linker by the acetyltransferase Tip60, followed by UHRF1-mediated ubiquitination, resulting in proteasomal degradation of DNMT1 at the end of DNA replication. In turn, histone deacetylase 1 (HDAC1) and deubiquitinase ubiquitin-specific peptidase 7 (USP7, also known as HAUSP) have an opposite effect and increase the stability of DNMT1 (Du et al. 2010; Qin et al. 2011). The crystal structure of DNMT1 in complex with USP7 revealed that this interaction is dependent on the KG linker of DNMT1, explaining why acetylation of this region impairs complex formation and promotes degradation of DNMT1 (Cheng et al. 2015). In addition, SIRT1 deacetylates DNMT1 at several sites and thereby regulates the activity and function of the methyltransferase (Peng et al. 2011).

4.3.6.3 Lysine Methylation of DNMT1

DNMT1 is methylated by SET7/9, both in vivo and in vitro. The monomethylation of human DNMT1 by SET7/9 occurs at K142 mainly during late S-phase and promotes proteasomal degradation of the enzyme in a cell cycle-dependent manner (Esteve et al. 2009). Recent work demonstrated that proteasomal targeting of DNMT1 is mediated by the L3MBTL3 methyl-binding protein that recruits CRL4(DCAF5) ubiquitin ligase (Leng et al. 2018). Methylation of DNMT1 is reversible and can be removed by LSD1 (Wang et al. 2009; Leng et al. 2018). In addition, it is antagonistic with phosphorylation of DNMT1 at S143 by AKT1 kinase described above (Esteve et al. 2011). The existence of these complex mechanisms to regulate DNMT1 stability underscores the biological requirement for tight regulation of cellular DNMT1 levels.

4.4 Structure, Function, and Mechanism of DNMT3 Enzymes

4.4.1 Domain Composition of DNMT3 Proteins

In most mammals, the DNMT3 family contains three members: DNMT3A, DNMT3B, and DNMT3L, which in humans comprise 912 aa, 853 aa, and 387 aa, respectively. In addition, a DNMT3B paralog called DNMT3C (739 aa) has been identified in rodents, where it has a specific role in transposon repression in the male germline (Barau et al. 2016; Jain et al. 2017) (Fig. 4.1). Several isoforms of DNMT3A and DNMT3B, resulting from alternative splicing or use of alternative start codons, have been identified both in mice and humans (Jurkowska et al. 2011a). In the case of DNMT3A, the DNMT3A2 isoform lacks the first 223 amino acid residues (Qiu et al. 2002). For DNMT3B, multiple isoforms have been found (Weisenberger et al. 2004); among them the inactive splicing isoform DNMT3B3, which lacks parts of the linker between the PWWP and ADD domains and a region of the catalytic domain containing the target recognition domain (Fig. 4.1). Besides the C-terminal domain required for catalysis, DNMT3A and DNMT3B possess an N-terminal part with domains involved in the targeting of the enzymes to chromatin and regulation of their function (Jurkowska et al. 2011a). In this part, three functional domains are

present: a UDR (ubiquitin-dependent recruitment region) which is present specifically in DNMT3A1, a PWWP domain in DNMT3A and DNMT3B, and an ADD (ATRX-DNMT3-DNMT3L) domain, also known as PHD (Plant homeodomain) domain that is present in all four proteins.

The ADD domain is a cysteine-rich region that binds zinc ions and creates a platform for protein–protein interactions. This domain mediates the interaction of DNMT3 enzymes with histone H3 tails unmethylated at lysine K4 (Ooi et al. 2007; Otani et al. 2009; Zhang et al. 2010; Guo et al. 2015). In addition, it is involved in the interaction of DNMT3A with various components of the epigenetic machinery, like protein lysine methyltransferases SUV39H1 (Fuks et al. 2003a), SETDB1 (Li et al. 2006), EZH2 (Vire et al. 2006), and deacetylase HDAC1, reading domain proteins, including HP1ß (Fuks et al. 2003a), Mbd3 (Datta et al. 2005), and MeCP2 (Kimura and Shiota 2003; Fuks et al. 2003b; Rajavelu et al. 2018), as well as transcription factors PU.1 (Suzuki et al. 2006), MYC (Brenner et al. 2005), and RP58 (Fuks et al. 2001), and chromatin remodeling factors hSNF2 (Geiman et al. 2004) and SMARCA4 (Datta et al. 2005). The ADD domain has been implicated in the allosteric control of DNMT3A, as it interacts with the catalytic domain of the methyltransferase and inhibits its activity (see below), indicating that ADD-mediated interactions with other proteins and chromatin could have direct regulatory effects on the catalytic activity of DNMT3A and DNMT3B.

The PWWP domain of DNMT3A and DNMT3B is a region of 100–150 amino acids, containing a conserved proline–tryptophan motif (hence the name PWWP). PWWP domains belong to the Royal domain superfamily, members of which interact with histone tails in various modification states (Qin and Min 2014). The PWWP domains of DNMT3A and DNMT3B specifically recognize the H3K36 di- and trimethylation mark (H3K36me2/3) (Dhayalan et al. 2010). This domain is essential for the targeting of DNMT3 enzymes to pericentromeric chromatin (Chen et al. 2004; Ge et al. 2004). The structures of the PWWP domains from both DNMT3A and DNMT3B have been solved (Qiu et al. 2002; Rondelet et al. 2016; Dukatz et al. 2019a). A biochemical study revealed that the PWWP domain synergistically binds the H3K36me2/3-modified histone tail and DNA through its conserved aromatic cage for H3K9me2/3 binding and a positively charged surface for DNA binding. Both interfaces were found to be necessary for chromatin targeting of DNMT3A1 (Dukatz et al. 2019a). In addition, the ZHX1 (zinc-finger and homeobox protein 1) interacts with the PWWP domain of DNMT3B and enhances DNMT3B-mediated transcriptional repression (Kim et al. 2007). Interestingly, although DNMT3C arose from a duplication of the DNM3B gene, it lost the PWWP domain. This may prevent targeting of the enzyme to H3K36me2/3-rich regions, potentially contributing to the specific localization of DNMT3C to retrotransposon promoters (Barau et al. 2016).

The part of DNMT3A and DNMT3B N-terminal to the PWWP domain is the least conserved region between both enzymes. This domain binds DNA (Suetake et al. 2011) and it is important for anchoring the enzymes to nucleosomes (Jeong et al. 2009; Baubec et al. 2015). In DNMT3A1, a small, folded domain called ubiquitin-dependent recruitment (UDR) domain has been recently identified. It is responsible for the interaction of DNMT3A1 with H2AK119ub1 (Weinberg et al. 2021). DNMT3A2 and DNMT3B enzymes lack this domain.

The C-terminal domains of DNMT3A and DNMT3B, which enclose the catalytic centers of the enzymes, share approximately 85% sequence homology. In contrast to the catalytic domain of DNMT1 they are active in an isolated form (Gowher and Jeltsch 2002) and have been used as a model system to study the catalytic mechanism and specificity of the DNMT3 proteins. Interestingly, isolated catalytic domains of DNMT3A and DNMT3B show higher enzymatic activity than the full-length proteins, indicating that the N-terminal domains allosterically inhibit the activity of the enzymes (Li et al. 2011). The

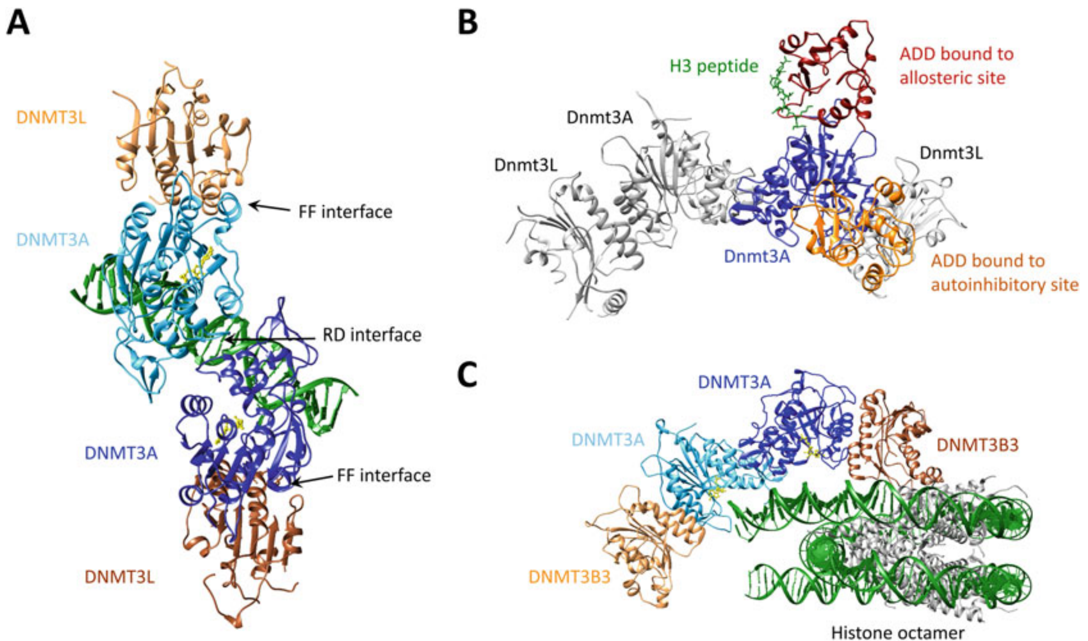


Fig. 4.4 Structure and allosteric regulation of DNMT3A. (**a**) Structure of the DNMT3A/DNMT3L complex with bound DNA (pdb 5YX2) (Zhang et al. 2018). Subunits and interfaces are annotated. AdoHcy is shown in yellow as a ball and stick model. (**b**) Allosteric regulation of DNMT3A. The ADD domain of the dark blue DNMT3A subunit is shown in both the autoinhibitory (orange) and in the active conformation (red) (pdb 4U7P and 4U7T) (Guo et al. 2015). The ADD domain of the second DNMT3A subunit (gray) has been omitted for clarity. Binding of the H3 peptide (green) to the ADD domain occurs with the residues involved in the autoinhibitory-binding interface. Therefore, H3 peptide binding is only possible in the active conformation and this conformation is consequently stabilized in the presence of the H3 peptide (Guo et al. 2015; Li et al. 2011). (**c**) Cryo-EM structure of the DNMT3A/DNMT3B3 nucleosome complex (pdb 6PA7) (Xu et al. 2020). Nucleosomal DNA is shown in green and AdoHcy is shown in yellow as ball and stick model

molecular mechanism underlying this observation was revealed by a structural study, which demonstrated that the ADD domain of DNMT3A, which directly interacts with the catalytic domain of the methyltransferase in two different binding modes (see below), is responsible for this inhibition in the absence of histones (Guo et al. 2015) (Fig. 4.4b). This model is further supported by kinetic experiments, showing that the binding of ADD domain of DNMT3A to H3 tail stimulates the activity of the enzyme (Li et al. 2011; Zhang et al. 2010).

DNMT3L, the third member of the DNMT3 family, lacks parts of the N-terminal region including the PWWP domain. Strikingly, it also carries amino acid exchanges and deletions within the conserved DNA-(cytosine C5)-MTase motifs, which contain the catalytic residues, indicating that while it still adopts the typical AdoMet-dependent MTase fold described above, it cannot have catalytic activity and is unable to bind AdoMet. The same is true for one splicing isoform of DNMT3B, DNMT3B3, which also contains a deletion in the C-terminal domain and lacks catalytic activity (Weisenberger et al. 2004; Zeng et al. 2020). While DNMT3L is mainly expressed in ES cells and the germline (Bourc'his and Bestor 2004; Bourc'his et al. 2001; Hata et al. 2002), DNMT3B3 shows expression in differentiated cells (Zeng et al. 2020). Despite being inactive, both DNMT3L and DNMT3B3 interact with the active members of the DNMT3 family and stimulate their catalytic activity.

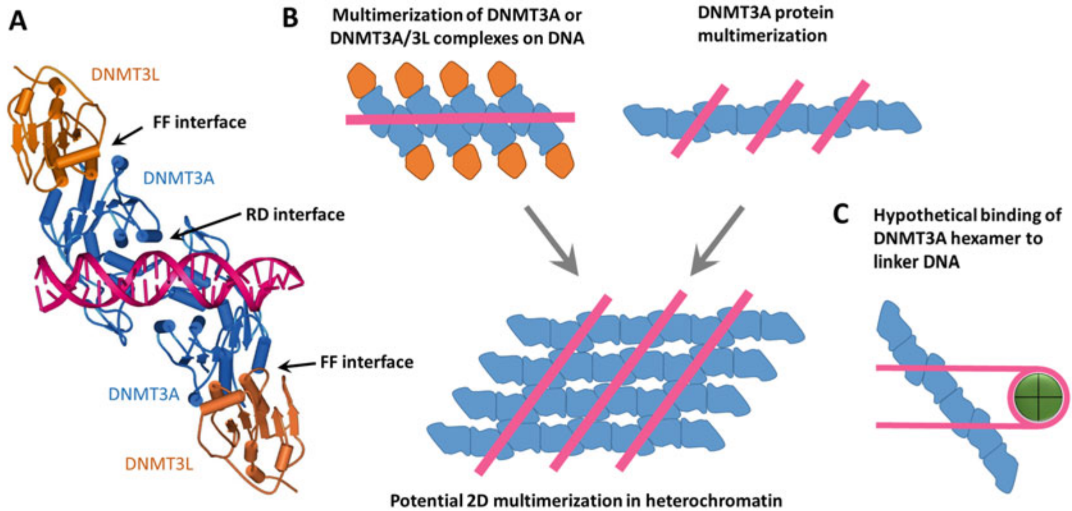


Fig. 4.5 Multimerization of DNMT3A and DNMT3A/DNMT3L complexes. (**a**) Structure of the DNMT3A/DNMT3L complex with bound DNA (pdb 5YX2) (Zhang et al. 2018). Subunits and interfaces are annotated. (**b**) Schematic models of DNMT3A multimerization on DNA, protein multimerization and binding to several DNA molecules, and the combination of both processes. (**c**) Hypothetical binding of a DNMT3A hexamer to the two linker DNAs emerging from one nucleosome

4.4.2 Structures of DNMT3A and DNMT3B

The structure of the complex of the C-terminal domains of DNMT3A/DNMT3L was solved in 2007 and represented the first structure published for a mammalian DNMT. It showed that the complex forms a linear heterotetramer consisting of two DNMT3L subunits (at the edges of the tetramer) and two DNMT3A subunits (in the center) (Jia et al. 2007) (Fig. 4.4a). The heterotetrameric structure of the complex was confirmed in solution (Jurkowska et al. 2008). The structure also revealed that the C-terminal domain of DNMT3A contains two interfaces for protein–protein contacts: a hydrophobic one generated by the stacking interaction of two phenylalanine residues (called FF interface), which mediates the DNMT3A/DNMT3L interaction, and a polar interface generated by a hydrogen bonding network between arginine and aspartate residues from both subunits (called RD interface), which can only mediate DNMT3A/DNMT3A interactions since the corresponding region is absent in DNMT3L (Fig. 4.5). DNA binding studies showed that the central DNMT3A/DNMT3A interface in the tetramer creates the DNA binding site, while both interfaces are essential for AdoMet binding and catalytic activity (Jurkowska et al. 2008). The dimerization of DNMT3A/DNMT3L complexes via the RD interface increases the size of the DNA interface and compensates for the small TRD of DNMT3A.

Later, the structure of the DNMT3A/DNMT3L C-terminal domain heterotetramer was solved in complex with a DNA molecule containing two CpG sites spaced in a distance of 12 base pairs (Zhang et al. 2018) (Fig. 4.4a). It provided the first mechanistic insights into the DNA interaction and specificity of DNMT3A. Zebularine was incorporated into the DNA instead of the target cytosines in the upper strand of the left CpG site and the lower strand of the right CpG site. This base analog leads to the formation of stable covalent complexes between the DNMT and the DNA because the nucleophilic attack of the active site cysteine residue is

catalyzed, but its later elimination is blocked. The complex showed base flipping of both zebularine bases, indicating that the heterotetramer could potentially co-methylate CpG sites at this distance. Biochemical studies confirmed that the 12 bp distance is the preferred one for covalent DNA complex formation of DNMT3A/DNMT3L and DNMT3B/DNMT3L (Gao et al. 2020a), further supported by strong peaks of co-methylation at CpG sites placed in this distance in substrates containing two CpG sites (Emperle et al. 2021). The DNMT3A–DNA interaction involves a target recognition domain (TRD) loop, a catalytic loop following the catalytic PCN motif, and a helix of the RD tetramer interface (Zhang et al. 2018). The TRD loop (which is unfolded in the DNA free complex) contains the R836 residue, which recognizes the guanine of the CpG sites, ensuring the preference of DNMT3A towards CpG observed in previous studies (Gowher and Jeltsch 2001; Aoki et al. 2001; Ramsahoye et al. 2000). V716 from the catalytic loop approaches the DNA from the minor groove and fills the DNA cavity generated by the flipping of the zebularine base. The RD interface loop contains R882, which is often mutated in acute myeloid leukaemia (AML) (see below). It contacts the DNA backbone at several phosphate residues on the 3′ side of the CpG site. The central part of the DNA shows about 40° bending and kinetic experiments demonstrated that enrichment of T in the region of bending stimulates methylation (Emperle et al. 2021).

Recent structures of DNMT3B/DNMT3L C-terminal domain heterotetramers with DNA revealed a very similar overall structure as the DNMT3A/DNMT3L complex (Gao et al. 2020b; Lin et al. 2020). Strikingly, despite similarities, the DNA recognition of both enzymes differs in the target recognition loop. In DNMT3B, N779 interacts specifically with the guanine in CpG sites, while in DNMT3A this interaction is mediated by R836. Moreover, DNMT3B contains a lysine residue (K777) which specifically interacts with the base at the +1 side of the CpG, mediating a strong preference for a G at this place, in particular during non-CpG methylation. The amino acid sequences and structures of DNMT3A and DNMT3B diverge most at the RD interface loop, as illustrated, for example, by a different conformation of R823 in DNMT3B, which corresponds to DNMT3A R882. These differences lead to distinct contacts to the DNA regions flanking the target CpG site and provide a mechanistic basis of the enzyme-specific flanking sequence preferences (see below) (Gao et al. 2020b).

In a seminal publication, a cryo-EM structure of a DNMT3A2/DNMT3B3 heterotetramer bound to a mononucleosome was reported (Xu et al. 2020) (Fig. 4.4c). The complex formed a similar linear heterotetramer as the DNMT3A/DNMT3L complexes, but DNMT3B3 replaced DNMT3L at the outer complex positions. This study confirms previous biochemical data (Li et al. 2007) showing that the DNMT3 binding interfaces support the interaction of different DNMT3 members, offering the unique potential for regulating methyltransferase activity depending on the complex composition. Unexpectedly, two arginine residues in the C-terminal domain of one DNMT3B3 subunit formed a direct contact with the H2A/H2B acidic patch on the disc face of the histone octamer. Thereby, the DNMT3A2/DNMT3B3 tetramer was anchored on the nucleosome core particle positioning the DNA binding region and active sites of the central DNMT3A subunits right above the linker DNA strand near the dyad axis (Xu et al. 2020). The detailed functional consequences of this unexpected architecture are still unknown, but biochemical data showed that binding of DNMT3A and DNMT3A/DNMT3B3 complexes to the acidic patch of histones contributes to the methylation preferences of CpG sites within the linker DNA (Bröhm et al. 2022).

4.4.3 Allosteric Regulation of DNMT3A

Additional structures of a longer DNMT3A C-terminal fragment also including the ADD domain in complex with DNMT3L were solved, providing seminal insights into the mechanism of this enzyme. They showed that the ADD domain

can bind to the catalytic domain at two distinct sites, creating two alternative conformations. ADD binding activates the enzyme in one conformation (allosteric binding), while it blocks access of the DNA to the active center and inhibits catalysis in the other (autoinhibitory binding) (Guo et al. 2015) (Fig. 4.4b). A modeling study suggested that a hinge-like property of the RD interface is important for the cooperative reorientations of the tetramer into the autoinhibitory or the active state (Liang et al. 2018).

These data indicate that the activity of DNMT3A, like DNMT1, is under precise allosteric control by domain rearrangements, illustrating a fascinating convergence of regulatory principles of these two enzymes. Similarly as in DNMT1, protein partners can influence the equilibrium of the active and inactive conformations, as it was shown that the stimulatory effect of H3 on DNMT3A depends on its binding to the ADD domain, leading to the stabilization of the ADD at the allosteric binding site (Li et al. 2011; Guo et al. 2015). Direct allosteric regulation of DNMT3A activity by the PWWP domain has not yet been shown. Moreover, so far it is not known if DNMT3B undergoes similar steps of allosteric regulation as DNMT3A.

4.4.4 Specificity of DNMT3 Enzymes

Consistent with their designation as de novo MTases, DNMT3A and DNMT3B do not display any significant preference between hemimethylated and unmethylated DNA (Okano et al. 1998; Gowher and Jeltsch 2001). However, in addition to their preference for the methylation of CpG sites, both DNMT3A and DNMT3B are very sensitive to the sequences flanking their target sites. This is illustrated by the finding that CpG sites in certain flanking sequences cannot be methylated by DNMT3A at all (Jurkowska et al. 2011c). It has been shown that purine bases are preferred at the 5′ side of the CpG sites, whereas pyrimidines are favored at their 3′ side (Lin et al. 2002; Handa and Jeltsch 2005; Jurkowska et al. 2011c). One further consequence of the strong flanking sequence preferences of DNMT3A and DNMT3B is that both DNA strands of a CpG site, which are embedded in an asymmetric flanking sequence context, usually differ strongly in their preference for DNMT3 methylation. This leads to the preferential methylation of one cytosine in each CpG site, meaning that DNMT3 enzymes tend to generate hemimethylated products. In vitro experiments showed that the products of DNMT3A methylation can be readily methylated by DNMT1 and that both enzymes can act synergistically in the efficient de novo methylation of unmethylated DNA (Fatemi et al. 2002). Mutational analysis of residues in the DNA binding site of DNMT3A demonstrated that exchanges of critical residues caused massive changes in flanking sequence preferences (Gowher et al. 2006). Interestingly, this includes the exchange at R882, a residue frequently mutated in acute myeloid leukemia (AML) cancer (Hamidi et al. 2015). An in-depth mechanistic understanding of the CpG recognition and flanking sequence preferences of DNMT3 enzymes was provided by the recent DNMT3 structures with bound substrate DNA (Zhang et al. 2018; Gao et al. 2020b; Lin et al. 2020), as described below.

Although DNMT3A and DNMT3B methylate cytosine residues predominantly in the context of CpG dinucleotides, they can also introduce methylation in a non-CpG context (CA > > CT > CC) (Gowher and Jeltsch 2001; Aoki et al. 2001; Ramsahoye et al. 2000). Consistently, methylated non-CpG sites (mainly CpA) were detected in embryonic stem (ES) cells and the brain, where DNMT3A and DNMT3B enzymes are highly expressed, but not in cells where DNMT3 enzymes are downregulated (Lister et al. 2009, 2013; Varley et al. 2013; Guo et al. 2014). However, another survey of the human body epigenomes identified low levels of non-CpG methylation in almost all human tissues (Schultz et al. 2015). Studies with DNMT KO cell lines confirmed that DNMT3 enzymes introduce the non-CpG methylation (Ziller et al. 2011; Arand et al. 2012). The exact mechanism for the propagation of DNA methylation outside of the CpG context is unknown, but it cannot be maintained

by DNMT1, which is very specific for CpG sites (Fatemi et al. 2001). First insights into the biological function of non-CpG methylation were provided with the observation that it can repress expression of long genes in the brain by recruiting MeCP2 (Guo et al. 2014; Gabel et al. 2015; Chen et al. 2015), disruption of which is implicated in the Rett syndrome.

Detailed biochemical studies demonstrated that the flanking sequence preferences of DNMT3A and DNMT3B differ (Gao et al. 2020b; Mallona et al. 2021; Jeltsch et al. 2021). This is due to the DNMT3B specific readout of the +1 flanking site by K777 and the differences in the conformation of the RD interface loop (Gao et al. 2020b; Lin et al. 2020). DNMT3B displays strong and characteristic preferences for CpG sites located in a sequence context that resembles the SatII minor satellite repeats, which lose methylation in the immunodeficiency-centromeric instability-facial anomalies (ICF) syndrome (Xu et al. 1999). This finding explains previous observations showing that 1) insufficient DNMT3B activity causes the ICF syndrome and 2) DNMT3A apparently cannot take over the function of DNMT3B in minor satellite methylation. Currently, specific details of how DNMT3A and DNMT3B interact with the CpG sites in different flanking sequence contexts are unknown. A recent DNMT3B mutational study indicated that the interaction with different flanking sequences involves complex and sequence-dependent contact networks of enzyme residues with the DNA (Dukatz et al. 2020). These adaptive interaction modes could help to balance the interaction of DNMT3B with different flanking sites, allowing a more equal methylation of CpG sites in different contexts, which is required for the function of DNA methylation as a system for storage and processing of epigenetic information (Jeltsch et al. 2021).

The differences in the flanking sequence preferences of DNMT3A and DNMT3B were even more pronounced in the context of non-CpG methylation (Gao et al. 2020b; Dukatz et al. 2020; Jeltsch et al. 2021). Here, DNMT3B showed a strong preference for a G at the +1 flanking site, mainly generating methylated CAG, while DNMT3A preferred a C instead, yielding predominantly methylated CAC. These preferences are in agreement with the cellular distribution of non-CpG methylation obtained in triple DNMT1/DNMT3A/DNMT3B KO cells reconstituted with DNMT3A- or DNMT3B (Gao et al. 2020b) and other cellular methylation data (Lister et al. 2009, 2013; Laurent et al. 2010; Xie et al. 2012; Lee et al. 2017).

Mutations of R882 in DNMT3A, most prominently R882H, are observed at a high frequency in the AML tumors (Hamidi et al. 2015). As described above, R882 is located in the RD loop at the DNA binding interface of DNMT3A and it is involved in flanking sequence DNA contacts on the 3′ side of the CpG site (Zhang et al. 2018). This loop shows conformational differences between DNMT3A and DNMT3B, which are related to the distinct flanking sequence preferences of both enzymes (Gao et al. 2020b). The structure of the DNMT3A(R882H)/DNMT3L heterotetramer bound to DNA showed enhanced dynamics of the TRD loop (Anteneh et al. 2020), suggesting that this loop recognizes CpG dinucleotides in a + 1 flanking site-dependent manner. Accordingly, the R882H mutation leads to strong changes in the flanking sequence preferences of DNMT3A (Emperle et al. 2018, 2019). Other mutations of R882 were shown to cause similar effects, indicating that the loss of the R882 side chain is responsible for the effect, rather than the amino acid side chain introduced instead (histidine 882 in case of the R882H mutation) (Emperle et al. 2019). Interestingly, detailed analyses revealed that the DNMT3A R882H flanking sequence preferences differ from wildtype DNMT3A mainly on the 3′-side of the CpG site, where they change into a DNMT3B-like pattern (Emperle et al. 2019; Norvil et al. 2020). Hence, the changes in flanking sequence preferences are one potential reason for the pathogenic effect of this mutation.

4.4.5 Kinetic Mechanism of DNMT3 Enzymes

Initial studies with the C-terminal domains of DNMT3A and DNMT3B showed an interesting difference in the catalytic mechanism of both enzymes. Whereas DNMT3B was able to methylate multiple CpG sites in a processive manner, DNMT3A was distributive (Gowher and Jeltsch 2002). Later, Reich and colleagues reported that DNMT3A methylates DNA in a processive manner (Holz-Schietinger and Reich 2010). However, at the same time, DNMT3A was shown to bind cooperatively to DNA forming large multimeric protein/DNA fibers (Jia et al. 2007; Jurkowska et al. 2008; Rajavelu et al. 2012) (Fig. 4.5). These properties appear mutually exclusive because the concept of a processive turnover is based on isolated enzyme complexes moving along a DNA substrate, which is not compatible with protein complexes multimerizing on DNA. Other biochemical studies did not detect processive DNA methylation by DNMT3A (Emperle et al. 2014).

4.4.6 Oligomerization of DNMT3 Enzymes

The DNMT3 enzymes exhibit a complex oligomerization and multimerization potential including two independent orthogonal multimerization reactions [for a review cf. (Jeltsch and Jurkowska 2013)]. First, DNMT3A multimerizes on DNA and cooperatively binds to DNA, and second, it can form protein oligomers able to bind to more than one DNA molecule. These two processes will be further described in the next sub-chapters (Fig. 4.5).

4.4.6.1 Protein Multimerization of DNMT3 Enzymes

Besides forming heterotetrameric complexes with DNMT3L, DNMT3A alone also forms homotetrameric structures and higher aggregates. The reason for this is that the FF interface of the DNMT3A/DNMT3L tetramer is symmetric so it also supports the homotypic interaction of two DNMT3A molecules in addition to the heterotypic interaction of DNMT3A with DNMT3L. Hence, each DNMT3A subunit contains two interfaces for homotypic interactions, the RD interface and the FF interface, explaining why it can form tetramers in which two additional DNMT3A subunits replace DNMT3L. Biochemical studies indeed demonstrated that DNMT3A catalytic domain and DNMT3A2 homotetramers are formed in the absence of DNMT3L (Jurkowska et al. 2011b; Nguyen et al. 2019). At higher protein concentrations further multimerization can occur and generate protein fibers (Fig. 4.5b), which can lead to reversible aggregation of DNMT3A as observed in different studies (Jurkowska et al. 2011b; Kareta et al. 2006). Notably, the addition of DNMT3L directs the preferential formation of defined DNMT3A/DNMT3L heterotetramers that cannot extend anymore, because DNMT3L does not contain an RD interface, and therefore functions as a cap in protein multimerization. As described below, this process has been implicated in the release of DNMT3A from heterochromatic sites by the addition of DNMT3L (Jurkowska et al. 2011b).

Since each RD interface of a multimeric DNMT3A oligomer constitutes a potential DNA binding site, the protein oligomers can bind to more than one DNA molecule, provided that they are oriented roughly in parallel, as shown by biophysical experiments (Jurkowska et al. 2011b). Strikingly, modeling suggests that a DNMT3A hexamer could simultaneously bind to the two linker DNAs emerging from one nucleosome (Fig. 4.5c). The ability to form protein oligomers apparently plays a central role in the heterochromatin localization of DNMT3A, as non-oligomerizing DNMT3A mutants affected at these interfaces lost the ability to bind to heterochromatin, despite the presence of intact PWWP and ADD domains. Since heterochromatic DNA is densely packed, it can provide several DNA strands for DNMT3A interaction in matching geometry, and this might contribute to guiding DNMT3A to pericentromeric chromatin.

Recently, a biochemical and molecular dynamics study revealed that the R882H cancer mutation also has a direct effect on the multimerization of DNMT3A as it was found that R882H/R822H RD interfaces were more preferred than WT/WT RD interfaces (Mack et al. 2022). Interestingly, one consequence of this finding is that R882H/WT FF-heterodimers preferentially assemble WT-R882H-R882H-WT heterotetramers, in which all enzymatic activity comes from the two central R882H subunits. Hence, this mechanism provides an elegant explanation why in R882H/WT heterozygous cells the R882H mutant behaves dominantly.

Despite significant progress in dissecting protein multimerization of the DNMT3 enzymes, many questions are still not resolved. For example, DNMT3A has been shown to form catalytically active heterodimers with DNMT3B that use the same interfaces as described above for DNMT3A (Li et al. 2007). In addition, a recent cryo-EM structure illustrated that DNMT3B3, a splice isoform of DNMT3B, can replace DNMT3L at the outer complex positions, forming linear DNMT3A2/DNMT3B3 heterotetramers (Xu et al. 2020). However, the relative affinities for the homotypic DNMT3A and DNMT3B, as compared to the heterotypic interaction of DNMT3A and DNMT3B at the two interfaces are currently unknown. Moreover, the relative preferences for binding DNMT3L at the FF interface are also unknown, although the formation of defined heterotetramers of DNMT3A and DNMT3L suggests that the DNMT3A/DNMT3L interaction is preferred over the DNMT3A/DNMT3A interaction. Finally, the direct proof for the existence of DNMT3 protein multimers in cells that are larger than the tetrameric structure observed in the DNMT3A/DNMT3L complex still needs to be provided. Nevertheless, the availale biochemical and structural data confirmed that the DNMT3 binding interfaces support the interaction of different DNMT3 members, offering a unique potential for regulating methyltransferase activity depending on the complex composition.

4.4.6.2 Multimerization of DNMT3A and DNMT3A/DNMT3L on DNA

As described above, DNMT3A forms a linear heterotetrameric complex with DNMT3L, in which two central DNMT3A subunits interact via the RD interface and generate the DNA binding pocket (Fig. 4.5a) (Jia et al. 2007; Jurkowska et al. 2008). DNA binding by DNMT3A is non-specific (Rajavelu et al. 2012) and DNMT3A (and DNMT3A/DNMT3L) complexes polymerize on DNA by binding next to each other and forming DNMT3A–DNA filaments (Jurkowska et al. 2008; Rajavelu et al. 2012) (Fig. 4.5). A productive interaction of DNMT3A complexes with neighboring CpG sites is possible if they are present approximately 10–12 bps apart, due to the spacing of the two active centers at the RD interface of the individual DNMT3 complex subunits. Indeed, in vitro methylation experiments demonstrated that there is a correlation of methylation between sites localized ~10 bps apart (Jia et al. 2007; Jurkowska et al. 2008). Interestingly, the enrichment of CpG sites in such distance is observed in the differentially methylated regions (DMRs) of 12 maternally imprinted mouse genes, which are the biological substrates of the DNMT3A/DNMT3L complex, suggesting that the favorable CpG spacing could make these sequences good substrates for the MTase complex (Jia et al. 2007). As mentioned above, co-methylation of CpG sites at a distance of 12 bps by DNMT3A and DNMT3A/DNMT3L complexes was experimentally shown. Kinetic experiments supported with atomic force microscopy provided evidence that dimers of DNMT3A homotetramers or DNMT3A/DNMT3L heterotetramers can also form and interact with the CpG sites in distances of 9 or 5–6 base pairs. These complexes induce stronger DNA bending and represent either a tetramer swap or a side-by-side binding structure (Emperle et al. 2021). These additional complex conformations may explain how DNMT3A can methylate natural DNA, which does not present CpG sites in regular 12 bp spacings.

Multimerization of DNMT3A or DNMT3A/DNMT3L tetramers on DNA leads to a

cooperative DNA binding, as confirmed by different methods, including cooperative binding detected in gel retardation assays, sigmoidal binding curves of DNA substrates observed in solution DNA binding experiments, and direct imaging of DNMT3–DNA filaments by atomic force microscopy (Jia et al. 2007; Jurkowska et al. 2008; Rajavelu et al. 2012; Emperle et al. 2014). The interface of adjacent DNMT3A complexes bound to DNA has been mapped to a loop within the TRD of DNMT3A and mutation of residues within this interface disrupted multimerization (Rajavelu et al. 2012). Interestingly, it also led to the loss of heterochromatic enrichment of DNMT3A, suggesting that cooperative DNA binding and multimerization of DNMT3A complexes on DNA contribute to the heterochromatic localization of the enzyme in cells. Biochemical studies have further shown that the cooperative binding of DNMT3A to long DNA substrates increases the rate of DNA methylation (Emperle et al. 2014), indicating that it is important for DNA methylation by DNMT3A. However, the exact role of cooperative DNA binding of DNMT3A in cells needs further investigation because the sizes of DNMT3A–DNA filaments in living cells are currently unknown; one may speculate that binding of up to 5 complexes would be possible in the linker DNA regions between neighboring nucleosomes. This is in agreement with biochemical data showing preferential methylation of linker DNA by DNMT3 enzymes in vitro (Gowher et al. 2005b; Takeshima et al. 2008; Felle et al. 2011; Bröhm et al. 2022). In vivo studies confirmed this observation, showing that DNMT3B expressed in yeast preferentially methylates linker DNA (Morselli et al. 2015) and a similar pattern was also observed after reintroduction of the DNMT3 enzymes into KO cell lines (Baubec et al. 2015). Longer filaments may form if DNMT3 binding is coupled to nucleosome remodeling. Consistently, DNMT3s form complexes with various chromatin remodelers, including SMARCA4 (Datta et al. 2005), CHD4 (Cai et al. 2014), hSNF2 (Geiman et al. 2004), and HELLS (Zhu et al. 2006; Myant and Stancheva 2008) and the interaction with HELLS is essential for DNA methylation (Muegge 2005). In line with this model, the remodeling activity has been shown to promote the methylation of nucleosomal DNA (Felle et al. 2011). One important functional aspect of the cooperative DNA binding of DNMT3A may be that it increases the DNA binding affinity and reduces the rate of dissociation, which may help to anchor the MTase on the DNA, in agreement with its strong binding to methylated chromatin (Jeong et al. 2009; Sharma et al. 2011).

4.4.7 Direct Chromatin Interaction of DNMT3 Enzymes

4.4.7.1 Binding of the DNMT3 ADD Domain to H3 Tails

The ADD domains of DNMT3A, DNMT3B, and DNMT3L proteins interact specifically with histone H3 tails unmethylated at lysine 4 (Fig. 4.4), and the binding is disrupted by H3K4me2, H3K4me3, H3K4ac or the acetylation of the N-terminus of H3 (Ooi et al. 2007; Otani et al. 2009; Zhang et al. 2010; Noh et al. 2015). Interestingly, H3K4me1, which is observed at enhancers, does not hinder the binding of ADD much, but the phosphorylation of T6 does (Zhang et al. 2010; Noh et al. 2015). The structures of the ADD domains from DNMT3A and DNMT3L in complex with histone H3 tail peptides were solved (Ooi et al. 2007; Otani et al. 2009). Notably, binding to H3 tails stimulates the methylation of chromatin-bound DNA by DNMT3A in vitro (Zhang et al. 2010) and directly activates DNMT3A by an allosteric mechanism (Li et al. 2011). This regulatory mechanism has been confirmed in a structural analysis by Xu and colleagues, which showed that the ADD domain could bind to the catalytic domain of DNMT3A at two sites, an allosteric site and an autoinhibitory site. H3 peptide binding stabilizes the active conformation, leading to allosteric activation of DNMT3A (Guo et al. 2015) (Fig. 4.4b). These results indicate that the ADD domain of DNMT3A can guide DNA methylation in response to specific histone modifications and provided the first evidence that DNA methyltransferases could be targeted to chromatin

carrying specific marks. Indeed, a strong correlation of DNA methylation with the absence of H3K4me3 was observed in several genome-wide studies (Hodges et al. 2009; Meissner et al. 2008; Weber et al. 2007; Zhang et al. 2009), suggesting that this mechanism plays an important role in the generation of the genomic DNA methylation pattern. This hypothesis was experimentally verified, when it was shown that 1) a DNMT3A enzyme with an engineered ADD domain able to tolerate K4 methylation or T6 phosphorylation generated abnormal DNA methylation patterns in cells (Noh et al. 2015), and 2) DNMT3B artificially introduced in yeast did not methylate genomic regions with high H3K4me3 content (Morselli et al. 2015). The stimulation of DNMT3A activity by histone H3 PTMs interaction is reminiscent of the DNMT1 interaction with double ubiquitinated H3 tails, H3K9me3 or H4K20me3, further illustrating the common principles of regulatory mechanisms of DNMT1 and DNMT3 enzymes.

4.4.7.2 Binding of DNMT3 PWWP Domain to H3 Methylated at K36

The PWWP domain is essential for the targeting of DNMT3A and DNMT3B to pericentromeric chromatin (Chen et al. 2004; Ge et al. 2004) and gene bodies via specific recognition of histone H3 tails di- or tri-methylated at lysine 36 (H3K36me2/me3) (Dhayalan et al. 2010). In addition, the interaction of DNMT3A with H3K36me3 increases the activity of DNMT3A on chromatin, which carries H3K36me2/3 (Dhayalan et al. 2010). These findings explain the genome-wide correlation of DNA methylation and H3K36me3 methylation in gene bodies. H3K36me3 accumulates in euchromatin in the body of active genes and its distribution is anti-correlated with H3K4me3 (Barski et al. 2007; Edmunds et al. 2008; Guenther et al. 2007; Larschan et al. 2007; Vakoc et al. 2006). DNA methylation of gene bodies mirrors that pattern, with gene bodies of active genes showing high and those of inactive genes low methylation (Ball et al. 2009; Hellman and Chess 2007). Additionally, a correlation between H3K36me3 and DNA methylation was observed at exon–intron boundaries, with exons showing increased levels of both H3K36me3 (Kolasinska-Zwierz et al. 2009) and DNA methylation (Hodges et al. 2009). Moreover, a subset of heterochromatin containing repetitive sequences with copy number variations is strongly enriched in H3K36me3 (Ernst et al. 2011), which may explain the role of the DNMT3A PWWP domain in the heterochromatic localization of the enzyme and the strong correlation of DNA methylation, absence of H3K4me3 and presence of H3K36me3 observed in genome-wide DNA methylation studies (Meissner et al. 2008; Hodges et al. 2009).

The central role of K36 methylation in targeting DNA methylation has been experimentally confirmed in yeast (Morselli et al. 2015) and in a study showing that the methylation of gene bodies by DNMT3B directly depends on H3K36me3 methylation and requires an intact DNMT3B PWWP domain (Baubec et al. 2015). In contrast, the PWWP domain of DNMT3A mediates binding and DNA methylation at H3K36me2-containing intergenic regions (Weinberg et al. 2019), in agreement with a preference of the DNMT3A PWWP domain for binding to H3K36me2 over H3K36me3 in vitro (Weinberg et al. 2019; Dukatz et al. 2019a).

Finally, mutations within the PWWP domain cause aberrant DNMT3A localization and genomic DNA methylation (Heyn et al. 2019; Dukatz et al. 2019a; Remacha et al. 2018; Weinberg et al. 2021), further emphasizing the critical role of H3K36me2/3 interaction for DNMT3A targeting and function. The PWWP domains of DNMT3A and DNMT3B were also shown to interact with DNA to a variable degree, with DNMT3B PWWP binding DNA more strongly (Qiu et al. 2002; Purdy et al. 2010). Moreover, a synergistic interaction of the DNMT3A PWWP domain with methylated H3K36 and DNA has been observed (Dukatz et al. 2019a). In this study, a basic surface region on the PWWP domain was identified which mediates DNA interaction and is essential for the cellular localization of DNMT3A. This finding is not unexpected, as the K36 side chain emerges from the nucleosome body next to the

exit site of the linker DNA. H3K36me2/3 and DNA binding by PWWP domains are mediated by two adjacent interfaces, one featuring an aromatic cage for peptide binding and the other one a basic region for DNA interaction, similarly as observed in complexes of the LEDGF PWWP domain bound to a nucleosome (Wang et al. 2020a).

4.4.7.3 H2AK119ub Binding of DNMT3A1

Interaction of the DNMT3A PWWP domain with H3K36me2/3 is believed to limit DNA methylation in Polycomb-marked regions. Mutations in the PWWP domain that disrupt K36me2/3 or DNA binding were identified in microcephalic dwarfism (Heyn et al. 2019) and paraganglioma, a rare neuroendocrine neoplasm (Remacha et al. 2018). In these diseases, hypermethylation of the DNA was observed at Polycomb regulated regions (Heyn et al. 2019; Weinberg et al. 2021), similarly as observed in mouse cells with knock-in of a DNMT3A PWWP carrying a mutation that blocks H3K36me2/3 binding (Sendzikaite et al. 2019). The aberrant DNA methylation was accompanied by the mistargeting of DNMT3A1 to Polycomb-marked chromatin. Further experiments showed that this localization is dependent on PRC1-deposited H2AK119ub, but not directly on H3K27me3 (Weinberg et al. 2021). The UDR domain in the N-terminus of DNMT3A1 has been identified as responsible for ubiquitin interaction and DNMT1A1 targeting to H2AK119-marked regions (Weinberg et al. 2021). The UDR domain of DNMT3A1 is similar to a region in 53BP1 that mediates the interaction with H2AK15ub-modified nucleosomes. H2AK119ub1 is a repressive Polycomb histone modification that occurs together with H3K27me3, hence the DNMT3A–H2AK119ub1 interaction provides a direct molecular link between DNA methylation and Polycomb silencing, two very important repressed chromatin states, explaining the association of both signals in somatic cells and cancer (Jeltsch et al. 2019). Moreover, the data indicating that the disruption of the PWWP domain promotes DNA methylation in Polycomb regions can be explained by an increased role of UDR domain mediated DNMT3A1 targeting after the functional loss of the PWWP domain. The fact that the UDR domain is absent in the DNMT3A2 splicing isoform may explain an earlier observation that the DNMT3A1 isoform is preferentially localized to bivalent CpG island promoters (Manzo et al. 2017), which likely contain H2AK119ub. Hence, a misbalance between DNMT3A recruitment mediated by distinct reader domains may contribute to the abnormal methylation patterns observed in diseases. Overall, the multivalent interaction of the DNMT3 enzymes with chromatin through multimerization, UBD, ADD, and PWWP domains can explain the strong binding of these enzymes to nucleosomal heterochromatic DNA (Jeong et al. 2009; Sharma et al. 2011).

4.4.8 Interaction Partners of DNMT3s

Up to date, the interaction of DNMT3 enzymes with DNMT3L and MeCP2 has been studied in detail, revealing important roles in targeting, allosteric regulation and control of DNMT3 multimerization. Unfortunately, for most other DNMT3 interacting proteins, detailed information about their function is not yet available.

4.4.8.1 DNMT3A/DNMT3L Interaction

DNMT3L co-localizes with both DNMT3A and DNMT3B in mammalian cells (Hata et al. 2002). It directly interacts with its C-terminal domain with the catalytic domains of DNMT3A and DNMT3B and stimulates the activity of both enzymes in vivo (Chedin et al. 2002; Chen et al. 2005) and in vitro (Suetake et al. 2004; Gowher et al. 2005a; Kareta et al. 2006). DNMT3L is expressed during gametogenesis and embryonic stages (Bourc'his et al. 2001; Hata et al. 2002; Bourc'his and Bestor 2004), where it functions as a stimulatory factor of DNTM3A needed to establish DNA methylation patterns in the developing germline cells. The structure of the complex of the C-terminal domains of DNMT3A and

DNMT3L (Fig. 4.4a) provided a mechanistic explanation for the observed stimulatory effect of DNMT3L. It revealed that the interaction of DNMT3A with DNMT3L through the FF interface influences the structure of DNMT3A via the α-helices C, D, and E. Residues from these helices directly interact with the key catalytic and AdoMet binding residues, which may explain the stimulatory effect DNMT3L exerts on AdoMet binding and the catalytic activity of DNMT3A (Jia et al. 2007). Systematic studies indicated that DNMT3L increased the activity of DNMT3A and DNMT3B without changing their flanking sequence preferences (Gao et al. 2020b; Mao et al. 2020), different from an earlier study that was based on the analysis of a much smaller number of CpG sites with different flanking context (Wienholz et al. 2010). Recently, it was shown that the DNMT3 splice isoform, DNMT3B3, can replace DNMT3L in heterotetramer formation and stimulate the activity of DNMT3A and DNMT3B as well (Zeng et al. 2020). Hence, different complex formation between DNMT3 family members and their splice variants provides an additional layer regulating the activity of DNMT3 enzymes.

As described above, binding of DNMT3L to DNMT3A leads to the disruption of DNMT3A protein oligomers and changes the sub-nuclear localization of DNMT3A in cells (Fig. 4.6). In vivo, DNMT3L was shown to release DNMT3A from heterochromatin, by disrupting large DNMT3A oligomers and converting them into defined tetramers, which are homogeneously distributed in the cell nucleus (Jurkowska et al. 2011b). The redistribution of DNMT3A may be important for the methylation of imprinted differentially methylated regions (DMRs) and other targets in gene promoters, which generally are euchromatic. This finding goes in line with the discovery that DNMT3L favors DNA methylation in gene bodies (Neri et al. 2013). Hence, DNMT3L, which was originally discovered as a stimulator of DNMT3A (Gowher et al. 2005a), also changes the sub-nuclear localization of this enzyme (Jurkowska et al. 2011b). It needs to be investigated whether DNMT3B3 has similar effects as well. Additional evidence indicates that the combined regulation of the activity and localization of DNMT3A also applies to other regulatory cues (see below for MeCP2 interaction and CK2-mediated phosphorylation of DNMT3A) and might be a general mechanism of regulation for this family of enzymes (Fig. 4.6).

4.4.8.2 Interaction of DNMT3A with MeCP2

The chromatin protein MeCP2, which binds methylated DNA with its methyl-binding domain (MBD), was identified as a direct and strong interactor of DNMT3A and DNMT3B (Rajavelu et al. 2018). The interaction was mapped to the transcriptional repression domain (TRD) of MeCP2 and the ADD domain of the DNMT3 enzymes. Binding of MeCP2 resulted in a strong reduction of the DNA methylation activity of DNMT3A in vitro, and overexpression of MeCP2 in human cells led to a global reduction of DNA methylation. Biochemical experiments revealed that the binding of MeCP2 allosterically stabilizes the autoinhibitory conformation of DNMT3A. Interestingly, this interaction and its resulting inhibition were relieved by the binding of histone H3 to DNMT3A. In addition, MeCP2 contributed to the heterochromatic targeting of DNMT3A. These findings led to a model of an allosteric control of the target site specificity of DNMT3A by the combined effects of its interacting partners, like MeCP2 and histone H3 tails (Fig. 4.6). In this model, MeCP2 binding inactivates DNMT3A, thereby preventing aberrant methylation of bulk DNA. At the same time, it helps to deliver DNMT3A to heterochromatin. After binding to chromatin, which presents H3 tails modified in a PTM pattern matching the specificity of DNMT3A, the MeCP2 is released from DNMT3A, the allosteric inhibition is relieved and the activated DNMT3A enzyme can methylate its target sites. By this mechanism, MeCP2 generates a self-enhancing feedback loop

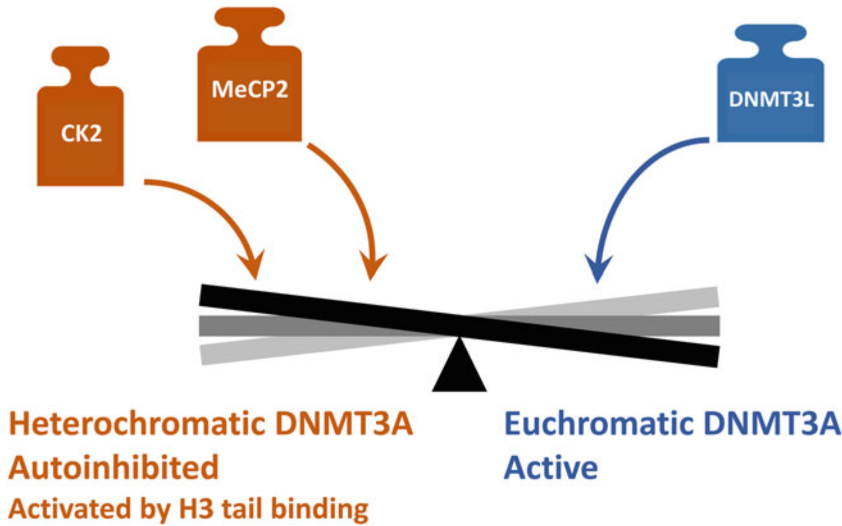


Fig. 4.6 Mechanisms regulating the activity and localization of DNMT3A. Different interactors and PTMs regulate the activity and localization in a concerted fashion. DNMT3L stimulates DNMT3A and promotes its euchromatic localization. Contrarily, MeCP2- and CK2-mediated phosphorylation downregulate the activity of DNMT3A and promote its heterochromatic localization, where the interaction with modified H3 tails allosterically stimulates the enzyme

that contributes to the deposition of DNA methylation at heterochromatic sites (Fig. 4.3b).

Interestingly, by this mechanism MeCP2 acts as a perfect antagonist of DNMT3L, which increases the activity of DNMT3A and leads to its release from heterochromatin. Moreover, the regulation of DNMT3A by CK2 (which is described below) resembles the MeCP2 effect, since it reduces the activity of DNMT3A and contributes to the heterochromatic sequestering of the methyltransferase. This illustrates an unexpected mechanistic convergence in the regulation and targeting of DNMT3A by interactors and posttranslational modifications (Fig. 4.6).

4.4.8.3 Other DNMT3A Interacting Proteins

Biochemical studies revealed that DNMT3A interacts with the tumor suppressor protein p53 in vitro and in cells (Wang et al. 2005; Sandoval and Reich 2019). This interaction reduces the activity of DNMT3A by interfering with tetramer formation in a DNMT3L competitive manner (Sandoval and Reich 2019). In turn, DNMT3A binding suppresses p53-mediated transcriptional activation in cells (Wang et al. 2005). The recent discovery of the SPOCD1 protein as an interactor of DNMT3A/DNMT3L and a key mediator in piRNA-directed DNA methylation has started to shed light on this fascinating process (Zoch et al. 2020).

4.4.9 Phosphorylation of DNMT3A

The regulation of the DNMT3 enzymes by phosphorylation has not been studied much, even though >70 phosphorylation sites have been identified in DNMT3A and DNMT3B in global proteomics studies (http://www.phosphosite.org). A unique example has been provided for casein kinase 2 (CK2) (Deplus et al. 2014). CK2 is a so-called survival protein kinase, which suppresses cancer cell death and is often upregulated in cancers. It was shown that CK2 phosphorylates DNMT3A at two sites, S386 and

S389, located next to the PWWP domain. CK2-mediated phosphorylation increased the heterochromatic targeting of DNMT3A and reduced its DNA methylation activity. This effect was reflected by changes in the cellular DNA methylation after CK2 knockout, which may explain global hypomethylation in cancer cells overexpressing CK2. These data further support the view that the combined regulation of enzymatic activity and localization is a general principle in the regulation of DNMT3A (as already described for DNMT3L and MeCP2 interaction above) (Fig. 4.6).

4.4.10 Binding of Regulatory DNA and RNA to DNMT3 Enzymes

Similar to DNMT1, additional nucleic acid binding sites have been identified in the N-terminal part of DNMT3 enzymes. As described above, the isolated PWWP domain of DNMT3B has a DNA binding activity (Qiu et al. 2002). In DNMT3A, an additional DNA binding site was detected and connected to the PWWP domain (Purdy et al. 2010; Dukatz et al. 2019a). Moreover, the very N-terminal part of DNMT3A had been shown to bind DNA (Suetake et al. 2011). Furthermore, it was observed that lncRNAs bind strongly to the catalytic domain of DNMT3A, causing inhibition of the enzyme (Holz-Schietinger and Reich 2012). The authors also detected binding of RNA to allosteric sites, which did not change the catalytic activity. In addition, it was shown that a ncRNA derived from the rDNA promoter binds to the promoter forming RNA/DNA triplex structures that are specifically recognized by DNMT3B, establishing a novel pathway of RNA-directed DNA methylation (Schmitz et al. 2010; Bierhoff et al. 2010). Conversely, the r-loop formation has been reported to protect promoters from DNMT3B engagement and DNA methylation in early development (Ginno et al. 2012). Future work will show whether the recruitment by regulatory DNA and RNA emerges as a new and general principle of DNMT targeting and regulation.

4.5 Outlook

After more than 40 years of intensive research in the DNA methylation field, we have learned many fascinating details regarding the biochemical, structural, and enzymatic properties of the mammalian DNA methyltransferases. With the availability of additional DNMT structures with bound substrate DNA, we begin to understand the molecular determinants of the enzymes' specificities. Strong effects of flanking sequences on the activity of DNMTs have been partially explained and structural studies of DNMT1 discovered a striking flanking sequence-dependent complex conformation. A better understanding of these effects will require the generation of additional DNMT structures with different DNA substrates, potentially combined with molecular dynamic simulations.

Despite progress regarding the mechanism of DNMTs, their regulation in cells has only begun to be uncovered. Importantly, it has been lately realized that the precise control of DNMT activity is critically involved in the generation and maintenance of the dynamic DNA methylation patterns in living cells. Crystallographic studies with DNMT1 and DNMT3A revealed that both enzymes unexpectedly undergo large domain rearrangements, which allosterically regulate their catalytic activity. This unforeseen discovery leads to the important conclusion that by influencing domain rearrangements any post-translational modification and interaction partner, be it a protein, allosteric DNA or non-coding RNA, could directly regulate the enzymatic activity, specificity, and localization of DNMTs via allosteric effects. This novel insight provides fascinating perspectives on the investigation of the effects of interactors and PTMs on these enzymes. More and more DNMTs interaction partners are discovered and investigated, emphasizing the key roles of other protein factors and chromatin modifications in targeting and regulation of DNMTs. To better understand the mechanisms and functions of these interactions, more cryo-EM structures of larger protein

complexes, including DNMTs and nucleosomes will be needed.

References

Adam S, Anteneh H, Hornisch M, Wagner V, Lu J, Radde NE et al (2020) DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. Nat Commun 11(1): 3723. https://doi.org/10.1038/s41467-020-17531-8

Allis CD, Jenuwein T (2016) The molecular hallmarks of epigenetic control. Nat Rev Genet 17(8):487–500. https://doi.org/10.1038/nrg.2016.59

Ambrosi C, Manzo M, Baubec T (2017) Dynamics and context-dependent roles of DNA methylation. J Mol Biol 429(10):1459–1475. https://doi.org/10.1016/j.jmb.2017.02.008

Anteneh H, Fang J, Song J (2020) Structural basis for impairment of DNA methylation by the DNMT3A R882H mutation. Nat Commun 11(1):2294. https://doi.org/10.1038/s41467-020-16213-9

Aoki A, Suetake I, Miyagawa J, Fujio T, Chijiwa T, Sasaki H et al (2001) Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases. Nucleic Acids Res 29(17):3506–3512

Arand J, Spieler D, Karius T, Branco MR, Meilinger D, Meissner A et al (2012) In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. PLoS Genet 8(6):e1002750. https://doi.org/10.1371/journal.pgen.1002750

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature 455(7214):818–821. https://doi.org/10.1038/nature07249

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C et al (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature 455(7214):822–825. https://doi.org/10.1038/nature07273

Bacolla A, Pradhan S, Roberts RJ, Wells RD (1999) Recombinant human DNA (cytosine-5) methyltransferase. II. Steady-state kinetics reveal allosteric activation by methylated dna. J Biol Chem 274(46):33011–33019

Bacolla A, Pradhan S, Larson JE, Roberts RJ, Wells RD (2001) Recombinant human DNA (cytosine-5) methyltransferase. III. Allosteric control, reaction order, and influence of plasmid topology and triplet repeat length on methylation of the fragile X CGG. CCG sequence. J Biol Chem 276(21):18605–18613. https://doi.org/10.1074/jbc.M100404200

Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH et al (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol 27(4):361–368. https://doi.org/10.1038/nbt.1533

Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Herault Y et al (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354(6314):909–912. https://doi.org/10.1126/science.aah5143

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z et al (2007) High-resolution profiling of histone methylations in the human genome. Cell 129(4): 823–837. https://doi.org/10.1016/j.cell.2007.05.009

Bashtrykov P, Jankevicius G, Smarandache A, Jurkowska RZ, Ragozin S, Jeltsch A (2012a) Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. Chem Biol 19(5):572–578. https://doi.org/10.1016/j.chembiol.2012.03.010

Bashtrykov P, Ragozin S, Jeltsch A (2012b) Mechanistic details of the DNA recognition by the Dnmt1 DNA methyltransferase. FEBS Lett 586(13):1821–1823. https://doi.org/10.1016/j.febslet.2012.05.026

Bashtrykov P, Jankevicius G, Jurkowska RZ, Ragozin S, Jeltsch A (2014a) The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. J Biol Chem 289(7):4106–4115. https://doi.org/10.1074/jbc.M113.528893

Bashtrykov P, Rajavelu A, Hackner B, Ragozin S, Carell T, Jeltsch A (2014b) Targeted mutagenesis results in an activation of DNA methyltransferase 1 and confirms an autoinhibitory role of its RFTS domain. Chembiochem 15(5):743–748. https://doi.org/10.1002/cbic.201300740

Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR et al (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520(7546): 243–247. https://doi.org/10.1038/nature14176

Bergman Y, Cedar H (2013) DNA methylation dynamics in health and disease. Nat Struct Mol Biol 20(3): 274–281. https://doi.org/10.1038/nsmb.2518

Berkyurek AC, Suetake I, Arita K, Takeshita K, Nakagawa A, Shirakawa M et al (2014) The DNA methyltransferase Dnmt1 directly interacts with the SET and RING finger-associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic center to hemi-methylated DNA. J Biol Chem 289(1):379–386. https://doi.org/10.1074/jbc.M113.523209

Bierhoff H, Schmitz K, Maass F, Ye J, Grummt I (2010) Noncoding transcripts in sense and antisense orientation regulate the epigenetic state of ribosomal RNA genes. Cold Spring Harb Symp Quant Biol 75:357–364. https://doi.org/10.1101/sqb.2010.75.060

Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317(5845):1760–1764. https://doi.org/10.1126/science.1147939

Bourc'his D, Bestor TH (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking

Dnmt3L. Nature 431(7004):96–99. https://doi.org/10.1038/nature02886
Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints. Science 294(5551):2536–2539. https://doi.org/10.1126/science.1065848
Brenner C, Deplus R, Didelot C, Loriot A, Vire E, De Smet C et al (2005) Myc represses transcription through recruitment of DNA methyltransferase corepressor. EMBO J 24(2):336–346. https://doi.org/10.1038/sj.emboj.7600509
Bröhm A, Schoch T, Dukatz M, Graf N, Dorscht F, Mantai E et al (2022) Methylation of recombinant mononucleosomes by DNMT3A demonstrates efficient linker DNA methylation and a role of H3K36me3. Commun Biol 5(1):192
Cai Y, Geutjes EJ, de Lint K, Roepman P, Bruurs L, Yu LR et al (2014) The NuRD complex cooperates with DNMTs to maintain silencing of key colorectal tumor suppressor genes. Oncogene 33(17):2157–2168. https://doi.org/10.1038/onc.2013.178
Chedin F, Lieber MR, Hsieh CL (2002) The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proc Natl Acad Sci U S A 99(26):16916–16921. https://doi.org/10.1073/pnas.262443999
Chen TP, Tsujimoto N, Li E (2004) The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Mol Cell Biol 24(20):9048–9058. https://doi.org/10.1128/Mcb.24.20.9048-9058.2004
Chen ZX, Mann JR, Hsieh CL, Riggs AD, Chedin F (2005) Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. J Cell Biochem 95(5):902–917. https://doi.org/10.1002/jcb.20447
Chen L, Chen K, Lavery LA, Baker SA, Shaw CA, Li W et al (2015) MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. Proc Natl Acad Sci U S A 112(17):5509–5514. https://doi.org/10.1073/pnas.1505909112
Cheng X, Blumenthal RM (2008) Mammalian DNA methyltransferases: a structural perspective. Structure 16(3):341–350. https://doi.org/10.1016/j.str.2008.01.004
Cheng X, Roberts RJ (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping. Nucleic Acids Res 29(18):3784–3795
Cheng J, Yang H, Fang J, Ma L, Gong R, Wang P et al (2015) Molecular mechanism for USP7-mediated DNMT1 stabilization by acetylation. Nat Commun 6:7023. https://doi.org/10.1038/ncomms8023
Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC et al (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science 325(5942):834–840. https://doi.org/10.1126/science.1175371
Christman JK, Sheikhnejad G, Marasco CJ, Sufrin JR (1995) 5-Methyl-2′-deoxycytidine in single-stranded DNA can act in cis to signal de novo DNA methylation. Proc Natl Acad Sci U S A 92(16):7347–7351
Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science 277(5334):1996–2000
Datta J, Majumder S, Bai S, Ghoshal K, Kutay H, Smith DS et al (2005) Physical and functional interaction of DNA methyltransferase 3A with Mbd3 and Brg1 in mouse lymphosarcoma cells. Cancer Res 65(23):10891–10900. https://doi.org/10.1158/0008-5472.CAN-05-1455
Deplus R, Blanchon L, Rajavelu A, Boukaba A, Defrance M, Luciani J et al (2014) Regulation of DNA methylation patterns by CK2-mediated phosphorylation of Dnmt3a. Cell Rep 8(3):743–753. https://doi.org/10.1016/j.celrep.2014.06.048
Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S et al (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem 285(34):26114–26120. https://doi.org/10.1074/jbc.M109.089433
Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA, Terragni J et al (2013) DNMT1-interacting RNAs block gene-specific DNA methylation. Nature 503(7476):371–376. https://doi.org/10.1038/nature12598
Du Z, Song J, Wang Y, Zhao Y, Guda K, Yang S et al (2010) DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination. Sci Signal 3(146):ra80. https://doi.org/10.1126/scisignal.2001462
Du Q, Wang Z, Schramm VL (2016) Human DNMT1 transition state structure. Proc Natl Acad Sci U S A 113:2916. https://doi.org/10.1073/pnas.1522491113
Dukatz M, Holzer K, Choudalakis M, Emperle M, Lungu C, Bashtrykov P et al (2019a) H3K36me2/3 binding and DNA binding of the DNA methyltransferase DNMT3A PWWP domain both contribute to its chromatin interaction. J Mol Biol 431(24):5063–5074. https://doi.org/10.1016/j.jmb.2019.09.006
Dukatz M, Requena CE, Emperle M, Hajkova P, Sarkies P, Jeltsch A (2019b) Mechanistic insights into cytosine-N3 methylation by DNA methyltransferase DNMT3A. J Mol Biol 431(17):3139–3145. https://doi.org/10.1016/j.jmb.2019.06.015
Dukatz M, Adam S, Biswal M, Song J, Bashtrykov P, Jeltsch A (2020) Complex DNA sequence readout mechanisms of the DNMT3B DNA methyltransferase. Nucleic Acids Res 48(20):11495–11509. https://doi.org/10.1093/nar/gkaa938
Easwaran HP, Schermelleh L, Leonhardt H, Cardoso MC (2004) Replication-independent chromatin loading of Dnmt1 during G2 and M phases. EMBO Rep 5(12):1181–1186. https://doi.org/10.1038/sj.embor.7400295

Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M et al (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 38(12): 1378–1385. https://doi.org/10.1038/ng1909

Edmunds JW, Mahadevan LC, Clayton AL (2008) Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. EMBO J 27(2):406–420. https://doi.org/10.1038/sj.emboj.7601967

Egger G, Jeong S, Escobar SG, Cortez CC, Li TW, Saito Y et al (2006) Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. Proc Natl Acad Sci U S A 103(38): 14080–14085. https://doi.org/10.1073/pnas.0604602103

Emperle M, Rajavelu A, Reinhardt R, Jurkowska RZ, Jeltsch A (2014) Cooperative DNA binding and protein/DNA fiber formation increases the activity of the Dnmt3a DNA methyltransferase. J Biol Chem 289(43):29602–29613. https://doi.org/10.1074/jbc.M114.572032

Emperle M, Rajavelu A, Kunert S, Arimondo PB, Reinhardt R, Jurkowska RZ et al (2018) The DNMT3A R882H mutant displays altered flanking sequence preferences. Nucleic Acids Res 46(6): 3130–3139. https://doi.org/10.1093/nar/gky168

Emperle M, Adam S, Kunert S, Dukatz M, Baude A, Plass C et al (2019) Mutations of R882 change flanking sequence preferences of the DNA methyltransferase DNMT3A and cellular methylation patterns. Nucleic Acids Res 47(21):11355–11367. https://doi.org/10.1093/nar/gkz911

Emperle M, Bangalore DM, Adam S, Kunert S, Heil HS, Heinze KG et al (2021) Structural and biochemical insight into the mechanism of dual CpG site binding and methylation by the DNMT3A DNA methyltransferase. Nucleic Acids Res 49(14): 8294–8308. https://doi.org/10.1093/nar/gkab600

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB et al (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473(7345):43–49. https://doi.org/10.1038/nature09906

Esteve PO, Chin HG, Benner J, Feehery GR, Samaranayake M, Horwitz GA et al (2009) Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells. Proc Natl Acad Sci U S A 106(13):5076–5081. https://doi.org/10.1073/pnas.0810362106

Esteve PO, Chang Y, Samaranayake M, Upadhyay AK, Horton JR, Feehery GR et al (2011) A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. Nat Struct Mol Biol 18(1):42–48. https://doi.org/10.1038/nsmb.1939

Esteve PO, Zhang G, Ponnaluri VK, Deepti K, Chin HG, Dai N et al (2016) Binding of 14-3-3 reader proteins to phosphorylated DNMT1 facilitates aberrant DNA methylation and gene expression. Nucleic Acids Res 44(4):1642–1656. https://doi.org/10.1093/nar/gkv1162

Fatemi M, Hermann A, Pradhan S, Jeltsch A (2001) The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. J Mol Biol 309(5):1189–1199. https://doi.org/10.1006/jmbi.2001.4709

Fatemi M, Hermann A, Gowher H, Jeltsch A (2002) Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA. Eur J Biochem 269(20): 4981–4984

Felle M, Hoffmeister H, Rothammer J, Fuchs A, Exler JH, Langst G (2011) Nucleosomes protect DNA from DNA methylation in vivo and in vitro. Nucleic Acids Res 39(16):6956–6969. https://doi.org/10.1093/nar/gkr263

Ferry L, Fournier A, Tsusaka T, Adelmant G, Shimazu T, Matano S et al (2017) Methylation of DNA ligase 1 by G9a/GLP recruits UHRF1 to replicating DNA and regulates DNA methylation. Mol Cell 67(4):550–65 e5. https://doi.org/10.1016/j.molcel.2017.07.012

Flynn J, Fang JY, Mikovits JA, Reich NO (2003) A potent cell-active allosteric inhibitor of murine DNA cytosine C5 methyltransferase. J Biol Chem 278(10): 8238–8243. https://doi.org/10.1074/jbc.M209839200

Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T (2000) DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet 24(1):88–91. https://doi.org/10.1038/71750

Fuks F, Burgers WA, Godin N, Kasai M, Kouzarides T (2001) Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. EMBO J 20(10):2536–2544. https://doi.org/10.1093/emboj/20.10.2536

Fuks F, Hurd PJ, Deplus R, Kouzarides T (2003a) The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res 31(9):2305–2312

Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T (2003b) The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem 278(6):4035–4040. https://doi.org/10.1074/jbc.M210256200

Gabel HW, Kinde B, Stroud H, Gilbert CS, Harmin DA, Kastan NR et al (2015) Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. Nature 522(7554):89–93. https://doi.org/10.1038/nature14319

Gao L, Anteneh H, Song J (2020a) Dissect the DNMT3A- and DNMT3B-mediated DNA co-methylation through a covalent complex approach. J Mol Biol 432(2): 569–575. https://doi.org/10.1016/j.jmb.2019.11.004

Gao L, Emperle M, Guo Y, Grimm SA, Ren W, Adam S et al (2020b) Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. Nat

Commun 11(1):3355. https://doi.org/10.1038/s41467-020-17109-4

Ge YZ, Pu MT, Gowher H, Wu HP, Ding JP, Jeltsch A et al (2004) Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. J Biol Chem 279(24):25447–25454. https://doi.org/10.1074/jbc.M312296200

Geiman TM, Sankpal UT, Robertson AK, Zhao Y, Robertson KD (2004) DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. Biochem Biophys Res Commun 318(2):544–555. https://doi.org/10.1016/j.bbrc.2004.04.058

Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell 45(6):814–825. https://doi.org/10.1016/j.molcel.2012.01.017

Glickman JF, Flynn J, Reich NO (1997a) Purification and characterization of recombinant baculovirus-expressed mouse DNA methyltransferase. Biochem Biophys Res Commun 230(2):280–284. https://doi.org/10.1006/bbrc.1996.5943

Glickman JF, Pavlovich JG, Reich NO (1997b) Peptide mapping of the murine DNA methyltransferase reveals a major phosphorylation site and the start of translation. J Biol Chem 272(28):17851–17857

Gowher H, Jeltsch A (2001) Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpA sites. J Mol Biol 309(5):1201–1208. https://doi.org/10.1006/jmbi.2001.4710

Gowher H, Jeltsch A (2002) Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. J Biol Chem 277(23):20409–20414. https://doi.org/10.1074/jbc.M202148200

Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A (2005a) Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. J Biol Chem 280(14):13341–13348. https://doi.org/10.1074/jbc.M413412200

Gowher H, Stockdale CJ, Goyal R, Ferreira H, Owen-Hughes T, Jeltsch A (2005b) De novo methylation of nucleosomal DNA by the mammalian Dnmt1 and Dnmt3A DNA methyltransferases. Biochemistry 44(29):9899–9904. https://doi.org/10.1021/bi047634t

Gowher H, Loutchanwoot P, Vorobjeva O, Handa V, Jurkowska RZ, Jurkowski TP et al (2006) Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase. J Mol Biol 357(3):928–941. https://doi.org/10.1016/j.jmb.2006.01.035

Goyal R, Reinhardt R, Jeltsch A (2006) Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. Nucleic Acids Res 34(4):1182–1188. https://doi.org/10.1093/nar/gkl002

Goyal R, Rathert P, Laser H, Gowher H, Jeltsch A (2007) Phosphorylation of serine-515 activates the mammalian maintenance methyltransferase Dnmt1. Epigenetics 2(3):155–160

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130(1):77–88. https://doi.org/10.1016/j.cell.2007.05.042

Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B et al (2014) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat Neurosci 17(2):215–222. https://doi.org/10.1038/nn.3607

Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X et al (2015) Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517(7536):640–644. https://doi.org/10.1038/nature13899

Haggerty C, Kretzmer H, Riemenschneider C, Kumar AS, Mattei AL, Bailly N et al (2021) Dnmt1 has de novo activity targeted to transposable elements. Nat Struct Mol Biol 28(7):594–603. https://doi.org/10.1038/s41594-021-00603-8

Hamidi T, Singh AK, Chen T (2015) Genetic alterations of DNA methylation machinery in human diseases. Epigenomics 7(2):247–265. https://doi.org/10.2217/epi.14.80

Handa V, Jeltsch A (2005) Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. J Mol Biol 348(5):1103–1112. https://doi.org/10.1016/j.jmb.2005.02.044

Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature 455(7214):826–829. https://doi.org/10.1038/nature07280

Hata K, Okano M, Lei H, Li E (2002) Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development 129(8):1983–1993

Hellman A, Chess A (2007) Gene body-specific methylation on the active X chromosome. Science 315(5815):1141–1143. https://doi.org/10.1126/science.1136352

Hermann A, Gowher H, Jeltsch A (2004a) Biochemistry and biology of mammalian DNA methyltransferases. Cell Mol Life Sci 61(19–20):2571–2587. https://doi.org/10.1007/s00018-004-4201-1

Hermann A, Goyal R, Jeltsch A (2004b) The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem 279(46):48350–48359. https://doi.org/10.1074/jbc.M403427200

Hervouet E, Lalier L, Debien E, Cheray M, Geairon A, Rogniaux H et al (2010) Disruption of Dnmt1/PCNA/UHRF1 interactions promotes tumorigenesis from

human and mice glial cells. PLoS One 5(6):e11333. https://doi.org/10.1371/journal.pone.0011333

Heyn P, Logan CV, Fluteau A, Challis RC, Auchynnikava T, Martin CA et al (2019) Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. Nat Genet 51(1):96–105. https://doi.org/10.1038/s41588-018-0274-x

Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M et al (2009) High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. Genome Res 19(9):1593–1605. https://doi.org/10.1101/gr.095190.109

Holoch D, Moazed D (2015) RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet 16(2): 71–84. https://doi.org/10.1038/nrg3863

Holz-Schietinger C, Reich NO (2010) The inherent processivity of the human de novo methyltransferase 3A (DNMT3A) is enhanced by DNMT3L. J Biol Chem 285(38):29091–29100. https://doi.org/10.1074/jbc.M110.142513

Holz-Schietinger C, Reich NO (2012) RNA modulation of the human DNA methyltransferase 3A. Nucleic Acids Res 40(17):8550–8557. https://doi.org/10.1093/nar/gks537

Hu L, Li Z, Wang P, Lin Y, Xu Y (2011) Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. Cell Res 21(9):1374–1378. https://doi.org/10.1038/cr.2011.124

Iida T, Suetake I, Tajima S, Morioka H, Ohta S, Obuse C et al (2002) PCNA clamp facilitates action of DNA cytosine methyltransferase 1 on hemimethylated DNA. Genes Cells 7(10):997–1007

Ishiyama S, Nishiyama A, Saeki Y, Moritsugu K, Morimoto D, Yamaguchi L et al (2017) Structure of the Dnmt1 reader module complexed with a unique two-mono-ubiquitin mark on histone H3 reveals the basis for DNA methylation maintenance. Mol Cell 68(2):350–60 e7. https://doi.org/10.1016/j.molcel.2017.09.037

Iwasaki YW, Siomi MC, Siomi H (2015) PIWI-interacting RNA: its biogenesis and functions. Annu Rev Biochem 84:405–433. https://doi.org/10.1146/annurev-biochem-060614-034258

Jain D, Meydan C, Lange J, Claeys Bouuaert C, Lailler N, Mason CE et al (2017) Rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. PLoS Genet 13(8):e1006964. https://doi.org/10.1371/journal.pgen.1006964

Jeltsch A (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chembiochem 3(4):275–293

Jeltsch A (2006) On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. Epigenetics 1(2):63–66

Jeltsch A (2008) Reading and writing DNA methylation. Nat Struct Mol Biol 15(10):1003–1004. https://doi.org/10.1038/nsmb1008-1003

Jeltsch A, Jurkowska RZ (2013) Multimerization of the dnmt3a DNA methyltransferase and its functional implications. Prog Mol Biol Transl Sci 117:445–464. https://doi.org/10.1016/B978-0-12-386931-9.00016-7

Jeltsch A, Jurkowska RZ (2014) New concepts in DNA methylation. Trends Biochem Sci 39(7):310–318. https://doi.org/10.1016/j.tibs.2014.05.002

Jeltsch A, Jurkowska RZ (2016) Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. Nucleic Acids Res 44(18): 8556–8575. https://doi.org/10.1093/nar/gkw723

Jeltsch A, Broche J, Bashtrykov P (2019) Molecular processes connecting DNA methylation patterns with DNA methyltransferases and histone modifications in mammalian genomes. Genes 10(5):388. https://doi.org/10.3390/genes10050388

Jeltsch A, Adam S, Dukatz M, Emperle M, Bashtrykov P (2021) Deep enzymology studies on DNA methyltransferases reveal novel connections between flanking sequences and enzyme activity. J Mol Biol 433(19):167186. https://doi.org/10.1016/j.jmb.2021.167186

Jeong S, Liang GN, Sharma S, Lin JC, Choi SH, Han H et al (2009) Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. Mol Cell Biol 29(19): 5366–5376. https://doi.org/10.1128/Mcb.00484-09

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449(7159):248–251. https://doi.org/10.1038/nature06146

Jurkowska RZ, Anspach N, Urbanke C, Jia D, Reinhardt R, Nellen W et al (2008) Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. Nucleic Acids Res 36(21):6656–6663. https://doi.org/10.1093/nar/gkn747

Jurkowska RZ, Jurkowski TP, Jeltsch A (2011a) Structure and function of mammalian DNA methyltransferases. Chembiochem 12(2):206–222. https://doi.org/10.1002/cbic.201000195

Jurkowska RZ, Rajavelu A, Anspach N, Urbanke C, Jankevicius G, Ragozin S et al (2011b) Oligomerization and binding of the Dnmt3a DNA methyltransferase to parallel DNA molecules: heterochromatic localization and role of Dnmt3L. J Biol Chem 286(27):24200–24207. https://doi.org/10.1074/jbc.M111.254987

Jurkowska RZ, Siddique AN, Jurkowski TP, Jeltsch A (2011c) Approaches to enzyme and substrate design of the murine Dnmt3a DNA methyltransferase. Chembiochem 12(10):1589–1594. https://doi.org/10.1002/cbic.201000673

Kareta MS, Botello ZM, Ennis JJ, Chou C, Chedin F (2006) Reconstitution and mechanism of the

stimulation of de novo methylation by human DNMT3L. J Biol Chem 281(36):25893–25902. https://doi.org/10.1074/jbc.M603140200
Kim SC, Sprung R, Chen Y, Xu Y, Ball H, Pei J et al (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. Mol Cell 23(4):607–618. https://doi.org/10.1016/j.molcel.2006.06.026
Kim SH, Park J, Choi MC, Kim HP, Park JH, Jung Y et al (2007) Zinc-fingers and homeoboxes 1 (ZHX1) binds DNA methyltransferase (DNMT) 3B to enhance DNMT3B-mediated transcriptional repression. Biochem Biophys Res Commun 355(2):318–323. https://doi.org/10.1016/j.bbrc.2007.01.187
Kimura H, Shiota K (2003) Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. J Biol Chem 278(7): 4806–4812. https://doi.org/10.1074/jbc.M209923200
Klimasauskas S, Kumar S, Roberts RJ, Cheng X (1994) HhaI methyltransferase flips its target base out of the DNA helix. Cell 76(2):357–369
Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet 41(3):376–381. https://doi.org/10.1038/ng.322
Kudithipudi S, Schuhmacher MK, Kebede AF, Jeltsch A (2017) The SUV39H1 protein lysine methyltransferase Methylates chromatin proteins involved in heterochromatin formation and VDJ recombination. ACS Chem Biol 12(4):958–968. https://doi.org/10.1021/acschembio.6b01076
Larschan E, Alekseyenko AA, Gortchakov AA, Peng S, Li B, Yang P et al (2007) MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. Mol Cell 28(1): 121–133. https://doi.org/10.1016/j.molcel.2007.08.011
Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT et al (2010) Dynamic changes in the human methylome during differentiation. Genome Res 20(3):320–331. https://doi.org/10.1101/gr.101907.109
Lavoie G, Esteve PO, Laulan NB, Pradhan S, St-Pierre Y (2011) PKC isoforms interact with and phosphorylate DNMT1. BMC Biol 9:31. https://doi.org/10.1186/1741-7007-9-31
Lee JH, Park SJ, Nakai K (2017) Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s. Sci Rep 7(1):11295. https://doi.org/10.1038/s41598-017-11800-1
Leng F, Yu J, Zhang C, Alejo S, Hoang N, Sun H et al (2018) Methylated DNMT1 and E2F1 are targeted for proteolysis by L3MBTL3 and CRL4(DCAF5) ubiquitin ligase. Nat Commun 9(1):1641. https://doi.org/10.1038/s41467-018-04019-9
Leonhardt H, Page AW, Weier HU, Bestor TH (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell 71(5):865–873
Li H, Rauch T, Chen ZX, Szabo PE, Riggs AD, Pfeifer GP (2006) The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. J Biol Chem 281(28):19489–19500. https://doi.org/10.1074/jbc.M513249200
Li JY, Pu MT, Hirasawa R, Li BZ, Huang YN, Zeng R et al (2007) Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. Mol Cell Biol 27(24): 8748–8759. https://doi.org/10.1128/MCB.01380-07
Li BZ, Huang Z, Cui QY, Song XH, Du L, Jeltsch A et al (2011) Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. Cell Res 21(8):1172–1181. https://doi.org/10.1038/cr.2011.92
Li T, Wang L, Du Y, Xie S, Yang X, Lian F et al (2018) Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation. Nucleic Acids Res 46(6): 3218–3231. https://doi.org/10.1093/nar/gky104
Li J, Wang R, Jin J, Han M, Chen Z, Gao Y et al (2020) USP7 negatively controls global DNA methylation by attenuating ubiquitinated histone-dependent DNMT1 recruitment. Cell Discov 6(1):58. https://doi.org/10.1038/s41421-020-00188-4
Liang Z, Hu J, Yan W, Jiang H, Hu G, Luo C (2018) Deciphering the role of dimer interface in intrinsic dynamics and allosteric pathways underlying the functional transformation of DNMT3A. Biochim Biophys Acta Gen Subj 1862(7):1667–1679. https://doi.org/10.1016/j.bbagen.2018.04.015
Lin IG, Han L, Taghva A, O'Brien LE, Hsieh CL (2002) Murine de novo methyltransferase Dnmt3a demonstrates strand asymmetry and site preference in the methylation of DNA in vitro. Mol Cell Biol 22(3): 704–723
Lin CC, Chen YP, Yang WZ, Shen JCK, Yuan HS (2020) Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. Nucleic Acids Res 48(7):3949–3961. https://doi.org/10.1093/nar/gkaa111
Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462(7271):315–322. https://doi.org/10.1038/nature08514
Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND et al (2013) Global epigenomic reconfiguration during mammalian brain development. Science 341(6146):1237905. https://doi.org/10.1126/science.1237905
Liu Y, Oakeley EJ, Sun L, Jost JP (1998) Multiple domains are involved in the targeting of the mouse DNA methyltransferase to the DNA replication foci. Nucleic Acids Res 26(4):1038–1045
Liu X, Gao Q, Li P, Zhao Q, Zhang J, Li J et al (2013) UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and

methylated H3K9. Nat Commun 4:1563. https://doi.org/10.1038/ncomms2562

Lukashevich OV, Cherepanova NA, Jurkovska RZ, Jeltsch A, Gromova ES (2016) Conserved motif VIII of murine DNA methyltransferase Dnmt3a is essential for methylation activity. BMC Biochem 17(1):7. https://doi.org/10.1186/s12858-016-0064-y

Mack A, Emperle M, Schnee P, Adam S, Pleiss J, Bashtrykov P et al (2022) Preferential self-interaction of DNA methyltransferase DNMT3A subunits containing the R882H cancer mutation leads to dominant changes of flanking sequence preferences. J Mol Biol 434(7):167482. https://doi.org/10.1016/j.jmb.2022.167482

Mallona I, Ilie IM, Karemaker ID, Butz S, Manzo M, Caflisch A et al (2021) Flanking sequence preference modulates de novo DNA methylation in the mouse genome. Nucleic Acids Res 49(1):145–157. https://doi.org/10.1093/nar/gkaa1168

Manzo M, Wirz J, Ambrosi C, Villasenor R, Roschitzki B, Baubec T (2017) Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands. EMBO J 36(23):3421–3434. https://doi.org/10.15252/embj.201797038

Mao SQ, Cuesta SM, Tannahill D, Balasubramanian S (2020) Genome-wide DNA methylation signatures are determined by DNMT3A/B sequence preferences. Biochemistry 59(27):2541–2550. https://doi.org/10.1021/acs.biochem.0c00339

Margot JB, Ehrenhofer-Murray AE, Leonhardt H (2003) Interactions within the mammalian DNA methyltransferase family. BMC Mol Biol 4:7. https://doi.org/10.1186/1471-2199-4-7

Markouli M, Strepkos D, Piperi C (2021) Structure, activity and function of the SETDB1 protein methyltransferase. Life (Basel) 11(8):817. https://doi.org/10.3390/life11080817

Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet 15(6):394–408. https://doi.org/10.1038/nrg3683

Meilinger D, Fellinger K, Bultmann S, Rothbauer U, Bonapace IM, Klinkert WE et al (2009) Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. EMBO Rep 10(11):1259–1264. https://doi.org/10.1038/embor.2009.201

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A et al (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454(7205):766–770. https://doi.org/10.1038/nature07107

Merry CR, Forrest ME, Sabers JN, Beard L, Gao XH, Hatzoglou M et al (2015) DNMT1-associated long non-coding RNAs regulate global gene expression and DNA methylation in colon cancer. Hum Mol Genet 24(21):6240–6253. https://doi.org/10.1093/hmg/ddv343

Ming X, Zhang Z, Zou Z, Lv C, Dong Q, He Q et al (2020) Kinetics and mechanisms of mitotic inheritance of DNA methylation and their roles in aging-associated methylome deterioration. Cell Res 30(11):980–996. https://doi.org/10.1038/s41422-020-0359-9

Mishima Y, Brueckner L, Takahashi S, Kawakami T, Arita K, Oka S et al (2017) RFTS-dependent negative regulation of Dnmt1 by nucleosome structure and histone tails. FEBS J 284(20):3455–3469. https://doi.org/10.1111/febs.14205

Mishima Y, Brueckner L, Takahashi S, Kawakami T, Otani J, Shinohara A et al (2020) Enhanced processivity of Dnmt1 by monoubiquitinated histone H3. Genes Cells 25(1):22–32. https://doi.org/10.1111/gtc.12732

Morselli M, Pastor WA, Montanini B, Nee K, Ferrari R, Fu K et al (2015) In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. elife 4:e06205. https://doi.org/10.7554/eLife.06205

Muegge K (2005) Lsh, a guardian of heterochromatin at repeat elements. Biochem Cell Biol 83(4):548–554. https://doi.org/10.1139/o05-119

Myant K, Stancheva I (2008) LSH cooperates with DNA methyltransferases to repress transcription. Mol Cell Biol 28(1):215–226. https://doi.org/10.1128/MCB.01073-07

Nady N, Lemak A, Walker JR, Avvakumov GV, Kareta MS, Achour M et al (2011) Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. J Biol Chem 286(27):24300–24311. https://doi.org/10.1074/jbc.M111.234104

Neri F, Krepelova A, Incarnato D, Maldotti M, Parlato C, Galvagni F et al (2013) Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. Cell 155(1):121–134. https://doi.org/10.1016/j.cell.2013.08.056

Nguyen TV, Yao S, Wang Y, Rolfe A, Selvaraj A, Darman R et al (2019) The R882H DNMT3A hot spot mutation stabilizes the formation of large DNMT3A oligomers with low DNA methyltransferase activity. J Biol Chem 294(45):16966–16977. https://doi.org/10.1074/jbc.RA119.010126

Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K et al (2013) Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature 502(7470):249–253. https://doi.org/10.1038/nature12488

Nishiyama A, Mulholland CB, Bultmann S, Kori S, Endo A, Saeki Y et al (2020) Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation. Nat Commun 11(1):1222. https://doi.org/10.1038/s41467-020-15006-4

Noh KM, Wang H, Kim HR, Wenderski W, Fang F, Li CH et al (2015) Engineering of a histone-recognition domain in Dnmt3a alters the epigenetic landscape and phenotypic features of mouse ESCs. Mol Cell 59(1):89–103. https://doi.org/10.1016/j.molcel.2015.05.017

Norvil AB, AlAbdi L, Liu B, Tu YH, Forstoffer NE, Michie AR et al (2020) The acute myeloid leukemia variant DNMT3A Arg882His is a DNMT3B-like enzyme. Nucleic Acids Res 48(7):3761–3775. https://doi.org/10.1093/nar/gkaa139

Okano M, Xie SP, Li E (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet 19(3):219–220

O'Keefe RT, Henderson SC, Spector DL (1992) Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. J Cell Biol 116(5):1095–1110

Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z et al (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448(7154):714–717. https://doi.org/10.1038/nature05987

Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M (2009) Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep 10(11):1235–1241. https://doi.org/10.1038/embor.2009.218

Peng L, Yuan Z, Ling H, Fukasawa K, Robertson K, Olashaw N et al (2011) SIRT1 deacetylates the DNA methyltransferase 1 (DNMT1) protein and alters its activities. Mol Cell Biol 31(23):4720–4734. https://doi.org/10.1128/MCB.06147-11

Petryk N, Bultmann S, Bartke T, Defossez PA (2021) Staying true to yourself: mechanisms of DNA methylation maintenance in mammals. Nucleic Acids Res 49(6):3020–3032. https://doi.org/10.1093/nar/gkaa1154

Pradhan S, Esteve PO (2003) Allosteric activator domain of maintenance human DNA (cytosine-5) methyltransferase and its role in methylation spreading. Biochemistry 42(18):5321–5332. https://doi.org/10.1021/bi034160+

Pradhan M, Esteve PO, Chin HG, Samaranayke M, Kim GD, Pradhan S (2008) CXXC domain of human DNMT1 is essential for enzymatic activity. Biochemistry 47(38):10000–10009. https://doi.org/10.1021/bi8011725

Purdy MM, Holz-Schietinger C, Reich NO (2010) Identification of a second DNA binding site in human DNA methyltransferase 3A by substrate inhibition and domain deletion. Arch Biochem Biophys 498(1):13–22. https://doi.org/10.1016/j.abb.2010.03.007

Qin S, Min J (2014) Structure and function of the nucleosome-binding PWWP domain. Trends Biochem Sci 39(11):536–547. https://doi.org/10.1016/j.tibs.2014.09.001

Qin W, Leonhardt H, Spada F (2011) Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. J Cell Biochem 112(2):439–444. https://doi.org/10.1002/jcb.22998

Qin W, Wolf P, Liu N, Link S, Smets M, La Mastra F et al (2015) DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. Cell Res 25(8):911–929. https://doi.org/10.1038/cr.2015.72

Qiu C, Sawada K, Zhang X, Cheng X (2002) The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nat Struct Biol 9(3):217–224. https://doi.org/10.1038/nsb759

Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S et al (2011) In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell 44(3):361–372. https://doi.org/10.1016/j.molcel.2011.08.032

Rajakumara E, Wang Z, Ma H, Hu L, Chen H, Lin Y et al (2011) PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. Mol Cell 43(2):275–284. https://doi.org/10.1016/j.molcel.2011.07.006

Rajavelu A, Jurkowska RZ, Fritz J, Jeltsch A (2012) Function and disruption of DNA methyltransferase 3a cooperative DNA binding and nucleoprotein filament formation. Nucleic Acids Res 40(2):569–580. https://doi.org/10.1093/nar/gkr753

Rajavelu A, Lungu C, Emperle M, Dukatz M, Brohm A, Broche J et al (2018) Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2. Nucleic Acids Res 46(17):9044–9056. https://doi.org/10.1093/nar/gky715

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A 97(10):5237–5242

Raurell-Vila H, Bosch-Presegue L, Gonzalez J, Kane-Goldsmith N, Casal C, Brown JP et al (2017) An HP1 isoform-specific feedback mechanism regulates Suv39h1 activity under stress conditions. Epigenetics 12(2):166–175. https://doi.org/10.1080/15592294.2016.1278096

Reither S, Li F, Gowher H, Jeltsch A (2003) Catalytic mechanism of DNA-(cytosine-C5)-methyltransferases revisited: covalent intermediate formation is not essential for methyl group transfer by the murine Dnmt3a enzyme. J Mol Biol 329(4):675–684

Remacha L, Curras-Freixes M, Torres-Ruiz R, Schiavi F, Torres-Perez R, Calsina B et al (2018) Gain-of-function mutations in DNMT3A in patients with paraganglioma. Genetics in medicine : official journal of the American College of Medical Genetics 20:1644. https://doi.org/10.1038/s41436-018-0003-y

Ren W, Fan H, Grimm SA, Guo Y, Kim JJ, Yin J et al (2020) Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. Proc Natl Acad Sci U S A 117(31):18439–18447. https://doi.org/10.1073/pnas.2009316117

Ren W, Fan H, Grimm SA, Kim JJ, Li L, Guo Y et al (2021) DNMT1 reads heterochromatic H4K20me3 to reinforce LINE-1 DNA methylation. Nat Commun 12(1):2490. https://doi.org/10.1038/s41467-021-22665-4

Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. Annu Rev Biochem 81:145–166. https://doi.org/10.1146/annurev-biochem-051410-092902

Robertson KD, Uzvolgyi E, Liang G, Talmadge C, Sumegi J, Gonzales FA et al (1999) The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. Nucleic Acids Res 27(11): 2291–2298

Rondelet G, Dal Maso T, Willems L, Wouters J (2016) Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. J Struct Biol 194(3): 357–367. https://doi.org/10.1016/j.jsb.2016.03.013

Rosic S, Amouroux R, Requena CE, Gomes A, Emperle M, Beltran T et al (2018) Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. Nat Genet 50(3):452–459. https://doi.org/10.1038/s41588-018-0061-8

Rothbart SB, Krajewski K, Nady N, Tempel W, Xue S, Badeaux AI et al (2012) Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat Struct Mol Biol 19(11):1155–1160. https://doi.org/10.1038/nsmb.2391

Rothbart SB, Dickson BM, Ong MS, Krajewski K, Houliston S, Kireev DB et al (2013) Multivalent histone engagement by the linked tandem Tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation. Genes Dev 27(11): 1288–1298. https://doi.org/10.1101/gad.220467.113

Rountree MR, Bachman KE, Baylin SB (2000) DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nat Genet 25(3): 269–277. https://doi.org/10.1038/77023

Sandoval JE, Reich NO (2019) The R882H substitution in the human de novo DNA methyltransferase DNMT3A disrupts allosteric regulation by the tumor supressor p53. J Biol Chem 294(48):18207–18219. https://doi.org/10.1074/jbc.RA119.010827

Schmitz KM, Mayer C, Postepska A, Grummt I (2010) Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev 24(20):2264–2269. https://doi.org/10.1101/gad.590910

Schneider K, Fuchs C, Dobay A, Rottach A, Qin W, Wolf P et al (2013) Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. Nucleic Acids Res 41(9):4860–4876. https://doi.org/10.1093/nar/gkt191

Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G et al (2004) A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev 18(11):1251–1262. https://doi.org/10.1101/gad.300704

Schubeler D (2015) Function and information content of DNA methylation. Nature 517(7534):321–326. https://doi.org/10.1038/nature14192

Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D et al (2015) Human body epigenome maps reveal noncanonical DNA methylation variation. Nature 523(7559):212–216. https://doi.org/10.1038/nature14465

Sendzikaite G, Hanna CW, Stewart-Morgan KR, Ivanova E, Kelsey G (2019) A DNMT3A PWWP mutation leads to methylation of bivalent chromatin and growth retardation in mice. Nat Commun 10(1): 1884. https://doi.org/10.1038/s41467-019-09713-w

Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA et al (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 450(7171):908–912. https://doi.org/10.1038/nature06397

Sharma S, De Carvalho DD, Jeong S, Jones PA, Liang G (2011) Nucleosomes containing methylated DNA stabilize DNA methyltransferases 3A/3B and ensure faithful epigenetic inheritance. PLoS Genet 7(2): e1001286. https://doi.org/10.1371/journal.pgen.1001286

Somasundaram S, Forrest ME, Moinova H, Cohen A, Varadan V, LaFramboise T et al (2018) The DNMT1-associated lincRNA DACOR1 reprograms genome-wide DNA methylation in colon cancer. Clin Epigenetics 10(1):127. https://doi.org/10.1186/s13148-018-0555-3

Song J, Rechkoblit O, Bestor TH, Patel DJ (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science 331(6020):1036–1040. https://doi.org/10.1126/science.1195380

Song J, Teplova M, Ishibe-Murakami S, Patel DJ (2012) Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science 335(6069):709–712. https://doi.org/10.1126/science.1214453

Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S (2004) DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J Biol Chem 279(26):27816–27823. https://doi.org/10.1074/jbc.M400181200

Suetake I, Hayata D, Tajima S (2006) The amino-terminus of mouse DNA methyltransferase 1 forms an independent domain and binds to DNA with the sequence involving PCNA binding motif. J Biochem 140(6): 763–776. https://doi.org/10.1093/jb/mvj210

Suetake I, Mishima Y, Kimura H, Lee YH, Goto Y, Takeshima H et al (2011) Characterization of DNA-binding activity in the N-terminal domain of the DNA methyltransferase Dnmt3a. Biochem J 437(1):141–148. https://doi.org/10.1042/BJ20110241

Sugiyama Y, Hatano N, Sueyoshi N, Suetake I, Tajima S, Kinoshita E et al (2010) The DNA-binding activity of

mouse DNA methyltransferase 1 is regulated by phosphorylation with casein kinase 1delta/epsilon. Biochem J 427(3):489–497. https://doi.org/10.1042/BJ20091856

Suva ML, Riggi N, Bernstein BE (2013) Epigenetic reprogramming in cancer. Science 339(6127):1567–1570. https://doi.org/10.1126/science.1230184

Suzuki M, Yamada T, Kihara-Negishi F, Sakurai T, Hara E, Tenen DG et al (2006) Site-specific DNA methylation by a complex of PU.1 and Dnmt3a/b. Oncogene 25(17):2477–2488. https://doi.org/10.1038/sj.onc.1209272

Svedruzic ZM, Reich NO (2005) DNA cytosine C5 methyltransferase Dnmt1: catalysis-dependent release of allosteric inhibition. Biochemistry 44(27):9472–9485. https://doi.org/10.1021/bi050295z

Syeda F, Fagan RL, Wean M, Avvakumov GV, Walker JR, Xue S et al (2011) The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. J Biol Chem 286(17):15344–15351. https://doi.org/10.1074/jbc.M110.209882

Takeshima H, Suetake I, Tajima S (2008) Mouse Dnmt3a preferentially methylates linker DNA and is inhibited by histone H1. J Mol Biol 383(4):810–821. https://doi.org/10.1016/j.jmb.2008.03.001

Takeshita K, Suetake I, Yamashita E, Suga M, Narita H, Nakagawa A et al (2011) Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proc Natl Acad Sci U S A 108(22):9055–9059. https://doi.org/10.1073/pnas.1019629108

Vakoc CR, Sachdeva MM, Wang H, Blobel GA (2006) Profile of histone lysine methylation across transcribed mammalian chromatin. Mol Cell Biol 26(24):9185–9195. https://doi.org/10.1128/MCB.01529-06

Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F et al (2013) Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res 23(3):555–567. https://doi.org/10.1101/gr.147942.112

Vilkaitis G, Suetake I, Klimasauskas S, Tajima S (2005) Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. J Biol Chem 280(1):64–72. https://doi.org/10.1074/jbc.M411126200

Vire E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C et al (2006) The Polycomb group protein EZH2 directly controls DNA methylation. Nature 439(7078):871–874. https://doi.org/10.1038/nature04431

Wang YA, Kamarova Y, Shen KC, Jiang Z, Hahn MJ, Wang Y et al (2005) DNA methyltransferase-3a interacts with p53 and represses p53-mediated gene expression. Cancer Biol Ther 4(10):1138–1143. https://doi.org/10.4161/cbt.4.10.2073

Wang J, Hevi S, Kurash JK, Lei H, Gay F, Bajko J et al (2009) The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. Nat Genet 41(1):125–129. https://doi.org/10.1038/ng.268

Wang C, Shen J, Yang Z, Chen P, Zhao B, Hu W et al (2011) Structural basis for site-specific reading of unmodified R2 of histone H3 tail by UHRF1 PHD finger. Cell Res 21(9):1379–1382. https://doi.org/10.1038/cr.2011.123

Wang H, Farnung L, Dienemann C, Cramer P (2020a) Structure of H3K36-methylated nucleosome-PWWP complex reveals multivalent cross-gyre binding. Nat Struct Mol Biol 27(1):8–13. https://doi.org/10.1038/s41594-019-0345-4

Wang Q, Yu G, Ming X, Xia W, Xu X, Zhang Y et al (2020b) Imprecise DNMT1 activity coupled with neighbor-guided correction enables robust yet flexible epigenetic inheritance. Nat Genet 52(8):828–839. https://doi.org/10.1038/s41588-020-0661-y

Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M et al (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 39(4):457–466. https://doi.org/10.1038/ng1990

Weinberg DN, Papillon-Cavanagh S, Chen H, Yue Y, Chen X, Rajagopalan KN et al (2019) The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. Nature 573(7773):281–286. https://doi.org/10.1038/s41586-019-1534-3

Weinberg DN, Rosenbaum P, Chen X, Barrows D, Horth C, Marunde MR et al (2021) Two competing mechanisms of DNMT3A recruitment regulate the dynamics of de novo DNA methylation at PRC1-targeted CpG islands. Nat Genet 53(6):794–800. https://doi.org/10.1038/s41588-021-00856-5

Weisenberger DJ, Velicescu M, Cheng JC, Gonzales FA, Liang G, Jones PA (2004) Role of the DNA methyltransferase variant DNMT3b3 in DNA methylation. Mol Cancer Res 2(1):62–72

Wienholz BL, Kareta MS, Moarefi AH, Gordon CA, Ginno PA, Chedin F (2010) DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo. PLoS Genet 6(9):e1001106. https://doi.org/10.1371/journal.pgen.1001106

Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J et al (2012) Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell 148(4):816–831. https://doi.org/10.1016/j.cell.2011.12.035

Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M et al (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 402(6758):187–191. https://doi.org/10.1038/46052

Xu TH, Liu M, Zhou XE, Liang G, Zhao G, Xu HE et al (2020) Structure of nucleosome-bound DNA methyltransferases DNMT3A and DNMT3B. Nature 586(7827):151–155. https://doi.org/10.1038/s41586-020-2747-1

Ye F, Kong X, Zhang H, Liu Y, Shao Z, Jin J et al (2018) Biochemical studies and molecular dynamic

simulations reveal the molecular basis of conformational changes in DNA Methyltransferase-1. ACS Chem Biol 13(3):772–781. https://doi.org/10.1021/acschembio.7b00890

Zeng Y, Ren R, Kaur G, Hardikar S, Ying Z, Babcock L et al (2020) The inactive Dnmt3b3 isoform preferentially enhances Dnmt3b-mediated DNA methylation. Genes Dev 34:1546. https://doi.org/10.1101/gad.341925.120

Zhang Y, Rohde C, Tierling S, Jurkowski TP, Bock C, Santacruz D et al (2009) DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. PLoS Genet 5(3):e1000438. https://doi.org/10.1371/journal.pgen.1000438

Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I et al (2010) Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res 38(13):4246–4253. https://doi.org/10.1093/nar/gkq147

Zhang G, Esteve PO, Chin HG, Terragni J, Dai N, Correa IR Jr et al (2015a) Small RNA-mediated DNA (cytosine-5) methyltransferase 1 inhibition leads to aberrant DNA methylation. Nucleic Acids Res 43(12):6112–6124. https://doi.org/10.1093/nar/gkv518

Zhang ZM, Liu S, Lin K, Luo Y, Perry JJ, Wang Y et al (2015b) Crystal structure of human DNA methyltransferase 1. J Mol Biol 427(15):2520–2531. https://doi.org/10.1016/j.jmb.2015.06.001

Zhang ZM, Lu R, Wang P, Yu Y, Chen D, Gao L et al (2018) Structural basis for DNMT3A-mediated de novo DNA methylation. Nature 554(7692):387–391. https://doi.org/10.1038/nature25477

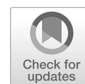
Zhao S, Allis CD, Wang GG (2021) The language of chromatin modification in human cancers. Nat Rev Cancer 21(7):413–430. https://doi.org/10.1038/s41568-021-00357-x

Zhou Q, Agoston AT, Atadja P, Nelson WG, Davidson NE (2008) Inhibition of histone deacetylases promotes ubiquitin-dependent proteasomal degradation of DNA methyltransferase 1 in human breast cancer cells. Mol Cancer Res 6(5):873–883. https://doi.org/10.1158/1541-7786.MCR-07-0330

Zhu H, Geiman TM, Xi S, Jiang Q, Schmidtmann A, Chen T et al (2006) Lsh is involved in de novo methylation of DNA. EMBO J 25(2):335–345. https://doi.org/10.1038/sj.emboj.7600925

Ziller MJ, Muller F, Liao J, Zhang Y, Gu H, Bock C et al (2011) Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. PLoS Genet 7(12):e1002389. https://doi.org/10.1371/journal.pgen.1002389

Zoch A, Auchynnikava T, Berrens RV, Kabayama Y, Schopp T, Heep M et al (2020) SPOCD1 is an essential executor of piRNA-directed de novo DNA methylation. Nature 584(7822):635–639. https://doi.org/10.1038/s41586-020-2557-5



5 Genetic Studies on Mammalian DNA Methyltransferases

Jiameng Dan and Taiping Chen

Abstract

Cytosine methylation at the C5-position—generating 5-methylcytosine (5mC)—is a DNA modification found in many eukaryotic organisms, including fungi, plants, invertebrates, and vertebrates, albeit its levels vary greatly in different organisms. In mammals, cytosine methylation occurs predominantly in the context of CpG dinucleotides, with the majority (60–80%) of CpG sites in their genomes being methylated. DNA methylation plays crucial roles in the regulation of chromatin structure and gene expression and is essential for mammalian development. Aberrant changes in DNA methylation and genetic alterations in enzymes and regulators involved in DNA methylation are associated with various human diseases, including cancer and developmental disorders. In mammals, DNA methylation is mediated by two families of DNA methyltransferases (Dnmts), namely Dnmt1 and Dnmt3 proteins. Over the last three decades, genetic manipulations of these enzymes, as well as their regulators, in mice have greatly contributed to our understanding of the biological functions of DNA methylation in mammals. In this chapter, we discuss genetic studies on mammalian Dnmts, focusing on their roles in embryogenesis, cellular differentiation, genomic imprinting, and human diseases.

Keywords

DNA methylation · Dnmt1 · Dnmt3a · Dnmt3b · Dnmt3c · Dnmt3L · Uhrf1 · Genomic imprinting

J. Dan (✉)
State Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Kunming, Yunnan, China

Yunnan Key Laboratory of Primate Biomedical Research, Kunming, Yunnan, China
e-mail: danjm@lpbr.cn

T. Chen (✉)
Department of Epigenetics and Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Programs in Genetics and Epigenetics, The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX, USA
e-mail: tchen2@mdanderson.org

5.1 Distinct Roles of Dnmt1 and Dnmt3 Families in DNA Methylation

In 1975, long before the identification of any mammalian DNA methyltransferase, Holliday and Pugh and Riggs independently proposed a theory that DNA methylation could serve as a heritable epigenetic mark for cellular memory. Recognizing that the CpG dinucleotide is

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_5

self-complementary, they postulated that methylated and unmethylated CpG sites could be copied when cells divide so that DNA methylation patterns would be replicated semi-conservatively like the base sequence of DNA itself (Holliday and Pugh 1975; Riggs 1975). A prediction of the theory was the existence of two DNA methyltransferase activities: de novo methyltransferase(s) would methylate unmodified DNA and establish DNA methylation patterns, and maintenance methyltransferase(s) would recognize hemi-methylated sites and "copy" the methylation patterns from the parental strand onto the daughter strand at each round of DNA replication.

5.1.1 Dnmt1: The Maintenance DNA Methyltransferase

The first mammalian DNA methyltransferase gene, *Dnmt1*, was cloned from murine cells (Bestor et al. 1988). The *Dnmt1* locus has several transcription start sites and produces two major protein products (Mertineit et al. 1998; Rouleau et al. 1992). Transcription initiation within a somatic cell-specific exon (exon 1 s) results in the Dnmt1s isoform (generally referred to as Dnmt1), which consists of 1620 amino acids (mouse). Transcription initiation within an oocyte-specific exon (exon 1o) produces a transcript that utilizes a downstream AUG as the translation initiation codon. As a result, the oocyte-specific isoform, Dnmt1o, lacks the N-terminal 118 amino acids of Dnmt1 (Fig. 5.1). Genetic evidence suggests no functional differences between the two isoforms, although Dnmt1o appears to be more stable (Ding and Chaillet 2002). Human DNMT1, consisting of 1616 amino acids, is 80% identical to mouse Dnmt1 at the amino acid level.

Dnmt1 contains a C-terminal catalytic domain with special sequence motifs that are homologous to bacterial DNA methyltransferases, and an N-terminal regulatory region that is not present in bacterial enzymes (Bestor et al. 1988). The N-terminal regulatory region contains several functional domains, including a proliferating cell nuclear antigen (PCNA) binding domain (PBD) responsible for the interaction with the DNA replication machinery, a nuclear localization signal (NLS), a replication foci-targeting sequence (RFTS) that mediates the association with late replicating heterochromatin, a zinc-finger CXXC domain that recognizes unmethylated CpG-containing DNA, and a pair of bromo-adjacent homology (BAH) domains (Fig. 5.1). Structural data revealed that the RFTS domain binds to the catalytic domain and blocks the catalytic center, suggesting an autoinhibitory role in the regulation of enzymatic activity (Takeshita et al. 2011).

In vitro biochemical assays revealed that, although Dnmt1 is capable of methylating both unmethylated and hemi-methylated CpG dinucleotides, its activity toward hemi-methylated substrates is far more efficient (Pradhan et al. 1999). Dnmt1 is ubiquitously expressed through development, with high levels in proliferating cells. Dnmt1 associates with the DNA replication machinery at S phase and with heterochromatin at late S and G2 phases (Chuang et al. 1997; Easwaran et al. 2004; Leonhardt et al. 1992; Schneider et al. 2013), suggesting that Dnmt1-mediated methylation is coupled to DNA replication. These findings supported the notion that Dnmt1 mainly functions as a maintenance enzyme (Fig. 5.2). However, because Dnmt1, the only known DNA methyltransferase at the time, also had de novo methylation activity in vitro, it was initially debated whether de novo methylation and maintenance methylation are carried out by Dnmt1 alone or by two or more distinct enzymes.

Genetic studies in mouse models and murine cells helped settling the debate. Several *Dnmt1* mutant alleles were generated by gene targeting. The *Dnmt1*^{n} allele (n stands for N-terminal disruption) was reported in 1992 (Li et al. 1992). This allele, in which a genomic region coding for 60 amino acids near the N-terminal end was replaced by a neomycin resistance cassette, is a partial loss-of-function (hypomorphic) mutation. *Dnmt1*$^{n/n}$ embryos have a ~ 70% reduction in global DNA methylation and show mid-gestation lethality (Li et al. 1992).

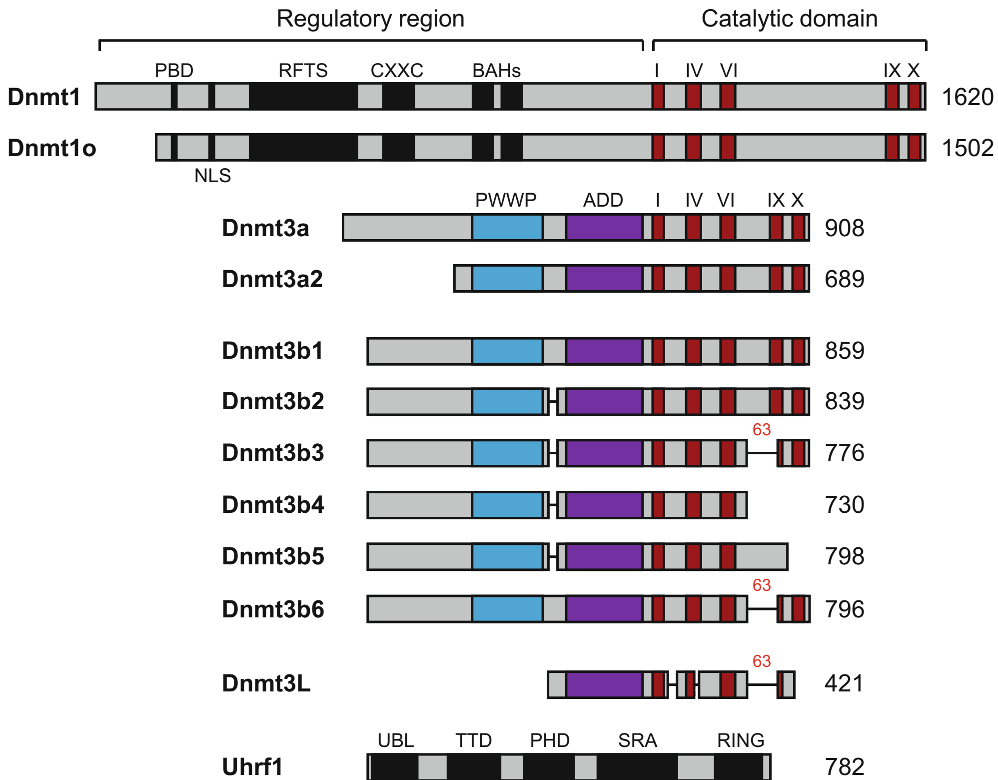


Fig. 5.1 DNMTs and major regulators. Schematic diagrams of major Dnmt1, Dnmt3a, and Dnmt3b isoforms, as well as Dnmt3L and Uhrf1, in mouse. The C-terminal catalytic domains of the Dnmt1 and Dnmt3 families are conserved (the highly conserved signature motifs I, IV, VI, IX, and X are shown in red), but their N-terminal regulatory regions are distinct. Note that Dnmt3L, Dnmt3b3, and Dnmt3b6 have the same 63-residue deletion in their C-terminal region. *PBD* PCNA-binding domain, *NLS* nuclear localization signal, *RFTS* replication foci-targeting sequence, *CXXC* a cysteine-rich domain implicated in binding CpG-containing DNA sequences, *BAHs* bromo-adjacent homology domains, *PWWP* proline-tryptophan-tryptophan-proline domain, *ADD* ATRX-Dnmt3-Dnmt3L domain, *UBL* ubiquitin-like domain, *TTD* tandem Tudor domain, *PHD* plant homeodomain, *SRA* SET and RING-associated domain, *RING* Really Interesting New Gene domain

Subsequently, the *Dnmt1*^{s} allele (s stands for *Sal*I site) was reported, which had a neomycin resistance cassette inserted into a *Sal*I site in exon 17, disrupting the RFTS (Li et al. 1993). The *Dnmt1*^{s} allele is functionally more severe than the *Dnmt1*^{n} allele, as *Dnmt1*$^{s/s}$ embryos show lower levels of DNA methylation and earlier lethality (Lei et al. 1996). However, it was unclear whether the *Dnmt1*^{s} allele was a null mutation, because the C-terminal catalytic domain was intact. To completely inactivate *Dnmt1*, Lei et al. generated the *Dnmt1*^{c} allele (c stands for C-terminal disruption) by disrupting the catalytic domain, including the highly conserved PC and ENV motifs that are essential for enzymatic activity (Lei et al. 1996). The development of *Dnmt1*$^{c/c}$ embryos is arrested prior to the 8-somite stage, significantly earlier than the developmental phenotype of *Dnmt1*$^{n/n}$ embryos, while the viability and proliferation of *Dnmt1* null mouse embryonic stem cells (mESCs) are not affected. DNA methylation analyses revealed that *Dnmt1* null embryos and mESCs contain low but stable levels of 5-methylcytosine (5mC) and methyltransferase activity. Moreover, the de novo methylation activity is not impaired by Dnmt1 loss, as integrated provirus DNA in MoMuLV-infected *Dnmt1* null mESCs becomes methylated at a

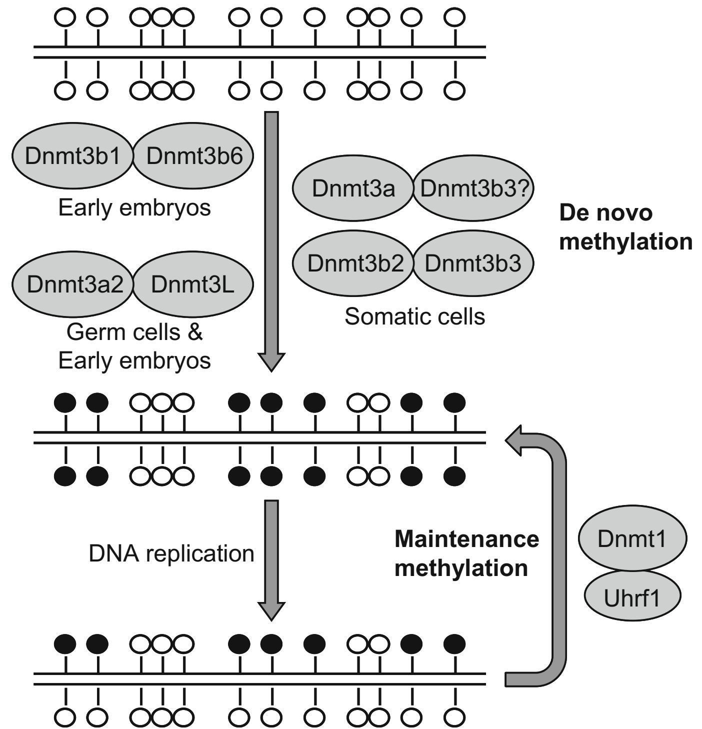


Fig. 5.2 De novo and maintenance methylation. The de novo methyltransferases Dnmt3a and Dnmt3b, in complex with their accessory factors Dnmt3L, Dnmt3b3, or Dnmt3b6, methylate unmodified DNA and establish methylation patterns. The major Dnmt3 isoforms in early embryos, germ cells, and somatic cells are shown. At each round of DNA replication, the maintenance methyltransferase Dnmt1, aided by its accessory factor Uhrf1, "copies" the methylation pattern from the parental strand onto the daughter strand. Open circles represent unmethylated CpG dinucleotides, and filled circles represent methylated CpG dinucleotides

similar rate as in wild-type (WT) mESCs (Lei et al. 1996). Taken together, these early studies provided compelling evidence for the existence of one or more DNA methyltransferases other than Dnmt1 that are important for de novo methylation.

5.1.2 Dnmt3 Family: Key Components of De Novo Methylation Machinery

Results from genetic studies of *Dnmt1* prompted the search for more DNA methyltransferase genes. In 1998, several groups reported the identification of a second putative DNA methyltransferase gene, named *Dnmt2*, which encodes a protein of 391 amino acids in human or 415 amino acids in mouse (Okano et al. 1998b; Van den Wyngaert et al. 1998; Yoder and Bestor 1998). Despite the presence of all the conserved motifs shared by known prokaryotic and eukaryotic DNA cytosine methyltransferases, Dnmt2 has no detectable DNA methyltransferase activity in standard in vitro assays. Furthermore, the inactivation of *Dnmt2* in mESCs by gene targeting has no effect on preexisting genomic methylation patterns or on the ability to methylate newly integrated retrovirus DNA de novo (Okano et al. 1998b). Indeed, a subsequent study demonstrated that Dnmt2 is a tRNA methyltransferase, specific for cytosine 38 in the anticodon loop of aspartic acid tRNA and has been renamed tRNA aspartic acid (D) methyltransferase 1 (Trdmt1) (Goll et al. 2006).

By searching an expressed sequence tag (EST) database using full-length bacterial type II cytosine-5 methyltransferase sequences as queries, Okano et al. identified two additional homologous genes, *Dnmt3a* and *Dnmt3b*, in both mouse and human. Their protein products contain the highly conserved DNA methyltransferase motifs in their C-terminal regions, but their N-terminal regulatory regions show little sequence similarity to that of Dnmt1 (Okano et al. 1998a). The N-terminal regions of Dnmt3a and Dnmt3b contain a variable region and two conserved domains, the

proline-tryptophan-tryptophan-proline (PWWP) domain and the ATRX-Dnmt3-Dnmt3L (ADD) domain (Fig. 5.1). Both domains are implicated in chromatin binding. The PWWP domain is required for heterochromatin localization and binds di- and tri-methylated histone H3 lysine 36 (H3K36me2/3) (Baubec et al. 2015; Chen et al. 2004; Dhayalan et al. 2010; Weinberg et al. 2019), and the ADD domain interacts with the N-terminal tail of histone H3 when H3K4 is unmethylated (Otani et al. 2009).

Dnmt3a produces two major isoforms, Dnmt3a and Dnmt3a2, driven by different promoters (Chen et al. 2002). The full-length Dnmt3a protein, consisting of 908 amino acids in mouse and 912 amino acids in human, is expressed ubiquitously at relatively low levels. The *Dnmt3a2* transcript is driven by a downstream intronic promoter and encodes a protein that lacks the N-terminal 219 (in mouse) or 223 (in human) amino acids of Dnmt3a. Dnmt3a2 is the predominant form in mESCs, early embryos, and developing germ cells, as well as human embryonal carcinoma cells, and is also detectable in spleen and thymus (Chen et al. 2002). While both Dnmt3a and Dnmt3a2 are catalytically active (Chen et al. 2002), recent studies suggest functional differences between them. Zeng et al. showed that in mESCs, Dnmt3a2 requires Dnmt3L (Dnmt3-like), an accessory factor (see below), for de novo methylation, whereas Dnmt3a has substantial activity by itself (Zeng et al. 2020). Moreover, Weinberg et al. showed that the very N terminus of Dnmt3a, which is absent in Dnmt3a2, contains a putative ubiquitin-dependent recruitment (UDR) domain that interacts with nucleosomes modified by monoubiquitylation of histone H2A lysine 119 (H2AK119ub) (Weinberg et al. 2021).

The *Dnmt3b* gene produces multiple alternatively spliced isoforms (>30 have been reported), many of which encode catalytically inactive protein products. The longest isoform, Dnmt3b1, consists of 859 amino acids in mouse and 853 amino acids in human, respectively. Active and inactive Dnmt3b isoforms appear to co-express in most, if not all, cell types. For example, Dnmt3b1, an active form, and Dnmt3b6, an inactive form, are the predominant forms in early embryos and mESCs, whereas Dnmt3b2, an active form, and Dnmt3b3, an inactive form, are expressed in differentiated somatic cells (Chen et al. 2002). There is evidence that catalytically inactive Dnmt3b protein products play regulatory roles in DNA methylation. For example, overexpression of human DNMT3B7, a truncated isoform frequently found in cancer cells, leads to higher levels of total genomic methylation and altered gene expression in both transgenic mice and human cancer cells (Ostler et al. 2012; Shah et al. 2010). Several catalytically inactive Dnmt3b isoforms were shown to facilitate DNA methylation in gene bodies in differentiated cells (Duymich et al. 2016). Interestingly, two Dnmt3b isoforms, Dnmt3b3 and Dnmt3b6, have the exact 63-residue deletion corresponding to that of Dnmt3L in the catalytic domain (Fig. 5.1). A recent study demonstrated that Dnmt3b3 plays a similar role as Dnmt3L but preferentially enhances Dnmt3b-mediated DNA methylation (Zeng et al. 2020).

Several lines of evidence suggest that Dnmt3a and Dnmt3b primarily function as de novo methyltransferases (Fig. 5.2). First, Dnmt3a and Dnmt3b are highly expressed in early embryos (and mESCs) and developing germ cells, where active de novo methylation takes place, but are downregulated in somatic tissues and when mESCs are induced to differentiate (Okano et al. 1998a). Second, recombinant Dnmt3a and Dnmt3b proteins methylate unmethylated and hemi-methylated DNA with equal efficiency (Okano et al. 1998a). Genetic studies provided definitive evidence that Dnmt3a and Dnmt3b were the long-sought de novo methyltransferases. The inactivation of both *Dnmt3a* and *Dnmt3b* by gene targeting blocks de novo methylation in mESCs and early embryos but has no effect on maintenance of imprinted methylation patterns (Okano et al. 1999). Dnmt3a deficiency also leads to failure to establish DNA methylation imprints in developing germ cells (Kaneda et al. 2004). Recently, the de novo methyltransferase activity of Dnmt3a was utilized for targeted genome de novo methylation by CRISPR-based epigenome editing (Nunez et al. 2021; Stepper

et al. 2017) Previously, different DNMTs including DNMT3A have been used for epigenome editing after fusion to zinc fingers (see Chap. 18 of this book). Based on the expression patterns of various Dnmt3 proteins and developmental phenotypes of knockout (KO) mice (Bourc'his and Bestor 2004; Bourc'his et al. 2001; Hata et al. 2002; Kaneda et al. 2004; Okano et al. 1999; Zeng et al. 2020), Dnmt3a2 and its accessory factor Dnmt3L are responsible for de novo methylation in developing germ cells, Dnmt3b1 and Dnmt3b6 (an inactive isoform similar to Dnmt3L), as well as Dnmt3a2 and Dnmt3L, likely play important roles in de novo methylation during early embryogenesis, and Dnmt3b2 and Dnmt3b3 (an inactive isoform similar to Dnmt3L), as well as Dnmt3a (either alone or complexed with Dnmt3b3), are likely involved in methylation during later stages of development and in somatic cells (Fig. 5.2).

It is worth noting that the de novo DNA methyltransferase activity of Dnmt3a and Dnmt3b is not only essential for the establishment of new methylation patterns, but also important for the faithful maintenance of these patterns. In culture, *Dnmt3a/3b* double KO (DKO) mESCs exhibit gradual loss of global DNA methylation and, after multiple passages, show severe hypomethylation (Chen et al. 2003), suggesting that Dnmt1 and Dnmt3 enzymes have distinct and non-redundant functions but act cooperatively in the maintenance of global DNA methylation. Recent evidence suggests competition between de novo methylation by Dnmt3 proteins and demethylation initiated by the ten-eleven translocation (TET) family of 5mC dioxygenases at many loci in mESCs (Charlton et al. 2020; Wang et al. 2020). There is also evidence that Dnmt1 has de novo methylation activity. For example, oocytes deficient for *Dnmt3a* and *Dppa3*, which encodes a maternal factor DPPA3 (also known as PGC7 and Stella), exhibit higher methylation levels than WT oocytes, suggesting a robust de novo methylation activity of Dnmt1 in mouse oocytes that is normally suppressed (Li et al. 2018). Dnmt1 has also been shown to be involved in de novo methylation of specific retrotransposons, such as intracisternal A particles (Haggerty et al. 2021).

A third member of the Dnmt3 family, *Dnmt3L*, was originally discovered by database analysis of the human genome sequence (Aapola et al. 2000). Its murine homolog was subsequently identified (Aapola et al. 2001; Hata et al. 2002). The human and mouse Dnmt3L proteins consist of 387 and 421 amino acids, respectively. Dnmt3L contains an ADD domain, but lacks a PWWP domain, in the N-terminal region. Its C-terminal region is highly related to the catalytic domains of Dnmt3a and Dnmt3b but lacks some motifs and regions essential for enzymatic activity (Aapola et al. 2000, 2001; Hata et al. 2002) (Fig. 5.1). Therefore, Dnmt3L has no methyltransferase activity. However, Dnmt3L has been shown to interact with Dnmt3a and Dnmt3b, stimulate their enzymatic activities, and target them to chromatin (Hata et al. 2002; Jia et al. 2007; Ooi et al. 2007; Suetake et al. 2004). The expression pattern of Dnmt3L during development is similar to that of Dnmt3a and Dnmt3b, including high expression in developing germ cells, early embryos, and mESCs (Hata et al. 2002). These findings indicate that Dnmt3L may regulate de novo DNA methylation (Fig. 5.2). However, genetic studies suggest that Dnmt3L functions mainly as an accessory factor for Dnmt3a (mainly Dnmt3a2) in the germline. *Dnmt3L* null mice are viable and grossly normal, but both male and female mice are infertile (Bourc'his et al. 2001; Hata et al. 2002). Male mice show the activation of retrotransposons in spermatogonia and spermatocytes, due to hypomethylation, and are azoospermic (Bourc'his and Bestor 2004). Female mice fail to establish maternal methylation imprints in oocytes and, as a result, embryos derived from these oocytes cannot survive beyond mid-gestation (Bourc'his et al. 2001; Hata et al. 2002). The phenotype is indistinguishable from that of mice with conditional *Dnmt3a* deletion in germ cells (Kaneda et al. 2004). Contrary to a previous report that Dnmt3L positively and negatively regulates DNA methylation depending on genomic regions (Neri et al. 2013), recent genome-wide methylation analysis revealed loss of methylation at many Dnmt3a target regions, but no gain of methylation at any loci in *Dnmt3L*-deficient mESCs, consistent with the widely accepted view that Dnmt3L is a

positive regulator of DNA methylation (Veland et al. 2019). Although Dnmt3L interacts with both Dnmt3a and Dnmt3b, its deficiency results in the degradation of Dnmt3a (especially Dnmt3a2), but not Dnmt3b, in mESCs, which explains, at least in part, its functional specificity in vivo (Veland et al. 2019).

In the rodent genome, a *Dnmt3b* duplicated gene, initially annotated as a pseudogene (Lees-Murdock et al. 2004), was subsequently identified as a new functional member of the Dnmt3 family. This gene, renamed *Dnmt3c,* encodes a protein of 709 amino acids which is similar to Dnmt3b but lacks the PWWP domain in the N-terminal regulatory region. It is specifically expressed in male germ cells. Genetic studies in mice demonstrated that *Dnmt3c* is not required for mouse development but is essential for normal spermatogenesis by methylating evolutionally young retrotransposons in the male germline (Barau et al. 2016; Jain et al. 2017).

5.1.3 Uhrf1: A Major Regulator of Maintenance DNA Methylation

A number of DNA methylation regulators have been identified, including the multi-domain protein Uhrf1 (ubiquitin-like with PHD and RING finger domains 1), also known as NP95 (mouse) and ICBP90 (human) (Fig. 5.1). Genetic studies demonstrated an essential role for Uhrf1 in maintaining DNA methylation (Fig. 5.2). Uhrf1 deficiency leads to embryonic lethality and global DNA hypomethylation (Bostick et al. 2007; Muto et al. 2002; Sharif et al. 2007), resembling the phenotype of Dnmt1 deficiency. Cellular and biochemical data indicate physical and functional interactions between Uhrf1 and Dnmt1. Uhrf1 co-localizes with Dnmt1 at DNA replication foci and heterochromatin, and Dnmt1 fails to enrich at these regions in the absence of Uhrf1 (Bostick et al. 2007; Liu et al. 2013; Sharif et al. 2007). These findings suggest that Uhrf1 is a key accessory factor of Dnmt1.

Uhrf1 harbors five known functional domains: a ubiquitin-like domain (UBL) at the N terminus, followed by a tandem Tudor domain (TTD), a plant homeodomain (PHD), a SET and RING-associated (SRA) domain, and a Really Interesting New Gene (RING) domain (Fig. 5.1). All the domains have been shown to be important for Dnmt1 subnuclear localization and maintenance of DNA methylation. Biochemical and structural evidence revealed that the SRA domain preferentially binds hemi-methylated DNA and is thought to play an important role in loading Dnmt1 onto newly synthesized DNA substrates (Arita et al. 2008; Avvakumov et al. 2008; Bostick et al. 2007; Hashimoto et al. 2008; Sharif et al. 2007). The association of Uhrf1 with heterochromatin is mediated by TTD, which contains an aromatic cage for binding of the heterochromatic H3K9me3 mark. The PHD acts in combination with TTD to read the H3K9me3 mark and, additionally, interacts with the unmethylated arginine 2 of histone H3 tail (H3R2me0) (Cheng et al. 2013; Liu et al. 2013; Rothbart et al. 2013; Rothbart et al. 2012; Rottach et al. 2010). There is also evidence that Uhrf1, via the E3 ligase activity of its RING domain, mediates ubiquitylation of histone H3 at several lysine residues on the N-terminal tail, creating binding sites for Dnmt1 (Ishiyama et al. 2017; Nishiyama et al. 2013; Qin et al. 2015). Recently, the UBL domain was shown to interact with the E2 ubiquitin conjugating enzyme UBE2D to facilitate H3 monoubiquitylation (DaRosa et al. 2018; Foster et al. 2018). It is worth noting that Uhrf1 also controls the ubiquitylation and stability of itself and Dnmt1 (Dan et al. 2017; Du et al. 2010; Qin et al. 2011). Indeed, transgenic Uhrf1 overexpression in zebrafish hepatocytes drives DNA hypomethylation by destabilizing Dnmt1 (Mudbhary et al. 2014). Thus, the role of Uhrf1 in the regulation of DNA methylation is complex.

Recent studies indicate that Uhrf1 is a major point of regulation in modulating DNA methylation. Methylation of DNA ligase 1 (LIG1) by GLP/G9a (also known as EHMT1/EHMT2) enhances the recruitment of Uhrf1 to replicating DNA (Ferry et al. 2017), while Uhrf1 acetylation by PCAF at K490 disrupts Uhrf1 binding to hemi-methylated DNA (Hahm et al. 2020). The DNA helicase HELLS (also known as LSH)

facilitates Uhrf1 chromatin association and Uhrf1-catalyzed histone H3 ubiquitylation, which in turn promotes Dnmt1 recruitment to replication forks and DNA methylation (Han et al. 2020). The protein arginine methyltransferase PRMT6 is overexpressed in multiple types of cancer (Yang and Bedford 2013). It has been reported that PRMT6-mediated asymmetric dimethylation of H3R2 (H3R2me2a) disrupts Uhrf1 association with chromatin and contributes to global DNA hypomethylation in cancer cells (Veland et al. 2017). In culture, pluripotent mESCs sporadically convert to a transient totipotent state, known as two-cell embryo-like ESCs (2CLCs), to extend shortened telomeres and repair damaged DNA (Macfarlan et al. 2012; Zalzman et al. 2010). Zscan4, a zinc-finger protein specifically and highly expressed in 2CLCs, has been shown to facilitate telomere elongation by inducing global DNA demethylation due to Uhrf1-dependent ubiquitylation and degradation of Uhrf1 itself and Dnmt1 (Dan et al. 2017).

5.2 Dnmts in Embryonic Development and Cellular Differentiation

5.2.1 Roles of Dnmts in Embryonic Development

DNA methylation is relatively stable in somatic tissues but exhibits dynamic changes in early embryos. During preimplantation development, both the maternal and paternal genomes undergo global DNA demethylation, albeit the mechanisms involved are distinct. Demethylation of the paternal genome involves both active and passive mechanisms. Shortly after fertilization and before the first cell division, the 5mC dioxygenase Tet3, which is highly expressed in oocytes and abundant in zygotes, converts the majority of 5mC in the male pronucleus to 5-hydroxymethylcytosine (5hmC) (Gu et al. 2011; Wossidlo et al. 2011). 5hmC can be further oxidized to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), which can be excised by thymine DNA glycosylase (TDG) and replaced by unmodified cytosine (He et al. 2011; Ito et al. 2011). The oxidized derivatives of 5mC can also persist in the paternal genome and undergo passive dilution during cleavage divisions (Inoue et al. 2011; Inoue and Zhang 2011). Even though exposed to an identical environment in the zygote as the paternal genome, the maternal genome is protected from Tet3-mediated active demethylation. DPPA3, a maternal factor, protects the maternal genome (as well as imprinted loci in the paternal genome) from Tet3-mediated oxidation of 5mC by binding to chromatin containing H3K9 methylation (Nakamura et al. 2007). Deletion of the H3K9me2 methyltransferase G9a or the H3K9me3 methyltransferase Setdb1 (also known as ESET and KMT1E) in growing oocytes results in significant reductions of the asymmetry of global 5mC oxidation (Zeng et al. 2019). The maternal genome is demethylated mainly through DNA replication-dependent passive dilution because of deficient maintenance methylation, presumably due to the exclusion of Dnmt1 from the nucleus (Hirasawa et al. 2008; Howell et al. 2001). As a result of active and passive demethylation, DNA methylation marks inherited from gametes are largely erased by the blastocyst stage in mice, with the exception of imprinting control regions (ICRs) and some retrotransposons, which resist this wave of global demethylation. Around the time of implantation, a wave of de novo methylation takes place to establish the embryonic methylation patterns, which are then maintained in a lineage-specific manner. In contrast to DNA demethylation in preimplantation embryos in mice, rhesus monkey embryos undergo remethylation at 8-cell stage, resulting in an increase in global DNA methylation, highlighting the difference in DNA methylation reprogramming during embryogenesis among different species (Gao et al. 2017).

Most of our knowledge about the significance of DNA methylation in mammalian development came from genetic manipulations of *Dnmt* genes in mice. Results from characterization of *Dnmt* mutant mice demonstrated that the establishment of embryonic methylation patterns requires both

de novo and maintenance Dnmts, and that maintaining genomic methylation above a threshold level is essential for embryonic development (Lei et al. 1996; Li et al. 1992; Okano et al. 1999). Complete inactivation of Dnmt1 results in the arrest of embryonic development between presomite and 8-somite stage around E9.5 (Lei et al. 1996). DNA methylation analysis showed that embryos deficient for Dnmt1 undergo dramatic loss of global DNA methylation (Lei et al. 1996; Li et al. 1992), in agreement with its role in maintenance methylation. The disruption of *Dnmt3b* also leads to embryonic lethality after E12.5, with multiple defects, including growth impairment and rostral neural tube defects (Okano et al. 1999). In contrast, *Dnmt3a*-deficient mice develop to term and appear normal at birth but die around 4 weeks of age (Okano et al. 1999). Consistent with the developmental phenotypes, DNA methylation analysis of E9.5 embryos revealed that germline-specific genes, pluripotency genes, hematopoietic genes, and eye genes are severely hypomethylated in the absence of Dnmt3b, but not Dnmt3a (Borgel et al. 2010). This suggests that Dnmt3b is the main enzyme responsible for de novo methylation during embryogenesis. Dnmt3b shows a dynamic expression change during pre- and early post-implantation development, with preferential expression in the trophectoderm at the mid-blastocyst stage and subsequent transition of expression in the embryonic lineage (Hirasawa and Sasaki 2009). Notably, DNA methylation at certain genes such as *Brdt, Dpep3, Cytip*, and *Crygd* is only partially reduced in *Dnmt3b* KO embryos (Borgel et al. 2010), suggesting that Dnmt3a cooperates with Dnmt3b to methylate some loci. Indeed, *Dnmt3a/3b* DKO embryos exhibit more severe defects than *Dnmt3b* KO embryos. Specifically, DKO embryos show smaller size, lack somites, do not undergo embryonic turning, and die before E11.5, indicating that their growth and morphogenesis are arrested shortly after gastrulation (Okano et al. 1999).

DNA methylation plays critical roles in gene repression and silencing of transposable elements (TEs). Thus, aberrant gene expression and reactivation of TEs likely contribute to the developmental defects observed in *Dnmt* mutant mice. Dnmt1, Dnmt3a, and Dnmt3b are all highly expressed in pluripotent mESCs, but the disruption of these genes individually, both *Dnmt3a* and *Dnmt3b*, or even all three *Dnmts* has no deleterious effects on mESCs in the undifferentiated state (Lei et al. 1996; Li et al. 1992; Okano et al. 1999; Tsumura et al. 2006). However, *Dnmt1* KO and *Dnmt3a/3b* DKO mESCs die upon induction of differentiation (Chen et al. 2003; Lei et al. 1996; Tucker et al. 1996). In contrast to mESCs, undifferentiated human ESCs (hESCs) require *DNMT1*, but not *DNMT3A* and *DNMT3B*, for survival (Liao et al. 2015). mESCs and hESCs represent different pluripotent states, with hESCs resembling the more mature epiblast state (Tesar et al. 2007), which could explain the sensitivity of hESCs to severe loss of DNA methylation.

5.2.2 Roles of Dnmts in Cellular Differentiation and Maintenance of Cell Identity.

The effects of DNA methylation deficiency become apparent during or after gastrulation, when the embryo differentiates to form the three germ layers (Lei et al. 1996; Li et al. 1992; Okano et al. 1999). Conditional inactivation of *Dnmt1* in mouse embryonic fibroblasts (MEFs) leads to severe hypomethylation and cell death, while *Dnmt3b*-deficient MEFs show modest hypomethylation, chromosomal instability, and abnormal cell proliferation (Dodge et al. 2005; Jackson-Grusby et al. 2001). Furthermore, although a hypomorphic mutation affecting the N-terminal region of human DNMT1 has no effect on the survival and proliferation of the colon cancer cell line HCT116 (Rhee et al. 2000), the disruption of the DNMT1 catalytic domain in HCT116 leads to mitotic catastrophe and cell death (Chen et al. 2007). Conditional KO studies in mice have also demonstrated that DNA methylation is essential in various organs and tissues. For example, Dnmt1 is important in the regulation of cell survival and neuronal maturation in the central nervous system and for T cell

development, function, and survival (Hutnick et al. 2009; Lee et al. 2001). These results suggest crucial roles for DNA methylation in cellular differentiation and in the viability and proper functioning of differentiated cells.

5.2.3 DNMT Mutations in Human Diseases

Consistent with the role of DNA methylation in cellular differentiation and maintenance of differentiated cell state, *DNMT* mutations have been identified in cancer and developmental disorders. *DNMT1* mutations in RFTS are reported in two related neurodegenerative diseases, hereditary sensory and autonomic neuropathy with dementia and hearing loss type IE (HSAN IE) and autosomal dominant cerebellar ataxia, deafness and narcolepsy (ADCA-DN) (Klein et al. 2011; Winkelmann et al. 2012). Somatic *DNMT3A* mutations are frequently observed in patients with hematologic malignancies including acute myeloid leukemia (AML), myelodysplastic syndrome (MDS), and T cell acute lymphoblastic leukemia (T-ALL), with almost all mutations being heterozygous (Yang et al. 2015). The DNMT3A R882 hotspot mutation, which is found in ~30% of normal karyotype AML, has been shown to dominantly inhibit WT DNMT3A and alter the flanking sequence preferences of DNMT3A, resulting in hypomethylation at specific CpGs throughout the genome (Emperle et al. 2018, 2019; Kim et al. 2013; Norvil et al. 2020; Russler-Germain et al. 2014). Changes in DNA methylation results in the inhibition of hematopoietic cell differentiation, facilitating leukemogenesis. Germline mutations in *DNMT3A* are associated with Tatton-Brown-Rahman syndrome, a congenital disorder characterized by overgrowth, macrocephaly, and intellectual disability (Tatton-Brown et al. 2014). A recent study identified heterozygous gain-of-function missense mutations (W330R and D333N) in the PWWP domain of DNMT3A in patients with microcephalic dwarfism, who exhibit hypermethylation of Polycomb-regulated regions (Heyn et al. 2019). *DNMT3B* mutations account for ~50% of cases with the immunodeficiency, centromeric instability, and facial anomalies (ICF) syndrome (Hansen et al. 1999; Okano et al. 1999; Xu et al. 1999), and the other ~50% of cases are caused by mutations in *ZBTB24*, *CDCA7*, *HELLS,* and one or more unknown genes (de Greef et al. 2011; Thijssen et al. 2015). Recent evidence indicates that the four known ICF-associated genes are functionally connected in regulating the specificity of DNA methylation. Specifically, ZBTB24 directly activates *CDCA7* transcription, and CDCA7 recruits the chromatin remodeler HELLS to (peri)centromeric heterochromatin to facilitate the chromatin accessibility to the DNA methylation machinery, including DNMT3B (Hardikar et al. 2020; Jenness et al. 2018; Ren et al. 2019; Thompson et al. 2018).

In recent years, progress has been made in modeling some of these diseases in mice. Mice heterozygous for W326R or D329A substitution (corresponding to human W330 and D333, respectively) in the Dnmt3a PWWP domain recapitulate the dwarfism phenotype observed in human patients. The Dnmt3a PWWP domain is important for the normal DNA methylation landscape, as the point mutations result in gain of DNA methylation in some hypomethylated regions, including those marked by H3K27me3 and H3K4me3/H3K27me3 (bivalent) chromatin (Heyn et al. 2019; Kibe et al. 2021; Sendzikaite et al. 2019). A recent study suggests that, when the Dnmt3a PWWP domain is mutated, the UDR domain in the Dnmt3a N terminus, which recognizes H2AK119ub, may target Dnmt3a to Polycomb-regulated regions (Weinberg et al. 2021). Mice carrying Dnmt3a R878H mutation (corresponding to the hotspot DNMT3A R882H mutation in AML patients) initiate leukemogenesis due to the activation of mTOR and loss of DNA methylation at specific genomic regions. Indeed, the mTOR inhibitor rapamycin elicits a significant therapeutic response in these mice (Dai et al. 2017). *Dnmt3b* mutations in mice recapitulate some aspects of ICF syndrome (e.g., facial anomalies), but not antibody deficiency, the major cause of infection and death occurring in

patients with ICF syndrome (Ueda et al. 2006). While germline deletion of *Zbtb24* leads to early embryonic lethality (Wu et al. 2016), conditional ablation of *Zbtb24* in B cells impairs nonhomologous end-joining and class-switch recombination (Helfricht et al. 2020). *Hells* KO mice display defects in B lymphocyte development and immunoglobulin class-switch recombination (He et al. 2020). The defects observed in *Zbtb24* KO and *Hells* KO mice could contribute to antibody deficiency observed in ICF syndrome.

5.3 Dnmts in Genomic Imprinting

In the 1980s, elegant nuclear transplantation experiments using pronuclear stage embryos showed that mouse embryos constructed to contain only maternal or paternal diploid genome complements fail to develop beyond mid-gestation. This suggested that the parental genomes are functionally non-equivalent and marked or "imprinted" differently during male and female gametogenesis (Barton et al. 1984; McGrath and Solter 1984; Surani et al. 1984). Separate experiments using chromosome translocations in mice showed that specific chromosome segments function differently depending on the parental origin (Cattanach and Kirk 1985). In the early 1990s, the first murine imprinted genes, *Igf2r*, *Igf2*, and *H19*, were discovered, which are expressed only from one parental allele (Barlow et al. 1991; Bartolomei et al. 1991; DeChiara et al. 1991). To date, hundreds of imprinted genes, which exhibit mono-allelic expression strictly according to the parental origin, have been identified in mouse and human, although only some of them are imprinted in both species. Imprinted genes are involved in diverse biological processes, including embryonic development, placental formation, fetal and postnatal growth, metabolism and cognitive behaviors (Tucci et al. 2019). In human, altered expression of imprinted genes, due to genetic and epigenetic changes, is linked to infertility, molar pregnancy, and various congenital disorders such as Prader–Willi syndrome, Angelman syndrome, Beckwith–Wiedemann syndrome, and Silver–Russell syndrome (Tomizawa and Sasaki 2012). Loss of imprinting (i.e., either biallelic expression or complete silencing of imprinted genes) is frequently observed in cancer (Jelinic and Shaw 2007).

The majority of imprinted genes are arranged in chromosomal clusters, which usually span hundreds to thousands of kilobases. Each of the imprinting clusters contains an imprinting control region (ICR), an essential regulatory sequence that contains one or more differentially methylated regions (DMRs)—i.e. regions marked by DNA methylation only on one of the two parental alleles. Thus, allele-specific DNA methylation is considered the primary epigenetic mark controlling the mono-allelic expression of most imprinted genes. Studies in recent years have also identified DNA methylation-independent (noncanonical) imprinting that occurs at a small number of loci (see below).

5.3.1 Establishment of DNA Methylation Imprints during Gametogenesis

The life cycle of DNA methylation imprints consists of three major steps: establishment, maintenance, and erasure (Fig. 5.3). DNA methylation imprints are acquired in the germline, with the majority being established during oogenesis (maternally imprinted) and only four known loci (*H19, Dlk1-Gtl2, Rasgrf1*, and *Zdbf2*) being established during spermatogenesis (paternally imprinted). Conditional deletion of *Dnmt3a* in primordial germ cells (PGCs) disrupts both maternal and paternal imprinting. Embryos from crosses between conditional *Dnmt3a* mutant females and WT males die around E10.5, and conditional *Dnmt3a* mutant males are sterile due to impaired spermatogenesis (Kaneda et al. 2004). *Dnmt3L* KO mice show an identical phenotype, with the exception of one paternally methylated locus, *Dlk1-Gtl2,* which does not require *Dnmt3L* for methylation (Bourc'his et al. 2001; Hata et al. 2002; Kaneda et al. 2004). In contrast, conditional deletion of *Dnmt3b* in PGCs shows no apparent phenotype (Kaneda et al.

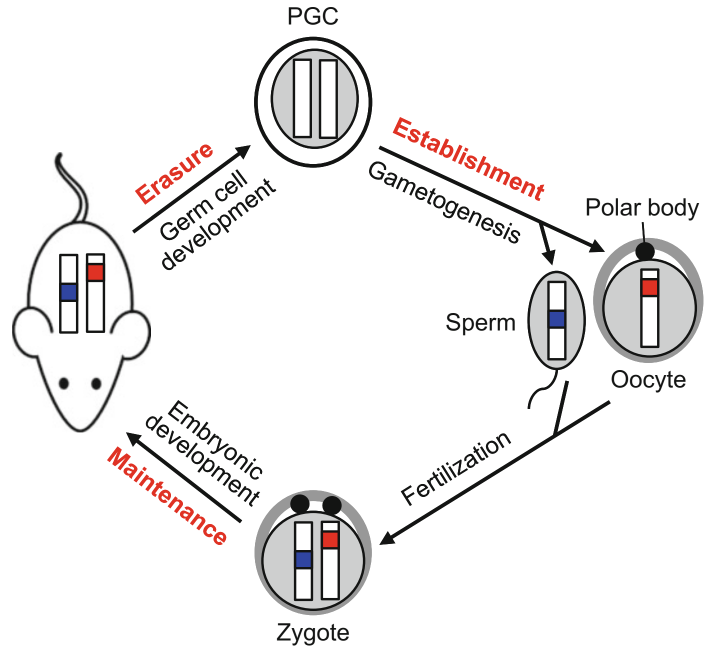


Fig. 5.3 Life cycle of DNA methylation imprints. The paternal (blue) and maternal (red) methylation imprints are established during gametogenesis and transmitted to the offspring through fertilization. These marks are maintained and control mono-allelic expression of imprinted genes during embryogenesis and in somatic cells throughout adult life. However, they are erased in primordial germ cells (PGCs) before sex-specific methylation imprints are re-established in later stages of germ cell development

2004). These results provide compelling genetic evidence that Dnmt3a (mainly Dnmt3a2) is responsible for the establishment of germline imprints and Dnmt3L is an essential cofactor for Dnmt3a in this process.

The mechanisms by which Dnmt3L facilitates Dnmt3a function are complex. Dnmt3L, via its C-terminal domain, forms a tetrameric complex with Dnmt3a (Jia et al. 2007). In vitro experiments revealed that Dnmt3L stimulates the enzymatic activity of Dnmt3a (Hata et al. 2002; Suetake et al. 2004). Via its ADD domain, Dnmt3L interacts with the N-terminal tail of histone H3 when H3K4 is unmethylated (Ooi et al. 2007). Indeed, mice homozygous for an engineered point mutation (D124A) in the Dnmt3L ADD domain exhibit defects in DNA methylation and spermatogenesis (Vlachogiannis et al. 2015), supporting a critical role of the ADD domain in Dnmt3L function in the male germline. It remains to be determined whether female mice homozygous for the D1124A mutation show defects in the establishment of maternal imprints. A recent study showed that Dnmt3L also plays a role in maintaining Dnmt3a (especially Dnmt3a2) stability (Veland et al. 2019).

The observation that H3K4 methylation disrupts the interaction between the Dnmt3a/3 L complex and histone H3 (Ooi et al. 2007; Otani et al. 2009) suggests that chromatin organization may be an important determinant of the sites of de novo DNA methylation in the germ line. Indeed, genetic evidence indicated that the H3K4 demethylase KDM1B (also known as LSD2 and AOF1) is essential for the establishment of a subset of maternal imprints (Ciccone et al. 2009). KDM1B is highly expressed in growing oocytes, where maternal imprints are acquired, but shows little expression in most somatic tissues. *Kdm1b* KO mice are viable and show no defects in spermatogenesis and oogenesis, and male mice are fertile. However, oocytes from KDM1B-deficient females exhibit global accumulation of H3K4me2 and fail to establish DNA methylation imprints at a subset of imprinted loci. Consequently, embryos derived from these oocytes die around mid-gestation (Ciccone et al. 2009), similar to embryos derived from Dnmt3L- or Dnmt3a-deficient female mice (Bourc'his et al. 2001; Hata et al. 2002; Kaneda et al. 2004). These results strongly suggest that

removal of H3K4 methylation is a prerequisite for de novo DNA methylation at imprinted loci.

Transcription is another requirement for DNA methylation, at least at some imprinted loci. In the mouse *Gnas* locus, transcription initiated at the promoter of *Nesp55*, a gene upstream of the DMRs of the *Gnas* locus, occurs in growing oocytes, placing a large genomic region, including the DMRs, within an active transcription unit. Deletion of the *Nesp55* promoter or insertion of a transcription termination cassette downstream of *Nesp55* to ablate transcription results in failure to establish DNA methylation at the ICR of the *Gnas* locus (Chotalia et al. 2009; Frohlich et al. 2010; Williamson et al. 2011). Methylation of the DMR at the *Snrpn* locus also depends upon transcription (Smith et al. 2011). The proposed model for transcription-dependent genomic imprinting is that transcription elongation causes an enrichment of H3K36me2/3 at the transcribed regions, which recruit the Dnmt3a/Dnmt3L complex to establish DNA methylation in oocytes. In support of this model, the depletion of the H3K36me3 methyltransferase Setd2 in oocytes causes genome-wide loss of H3K36me3, aberrant acquisition of H3K4me3 instead of DNA methylation at ICR of maternally imprinted loci (Xu et al. 2019). In contrast, Setd2 is dispensable for de novo DNA methylation in male germline. Instead, the H3K36me2 methyltransferase Nsd1 plays a critical role in de novo DNA methylation in prospermatogonia, including at imprinted genes (Shirane et al. 2020).

5.3.2 Maintenance of DNA Methylation Imprints during Development

The second step of the life cycle of genomic imprints is the maintenance of mono-allelic DNA methylation marks in the offspring. The paternal and maternal imprints are transmitted to the zygote through fertilization and, despite extensive demethylation during preimplantation development, parental allele-specific DNA methylation imprints are faithfully maintained through development and adult life. Notably, genome-wide DNA methylation analyses revealed far more differentially methylated loci in oocytes and sperm than the number of imprinted genes. Therefore, the previous notion that imprinted loci are determined by distinct methylation patterns in gametes has been revised to the current view that genomic imprinting results from selective maintenance of germline-derived allele-specific methylation. Genetic studies using conditional KO mice demonstrated that Dnmt1, but not Dnmt3a or Dnmt3b, is responsible for maintaining methylation marks at imprinted loci during preimplantation development (Hirasawa et al. 2008). The oocyte-specific variant, Dnmt1o, is the predominant Dnmt1 isoform in preimplantation embryos (Hirasawa et al. 2008; Kurihara et al. 2008). However, offspring of females lacking Dnmt1o exhibit only a ~ 50% reduction of methylation at certain imprinted loci (Howell et al. 2001). While initial evidence suggested that the somatic form, Dnmt1, is not expressed until the blastocyst stage (Ratnam et al. 2002), subsequent work showed that Dnmt1 is present at very low levels in the nucleus of oocytes and preimplantation embryos (Hirasawa et al. 2008; Kurihara et al. 2008). Conditional deletion of *Dnmt1* (both Dnmt1o and Dnmt1) in growing oocytes leads to a partial loss of methylation imprints in the offspring (Hirasawa et al. 2008), resembling the effect of Dnmt1o loss (Howell et al. 2001). However, the ablation of both maternal and zygotic Dnmt1 results in a complete loss of methylation at paternally and maternally methylated DMRs in embryos (Hirasawa et al. 2008). Therefore, both maternal and zygotic Dnmt1 proteins are necessary for the maintenance of methylation imprints in preimplantation embryos. Dnmt1 is also responsible for the maintenance of methylation imprints in post-implantation embryos (Li et al. 1993) and likely in adult somatic cells as well.

It is not well understood what confers the specificity of Dnmt1, such that methylation is maintained at imprinted genes, but not at other sequences in preimplantation embryos. Genetic and epigenetic features may distinguish imprinted loci from other regions. Several factors have been shown to be essential for the maintenance of DNA methylation imprints. DPPA3, a

DNA-binding protein, is highly expressed in oocytes and persists in preimplantation embryos. Genetic evidence suggests that, in early embryos, maternal DPPA3 plays a crucial role in protecting the maternal genome against DNA demethylation. DPPA3 also protects the paternally imprinted *H19* and *Rasgrf1* against demethylation (Nakamura et al. 2007). While the mechanisms involved remain to be determined, DPPA3 has been shown to play a role in chromatin condensation during oogenesis and to protect the maternal genome against Tet3-mediated conversion of 5mC to 5hmC in early embryos (Bian and Yu 2014; Liu et al. 2012; Nakamura et al. 2012). The *Krüppel* associated box (KRAB)-containing zinc-finger proteins ZFP57 and ZFP445 are also involved in the maintenance of genomic imprints (Li et al. 2008; Takahashi et al. 2019), as mouse embryos lacking ZFP57 or ZFP445 fail to maintain DNA methylation at most ICRs. Human *ZFP57* mutations are associated with hypomethylation at multiple imprinted loci in patients with transient neonatal diabetes (Mackay et al. 2008), suggesting conserved roles of ZFP57 in evolution. Mechanistic studies indicate that ZFP57 and ZFP445 act together to specifically bind the methylated allele of ICRs, recognizing a hexanucleotide sequence (TGC^{m}CGC) shared by all murine ICRs and most human putative ICRs (Quenneville et al. 2011; Takahashi et al. 2019). ZFP57 and ZFP445 recruit KAP1 (also known as Trim28) to ICRs, which acts as a scaffold protein for various heterochromatin proteins, including heterochromatin protein 1 (HP1), the histone H3K9me3 methyltransferase Setdb1, the nuclear remodeling and histone deacetylation (NuRD) complex, and the DNA methylation machinery (Nielsen et al. 1999; Quenneville et al. 2011; Ryan et al. 1999; Schultz et al. 2001, 2002; Zuo et al. 2012). The ablation of either maternal or zygotic KAP1 causes partial loss of DNA methylation imprints, and the ablation of both maternal and zygotic KAP1 leads to a complete loss of imprinting (Lorthongpanich et al. 2013; Messerschmidt et al. 2012; Quenneville et al. 2011). The depletion of the NuRD components MBD3 or MTA2 also results in the reduction of methylation at some imprinted loci in preimplantation embryos (Ma et al. 2010; Reese et al. 2007). N-α-acetyltransferase 10 protein (Naa10p) is another protein that has been shown to facilitate Dnmt1 binding to ICRs of imprinted loci during S phase (Lee et al. 2017). In addition to Dnmt1-mediated maintenance of DNA methylation marks at ICRs, it is equally important to prevent ICRs from being demethylated. DPPA3 has been shown to prevent erasure of DNA methylation marks at maternally imprinted genes, as well as two paternally imprinted genes, *H19* and *Rasgrf1*, presumably by binding to H3K9me2/3 and blocking the recruitment of Tet3 (Nakamura et al. 2007; Zeng et al. 2019).

5.3.3 Erasure of DNA Methylation Imprints in Primordial Germ Cells

The last step of the imprint life cycle is the erasure of methylation imprints in PGCs, which ensures the establishment of sex-specific imprints in later stages of germ cell development. In mice, PGCs are specified around E7.25 in the epiblast of the developing embryo. Shortly afterwards, PGCs begin migrating along the embryonic–extraembryonic interface and eventually arrive at the genital ridge by E12.5. Genome-wide DNA methylation analyses reveal that PGCs undergo demethylation in two major phases (Guibert et al. 2012; Kobayashi et al. 2013; Popp et al. 2010; Seisenberger et al. 2012). The first phase takes place during PGC expansion and migration from ~E8.5, which leads to a global demethylation affecting almost all genomic regions. Passive demethylation likely plays a major role in this phase, as a result of repression of Uhrf1, as well as Dnmt3a and Dnmt3b, in PGCs (Kagiwada et al. 2013; Kurimoto et al. 2008). The second phase occurs from E9.5 to E13.5 and affects specific loci including ICRs, germline-specific genes, and CpG islands on the X chromosome (Guibert et al. 2012; Hackett et al. 2013; Popp et al. 2010; Seisenberger et al. 2012; Yamaguchi et al. 2013a). Tet1- and Tet2-mediated 5mC oxidation is important in the second phase of

demethylation (Zeng and Chen 2019; Zhao and Chen 2013). Genetic studies in mouse indicate that Tet1 deficiency leads to aberrant DNA hypomethylation at a subset of ICRs in germ cells and somatic tissues and results in the activation of germline genes involved in gametogenesis and meiosis (Hill et al. 2018; SanMiguel et al. 2018; Yamaguchi et al. 2012, 2013b). Thus, both passive and active demethylation pathways are involved in the erasure of parental imprints in PGCs.

5.3.4 Noncanonical Genomic Imprinting

In addition to DNA methylation-dependent "canonical" imprinting, recent studies in mice have shown that maternal H3K27me3 can lead to DNA methylation-independent "noncanonical" imprinting, which occurs at several placenta-specific genes (e.g., *Gab1*, *Sfmbt2*, *Slc38a4*, and *Phf17*) (Inoue et al. 2017a; Okae et al. 2012). These genes, which are not associated with germline DMRs, are paternally expressed in preimplantation embryos. After implantation, they become biallelically expressed or repressed in the epiblast (which gives rise to the embryonic lineages), while retaining imprinted expression in the extraembryonic lineages and placenta (Inoue et al. 2017a). Loss of H3K27me3-mediated imprinting in mice cloned by somatic cell nuclear transfer (SCNT) could contribute to placental hyperplasia and postnatal developmental defects (Matoba et al. 2018; Okae et al. 2014), highlighting the importance of noncanonical imprinting. Oocyte-derived H3K27me3 also serves as a maternal imprint for the long non-coding RNA *Xist*, leading to paternal X chromosome inactivation (XCI) in female preimplantation embryos and extraembryonic tissues (i.e., imprinted XCI). The H3K27me3 imprinting mark is established during oocyte growth. Oocyte-specific deletion of *Eed*, which encodes an essential component of Polycomb repressive complex 2 (PRC2), abrogates noncanonical imprinting and imprinted XCI (Inoue et al. 2017b, 2018). In contrast, embryos derived from DNA methylation-deficient oocytes show intact noncanonical imprinting (Chen et al. 2019; Hanna et al. 2019). Unlike DNA methylation-dependent canonical imprinting, which is maintained throughout development, H3K27me3-mediated noncanonical imprinting is transient in preimplantation embryos and absent after implantation (Chen et al. 2019; Hanna et al. 2019; Inoue et al. 2017a). Instead, maintenance of noncanonical imprinting in extraembryonic cells requires de novo DNA methylation of the maternal allele (i.e., acquisition of somatic DMRs) at implantation, which is mediated by zygotic Dnmt3a/3b (Chen et al. 2019; Hanna et al. 2019). Recent evidence suggests that the H3K9me2 methyltransferase G9a is also critical in the maintenance of noncanonical imprinting, likely by regulating the acquisition of somatic DMRs (Andergassen et al. 2021; Zeng et al. 2021).

Imprinted XCI is absent in humans (Looijenga et al. 1999; Migeon and Do 1979), and the *Xist* locus is devoid of H3K27me3 in human preimplantation embryos (Xia et al. 2019). Whether noncanonical imprinting by H3K27me3 exists in humans is controversial, although candidate genes have been reported (Hanna and Kelsey 2021).

5.4 Concluding Remarks

DNA methylation, a relatively stable epigenetic mark, acts in concert with other epigenetic mechanisms such as histone modifications to stably maintain gene silencing and chromatin structure. Aberrant DNA methylation patterns are associated with various human diseases, including cancer and developmental disorders. Since the discovery of mammalian Dnmts (Bestor et al. 1988; Okano et al. 1998a), great progress has been made in understanding the biological functions of DNA methylation in mammals. Genetic studies using *Dnmt* mutant mice and murine cells have provided important insights into the roles of DNA methylation in various biological processes and the mechanisms by which mutations in DNMTs contribute to disease

Table 5.1 Phenotypes of *Dnmt* mutant mice

Gene	Mutations	Major Phenotypes	References
Dnmt1	*Dnmt1*$^{n/n}$	~70% reduction in global DNA methylation, embryonic lethality at E12.5–15.5	Li et al. (1992)
	Dnmt1$^{c/c}$	~90% reduction in global DNA methylation, embryonic lethality at ~E9.5. Unstable XCI	Lei et al. (1996), Sado et al. (2000)
	Dnmt1$^{s/s}$	Similar to *Dnmt1*$^{c/c}$	Beard et al. (1995), Lei et al. (1996)
	Dnmt1o$^{-/-}$	Maternal-effect phenotype: Partial loss of DNA methylation imprints, defects in imprinted XCI, embryonic lethality at mid-gestation	Howell et al. (2001), McGraw et al. (2013)
	Both maternal and zygotic *Dnmt1*$^{-/-}$	Complete loss of DNA methylation imprints, embryonic lethality at mid-gestation	Hirasawa et al. (2008)
Dnmt3a	*Dnmt3a*$^{-/-}$	Gut malfunction, spermatogenesis defects, death at ~4 weeks of age	Okano et al. (1999)
	Dnmt3a$^{-/-}$ in PGCs	Failure to establish DNA methylation imprints, spermatogenesis defects	Kaneda et al. (2004)
	Heterozygous W326R or D329A KI in PWWP domain (corresponding to human W330 and D333)	Hypermethylation of regions marked by H3K27me3 and H3K4me3/H3K27me3 (bivalent) chromatin, dwarfism	Heyn et al. (2019), Kibe et al. (2021), Sendzikaite et al. (2019)
Dnmt3b	*Dnmt3b*$^{-/-}$	Hypomethylation of some regions, including minor satellite DNA, neural tube defects, embryonic lethality at E14.5–18.5	Okano et al. (1999)
Dnmt3a & 3b	*Dnmt3a*$^{-/-}$, *Dnmt3b*$^{-/-}$	Failure to initiate de novo methylation after implantation, embryonic lethality before E11.5.	Okano et al. (1999)
Dnmt3L	*Dnmt3L*$^{-/-}$	Failure to establish DNA methylation imprints, spermatogenesis defects	Bourc'his et al. (2001), Hata et al. (2002)
	Homozygous D124A KI in ADD domain	Hypomethylation in male germ cells, spermatogenesis defects	Vlachogiannis et al. (2015)
Dnmt3c	*Dnmt3c*$^{-/-}$	Failure to silence evolutionally young retrotransposons, spermatogenesis defects	Barau et al. (2016)

phenotypes (Tables 5.1 and 5.2). Nevertheless, some key questions remain to be answered. For example, what retains Dnmt1 in the cytoplasm in oocytes and early embryos that presumably leads to passive demethylation during preimplantation development? How are genes selected for the establishment of canonical or noncanonical imprinting during gametogenesis? In the coming years, we expect to see the generation of mouse models that better recapitulate the major features of human diseases associated with *DNMT* mutations. This becomes more feasible with the development of new technologies such as CRISPR/Cas9-mediated gene editing. Genomic, epigenomic, transcriptomic, and proteomic analyses of these models will be powerful approaches for defining the molecular mechanisms and pathways involved in pathogenesis. These studies will likely lead to novel therapeutic strategies, and the genetically engineered mouse models would be valuable tools for testing therapeutics.

Table 5.2 Human diseases caused by *DNMT* mutations

Gene	Mutations	Diseases	References
DNMT1	Heterozygous missense mutations (D490E, P491Y, Y495C) in RFTS	Hereditary sensory and autonomic neuropathy with dementia and hearing loss type IE (HSAN IE)	Klein et al. (2011)
	Heterozygous missense mutations (A570V, G605A, V606F) in RFTS	Autosomal dominant cerebellar ataxia, deafness and narcolepsy (ADCA-DN)	Winkelmann et al. (2012)
DNMT3A	Somatic heterozygous mutations (multiple), with R882 as hotspot	Acute myeloid leukemia (AML), myelodysplastic syndrome (MDS), T cell acute lymphoblastic leukemia (T-ALL)	Yang et al. (2015)
	De novo heterozygous mutations (multiple)	Tatton-Brown-Rahman syndrome	Tatton-Brown et al. (2014)
	Gain-of-function heterozygous mutations (W330R, D333N) in PWWP domain	Microcephalic dwarfism	Heyn et al. (2019)
DNMT3B	Hypomorphic compound heterozygous or homozygous mutations (multiple)	Immunodeficiency, centromeric instability, and facial anomalies (ICF) syndrome	Hansen et al. (1999), Okano et al. (1999), Xu et al. (1999)

Acknowledgments Work in the Chen laboratory is supported by grants from the National Institutes of Health (R01DK106418 and R01AI12140301). Work in the Dan laboratory is supported by grants from the Natural Science Foundation of Yunnan Province (202001BC070001 and 202102AA100053).

References

Aapola U, Kawasaki K, Scott HS, Ollila J, Vihinen M, Heino M, Shintani A, Minoshima S, Krohn K, Antonarakis SE et al (2000) Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics 65:293–298. https://doi.org/10.1006/geno.2000.6168

Aapola U, Lyle R, Krohn K, Antonarakis SE, Peterson P (2001) Isolation and initial characterization of the mouse Dnmt3l gene. Cytogenet Cell Genet 92(122–126):56881

Andergassen D, Smith ZD, Kretzmer H, Rinn JL, Meissner A (2021) Diverse epigenetic mechanisms maintain parental imprints within the embryonic and extraembryonic lineages. Dev Cell 56(2995–3005):e2994. https://doi.org/10.1016/j.devcel.2021.10.010

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature 455:818–821. https://doi.org/10.1038/nature07249

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, Arrowsmith CH, Dhe-Paganon S (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature 455:822–825. https://doi.org/10.1038/nature07273

Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Herault Y, Guillou F, Bourc'his D (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354:909–912. https://doi.org/10.1126/science.aah5143

Barlow DP, Stoger R, Herrmann BG, Saito K, Schweifer N (1991) The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. Nature 349:84–87. https://doi.org/10.1038/349084a0

Bartolomei MS, Zemel S, Tilghman SM (1991) Parental imprinting of the mouse H19 gene. Nature 351:153–155. https://doi.org/10.1038/351153a0

Barton SC, Surani MA, Norris ML (1984) Role of paternal and maternal genomes in mouse development. Nature 311:374–376. https://doi.org/10.1038/311374a0

Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, Akalin A, Schubeler D (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520:243–247. https://doi.org/10.1038/nature14176

Beard C, Li E, Jaenisch R (1995) Loss of methylation activates Xist in somatic but not in embryonic cells. Genes Dev 9:2325–2334. https://doi.org/10.1101/gad.9.19.2325

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

Bian C, Yu X (2014) PGC7 suppresses TET3 for protecting DNA methylation. Nucleic Acids Res 42:2893–2905. https://doi.org/10.1093/nar/gkt1261

Borgel J, Guibert S, Li Y, Chiba H, Schubeler D, Sasaki H, Forne T, Weber M (2010) Targets and dynamics of promoter DNA methylation during early mouse

development. Nat Genet 42:1093–1100. https://doi.org/10.1038/ng.708

Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317:1760–1764. https://doi.org/10.1126/science.1147939

Bourc'his D, Bestor TH (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature 431:96–99. https://doi.org/10.1038/nature02886

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints. Science 294:2536–2539. https://doi.org/10.1126/science.1065848

Cattanach BM, Kirk M (1985) Differential activity of maternally and paternally derived chromosome regions in mice. Nature 315:496–498. https://doi.org/10.1038/315496a0

Charlton J, Jung EJ, Mattei AL, Bailly N, Liao J, Martin EJ, Giesselmann P, Brandl B, Stamenova EK, Muller FJ et al (2020) TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. Nat Genet 52:819–827. https://doi.org/10.1038/s41588-020-0639-9

Chen T, Ueda Y, Xie S, Li E (2002) A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. J Biol Chem 277:38746–38754. https://doi.org/10.1074/jbc.M205312200

Chen T, Ueda Y, Dodge JE, Wang Z, Li E (2003) Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol 23:5594–5605

Chen T, Tsujimoto N, Li E (2004) The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Mol Cell Biol 24:9048–9058. https://doi.org/10.1128/MCB.24.20.9048-9058.2004

Chen T, Hevi S, Gay F, Tsujimoto N, He T, Zhang B, Ueda Y, Li E (2007) Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. Nat Genet 39:391–396. https://doi.org/10.1038/ng1982

Chen Z, Yin Q, Inoue A, Zhang C, Zhang Y (2019) Allelic H3K27me3 to allelic DNA methylation switch maintains noncanonical imprinting in extraembryonic cells. Sci Adv 5:eaay7246. https://doi.org/10.1126/sciadv.aay7246

Cheng J, Yang Y, Fang J, Xiao J, Zhu T, Chen F, Wang P, Li Z, Yang H, Xu Y (2013) Structural insight into coordinated recognition of trimethylated histone H3 lysine 9 (H3K9me3) by the plant homeodomain (PHD) and tandem tudor domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein. J Biol Chem 288:1329–1339. https://doi.org/10.1074/jbc.M112.415398

Chotalia M, Smallwood SA, Ruf N, Dawson C, Lucifero D, Frontera M, James K, Dean W, Kelsey G (2009) Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev 23:105–117. https://doi.org/10.1101/gad.495809

Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science 277:1996–2000. https://doi.org/10.1126/science.277.5334.1996

Ciccone DN, Su H, Hevi S, Gay F, Lei H, Bajko J, Xu G, Li E, Chen T (2009) KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. Nature 461:415–418. https://doi.org/10.1038/nature08315

Dai YJ, Wang YY, Huang JY, Xia L, Shi XD, Xu J, Lu J, Su XB, Yang Y, Zhang WN et al (2017) Conditional knockin of Dnmt3a R878H initiates acute myeloid leukemia with mTOR pathway involvement. Proc Natl Acad Sci U S A 114:5237–5242. https://doi.org/10.1073/pnas.1703476114

Dan J, Rousseau P, Hardikar S, Veland N, Wong J, Autexier C, Chen T (2017) Zscan4 inhibits maintenance DNA methylation to facilitate telomere elongation in mouse embryonic stem cells. Cell Rep 20:1936–1949. https://doi.org/10.1016/j.celrep.2017.07.070

DaRosa PA, Harrison JS, Zelter A, Davis TN, Brzovic P, Kuhlman B, Klevit RE (2018) A bifunctional role for the UHRF1 UBL domain in the control of hemi-methylated DNA-dependent histone ubiquitylation. Mol Cell 72(753–765):e756. https://doi.org/10.1016/j.molcel.2018.09.029

de Greef JC, Wang J, Balog J, den Dunnen JT, Frants RR, Straasheijm KR, Aytekin C, van der Burg M, Duprez L, Ferster A et al (2011) Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. Am J Hum Genet 88:796–804. https://doi.org/10.1016/j.ajhg.2011.04.018

DeChiara TM, Robertson EJ, Efstratiadis A (1991) Parental imprinting of the mouse insulin-like growth factor II gene. Cell 64:849–859. https://doi.org/10.1016/0092-8674(91)90513-x

Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem 285:26114–26120. https://doi.org/10.1074/jbc.M109.089433

Ding F, Chaillet JR (2002) In vivo stabilization of the Dnmt1 (cytosine-5)- methyltransferase protein. Proc Natl Acad Sci U S A 99:14861–14866. https://doi.org/10.1073/pnas.232565599

Dodge JE, Okano M, Dick F, Tsujimoto N, Chen T, Wang S, Ueda Y, Dyson N, Li E (2005) Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. J Biol Chem 280:17986–17991. https://doi.org/10.1074/jbc.M413246200

Du Z, Song J, Wang Y, Zhao Y, Guda K, Yang S, Kao HY, Xu Y, Willis J, Markowitz SD et al (2010) DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination. Sci Signal 3:ra80. https://doi.org/10.1126/scisignal.2001462

Duymich CE, Charlet J, Yang X, Jones PA, Liang G (2016) DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. Nat Commun 7:11453. https://doi.org/10.1038/ncomms11453

Easwaran HP, Schermelleh L, Leonhardt H, Cardoso MC (2004) Replication-independent chromatin loading of Dnmt1 during G2 and M phases. EMBO Rep 5:1181–1186. https://doi.org/10.1038/sj.embor.7400295

Emperle M, Rajavelu A, Kunert S, Arimondo PB, Reinhardt R, Jurkowska RZ, Jeltsch A (2018) The DNMT3A R882H mutant displays altered flanking sequence preferences. Nucleic Acids Res 46:3130–3139. https://doi.org/10.1093/nar/gky168

Emperle M, Adam S, Kunert S, Dukatz M, Baude A, Plass C, Rathert P, Bashtrykov P, Jeltsch A (2019) Mutations of R882 change flanking sequence preferences of the DNA methyltransferase DNMT3A and cellular methylation patterns. Nucleic Acids Res 47:11355–11367. https://doi.org/10.1093/nar/gkz911

Ferry L, Fournier A, Tsusaka T, Adelmant G, Shimazu T, Matano S, Kirsh O, Amouroux R, Dohmae N, Suzuki T et al (2017) Methylation of DNA ligase 1 by G9a/GLP recruits UHRF1 to replicating DNA and regulates DNA methylation. Mol Cell 67(550–565):e555. https://doi.org/10.1016/j.molcel.2017.07.012

Foster BM, Stolz P, Mulholland CB, Montoya A, Kramer H, Bultmann S, Bartke T (2018) Critical role of the UBL domain in stimulating the E3 ubiquitin ligase activity of UHRF1 toward chromatin. Mol Cell 72(739–752):e739. https://doi.org/10.1016/j.molcel.2018.09.028

Frohlich LF, Mrakovcic M, Steinborn R, Chung UI, Bastepe M, Juppner H (2010) Targeted deletion of the Nesp55 DMR defines another Gnas imprinting control region and provides a mouse model of autosomal dominant PHP-Ib. Proc Natl Acad Sci U S A 107:9275–9280. https://doi.org/10.1073/pnas.0910224107

Gao F, Niu Y, Sun YE, Lu H, Chen Y, Li S, Kang Y, Luo Y, Si C, Yu J et al (2017) De novo DNA methylation during monkey pre-implantation embryogenesis. Cell Res 27:526–539. https://doi.org/10.1038/cr.2017.25

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golic KG, Jacobsen SE, Bestor TH (2006) Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science 311:395–398. https://doi.org/10.1126/science.1120976

Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG et al (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 477:606–610. https://doi.org/10.1038/nature10443

Guibert S, Forne T, Weber M (2012) Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Res 22:633–641. https://doi.org/10.1101/gr.130997.111

Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, Surani MA (2013) Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science 339:448–452. https://doi.org/10.1126/science.1229277

Haggerty C, Kretzmer H, Riemenschneider C, Kumar AS, Mattei AL, Bailly N, Gottfreund J, Giesselmann P, Weigert R, Brandl B et al (2021) Dnmt1 has de novo activity targeted to transposable elements. Nat Struct Mol Biol 28:594–603. https://doi.org/10.1038/s41594-021-00603-8

Hahm JY, Park JW, Kang JY, Park J, Kim CH, Kim JY, Ha NC, Kim JW, Seo SB (2020) Acetylation of UHRF1 regulates hemi-methylated DNA binding and maintenance of genome-wide DNA methylation. Cell Rep 32:107958. https://doi.org/10.1016/j.celrep.2020.107958

Han M, Li J, Cao Y, Huang Y, Li W, Zhu H, Zhao Q, Han JJ, Wu Q, Li J et al (2020) A role for LSH in facilitating DNA methylation by DNMT1 through enhancing UHRF1 chromatin association. Nucleic Acids Res 48:12116–12134. https://doi.org/10.1093/nar/gkaa1003

Hanna CW, Kelsey G (2021) Features and mechanisms of canonical and noncanonical genomic imprinting. Genes Dev 35:821–834. https://doi.org/10.1101/gad.348422.121

Hanna CW, Perez-Palacios R, Gahurova L, Schubert M, Krueger F, Biggins L, Andrews S, Colome-Tatche M, Bourc'his D, Dean W, Kelsey G (2019) Endogenous retroviral insertions drive non-canonical imprinting in extra-embryonic tissues. Genome Biol 20:225. https://doi.org/10.1186/s13059-019-1833-x

Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM, Gartler SM (1999) The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci U S A 96:14412–14417

Hardikar S, Ying Z, Zeng Y, Zhao H, Liu B, Veland N, McBride K, Cheng X, Chen T (2020) The ZBTB24-CDCA7 axis regulates HELLS enrichment at centromeric satellite repeats to facilitate DNA methylation. Protein Cell 11:214–218. https://doi.org/10.1007/s13238-019-00682-w

Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature 455:826–829. https://doi.org/10.1038/nature07280

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

He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L et al (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in

mammalian DNA. Science 333:1303–1307. https://doi.org/10.1126/science.1210944

He Y, Ren J, Xu X, Ni K, Schwader A, Finney R, Wang C, Sun L, Klarmann K, Keller J et al (2020) Lsh/HELLS is required for B lymphocyte development and immunoglobulin class switch recombination. Proc Natl Acad Sci U S A 117:20100–20108. https://doi.org/10.1073/pnas.2004112117

Helfricht A, Thijssen PE, Rother MB, Shah RG, Du L, Takada S, Rogier M, Moritz J, Hanna IJ, Stoepker C et al (2020) Loss of ZBTB24 impairs nonhomologous end-joining and class-switch recombination in patients with ICF syndrome. J Exp Med 217(11):e20191688. https://doi.org/10.1084/jem.20191688

Heyn P, Logan CV, Fluteau A, Challis RC, Auchynnikava T, Martin CA, Marsh JA, Taglini F, Kilanowski F, Parry DA et al (2019) Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. Nat Genet 51:96–105. https://doi.org/10.1038/s41588-018-0274-x

Hill PWS, Leitch HG, Requena CE, Sun Z, Amouroux R, Roman-Trufero M, Borkowska M, Terragni J, Vaisvila R, Linnett S et al (2018) Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. Nature 555:392–396. https://doi.org/10.1038/nature25964

Hirasawa R, Sasaki H (2009) Dynamic transition of Dnmt3b expression in mouse pre- and early post-implantation embryos. Gene Expr Patterns 9:27–30. https://doi.org/10.1016/j.gep.2008.09.002

Hirasawa R, Chiba H, Kaneda M, Tajima S, Li E, Jaenisch R, Sasaki H (2008) Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. Genes Dev 22:1607–1616. https://doi.org/10.1101/gad.1667008

Holliday R, Pugh JE (1975) DNA modification mechanisms and gene activity during development. Science 187:226–232

Howell CY, Bestor TH, Ding F, Latham KE, Mertineit C, Trasler JM, Chaillet JR (2001) Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell 104:829–838

Hutnick LK, Golshani P, Namihira M, Xue Z, Matynia A, Yang XW, Silva AJ, Schweizer FE, Fan G (2009) DNA hypomethylation restricted to the murine forebrain induces cortical degeneration and impairs postnatal neuronal maturation. Hum Mol Genet 18:2875–2888. https://doi.org/10.1093/hmg/ddp222

Inoue A, Zhang Y (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science 334:194. https://doi.org/10.1126/science.1212483

Inoue A, Shen L, Dai Q, He C, Zhang Y (2011) Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. Cell Res 21:1670–1676. https://doi.org/10.1038/cr.2011.189

Inoue A, Jiang L, Lu F, Suzuki T, Zhang Y (2017a) Maternal H3K27me3 controls DNA methylation-independent imprinting. Nature 547:419–424. https://doi.org/10.1038/nature23262

Inoue A, Jiang L, Lu F, Zhang Y (2017b) Genomic imprinting of Xist by maternal H3K27me3. Genes Dev 31:1927–1932. https://doi.org/10.1101/gad.304113.117

Inoue A, Chen Z, Yin Q, Zhang Y (2018) Maternal Eed knockout causes loss of H3K27me3 imprinting and random X inactivation in the extraembryonic cells. Genes Dev 32:1525–1536. https://doi.org/10.1101/gad.318675.118

Ishiyama S, Nishiyama A, Saeki Y, Moritsugu K, Morimoto D, Yamaguchi L, Arai N, Matsumura R, Kawakami T, Mishima Y et al (2017) Structure of the Dnmt1 reader module complexed with a unique two-mono-ubiquitin mark on histone H3 reveals the basis for DNA methylation maintenance. Mol Cell 68(350–360):e357. https://doi.org/10.1016/j.molcel.2017.09.037

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

Jackson-Grusby L, Beard C, Possemato R, Tudor M, Fambrough D, Csankovszki G, Dausman J, Lee P, Wilson C, Lander E, Jaenisch R (2001) Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat Genet 27:31–39. https://doi.org/10.1038/83730

Jain D, Meydan C, Lange J, Claeys Bouuaert C, Lailler N, Mason CE, Anderson KV, Keeney S (2017) Rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. PLoS Genet 13:e1006964. https://doi.org/10.1371/journal.pgen.1006964

Jelinic P, Shaw P (2007) Loss of imprinting and cancer. J Pathol 211:261–268. https://doi.org/10.1002/path.2116

Jenness C, Giunta S, Muller MM, Kimura H, Muir TW, Funabiki H (2018) HELLS and CDCA7 comprise a bipartite nucleosome remodeling complex defective in ICF syndrome. Proc Natl Acad Sci U S A 115:E876–E885. https://doi.org/10.1073/pnas.1717509115

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449:248–251. https://doi.org/10.1038/nature06146

Kagiwada S, Kurimoto K, Hirota T, Yamaji M, Saitou M (2013) Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. EMBO J 32:340–353. https://doi.org/10.1038/emboj.2012.331

Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal

imprinting. Nature 429:900–903. https://doi.org/10.1038/nature02633

Kibe K, Shirane K, Ohishi H, Uemura S, Toh H, Sasaki H (2021) The DNMT3A PWWP domain is essential for the normal DNA methylation landscape in mouse somatic cells and oocytes. PLoS Genet 17:e1009570. https://doi.org/10.1371/journal.pgen.1009570

Kim SJ, Zhao H, Hardikar S, Singh AK, Goodell MA, Chen T (2013) A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. Blood 122:4086–4089. https://doi.org/10.1182/blood-2013-02-483487

Klein CJ, Botuyan MV, Wu Y, Ward CJ, Nicholson GA, Hammans S, Hojo K, Yamanishi H, Karpf AR, Wallace DC et al (2011) Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. Nat Genet 43:595–600. https://doi.org/10.1038/ng.830

Kobayashi H, Sakurai T, Miura F, Imai M, Mochiduki K, Yanagisawa E, Sakashita A, Wakai T, Suzuki Y, Ito T et al (2013) High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. Genome Res 23:616–627. https://doi.org/10.1101/gr.148023.112

Kurihara Y, Kawamura Y, Uchijima Y, Amamo T, Kobayashi H, Asano T, Kurihara H (2008) Maintenance of genomic methylation patterns during preimplantation development requires the somatic form of DNA methyltransferase 1. Dev Biol 313:335–346. https://doi.org/10.1016/j.ydbio.2007.10.033

Kurimoto K, Yabuta Y, Ohinata Y, Shigeta M, Yamanaka K, Saitou M (2008) Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev 22:1617–1635. https://doi.org/10.1101/gad.1649908

Lee PP, Fitzpatrick DR, Beard C, Jessup HK, Lehar S, Makar KW, Perez-Melgosa M, Sweetser MT, Schlissel MS, Nguyen S et al (2001) A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. Immunity 15:763–774. https://doi.org/10.1016/s1074-7613(01)00227-8

Lee CC, Peng SH, Shen L, Lee CF, Du TH, Kang ML, Xu GL, Upadhyay AK, Cheng X, Yan YT et al (2017) The role of N-alpha-acetyltransferase 10 protein in DNA methylation and genomic Imprinting. Mol Cell 68(89–103):e107. https://doi.org/10.1016/j.molcel.2017.08.025

Lees-Murdock DJ, McLoughlin GA, McDaid JR, Quinn LM, O'Doherty A, Hiripi L, Hack CJ, Walsh CP (2004) Identification of 11 pseudogenes in the DNA methyltransferase gene family in rodents and humans and implications for the functional loci. Genomics 84:193–204. https://doi.org/10.1016/j.ygeno.2004.02.004

Lei H, Oh SP, Okano M, Juttermann R, Goss KA, Jaenisch R, Li E (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development 122:3195–3205

Leonhardt H, Page AW, Weier HU, Bestor TH (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell 71:865–873

Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69:915–926

Li E, Beard C, Jaenisch R (1993) Role for DNA methylation in genomic imprinting. Nature 366:362–365. https://doi.org/10.1038/366362a0

Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, Ferguson-Smith AC (2008) A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Dev Cell 15:547–557. https://doi.org/10.1016/j.devcel.2008.08.014

Li Y, Zhang Z, Chen J, Liu W, Lai W, Liu B, Li X, Liu L, Xu S, Dong Q et al (2018) Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. Nature 564:136–140. https://doi.org/10.1038/s41586-018-0751-5

Liao J, Karnik R, Gu H, Ziller MJ, Clement K, Tsankov AM, Akopian V, Gifford CA, Donaghey J, Galonska C et al (2015) Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat Genet 47:469–478. https://doi.org/10.1038/ng.3258

Liu YJ, Nakamura T, Nakano T (2012) Essential role of DPPA3 for chromatin condensation in mouse oocytogenesis. Biol Reprod 86:40. https://doi.org/10.1095/biolreprod.111.095018

Liu X, Gao Q, Li P, Zhao Q, Zhang J, Li J, Koseki H, Wong J (2013) UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nat Commun 4:1563. https://doi.org/10.1038/ncomms2562

Looijenga LH, Gillis AJ, Verkerk AJ, van Putten WL, Oosterhuis JW (1999) Heterogeneous X inactivation in trophoblastic cells of human full-term female placentas. Am J Hum Genet 64:1445–1452. https://doi.org/10.1086/302382

Lorthongpanich C, Cheow LF, Balu S, Quake SR, Knowles BB, Burkholder WF, Solter D, Messerschmidt DM (2013) Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. Science 341:1110–1112. https://doi.org/10.1126/science.1240617

Ma P, Lin S, Bartolomei MS, Schultz RM (2010) Metastasis tumor antigen 2 (MTA2) is involved in proper imprinted expression of H19 and Peg3 during mouse preimplantation development. Biol Reprod 83:1027–1035. https://doi.org/10.1095/biolreprod.110.086397

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL (2012) Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature 487:57–63. https://doi.org/10.1038/nature11244

Mackay DJ, Callaway JL, Marks SM, White HE, Acerini CL, Boonen SE, Dayanikli P, Firth HV, Goodship JA, Haemers AP et al (2008) Hypomethylation of multiple imprinted loci in individuals with transient neonatal

diabetes is associated with mutations in ZFP57. Nat Genet 40:949–951. https://doi.org/10.1038/ng.187
Matoba S, Wang H, Jiang L, Lu F, Iwabuchi KA, Wu X, Inoue K, Yang L, Press W, Lee JT et al (2018) Loss of H3K27me3 Imprinting in somatic cell nuclear transfer embryos disrupts post-implantation development. Cell Stem Cell 23(343–354):e345. https://doi.org/10.1016/j.stem.2018.06.008
McGrath J, Solter D (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell 37:179–183. https://doi.org/10.1016/0092-8674(84)90313-1
McGraw S, Oakes CC, Martel J, Cirio MC, de Zeeuw P, Mak W, Plass C, Bartolomei MS, Chaillet JR, Trasler JM (2013) Loss of DNMT1o disrupts imprinted X chromosome inactivation and accentuates placental defects in females. PLoS Genet 9:e1003873. https://doi.org/10.1371/journal.pgen.1003873
Mertineit C, Yoder JA, Taketo T, Laird DW, Trasler JM, Bestor TH (1998) Sex-specific exons control DNA methyltransferase in mammalian germ cells. Development 125:889–897
Messerschmidt DM, de Vries W, Ito M, Solter D, Ferguson-Smith A, Knowles BB (2012) Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. Science 335:1499–1502. https://doi.org/10.1126/science.1216154
Migeon BR, Do TT (1979) In search of non-random X inactivation: studies of fetal membranes heterozygous for glucose-6-phosphate dehydrogenase. Am J Hum Genet 31:581–585
Mudbhary R, Hoshida Y, Chernyavskaya Y, Jacob V, Villanueva A, Fiel MI, Chen X, Kojima K, Thung S, Bronson RT et al (2014) UHRF1 overexpression drives DNA hypomethylation and hepatocellular carcinoma. Cancer Cell 25:196–209. https://doi.org/10.1016/j.ccr.2014.01.003
Muto M, Kanari Y, Kubo E, Takabe T, Kurihara T, Fujimori A, Tatsumi K (2002) Targeted disruption of Np95 gene renders murine embryonic stem cells hypersensitive to DNA damaging agents and DNA replication blocks. J Biol Chem 277:34549–34555. https://doi.org/10.1074/jbc.M205189200
Nakamura T, Arai Y, Umehara H, Masuhara M, Kimura T, Taniguchi H, Sekimoto T, Ikawa M, Yoneda Y, Okabe M et al (2007) PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol 9:64–71. https://doi.org/10.1038/ncb1519
Nakamura T, Liu YJ, Nakashima H, Umehara H, Inoue K, Matoba S, Tachibana M, Ogura A, Shinkai Y, Nakano T (2012) PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature 486:415–419. https://doi.org/10.1038/nature11093
Neri F, Krepelova A, Incarnato D, Maldotti M, Parlato C, Galvagni F, Matarese F, Stunnenberg HG, Oliviero S (2013) Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. Cell 155:121–134. https://doi.org/10.1016/j.cell.2013.08.056
Nielsen AL, Ortiz JA, You J, Oulad-Abdelghani M, Khechumian R, Gansmuller A, Chambon P, Losson R (1999) Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. EMBO J 18:6385–6395. https://doi.org/10.1093/emboj/18.22.6385
Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K, Shimamura S, Arita K, Kodama T, Ishikawa F et al (2013) Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature 502:249–253. https://doi.org/10.1038/nature12488
Norvil AB, AlAbdi L, Liu B, Tu YH, Forstoffer NE, Michie AR, Chen T, Gowher H (2020) The acute myeloid leukemia variant DNMT3A Arg882His is a DNMT3B-like enzyme. Nucleic Acids Res 48(7):3761–3775. https://doi.org/10.1093/nar/gkaa139
Nunez JK, Chen J, Pommier GC, Cogan JZ, Replogle JM, Adriaens C, Ramadoss GN, Shi Q, Hung KL, Samelson AJ et al (2021) Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. Cell 184(2503–2519):e2517. https://doi.org/10.1016/j.cell.2021.03.025
Okae H, Hiura H, Nishida Y, Funayama R, Tanaka S, Chiba H, Yaegashi N, Nakayama K, Sasaki H, Arima T (2012) Re-investigation and RNA sequencing-based identification of genes with placenta-specific imprinted expression. Hum Mol Genet 21:548–558. https://doi.org/10.1093/hmg/ddr488
Okae H, Matoba S, Nagashima T, Mizutani E, Inoue K, Ogonuki N, Chiba H, Funayama R, Tanaka S, Yaegashi N et al (2014) RNA sequencing-based identification of aberrant imprinting in cloned mice. Hum Mol Genet 23:992–1001. https://doi.org/10.1093/hmg/ddt495
Okano M, Xie S, Li E (1998a) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet 19:219–220. https://doi.org/10.1038/890
Okano M, Xie S, Li E (1998b) Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. Nucleic Acids Res 26:2536–2540
Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99:247–257
Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD et al (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448:714–717. https://doi.org/10.1038/nature05987
Ostler KR, Yang Q, Looney TJ, Zhang L, Vasanthakumar A, Tian Y, Kocherginsky M, Raimondi SL, DeMaio JG, Salwen HR et al (2012)

Truncated DNMT3B isoform DNMT3B7 suppresses growth, induces differentiation, and alters DNA methylation in human neuroblastoma. Cancer Res 72:4714–4723. https://doi.org/10.1158/0008-5472.CAN-12-0886

Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M (2009) Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep 10:1235–1241. https://doi.org/10.1038/embor.2009.218

Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 463: 1101–1105. https://doi.org/10.1038/nature08829

Pradhan S, Bacolla A, Wells RD, Roberts RJ (1999) Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. J Biol Chem 274:33002–33010

Qin W, Leonhardt H, Spada F (2011) Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. J Cell Biochem 112: 439–444. https://doi.org/10.1002/jcb.22998

Qin W, Wolf P, Liu N, Link S, Smets M, La Mastra F, Forne I, Pichler G, Horl D, Fellinger K et al (2015) DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. Cell Res 25:911–929. https://doi.org/10.1038/cr.2015.72

Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A, Trono D (2011) In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell 44:361–372. https://doi.org/10.1016/j.molcel.2011.08.032

Ratnam S, Mertineit C, Ding F, Howell CY, Clarke HJ, Bestor TH, Chaillet JR, Trasler JM (2002) Dynamics of Dnmt1 methyltransferase expression and intracellular localization during oogenesis and preimplantation development. Dev Biol 245:304–314. https://doi.org/10.1006/dbio.2002.0628

Reese KJ, Lin S, Verona RI, Schultz RM, Bartolomei MS (2007) Maintenance of paternal methylation and repression of the imprinted H19 gene requires MBD3. PLoS Genet 3:e137. https://doi.org/10.1371/journal.pgen.0030137

Ren R, Hardikar S, Horton JR, Lu Y, Zeng Y, Singh AK, Lin K, Coletta LD, Shen J, Lin Kong CS et al (2019) Structural basis of specific DNA binding by the transcription factor ZBTB24. Nucleic Acids Res 47:8388–8398. https://doi.org/10.1093/nar/gkz557

Rhee I, Jair KW, Yen RW, Lengauer C, Herman JG, Kinzler KW, Vogelstein B, Baylin SB, Schuebel KE (2000) CpG methylation is maintained in human cancer cells lacking DNMT1. Nature 404:1003–1007. https://doi.org/10.1038/35010000

Riggs AD (1975) X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet 14:9–25

Rothbart SB, Krajewski K, Nady N, Tempel W, Xue S, Badeaux AI, Barsyte-Lovejoy D, Martinez JY, Bedford MT, Fuchs SM et al (2012) Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat Struct Mol Biol 19: 1155–1160. https://doi.org/10.1038/nsmb.2391

Rothbart SB, Dickson BM, Ong MS, Krajewski K, Houliston S, Kireev DB, Arrowsmith CH, Strahl BD (2013) Multivalent histone engagement by the linked tandem Tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation. Genes Dev 27:1288–1298. https://doi.org/10.1101/gad.220467.113

Rottach A, Frauer C, Pichler G, Bonapace IM, Spada F, Leonhardt H (2010) The multi-domain protein Np95 connects DNA methylation and histone modification. Nucleic Acids Res 38:1796–1804. https://doi.org/10.1093/nar/gkp1152

Rouleau J, Tanigawa G, Szyf M (1992) The mouse DNA methyltransferase 5′-region. A unique housekeeping gene promoter. J Biol Chem 267:7368–7377

Russler-Germain DA, Spencer DH, Young MA, Lamprecht TL, Miller CA, Fulton R, Meyer MR, Erdmann-Gilmore P, Townsend RR, Wilson RK, Ley TJ (2014) The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. Cancer Cell 25:442–454. https://doi.org/10.1016/j.ccr.2014.02.010

Ryan RF, Schultz DC, Ayyanathan K, Singh PB, Friedman JR, Fredericks WJ, Rauscher FJ 3rd (1999) KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Kruppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. Mol Cell Biol 19:4366–4378. https://doi.org/10.1128/MCB.19.6.4366

Sado T, Fenner MH, Tan SS, Tam P, Shioda T, Li E (2000) X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. Dev Biol 225: 294–303. https://doi.org/10.1006/dbio.2000.9823

SanMiguel JM, Abramowitz LK, Bartolomei MS (2018) Imprinted gene dysregulation in a Tet1 null mouse model is stochastic and variable in the germline and offspring. Development 145(7):dev160622. https://doi.org/10.1242/dev.160622

Schneider K, Fuchs C, Dobay A, Rottach A, Qin W, Wolf P, Alvarez-Castro JM, Nalaskowski MM, Kremmer E, Schmid V et al (2013) Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. Nucleic Acids Res 41: 4860–4876. https://doi.org/10.1093/nar/gkt191

Schultz DC, Friedman JR, Rauscher FJ 3rd (2001) Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel

isoform of the mi-2alpha subunit of NuRD. Genes Dev 15:428–443. https://doi.org/10.1101/gad.869501
Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ 3rd (2002) SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev 16:919–932. https://doi.org/10.1101/gad.973302
Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W (2012) The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell 48:849–862. https://doi.org/10.1016/j.molcel.2012.11.001
Sendzikaite G, Hanna CW, Stewart-Morgan KR, Ivanova E, Kelsey G (2019) A DNMT3A PWWP mutation leads to methylation of bivalent chromatin and growth retardation in mice. Nat Commun 10: 1884. https://doi.org/10.1038/s41467-019-09713-w
Shah MY, Vasanthakumar A, Barnes NY, Figueroa ME, Kamp A, Hendrick C, Ostler KR, Davis EM, Lin S, Anastasi J et al (2010) DNMT3B7, a truncated DNMT3B isoform expressed in human tumors, disrupts embryonic development and accelerates lymphomagenesis. Cancer Res 70:5840–5850. https://doi.org/10.1158/0008-5472.CAN-10-0847
Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K et al (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 450:908–912. https://doi.org/10.1038/nature06397
Shirane K, Miura F, Ito T, Lorincz MC (2020) NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing. Nat Genet 52:1088–1098. https://doi.org/10.1038/s41588-020-0689-z
Smith EY, Futtner CR, Chamberlain SJ, Johnstone KA, Resnick JL (2011) Transcription is required to establish maternal imprinting at the Prader-Willi syndrome and Angelman syndrome locus. PLoS Genet 7: e1002422. https://doi.org/10.1371/journal.pgen.1002422
Stepper P, Kungulovski G, Jurkowska RZ, Chandra T, Krueger F, Reinhardt R, Reik W, Jeltsch A, Jurkowski TP (2017) Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. Nucleic Acids Res 45:1703–1713. https://doi.org/10.1093/nar/gkw1112
Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S (2004) DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J Biol Chem 279:27816–27823. https://doi.org/10.1074/jbc.M400181200
Surani MA, Barton SC, Norris ML (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature 308:548–550. https://doi.org/10.1038/308548a0
Takahashi N, Coluccio A, Thorball CW, Planet E, Shi H, Offner S, Turelli P, Imbeault M, Ferguson-Smith AC, Trono D (2019) ZNF445 is a primary regulator of genomic imprinting. Genes Dev 33:49–54. https://doi.org/10.1101/gad.320069.118
Takeshita K, Suetake I, Yamashita E, Suga M, Narita H, Nakagawa A, Tajima S (2011) Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proc Natl Acad Sci U S A 108:9055–9059. https://doi.org/10.1073/pnas.1019629108
Tatton-Brown K, Seal S, Ruark E, Harmer J, Ramsay E, Del Vecchio Duarte S, Zachariou A, Hanks S, O'Brien E, Aksglaede L et al (2014) Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. Nat Genet 46:385–388. https://doi.org/10.1038/ng.2917
Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RD (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 448:196–199. https://doi.org/10.1038/nature05972
Thijssen PE, Ito Y, Grillo G, Wang J, Velasco G, Nitta H, Unoki M, Yoshihara M, Suyama M, Sun Y et al (2015) Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. Nat Commun 6:7870. https://doi.org/10.1038/ncomms8870
Thompson JJ, Kaur R, Sosa CP, Lee JH, Kashiwagi K, Zhou D, Robertson KD (2018) ZBTB24 is a transcriptional regulator that coordinates with DNMT3B to control DNA methylation. Nucleic Acids Res 46: 10034–10051. https://doi.org/10.1093/nar/gky682
Tomizawa S, Sasaki H (2012) Genomic imprinting and its relevance to congenital disease, infertility, molar pregnancy and induced pluripotent stem cell. J Hum Genet 57:84–91. https://doi.org/10.1038/jhg.2011.151
Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M, Matsuoka C, Shimotohno K, Ishikawa F, Li E, Ueda HR et al (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells 11:805–814. https://doi.org/10.1111/j.1365-2443.2006.00984.x
Tucci V, Isles AR, Kelsey G, Ferguson-Smith AC, Erice Imprinting G (2019) Genomic Imprinting and physiological processes in mammals. Cell 176:952–965. https://doi.org/10.1016/j.cell.2019.01.043
Tucker KL, Talbot D, Lee MA, Leonhardt H, Jaenisch R (1996) Complementation of methylation deficiency in embryonic stem cells by a DNA methyltransferase minigene. Proc Natl Acad Sci U S A 93:12920–12925
Ueda Y, Okano M, Williams C, Chen T, Georgopoulos K, Li E (2006) Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. Development 133:1183–1192. https://doi.org/10.1242/dev.02293
Van den Wyngaert I, Sprengel J, Kass SU, Luyten WH (1998) Cloning and analysis of a novel human putative

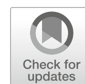
DNA methyltransferase. FEBS Lett 426:283–289. https://doi.org/10.1016/s0014-5793(98)00362-7
Veland N, Hardikar S, Zhong Y, Gayatri S, Dan J, Strahl BD, Rothbart SB, Bedford MT, Chen T (2017) The arginine methyltransferase PRMT6 regulates DNA methylation and contributes to global DNA hypomethylation in cancer. Cell Rep 21:3390–3397. https://doi.org/10.1016/j.celrep.2017.11.082
Veland N, Lu Y, Hardikar S, Gaddis S, Zeng Y, Liu B, Estecio MR, Takata Y, Lin K, Tomida MW et al (2019) DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells. Nucleic Acids Res 47:152–167. https://doi.org/10.1093/nar/gky947
Vlachogiannis G, Niederhuth CE, Tuna S, Stathopoulou A, Viiri K, de Rooij DG, Jenner RG, Schmitz RJ, Ooi SK (2015) The Dnmt3L ADD domain controls cytosine methylation establishment during spermatogenesis. Cell Rep 10(6):944–956. https://doi.org/10.1016/j.celrep.2015.01.021
Wang Q, Yu G, Ming X, Xia W, Xu X, Zhang Y, Zhang W, Li Y, Huang C, Xie H et al (2020) Imprecise DNMT1 activity coupled with neighbor-guided correction enables robust yet flexible epigenetic inheritance. Nat Genet 52:828–839. https://doi.org/10.1038/s41588-020-0661-y
Weinberg DN, Papillon-Cavanagh S, Chen H, Yue Y, Chen X, Rajagopalan KN, Horth C, McGuire JT, Xu X, Nikbakht H et al (2019) The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. Nature 573: 281–286. https://doi.org/10.1038/s41586-019-1534-3
Weinberg DN, Rosenbaum P, Chen X, Barrows D, Horth C, Marunde MR, Popova IK, Gillespie ZB, Keogh MC, Lu C et al (2021) Two competing mechanisms of DNMT3A recruitment regulate the dynamics of de novo DNA methylation at PRC1-targeted CpG islands. Nat Genet 53:794–800. https://doi.org/10.1038/s41588-021-00856-5
Williamson CM, Ball ST, Dawson C, Mehta S, Beechey CV, Fray M, Teboul L, Dear TN, Kelsey G, Peters J (2011) Uncoupling antisense-mediated silencing and DNA methylation in the imprinted Gnas cluster. PLoS Genet 7:e1001347. https://doi.org/10.1371/journal.pgen.1001347
Winkelmann J, Lin L, Schormair B, Kornum BR, Faraco J, Plazzi G, Melberg A, Cornelio F, Urban AE, Pizza F et al (2012) Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. Hum Mol Genet 21:2205–2210. https://doi.org/10.1093/hmg/dds035
Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, Arand J, Nakano T, Reik W, Walter J (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun 2:241. https://doi.org/10.1038/ncomms1240
Wu H, Thijssen PE, de Klerk E, Vonk KK, Wang J, den Hamer B, Aytekin C, van der Maarel SM, Daxinger L (2016) Converging disease genes in ICF syndrome: ZBTB24 controls expression of CDCA7 in mammals. Hum Mol Genet 25:4041–4051. https://doi.org/10.1093/hmg/ddw243
Xia W, Xu J, Yu G, Yao G, Xu K, Ma X, Zhang N, Liu B, Li T, Lin Z et al (2019) Resetting histone modifications during human parental-to-zygotic transition. Science 365:353–360. https://doi.org/10.1126/science.aaw5118
Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, Viegas-Pequignot E (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 402:187–191. https://doi.org/10.1038/46052
Xu Q, Xiang Y, Wang Q, Wang L, Brind'Amour J, Bogutz AB, Zhang Y, Zhang B, Yu G, Xia W et al (2019) SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. Nat Genet 51:844–856. https://doi.org/10.1038/s41588-019-0398-7
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 492:443–447. https://doi.org/10.1038/nature11709
Yamaguchi S, Hong K, Liu R, Inoue A, Shen L, Zhang K, Zhang Y (2013a) Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. Cell Res 23:329–339. https://doi.org/10.1038/cr.2013.22
Yamaguchi S, Shen L, Liu Y, Sendler D, Zhang Y (2013b) Role of Tet1 in erasure of genomic imprinting. Nature 504:460–464. https://doi.org/10.1038/nature12805
Yang Y, Bedford MT (2013) Protein arginine methyltransferases and cancer. Nat Rev Cancer 13: 37–50. https://doi.org/10.1038/nrc3409
Yang L, Rau R, Goodell MA (2015) DNMT3A in haematological malignancies. Nat Rev Cancer 15: 152–165. https://doi.org/10.1038/nrc3895
Yoder JA, Bestor TH (1998) A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum Mol Genet 7:279–284. https://doi.org/10.1093/hmg/7.2.279
Zalzman M, Falco G, Sharova LV, Nishiyama A, Thomas M, Lee SL, Stagg CA, Hoang HG, Yang HT, Indig FE et al (2010) Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature 464:858–863. https://doi.org/10.1038/nature08882
Zeng Y, Chen T (2019) DNA methylation reprogramming during mammalian development. Genes (Basel) 10: 257. https://doi.org/10.3390/genes10040257
Zeng TB, Han L, Pierce N, Pfeifer GP, Szabo PE (2019) EHMT2 and SETDB1 protect the maternal pronucleus from 5mC oxidation. Proc Natl Acad Sci U S A 116: 10834–10841. https://doi.org/10.1073/pnas.1819946116
Zeng Y, Ren R, Kaur G, Hardikar S, Ying Z, Babcock L, Gupta E, Zhang X, Chen T, Cheng X (2020) The inactive Dnmt3b3 isoform preferentially enhances

Dnmt3b-mediated DNA methylation. Genes Dev 34: 1546–1558. https://doi.org/10.1101/gad.341925.120

Zeng TB, Pierce N, Liao J, Szabo PE (2021) H3K9 methyltransferase EHMT2/G9a controls ERVK-driven noncanonical imprinted genes. Epigenomics 13:1299–1314. https://doi.org/10.2217/epi-2021-0168

Zhao H, Chen T (2013) Tet family of 5-methylcytosine dioxygenases in mammalian development. J Hum Genet 58:421–427. https://doi.org/10.1038/jhg.2013.63

Zuo X, Sheng J, Lau HT, McDonald CM, Andrade M, Cullen DE, Bell FT, Iacovino M, Kyba M, Xu G, Li X (2012) Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. J Biol Chem 287: 2107–2118. https://doi.org/10.1074/jbc.M111.322644



6 Structure and Mechanism of Plant DNA Methyltransferases

Sarah M. Leichter, Jiamu Du, and Xuehua Zhong

Abstract

DNA methylation is an important epigenetic mark conserved in eukaryotes from fungi to animals and plants, where it plays a crucial role in regulating gene expression and transposon silencing. Once the methylation mark is established by de novo DNA methyltransferases, specific regulatory mechanisms are required to maintain the methylation state during chromatin replication, both during meiosis and mitosis. Plant DNA methylation is found in three contexts; CG, CHG, and CHH (H = A, T, C), which are established and maintained by a unique set of DNA methyltransferases and are regulated by plant-specific pathways. DNA methylation in plants is often associated with other epigenetic modifications, such as noncoding RNA and histone modifications. This chapter focuses on the structure, function, and regulatory mechanism of plant DNA methyltransferases and their crosstalk with other epigenetic pathways.

Keywords

Chromodomain · MTase activity · CXXC domain · MTase domain · DRM1 DRM2

Abbreviations

5mC	5-Methyl-cytosine
6 mA	6-Methyl-adenine
AdoHcy	S-Adenosyl-L-homocysteine
AdoMet	S-Adenosyl-L-methionine
AGO4	ARGONAUTE 4
BAH domain	Bromo-adjacent homology domain
CMT	CHROMOMETHYLASE
DCL3	DICER-LIKE 3
DDM1	DECREASED IN DNA METHYLATION 1
DRM2	DOMAINS REARRANGED METHYLTRANSFERASE 2
ITC	Isothermal titration calorimetry
KYP	KRYPTONITE
MET1	DNA METHYLTRANSFERASE 1
MTase	Methyltransferase
PKMT	Protein lysine methyltransferase
Pol II/IV/V	RNA polymerase II/IV/V
RdDM	RNA-directed DNA methylation

S. M. Leichter · X. Zhong (✉)
Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI, USA

Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI, USA
e-mail: xuehua.zhong@wisc.edu

J. Du
Key Laboratory of Molecular Design for Plant Cell Factory of Guangdong Higher Education Institutes, Institute of Plant and Food Science, School of Life Science, Southern University of Science and Technology, Shenzhen, Guangdong, China

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_6

RDR2	RNA-DEPENDENT RNA POLYMERASE 2
RFTD	Replication foci targeting domain
SET domain	Su(var)3–9, enhancer of zeste, trithorax domain
SHH1	SAWADEE HOMEODOMAIN HOMOLOG 1
SRA domain	SET and RING finger-associated domain
SUVH	SUPPRESSOR OF VARIEGATION 3–9 HOMOLOG
TE	Transposable elements
TRD	Target recognition domain
UBA	Ubiquitin-associated domain
UHRF1	Ubiquitin-like PHD and RING finger domains 1
VIM	VARIANT IN METHYLATION
ZMET2	Zea methyltransferase 2

6.1 Introduction

DNA methylation modification refers to the addition of a methyl group to DNA. In plants, DNA methylation can occur both at the C5-position of the cytosine base (5mC) and at the N6-position of the adenine base (6 mA) (Vanyushin and Ashapkin 2011). 5mC has been extensively studied as an important part of bacterial defense systems and eukaryotic epigenetic regulation systems ranging from fungi to humans, even though several species, such as yeast and *Caenorhabditis elegans*, lack this type of epigenetic mark (Goll and Bestor 2005). In contrast, the 6 mA has been less studied in plants. Thus, in a narrower sense, DNA methylation refers to the 5mC mark, and this chapter will only focus on 5mC-related studies in higher plants. Studies from fungi, plants, and higher mammalian systems have established the conserved function of DNA methylation in gene silencing, genome imprinting, and the repression of transposable elements (TEs) and repeat sequences (Law and Jacobsen 2010; Castel and Martienssen 2013; Zhang et al. 2018a). DNA methylation requires specific enzymes, namely, DNA methyltransferases (MTases), which share a common catalytic mechanism that enables the transfer of a methyl group from the methyl donor S-adenosyl-L-methionine (AdoMet) to the 5-position of the cytosine base. Nevertheless, DNA MTases of plants and mammals differ in their sequence specificity and regulatory mechanisms. Plants have evolved distinctly from animals, and hence invoke some specific regulatory pathways controlling DNA methylation. This chapter focuses on the structure and mechanism of plant DNA MTases, thereby highlighting the unique plant DNA methylation system. As most plant DNA methylation studies are carried out using the model system *Arabidopsis thaliana*, the emphasis will be on the use of *Arabidopsis* genes to present current studies on plant DNA methylation, unless specified otherwise.

Distinct from symmetrical DNA methylation at CG dinucleotide sites that dominates in mammals, plant DNA methylation is much more complex, because it occurs symmetrically and asymmetrically in three different sequence contexts, CG, CHG (H denotes A, T, or C), and CHH, with methylation levels of about 24%, 6.7%, and 1.7% in these specific sequences in *Arabidopsis*, respectively (Cokus et al. 2008) (Fig. 6.1). In pericentromeric heterochromatin and some small patches in euchromatin regions, all three types of DNA methylation are heavily enriched at TEs and repeat sequences (Lister et al. 2008; Cokus et al. 2008; Wendte and Schmitz 2018). The heterochromatic DNA methylation has extensive internal crosstalk and highly correlates with repressive histone marks such as histone 3 lysine 9 dimethylation (H3K9me2) (Du et al. 2015; Stoddard et al. 2019). It serves to silence TEs to preserve genome integrity and to act as a genomic immune system (Law and Jacobsen 2010; Kim and Zilberman 2014; Matzke and Mosher 2014). In contrast, CG methylation is observed not only abundantly in heterochromatic regions for repression of TEs and repeat sequences, but also though to a lesser extent in euchromatic genic regions in one-third of the transcribed genes (Castel and Martienssen 2013; Law and Jacobsen 2010; Cokus et al. 2008; Lister et al. 2008; Wendte and Schmitz 2018). In

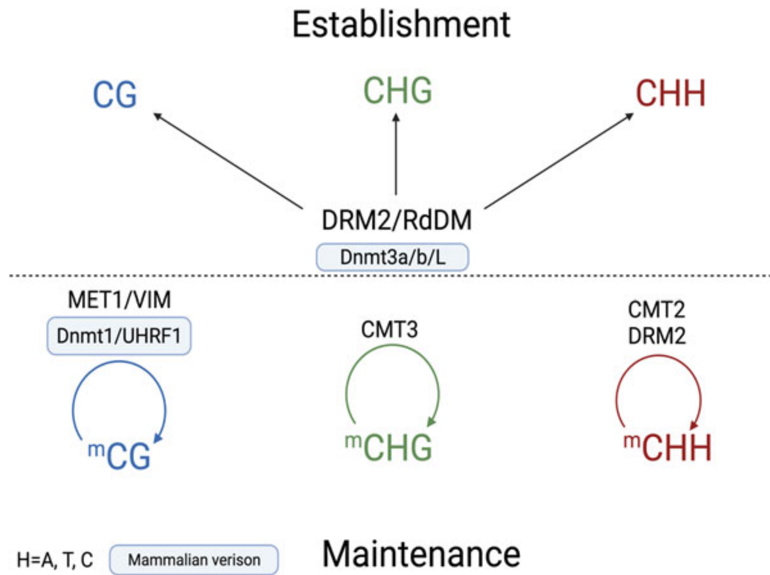


Fig. 6.1 Establishment and maintenance of plant DNA methylation and the corresponding MTases. Plants possess three types of DNA methylation patterns at CG, CHG, and CHH, all of which are established by DRM2 (an ortholog of mammalian Dnmt3a) under the control of the RNA-directed DNA methylation (RdDM) pathway. The maintenance of CG methylation requires MET1 and its cofactor VIM proteins, which are orthologs of mammalian Dnmt1 and UHRF1, respectively. CHG methylation is mainly maintained by plant-specific DNA MTase CMT3. CHH methylation is maintained through two pathways: the DRM2 pathway driven by RdDM and CMT2-mediated CHH methylation

plants, CG methylation is independent of the silencing mark H3K9me2 and shows limited crosstalk with non-CG methylation (Du et al. 2015).

All de novo methylation in plants uses a plant-specific RNA-directed DNA methylation (RdDM) pathway to guide the de novo DNA MTase DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) (Fig. 6.1). DRM2 is an ortholog of mammalian Dnmt3a but with a rearrangement of the MTase signature motifs (Cao and Jacobsen 2002a, b; Fang et al. 2021). The maintenance of different DNA methylation patterns in plants requires the employment of different pathways. CG methylation is maintained through a similar pathway as in mammals. DNA METHYLTRANSFERASE 1 (MET1), which is an ortholog of the mammalian Dnmt1, maintains plant CG methylation in cooperation with the VARIANT IN METHYLATION (VIM) proteins, which are orthologs of mammalian ubiquitin-like PHD and RING finger domains 1 (UHRF1) (Fig. 6.1) (Finnegan and Kovac 2000; Woo et al. 2008). Most CHG methylation is maintained by a self-reinforcing loop between a plant-specific CHG DNA MTase CHROMOMETHYLASE 3 (CMT3) and the H3K9 MTase KRYPTONITE [KYP, also known as SUPPRESSOR OF VARIEGATION 3–9 HOMOLOG 4 (SUVH4)] and its close homologs SUVH5 and SUVH6 (Du et al. 2012, 2014, 2015; Jackson et al. 2002; Lindroth et al. 2001). CHH methylation is asymmetric; therefore, its maintenance biochemically corresponds to de novo methylation. CHH methylation maintenance is mediated by DRM2 through RdDM (Law and Jacobsen 2010). In addition, an alternative pathway, CMT2-controlled CHH methylation, was shown to be responsible for the majority of heterochromatic CHH methylation in *Arabidopsis* (Fig. 6.1). CMT2 can form a similar self-reinforcing loop with KYP/SUVH5/SUVH6 as CMT3 (Stroud et al. 2014; Zemach et al. 2013). However, plant DNA methylation is not

a simple system where each enzyme performs its own exclusive function. There is extensive crosstalk between different methylation patterns, especially between the two types of non-CG methylation. CMT2 functions in the maintenance of CHG methylation at some loci supporting CMT3, while CMT3 also functions in CHH methylation together with CMT2 and DRM2 in a redundant manner (Stroud et al. 2014; He et al. 2021; Nozawa et al. 2021). Altogether, there are five types of functionally active DNA MTases in *Arabidopsis*: DRM2, DRM3 (Zhong et al. 2015), MET1, CMT3, and CMT2. The structural features and molecular mechanisms of each of these plant DNA MTases will be discussed in the following sections. For details about mammalian and bacterial enzymes, refer to other chapters in this book.

6.2 Structure and Mechanism of Plant DNA MTases

6.2.1 Structural Mechanism of the Maintenance of CHG Methylation in Plants

6.2.1.1 Overview of Plant CHG DNA Methylation

CMT-type MTases are evolutionarily conserved plant-specific DNA MTases that have not been identified in other species. They feature an N-terminal bromo-adjacent homology (BAH) domain and a conserved chromodomain embedded within the C-terminal MTase domain (Fig. 6.2a). *Arabidopsis* CMT3 was first identified in a forward genetic screen of suppressor genes in the hypermethylated *clark kent* mutant background, with the *cmt3* mutants showing a significant decrease of genome-wide CHG methylation levels (Lindroth et al. 2001). The chromodomain has been studied extensively as a histone methyl lysine reader module (Blus et al. 2011; Du et al. 2012; Scheid et al. 2021), providing a potential linkage between the readout of methylated histones and the establishment of DNA methylation. Shortly after the discovery of CMT3, another gene was identified to be critical for CHG methylation using the same screening system, which turned out to be a SUVH family histone-specific protein lysine methyltransferase (PKMT) and was subsequently named KRYPTONITE (KYP) (Jackson et al. 2002). The SUVH family histone PKMTs share a common architecture consisting of an N-terminal SET and RING finger-associated (SRA) domain, which can recognize methylated DNA. They have a C-terminal Su(var)3–9, enhancer of zeste, trithorax (SET) domain, which harbors the catalytic PKMT active site (Fig. 6.2b), indicating that methylated DNA may play a role in histone methylation (Johnson et al. 2007, 2008). Altogether, a self-reinforcing loop model was postulated to link H3K9me2 and DNA methylation by cycling between CMT3 and KYP (Law and Jacobsen 2010) and shown to exist by the biochemical and structural work described in the following paragraphs (Fig. 6.2c).

6.2.1.2 Structure and Mechanism of CMT3

The co-occurrence between CHG methylation and the H3K9me2 histone mark has been observed throughout the genome in plants (Bernatavichute et al. 2008), raising the possibility of the genome-wide association between CMT3 and H3K9me2. This was subsequently confirmed by both in vitro pulldown and whole-genome chromatin immunoprecipitation coupled with sequencing (ChIP-seq) experiments (Du et al. 2012). The direct binding of CMT3 to H3K9me2 was postulated, because of the existence of the chromodomain and was confirmed by in vitro peptide chip array containing hundreds of different combinations of histone modifications (Du et al. 2012). Interestingly, quantitative measurements of the binding between CMT3 or its maize equivalent Zea mays methyltransferase 2 (ZMET2), with an H3(1–15)K9me2 peptide by isothermal titration calorimetry (ITC), yielded a stoichiometry of around 2, indicating that two H3K9me2 binding sites exist on each CMT3 or ZMET2 molecules (Du et al. 2012; Stoddard et al. 2019). The ITC with individual domains of ZMET2 further revealed that the BAH and chromodomains contain these two binding sites

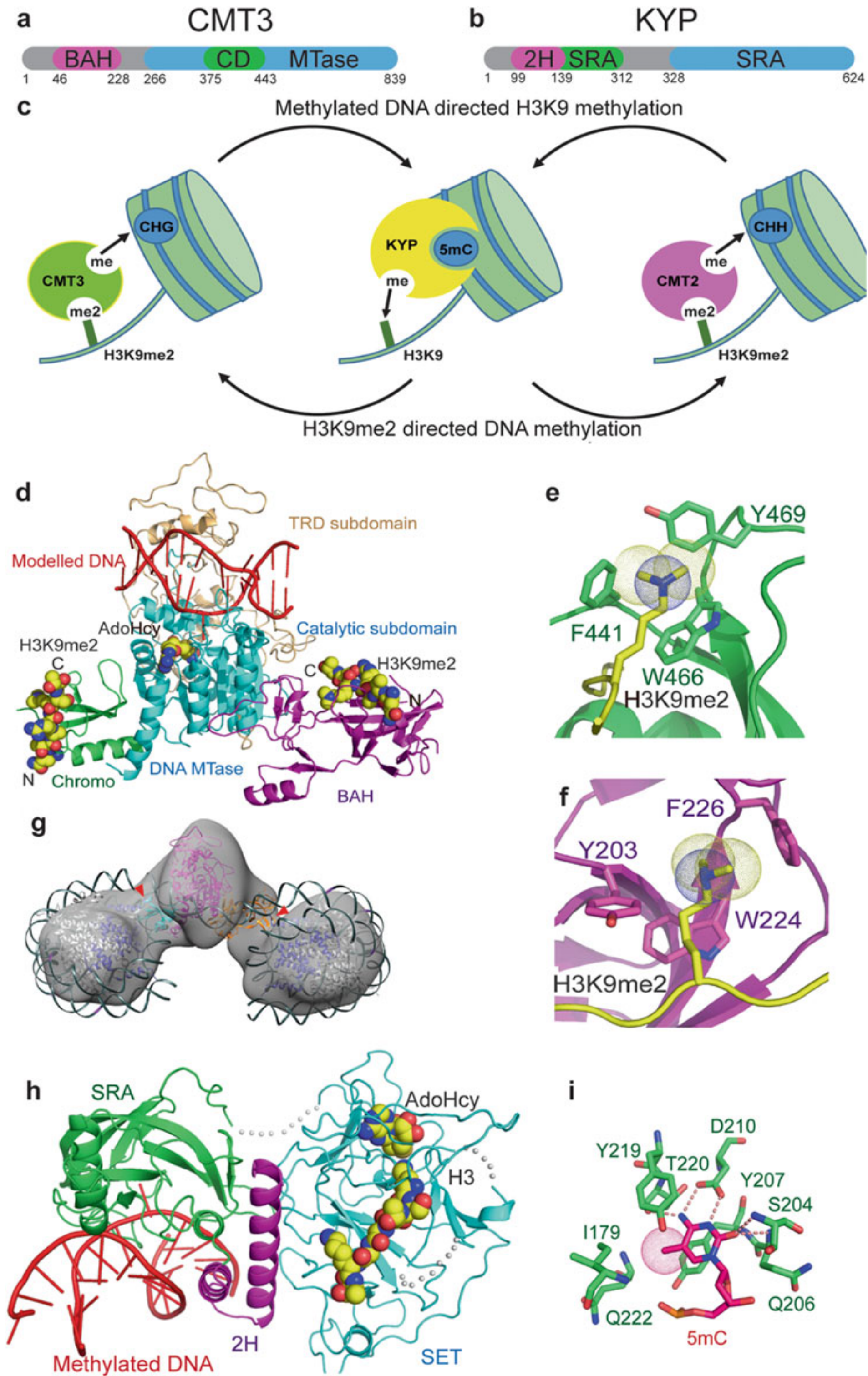


Fig. 6.2 Structure and mechanism of maintenance of CHG methylation in plants. (**a**) Schematic representation of the domain architecture of CMT3. CD, chromodomain. (**b**) Schematic representation of the domain architecture of KYP. 2H, two-helix bundle. (**c**) A schematic model showing that methylated DNA-directed H3K9

for H3K9me2 on CMT3/ZMET2 (Du et al. 2012).

Although crystallization attempts with CMT3 failed, the crystal structure of ZMET2 in complex with bound cofactor S-adenosyl-L-homocysteine (AdoHcy) was determined and provided the first structure of a plant DNA MTase (Fig. 6.2d) (Du et al. 2012). The structure of the AdoHcy-bound form of ZMET2 containing all the functional domains (BAH, chromo, and DNA MTase) adopts a triangular-shaped fold (Fig. 6.2d). The BAH and chromodomains are positioned on the two triangular-like edges of the MTase domain, despite the chromodomain being embedded inside the MTase domain in the primary sequence (Fig. 6.2a, d) (Du et al. 2012). The BAH and chromodomains, together with the target recognition domain (TRD) of the MTase domain, form the three vertices of a triangular-shaped topology, with the catalytic pocket located in the center of the triangle (Du et al. 2012). This reveals a potential regulatory mechanism of MTase activity by the surrounding domains. It is worth noting that the relative positioning of the ZMET2 BAH domain against the MTase domain resembles the relative positioning observed for the first BAH domain and the MTase domain of mouse Dnmt1, suggesting a common evolutionary origin and plausibly a similar regulatory role of the BAH domain (Du et al. 2012; Song et al. 2011). The MTase domain adopts a classical type I MTase fold composed of a catalytic subdomain and a TRD subdomain (Fig. 6.2d). The catalytic subdomain adopts the classic sandwich topology with a central seven-stranded β-sheet flanked by two layers of α-helices on either side (Fig. 6.2d), which resembles other reported structures of type I DNA MTases, such as M.HhaI, Dnmt3a, and Dnmt1 (Jia et al. 2007; Song et al. 2011; Du et al. 2012; Cheng et al. 1993). The TRD subdomain adopts a novel folding topology enriched with loops, which has never been observed in structures of other DNA MTases (Fig. 6.2d). This suggests a plausible novel DNA recognition mechanism, although the structure with bound DNA is currently lacking (Du et al. 2012). The BAH and chromodomains are firmly anchored on the two edges of the MTase domain by extensive interdomain interactions (Du et al. 2012). Two β-strands of the BAH domain form a continuous nine-stranded β-sheet with the central seven-stranded β-sheet of the MTase domain, which both stabilizes and defines the relative position of the BAH domain (Du et al. 2012). The chromodomain has several hydrophobic residues to form a hydrophobic core with the MTase domain, thereby stabilizing their relative positions (Du et al. 2012). Therefore, the BAH

Fig. 6.2 (continued) methylation by KYP can form self-reinforcing loops with H3K9me2-directed CHG methylation by CMT3 or CHH methylation by CMT2. (**d**) A structural model of ZMET2 in complex with the cofactor AdoHcy, two H3K9me2 peptides, and a modeled DNA (PDB codes: 4FSX, 4FT2, and 4FT4) (Du et al. 2012). The DNA was modeled based on PDB: 4DA4 (Song et al. 2012). The BAH domain, catalytic subdomain, TRD subdomain, chromodomain, and the modeled DNA are colored in *magenta*, *cyan*, *orange*, *green*, and *red*, respectively. The two peptides are shown in space-filling representation with their C-termini directed toward the catalytic center. (**e**) Three aromatic residues (F441, W466, and Y469) of the ZMET2 chromodomain form an aromatic cage to specifically recognize the methyl lysine of H3K9me2. (**f**) Three aromatic residues (Y203, W224, and F226) of the ZMET2 BAH domain form an aromatic cage to specifically recognize the methyl lysine of H3K9me2. (**g**) Three-dimensional reconstruction of ZMET2(130–912) (PDB:4FSX, Du et al. 2012) bridging across two nucleosomes (Stoddard et al. 2019). Red arrows indicate the emergence of the H3 tail from the globular portion of histone H3. ZMET2 (130–912), histone H3, Catalytic domain (CD) and BAH domain are colored in *magenta*, *purple*, *blue*, and *orange*, respectively. Reprinted from Molecular Cell, 73, C. I. Stoddard et al., A Nucleosome Bridging Mechanism for Activation of a Maintenance DNA Methyltransferase, 73–83, Copyright (2019), with permission from Elsevier. (**h**) Crystal structure of KYP in complex with methylated DNA, AdoHcy, and H3 peptide. The two-helix bundle (2H), SRA domain, SET domain, and the methylated DNA are colored in *magenta*, *green*, *cyan*, and *red*, respectively. The AdoHcy and H3 peptides are shown in space-filling representation. (**i**) Structural basis for the specific recognition of 5mC by a small pocket within the SRA domain

and chromodomains adopt a rigid alignment against the MTase domain and possess a potential regulatory role to modulate the MTase domain function.

Further structural studies of ZMET2-H3K9me2 complexes revealed a regulatory mechanism that directs CHG methylation in plants by the silencing mark H3K9me2 (Du et al. 2012). The crystal structure of the ZMET2-AdoHcy-H3(1–15)K9me2 complex adopts a conformation whereby the H3K9me2 peptide binds to the chromodomain (Du et al. 2012). After binding the peptide, the ZMET2 protein shows a nearly identical conformation as in its free form, revealing rigid binding without allosteric regulation. The chromodomain recognizes the peptide using a classic recognition mode in which the peptide adopts an extended β-strand-like conformation stabilized through intermolecular main chain hydrogen-bonding interactions with the chromodomain (Du et al. 2012; Blus et al. 2011). The dimethyl lysine side-chain inserts into a classic aromatic cage formed by three aromatic residues like that observed in other chromodomains where it is stabilized through both hydrophobic and cation-π interactions (Fig. 6.2e) (Du et al. 2012; Blus et al. 2011; Patel and Wang 2013). The peptide possesses a specific directionality with its C-terminus directed toward the catalytic center of the MTase domain (Fig. 6.2d), revealing a plausible mechanism that positions the catalytic center of the MTase domain toward the inner core region of the nucleosome. The chromodomain appears to be necessary for ZMET2's binding to H3K9me2 because mutations of its aromatic cage residues abolished ZMET2's binding affinity to H3K9me2 in vitro (Stoddard et al. 2019). Loss of function of the chromodomain's H3K9me2 binding affinity does not affect ZMET2's MTase activity in vitro, suggesting that the chromodomain is important for the initial binding of H3K9me2 and recruitment of ZMET2 to genomic target loci.

In the ZMET2-AdoHcy-H3(1–32)K9me2 complex, the peptide is bound to the BAH domain (Du et al. 2012) and it stimulates MTase activity in vitro (Stoddard et al. 2019). The BAH domain has been extensively studied both structurally and functionally as a histone methyl lysine reader module (Yang and Xu 2013; Kuo et al. 2012), though this is the first report of a BAH domain recognition of the H3K9me2 mark (Yang and Xu 2013; Du et al. 2012). The ZMET2 BAH domain forms fewer main chain interactions with the H3K9me2 peptide when compared to the chromodomain. However, it also contains an aromatic cage formed by three aromatic residues which accommodate the dimethyl lysine side chain through hydrophobic and cation-π interactions, similar to other methyl lysine reader modules (Fig. 6.2f) (Du et al. 2012; Patel and Wang 2013). Interestingly, the BAH domain-bound peptide has the same directionality as that observed for the chromodomain-bound peptide. Its C-terminus is directed toward the catalytic center of the MTase domain (Fig. 6.2d), revealing a similar regulation mechanism as the chromodomain, thereby positioning the MTase domain toward the inner core region of the nucleosome. The BAH domain does not appear to be necessary for H3K9me2 binding, because disruption of BAH-H3K9me2 binding does not abolish ZMET2's ability to bind H3K9me2, but severely decreases its MTase activity (Stoddard et al. 2019). This suggests that the chromodomain is important for the initial binding to H3K9me2 and the BAH domain aids in enzymatic catalysis. Interestingly, loss-of-function of either the BAH or chromodomains of CMT3 fails to complement the in vivo function of CMT3 (Du et al. 2012) suggesting that the H3K9me2 recognition capacity of both domains is essential for its in vivo function.

Considering ZMET2 has two H3K9me2 binding domains, ZMET2 has been proposed to either span the diameter of a single nucleosome or bind to two adjacent H3K9me2 containing nucleosomes (Du et al. 2012). In vitro MTase activity assays performed on mononucleosomes and dinucleosomes containing the H3K9me2 showed that ZMET2 has greater activity on dinucleosomes than mononucleosomes (Stoddard et al. 2019). With a preference for dinucleosome substrates (Fig. 6.2g), it was found that the ideal spacing of the nucleosomes is 30 bp and that

ZMET2 preferentially methylates linker DNA. This is consistent with the average nucleosome spacing in *Arabidopsis* heterochromatin (Chodavarapu et al. 2010). The positioning of the CHG sites is also an important factor for activity. ZMET2 shows markedly higher activity on a CHG site in the center of the linker DNA than on a CHG site located on either end of the linker DNA (Stoddard et al. 2019).

In this model, CMT3 preferentially methylates linker DNA between two adjacent H3K9me2-containing nucleosomes (Stoddard et al. 2019). The chromodomain of CMT3 must first bind to one H3K9me2 peptide and subsequently, the BAH domain binds to H3K9me2 on the second nucleosome to facilitate the catalytic activity of CMT3. CHG methylation is highly enriched in heterochromatic, transposon-rich regions of the plant genome. Working in concert with the silencing histone mark H3K9me2, CMT3 ensures the silencing of transposons in plants. The requirement of H3K9me2 for CMT3 methylation activity ensures proper targeting of CHG methylation to heterochromatic regions and does not lead to aberrant silencing of other CHG sites.

Through in vitro MTase activity assays (Du et al. 2012; Stoddard et al. 2019) and in vivo knockout studies (Stroud et al. 2013), it has been shown that CMT3 is responsible for methylation on CHG substrates, which are comprised of three contexts: CAG, CTG, and CCG. CMT3 prefers CAG and CTG (referred to as CWG) sites over CCG sites (Cokus et al. 2008; Gouil and Baulcombe 2016) and CWG methylation level is greater than that of CCG methylation in *Arabidopsis*. The preference for methylation of CWG sites has been observed in *Arabidopsis*, tomato, rice, and maize (Gouil and Baulcombe 2016). It has been noted that specific loss of CCG methylation occurs in *met1* mutants (Yaari et al. 2015). This disparity is not due to fewer CCG sites being methylated compared to CWG sites, but rather an overall lower methylation level at CCG sites (Gouil and Baulcombe 2016). This indicates that CCG methylation may be regulated by both MET1 and CMT3. The evidence supporting this notion is that the external cytosine residues in CCG sites can only be methylated when internal cytosines are methylated (Zabet et al. 2017). Interestingly, in *cmt3* mutants, all three contexts of CHG methylation are equally reduced (Gouil and Baulcombe 2016), indicating a mechanism beyond CMT3 that may be regulating its context preference.

While both *Arabidopsis* and maize only encode one CMT3 protein, rice (*Oryza sativa*) encodes two CMT3 proteins, OsCMT3a and OsCMT3b (Cheng et al. 2015). Loss of function of OsCMT3a leads to multiple defects, such as the activation of multiple transposons and sterility (Hu et al. 2021), suggesting that CHG methylation plays an important role in development and reproduction. Interestingly, unlike the *Arabidopsis* triple mutant *drm2 cmt2 cmt3* (Stroud et al. 2014), the triple mutant of rice *Osdrm2 cmt2 cmt3a* does not eliminate all non-CG methylation but retains non-CG methylation in high GC-rich regions of the genome (Hu et al. 2021). Simultaneous loss-of-function of CMT3b leads to the complete loss of non-CG methylation within GC-rich regions, indicating that OsCMT3a and OsCMT3b have non-overlapping roles to maintain non-CG methylation within different regions of the genome.

Although CMT3 typically acts within the H3K9me2-heterochromatic parts of the genome (Du et al. 2012; Stroud et al. 2013), CMT3 has been associated with the presence of gene body methylation as well. Genome-wide methylation analysis of multiple angiosperms showed that two species (*Conringia planisiliqua* and *Eutrema salsugineum*) that do not encode CMT3 also lack gene body methylation (Bewick et al. 2016, 2017). This shows an interesting connection between CMT3 and gene body methylation that had previously not been reported. Furthermore, the ectopic expression of CMT3 led to the gain of gene body methylation in *Eutrema salsugineum* (Wendte et al. 2019). This observation indicates that CMT3 can perform de novo DNA methylation, a new function for this enzyme.

6.2.1.3 Structure and Mechanism of KRYPTONITE

Since CHG methylation in plants solely depends on H3K9me2, the maintenance of H3K9me2 is

necessary for proper maintenance of CHG methylation during mitosis and meiosis. Thus, it is indispensable to study H3K9 methylation during an investigation of CHG methylation in plants. In *Arabidopsis*, H3K9 methylation relies on the SUVH family H3K9 MTase KYP and its homologs SUVH5 and SUVH6. A *kyp suvh5 suvh6* triple mutant strain which has lost most of the H3K9me2 mark on a genome-wide scale showed a significant loss of CHG methylation similar to the *cmt3* mutant strain (Stroud et al. 2013), confirming that CHG methylation is controlled by H3K9 MTases. The crystal structure of KYP in complex with its cofactor product AdoHcy, DNA containing a methylated CHH or CHG site (mCHH/mCHG), and the substrate H3 (1–15) peptide highlighted how the H3K9 MTase is regulated by methylated DNA (Fig. 6.2g) (Du et al. 2014). KYP contains an N-terminal two-helix bundle, a middle SRA domain that can recognize methylated CHH or CHG DNA (Johnson et al. 2007), and a C-terminal SET domain (including pre-SET, SET, and post-SET subdomains) which confers the H3K9 MTase activity (Jackson et al. 2004). The two-helix bundle is located within the middle of the structure holding together the whole architecture of the protein (Du et al. 2014). The SRA domain is aligned on one side of the two-helix bundle, while the SET domain is aligned on the other side (Fig. 6.2g) (Du et al. 2014). There are extensive interactions between the three segments of the structure, suggesting a rigid alignment of the entire topology (Du et al. 2014). The KYP SRA domain possesses a positively charged surface cleft that holds the bound DNA, similar to structures of other reported DNA-SRA domain complexes (Hashimoto et al. 2008; Arita et al. 2008; Avvakumov et al. 2008; Rajakumara et al. 2011). In addition, positively charged residues of the two-helix bundle are involved in the recognition of the backbone of the methylated DNA (Du et al. 2014). The 5mC base is flipped out of the DNA duplex and inserted into a small pocket of the SRA domain, thereby forming extensive intermolecular interactions with the surrounding residues (Fig. 6.2h) (Du et al. 2014). Mutations of the 5mC-binding pocket residues led to a loss of the capacity to bind methylated DNA while retaining the in vitro histone MTase activity. Interestingly, these mutations also impaired the in vivo function of KYP, suggesting that specific binding to methylated DNA is required for the proper targeting of KYP (Du et al. 2014). The cofactor product AdoHcy and the substrate peptide are bound in between the SET and post-SET subdomains (Du et al. 2014). Several important tyrosine residues are positioned around the target lysine and form hydrogen-bonding interactions that can facilitate the catalytic reaction and restrict the enzyme to be an H3K9me2 MTase (Du et al. 2014). Overall, the structures of the KYP-mCHH/CHG DNA-AdoHcy-H3 peptide complex and the DNA-free form of the KYP homolog SUVH9 are nearly identical, indicating that neither conformational change nor allosteric regulation occurs upon DNA binding (Du et al. 2014; Johnson et al. 2014). Therefore, the KYP protein is recruited by the methylated CHH or CHG DNA and then methylates nearby histone tails.

The structures of the SRA domain of SUVH5 (Rajakumara et al. 2011) and full-length protein of SUVH6 (Li et al. 2018) showed strong homology to KYP. SUVH5 and SUVH6 both contain a two-helix bundle, SRA, and SET domains that aid in binding methylated non-CG sites and MTase activity. It has been hypothesized that each H3K9 methyltransferase prefers specific DNA substrates (Gouil and Baulcombe 2016), which may help explain the patterning of CHG methylation. KYP shows the strongest binding to CAG sites, while SUVH5 shows the greatest binding to CCG sites in vitro. Interestingly, the SRA domain of SUVH5 has an affinity for both methylated CG sites and hydroxymethylated CG sites in vitro (Rajakumara et al. 2016). Therefore, some of the observed sequence bias of CMT3 for CWG sites can be attributed to the histone methyltransferase establishing H3K9me2 bindings sites (Wendte and Schmitz 2018).

In summary, CMT3 uses its BAH and chromodomains to bind two adjacent H3K9me2-containing nucleosomes and target its DNA MTase domain to achieve H3K9me2-directed CHG DNA methylation. Conversely, KYP uses its SRA domain to target its histone MTase

domain to mCHG/mCHH-containing DNA to carry out the methylated DNA-directed H3K9 methylation. Taken together, the two proteins form a simple self-reinforcing feedback loop to strengthen H3K9 and CHG methylation at heterochromatin to maintain the silenced state of TEs and repeat sequences.

6.2.2 Mechanism of CMT2-Mediated CHH Methylation

CMT2 was first identified as a downstream effector of the chromatin remodeler DECREASED IN DNA METHYLATION 1 (DDM1), mediating CHH methylation of TEs independently of the classical RdDM-dependent CHH methylation pathway (Zemach et al. 2013). Further biochemical characterization of CMT2 revealed that this MTase can de novo methylate both CHH and CHG sites in vitro, in contrast to CMT3 which prefers to maintain CHG site methylation (Stroud et al. 2014). A *cmt2 cmt3* double mutant showed a stronger loss of CHG methylation than the *cmt3* mutant alone, confirming that the in vitro CHG methylation activity of CMT2 is also functionally relevant in vivo (Stroud et al. 2014). In addition, the *drm1 drm2 cmt2* triple mutant can nearly eliminate CHH methylation genome-wide, while the *drm1 drm2* double mutant or *cmt2* single mutant can only partially reduce CHH methylation, indicating that the RdDM pathway and CMT2 are complementary in controlling almost all the CHH methylation (Stroud et al. 2014). Like CMT3, the *kyp suvh5 suvh6* triple mutant, which can eliminate H3K9me2, shows 86% loss of CHH methylation controlled by CMT2, revealing a connection between H3K9me2 and CMT2 (Stroud et al. 2014). CMT2 contains a BAH domain and C-terminal chromodomain inserted within the MTase domain, with a similar domain architecture as CMT3. Thus, it is reasonable to speculate that CMT2 acts in a similar way as CMT3. ITC binding experiments established that CMT2 can directly bind H3K9me with a preference for H3K9me2 and reduced binding of H3K9me1 and H3K9me3, which is consistent with the observation that CMT2-controlled targets have a higher H3K9me2 level than H3K9me1 (Stroud et al. 2014). The ITC binding yields a protein-to-peptide molar ratio of 1:2, establishing that the dual recognition mode observed for CMT3 is also conserved for CMT2 (Stroud et al. 2014). Due to the high-sequence homology between CMT3 and CMT2 (46% identity over the functional BAH, chromo, and MTase domains), it is speculated that CMT2 can use its BAH and chromodomains to target the MTase domain to H3K9me2-containing nucleosomes to achieve position-specific CHH methylation, forming a self-reinforcing loop with KYP, similar to that observed for CMT3 (Fig. 6.2c). CMT2 has an extra-long N-terminal extension of about 500 residues without any homology to known domains or predictable secondary structure. It remains to be determined whether this N-terminal segment has any functional role within CMT2. Unlike CMT3, which exists in almost all plant species, CMT2 exists in some but not all angiosperm plant species. For example, maize does not have a homologous gene of CMT2, while rice possesses a single CMT2 ortholog (Zemach et al. 2013). Therefore, CMT2-mediated CHH methylation appears not to be a strictly conserved plant DNA methylation pathway, in contrast to DRM2 driven by RdDM. Moreover, the CMT2 pathway is functionally and partially redundant with RdDM. Thus, CMT2 may function in an RdDM redundant way in some plant species, while in others, it could have been superseded by RdDM during evolution. Functionally, CMT2 was identified to associate with the heat response by comparison with the genome of different ecotypes of *Arabidopsis* (Shen et al. 2014). The *cmt2* mutant has a stronger tolerance to heat stress, indicating an epigenetic basis for the adaptation to environmental stress (Shen et al. 2014). Furthermore, in the *cmt3* mutant, more CMT2 binding is observed on the heat-sensitive transposon *ONSEN* than in the wild type (Nozawa et al. 2021). Consistently, there is an observed ectopic gain of CHH methylation on *ONSEN* in the *cmt3* mutant, but CHH methylation is almost lost in *cmt2*, indicating that CMT2 is a major DNA methyltransferase targeting *ONSEN*. Under heat stress, the H3K9me2 level is reduced

in wild-type plants, but maintained at high level in *cmt3*, likely through the H3K9me2-DNA methylation feedback loop. The higher CHH methylation and H3K9me2 levels are associated with lower *ONSEN* transcription in *cmt3* under heat (Nozawa et al. 2021).

Like CMT3-mediated DNA methylation, CMT2-mediated DNA methylation is also seen in a specific pattern. CMT2-dependent CHH methylation is primarily found in the CWA context (W = A or T), which is significant because CHH methylation can be composed of nine different contexts (Gouil and Baulcombe 2016). CCC and CCT are the lowest methylated contexts seen in *Arabidopsis*, while CAA and CTA are the highest. The CHH subcontext specificity is most pronounced in the pericentric heterochromatin regions of the *Arabidopsis* genome, enriched with KYP/SUVH5/SUVH6 dependent H3K9me2 marks. The histone MTases play a role in context-specific non-CG methylation; there appears to be a multi-layered regulation system within the heterochromatin that involves CMTs, H3K9me2, and DNA sequence.

6.2.3 RNA-Directed DNA Methylation (RdDM)

6.2.3.1 Overview of RdDM

RdDM is a plant-specific de novo DNA methylation pathway, which conducts all the de novo DNA methylation in plants and is required for the maintenance of CHH methylation within small euchromatic TEs (He et al. 2014; Matzke and Mosher 2014; Zhao and Chen 2014). The pathway starts with a plant-specific RNA polymerase IV (Pol IV), which is a specialized RNA polymerase that evolved from Pol II. Pol IV is targeted to certain loci to generate single-stranded RNA transcripts (Fig. 6.3a) (Haag and Pikaard 2011). The RNA transcripts are subsequently used as templates by Pol IV-associated RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) to generate double-stranded RNA (Fig. 6.3a) (Haag et al. 2012). The double-stranded RNA is then cut into 24 bp siRNAs by DICER-LIKE 3 (DCL3), and one 24-nt strand is loaded into ARGONAUTE 4 (AGO4) (Fig. 6.3a) (Law and Jacobsen 2010; Matzke and Mosher 2014). The structure of DCL3 in complex with pre-siRNA shows the preference for 24 nt siRNA containing an adenine in the 5′ position of the guide strand, which has also been observed for AGO4's binding preference in vivo (Wang et al. 2021a; Zhai et al. 2015). Meanwhile, another plant-specific RNA polymerase Pol V can be targeted to certain loci and produce long noncoding scaffold RNA transcript. Pol V transcripts are dependent on the action of a chromatin remodeling complex, which is composed of three proteins: DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), DEFECTIVE IN MERISTEM SILENCING 3 (DMS3), and RNA DIRECTED DNA METHYLATION1 (RDM1) (Law et al. 2010), referred to as the DDR complex. The AGO4-bound siRNA can interact with Pol V transcripts by base-pairing as well as protein interaction and further target the plant-specific DNA MTase DRM2 to facilitate site-specific DNA methylation (Fig. 6.3a) (Law and Jacobsen 2010; Matzke and Mosher 2014). Thus, in this pathway, DNA methylation sites on the chromatin are determined by the targeting of Pol IV, Pol V, and DRM2. Studies have established additional mechanisms in which SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1, also called DNA-binding transcription factor 1, DTF1) can target Pol IV to H3K9me2 containing chromatin regions, and the catalytically inactive SUVH family proteins SUVH2/9 can direct Pol V to methylated DNA-containing loci (Fig. 6.3a) (Law et al. 2013; Johnson et al. 2014; Zhang et al. 2013; Liu et al. 2014). The *Arabidopsis* Pol V-binding protein RDM15 (Niu et al. 2021) and maize Pol V-binding protein SHH2 (Haag et al. 2014; Wang et al. 2021b) were reported to specifically recognize H3K4me1 and H3K9me1, respectively, suggesting that the targeting of Pol V is influenced by diverse histone marks. The hierarchy of action of each RdDM component was determined by fusing each component to a zinc-finger endonuclease and testing their abilities to recruit the rest of the RdDM pathway to de novo

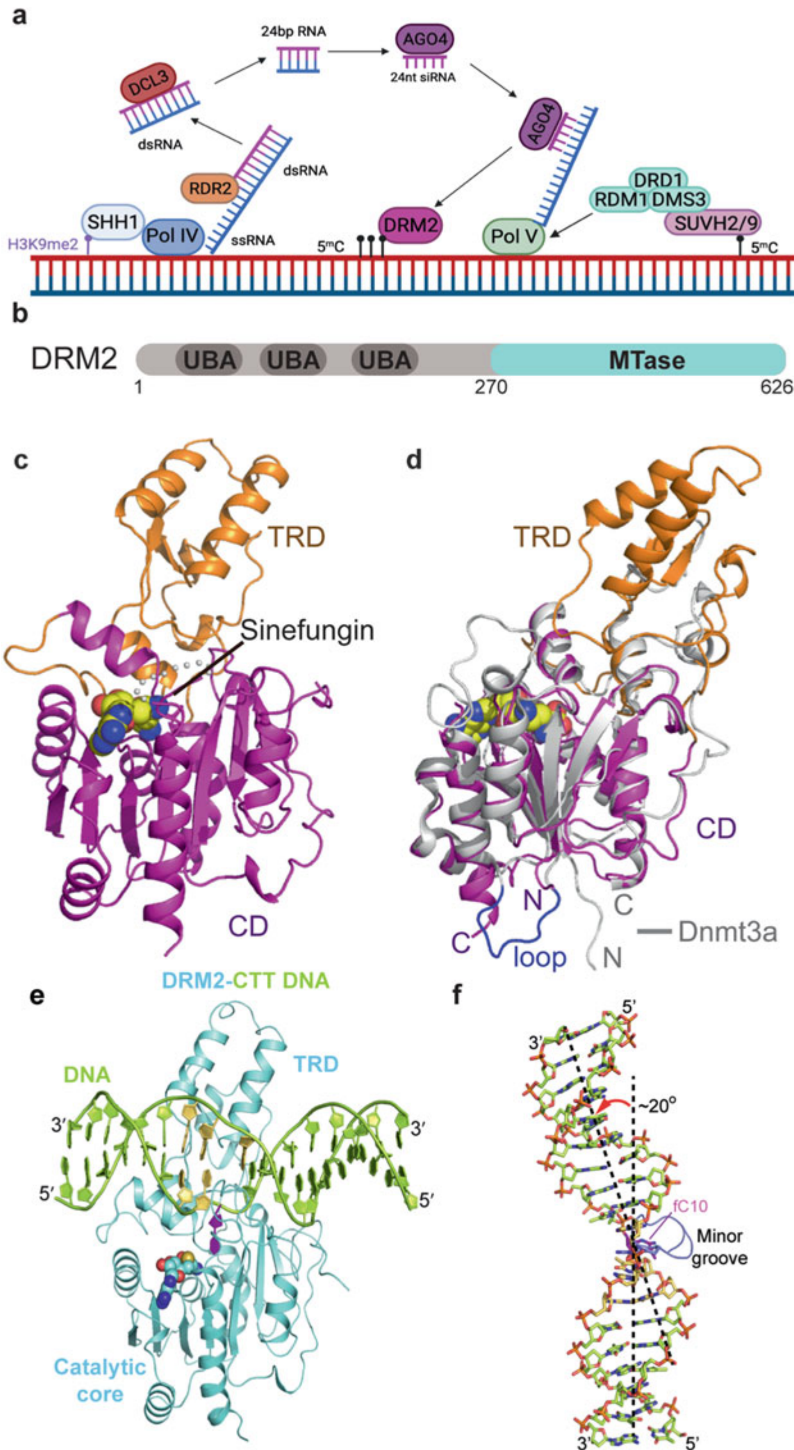


Fig. 6.3 Structure and mechanism of DRM2 function driven by RdDM. (**a**) Schematic model of RdDM pathway. Pol IV is targeted by SHH2 to H3K9me2-containing loci to produce ssRNA transcripts. Using the ssRNA as a template dsRNA is generated by Pol IV-associated RDR2 and cleaved by DCL3 to generate 24-bp dsRNA. The 24-nt siRNA is then loaded onto AGO4. SUVH2/SUVH9 binds to methylated cytosines and recruits the DDR complex (DRD1, DMS3, RDM1), which recruits Pol V. Pol V generates a scaffold noncoding RNA

methylate a site. It was found that the components of the DDR complex can robustly recruit other RdDM components to a target site for methylation (Gallego-Bartolomé et al. 2019). Co-IP experiments indicated that DRM2 occurs in the same complex with AGO4, suggesting that AGO4 plays an important role in the targeting of DRM2 (Zhong et al. 2014). However, the molecular mechanism of the AGO4-DRM2 interaction still remains unclear.

6.2.3.2 Structure and Mechanism of DRM2

Arabidopsis DRM2 was identified as the de novo MTase in plants in a genetic screen and showed sequence conservation with the mammalian de novo MTase Dnmt3a (Cao and Jacobsen 2002a, b). From a sequence perspective, DRM family proteins have N-terminal 2–3 ubiquitin-associated domains (UBAs) and a C-terminal DNA MTase domain, which possesses all the signature motifs of type I DNA MTase, except that an alignment revealed a circular permutation of these motifs (Fig. 6.3b) (Cao and Jacobsen 2002a). This type of rearrangement only exists in plants and is considered a plant-specific feature. The crystal structure of the MTase domain of *Nicotiana tabacum* (NtDRM) shows a classic type I DNA MTase fold with a central seven-stranded β-sheet flanked by two layers of α-helices positioned on both sides (Fig. 6.3c) (Zhong et al. 2014). The TRD subdomain of NtDRM is composed of a two-stranded antiparallel β-sheet and two antiparallel α-helices, which are different from structures of other reported DNA MTases, indicative of a unique DNA recognition and selection mode (Fig. 6.3c) (Zhong et al. 2014). The catalytic subdomain resembles other DNA MTase domains, such as the Dnmt3a MTase domain (Jia et al. 2007; Zhong et al. 2014). In the Dnmt3a MTase domain crystal structure, the N- and C-termini are close to each other (Jia et al. 2007). If the N- and C-termini of Dnmt3a MTase were fused together and a break was incorporated at the loop between Pro739 and Pro746 (highlighted in blue in Fig. 6.3d) (Jia et al. 2007), the resulting sequence-based folding topology would be identical to that of the NtDRM MTase domain. As a result, the nearly identical folding topology between NtDRM and other type I DNA MTases indicates both a common catalytic mechanism and a common evolutionary ancestor.

The NtDRM MTase domain forms a homodimeric arrangement as observed both in solution and from packing alignments in the crystal (Fig. 6.3e). The dimer interface is composed of a hydrophobic core together with several pairs of intermolecular salt bridges (Zhong et al. 2014). It is interesting that the NtDRM homodimer interface mimics the mammalian Dnmt3a-Dnmt3L heterodimer interface (Fig. 6.3e) (Jia et al. 2007; Zhong et al. 2014). Although *Arabidopsis* has an inactive DRM homolog DRM3, which is essential for the establishment of DNA methylation by RdDM (Henderson et al. 2010), DRM3 appears not to engage in interactions with DRM2, given that DRM3 cannot be detected in immunoprecipitation followed by mass spectrometry assays using epitope-tagged DRM2 (Zhong et al. 2015). In the *drm3* mutant, Pol V occupancy extends to additional loci, but the Pol V transcript

Fig. 6.3 (continued) transcript that can interact with the AGO4-bound siRNA to further direct DRM2 to perform de novo DNA methylation. (**b**) Domain architecture of *Arabidopsis* DRM2. (**c**) Crystal structure of the catalytic domain of NtDRM with the catalytic subdomain (CD) and TRD subdomain colored in *magenta* and *orange*, respectively (PDB code: 4ONJ) (Zhong et al. 2014). (**d**) Superposition of the NtDRM catalytic domain with the Dnmt3a catalytic domain, which is colored in *silver* (PDB codes: 4ONJ and 2QRV) (Zhong et al. 2014; Jia et al. 2007). If the N and C-termini of Dnmt3a are fused together and a break introduced within the loop marked in blue, the overall topology of Dnmt3a becomes similar to DRM. (**e**) Structure of the MTase domain of *Arabidopsis* DRM2 in complex with CTT DNA (PDB code: 7L4C) (Fang et al. 2021). (**f**) Model of the deformed structure of DNA upon DRM2 binding. Panels e and f were modified from Fang et al., *Sci. Adv.* 2021; 7: eabd9224. © The Authors, some rights reserved, exclusive licensee AAAS. Distributed under a CC BY-NC 4.0 License (http://creativecommons.org/licenses/by-nc/4.0/)

abundance is reduced, revealing a potential role of DRM3 in the stabilization of Pol V and/or mediating Pol V transcript elongation (Zhong et al. 2015). Thus, DRM3 in plants appears not to function like Dnmt3L by acting as a cofactor of the active DNA MTase, but likely functions in regulating RdDM by association with Pol V (Zhong et al. 2015). The disruption of the dimerization interface by five mutations of NtDRM results in perturbation of its enzymatic activity, indicating that dimerization is biochemically essential (Zhong et al. 2014). A plausible explanation of the homodimerization of DRM MTase domains is that dimerization could help to maintain the conformation of the catalytic loop, because the C-terminal portion of the catalytic loop is involved in the dimerization interface formation. The mammalian de novo DNA MTase requires a Dnmt3L-Dnmt3a-Dnmt3a-Dnmt3L tetrameric arrangement (Jia et al. 2007). Although the DRM homodimer mimics the Dnmt3a-Dnmt3L dimer interface, there is no interface in DRM that corresponds to the Dnmt3a-Dnmt3a interface. The Dnmt3a-Dnmt3a interface is essential for its activity and is believed to double the DNA-binding surface to overcome the shortage of the DNA-binding surface, as a result of the small TRD subdomain of Dnmt3a (Jia et al. 2007; Jurkowska et al. 2008). The TRD subdomain of DRM is bigger than that of Dnmt3a and can form a continuous large negatively charged surface, which can access the DNA substrate (Zhong et al. 2014). Thus, the Dnmt3a-Dnmt3a-like interface appears not to be required in the DRM case.

Although DRM2 shares structural homology with the mammalian Dnmta3, the two methyltransferases have distinct substrate preferences. Mammalian Dnmt3a primarily acts upon CpG substrates (Zhang et al. 2018b; Gao et al. 2020), while DRM2 primarily acts upon CHH substrates (Zhong et al. 2014). To dissect the substrate preference difference, the structure of *Arabidopsis* DRM2 was solved in complex with a representative CHH DNA substrate (Fang et al. 2021) (Fig. 6.3e). The interaction between DRM2 and its DNA substrate spans 13 bps involving contacts between the major and minor grooves. The major groove is contacted by a loop-helix-helix (LHH) motif and a helix from the target recognition domain, while the minor groove is contacted by a subset of residues within the catalytic loop and a loop containing a rearranged motif IV. Upon binding to the DNA, residues from the catalytic core of the enzyme intercalate with the non-target DNA strand, resulting in a large DNA deformation at the major groove of the DNA strand (Fig. 6.3f). The target recognition domain of DRM2 then stabilizes the deformed major groove via shape complementarity rather than the base-specific mechanism observed in Dnmt3a (Zhang et al. 2018b). This is an unprecedented substrate recognition mechanism, which allows for DRM2 to have the capabilities to methylate a broad range of DNA substrates in comparison to its mammalian homolog, Dnmt3a. Notably, the substrate preference DRM2 can be engineered to shift toward a CHG context through a single point mutation in the target recognition domain of the enzyme (Fang et al. 2021).

The UBA domains of DRM have been shown to be critical for its in vivo but not in vitro DNA MTase activity, revealing an important regulatory but not catalytic role (Henderson et al. 2010; Zhong et al. 2014). So far, it is still unclear how the UBA domains participate in DRM function. A report revealed that the UBA domains of rice OsDRM2 can interact with the ATP-dependent RNA helicase eIF4A, which links RNA higher structure to DRM (Dangwal et al. 2013). Recently, UBA domains of *Arabidopsis* DRM2 interact with the UVB photoreceptor UVR8 in a UVB-promoted manner (Jiang et al. 2021). Another plausible connection might be in the recognition of some ubiquitination modifications by the UBA domains, which is a common function of UBA domains. Supporting this notion, DRM2 was found to interact with CFK1, an F-box E3 ligase in the SCF complex that regulates DRM2 protein stability via the 26S proteasome degradation pathway (Chen et al. 2021). Further investigation is required to directly understand

the role of UBA domains in regulating DRM2 function.

6.2.4 Potential Mechanism of MET1 in CG Methylation Maintenance

Unlike the extensive crosstalk between non-CG methylation and H3K9me2, CG methylation seems to be independent of H3K9me2 (Du et al. 2015). In the *kyp suvh5 suvh6* triple mutant, which eliminates most of the H3K9me2, only a limited reduction in CG methylation is observed (Stroud et al. 2013). Once established, CG methylation is subsequently maintained by MET1 (Kankel et al. 2003) with the aid of VIM proteins. MET1 is an ortholog of mammalian Dnmt1 with several distinct features in its sequence. By comparing the sequence motifs of mouse Dnmt1, whose structure has been extensively studied, and *Arabidopsis* MET1 (Fig. 6.4a), it becomes apparent that the two C-terminal BAH domains (BAH1 and BAH2) and the DNA MTase domains exhibit similarities with a sequence identity of 36%. The high sequence homology between the catalytic portion of MET1 and Dnmt1 suggests that MET1 may share a similar type I DNA MTase domain fold and a common catalytic mechanism as Dnmt1. The BAH2 domain of Dnmt1 has a long loop extending outward whose tip contacts the TRD subdomain of the MTase fold (Fig. 6.4b) (Song et al. 2011), suggestive of a regulatory role for substrate DNA binding through adjustment of the TRD position by the BAH2 domain. Similarly, the BAH1 domain of mammalian Dnmt1 has been shown to bind H4K20me3 (Ren et al. 2021). However, neither the interacting region on the TRD (highlighted in red in Fig. 6.4b) nor the BAH2 loop (highlighted in blue in Fig. 6.4b) of Dnmt1 is conserved in MET1 (Fig. 6.4c, d), indicating that MET1 may lack this type of BAH domain-mediated regulation of the TRD subdomain. Toward the N-terminus, Dnmt1 has a CXXC domain (highlighted in yellow in Fig. 6.4b), which can specifically recognize unmethylated CG sites and subsequently position a loop (highlighted in green in Fig. 6.4b) between the DNA and the active site of Dnmt1 to block the MTase activity (Song et al. 2011). This auto-inhibition mechanism can protect Dnmt1's function as a maintenance MTase by ensuring cytosine methylation of the daughter strand on hemimethylated CG DNA but not on unmethylated CG DNA, which can be captured by the CXXC domain (Song et al. 2011). However, this type of auto-inhibitory effect did not biochemically show an effect in the full-length protein, revealing that there are domain rearrangements that are dependent also on all other domains as illustrated by the different effect of the CXXC domain observed with truncated and full-length Dnmt1 (Bashtrykov et al. 2012; Song et al. 2011). In contrast, plant MET1 does not have a CXXC domain (Fig. 6.4a), suggesting that such an auto-inhibitory mechanism is not conserved in plants. In addition, Dnmt1 contains a replication foci targeting domain (RFTD), which has been shown to interact with the catalytic domain in the absence of DNA, thereby achieving an additional layer of auto-inhibition (Takeshita et al. 2011; Bashtrykov et al. 2014; Syeda et al. 2011). The RFTD domain of Dnmt1 has also been shown to bind to H3K18Ub and H3K23Ub (Ishiyama et al. 2017), leading to Dnmt1 allosteric activation. Binding to H3K9me3 by the RFTD further enhances enzymatic stimulation of Dnmt1 by H3K18Ub and H3K23Ub and mediates Dnmt1 colocalization with H3K9me3 (Ren et al. 2020). In contrast, plant MET1 has two putative RFTDs as predicted by the Pfam server (Finn et al. 2014) (Fig. 6.4a), but their exact function remains unclear. Further structural and functional studies may shed light on the function and regulation of the plant MET1 DNA MTase.

6.3 Conclusion and Perspective

In plants, DNA methylation has important functions in the suppression of TEs and repeat sequences. It acts as a genomic immune response necessary to silence the abundantly distributed TEs across plant genomes (Kim and Zilberman

Fig. 6.4 Potential mechanistic insights into MET1 function. (**a**) Comparison between the domain architectures of *Arabidopsis* MET1 and its ortholog mouse Dnmt1. MET1 contains one additional putative RFTD but lacks the CXXC domain. (**b**) Structure of Dnmt1-DNA complex in an auto-inhibition mode (PDB code: 3PT6) (Song et al. 2011). The CXXC domain, auto-inhibition loop, BAH1, BAH2, MTase domains, and the bounded DNA are colored in *yellow*, *green*, *orange*, *light cyan*, *light magenta*, and *silver*, respectively. A tip within the loop of the BAH2 domain can interact with the TRD subdomain of the MTase to form a novel regulatory mechanism, with the interaction highlighted in blue and red, respectively. (**c**) Sequence alignment of the TRD subdomain of MTase between Dnmt1 and MET1 indicates that MET1 lacks the interaction region within the TRD. Strictly conserved residues are highlighted in red with yellow background, and the moderately conserved residues are colored in *green*. (**d**) Sequence alignment of the BAH2 loop region of Dnmt1 and MET1 showing that MET1 lacks the BAH2 loop. Therefore, MET1 does not have the BAH2-TRD interaction and this type of enzymatic regulation

2014; Wendte and Schmitz 2018). In contrast to the mammalian DNA methylation system, plant DNA methylation is both more diversified and complex and includes methylation of cytosines in all three sequence contexts (CG, CHG, and CHH). Notably, there are four active DNA MTases (DRM2, MET1, CMT3, CMT2), in which the DRM and CMT3/2 are plant-specific DNA MTases, while MET1 has similarities with mammalian Dnmt1. In general, all the available

structures of the catalytic domains of known plant DNA MTases adopt typical type I DNA MTase folds (Zhong et al. 2014; Du et al. 2012), sharing the same catalytic mechanism as mammalian DNA MTases such as Dnmt3a and Dnmt1 (Jia et al. 2007; Song et al. 2011, 2012). However, there is diversity in the regulation of the plant DNA MTases and they are enriched with specific additional features. The CMT family MTases can be regulated through their BAH and chromodomains by the recognition of H3K9me2 (Stroud et al. 2014; Du et al. 2012). The RdDM-driven DRM2 function is regulated by H3K9me2 through SHH1 and by pre-methylated DNA through SUVH2/9 (Johnson et al. 2014; Law et al. 2013). The UBA domains of DRM2 may also play regulatory roles through a so far unknown pathway. MET1 may be regulated similarly to Dnmt1, but also with its own distinct features, because it lacks both the CXXC domain and the regulatory loop within the BAH2 domain.

The current structural studies on plant DNA MTases have established regulatory mechanisms for CMTs and DRM2. Further studies on the UBA domains of DRM2 and MET1 may resolve additional details about the regulation of RdDM and the maintenance of CG methylation in plants. Moreover, further investigations with structures of plant DNA MTase in complex with DNA substrates are required to dissect the molecular mechanisms underlying the observed sequence specificity associated with each plant DNA MTase. The DRM2-DNA structure uncovered DNA deformation as a novel substrate-recognition mechanism for a DNA methyltransferase to establish group-specific DNA methylation (Fang et al. 2021). It will be important to conduct similar investigations on the CMT MTases to investigate whether the substrate deformation is a general mechanism for non-CG methylation in plants.

It is worth noting that in addition to 5mC, the 6 mA mark was reported to be present at transcription start sites that mark the location of active genes in green algae *Chlamydomonas reinhardtii* (Fu et al. 2015). In addition to the algae, 6 mA was also detected in higher plants (Vaniushin et al. 1971). However, the functional role of 6 mA in higher plants is still controversial and requires additional investigation. The enzymes responsible for 6 mA deposition and elimination in plants have not yet been identified. Thus, further investigation of 6 mA in higher plants may open a new window for plant DNA methylation studies.

Acknowledgments We apologize to those whose work was not discussed due to the space limitation. We would like to thank Zhong lab members Dr. Jianjun Jiang and Ray Scheid for the critical reading of this manuscript. This work was supported by NSF (MCB-1552455 and MCB-2043544) and the NIH-MIRA (R35GM124806) awards to XZ, and the Shenzhen Science and Technology Program (JCYJ20200109110403829) to JD.

References

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature 455(7214):818–821. https://doi.org/10.1038/nature07249

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C et al (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature 455(7214):822–825. https://doi.org/10.1038/nature07273

Bashtrykov P, Jankevicius G, Smarandache A, Jurkowska RZ, Ragozin S, Jeltsch A (2012) Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. Chem Biol 19(5):572–578. https://doi.org/10.1016/j.chembiol.2012.03.010

Bashtrykov P, Rajavelu A, Hackner B, Ragozin S, Carell T, Jeltsch A (2014) Targeted mutagenesis results in an activation of DNA methyltransferase 1 and confirms an autoinhibitory role of its RFTS domain. Chembiochem 15(5):743–748. https://doi.org/10.1002/cbic.201300740

Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS One 3(9):e3156. https://doi.org/10.1371/journal.pone.0003156

Bewick AJ, Ji L, Niederhuth CE, Willing EM, Hofmeister BT, Shi X et al (2016) On the origin and evolutionary consequences of gene body DNA methylation. Proc Natl Acad Sci U S A 113(32):9111–9116. https://doi.org/10.1073/pnas.1604666113

Bewick AJ, Niederhuth CE, Ji L, Rohr NA, Griffin PT, Leebens-Mack J et al (2017) The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. Genome Biol 18(1):65. https://doi.org/10.1186/s13059-017-1195-1

Blus BJ, Wiggins K, Khorasanizadeh S (2011) Epigenetic virtues of chromodomains. Crit Rev Biochem Mol Biol 46(6):507–526. https://doi.org/10.3109/10409238.2011.619164

Cao X, Jacobsen SE (2002a) Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A 99(Suppl 4):16491–16498. https://doi.org/10.1073/pnas.162371599

Cao X, Jacobsen SE (2002b) Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Curr Biol 12(13):1138–1144

Castel SE, Martienssen RA (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat Rev Genet 14(2): 100–112. https://doi.org/10.1038/nrg3355

Chen J, Liu J, Jiang J, Qian S, Song J, Kabara R et al (2021) F-box protein CFK1 interacts with and degrades de novo DNA methyltransferase in Arabidopsis. New Phytol 229(6):3303–3317. https://doi.org/10.1111/nph.17103

Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ (1993) Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell 74(2):299–307

Cheng C, Tarutani Y, Miyao A, Ito T, Yamazaki M, Sakai H et al (2015) Loss of function mutations in the rice chromomethylase OsCMT3a cause a burst of transposition. Plant J 83(6):1069–1081. https://doi.org/10.1111/tpj.12952

Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y et al (2010) Relationship between nucleosome positioning and DNA methylation. Nature 466(7304):388–392. https://doi.org/10.1038/nature09147.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD et al (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452(7184):215–219. https://doi.org/10.1038/nature06745

Dangwal M, Malik G, Kapoor S, Kapoor M (2013) De novo methyltransferase, OsDRM2, interacts with the ATP-dependent RNA helicase, OseIF4A, in rice. J Mol Biol 425(16):2853–2866. https://doi.org/10.1016/j.jmb.2013.05.021

Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E et al (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell 151(1):167–180. https://doi.org/10.1016/j.cell.2012.07.034

Du J, Johnson LM, Groth M, Feng S, Hale CJ, Li S et al (2014) Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. Mol Cell 55(3): 495–504. https://doi.org/10.1016/j.molcel.2014.06.009

Du J, Johnson LM, Jacobsen SE, Patel DJ (2015) DNA methylation pathways and their crosstalk with histone methylation. Nat Rev Mol Cell Biol 16(9):519–532. https://doi.org/10.1038/nrm4043

Fang J, Leichter SM, Jiang J, Biswal M, Lu J, Zhang ZM et al (2021) Substrate deformation regulates DRM2-mediated DNA methylation in plants. Sci Adv 7(23). https://doi.org/10.1126/sciadv.abd9224

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR et al (2014) Pfam: the protein families database. Nucleic Acids Res 42(Database issue): D222–D230. https://doi.org/10.1093/nar/gkt1223

Finnegan EJ, Kovac KA (2000) Plant DNA methyltransferases. Plant Mol Biol 43(2–3):189–201

Fu Y, Luo GZ, Chen K, Deng X, Yu M, Han D et al (2015) N(6)-methyldeoxyadenosine marks active transcription start sites in chlamydomonas. Cell 161(4): 879–892. https://doi.org/10.1016/j.cell.2015.04.010

Gallego-Bartolomé J, Liu W, Kuo PH, Feng S, Ghoshal B, Gardiner J et al (2019) Co-targeting RNA polymerases IV and V promotes efficient de novo DNA methylation in Arabidopsis. Cell 176(5):1068–82.e19. https://doi.org/10.1016/j.cell.2019.01.029

Gao L, Emperle M, Guo Y, Grimm SA, Ren W, Adam S et al (2020) Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. Nat Commun 11(1):3355. https://doi.org/10.1038/s41467-020-17109-4

Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. Annu Rev Biochem 74:481–514. https://doi.org/10.1146/annurev.biochem.74.010904.153721

Gouil Q, Baulcombe DC (2016) DNA methylation signatures of the plant Chromomethyltransferases. PLoS Genet 12(12):e1006526. https://doi.org/10.1371/journal.pgen.1006526

Haag JR, Pikaard CS (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nat Rev Mol Cell Biol 12(8):483–492. https://doi.org/10.1038/nrm3152

Haag JR, Ream TS, Marasco M, Nicora CD, Norbeck AD, Pasa-Tolic L et al (2012) In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. Mol Cell 48(5):811–818. https://doi.org/10.1016/j.molcel.2012.09.027

Haag JR, Brower-Toland B, Krieger EK, Sidorenko L, Nicora CD, Norbeck AD et al (2014) Functional diversification of maize RNA polymerase IV and V subtypes via alternative catalytic subunits. Cell Rep 9(1):378–390. https://doi.org/10.1016/j.celrep.2014.08.067

Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature 455(7214):826–829. https://doi.org/10.1038/nature07280

He XJ, Ma ZY, Liu ZW (2014) Non-coding RNA transcription and RNA-directed DNA methylation in Arabidopsis. Mol Plant 7(9):1406–1414. https://doi.org/10.1093/mp/ssu075

He L, Zhao C, Zhang Q, Zinta G, Wang D, Lozano-Durán R et al (2021) Pathway conversion enables a double-lock mechanism to maintain DNA methylation and genome stability. Proc Natl Acad Sci U S A 118(35). https://doi.org/10.1073/pnas.2107320118

Henderson IR, Deleris A, Wong W, Zhong X, Chin HG, Horwitz GA et al (2010) The de novo cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA-directed DNA methylation in Arabidopsis thaliana. PLoS Genet 6(10):e1001182. https://doi.org/10.1371/journal.pgen.1001182

Hu D, Yu Y, Wang C, Long Y, Liu Y, Feng L et al (2021) Multiplex CRISPR-Cas9 editing of DNA methyltransferases in rice uncovers a class of non-CG methylation specific for GC-rich regions. Plant Cell 33(9):2950–2964. https://doi.org/10.1093/plcell/koab162

Ishiyama S, Nishiyama A, Saeki Y, Moritsugu K, Morimoto D, Yamaguchi L et al (2017) Structure of the Dnmt1 reader module complexed with a unique two-mono-ubiquitin mark on histone H3 reveals the basis for DNA methylation maintenance. Mol Cell 68(2):350–60.e7. https://doi.org/10.1016/j.molcel.2017.09.037

Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature 416(6880):556–560. https://doi.org/10.1038/nature731

Jackson JP, Johnson L, Jasencakova Z, Zhang X, PerezBurgos L, Singh PB et al (2004) Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. Chromosoma 112(6):308–315. https://doi.org/10.1007/s00412-004-0275-7

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449(7159):248–251. https://doi.org/10.1038/nature06146

Jiang J, Liu J, Sanders D, Qian S, Ren W, Song J et al (2021) UVR8 interacts with de novo DNA methyltransferase and suppresses DNA methylation in Arabidopsis. Nat Plants 7(2):184–197. https://doi.org/10.1038/s41477-020-00843-4

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J et al (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol 17(4):379–384. https://doi.org/10.1016/j.cub.2007.01.009

Johnson LM, Law JA, Khattar A, Henderson IR, Jacobsen SE (2008) SRA-domain proteins required for DRM2-mediated de novo DNA methylation. PLoS Genet 4(11):e1000280. https://doi.org/10.1371/journal.pgen.1000280

Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK et al (2014) SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature 507(7490):124–128. https://doi.org/10.1038/nature12931

Jurkowska RZ, Anspach N, Urbanke C, Jia D, Reinhardt R, Nellen W et al (2008) Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. Nucleic Acids Res 36(21):6656–6663. https://doi.org/10.1093/nar/gkn747

Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA et al (2003) Arabidopsis MET1 cytosine methyltransferase mutants. Genetics 163(3):1109–1122

Kim MY, Zilberman D (2014) DNA methylation as a system of plant genomic immunity. Trends Plant Sci 19(5):320–326. https://doi.org/10.1016/j.tplants.2014.01.014

Kuo AJ, Song J, Cheung P, Ishibe-Murakami S, Yamazoe S, Chen JK et al (2012) The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. Nature 484(7392):115–119. https://doi.org/10.1038/nature10956

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11(3):204–220. https://doi.org/10.1038/nrg2719

Law JA, Ausin I, Johnson LM, Vashisht AA, Zhu JK, Wohlschlegel JA et al (2010) A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in Arabidopsis. Curr Biol 20(10):951–956. https://doi.org/10.1016/j.cub.2010.03.062

Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AM et al (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature 498(7454):385–389. https://doi.org/10.1038/nature12178

Li X, Harris CJ, Zhong Z, Chen W, Liu R, Jia B et al (2018) Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. Proc Natl Acad Sci U S A 115(37):E8793–EE802. https://doi.org/10.1073/pnas.1809841115

Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S et al (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science 292(5524):2077–2080. https://doi.org/10.1126/science.1059745

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH et al (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133(3):523–536. https://doi.org/10.1016/j.cell.2008.03.029

Liu ZW, Shao CR, Zhang CJ, Zhou JX, Zhang SW, Li L et al (2014) The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. PLoS Genet 10(1):e1003948. https://doi.org/10.1371/journal.pgen.1003948

Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet 15(6):394–408. https://doi.org/10.1038/nrg3683

Niu Q, Song Z, Tang K, Chen L, Wang L, Ban T et al (2021) A histone H3K4me1-specific binding protein is required for siRNA accumulation and DNA methylation at a subset of loci targeted by RNA-directed DNA methylation. Nat Commun 12(1):3367. https://doi.org/10.1038/s41467-021-23637-4

Nozawa K, Chen J, Jiang J, Leichter SM, Yamada M, Suzuki T et al (2021) DNA methyltransferase CHROMOMETHYLASE3 prevents ONSEN transposon silencing under heat stress. PLoS Genet 17(8): e1009710. https://doi.org/10.1371/journal.pgen.1009710

Patel DJ, Wang Z (2013) Readout of epigenetic modifications. Annu Rev Biochem 82:81–118. https://doi.org/10.1146/annurev-biochem-072711-165700

Rajakumara E, Law JA, Simanshu DK, Voigt P, Johnson LM, Reinberg D et al (2011) A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. Genes Dev 25(2): 137–152. https://doi.org/10.1101/gad.1980311

Rajakumara E, Nakarakanti NK, Nivya MA, Satish M (2016) Mechanistic insights into the recognition of 5-methylcytosine oxidation derivatives by the SUVH5 SRA domain. Sci Rep 6:20161. https://doi.org/10.1038/srep20161

Ren W, Fan H, Grimm SA, Guo Y, Kim JJ, Yin J et al (2020) Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. Proc Natl Acad Sci U S A 117(31): 18439–18447. https://doi.org/10.1073/pnas.2009316117

Ren W, Fan H, Grimm SA, Kim JJ, Li L, Guo Y et al (2021) DNMT1 reads heterochromatic H4K20me3 to reinforce LINE-1 DNA methylation. Nat Commun 12(1):2490. https://doi.org/10.1038/s41467-021-22665-4

Scheid R, Chen J, Zhong X (2021) Biological role and mechanism of chromatin readers in plants. Curr Opin Plant Biol 61:102008. https://doi.org/10.1016/j.pbi.2021.102008

Shen X, De Jonge J, Forsberg SK, Pettersson ME, Sheng Z, Hennig L et al (2014) Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality. PLoS Genet 10(12):e1004842. https://doi.org/10.1371/journal.pgen.1004842

Song J, Rechkoblit O, Bestor TH, Patel DJ (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science 331(6020):1036–1040. https://doi.org/10.1126/science.1195380

Song J, Teplova M, Ishibe-Murakami S, Patel DJ (2012) Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science 335(6069):709–712. https://doi.org/10.1126/science.1214453

Stoddard CI, Feng S, Campbell MG, Liu W, Wang H, Zhong X et al (2019) A nucleosome bridging mechanism for activation of a maintenance DNA methyltransferase. Mol Cell 73(1):73–83.e6. https://doi.org/10.1016/j.molcel.2018.10.006

Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell 152(1–2):352–364. https://doi.org/10.1016/j.cell.2012.10.054

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L et al (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol 21(1):64–72. https://doi.org/10.1038/nsmb.2735

Syeda F, Fagan RL, Wean M, Avvakumov GV, Walker JR, Xue S et al (2011) The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. J Biol Chem 286(17):15344–15351. https://doi.org/10.1074/jbc.M110.209882

Takeshita K, Suetake I, Yamashita E, Suga M, Narita H, Nakagawa A et al (2011) Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proc Natl Acad Sci U S A 108(22): 9055–9059. https://doi.org/10.1073/pnas.1019629108

Vaniushin BF, Kadyrova D, Karimov K, Belozerskii AN (1971) Minor bases in DNA of higher plants. Biokhimiia 36(6):1251–1258

Vanyushin BF, Ashapkin VV (2011) DNA methylation in higher plants: past, present and future. Biochim Biophys Acta 1809(8):360–368. https://doi.org/10.1016/j.bbagrm.2011.04.006

Wang Q, Xue Y, Zhang L, Zhong Z, Feng S, Wang C et al (2021a) Mechanism of siRNA production by a plant Dicer-RNA complex in dicing-competent conformation. Science 374(6571):1152–1157. https://doi.org/10.1126/science.abl4546

Wang Y, Zhou X, Luo J, Lv S, Liu R, Du X et al (2021b) Recognition of H3K9me1 by maize RNA-directed DNA methylation factor SHH2. J Integr Plant Biol 63(6):1091–1096. https://doi.org/10.1111/jipb.13103

Wendte JM, Schmitz RJ (2018) Specifications of targeting heterochromatin modifications in plants. Mol Plant 11(3):381–387. https://doi.org/10.1016/j.molp.2017.10.002

Wendte JM, Zhang Y, Ji L, Shi X, Hazarika RR, Shahryary Y et al (2019) Epimutations are associated with CHROMOMETHYLASE 3-induced de novo DNA methylation. elife 8:e47891. https://doi.org/10.7554/eLife.47891

Woo HR, Dittmer TA, Richards EJ (2008) Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. PLoS Genet 4(8):e1000156. https://doi.org/10.1371/journal.pgen.1000156

Yaari R, Noy-Malka C, Wiedemann G, Auerbach Gershovitz N, Reski R, Katz A et al (2015) DNA

METHYLTRANSFERASE 1 is involved in (m)CG and (m)CCG DNA methylation and is essential for sporophyte development in Physcomitrella patens. Plant Mol Biol 88(4–5):387–400. https://doi.org/10.1007/s11103-015-0328-8

Yang N, Xu RM (2013) Structure and function of the BAH domain in chromatin biology. Crit Rev Biochem Mol Biol 48(3):211–221. https://doi.org/10.3109/10409238.2012.742035

Zabet NR, Catoni M, Prischi F, Paszkowski J (2017) Cytosine methylation at CpCpG sites triggers accumulation of non-CpG methylation in gene bodies. Nucleic Acids Res 45(7):3777–3784. https://doi.org/10.1093/nar/gkw1330

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K et al (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153(1):193–205. https://doi.org/10.1016/j.cell.2013.02.033

Zhai J, Bischof S, Wang H, Feng S, Lee TF, Teng C et al (2015) A one precursor one siRNA model for pol IV-dependent siRNA biogenesis. Cell 163(2):445–455. https://doi.org/10.1016/j.cell.2015.09.032

Zhang H, Ma ZY, Zeng L, Tanaka K, Zhang CJ, Ma J et al (2013) DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. Proc Natl Acad Sci U S A 110(20):8290–8295. https://doi.org/10.1073/pnas.1300585110

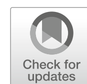
Zhang H, Lang Z, Zhu JK (2018a) Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol 19(8):489–506. https://doi.org/10.1038/s41580-018-0016-z

Zhang ZM, Lu R, Wang P, Yu Y, Chen D, Gao L et al (2018b) Structural basis for DNMT3A-mediated de novo DNA methylation. Nature 554(7692):387–391. https://doi.org/10.1038/nature25477

Zhao Y, Chen X (2014) Noncoding RNAs and DNA methylation in plants. Natl Sci Rev 1(2):219–229. https://doi.org/10.1093/nsr/nwu003

Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA et al (2014) Molecular mechanism of action of plant DRM de novo DNA methyltransferases. Cell 157(5):1050–1060. https://doi.org/10.1016/j.cell.2014.03.056

Zhong X, Hale CJ, Nguyen M, Ausin I, Groth M, Hetzel J et al (2015) Domains rearranged methyltransferase3 controls DNA methylation and regulates RNA polymerase V transcript abundance in Arabidopsis. Proc Natl Acad Sci U S A 112(3):911–916. https://doi.org/10.1073/pnas.1423603112



7 DNA Methylation in Honey Bees and the Unresolved Questions in Insect Methylomics

Laura Wedd, Robert Kucharski, and Ryszard Maleszka

Abstract

DNA methylation has been found in most invertebrate lineages except for *Diptera*, *Placozoa* and the majority of *Nematoda*. In contrast to the mammalian methylation toolkit that consists of one DNMT1 and several DNMT3s, some of which are catalytically inactive accessory isoforms, invertebrates have different combinations of these proteins with some using just one DNMT1 and the others, like the honey bee, two DNMT1s one DNMT3. Although the insect DNMTs show sequence similarity to mammalian DNMTs, their in vitro and in vivo properties are not well investigated. In contrast to heavily methylated mammalian genomes, invertebrate genomes are only sparsely methylated in a 'mosaic' fashion with the majority of methylated CpG dinucleotides found across gene bodies that are frequently associated with active transcription. Additional work also highlights that obligatory methylated epialleles influence transcriptional changes in a context-specific manner. We argue that some of the lineage-specific properties of DNA methylation are the key to understanding the role of this genomic modification in insects. Future mechanistic work is needed to explain the relationship between insect DNMTs, genetic variation, differential DNA methylation, other epigenetic modifications, and the transcriptome in order to fully understand the role of DNA methylation in converting genomic sequences into phenotypes.

Keywords

Social insects · Epigenome · Epigenetics · Epialleles · DNMT isoforms · TET

Abbreviations

ALK	anaplastic lymphoma kinase
CpG	cytosine and guanidine dinucleotide separated by one phosphate in DNA
DBP	DNA-binding protein
DNMT	DNA methyltransferase
LAM	lysosomal alpha-mannosidase
MeCP	methyl-CpG-binding factor
PTM	post-translational modification
RdDM	RNA-directed DNA methylation system
TET	Ten-eleven translocation enzyme

L. Wedd · R. Kucharski · R. Maleszka (✉)
Research School of Biology, The Australian National University, Canberra, Australia
e-mail: robert.kucharski@anu.edu.au; ryszard.maleszka@anu.edu.au

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_7

7.1 Introduction

Amongst all fields of biomedical research, it is ‘epigenetics’ that is emerging as a principal discipline bridging the gap between genotype and phenotype. One reason for such a prominent standing of epigenetic concepts in modern biology is the flexible and ever-expanding definition of what the term ‘epigenetics’ actually means. Over 70 years ago Waddington originally introduced the word ‘epigenetics’ (derived from the word ‘epigenesis’) as ‘a suitable name for the branch of biology which studies the causal interactions between genes and their products which bring phenotype into being’ (Jablonka and Lamb 2002; Waddington 1942). This new definition implied that translating the genetic blueprint into a functional organism requires a control system whose mode of action is over and above, or in addition to, the classical genotype (Waddington 1942). His ideas were incited by the realisation that phenotypes are remarkably stable despite the environmental pressure, a phenomenon that he referred to as developmental canalisation, whereby development is buffered against environmental or genetic variation by evolutionarily selected gene networks. He illustrated this concept by the famed epigenetic landscape with systems of connected valleys or channels representing the optimal developmental trajectories. The key idea behind this imaginary topography is that a set of instructions carving the trajectory towards an optimal phenotype is selected by responses to recurring environmental insults in an organism’s adaptive niche (Waddington 1957). Distinct from canalisation, phenotypic plasticity is another feature independent of the underlying DNA sequence, that is equally important for the living world. It is a phenomenon whereby contrasting organismal outcomes are produced from one genotype using intricate developmental cues. Cellular differentiation in multicellular organisms or phenotypic polymorphisms in social insects is generated by epigenetic mechanisms, not by genetic differences. These two contrasting aspects of phenotypic changeability that are uncoupled from genetic variation are central to the study of epigenetics. At present, epigenetics is a very wide field of study, covering virtually all aspects of biology ranging from morphogenesis to transgenerational epigenetic inheritance. For more detailed discussions on the origins and changing concept of epigenetics, see recent review and opinion articles (Jablonka and Lamm 2012; Haig 2004, 2012).

In modern times the distinction between genetic and epigenetic control system has been associated with specific biochemical processes. Typically, DNA methylation and histone modifications are itemised as epigenetic mechanisms, but non-coding RNAs, chromatin remodelling and even prions are often included (Halfmann and Lindquist 2010). In this context, it is worth noting that only DNA methylation is chemically a direct part of the DNA molecule. It is thus clearly differentiated and distinguished from other epigenetic mechanisms, such as chromatin modification and non-coding RNAs, which are associated with but separate from DNA. For over 50 years, DNA methylation or more specifically cytosine methylation has been studied extensively in mammals, initially as a mechanism of gene silencing via hyper-methylation of promoters associated with the CpG islands and later as a genome-wide modification. In contrast, only a couple of studies in the late 1990s have shown that 5-methyl-cytosine can be detected in DNAs extracted from various invertebrate species (Tweedie et al. 1997; Regev et al. 1998), and it is largely through the honey bee genomic research that methylomics took center stage as an important mechanism of gene regulation in insects/invertebrates. While the progress in this field has been inspiring, it also led to a widely accepted idea that social insects, in particular honey bees, wasps and ants have a DNA methylation toolkit that is functionally similar to that in mammals. However, many features of DNA methylation in invertebrates suggest that this level of genomic modifications is driven by several quite distinct mechanisms that are lineage specific (Maleszka and Kucharski 2022).

In this article, we discuss DNA methylation as part of gene regulatory systems in invertebrates,

with special emphasis on the honey bee, *Apis mellifera*, and bring into focus several unresolved aspects of this important regulatory mechanism that need to be addressed experimentally to better understand how epigenomic modifications link genotype to phenotype.

7.2 Genotype to Phenotype

The existence of a multicellular organism depends upon the transformation of an apparently simple, static genetic 'code' into variable functional states. With perfect timing, the genetic information contained within that first single cell is translated into a series of complex cellular signals that guide development. As each cell differentiates uniquely, transcriptional profiles are established, and their functional roles are specified. This transition from genotype to phenotype, termed *epigenesis*, is infinitely complex and results from interactions between the underlying genetic sequence, chemical modifications on DNA and chromatin, and environmental cues. Yet, how exactly do these layers of information contribute to establishing phenotype?

Foremost, the DNA sequence itself interacts with the transcriptional machinery to produce the multitude of proteins necessary for a functioning organism. The human genome, for instance, encodes approximately 21,000 protein-coding genes whose sequences are transcribed to generate our proteome (Clamp et al. 2007). The honey bee has at least 18,000 genes, but this number is expected to be evaluated upward following upgraded genomic assemblies (Elsik et al. 2014). Gene transcription occurs when RNA Polymerase II (RNA Pol II) interacts with the genomic regions to produce a transcript. This is governed by numerous functional elements such as enhancers, or promoters associated with coding regions, that interact with various activators and transcription factors to facilitate the assembly of the pre-initiation complex and subsequent transcription by RNA Pol II (Fig. 7.1).

In order to generate the full repertoire of proteins required to build a complex and fully functioning organism, each gene must be expressed in a precise spatio-temporal pattern. To achieve this, numerous control mechanisms have evolved to tightly regulate transcription. These range from modulating transcription initiation and elongation, which can alter whether a gene is expressed or not and how a transcript is processed, to post-translational processes that can, for instance, fine-tune the level of a transcript through the degradation of an mRNA product. In addition, epigenetic control systems have evolved to coordinate the action of thousands of genes and to provide an interface between the genome and environment. Each of these layers represents an important mechanism through which gene expression can be controlled.

7.3 The Epigenetic Control of Gene Expression

Epigenetics encompasses those mechanisms and processes that are involved in facilitating transcriptional changes via various covalent modifications made to DNA itself or the histone proteins around which DNA is wrapped. The post-translational modifications (PTMs) of histone proteins are numerous, and include methylation, acetylation, phosphorylation, ubiquitination, and many other PTMs. The manner in which these modifications can influence transcription is complex, and it is likely that a given mark can lead to both gene activation and repression, depending on its location and the genomic context (Kouzarides 2007; Jones 2012). Importantly, none of these modifications works in isolation, instead they interact together forming an epigenome modification network and produce a unique epigenetic cellular 'signature'.

This signature, termed a cell's *epigenome*, describes all epigenome modifications found across the genome. For each cell type within an organism, and across different developmental stages and disease states the epigenome will vary, leading to a vast number of possible epigenomes (Bernstein et al. 2007). Through analysing patterns of DNA methylation, histone modifications, chromatin accessibility and RNA expression across diverse cell lineages, human

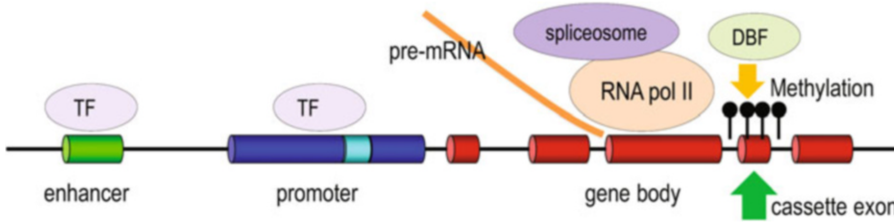


Fig. 7.1 General model of transcription in the honey bee showing the key elements involved in the initiation and elongation of pre-mRNA. Splicing is assumed to occur co-transcriptionally with DNA methylation affecting conditional usage of weak exons. In insects, methylated cytosines have been found predominantly in intragenic regions, often in proximity of splice sites. DNA methylation may affect binding kinetics of sequence-specific factors (DBP) either by disrupting their binding, as in the CTCF model (Shukla et al. 2011), or by recruiting methyl-CpG-binding factors such as MeCP2 (Maunakea et al. 2013). In both models, such factors promote exon recognition either by pausing the spliceosome/RNA pol II complex (Shukla et al. 2011), or by recruiting histone deacetylases to maintain low acetylation levels of alternatively spliced exons, which could reduce transcription elongation (Maunakea et al. 2013). *TF* transcription factor

epigenomic maps are being catalogued (Kundaje et al. 2015; Bernstein et al. 2010). These large-scale pursuits are leading to a greater understanding of how the epigenome contributes to cell specification and development, and how alterations to the epigenome contribute to disease and phenotypic variation. However, defining the exact role that a given epigenome modification plays in directing transcriptional changes remains a challenge. This difficulty will be discussed here in the context of one of the extensively studied epigenomic modifications, namely DNA methylation.

7.4 DNA Methylation

In all vertebrates and in most invertebrates the cytosine nucleotide in DNA can be modified by the addition of a methyl group to its fifth carbon atom. This modification typically occurs at cytosines in CpG dinucleotides and is prevalent in mammalian genomes where up to 80% (over 20 million) (Smith and Meissner 2013) of cytosines in the CpG context are methylated (Bird 2002). In comparison, the proportion of methylated CpGs in insect genomes is a few orders of magnitude lower, ranging from approximately 0.1 to 1% (Lyko and Maleszka 2011). The realisation that this covalent modification could stably propagate information throughout development led to considerable research into its role as an epigenetic mark (Holliday and Pugh 1975; Riggs 1975). It has since been shown that DNA methylation plays a key role in biological processes such as X-inactivation, genomic imprinting, and transposon silencing across several organisms (chapter cross reference Genetic studies on mammalian DNA methylation). In each of these instances, DNA methylation is engaged in gene silencing, and this remains one of the more commonly considered functions of DNA methylation. This is evident in mammals where the methylation of promoter regions has long since been associated with gene repression (Jones and Takai 2001). However, considerable variation in the distribution of methylation across organisms and additional evidence that methylation also associates with active transcription indicates that the relationship between DNA methylation and transcription is complex, and not fully understood. Insight into its functionality can be gained through understanding how DNA methylation patterns are established and by analysing the evolutionary

origins of this mark and its genomic location (Regev et al. 1998; Suzuki and Bird 2008).

7.4.1 Conserved and Non-Conserved Features of DNA Methylation Enzymology in Animals

The overall level and patterns of DNA methylation across a genome, what is termed a *methylome*, are set up in a cell- and tissue-specific manner. Throughout embryogenesis and cell differentiation, DNA methyltransferases (DNMTs) establish DNA methylation patterns and then maintain these patterns across cell divisions. Two families of DNMTs exist; one referred to as DNMT1, which preferentially methylates hemimethylated cytosine residues, and the other as DNMT3 family, which catalyses the de novo methylation of DNA (Goll and Bestor 2005) (chapter cross references Enzymology of mammalian DNA MTase). Other important proteins relevant for DNA methylation is the family of Ten-eleven translocation (TET) dioxygenase enzymes that are involved in the active DNA demethylation process (see a section below) (chapter reference TET enzymes). All these enzymes are well conserved and present across a wide variety of vertebrates and invertebrates. However, as shown in Table 7.1 there are two salient aspects of DNA methylation across the animal kingdom. One is its dispensability in an evolutionary context, and second, the apparent differences in the allocation of DNMTs in various lineages. The distribution of the DNA methylation toolkit in *Metazoa* is mosaic, with a variety of distinct patterns ranging from a total loss of both DNMT and TET enzymes to partial gene losses and duplications. Many species, including most nematodes, advanced Dipteran insects and *Placozoa* have lost DNA methylating enzymes and apparently expanded other epigenetic mechanisms or even recruited new ones to regulate gene expression. Given this contrasting utilisation of an ancient biochemical modification, the extent to which DNA methylation is advantageous if maintained in a given lineage remains unclear (Lyko and Maleszka 2011; Zemach et al. 2010; Miklos and Maleszka 2011).

Initially, it was considered that DNA methylation patterns were established by the de novo activity of the DNMT3 family during embryogenesis and then these patterns were maintained by DNMT1, the 'maintenance' DNMT. However, recent work indicates that the classical roles of DNMT1 and DNMT3s in establishing and maintaining methylation patterns need to be redefined to include the apparent de novo activity of DNMT1 and DNMT3s' contribution to DNA methylation maintenance (Jeltsch and Jurkowska 2014). Another interesting feature of the mammalian methylation toolkit is the catalytically inactive paralog DNMT3L that evolved to function as an accessory protein interacting with the active DNMT3s. In mammals, DNMT3A has two different isoforms, while DNMT3B has more than 30 isoforms showing highly conserved patterns of expression in humans and rodents (Duymich et al. 2016; Gao et al. 2020; Zeng et al. 2020). DNMT3B isoforms modulate gene body methylation and re-methylation, and even isoforms without catalytic activity, like DNMT3B3, have roles in this process, thus resembling a functional role of the partner protein DNMT3L that recruits DNMT3A in undifferentiated cells to initiate DNA methylation. Recent evidence suggests that DNMT3B3 might substitute for the accessory protein DNMT3L to recruit DNMT3A in somatic cells (Duymich et al. 2016; Gao et al. 2020).

The variation of DNMTs across invertebrates is also suggestive of diverse roles for these enzymes (Table 7.1). Many insect lineages lack DNMT3 but still methylate their genomes and it is therefore likely that DNMT1 present in these organisms has de novo activity (Lyko and Maleszka 2011; Xiang et al. 2010). Honey bees and other Hymenoptera have a very characteristic methylation toolkit comprising two or even three copies of DNMT1 and one copy of DNMT3 (Maleszka and Kucharski 2022). This distinct expansion of the DNA methylation enzymology is found in insects with high level of phenotypic plasticity, suggesting that epigenomic control systems are prime movers of developmental networks controlling organismal plasticity

Table 7.1 Examples of mosaic distribution of DNA methylation toolkit in selected metazoan species

Lineage	Species	DNMT 1	DNMT 3	TET	methyl CpGs
Mammalia	All	●	●●●	●●●	Yes
Hymenoptera	*Apis mellifera* (honey bee)	●●	●	●	Yes
	Nasonia spp (parasitic wasps)	●●●	●	●	Yes
Homoptera	*Acyrthosiphon pisum* (pea aphid)	●	●	●	Yes
Coleoptera	*Tribolium castaneum* (flour beetle)	●	–	●	Yes
Lepidoptera	*Bombyx mori* *(silk worm)*	●	–	●	Yes
Diptera	Flies, mosquitos	–	–	●[a]	No
Nematoda	*Caenorhabditis elegans* (free living roundworm)	–	–	–	No
	Trichuris trichiura (whipworm, parasitic)	–	●[b]	–	?
Placozoa	*Trichoplax adhaerens* (the only extant member of this phylum)	–	–	●	No
Cnidaria	*Nematostella vectensis* (sea anemone)	●	●	●	Yes

Black dots indicate the number of genes encoding DNMT paralogs, and the presence of TET in various species

[a] It has been suggested that in Drosophila, TET demethylates N6-methyl-adenine (6 mA) and this process correlates with transposon expression (Zhang et al. 2015)

[b] Highly conserved relatives of DNMT3 in *T. suis* and in a closely related species *T. trichiura* available in GenBank (KFD71641.1; CDW57637.1) suggest that certain nematodes have genes encoding DNMT3, but no DNMT1 and TET

(Miklos and Maleszka 2011; Maleszka 2016, 2018). Interference with DNA methylation by knocking down DNMT3 in larvae reared in vitro mimics the effect of royal jelly in post-embryonic female development, suggesting that dietary ingredients have the capacity to affect epigenomic settings of a complex nutritionally-driven developmental process (Maleszka 2018; Kucharski et al. 2008). In a colony, newly hatched larvae fed royal jelly develop into long-lived reproductive queens, whereas larvae fed less potent worker jelly develop into functionally sterile short-lived workers (Maleszka 2014). Although this result clearly implicates DNA methylation in a diet-controlled developmental division between long-lived highly fertile queens and short-lived functionally sterile workers, it does not provide an unambiguous mechanistic explanation of this process based on one enzyme whose properties remain to be determined. Given the emerging view that all DNMTs are not only functionally interweaved, but also cooperate with histone modifiers, the impact of DNMT3 silencing on honey bee phenotypes is more likely the outcome of a global effect that creates a disturbance in a highly interconnected epigenomic regulatory system. To move forward, insect epigenetics needs to focus on unravelling the functions of both DNA and histone modifiers before a model of this intriguing epigenetic phenomenon can be generated (Maleszka and Kucharski 2022). In this context, the queen bee food, royal jelly, may well hold clues to this problem. It is a complex mixture of unique compounds with intriguing but poorly understood biochemical characteristics (Maleszka 2014; Mandacaru et al. 2017; Buttstedt et al. 2018). Some of these compounds have been identified as histone deacetylase inhibitors and it is reasonable to assume that many other ingredients in royal jelly also have potent epigenetic qualities, possibly affecting DNA/RNA modifying enzymology (Maleszka 2014; Spannhoff et al. 2011; Hattori et al. 2007). Cloning and in vitro characterisation

of the honey bee DNMTs and other genes combined with examining the effects of royal jelly components will provide unprecedented clues to the nature of dietary impacts on epigenetic machinery.

7.4.2 DNMTs and Establishing DNA Methylation Patterns in the Honey Bee

Currently, little is known about DNMTs in insects and other invertebrates and our ideas and explanations carry the proviso that their *modus operandi* is comparable to that of mammalian DNMTs. All DNMTs in honey bees are ubiquitously expressed in various tissues and throughout development but their functionality could be distinctive depending on the context, the level of expression, specific splice variants, substrate preference, etc. The DNMT1 paralogs AmDNMT1a and AmDNMT1b encode almost identical proteins whose specific roles have not been studied in detail. Although they might be partially redundant, they also could have distinctive cellular roles resulting from divergent substrate preference as shown for DNMT3A and DNMT3B in mammals (Guo et al. 2014). Interestingly, recent sequencing data have revealed the potential of the gene encoding DNMT3 in honey bees to generate multiple splice variants whose biochemical properties are presently unknown. Improved gene models of AmDNMT1a, AmDNMT1b and AmDNMT3 based on transcriptome sequencing from several studies (Wojciechowski et al. 2018; Ashby et al. 2016) are shown in Fig. 7.2. Some of the newly uncovered variants are unlikely to be catalytically active and may serve as regulatory or accessory proteins. For example, a transcript with deleted exons 4–5 does not have the PWWP domain that is important for histone binding, and therefore is expected to act in a different manner than DNMT3A or DNMT3B. This variant resembles the accessory isoform DNMT3L in mammals that has no PWWP domain but contains the ADDz-DNMT3L domain with a PHD-like zinc finger motif and a C2-C2 (Lu et al. 2020). Although DNMT3L does not have DNA methyltransferase activity, it acts as an accessory protein to regulate the functionality of DNMT3A and DNMT3B (Ooi et al. 2007). These improved AmDNMT genes models suggest that alternative splicing rather than gene duplication is used to expand the cellular utility of DNMT3 in honey bees. Another intriguing feature of AmDNMT3 is a duplicated PWWP domain that is critical for DNMT3s' interactions with chromatin and other proteins (Maleszka and Kucharski 2022). It is one of four currently known domains that not only bind modified histone tails but can additionally bind to DNA (Wang et al. 2020). Its key function is to read methylated DNA and methylated lysine in histones (Rona et al. 2016). So far, this is the only case of a double PWWP domain found in DNMT3.

7.4.3 How do TET Enzymes Contribute to Gene Regulation in the Honey Bees and Other Insects?

Caste determination in *A. mellifera* is an illustrative example of Waddington's developmental canalisation, whereby epigenetic processes have been recruited to conditionally modulate the expression of one genome using an environmental cue. Although various cellular elements have the ability to respond to environmental change, their combined and coordinated action has evolved in honey bees as a controlling mechanism for reprograming the entire larval development with critical consequences for cellular and organismal phenotypes. A limited number of genes can be epigenetically programmed to yield more than one organismal outcome, suggesting that epigenomic modifiers are able to relax evolutionary constrains on development. These modifiers operate by recruiting only a subset of an organism's gene repertoire and reusing it in a combinatorial manner to remodel multiple sub-networks (Mattick et al. 2010; Maleszka et al. 2014; Erwin and Davidson 2009).

Importantly, DNA methylation patterns are governed not only by the de novo and

Fig. 7.2 Improved models of the honey bee DNMTs based on new sequencing data. (**a**) Domain architecture of two very closely related DNMT1s, designated AmDNMT1a and 1b, and one ortholog of DNMT3. The positions of the conserved domains and motifs are labelled in colour. AmDNMT3 has an unusual duplicated PWWP domain not seen in other DNMTs. (**b**) The gene encoding AmDNMT3 has the capacity to generate multiple spliced variants suggesting that various protein isoforms, including some with no catalytic activity, are used by the honey bee methylation system. It is noteworthy that a long intron between exons 15 and 16 with lots of repeats is methylated (indicated by the black lollypops). These models have been improved using the honey bee genome assembly version Amel4.5. *DNMT1-RFD* cytosine specific DNA methyltransferase replication foci domain, *BAH* Bromo-adjacent homology domain, *NLS* predicted nuclear localisation signal, *PHD* finger -a Cys4-His-Cys3 motif in the plant homeodomain, *Zf - CXXC domain* the classical zf-CXXC domain that binds nonmethyl-CpG dinucleotides, *PWWP domain* named after its central core 'Pro-Trp-Trp-Pro'

maintenance activity of DNMTs, but by the passive loss of methylation during replication and active demethylation by the family of ten-eleven translocation (TET1–3) dioxygenases (Pastor et al. 2013). Demethylation by TETs is critical to ensure the flexibility of methyl marks whose responsive nature is contingent on their reversibility. These enzymes oxidise 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) that can be further converted to 5-formyl-cytosine (5fC) and 5-carboxyl-cytosine (5caC). Recent findings, including genome-wide mapping of 5hmC at a single-base resolution in mammalian brain (Lister et al. 2013), reveal a complex picture consistent with the idea that both 5mC and 5hmC can act as independent epigenetic marks. In

addition to their role as methylcytosine dioxygenases, mammalian TETs perform other functions, including interactions with metabolic enzymes and other proteins, participation in transcriptional regulation, telomere elongation and conveying cellular signals (Pastor et al. 2013; Lu et al. 2014). TETs have also been implicated in histone modifications. For example, TET2 promotes histone O-GlcNAcylation during TET2-dependent gene transcription by recruiting O-GlcNAc transferase to chromatin (Chen et al. 2013).

Yet, while active DNA demethylation has been relatively well investigated in mammals, little is known about this process in invertebrates. Recent work in *A. mellifera* has highlighted that a conserved TET enzyme in this organism is capable of converting methylated cytosine residues to 5hmC, as in mammals, and is likely to play an important role in establishing DNA methylation patterns, although the exact nature of this role remains unclear (Wojciechowski et al. 2014). In contrast to mammals that have three TET paralogs, only one TET has been found in *A. mellifera* and in most of other invertebrates. TET proteins are absent in organisms such as nematodes, which have lost the genes encoding DNMT1 and DNMT3, as well as a tRNA methylating enzyme (known as DNMT2). One puzzling feature of the honey bee TET is its very high level of expression, especially in the brain, that does not translate into a substantial number of 5hmCs. In most tissues fewer than 3000 5mhCs have been detected corresponding to 4–5% of the methylated cytosines. Only in testes and ovaries the levels of 5hmC in *A. mellifera* (7–10%) appear to be more comparable to those in some mammalian systems. Given this discrepancy between TET expression and the scarcity of 5hmC in the honey bee, it is likely that this protein performs other roles not related to cytosine demethylation. Alternatively, the honey bee TET may be very efficient in converting 5hmC into 5fC and then into 5caC, which then could be removed by terminal deoxynucleotidyl transferase. In this scenario, little 5hmC would be found despite high level of TET expression and activity.

An attempt to produce a specific honey bee antibody to better understand this incongruity yielded an unexpected binding to an unfamiliar nuclear structure in intrinsic neurons (Kenyon cells) in brains of adult honey bees (Hurd et al. 2021). Although the relevance of this structure to AmTET has not been established, its intriguing novelty suggests that a possible link to another TET functionality cannot be ruled out.

A broader functionality of invertebrate TETs is consistent with its expression in *Drosophila* that has no DNA methylation toolkit and does not require demethylation of genomic cytosines. It is noteworthy that a few studies have reported the presence of low levels of highly localised and asymmetrical methylated cytosines in *Drosophila* genome (Takayama et al. 2014; Capuano et al. 2014). If confirmed, these results would hint towards a novel enzymatic machinery capable of modifying DNA in certain contexts.

A study in *Drosophila* has shown that removal of the TET gene increases N6-methyladenie (m6A) levels in DNA (chapter reference Adenine methylation), but has no effects on RNA in which m6A is common (Zhang et al. 2015). *Drosophila* TET null mutants show a lethal phenotype beginning at the pupal stage with the last survivors dying three days post-eclosion. This finding suggests that a single relative of TETs in insects has the potential to catalyse a variety of chemical modifications not necessarily involving 5mC. Even more intriguing is a new study showing that in *Drosophila,* TET has the capacity to add hydroxymethylcytosine to RNA and that this modification is most prominent in the brain where it is important for promoting the translation of mRNAs (Delatte et al. 2015). Whether or not a similar TET activity exists in other insects remains to be determined, but it is conceivable that high levels of TET transcripts in *A. mellifera* brain are needed for RNA hydroxymethylation. Further exploration of demethylation and hydroxymethylation dynamics is required to fully understand both the role of TET in DNA/RNA methylation patterning and in other hitherto unidentified functions amongst invertebrate organisms.

How are DNA methyltransferases and TET dioxygenases guided to a given genomic region? Recent models indicate that these enzymes are recruited to specific genomic locations via their interactions with other epigenetic modifiers, and that these interactions are dependent on the chromatin environment (Jeltsch and Jurkowska 2014; Jones and Liang 2009). Work in mammals, for instance, has highlighted that various histone post-translational modifications (PTMs) and chromatin remodelling factors are involved in recruiting DNMTs to a genomic location (Jones and Liang 2009). It is likely that similar mechanisms operate in insects, but until these are explored in the context of differential DNA methylation, the role these factors play in establishing DNA methylation patterns will remain unknown. To this end, extensive PTMs have been described in *A. mellifera* using mass spectrometry (Dickman et al. 2013), and genome-wide chromatin immunoprecipitation-based approach (ChiPseq) has been applied to map specific histone PTMs in the ant *C. floridanus* (Simola et al. 2013) and in the honey bee (Wojciechowski et al. 2018). Although the exact recruitment mechanism of DNA methylation and demethylation machinery in insects is still unknown, the distribution of methylated cytosines across the genome of *A. mellifera* has been mapped at single base resolution, providing initial impetus for studies on how DNA methylation might be contributing to a variety of cellular processes.

7.4.4 DNA Methylation Patterns Across Invertebrates

The targeting of DNA methylation across the genome appears to take on two distinct types of patterns: global DNA methylation and mosaic DNA methylation (Suzuki and Bird 2008). In the case of global DNA methylation, which is typical of vertebrates, the genome is densely methylated throughout most tissues and developmental stages. In humans, for instance, the majority of CpG sites across the genome are methylated with the exception of small unmethylated domains, typically near regulatory regions such as promoters, termed CpG islands (Ehrlich et al. 1982). In contrast, in invertebrate genomes, distinct domains are methylated and unmethylated in a 'mosaic' fashion (Tweedie et al. 1997; Suzuki and Bird 2008). The overall level of this mosaic methylation varies; for instance, *A. mellifera* presents with a much lower level of methylation when compared to invertebrates such as *Ciona intestinalis* (Feng et al. 2010). However, several features of DNA methylation patterning are conserved.

Being an important epigenetic mark, DNA methylation has long been associated with the transcriptional silencing of genomic regions. Across a wide range of organisms it is frequently targeted to repetitive elements and transposons; *A. thaliana*, together with the mouse and zebrafish all exhibit such methylation, and disruption of DNA methylation leads to the reactivation of these elements, showing that the methylation of these regions is critical to their silencing (Feng et al. 2010; Bourc'his and Bestor 2004; Kato et al. 2003). Additionally, genomic imprinting, whereby a gene or chromosomal region is transcriptionally controlled in a parent-of-origin manner, is frequently associated with DNA methylation and gene silencing. In mammals and plants imprinted genes are differentially methylated and disruption of DNA methyltransferase activity leads to the aberrant expression of the maternal and/or paternal transcript (Li et al. 1993; Jullien et al. 2006).

Yet, while it is clear that DNA methylation has been utilised by many organisms to maintain genome stability and regulate imprinted regions, this does not extend to all species. Whereas moderate levels of methylation have been observed across transposons in *C. intestinalis*, other invertebrate genomes do not exhibit such methylation; in *A. mellifera*, for example, transposons are not targeted for methylation (Zemach et al. 2010; Feng et al. 2010). However, transposons are not as frequent in *A. mellifera* as they are in some other species, suggesting that a mechanism controlling mobile genomic elements may be less critical for this organism. Additionally, not all species utilise DNA methylation to establish

imprints; organisms such as *C. elegans* do not have a functional methylation system, yet still contain imprinted genes, and there is no current evidence for genomic imprinting via DNA methylation in *A. mellifera* (Sha and Fire 2005). It thus appears that while DNA methylation can be critical for transcriptional silencing in many contexts, this is not always the case.

The better-conserved feature of eukaryotic DNA methylation is the distribution of methylation at gene bodies. Gene body methylation has been identified across numerous organisms, and it is likely that its occurrence predates the last common ancestor of plants and animals (Feng et al. 2010). It is characterised by methylation across the introns and exons of protein-coding genes. In vertebrates and plants, this gene body methylation is prevalent and has been shown to frequently occur in the regions that are highly expressed (Ball et al. 2009; Zilberman et al. 2007).

In invertebrates, DNA methylation predominantly occurs across gene bodies. The vase tunicate *C. intestinalis*, the parasitoid wasp *Nasonia vitripennis,* the carpenter ant *Camponotus floridanus*, *A.mellifera* and lower *Metazoa* such as the sea anemone *Nematostella vectensis,* all display high levels of gene body methylation, and the regions with high gene body methylation also correlate with genomic regions that are actively transcribed (Tweedie et al. 1997; Zemach et al. 2010; Feng et al. 2010; Wang et al. 2013; Foret et al. 2009; Bonasio et al. 2012). To date, in all insects in which methylation has been analysed on the genome-wide scale, 5mC appears to be limited almost exclusively to CpG dinucleotides with only marginal levels found at non-CpG sites, and very few CpGs methylated asymmetrically (Welsh et al. 2017). Whether this pattern is a universal feature of all insects awaits sequencing of methylomes in more species representing diverse evolutionary lineages. In *A. mellifera,* intragenic DNA methylation is higher in exons than introns and those genes, which are highly methylated are often ubiquitously expressed, with evidence that this type of methylation modulates the expression of these transcripts (Foret et al. 2009; Lyko et al. 2010; Wedd et al. 2016). These common findings suggest that intragenic DNA methylation is associated with active transcription and have important implications for understanding how this mark might direct transcription.

The correlation between intragenic DNA methylation and the expression of 'housekeeping' genes has led to the suggestion that DNA methylation functions by preventing spurious transcription. By methylating intragenic regions, the initiation of transcription at cryptic sites within coding regions can be prevented (Weber and Schubeler 2007). For frequently transcribed genes, such as the ubiquitously expressed housekeeping genes, this would represent an important mechanism, whereby transcription can be controlled and transcriptional noise reduced (Zilberman et al. 2007; Suzuki et al. 2007). This hypothesis is supported by a recent work, which demonstrated that gene body methylation is negatively correlated with transcriptional noise (Huh et al. 2013). However, alternative suggestions, based on work in *A. thaliana*, indicate that it is the rate of transcription of these moderately expressed transcripts that induces intragenic methylation, as opposed to methylation itself influencing transcription (Zilberman et al. 2007).

Yet, given that key differences in the DNA methylation system exist between *A. thaliana* and other eukaryotic model organisms, this relationship between intragenic methylation and the rate of transcription may not be universal. DNA methylation patterns in *A. thaliana* and other flowering plants are established via an RNA-directed DNA methylation system (RdDM), where RNA molecules, small interfering RNAs (siRNAs), guide de novo methylation (Chan et al. 2005; Bao et al. 2004). The proposed model that transcription itself causes intragenic methylation, is based upon the spurious transcription of these siRNAs (Zilberman et al. 2007). Amongst invertebrate models, there is no current evidence for an RdDM-like system and therefore the intragenic DNA methylation seen in these species may have alternate functions.

Indeed, several lines of evidence indicate a number of functional roles for intragenic DNA methylation in mammals, including the regulation

of non-coding RNAs and transcription elongation. Both microRNAs and long non-coding RNA (lncRNA) transcripts have been shown to be influenced by intragenic DNA methylation, with important transcriptional consequences (Kulis et al. 2013; Cheung et al. 2011; Lujambio et al. 2007). By altering chromatin structure, DNA methylation can reduce the efficiency of transcription elongation (Lorincz et al. 2004). Intragenic DNA methylation has been shown to influence elongation; exonic differential methylation modulates binding of the CTCF transcription factor, altering RNA Pol II processing and the alternative splicing of a transcript (Shukla et al. 2011). Further evidence that intragenic DNA methylation modulates alternative splicing (Maunakea et al. 2013; Foret et al. 2012) and leads to the cell- and tissue-specific expression of alternative transcripts (Maunakea et al. 2010) across a number of species suggests that this function of intragenic DNA methylation is commonplace, and likely to have important biological consequences (Kulis et al. 2013).

In *A. mellifera,* differential intragenic DNA methylation has similarly been correlated to alternative splicing. High intragenic DNA methylation has been observed near alternatively spliced sites, and the differential methylation of these regions has been linked to the expression of condition-specific alternatively spliced transcript variants (Table 7.2) (Lyko and Maleszka 2011; Foret et al. 2012; Kucharski et al. 2016). Some of these cases support the Shukla et al. (2011) model, whereby methylation inhibits binding of a CTCF factor that can promote inclusion of weak upstream exons by mediating local RNA polymerase II pausing. In the case of Anaplastic Lymphoma Kinase (Table 7.2), low methylation correlates with exon 25 inclusion at high frequency. In the adult brain, where the methylated region shows much lower methylation relative to larvae, the majority of the available ALK transcripts (81%) have been found to contain exon 25 (Foret et al. 2012). Thus, differential methylation of this sequence seems to affect the dynamics of exon 25 inclusion and generates ALK isoforms with different amino termini in the intracellular domain that could interact with distinct partners. This suggests that the DNA methylation seen across the genes of *A. mellifera* can control the choice of context-dependent transcript variants.

7.4.5 Does Gene Body Methylation Direct Gene Expression in Insects?

Although several clear relationships between DNA methylation and transcription have emerged over the past decades, including the link between intragenic DNA methylation and active transcription, this relationship is not entirely straightforward. While there are a number of cases where DNA methylation has been shown to directly influence transcription, for instance Shukla et al. (2011) provide a direct relationship between methylation and alternative splicing in the mouse model, the situation in invertebrates is less clear. In the case of invertebrate models, the few examples seen in *A. mellifera* indicate that intragenic DNA methylation is likely to play an important role in eliciting transcriptional changes. Yet, these studies also highlight that such changes are context-dependent and do not necessarily eliminate the possibility that they result as a function of other processes that are simply associated with differential DNA methylation.

For some time it has been acknowledged that DNA methylation does not stand alone in directing transcriptional changes, but is coupled with other epigenetic marks, such as histone modifications (Fuks 2005; Cedar and Bergman 2009). These epigenetic modifications and DNA methylation itself are also influenced by genetic variation (Kilpinen et al. 2013; McVicker et al. 2013; Kasowski et al. 2013; Furey and Sethupathy 2013). The complexity of these interactions makes it difficult to ascertain whether DNA methylation itself directs transcriptional changes, or if other factors, such as the underlying DNA sequence, play a more substantive role, and this has become an important avenue of research across the epigenetics field (Schubeler 2012).

Table 7.2 Examples of genes with cassette exons whose expression correlates with differential methylation

Gene Id	Exon	Effect on protein	Putative function
GB43824 Trans-membrane lysoplasmalogenase (Lyko et al. 2010)	3	Creates a truncated protein by introducing a premature STOP codon	In a heterodimer, the truncated protein is predicted to inhibit trans-membrane activity of the full-length protein
GB43446 Anaplastic lymphoma kinase (ALK) (Foret et al. 2012)	25	Creates a different interacting domain by changing the N-terminal sequence of ALK	Modifies selection of protein partners interacting with ALK
GB43543 Glycine Receptor (Foret et al. 2012)	8	Extends the intracellular TM3–4 loop by 36 bp (12aa)	New gating properties of the glycine channel with the insertion. The cassette exon is found mostly in GlyR expressed in the brain and sensory organs

Cis- and *trans*-acting polymorphisms have been shown to lead to differentially methylated regions across a number of organisms (Schubeler 2012; Richards 2006). In mammals, these obligatory methylated epialleles have been well-documented; numerous studies have established linkage between allelic variation and a methylated state across many cell types (Kerkel et al. 2008; Schilling et al. 2009; Shoemaker et al. 2010). Importantly, epiallelic variation is a key contributor to phenotypic variation, where these epialleles influence transcription across different cellular contexts (Schubeler 2012; Gutierrez-Arcelus et al. 2013). In *A. mellifera* and other invertebrate models, although DNA methylation is critical, there has been little investigation into how genetic variability might be contributing to differential DNA methylation.

In a recent study (Wedd et al. 2015), differential intragenic DNA methylation of the gene encoding for lysosomal-alpha mannosidase (LAM) was correlated with sequence variation, providing the first evidence for an obligatory methylated epiallele within the *A. mellifera* population (Fig. 7.3). In contrast to pure epialleles that arise from stochastic genotype-independent events, obligatory differentially methylated epialleles are generated by sequence variants (Richards 2006). The methylated status of the LAM epiallele in *A. mellifera* has been found to increase LAM expression in a context-specific manner, along with the expression of a long non-coding RNA transcript. LAM epialleles are inherited in the classical Mendelian manner with no apparent evidence of imprinting (Wedd et al. 2016). This result not only further supports the link between intragenic DNA methylation and active transcription, but also indicates that obligatory epialleles in *A. mellifera* will likely contribute to phenotypic variation in a context-specific manner, as has been demonstrated across other species.

Given that phenotypic differences in *A. mellifera* are generated by tightly controlled epigenetic changes, any impact that genetic variation might have on the epigenetic layer of information could be profound. In the context of DNA methylation, substantial work in social insects like *A. mellifera* has shown that differential DNA methylation patterns correlate with transcriptional changes and phenotypic variation, but fewer studies interpret these changes in the context of any underlying sequence variation. The discovery of obligatory epialleles in the *A. mellifera* population highlights the importance of interpreting differential DNA methylation patterns more carefully and investigating the extent to which epiallelic variation influences phenotype in this organism. The availability of ultra-deep next-generation bisulfite sequencing technologies, as used in the LAM study, will facilitate interpreting these patterns accurately in the context of developmental, tissue-specific, and stochastic effects that are known to influence methylation patterns (Wagner et al. 2014; Landan et al. 2012).

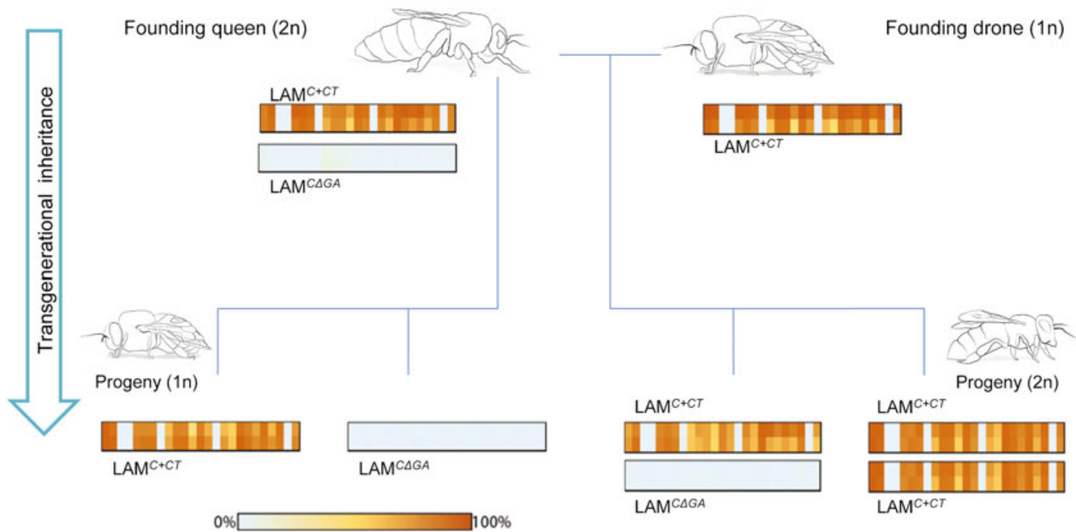


Fig. 7.3 Transgenerational inheritance of obligatory methylated epialleles of the gene coding for lysosomal alpha-mannosidase (LAM) in *A. mellifera*. A single male drone was mated with a queen; haploid drones develop from unfertilised eggs and diploid workers/queens develop from fertilised eggs. The level of methylation of the two alleles, LAM^{C+CT} and $LAM^{C\Delta GA}$, across an intragenic region of LAM (exons 16 to exons 18) is indicated; 0% indicating a CpG site was never found to methylated, 100% indicating that a CpG site was always found to be methylated. The transgenerational inheritance of methylation patterns, irrespective of factors such as caste or developmental stage and without parent-of-origin effects, is indicative of an obligatory methylated epiallele (see Wedd et al. 2016 for more details)

7.5 Conclusion

Invertebrates, such as the honey bee *A. mellifera,* represent an important model from which a broader understanding of DNA methylation, and its role in directing transcription, can be drawn. Importantly, this is reliant on the accurate interpretation of insect methylomes, and without performing in-depth analyses of differential methylation patterns in *A. mellifera* there is a risk of misconstruing the biological significance of such marks. Equally important is advancing efforts to unravel mechanistic properties of non-mammalian DNMTs within living organisms, and to comprehend the molecular role of their various isoforms. To achieve a full understanding of how this epigenetic modification links genotype to phenotype, the relationship between DNA methylation patterns, genetic variability and other epigenetic marks must be understood and interpreted in the context of additional factors, such as developmental and tissue-specific effects, and environmental influences.

References

Ashby R, Foret S, Searle I, Maleszka R (2016) MicroRNAs in honey bee caste determination. Sci Rep 6:18794

Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH et al (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells (vol 27, pg 361, 2009). Nat Biotechnol 27(5):485

Bao N, Lye KW, Barton MK (2004) MicroRNA binding sites in Arabidopsis class IIIHD-ZIP mRNAs are required for methylation of the template chromosome. Dev Cell 7(5):653–662

Bernstein BE, Meissner A, Lander ES (2007) The mammalian epigenome. Cell 128(4):669–681

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A et al (2010) The NIH roadmap Epigenomics mapping consortium. Nat Biotechnol 28(10):1045–1048

Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16(1):6–21

Bonasio R, Li QY, Lian JM, Mutti NS, Jin LJ, Zhao HM et al (2012) Genome-wide and caste-specific DNA Methylomes of the ants Camponotus floridanus and Harpegnathos saltator. Curr Biol 22(19):1755–1764

Bourc'his D, Bestor TH (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature 431(7004):96–99

Buttstedt A, Mureşan CI, Lilie H, Hause G, Ihling CH, Schulze SH et al (2018) How honeybees defy gravity with Royal Jelly to raise Queens. Curr Biol 28(7): 1095–100.e3

Capuano F, Mulleder M, Kok R, Blom HJ, Ralser M (2014 Apr 15) Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. Anal Chem 86(8):3697–3702. https://doi.org/10.1021/ac500447w. Epub 2014 Mar 25

Cedar H, Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet 10(5):295–304

Chan SWL, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat Rev Genet 6(5):351–360

Chen Q, Chen Y, Bian C, Fujiki R, Yu X (2013) TET2 promotes histone O-GlcNAcylation during gene transcription. Nature 493(7433):561–564

Cheung HH, Davis AJ, Lee TL, Pang AL, Nagrani S, Rennert OM et al (2011) Methylation of an intronic region regulates miR-199a in testicular tumor malignancy. Oncogene 30(31):3404–3415

Clamp M, Fry B, Kamal M, Xie XH, Cuff J, Lin MF et al (2007) Distinguishing protein-coding and noncoding genes in the human genome. P Natl Acad Sci USA 104(49):19428–19433

Delatte B, Jeschke J, Defrance M, Bachman M, Creppe C, Calonne E et al (2015) Genome-wide hydroxymethylcytosine pattern changes in response to oxidative stress. Sci Rep 5:12714

Dickman MJ, Kucharski R, Maleszka R, Hurd PJ (2013) Extensive histone post-translational modification in honey bees. Insect Biochem Mol Biol 43(2):125–137

Duymich CE, Charlet J, Yang X, Jones PA, Liang G (2016) DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. Nat Commun 7(1):11453

Ehrlich M, Gamasosa MA, Huang LH, Midgett RM, Kuo KC, Mccune RA et al (1982) Amount and distribution of 5-Methylcytosine in human DNA from different types of tissues or cells. Nucleic Acids Res 10(8): 2709–2721

Elsik CG, Worley KC, Bennett AK, Beye M, Camara F, Childers CP et al (2014) Finding the missing honey bee genes: lessons learned from a genome upgrade. BMC Genomics 15:86

Erwin DH, Davidson EH (2009) The evolution of hierarchical gene regulatory networks. Nat Rev Genet 10(2): 141–148

Feng SH, Cokus SJ, Zhang XY, Chen PY, Bostick M, Goll MG et al (2010) Conservation and divergence of methylation patterning in plants and animals. P Natl Acad Sci USA. 107(19):8689–8694

Foret S, Kucharski R, Pittelkow Y, Lockett GA, Maleszka R (2009) Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. BmcGenomics 10:472

Foret S, Kucharski R, Pellegrini M, Feng SH, Jacobsen SE, Robinson GE et al (2012) DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proc Natl Acad Sci U S A 109(13):4968–4973

Fuks F (2005) DNA methylation and histone modifications: teaming up to silence genes. Curr Opin Genet Dev 15(5):490–495

Furey TS, Sethupathy P (2013) Genetics driving epigenetics. Science (New York, NY) 342(6159): 705–706

Gao L, Emperle M, Guo Y, Grimm SA, Ren W, Adam S et al (2020) Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. Nat Commun 11(1):3355

Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. Annu Rev Biochem 74:481–514

Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B et al (2014) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat Neurosci 17(2):215–222

Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A et al (2013) Passive and active DNA methylation and the interplay with genetic variation in gene regulation. elife 2:e00523

Haig D (2004) The (dual) origin of epigenetics. Cold Spring Harb Sym 69:67–70

Haig D (2012) Commentary: the epidemiology of epigenetics. Int J Epidemiol 41(1):13–16

Halfmann R, Lindquist S (2010) Epigenetics in the extreme: prions and the inheritance of environmentally acquired traits. Science 330(6004):629–632

Hattori N, Nomoto H, Fukumitsu H, Mishima S, Furukawa S (2007) Royal jelly and its unique fatty acid, 10-hydroxy-trans-2-decenoic acid, promote neurogenesis by neural stem/progenitor cells in vitro. Biomed Res 28(5):261–266

Holliday R, Pugh JE (1975) DNA modification mechanisms for control of gene activity during development. Heredity 35(Aug):149

Huh I, Zeng J, Park T, Yi SV (2013) DNA methylation and transcriptional noise. Epigenet Chromatin 6:9

Hurd PJ, Grübel K, Wojciechowski M, Maleszka R, Rössler W (2021) Novel structure in the nuclei of honey bee brain neurons revealed by immunostaining. Sci Rep 11(1):6852

Jablonka E, Lamb MJ (2002) The changing concept of epigenetics. From Epigenesis to Epigenetics: The Genome in Context 981:82–96

Jablonka E, Lamm E (2012) Commentary: the epigenotype-a dynamic network view of development. Int J Epidemiol 41(1):16–20

Jeltsch A, Jurkowska RZ (2014) New concepts in DNA methylation. Trends Biochem Sci 39(7):310–318

Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13(7):484–492

Jones PA, Liang GN (2009) OPINION rethinking how DNA methylation patterns are maintained. Nat Rev Genet 10(11):805–811

Jones PA, Takai D (2001) The role of DNA methylation in mammalian epigenetics. Science 293(5532): 1068–1070

Jullien PE, Kinoshita T, Ohad N, Berger F (2006) Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. Plant Cell 18(6):1360–1372

Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu YL et al (2013) Extensive variation in chromatin states across humans. Science (New York, NY) 342(6159):750–752

Kato M, Miura A, Bender J, Jacobsen SE, Kakutani T (2003) Role of CG and non-CG methylation in immobilization of transposons in arabidopsis. Curr Biol 13(5):421–426

Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E et al (2008) Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. Nat Genet 40(7):904–908

Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A et al (2013) Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. Science 342(6159): 744–747

Kouzarides T (2007) Chromatin modifications and their function. Cell 128(4):693–705

Kucharski R, Maleszka J, Foret S, Maleszka R (2008) Nutritional control of reproductive status in honeybees via DNA methylation. Science (New York, NY) 319(5871):1827–1830

Kucharski R, Maleszka J, Maleszka R (2016) A possible role of DNA methylation in functional divergence of a fast evolving duplicate gene encoding odorant binding protein 11 in the honeybee. Proc Biol Sci 283(1833): 20160558

Kulis M, Queiros AC, Beekman R, Martin-Subero JI (2013) Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. Bba-Gene Regul Mech 1829(11):1161–1174

Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A et al (2015) Integrative analysis of 111 reference human epigenomes. Nature 518(7539): 317

Landan G, Cohen NM, Mukamel Z, Bar A, Molchadsky A, Brosh R et al (2012) Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. Nat Genet 44(11):1207–1214

Li E, Beard C, Jaenisch R (1993) Role for DNA methylation in genomic imprinting. Nature 366(6453): 362–365

Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND et al (2013) Global epigenomic reconfiguration during mammalian brain development. Science (New York, NY) 341(6146):629–640

Lorincz MC, Dickerson DR, Schmitt M, Groudine M (2004) Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat Struct Mol Biol 11(11):1068–1075

Lu FL, Liu YT, Jiang L, Yamaguchi S, Zhang Y (2014) Role of Tet proteins in enhancer activity and telomere elongation. Genes Dev 28(19):2103–2119

Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR et al (2020) CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res 48(D1):D265–D2d8

Lujambio A, Ropero S, Ballestar E, Fraga MF, Cerrato C, Setien F et al (2007) Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. Cancer Res 67(4):1424–1429

Lyko F, Maleszka R (2011) Insects as innovative models for functional studies of DNA methylation. Trends Genet 27(4):127–131

Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R (2010) The honey bee epigenomes: differential methylation of brain DNA in Queens and workers. PLoS Biol 8(11):e1000506

Maleszka R (2014) The social honey bee in biomedical research: realities and expectations. Drug Discov Today Dis Models 12:7–13

Maleszka R (2016) Epigenetic code and insect behavioural plasticity. Current Opinion in Insect Science 15:45–52

Maleszka R (2018) Beyond Royalactin and a master inducer explanation of phenotypic plasticity in honey bees. Commun Biol 1:8

Maleszka R, Kucharski R (2022) Without mechanisms theories and models in insect epigenetics remain a black box. Trends Genet. https://doi.org/10.1016/j.tig.2022.05.004. S0168952522001123

Maleszka R, Mason PH, Barron AB (2014) Epigenomics and the concept of degeneracy in biological systems. Brief Funct Genomics 13(3):191–202

Mandacaru SC, do Vale LH, Vahidi S, Xiao Y, Skinner OS, Ricart CA et al (2017) Characterizing the structure and oligomerization of major Royal Jelly Protein 1 (MRJP1) by mass spectrometry and complementary biophysical tools. Biochemistry 56(11):1645–1655

Mattick JS, Taft RJ, Faulkner GJ (2010) A global view of genomic information--moving beyond the gene and the master regulator. Trends Genet 26(1):21–28

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD et al (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466(7303):253–U131

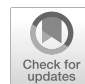
Maunakea AK, Chepelev I, Cui KR, Zhao KJ (2013) Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res 23(11):1256–1269

McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A et al (2013) Identification of genetic variants that affect histone modifications in human cells. Science 342(6159):747–749

Miklos GLG, Maleszka R (2011) Epigenomic communication systems in humans and honey bees: from molecules to behavior. Horm Behav 59(3):399–406
Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z et al (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448(7154):714–717
Pastor WA, Aravind L, Rao A (2013) TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol 14(6): 341–356
Regev A, Lamb MJ, Jablonka E (1998) The role of DNA methylation in invertebrates: developmental regulation or genome defense? Mol Biol Evol 15(7):880–891
Richards EJ (2006) Opinion - inherited epigenetic variation - revisiting soft inheritance. Nat Rev Genet 7(5): 395–3U2
Riggs AD (1975) X-inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet 14(1):9–25
Rona GB, Eleutherio ECA, Pinheiro AS (2016) PWWP domains and their modes of sensing DNA and histone methylated lysines. Biophys Rev 8(1):63–74
Schilling E, El Chartouni C, Rehli M (2009) Allele-specific DNA methylation in mouse strains is mainly determined by cis-acting sequences. Genome Res 19(11):2028–2035
Schubeler D (2012) Epigenetic Islands in a Genetic Ocean. Science (New York, NY) 338(6108):756–757
Sha K, Fire A (2005) Imprinting capacity of gamete lineages in Caenorhabditis elegans. Genetics 170(4): 1633–1652
Shoemaker R, Deng J, Wang W, Zhang K (2010) Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. Genome Res 20(7): 883–889
Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M et al (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 479(7371):74–79
Simola DF, Ye C, Mutti NS, Dolezal K, Bonasio R, Liebig J et al (2013) A chromatin link to caste identity in the carpenter ant Camponotus floridanus. Genome Res 23(3):486–496
Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. Nat Rev Genet 14(3): 204–220
Spannhoff A, Kim YK, Raynal NJM, Gharibyan V, Su MB, Zhou YY et al (2011) Histone deacetylase inhibitor activity in royal jelly might facilitate caste switching in bees. EMBO Rep 12(3):238–243
Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9(6):465–476
Suzuki MM, Kerr ARW, De Sousa D, Bird A (2007) CpG methylation is targeted to transcription units in an invertebrate genome. Genome Res 17(5):625–631
Takayama S, Dhahbi J, Roberts A, Mao G, Heo SJ, Pachter L et al (2014) Genome methylation in D. melanogaster is found at specific short motifs and is independent of DNMT2 activity. Genome Res 24(5): 821–830
Tweedie S, Charlton J, Clark V, Bird A (1997) Methylation of genomes and genes at the invertebrate-vertebrate boundary. Mol Cell Biol 17(3):1469–1475
Waddington CH (1942) The Epigenotype. Endeavour 1: 18–20
Waddington CH (1957) The strategy of the genes. A discussion of some aspects of theoretical biology. George Allen & Unwin, London
Wagner JR, Busche S, Ge B, Kwan T, Pastinen T, Blanchette M (2014) The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol 15(2):R37
Wang X, Wheeler D, Avery A, Rago A, Choi JH, Colbourne JK et al (2013) Function and evolution of DNA methylation in Nasonia vitripennis. PLoS Genet 9(10):e1003872
Wang H, Farnung L, Dienemann C, Cramer P (2020) Structure of H3K36-methylated nucleosome-PWWP complex reveals multivalent cross-gyre binding. Nat Struct Mol Biol 27(1):8–13
Weber M, Schubeler D (2007) Genomic patterns of DNA methylation: targets and function of an epigenetic mark. Curr Opin Cell Biol 19(3):273–280
Wedd L, Kucharski R, Maleszka R (2015) Differentially methylated obligatory epialleles modulate context-dependent LAM gene expression in the honey bee Apis mellifera. Epigenetics 11(1):1–10
Wedd L, Kucharski R, Maleszka R (2016) Differentially methylated obligatory epialleles modulate context-dependent LAM gene expression in the honeybee Apis mellifera. Epigenetics 11(1):1–10
Welsh L, Maleszka R, Foret S (2017) Detecting rare asymmetrically methylated cytosines and decoding methylation patterns in the honeybee genome. Royal Society open. Science 4(9):170248
Wojciechowski M, Rafalski D, Kucharski R, Misztal K, Maleszka J, Bochtler M et al (2014) Insights into DNA hydroxymethylation in the honeybee from in-depth analyses of TET dioxygenase. Open Biol 4(8):140110
Wojciechowski M, Lowe R, Maleszka J, Conn D, Maleszka R, Hurd PJ (2018) Phenotypically distinct female castes in honey bees are defined by alternative chromatin states during larval development. Genome Res 28(10):1532–1542
Xiang H, Zhu JD, Chen QA, Dai FY, Li X, Li MW et al (2010) Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. Nat Biotechnol 28(7):756
Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science (New York, NY) 328(5980):916–919
Zeng Y, Ren R, Kaur G, Hardikar S, Ying Z, Babcock L et al (2020) The inactive Dnmt3b3 isoform preferentially enhances Dnmt3b-mediated DNA methylation. Genes Dev 34(21–22):1546–1558

Zhang GQ, Huang H, Liu D, Cheng Y, Liu XL, Zhang WX et al (2015) N-6-Methyladenine DNA modification in drosophila. Cell 161(4):893–906

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet 39(1):61–69



N6-methyladenine: A Rare and Dynamic DNA Mark

8

Zach Klapholz O'Brown and Eric Lieberman Greer

Abstract

Chromatin, consisting of deoxyribonucleic acid (DNA) wrapped around histone proteins, facilitates DNA compaction and allows identical DNA code to confer many different cellular phenotypes. This biological versatility is accomplished in large part by post-translational modifications to histones and chemical modifications to DNA. These modifications direct the cellular machinery to expand or compact specific chromatin regions and mark certain regions of the DNA as important for cellular functions. While each of the four bases that make up DNA can be modified (Iyer et al., Prog Mol Biol Transl Sci. 101:25–104, 2011), this chapter will focus on methylation of the 6th position on adenines (6mA). 6mA is a prevalent modification in unicellular organisms and until recently was thought to be restricted to them. A flurry of conflicting studies have proposed that 6mA either does not exist, is present at low levels, or is present at relatively high levels and regulates complex processes in different multicellular eukaryotes. Here, we will briefly describe the history of 6mA, examine its evolutionary conservation, and evaluate the current methods for detecting 6mA. We will discuss the proteins that have been reported to bind and regulate 6mA and examine the known and potential functions of this modification in eukaryotes. Finally, we will close with a discussion of the ongoing debate about whether 6mA exists as a directed DNA modification in multicellular eukaryotes.

Z. K. O'Brown · E. L. Greer (✉)
Division of Newborn Medicine, Boston Children's Hospital, Boston, MA, USA

Department of Pediatrics, Harvard Medical School, Boston, MA, USA
e-mail: eric.greer@childrens.harvard.edu

Keywords

N6-methyladenine · N6-methyl-2′-deoxyadenosine · 6mA · 6mdA · m6dA · Directed DNA methylation · Epigenetics · MT-A70 · ALKB · METTL4 · ALKBH1 · ALKBH4

8.1 Introduction

DNA must faithfully transmit the blueprints of life from generation to generation. However, it is also necessary that different cell types have access to different portions of the genome, and that specific cell types can respond appropriately to changes in the environment. Such dynamic responses are mediated in part by transcription factor complexes, and by chemical modifications to chromatin. DNA is not as heavily modified as RNA, which has over 170 different modifications identified to date (Frye et al. 2018). The limited number of DNA modifications (relative to RNA) is presumably evolutionarily selected to protect

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_8

the DNA code from mutations, and to enable the formation of a stable double helix. Nevertheless, several DNA modifications occur in different species distributed across the tree of life and are important as both signals of DNA lesions and as epigenetic regulators of diverse biological processes. Importantly, DNA modifications increase the repertoire of cellular phenotypes that can be encoded by a single DNA sequence, without directly altering the integrity of the genetic code. Soon after DNA was discovered, variants of each base were identified. However, the role of DNA methylation in the context of normal biological processes and disease pathogenesis remains an active area of study.

Although 6mA was discovered in 1955 (Dunn and Smith 1955, 1958) soon after cytosine methylation (5mC) which was confirmed in 1950 (Johnson and Coghill 1925; Hotchkiss 1948; Wyatt 1950), 6mA was thought to exist predominantly in prokaryotes and was therefore not given the same amount of research attention in eukaryotes as 5mC. The discovery that 6mA exists in more recently evolved eukaryotes has revived interest in this DNA modification. To understand the dynamic regulation of and by adenine methylation, it is useful to view the role of 6mA across evolution. Here, we aim to provide a broad overview of the historical research on 6mA across the evolutionary spectrum and discuss the mechanisms by which N6-adenine methylation is established, reversed, and recognized. We examine the role of 6mA in biology, discuss the possibility of 6mA playing a functional role in multicellular eukaryotes as well as contradictory evidence regarding its existence, and summarize exciting areas for future research.

8.2 Types of DNA Modifications

Each DNA base is modified to varying degrees in different organisms. DNA methylation occurs either as non-enzymatic DNA damaging lesions or as directed modifications with signaling function, which are actively introduced by specific methyltransferase enzymes. DNA lesions include N1-methyladenine (1mA), N3-methyladenine (3mA), N7-methyladenine (7mA), N3-methylcytosine (3mC), N2-methylguanine (2mG), O6-methylguanine (6mG), N7-methylguanine (7mG), N3-methylthymine (3mT), and O4-methylthymine (4mT), while directed methylation includes N6-methyladenine (6mA), N4-methylcytosine (4mC), and C5-methylcytosine (5mC) (Sedgwick et al. 2007; Iyer et al. 2011; Grosjean 2009). Other DNA modifications include deaminated cytosines (Shapiro and Klein 1966; Lindahl and Nyberg 1974), oxidized derivatives of 5mC (5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC)) (Wyatt and Cohen 1952; Privat and Sowers 1996; Shen et al. 2014) and the hypermodified thymine base J (Gommers-Ampt et al. 1993). These modifications are discussed in greater detail in other reviews; we will focus on 6mA, a relatively uncharacterized DNA modification in eukaryotes with potential epigenetic function.

Of the directed DNA methylation events, 5mC is the most extensively studied. 5mC occurs at a higher frequency in more recently evolved organisms and its abundance in the genome ranges from 0.002% to 27% of cytosines, depending on the organism (Fig. 8.1). In mammals and plants, 5mC is the most abundant DNA modification (Iyer et al. 2011), and functions in the regulation of gene expression and maintenance of epigenetic memory (Bird 2002). 5mC in promoter regions typically leads to transcriptional gene silencing and therefore plays important roles in diverse cellular and developmental processes, including X-chromosome inactivation, genomic imprinting, stem cell pluripotency and differentiation (Bird 2002). Other directed DNA methylation events include 4mC and 6mA. 4mC has been identified mainly in thermophilic bacteria and archaea (Janulaitis et al. 1983; Ehrlich et al. 1985, 1987; Grosjean 2009; O'Brown et al. 2019). Until recently, 6mA was also thought to be restricted to bacteria, archaea, and protists. However, its recent identification in several eukaryotes raises the possibility that 6mA serves as an epigenetic

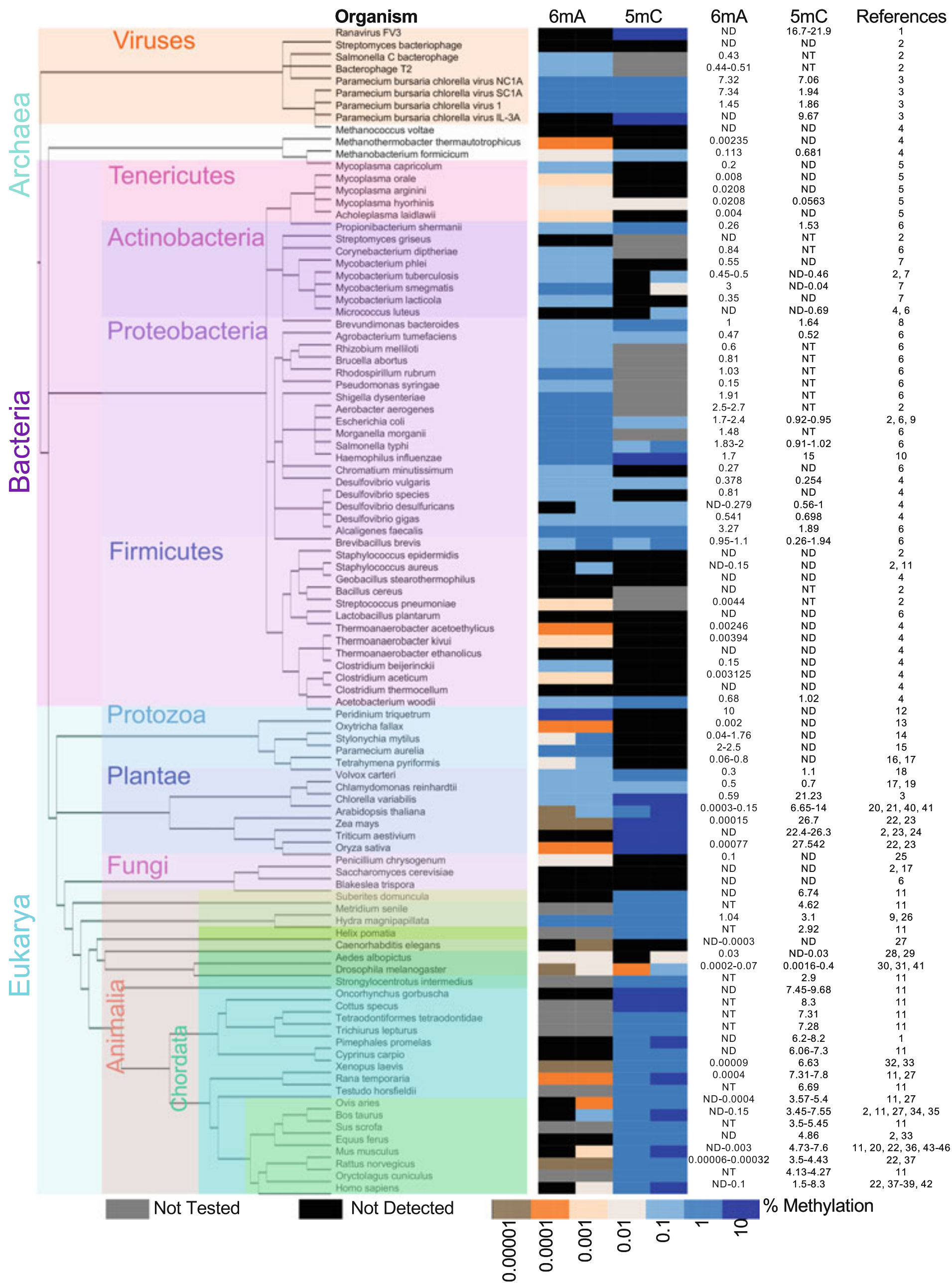


Fig. 8.1 Abundance of 6mA and 5mC across the tree of life. The relative abundance of 6mA and 5mC are displayed in a heat map. The first column of the heat map displays the percentage of adenines that are N6-methylated (%6mA/A) and the second column displays the percentage of cytosines that are C5-methylated (%5mC/C) for the organism indicated in each row. Blue color represents lower 6mA or 5mC abundance and red color represents higher 6mA or 5mC abundance. Gray color indicates that the methylation mark was not tested in that organism. Dark blue color indicates that the methylation mark was not detected in that organism, and therefore may or may not be present at levels below the limit of detection for the technique used. For some organisms, the level of methylation has been shown to vary across multiple measurements, between different studies or between different cell types within the same organism. In such cases, a range is presented where the left half of the column reflects the lowest detected level (or not detected in some cases) and the right half of the column shows the highest detected level. Methylation values are presented on the right along with citations.

signaling modification within an organism and potentially across generations.

8.3 Discovery of 6mA in Various Eukaryotes

DNA N6-methyladenine (6mA) is a widespread modification in prokaryotes. Although 6mA is not necessary for viability in prokaryotes (Marinus and Morris 1973; Russell and Hirata 1989), it plays crucial roles in regulating DNA replication (Campbell and Kleckner 1990; Yamaki et al. 1988), repair (Pukkila et al. 1983), transposition (Roberts et al. 1985), transcription (Wallecha et al. 2002; Robbins-Manke et al. 2005), and cellular defense (Luria and Human 1952; Meselson and Yuan 1968; Linn and Arber 1968; Smith et al. 1972). For reviews on 6mA in prokaryotes, please see (Marinus and Lobner-Olesen 2014; Wion and Casadesus 2006; Murray 2002) and Chapter 2. At the time of 6mA discovery, an unknown base was initially identified in *E. coli* and, using several techniques, this base was compared to synthesized nucleotides to identify 6mA. Hydrolyzed bases were separated by two-dimensional paper chromatography in different solvents, ultraviolet absorption spectrum maximums and minimums were measured, and electrophoretic mobility of this unknown base all confirmed the detection of 6mA (Dunn and Smith 1955, 1958). The existence of 6mA was subsequently confirmed in a variety of different bacterial species (Vanyushin et al. 1968). These initial detection techniques were capable of detecting 6mA at ~0.01% of total adenines (Vanyushin et al. 1970). This detection limit, combined with the confounding presence of commensal symbionts, technical variability, tissue-specific differences, development/stage-specific variability, or subtle environmental effects on 6mA levels initially led to contradictory reports of the identification of 6mA in eukaryotes. Indeed, 6mA was reported by one group to occur in bull and human sperm (Unger and Venner 1966), but other groups were unable to replicate this result or detect 6mA in other metazoa (Dunn and Smith 1958; Vanyushin et al. 1970). 6mA was reported to occur in some unicellular eukaryotes including *Paramecium aurelia* (Cummings et al. 1974), *Stylonychia mytilus* (Ammermann et al. 1981), *Oxytricha fallax* (Rae and Spear 1978), *Chlorella variabilis* (Van Etten et al. 1985), *Tetrahymena pyriformis* (Gorovsky et al. 1973) and *Chlamydomonas reinhardi* (Hattman et al. 1978). Two reports also identified 6mA in multicellular eukaryotes, including the mosquito *Aedes albopictus* (Adams et al. 1979) and the sponge *Suberites domuncula* (Vanyushin et al. 1970). However, the detection

Fig. 8.1 (continued) The phylogenetic tree was generated using the PhyloT web server (http://phylot.biobyte.de/index.html) and visualized using the Interactive Tree Of Life web server (http://itol.embl.de/). The phylogenetic tree ("rooted" setting) displays the inferred evolutionary relationships between the indicated genera based on their genetic similarity (Letunic and Bork 2011). The tree was created using FigTree v1.4.2. The different organisms are subdivided into different colored boxes to represent different kingdoms and phyla. For some phyla only one organism has been examined. 1: (Willis and Granoff 1980), 2: (Dunn and Smith 1958), 3: (Van Etten et al. 1985), 4: (Ehrlich et al. 1985), 5: (Razin and Razin 1980), 6: (Vanyushin et al. 1968), 7: (Srivastava et al. 1981), 8: (Degnen and Morris 1973), 9: (Yuki et al. 1979), 10: (Drozdz et al. 2012), 11: (Vanyushin et al. 1970), 12: (Rae 1976), 13: (Rae and Spear 1978), 14: (Ammermann et al. 1981), 15: (Cummings et al. 1974), 16: (Gorovsky et al. 1973), 17: (Hattman et al. 1978), 18: (Babinger et al. 2001), 19: (Fu et al. 2015), 20: (Capuano et al. 2014), 21: (Kakutani et al. 1999), 22: (Huang et al. 2015), 23: (Wagner and Capesius 1981), 24: (Montero et al. 1992), 25: (Rogers et al. 1986), 26: (Hassel et al. 2010), 27: (O'Brown et al. 2019), 28: (Adams et al. 1979), 29: (Proffitt et al. 1984), 30: (Zhang et al. 2015), 31: (Lyko et al. 2000), 32: (Koziol et al. 2016), 33: (Jabbari et al. 1997), 34: (Unger and Venner 1966), 35: (Romanov and Vanyushin 1981), 36: (Wu et al. 2016), 37: (Gama-Sosa et al. 1983), 38: (Tawa et al. 1992), 39: (Ehrlich et al. 1982), 40: (Liang et al. 2018), 41: (Kong et al. 2022), 42: (Xie et al. 2018), 43: (Yao et al. 2017), 44: (Douvlataniotis et al. 2020), 45: (Hao et al. 2020), 46: (Schiffers et al. 2017)

of 6mA in mosquitos was not reproduced (Proffitt et al. 1984), and its detection in the sponge was dismissed as potentially coming from symbiotic prokaryotes or algae (Vanyushin et al. 1970). Therefore, until recently, 6mA was thought to be restricted to prokaryotes and unicellular eukaryotes (Casadesus and Low 2006).

With the advent of more sensitive detection techniques (discussed below), 6mA has been identified in multicellular eukaryotes including *Caenorhabditis elegans* and *Drosophila melanogaster* (Greer et al. 2015; Zhang et al. 2015). Several other papers reported low levels of 6mA in more recently evolved eukaryotes, but each of these has caveats that we must acknowledge. 6mA was detected in *Drosophila,* calf thymus, and human placental samples by dot blots (Achwal et al. 1983). 6mA was also detected by immunofluorescence in mouse heart tissues (Sun et al. 2015). Another group identified 6mA in the plants *Oryza sativa* and *Zea mays,* rat tissues, and human cells by high-performance liquid chromatography coupled with mass spectrometry (HPLC-ms/ms) (Huang et al. 2015). Furthermore, 6mA was found by dot blots, HPLC, and methyl DNA immunoprecipitation followed by sequencing (MeDIPseq) in *Xenopus laevis* and mouse kidney (Koziol et al. 2016), and by dot blots, MeDIPseq, HPLC and SMRT-seq in mouse embryonic stem (ES) cells (Wu et al. 2016). A number of studies have also reported 6mA occurring in human cell lines as well as in human tissues (Xiao et al. 2018; Xie et al. 2018; Pacini et al. 2019; Hao et al. 2020). While these papers raise the exciting possibility that 6mA may indeed be present across the tree of life, it is difficult to discount potential contaminating microbiota and to confirm that the detection of 6mA is real when the reported levels of 6mA are at the limit of detection. In fact, several studies have reported that detected 6mA in each of these multicellular eukaryotes is the consequence of artifacts introduced during tissue or genomic DNA (gDNA) sample preparation, or methodological flaws in 6mA detection or mapping techniques (O'Brown et al. 2019; Schiffers et al. 2017; Liu et al. 2017; Douvlataniotis et al. 2020; Musheev et al. 2020; Lentini et al. 2018). RNA m6A (discussed below) could also account for contaminating signal in dot blots and immunofluorescence if not properly removed. It has been proposed that the presence of 6mA in genomic DNA is unlikely because injection of N6-adenine methylated oligos into mice induces a greater immune response than unmethylated oligos, as measured by the production of IL-12 (Tsuchiya et al. 2005). But this does not necessarily confirm that 6mA is a foreign base in mice, as unmethylated CpG motifs also induce a more substantial immune response (Tsuchiya et al. 2005). These results raise the possibility that 6mA is either not present in mammals, or present in sufficiently small quantities to keep it as an immunogenic species in the mammalian repertoire. To confirm the existence of 6mA across eukaryotes, it will be necessary to identify the enzymes that regulate 6mA and specific biological conditions under which the modification changes.

The studies suggesting that 6mA might be a conserved DNA modification raise several fundamental and largely unexplored questions about the evolutionary importance of 6mA across the tree of life. From an evolutionary perspective, why did higher eukaryotes shift from 6mA (the most pervasive DNA modification in prokaryotes), toward using 5mC as the more dominant DNA modification? To what extent are the ancient functions of 6mA and its modifying enzymes conserved from prokaryotes to more recent eukaryotes?

In contrast to DNA adenine methylation, RNA adenine methylation (m6A) has long been recognized as the most abundant post-transcriptional modification of prokaryotic and eukaryotic mRNAs (Niu et al. 2013). In humans, there are over 18,000 m6A sites representing approximately 7,000 unique mRNA transcripts (Jia et al. 2011; Meyer et al. 2012; Dominissini et al. 2012). Furthermore, m6A is enriched in 3'UTRs in highly conserved regions (Meyer et al. 2012; Dominissini et al. 2012; Deng et al. 2015), suggesting a shared function for m6A in evolutionarily distant species. N6-methyladenosine regulates multiple aspects of RNA metabolism, including mRNA stability/

decay, translation, splicing, and localization (Wang et al. 2014, 2015; Zhou et al. 2015; Niu et al. 2013), and participates in diverse cellular and biological processes including meiosis and embryonic stem cell differentiation (Yue et al. 2015; Batista et al. 2014; Hongay and Orr-Weaver 2011; Bodi et al. 2012). The prevalence of RNA m6A raises the possibility that DNA adenine methylation could be a consequence of methylated adenines in RNA recycled via the nucleotide salvage pathway. Another possibility is that DNA adenine methylation is catalyzed by RNA methyltransferases, either as an off-target effect of these enzymes or as a biologically regulated process. Unlike the better-characterized RNA m6A, relatively little is known about the functional importance of DNA 6mA in metazoan genomes, and whether 6mA plays a similarly conserved role in the dynamic regulation of biological processes. The effects that RNA m6A have on RNA structure and function might provide clues to the roles of N6-adenine methylation on DNA.

8.4 Abundance of 6mA

The relative genomic abundance of 6mA can provide clues to its biological function across evolutionarily distinct organisms. 6mA and 5mC appear to have a large range of abundance in the genomes of different organisms across evolution (Gommers-Ampt and Borst 1995). 5mC is undetectable in many bacterial species, as well as the genome of *S. cerevisiae,* and ranges from 0.0016% of cytosines in *D. melanogaster* to as high as 10% in some mammals and 30% in certain plant species (Gommers-Ampt and Borst 1995; Capuano et al. 2014; Wagner and Capesius 1981). If we accept that published literature documenting the presence of 6mA in different organisms is in fact detecting 6mA in the reported organism (rather than in contaminating symbionts or technical artifacts), the genomic abundance of 6mA varies by several orders of magnitude across the tree of life as well (Fig. 8.1). Generally, organisms with higher levels of 6mA such as bacteria and single-celled eukaryotes tend to have lower levels of 5mC, while organisms with higher levels of 5mC such as plants and mammals tend to have lower levels of 6mA. The detected level of 6mA ranges from ~0.0001 to 0.0003% of adenines in plants and mammals to as high as 3% of adenines in some species of bacteria, and up to 10% of adenines in the dinoflagellate *Peridinium triquetrum* (Rae 1976). Early studies of nucleic acid composition in the 1950s examined the base composition of DNA in different strains of bacteria using 2D paper chromatography (Dunn and Smith 1958). It was found that 6mA comprised 1.75% of all adenines in *E. coli* and 2.5% of adenines in *Aerobacter aerogenes* (Dunn and Smith 1958). Subsequent studies examined the content of 6mA in the DNA of unicellular eukaryotes, such as the ciliate *Tetrahymena pyriformis* (0.65–0.8% of adenines) (Gorovsky et al. 1973), *Paramecium aurelia* (2.5%) (Cummings et al. 1974), and *Stylonychia mytilus* (0.176%) (Ammermann et al. 1981). The level of 6mA in these unicellular eukaryotes is comparable to the 6mA abundance in many species of bacteria. Interestingly *Tetrahymena* and *Stylonychia mytilus* have 4–13 fold lower 6mA levels in their micronucleus than their macronucleus (Gorovsky et al. 1973; Ammermann et al. 1981), suggesting that this modification plays an important role in determining the differences between the two nuclei in these species, which are separated by ~1159 million years of evolution (Parfrey et al. 2011).

6mA was initially identified in the DNA of *C. elegans,* using both antibody-based approaches and antibody-independent methods of quantitation, including single molecule real time (SMRT) sequencing and ultra-high performance liquid chromatography followed by mass spectrometry (UHPLC-ms/ms) (Greer et al. 2015). Based on the UHPLC-ms/ms data, the levels of 6mA ranged from 0.013% to 0.39% of adenines. However, more recent measurements in *C. elegans* have quantified that 6mA is either undetectable or only occurs at 0.0003% of adenines (O'Brown et al. 2019). The initial higher quantifications appear to be due to artifacts introduced because of the presence of bacteria in the guts of *C. elegans*, exogenous methylated

adenines introduced to the samples by recombinant bacterial enzymes used to digest gDNA samples prior to UHPLC-ms/ms analysis, as well as by limitations of 6mA sequencing techniques (O'Brown et al. 2019).

6mA abundance was quantified in plants, rat tissues, and human cells using HPLC-ms/ms (Huang et al. 2015). These data must be viewed with caution, as there was no independent validation that the 6mA modification was occurring in the reported organisms, rather than contaminating symbionts. In that study, the abundance of 6mA in plant and mammalian genomes ranged from 0.00008% of adenines in rat lung DNA to as high as 0.0007% of adenines in plant DNA. The human cell lines had 0.0017% and 0.0023% 6mA (in Jurkat and 293T cells, respectively). Another group identified 6mA in 0.00009% of adenines in *Xenopus laevis* by HPLC and MeDIPseq (Koziol et al. 2016). More recently 6mA was identified in mouse ES cells at 0.0006–0.0007% (or 6–7 parts per million) of adenines (Wu et al. 2016). However, each of these quantifications has been called into question by conflicting reports which have questioned whether the modification exists at all in mammals (Schiffers et al. 2017; Douvlataniotis et al. 2020). The large range of reported 6mA levels in mammals, either not occurring (Schiffers et al. 2017; Douvlataniotis et al. 2020), occurring at the lower range of around 0.1–1 part per million bases (Huang et al. 2015) to the higher range of ~400 parts per million in mitochondrial DNA (Hao et al. 2020) or even as high as ~1000 parts per million in human glioblastoma derived stem cells (Xie et al. 2018), suggests that these differences are not biological but rather methodological. It will be important in future studies to ensure that when making direct comparisons culturing conditions as well as the methods used for detecting and quantifying 6mA are comparable. In summary, these findings suggest that if 6mA occurs in plants and mammalian genomes it is ~1,000–40,000-fold lower than in some bacteria and single-celled eukaryotes. The large degree of variability in 6mA abundance between eukaryotes motivates further exploration into the environmental factors and evolutionary pressures that led to a decline in 6mA levels and an increase in 5mC levels during eukaryotic evolution. These differences could also indicate that at very low 6mA levels, 6mA is at the limit of detection. Therefore, quantitative differences between different samples could be attributed to technical errors, rather than true biological variability. Moreover, these modifications are typically detected under basal conditions. It is possible that 6mA levels are dramatically altered under specific environmental conditions. Finally, we should note that even if a relatively rare percentage of adenines are methylated, the presence of a single methylated adenine at a critical genomic location could have dramatic phenotypic consequences by affecting the binding of specific regulatory proteins (see cell cycle regulation below).

8.5 Methods of Detecting 6mA

Detection of DNA methylation has evolved over the years to become increasingly sensitive and accurate. Detecting different DNA modifications started with a technique of combining the cytosine fraction with picric acid to form crystalline picrate. After purification by crystallization, salt crystals were compared to synthetic pyrimidines of known structure. By this method, the authors reported the identification of 5mC in *Mycobacterium tuberculosis* in 1925 (Johnson and Coghill 1925). Detection techniques shifted to paper chromatography (Hotchkiss 1948), which had a limit of detection of 1%, and was used to compare synthetically generated 5mC to the content of 5mC in animal, plant, viral, and bacterial DNA (Wyatt 1950). By the time, 6mA was first identified in 1955, its presence was confirmed by a combination of ultraviolet absorption spectrum (Mason 1954), electrophoretic mobility, and its paper chromatographic movement in different solvents (Dunn and Smith 1955). Because these early methods were relatively insensitive, the presence of 6mA in a number of animal species was undetectable. Researchers quickly realized that they could take advantage of restriction enzymes to identify methylated residues (Bird and Southern 1978; Geier and Modrich 1979). A

limitation of this approach is that detection of methylation sites is dependent on the methylated residue occurring in the appropriate restriction enzyme target motif, and whether the restriction enzyme preferentially recognizes un-, hemi-, or fully-methylated substrates. Therefore, not all sequence contexts can be addressed with this method.

High-performance liquid chromatography was subsequently used to determine that *E. coli* has 1.4% 6mA (Yuki et al. 1979). Liquid chromatography has become increasingly sensitive and, recently, ultra-high performance liquid chromatography coupled with mass spectrometry (UHPLC-ms/ms) has been used to detect concentrations of 6mA in the order of 0.00001% (Huang et al. 2015). However, UHPLC-ms/ms and other quantitative techniques cannot discriminate from which species the genomic DNA originates. This can cause problems if the gDNA is contaminated with microbiota or other species which could have substantially higher levels of 6mA than the species being queried. If the levels of 6mA are low, the contaminating prokaryotic DNA could cause an artificially elevated signal. Additionally, the enzymes used to digest DNA for UHPLC-ms/ms could be contaminated with methylated DNA from their recombinant production and therefore add abundant 6mA into the sample, which must be avoided if possible or subtracted from final concentrations when quantifying 6mA levels (O'Brown et al. 2019; Boulias and Greer 2021; Douvlataniotis et al. 2020). An alternative technique, called capillary electrophoresis and laser-induced fluorescence (CE-LIF), uses the fluorescent dye boron-dipyrromethene (BODIPY), to specifically bind to 6mA, followed by capillary electrophoresis combined with laser-induced fluorescence to detect 6mA levels (Krais et al. 2010). This technique has a lower limit of detection of 0.01% 6mA and was used to confirm the presence of 6mA in Bacteriophage λ, *E. coli,* and to identify 6mA's presence in *Hydra magnipapillata* (1.04% of adenines) (Krais et al. 2010). At this limit of detection, the authors could not detect 6mA in calf thymus or human kidney samples.

While the aforementioned techniques have proven useful for detecting whether 6mA is present in a particular organism, they do not provide information on the genomic location of this modification. To determine the genomic locations of 6mA, several methylation-sensitive sequencing techniques have been developed. Methylated DNA immunoprecipitation (MeDIP) coupled with microarray analysis (Weber et al. 2005) has evolved into MeDIP sequencing (MeDIP-seq) (Pomraning et al. 2009). MeDIP-seq has been optimized by a combination of photo-crosslinking, exonuclease digestion, and restriction enzyme digestion to achieve near single-nucleotide resolution of 6mA (Chen et al. 2015; Fu et al. 2015). MeDIP-seq, however, is dependent on the antibody specifically recognizing 6mA. While the most commonly used 6mA antibody displays a greater than 1000-fold affinity for methylated adenines relative to unmethylated adenines (Greer et al. 2015), if 6mA is rare, as is the case in most multicellular eukaryotes, non-specific binding can still confound analyses. Sequencing methods have an inherent error rate which can be further exacerbated by the non-specific binding of IgG to unmodified repetitive DNA sequences (Lentini et al. 2018; Douvlataniotis et al. 2020). Alternative techniques have also been developed to identify where throughout the genome 6mA occurs. One such technique consists of radioactive methylation of DNA followed by restriction digest, electrophoresis, and sequencing (Posfai and Szybalski 1988). Single-molecule real-time sequencing (SMRT-seq) is a next-generation sequencing technique that provides accurate sequence reads and measures the kinetic rate of nucleotide incorporation during sequencing (Flusberg et al. 2010). Since different DNA modifications result in different kinetic signatures, SMRT-seq can identify every DNA modification at single-base resolution. This technology, however, does have troubles distinguishing several closely related modifications from each other, including 1mA from 6mA. While SMRT-seq provides an antibody-independent manner of detecting, at nucleoside resolution, every different DNA

modification which produces a unique kinetic signature, this method requires high sequence depth and loses accuracy when 6mA levels are lower than 10 parts per million (Mondo et al. 2017; Ye et al. 2017; O'Brown et al. 2019; Douvlataniotis et al. 2020; Zhu et al. 2018). The earliest SMRT-seq analyses were performed using mapping algorithms that were designed for bacterial species where 6mA occurs at high abundance in specific motifs (Zhang et al. 2018), it is important to confirm with higher sequencing depth that detected methylated bases are not false positives (Zhu et al. 2018; O'Brown et al. 2019; Kong et al. 2022). Oxford Nanopore sequencing is an alternative long read sequencing technology that reads out disruption of ionic current as a DNA molecule passes through a nanopore present in a lipid bilayer (Bayley 2015). This sequencing method has been used to examine 6mA (McIntyre et al. 2019; Shah et al. 2019) but is subject to many of the same limitations as SMRT-seq. One of the most promising new 6mA sequencing technologies, nitrite sequencing, uses sodium nitrite under acidic conditions to selectively deaminate unmethylated adenines while not affecting N6-methylated adenines. This deamination converts unmethylated adenines to hypoxanthines, which pairs with cytosine rather than thymine. Therefore when sodium nitrite treated DNA is subjected to polymerase chain reactions all unmethylated adenines are converted to guanines during sequencing (Mahdavi-Amiri et al. 2020). While it still remains to be determined what the limit of detection of nitrite sequencing is, and whether it can accurately detect 6mA at lower concentrations than 10 parts per million, this chemical-based sequencing method will be a powerful tool for accurate mapping of 6mA in genomic DNA. Methylated residues can be confirmed by restriction digest coupled with real-time RT PCR to determine the methylation at a specific locus (Fu et al. 2015). Alternatively, sequence-specific probes have been developed that can selectively bind to 6mA or unmodified adenines in specific sequence contexts (Dohno et al. 2010).

To convincingly identify rare modifications, such as 6mA, a combination of multiple complimentary techniques is ideal since each technique has its own set of limitations (Table 8.1). UHPLC-ms/ms can be complemented by restriction enzyme digestion confirmation (as long as 6mA occurs in the appropriate motif), dot blots and MeDIP with a 6mA-specific antibody, and SMRT-seq. For a complementary discussion of the methods for detection of 5mC see chapter 16.

8.6 6mA Regulating Enzymes

8.6.1 DNA Methyltransferases

An important step in the confirmation of 6mA as a regulated mark of biological significance has been the identification of enzymes that deposit and remove this mark. It was previously thought that methylated adenines were incorporated premade into genomic DNA. This assumption likely hampered initial efforts to identify 6mA in eukaryotes. A study in the early 1970s concluded that 6mA did not exist in eukaryotes, because radioactively labeled adenines, but not methylated adenines were incorporated into DNA when added exogenously (Vanyushin et al. 1970). However, several groups demonstrated that DNA could be glycosylated and RNA could be methylated at the N6 position of adenines after incorporation into polynucleotides, rather than pre-methylated nucleotides being incorporated during the biosynthesis of polynucleotide (Kornberg et al. 1959; Kornberg et al. 1961; Fleissner and Borek 1962). These findings led to the hypothesis that methylation occurs after DNA synthesis (Theil and Zamenhof 1963), rather than on unincorporated nucleotides, and spurred attempts to identify the DNA methylating enzymes. The first biochemical studies aiming to identify DNA methyltransferases were conducted in *E. coli* by fractionation of total protein lysates followed by methylation assays with each fraction. Early studies identified a single fraction that methylated DNA at the C5 position of cytosines and the N6 position of adenines, but this fraction was only efficient at methylating foreign DNA (Gold et al.

Table 8.1 Recent methods for detecting and quantifying 6mA

Detection/ Quantitation Method	Description/Limitations	6mA Specificity	Sensitivity(lower limit of detection)	Genomic sequence information	Reference(s)
6mA-sensitive restriction enzymes	Restriction endonuclease cleavage of methylated motifs. Cannot detect 6mA outside of restriction recognition sites.	High (enzyme-dependent)	Can identify single methylated adenine so long as it occurs within specific recognition motifs (e.g., GATC) Preference for hemimethylated or dually methylated depending on the enzyme	None normally Can be used, in combination with real-time RT PCR, to validate sites identified by sequencing methods	Bird and Southern (1978), Geier and Modrich (1979)
6mA Dot Blotting	Antibody-dependent semi-quantitative detection of 6mA levels in genomic DNA samples. If samples have contamination with other nucleosides can be mistaken for 6mA	High (antibody-dependent) If DNA is not single-stranded 6mA antibodies generally recognize 1mA as well	Moderate (can not distinguish lowly methylated samples from each other)	None	Achwal et al. (1983)
Immunofluorescence	Antibody-dependent method for detecting 6mA in whole animals or tissues at cell-level resolution. Very difficult to validate that signal is coming from 6mA rather than background (ideally need to manipulate the 6mA regulating enzymes)	Moderate (antibody-dependent). Can recognize RNA or RNA:DNA or 1mA Necessary to eliminate all RNA and RNA:DNA hybrids by treatment with RNAses	Moderate. Immunofluorescence is not good for assessing relative changes in 6mA. (antibody-dependent)	None Could be used in combination with DNA probes to confirm 6mA localization in specific genomic regions	Greer et al. (2015), Sun et al. (2015), Liu et al. (2016b)
MeDIP-seq	Antibody-dependent method for identifying genomic regions harboring 6mA. Typical problems of antibody pull downs including IgG affinity for non-methylation regions	Moderate (antibody-dependent)	Moderate (antibody- and organism-dependent)	Genome-wide 6mA localization at near base pair resolution	Pomraning et al. (2009), Chen et al. (2015), Lentini et al. (2018)

SMRT-seq	Long-read sequencing, kinetics of sequencing is altered when bases are methylated. Provides base modifications at single-nucleotide resolution 6mA and 1mA are indistinguishable and quite expensive.	Single nucleotide (but 1mA and 6mA are indistinguishable)	Can identify single methylated adenine with sufficient coverage (if 6mA is > 10 ppm)	Single base resolution 6mA detection genome-wide	Flusberg et al. (2010), O'Brown et al. (2019)
UHPLC-MS/MS	Chemical separation and detection by mass spectrometry. Mycoplasma and digestion enzymes mixes can contaminate samples with methylated bases and cause artificially high readings	Highest	High (As low as 0.0001% 6mA/A)	None	Yuki et al. (1979), Zhang et al. (2015), Fu et al. (2015), Greer et al. (2015), Huang W et al. (2015), O'Brown et al. (2019)
6mA-specific probes	DNA probe containing a formyl group on the O6 position of a G base discriminates between adenine and 6mA via formation of an interstrand cross-link (ICL). 6mA can not form ICL. ICLs detected by PAGE or HPLC	Single site. Untested for other modifications (such as 1mA)	Detection by electrophoresis (6mA has no ICLs and will be single-stranded on the gel)	Can confirm 6mA in specific genomic locations, not a discovery tool	Dohno et al. (2010)
CE-LIF	BODIPY FL EDA binds covalently to the phosphate group of deoxyribonucleotide after activated by carbodiimide reagent. Run by CE-LIF which distinguishes different bases from each other and methylated bases based on migration time.	Can distinguish from 5mC or other bases but untested with 1mA	Moderate (0.01% limit)	None	Krais et al. (2010)

(continued)

Table 8.1 (continued)

Detection/ Quantitation Method	Description/Limitations	6mA Specificity	Sensitivity(lower limit of detection)	Genomic sequence information	Reference(s)
Nitrite sequencing	Treatment of DNA with sodium nitrite under acidic conditions selectively deaminates unmethylated As which are read out as guanosines after OCR amplification. This chemical-based conversion should be extremely specific but lower limits of sensitivity have not been tested.	Single nucleotide (it is unclear how sodium nitrite treatment behaves with other methylated adenines)	Undetermined	Single base resolution 6mA detection genome-wide	Mahdavi-Amiri et al. (2020)

[en]
Recent methods for detecting and quantifying 6mA are summarized in this table. Relative limitations and sensitivity of each method are discussed and references to the primary papers are cited. Abbreviations: *MeDIP-seq* Methylated DNA immunoprecipitation followed by high throughput DNA-sequencing, *SMRT-seq* Single-Molecule Real-Time sequencing, *UHPLC-MS/MS* Ultra-high performance liquid chromatography coupled to tandem mass spectrometry, *BODIPY FL EDA* fluorescent dye 4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-propionyl ethylene diamine hydrochloride, *CE-LIF* Capillary electrophoresis with laser-induced fluorescence

1963; Gold and Hurwitz 1964). Subsequent studies using increasingly subdivided fractions were able to identify multiple adenine and cytosine methyltransferases in *E. coli* (Nikolskaya et al. 1976; Nikolskaya et al. 1981). However, the identification and characterization of active DNA methyltransferases does not preclude that premethylated RNA or DNA nucleosides could be incorporated through the nucleotide salvage pathway or DNA polymerases. Several groups have demonstrated that administering exogenous premethylated adenines to mammalian cells leads to the incorporation of these N6-adenine methylated bases into the mammalian DNA (Schiffers et al. 2017; Charles et al. 2004; O'Brown et al. 2019; Musheev et al. 2020; Liu et al. 2021). Using exogenous heavy isotopes has revealed that exogenous DNA N6-methyladenine and even RNA N6-methyladenosine can both be incorporated into mammalian DNA (Schiffers et al. 2017; Musheev et al. 2020; Liu et al. 2021).

Additional evidence for the widespread presence and functional importance of 6mA in eukaryotic genomes comes from the observation that members of the MT-A70 family of known or putative N6-adenine methyltransferases exist in most organisms, ranging from bacteria to humans (Luo et al. 2015). Based on structural similarity to other members of the MT-A70 family of methyltransferases, the candidate DNA adenine methyltransferase enzymes in multicellular organisms likely evolved from the bacterial M. MunI-like 6mA methyltransferase, which functions in the host restriction modification system (Iyer et al. 2011). The MT-A70 family includes both RNA and DNA methyltransferases, including IME4 (also called SPO8) in *S. cerevisiae* (Clancy et al. 2002), DAMT-1 in *C. elegans* (Greer et al. 2015), and members of the methyltransferase-like (METTL) family in mammals, including METTL3 (an N6-adenosine RNA methyltransferase) (Liu et al. 2014), and METTL4 (a homolog of DAMT-1) (Greer et al. 2015). Whether the same enzymes catalyze both RNA and DNA adenine methylation in different organisms remains an open question. Notably, biochemical in vitro studies have suggested that the mammalian RNA methyltransferase METTL3 also methylates DNA (Woodcock et al. 2019), suggesting that the same enzymes can be capable of methylating both RNA and DNA in certain contexts, but the substrate specificity (i.e. RNA, DNA or both) for each member of the different MT-A70 family members remains incompletely characterized. Recent research has suggested that METTL4 is present in the mitochondria and is necessary for 6mA (Hao et al. 2020) which is highly enriched on mitochondrial DNA in humans (Koh et al. 2018; Hao et al. 2020). Knock-down of METTL4 caused an increase in the expression of mitochondrial DNA genes and an increase in mtDNA copy number (Hao et al. 2020). It was suggested that these effects were mediated by 6mA repelling the mitochondrial transcription factor TFAM (Hao et al. 2020). METTL4 was shown to be active in vitro against mitochondrial DNA (Hao et al. 2020), raising the possibility that 6mA could be a directed active epigenetic modification. However, METTL4 has also been reported to catalyze m^6Am on U2 snRNAs (Goh et al. 2020; Chen et al. 2020; Gu et al. 2020). It will be important for future experiments to determine the physiologically relevant substrate of METTL4. Moreover, several groups have also been unable to detect 6mA in appreciable levels in mitochondrial DNA (Ratel et al. 2006; Chen et al. 2020). It will be important for future experiments to determine whether techniques to isolate mitochondrial DNA could explain differences in detecting 6mA, whether 6mA is only present on mammalian mitochondrial DNA under specific stress conditions, or whether 6mA is absent from mammalian mitochondrial DNA. At the structural level, all of the MT-A70 containing enzymes are characterized by a 7-ß-strand methyltransferase domain at their C-terminus, fused to a predicted alpha-helical domain at their N-terminus and require S-adenosyl-L-methionine (SAM) as a methyl donor (Iyer et al. 2011). The high degree of amino acid sequence conservation among the predicted N6-adenine methyltransferases motivates further exploration into their potential functional conservation.

How adenine methyltransferases of recently evolved eukaryotes recognize their substrates

still remains to be determined. The utilization of adenine methylation by the restriction-modification system suggests that bacterial 6mA methyltransferases evolved to recognize specific sequences for methylation. In bacteria and the unicellular eukaryote *Tetrahymena,* DNA adenine methylation occurs in a palindromic sequence-specific manner in vitro and in vivo (Geier and Modrich 1979; Zelinkova et al. 1990; Bromberg et al. 1982). However, sequence-specific adenine methylation is not observed in all organisms and some bacterial DNA adenine methyltransferases show no sequence specificity (Drozdz et al. 2012). Similarly, 6mA sites in multicellular eukaryotes appear modestly enriched in specific sequence contexts (Greer et al. 2015; O'Brown et al. 2019; Pacini et al. 2019; Wu et al. 2016; Yao et al. 2017; Koh et al. 2018; He et al. 2019; Zhu et al. 2018; Li et al. 2019; Xiao et al. 2018) suggesting that targeted adenines might be selected by more complicated metrics than simply sequence codes. 6mA has been reported to correlate with chromatin boundaries (Li et al. 2020), the histone variant H2A.X (Wu et al. 2016), and various histone modifications (including histone H3 lysine 4 dimethylation (H3K4me2) (Greer et al. 2015), H3K9me3 and H3K27me3 (Xie et al. 2018; Yao et al. 2018)), leading to the supposition that these modifications could communicate with 6mA to help direct 6mA to specific locations beyond a sequence-specific pattern. However, since methods used to map 6mA can be prone to false positives (Lentini et al. 2018; Douvlataniotis et al. 2020; Zhu et al. 2018; O'Brown et al. 2019), accurate mapping and the existence of 6mA in multicellular eukaryotes must be confirmed before conclusions can be drawn on how this modification is localized.

8.6.2 Mechanism of 6mA Methyltransferases

Substantial work in prokaryotes has identified the mechanism of action, the preferred methyl donor, and the kinetics of 6mA methyltransferases. Whether these regulatory principles are conserved in eukaryotes remains to be seen. There was an initial debate as to whether N6 was directly methylated, or if adenines were first methylated on the N1 position and then, following a Dimroth rearrangement, the methyl group would be transferred to the N6 position. However, the enzyme *EcoRI* had been shown to methylate N6 directly rather than through an initial N1 methylation (Pogolotti et al. 1988). This result, combined with the slow rate of Dimroth reactions at physiological pH (Macon and Wolfenden 1968), suggests that N6 is the direct target of methyltransferases. This conclusion has been confirmed by the structures of different adenine-N6 methyltransferases in complex with DNA, showing a direct approximation of the N6 atom toward the methyl-donor (Goedecke et al. 2001; Horton et al. 2005; Horton et al. 2006).

Early reports identifying that DNA was methylated suggested that S-adenosyl-L-methionine (SAM) was the primary methyl donor (Gold et al. 1963), and subsequent work has confirmed that SAM is the predominant methyl donor for not only DNA and RNA methylation, but also for proteins and lipids (Chiang et al. 1996). However, 5,10-methylene tetrahydrofolate has been identified as the methyl donor for tRNAs in *Streptococcus faecalis* and *Bacillus subtilis* (Delk and Rabinowitz 1975; Delk et al. 1976; Urbonavicius et al. 2005). While the enzyme that utilizes 5,10-methylene tetrahydrofolate in *B. subtilis*, *GidA*, is absent in eukaryotes (Urbonavicius et al. 2005), this finding raises the possibility that some DNA methyltransferases might use alternative methyl donors.

Kinetic rates have been measured for the T4 bacteriophage DNA adenine methyltransferase, Dam (Malygin et al. 2000) and the EcoRI adenine methyltransferase (Reich and Mashhoon 1991). For Dam the methylation rate constant (k_{meth}) was significantly faster than the overall reaction rate constant (k_{cat}) (0.56 and 0.47 s^{-1} vs 0.023 s^{-1}), suggesting that product dissociation is the rate-limiting step. Similar, but faster results were observed with EcoRI (Reich and Mashhoon 1991). These enzymes function by binding, flipping out the adenine, methylating, and restacking of the modified base (Allan et al. 1998). Whether

these hold true for M.MunI-like methyltransferases remains to be determined. Reducing the double strand duplex stability did not alter the k_{meth}, suggesting that base-flipping is not a rate limiting step in the methylation reaction (Malygin et al. 2000). Additionally, EcoRI enzyme-DNA complexes were less efficient compared to enzyme-SAM complexes, suggesting that the enzyme first binds SAM before methylating its substrates (Reich and Mashhoon 1991). This is opposite to what has been observed with *Dam* and the bacterial 5mC methyltransferase HhaI, where the methyltransferase first binds DNA, followed by SAM (Urig et al. 2002; Wu and Santi 1987), suggesting that the sequence of binding events in the DNA methylation reaction is enzyme-dependent.

An important step for the confirmation of the presence and role of 6mA in more recently evolved eukaryotes will be the identification of genuine 6mA methyltransferases. The conservation of MT-A70 domain-containing proteins in conjunction with the identification of 6mA in many eukaryotes suggests that this modification is conserved. Whether eukaryotic DNA methyltransferases function in a similar manner to prokaryotic methyltransferases remains to be seen. Interestingly, the RNA m6A methyltransferase, METTL3, functions in complex with METTL14 (Liu et al. 2014), raising the possibility that DNA methyltransferase enzymes, like many other chromatin regulating enzymes, function in multi-protein complexes. These multi-protein complexes could help the enzymes achieve their specificity.

8.7 DNA Adenine Demethylation

The identification of the enzymes that catalyze the removal of 6mA from DNA strongly suggests that 6mA is a regulated and dynamic epigenetic mark. Examination of the enzymes responsible for the removal of DNA base damage fostered the identification and characterization of the DNA demethylation processes. DNA base damage, in the form of 1mA and 3mC, was shown to be removed by the Fe(II)- and α-ketoglutarate-dependent dioxygenase AlkB in *E. coli* (Trewick et al. 2002). The AlkB family of dealkylating enzymes is highly conserved from bacteria to humans (Fedeles et al. 2015; Wei et al. 1996). AlkB enzymes can demethylate many DNA substrates, including the DNA lesions 1mA, 3mC, and 3mT (Kamat et al. 2011; Chen et al. 2015). Notably, humans have nine AlkB family members (ALKBH1-8 and FTO). Like *E. coli* AlkB enzymes, the mammalian enzymes ALKBH2 and ALKBH3 function in the repair of DNA alkylation damage (Duncan et al. 2002). In addition to their DNA demethylase activity, AlkB members catalyze oxidative demethylation of RNA (Aas et al. 2003). Interestingly, AlkB enzymes in RNA viruses preferentially demethylate RNA substrates, suggesting these AlkBs are necessary for maintaining the integrity of the viral RNA genome (van den Born et al. 2008). More recently, it was found that AlkB family members function in the oxidative demethylation of N6-methyladenosine in RNA, catalyzed by ALKBH5 and FTO in mammals (Jia et al. 2011; Zheng et al. 2013), and that the AlkB family member NMAD-1 in *C. elegans* demethylates 6mA in DNA (Greer et al. 2015), although whether alternative substrates are more physiologically relevant remains to be determined (Wang et al. 2019). FTO was also shown to demethylate 6mA in single-stranded DNA in vitro (Jia et al. 2011), raising the possibility that these enzymes might regulate both DNA and RNA 6mA. ALKBH1 and ALKBH4 have also been proposed to demethylate 6mA (Wu et al. 2016; Xie et al. 2018; Xiao et al. 2018). ALKBH4 demethylates 6mA in in vitro demethylation assays (Kweon et al. 2019) and ALKBH1 was also shown to demethylate 6mA in single-stranded DNA in vitro (Wu et al. 2016). Additionally, ALKBH1 knockout was reported to cause an increase in global 6mA levels in mouse embryonic stem cells and this increase can be rescued by a wildtype, but not a catalytic domain mutant of ALKBH1 (Wu et al. 2016), suggesting that ALKBH1 functions as a 6mA demethylase in mammals. *Alkbh1* knockout leads to embryonic

lethality and significantly more males born than females due to ALKBH1 regulating gene expression during spermatogenesis (Nordstrand et al. 2010). However, in vitro assays with ALKBH1 and a variety of potential substrates reveals that it preferentially demethylates m1A on tRNAs (Liu et al. 2016a) or m5C on tRNAs (Haag et al. 2016), again suggesting that determination of the physiologically relevant substrates of ALKBH1 and ALKBH4 must be determined.

Several studies have begun to dissect the mechanism of action of AlkB demethylases. In the presence of their essential cofactors α-ketoglutarate and Fe(II), AlkB demethylases use molecular oxygen to oxidize the methyl group of 6mA, forming the unstable intermediate 6-hydroxymethyladenine (6hmA), which spontaneously releases its aldehyde group, regenerating the unmodified adenine base (Fig. 8.2) (Fedeles et al. 2015). Whether the same mechanism occurs for the demethylation of 6mA in eukaryotes and if so, whether 6hmA has any additional function remains to be seen. 6hmA was detected in both rat tissues and human cell lines (Xiong et al. 2019). In mammals, FTO was recently shown to oxidize m6A on RNA to N6-hydroxymethyladenosine (hm6A) and N6-formyladenosine (f6A) (Fu et al. 2013). These mRNA derivatives have half-lives of ~3 hours (Fu et al. 2013), suggesting that if 6hmA does have additional functions, they would require a 6hmA-specific binding protein that could stabilize the intermediate. ALKBH1 was shown to generate 6hmA in vitro and ex vivo (Xiong et al. 2019), raising the possibility that this mechanism of 6mA demethylation is conserved. The same oxidation reaction mechanism is used by AlkB enzymes to demethylate 1mA and 3mC during the cellular response to DNA alkylation damage (Falnes et al. 2002; Trewick et al. 2002).

In addition to demethylation of 6mA by the AlkB demethylase family, 6mA can also be converted to hypoxanthine by a 6mA deaminase (Kamat et al. 2011). This modified base can then undergo base excision repair by hypoxanthine DNA glycosylases of the AlkA family (Saparbaev and Laval 1994) (Fig. 8.2). If hypoxanthine is not removed, it can cause a transition mutation (AT pairs would be converted to GC pairs), since hypoxanthine pairs with cytosine instead of thymine. Recently, 6mA was found to be correlated with increased point mutations in *Neisseria meningitidis* (Sater et al. 2015), suggesting that this modified base might be mutagenic, potentially as a consequence of unrepaired 6mA deamination events. However, 6mA deaminases in *Neisseria meningiditis* have not yet been identified. In contrast to 6mA deamination, which is only mutagenic if not removed, 5mC is converted to thymine when deaminated, which leads to a transition mutation in a single step (Lindahl and Nyberg 1974; Heindell et al. 1978). Deamination of adenine, 6mA, or cytosine all leads to non-natural bases, which can easily be identified by specific glycosylases. Deamination of 5mC, on the other hand, leads to thymine which requires a more complicated repair process. This more direct mutational path might explain why 5mC is more prone to mutation than 6mA. This divergence begs the question as to why evolution has selected for a higher prevalence of the more mutagenic DNA modification in more recently evolved species.

In *E. coli,* AlkB expression is induced by DNA damage and the enzyme functions in DNA repair via direct removal of base alkylation damage (Trewick et al. 2002). *AlkB* mutant *E. coli* are sensitized to cell death induced by the alkylating agent methyl methanesulfonate (MMS), and the predicted human ortholog of AlkB is sufficient to partially rescue the MMS-induced cytotoxicity seen in AlkB mutants (Wei et al. 1996). Interestingly, MMS treatment of human skin fibroblasts did not result in the same induction of AlkB seen in *E. coli*, suggesting that the regulation of AlkB expression may have diverged during the evolution of more recent eukaryotes (Wei et al. 1996), or that one of the other 8 AlkB family members in humans has taken on this role or that the induction by different alkylating agents is cell-type specific, and may only occur in certain cell types. In *Penicillium chrysogenum* mutants lacking DNA adenine methyltransferase mutate more readily and display increased sensitivity to DNA damaging agents, suggesting that 6mA could regulate DNA

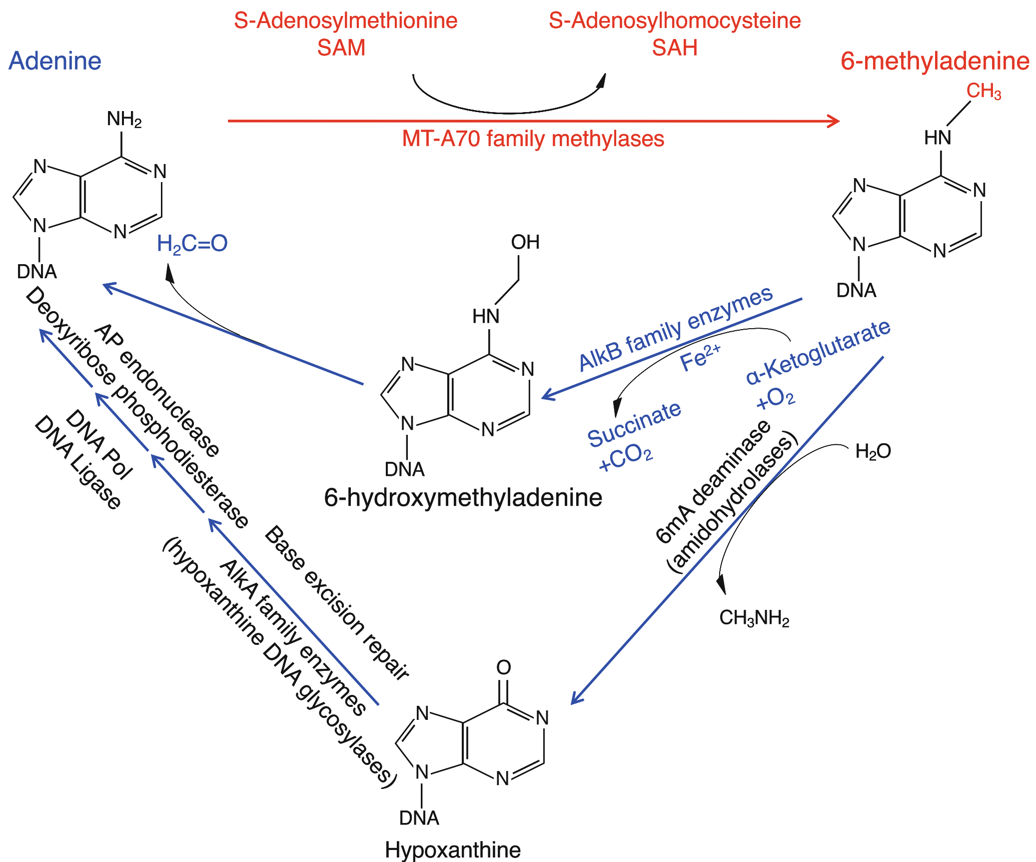


Fig. 8.2 Mechanisms of N6-adenine methylation and demethylation. MT-A70 family methylases catalyze the methylation of adenine at the sixth position of the purine ring. MT-A70 methylases use S-adenosylmethionine (SAM) as their methyl donor to generate 6-methyladenine and S-adenosylhomocysteine (SAH). Adenine could be regenerated from 6mA by several different enzymatic mechanisms: AlkB family enzymes catalyze the oxidative demethylation of 6mA. AlkB enzymes require α-ketoglutarate and Fe^{2+} and use oxygen to oxidize the methyl group. This oxidative demethylation reaction first generates 6-hydroxymethyladenine, which releases its formaldehyde group to generate adenine. Alternatively, 6mA can be deaminated and subsequently removed via the base excision repair pathway. First, 6mA deaminase hydrolyzes the methylamine to generate hypoxanthine. Hypoxanthine is recognized as a damaged base by AlkA family enzymes, which cleave the glycosyl bond to remove the base. Apurinic (AP) endonuclease cleaves the phosphodiester backbone at the abasic site, exposing the residual 5′ deoxyribose phosphate group, which is then removed by deoxyribose phosphodiesterase. Finally, DNA polymerase I incorporates the unmodified adenine and DNA ligase catalyzes the formation of the phosphodiester bond

damage or DNA repair in fungus as well (Rogers et al. 1986). In *C. elegans,* mutation of the putative 6mA demethylase *nmad-1*, causes increased DNA damage and defective expression of DNA repair genes (Wang et al. 2019), raising the possibility that some aspects of the prokaryotic DNA repair function of 6mA could be conserved in eukaryotes. However, NMAD-1 could function by demethylating other residues; therefore, NMAD-1's physiologically relevant substrates need to be identified before broader conclusions can be drawn about a conservation of 6mA's role in multicellular eukaryotes.

Interestingly, a different family of enzymes, ten-eleven translocation (Tet) proteins, has been shown to demethylate 5mC in many organisms

(Tahiliani et al. 2009; Ito et al. 2010, 2011). Unlike AlkB proteins, whose crystal structures have revealed that the enzymes flip out the base to facilitate demethylation (Yang et al. 2008; Sundheim et al. 2008), crystal structure of the TET enzymes demonstrated that TET catalytic domains are not suitable for accommodating flipped out purines (Aravind et al. 2015), suggesting that they cannot act on 6mA. The TET family has a good phyletic correlation with DNA cytosine methyltransferases, but not with DAMT-1 or other Dam family methylases (Aravind et al. 2015). Additionally in bacteria, there is little evidence that TET-related enzymes are capable of demethylating purines (Aravind et al. 2015). Given these findings, it is surprising that the *D. melanogaster* ortholog of Tet (named DMAD) was reported to function as a 6mA demethylase on DNA (Zhang et al. 2015). Nuclear extracts from DMAD mutant flies showed reduced in vitro demethylation activity compared to nuclear extracts from wild-type flies, while the addition of purified DMAD was sufficient to increase adenine demethylation in these assays (Zhang et al. 2015). It remains to be seen whether this 6mA demethylase activity can be biochemically confirmed using purified DMAD, and whether Tet proteins play a conserved role as 6mA demethylases.

8.8 6mA Binding Proteins

Beyond the machinery that catalyzes the addition and removal of 6mA, cells have evolved mechanisms to recognize 6mA as a regulatory signal that can be translated into different biological consequences (see Biological functions of 6mA). We will discuss later in this chapter the direct chemical consequences of adenine methylation, but 6mA can be recognized by specific effector molecules or complexes that alter chromatin architecture and/or transcriptional states. Alternatively, methylation could function by preventing the binding of proteins. Methyladenine-binding proteins have evolved to recognize and transduce 6mA signals into specific biological outcomes. For example, in *E. coli* the MutS enzyme binds to mismatch base pairs as a homodimer, facilitating recruitment of the MutL protein, which binds MutS. The MutS-MutL-DNA complex then loops out until it finds the nearest hemimethylated GATC site, which is bound by the endonuclease MutH. Upon binding of MutL-MutS to the MutH-DNA complex, MutH is activated and nicks the unmethylated daughter strand, allowing helicase and exonucleases to excise the single-stranded mismatch region (Su and Modrich 1986). Thus, hemimethylated GATC sites are used to specifically direct mismatch repair of the daughter strand (Lahue et al. 1987) Similarly, the *oriC* region of *E. coli* is hemimethylated to prevent premature replication before the cell has divided. These hemimethylated adenine sites are recognized and bound by the SeqA protein (Brendler et al. 1995; Slater et al. 1995), which prevents assembly of the DNA replication machinery at this region (von Freiesleben et al. 1994; Wold et al. 1998). The crystal structure for SeqA has revealed why SeqA binds preferentially to hemimethylated over fully methylated DNA (Guarne et al. 2002; Fujikawa et al. 2004), highlighting the importance of determining the crystal structure of 6mA binding proteins for deciphering the chemical and biological consequences of their binding. Several eukaryotic 6mA binding proteins have also been identified. The *D. melanogaster* transcription factor Jumu has a slight preference for binding to N6-adenine methylated DNA and might play a role in regulating the maternal-to-zygotic transition through binding to and transcribing N6-adenine methylated genes (He et al. 2019). The mitochondrial single-stranded DNA binding protein 1 (SSBP1) displays a ~2.5-fold higher affinity for N6-adenine methylated DNA (Koh et al. 2018). To fully understand the potential biological roles of 6mA it will be important to further identify and characterize 6mA binding proteins in eukaryotes.

8.9 Biological Functions of 6mA

The direct effects of adenine methylation on the structure of DNA and its roles in prokaryote biology have been well characterized (see also chapter 2). The functional role that 6mA plays in eukaryotes is actively being deciphered. Discussing 6mAs functional effects in prokaryotes raises several interesting potential functions which will need to be further explored in eukaryotes.

8.9.1 Effects of Adenine Methylation on DNA Structure

One possible role for adenine methylation, beyond providing a binding site for effector proteins, is to directly alter the overall structure of DNA. An early crystal structure suggested that 6mA might alter the secondary structure of DNA (Sternglanz and Bugg 1973). Adenine methylation is thought to affect DNA double helix formation through altering both base pair stability and base stacking. Ultraviolet photoelectron studies suggested that adenine methylation would lower the ionization potentials and cause the destabilization of valence electrons to increase base stacking in methylated adenines (Peng et al. 1976). This increased base stacking would be offset by a slight destabilization of base pairing ranging from ~0.35 to 0.95 kcal/mol (Engel and von Hippel 1978b). Interestingly, 5mC behaves oppositely to 6mA in these regards. Hence, 5mC causes an increase in helix stability, while adenine methylation destabilizes the DNA, as measured by denaturing gradient gel electrophoresis (Collins and Myers 1987). Moreover, 6mA within GATC sequences causes slight DNA unwinding of 0.5°/methyl group (Cheng et al. 1985), but two-dimensional NMR studies revealed that, in almost all cases, 6mA has only minor effects on the overall helix conformation, as it retains the canonical B-form (Fazakerley et al. 1985; Quignard et al. 1985; Fazakerley et al. 1987). The effects of 6mA on the thermodynamic stability and folding of DNA appear to be sequence-specific (Fazakerley et al. 1987). Indeed, when 6mA occurs directly after a T this can cause a highly altered structure that is overwound and bent (Fazakerley et al. 1989). However, 6mA lowers melting temperatures and slows the rate of helix formation, as demonstrated by the enthalpy of dissociation studies (Quignard et al. 1985; Fazakerley et al. 1985). While 6mA does not dramatically alter helix rigidity (Hagerman and Hagerman 1996; Mills and Hagerman 2004), it can increase DNA curvature to variable degrees, depending on sequence context (Diekmann 1987). These studies suggest that methylated adenines are associated with DNA regions that spend prolonged periods in the open state. These effects were confirmed by cruciform extrusion assays where 5mC inhibits extrusion and 6mA facilitates the initial opening of DNA (Murchie and Lilley 1989). These consequences seem to be in line with the reported effects of 5mC and 6mA on gene transcription; 5mC is generally believed to be a repressor of gene transcription when it occurs at promoters, while 6mA is generally associated with gene activation (Rogers and Rogers 1995; Graham and Larkin 1995; Allamane et al. 2000; Liu et al. 2016b; Zhang et al. 2018; Liang et al. 2018; Shah et al. 2019). However, the correlation between 5mC and gene transcription is dependent on the genomic context in which it occurs. When 5mC occurs within gene bodies, rather than promoters, it is correlated with gene transcription (Reviewed in (Jones 2012)). Similarly, 6mA has also been correlated with repression of gene expression (Zhang et al. 2015; Yao et al. 2018; Lizarraga et al. 2020) as well as repression of transposons (Wu et al. 2016; Yao et al. 2017; Koh et al. 2018; He et al. 2019). Thus, the effects of 6mA on gene transcription may depend on its location in the genome.

8.9.2 Restriction-Modification Systems

In prokaryotes, DNA N6-adenine methylation is oftentimes used to discriminate self from foreign DNA, as part of restriction modification systems;

a bacterial immune system by which pathogenic DNA from bacteriophages is recognized by endonucleases that selectively cleave unmethylated DNA at specific restriction sites that are methylated in the host's genome, and thus protected from endonuclease digestion (Low et al. 2001; Iyer et al. 2011). Interestingly, enterobacteriophages appear to have evolved to contain fewer GATCs to avoid the GATC R-M system of their hosts (McClelland 1984). However, GATC methylation is not always involved in the R-M system as discussed in more detail below (Marinus and Lobner-Olesen 2014). This system does not appear to be conserved in eukaryotes that have evolved more complex immune systems. However, 6mA has been suggested to correlate with long interspersed element (LINE) retrotransposons suppression (Wu et al. 2016; Yao et al. 2017; Koh et al. 2018; He et al. 2019; Zhu et al. 2018), raising the possibility that 6mA could recognize and inhibit foreign DNA through an independent mechanism when it is integrated into the host genome. But since the enrichment of 6mA at LINE elements is not always observed (Li et al. 2019; Xiao et al. 2018), further studies using alternative 6mA mapping methods are required to determine whether 6mA could play a role in suppressing foreign DNA in eukaryotes.

8.9.3 DNA Damage Control

Early reports indicated that *dam* mutant *E. coli* had higher mutation rates and were more sensitive to UV and mitomycin C, suggesting that 6mA could protect against DNA damage (Marinus and Morris 1974). It was subsequently suggested that 6mA could help to distinguish the parental DNA strand from the mutated daughter strand (Glickman et al. 1978; Glickman 1979). Similarly, *Penicillium chrysogenum* mutants deficient in 6mA had higher sensitivity to mutagenic agents without changes in the number of mutations (Rogers et al. 1986). Additionally, mutation of the putative DNA demethylase, *nmad-1,* in *C. elegans* leads to elevated levels of DNA damage (Wang et al. 2019). However, as stated above, NMAD-1 could regulate DNA damage through the regulation of substrates other than 6mA. Since deletion of *nmad-1* is correlated with defects in the expression of DNA repair genes (Wang et al. 2019), NMAD-1 could also regulate DNA damage repair in eukaryotes through indirect mechanisms.

In *E. coli* and other Gram-negative bacteria, DNA adenine methylation plays an important role in the DNA mismatch repair pathway, a strand-specific repair pathway that relies on the transient post-replicative hemimethylation of DNA. The DNA adenine methylase, Dam, binds selectively to hemimethylated DNA substrates and methylates GATC sites after DNA replication. The delay between DNA synthesis and methylation of the newly synthesized daughter strand is crucial for the fidelity of DNA mismatch repair (Pukkila et al. 1983). When DNA replication errors lead to base pair mismatches, the DNA repair machinery uses adenine methylation to distinguish the already methylated template strand from the newly synthesized unmethylated daughter strand. As described above (6mA binding proteins) hemimethylated DNA allows MutL, MutS, and MutH to identify and specifically cleave the daughter strand, allowing helicase and exonucleases to excise the single-stranded mismatch region. Subsequently, DNA polymerase III re-synthesizes the mismatch region of single-stranded DNA using the methylated parental strand as a template (Pukkila et al. 1983). However, mechanisms of DNA mismatch repair appear to be different in eukaryotes (Fukui 2010).

8.9.4 Effect on Transcription

Several studies listed below have suggested that N6-adenine methylation correlates with increased gene expression in different more recently evolved eukaryotes. Whether this is due to the direct effect on relaxing DNA structure (as discussed above), recruitment of 6mA-specific binding proteins, or both, remains unknown. It is still also unclear whether this phenomenon is conserved across all organisms that contain 6mA. While 5mC CpG methylation

had little effect on transcription in barley, 6mA methylation increased transcription two to five-fold (Rogers and Rogers 1995). Similarly, 6mA but not 5mC methylation increased gene expression by 3–50 fold of reporter constructs in tobacco or wheat protoplast, or intact wheat tissues (Graham and Larkin 1995). Luciferase reporter constructs purified from dam+dcm+ bacteria (with 5mC and 6mA methylation) had 2–6 fold increased luciferase production compared to constructs purified from dam-dcm- bacteria in rat or mouse cell lines, or when electroporated into mice (Allamane et al. 2000). Together, these results suggest that 6mA promotes gene expression.

6mA can also directly affect binding of transcription factors. Methylation of a HNF1 binding site reduces HNF1 binding affinity, but this only causes a minor reduction in gene transcription (Tronche et al. 1989; Lichtsteiner and Schibler 1989). Conversely, 6mA increases binding affinity for the transcription factor AGP1 in tobacco (Sugimoto et al. 2003). These results suggest that the effects of adenine methylation on transcription will be sequence- and transcription factor specific. Interestingly, 6mA was shown to reduce the incorporation rate of uridines by inducing a stalling of RNA polymerase II in *S. cerevisiae* in vitro experiments (Wang et al. 2017). This finding suggests that an increase in transcription would have to overcome a physical pausing of the polymerase, however, it is important to perform directed adenine methylation to determine what 6mA's causal effects are on transcription.

Similar to DNA cytosine methylation in metazoa, bacterial DNA adenine methylation regulates gene expression programs, including those related to virulence and phase variation (Low et al. 2001; Wallecha et al. 2002; Zaleski et al. 2005; Sarnacki et al. 2013), suggesting that 6mA levels might be sensitive to changes in environmental conditions. By directed manipulation of the Dam methyltransferase it was shown that 6mA in *Salmonella enterica* predominantly leads to activation of transcription (Sanchez-Romero et al. 2020). Similarly, recent data suggest that 6mA may play a role in transcriptional regulation in the single-celled eukaryote *Chlamydomonas reinhardti,* where 6mA occurs preferentially near actively transcribed genes (Fu et al. 2015). As preliminary evidence that 6mA levels might be relevant to human physiology and disease, it was reported that human patients with type 2 diabetes have reduced levels of m6A on RNA and 6mA on DNA, as measured by HPLC-ms/ms. It was proposed that these differences might be regulated by the cellular fat mass and obesity associated protein (FTO) (Huang et al. 2015), which was shown to function as an RNA m6A and single-stranded DNA 6mA demethylase (Jia et al. 2011) and DNA 3mT demethylase (Gerken et al. 2007). 6mA was found to be significantly enriched in the mitochondria where it was demonstrated that the mitochondrial transcription factor TFAM was repelled by N6-adenine methylated DNA and 6mA suppressed in vitro transcription of mitochondrial DNA (Hao et al. 2020). Future studies will be required to definitively determine whether 6mA exists in human DNA using independent detection methods.

8.9.5 Nucleosome Positioning

In the protists *Tetrahymena thermophilia*, *Chlamydomonas reinhardtii,* and *Oxytricha trifallax,* 6mA is preferentially located in the linker regions between nucleosomes (Karrer and VanNuland 2002; Fu et al. 2015; Pratt and Hattman 1983; Beh et al. 2019), raising the possibility that 6mA could help to direct nucleosome positioning. Alternatively, enrichment of 6mA in linker regions may reflect increased accessibility, or recruitment of the methyltransferase at regions of open chromatin. Interestingly, in rice, deletion of the nucleosome remodeler, DDM1, causes a 2.5-fold reduction in 6mA (Zhang et al. 2018). In future studies, it will be interesting to examine whether 6mA directs nucleosome positioning and whether it does so in a conserved manner, or whether other open chromatin modifications can direct N6-adenine methylation at those sites.

8.9.6 Cell Cycle Regulation

N6-adenine methylation marks regions for DNA replication initiation in prokaryotes and has been shown to alter the rate of cell cycle progression (see chapter 2). In *E. coli*, the Dam methyltransferase is necessary for precise timing between DNA replication events (Bakker and Smith 1989; Boye and Lobner-Olesen 1990). The hemimethylation of DNA plays an important role in modulating the initiation of DNA replication. The SeqA protein binds to hemimethylated DNA adjacent to the origin of replication *OriC*, preventing its methylation by Dam, and leading to a delay in DNA replication before the cell has divided, which is only initiated from a fully methylated promoter (Low et al. 2001; Lu et al. 1994). When DNA replication is desired, adenine methylation at the oriC region lowers the thermal melting temperature which could facilitate the unwinding at the origin of replication (Yamaki et al. 1988). Interestingly, 6mA also slows the rate of DNA polymerase I catalysis, presumably due to the effects of 6mA on base pairing (discussed above) (Engel and von Hippel 1978a).

In *Caulobacter crescentus,* the cell cycle-regulated DNA adenine methylase (CcrM) controls the timing of DNA replication and progression through the cell cycle (Collier et al. 2007). In contrast to *E. Coli* Dam methylase, which does not have a preference for hemimethylated sites, *C. crescentus* CcrM preferentially methylates hemimethylated DNA after replication (Berdis et al. 1998) and is essential for cell viability (Stephens et al. 1996). In *C. crescentus,* 6mA levels change throughout the cell cycle from fully to hemimethylated as the replication forks progress (Kozdon et al. 2013). The promoter of the replication initiation factor DnaA is preferentially activated when its promoter is fully methylated, leading to DnaA accumulation and progression through the cell cycle (Collier et al. 2007). In *C. elegans* deletion of *nmad-1* causes delayed DNA replication (Wang et al. 2019). Whether this change in DNA replication is due to a misregulation of cell cycle gene expression or through a direct consequence to DNA methylation remains to be seen. Mitochondrial DNA replication in humans could also be regulated by 6mA, as SSBP1, a mitochondrial DNA replication factor, is a 6mA binding protein (Koh et al. 2018). In vitro kinetic experiments with the human DNA polymerase η suggest that 6mA directly decreases replication efficiency (Du et al. 2019). To determine a definitive role for 6mA in cell cycle regulation in eukaryotes it will be necessary to perform directed N6-adenine methylation or demethylation and measure the consequences on cell cycle progression.

8.9.7 Transgenerational Inheritance

DNA methylation at palindromic sites provides the most parsimonious method by which epigenetic information could be transmitted across generations. Because of the semi-conservative nature of DNA replication, methylation events on the parental strand can be replicated on the newly synthesized daughter strand. In mammals, 5mC methylation patterns are established by the *de novo* methyltransferases Dnmt3a and Dnmt3b during early embryonic development (Okano et al. 1999). Inheritance of cytosine methylation patterns through cell division is mediated by the maintenance methyltransferase Dnmt1 (Bestor et al. 1988). Dnmt1 preferentially binds hemimethylated DNA at the replication fork and copies parental-strand methylation patterns onto the unmethylated daughter strand (Stein et al. 1982; Yoder et al. 1997; Bestor 2000; Bashtrykov and Jeltsch 2018). Whether adenine methylation propagates non-genetic information through cell divisions, or from parents to their offspring remains to be seen. However, there are some hints that 6mA could transmit non-genetic information. Labeling experiments showed that newly synthesized *E. coli* DNA in Okazaki fragments were quickly N6-adenine methylated (Marinus 1976), consistent with the idea that parental methylation patterns might be passed on to their descendants during DNA replication. In some bacteria, DNA adenine methylation is tightly coordinated with cell division (Casadesus and

Low 2006)(see cell cycle regulation above), enabling the inheritance of parental methylation patterns. Thus, a key unanswered question is whether there exists a mode of inheritance of adenine methylation in eukaryotes, or whether different organisms have evolved different mechanisms for the inheritance of parental DNA methylation through somatic nuclear divisions and across generations. In the ciliate *Tetrahymena thermophila* macronucleus, analysis of methylation patterns using methylation-sensitive restriction enzymes showed that both actively replicating and non-replicating DNA contained hemimethylated sites, and that the vegetatively growing macronucleus contained a combination of partially methylated and fully methylated sites (Capowski et al. 1989). These findings are inconsistent with a simple semi-conservative 6mA inheritance mechanism and suggest that inheritance of 6mA in some organisms may rely on hemi-methylation-independent mechanisms of 6mA maintenance through cell division (Capowski et al. 1989).

In *C. elegans*, loss of the histone H3 lysine 4 dimethyl (H3K4me2) demethylase *spr-5* causes a progressive transgenerational loss of fertility (Katz et al. 2009) and a transgenerational extension in lifespan (Greer et al. 2016). This is accompanied by a progressive decline in H3K9me3 and accumulation of H3K4me2 and 6mA (Greer et al. 2014; Greer et al. 2015). Deletion of the putative 6mA demethylase, *nmad-1*, accelerates the progressive fertility decline, while deletion of the putative 6mA methyltransferase, *damt-1*, suppresses the transgenerational H3K4me2 accumulation, fertility, and longevity phenotypes (Greer et al. 2015; Greer et al. 2016), raising the possibility that 6mA might transmit epigenetic information across generations. 6mA also increases transgenerationally in response to electron transport chain stress, and deletion of *damt-1* eliminates the transgenerationally phenotype (Ma et al. 2019). However, the physiologically relevant substrates of NMAD-1 and DAMT-1 must be identified before it can be determined whether these transgenerational phenotypes are truly regulated by 6mA or some other modification. Future studies will be needed to reveal whether 6mA can regulate transgenerational inheritance in multicellular eukaryotes.

Many years of research have shown that chromatin modifications do not occur in isolation, but rather actively communicate with each other. For example, 5mC and H3K9me3 are coordinately regulated in mammals and plants (see chapters 11). The H3K9 methyltransferase binds to 5mC methylated DNA (Jackson et al. 2002; Johnson et al. 2007; Malagnac et al. 2002) and the DNA methyltransferase binds to H3K9me-containing nucleosomes (Du et al. 2012). It is possible that a similar reciprocal cross-talk occurs between 6mA and chromatin modifications. 6mA correlates with chromatin modifications in several eukaryotic species (Li et al. 2020; Wu et al. 2016; Greer et al. 2015; Xie et al. 2018; Yao et al. 2018). In *D. melanogaster,* Dmad binds to Wds, an H3K4 trimethyltransferase complex component, and deletion of Dmad causes a decrease in H3K4me3 (Yao et al. 2018). Future work should reveal whether 6mA methyltransferases can bind to specific methylated histones to direct DNA methylation to particular loci. However, this coordinate cross-talk between 6mA and chromatin modifications is predicated on accurate mapping of 6mA. The correlation between 6mA and chromatin modifications must first be confirmed in eukaryotes by alternative mapping techniques (Lentini et al. 2018; Douvlataniotis et al. 2020; Zhu et al. 2018; O'Brown et al. 2019), before any conclusions about cross-talk between 6mA and other modifications can be drawn.

8.10 Conclusions and Future Directions

As detection techniques are becoming increasingly sensitive, 6mA has begun to be convincingly observed in several metazoa. However, several groups have pointed to errors in these methods which could lead to high false positives (Schiffers et al. 2017; Liu et al. 2017; O'Brown et al. 2019; Douvlataniotis et al. 2020; Zhu et al. 2018; Lentini et al. 2018). Due to the relative paucity of 6mA in multicellular eukaryotes, at or near the limit of detection for multiple techniques,

changes in 6mA could help confirm or negate the conserved presence of this modification. 6mA might only occur under specific conditions of stress or in the mitochondria, which could be difficult to detect under basal conditions (Zhang et al. 2018; Ma et al. 2019; Yao et al. 2017; Li et al. 2019). The conservation of active 6mA methyltransferases, demethylases, and binding proteins, coupled with alternative detection techniques could confirm that N6-adenine methylation is a conserved signaling modification. However, it will be important to rigorously examine whether 6mA is present across the tree of life using a combination of rapidly evolving detection techniques (discussed in this review and others that are actively being developed). For metazoa that are confirmed to have 6mA in their DNA, it will be important to define the biological functions of 6mA and its genomic localization patterns in different cell types. A fundamental question is whether the biological functions of 6mA in bacteria are conserved in higher eukaryotes or whether 6mA has evolved new biological functions in these organisms. As 6mA occurs less frequently in more recently evolved organisms, this might reflect a more specialized functional role.

A growing body of work has revealed an important role for m6A on mRNAs in the regulation of gene expression and cellular differentiation in eukaryotes (Peer et al. 2017; He and He 2021; Zaccara et al. 2019). Therefore, another open question is whether N6-adenine methylation of DNA is coordinately regulated with N6-adenine methylation on RNA. Given that substrates of the AlkB family of demethylases and MT-A70 family of methyltransferases can include both RNA and DNA, it will be of interest to better characterize the substrate specificity of these enzymes in different organisms and to examine whether the same enzymes regulate both RNA and DNA N6-adenine methylation in different organisms. Moreover, it will be relevant to find out if in cases of overlapping substrate specificities, whether methylation of DNA or RNA (or both) is the biologically relevant signal under different physiological conditions. Additionally, RNA m6A, or methylated DNA from foreign organisms, could be incorporated into genomic DNA through the nucleotide salvage pathway (Schiffers et al. 2017; Charles et al. 2004; O'Brown et al. 2019; Musheev et al. 2020; Liu et al. 2021). While this indirect incorporation of 6mA into eukaryotic DNA would be less directed, it could still have an effect on biological processes in multicellular eukaryotes.

Given the dynamic nature of 5mC in mammalian development and cell differentiation (Okano et al. 1999; Chen and Zhang 2020) (see chapter 1 + 5), it will be of interest to define the dynamics and potential functions of 6mA during mammalian development, if its presence in mammals can be rigorously confirmed. 6mA has been proposed to change in development in *D. melanogaster, A. thaliana, D. rerio, M. musculus,* and *S. domesticus* (Liu et al. 2016b; Fernandes et al. 2021; Liang et al. 2018; Shah et al. 2019) however, it will be important to confirm that these changes are not due to changes in the relative contribution of foreign DNA with high levels of 6mA (O'Brown et al. 2019; Kong et al. 2022). Future studies should also reveal the environmental factors that regulate the levels of 6mA and its modifying enzymes in eukaryotes, which should provide clues to its evolutionary conservation and biological relevance. The diversity of methods for detection of 6mA in DNA will allow for a comprehensive and detailed examination of 6mA's presence, localization patterns, and potential functions in the genomes of diverse organisms. All in all, the newly developed and more sensitive tools for detection, along with the recent discovery of 6mA in metazoa tentatively open an exciting new chapter of discovery in the field of adenine methylation.

Acknowledgments We thank S. Burger, N. O'Brown, and E. Pollina for the critical reading of the manuscript. We thank M.H. Rothi for help generating heat maps in Fig. 8.1. We thank C. He for helpful discussions. The work from the Greer laboratory is supported by grants from the NIH (DP2AG055947 and R01AI151215). Z.K.O. was supported by 5T32HD7466-19. We apologize for literature omitted owing to space limitations.

References

Aas PA, Otterlei M, Falnes PO, Vagbo CB, Skorpen F, Akbari M et al (2003) Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. Nature. 421(6925):859–863. https://doi.org/10.1038/nature01363

Achwal CW, Iyer CA, Chandra HS (1983) Immunochemical evidence for the presence of 5mC, 6mA and 7mG in human, Drosophila and mealybug DNA. FEBS Lett. 158(2):353–358

Adams RL, McKay EL, Craig LM, Burdon RH (1979) Methylation of mosquito DNA. Biochim Biophys Acta. 563(1):72–81

Allamane S, Jourdes P, Ratel D, Vicat JM, Dupre I, Laine M et al (2000) Bacterial DNA methylation and gene transfer efficiency. Biochem Biophys Res Commun. 276(3):1261–1264. https://doi.org/10.1006/bbrc.2000.3603

Allan BW, Beechem JM, Lindstrom WM, Reich NO (1998) Direct real time observation of base flipping by the EcoRI DNA methyltransferase. J Biol Chem 273(4):2368–2373

Ammermann D, Steinbruck G, Baur R, Wohlert H (1981) Methylated bases in the DNA of the ciliate Stylonychia mytilus. Eur J Cell Biol. 24(1):154–156

Aravind L, Zhang D, Iyer LM (2015) The TET/JBP family of nucleic acid base-modifying 2-oxoglutarate and iron-dependent dioxygenases In: Hausinger R, Schofield C, editors. 2-Oxoglutarate-dependent oxygenases. Royal Society of Chemistry 3(12):289

Babinger P, Kobl I, Mages W, Schmitt R (2001) A link between DNA methylation and epigenetic silencing in transgenic Volvox carteri. Nucleic Acids Res. 29(6): 1261–1271

Bakker A, Smith DW (1989) Methylation of GATC sites is required for precise timing between rounds of DNA replication in Escherichia coli. J Bacteriol. 171(10): 5738–5742

Bashtrykov P, Jeltsch A (2018) DNA Methylation Analysis by Bisulfite Conversion Coupled to Double Multiplexed Amplicon-Based Next-Generation Sequencing (NGS). Methods Mol Biol. 1767:367–382. https://doi.org/10.1007/978-1-4939-7774-1_20

Batista PJ, Molinie B, Wang J, Qu K, Zhang J, Li L et al (2014) m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell. 15(6):707–719. https://doi.org/10.1016/j.stem.2014.09.019

Bayley H (2015) Nanopore sequencing: from imagination to reality. Clin Chem. 61(1):25–31. https://doi.org/10.1373/clinchem.2014.223016

Beh LY, Debelouchina GT, Clay DM, Thompson RE, Lindblad KA, Hutton ER et al (2019) Identification of a DNA N6-adenine methyltransferase complex and its impact on chromatin organization. Cell. 177(7): 1781–1796. e25. https://doi.org/10.1016/j.cell.2019.04.028

Berdis AJ, Lee I, Coward JK, Stephens C, Wright R, Shapiro L et al (1998) A cell cycle-regulated adenine DNA methyltransferase from Caulobacter crescentus processively methylates GANTC sites on hemimethylated DNA. Proc Nat Acad Sci U S A 95(6):2874–2879

Bestor TH (2000) The DNA methyltransferases of mammals. Hum Mol Genet. 9(16):2395–2402

Bestor T, Laudano A, Mattaliano R, Ingram V (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol. 203(4):971–983

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

Bird AP, Southern EM (1978) Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis. J Mol Biol. 118(1):27–47

Bodi Z, Zhong S, Mehra S, Song J, Graham N, Li H et al (2012) Adenosine methylation in arabidopsis mRNA is associated with the 3′ end and reduced levels cause developmental defects. Front Plant Sci. 3:48. https://doi.org/10.3389/fpls.2012.00048

Boulias K, Greer EL (2021) Detection of DNA Methylation in Genomic DNA by UHPLC-MS/MS. Methods Mol Biol. 2198:79–90. https://doi.org/10.1007/978-1-0716-0876-0_7

Boye E, Lobner-Olesen A (1990) The role of dam methyltransferase in the control of DNA replication in E. coli. Cell. 62(5):981–989

Brendler T, Abeles A, Austin S (1995) A protein that binds to the P1 origin core and the oriC 13mer region in a methylation-specific fashion is the product of the host seqA gene. The EMBO journal. 14(16):4083–4089

Bromberg S, Pratt K, Hattman S (1982) Sequence specificity of DNA adenine methylase in the protozoan Tetrahymena thermophila. J Bacteriol. 150(2): 993–996

Campbell JL, Kleckner N (1990) E. coli oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. Cell. 62(5):967–979

Capowski EE, Wells JM, Harrison GS, Karrer KM (1989) Molecular analysis of N6-methyladenine patterns in Tetrahymena thermophila nuclear DNA. Mol Cell Biol. 9(6):2598–2605

Capuano F, Mulleder M, Kok R, Blom HJ, Ralser M (2014) Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. Anal Chem. 86(8):3697–3702. https://doi.org/10.1021/ac500447w

Casadesus J, Low D (2006) Epigenetic gene regulation in the bacterial world. Microbiol Mol Biol Rev. 70(3): 830–856. https://doi.org/10.1128/MMBR.00016-06

Charles MP, Ravanat JL, Adamski D, D'Orazi G, Cadet J, Favier A et al (2004) N(6)-Methyldeoxyadenosine, a nucleoside commonly found in prokaryotes, induces C2C12 myogenic differentiation. Biochem Biophys Res Commun. 314(2):476–482

Chen Z, Zhang Y (2020) Role of mammalian DNA methyltransferases in development. Annu Rev Biochem. 89:135–158. https://doi.org/10.1146/annurev-biochem-103019-102815

Chen K, Luo GZ, He C (2015) High-resolution mapping of N(6)-methyladenosine in transcriptome and genome using a photo-crosslinking-assisted strategy. Methods Enzymol. 560:161–185. https://doi.org/10.1016/bs.mie.2015.03.012

Chen H, Gu L, Orellana EA, Wang Y, Guo J, Liu Q et al (2020) METTL4 is an snRNA m(6)Am methyltransferase that regulates RNA splicing. Cell Res. 30(6):544–547. https://doi.org/10.1038/s41422-019-0270-4

Cheng SC, Herman G, Modrich P (1985) Extent of equilibrium perturbation of the DNA helix upon enzymatic methylation of adenine residues. J Biol Chem 260(1):191–194

Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, Pardhasaradhi K et al (1996) S-Adenosylmethionine and methylation. FASEB J. 10(4):471–480

Clancy MJ, Shambaugh ME, Timpte CS, Bokar JA (2002) Induction of sporulation in Saccharomyces cerevisiae leads to the formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. Nucleic Acids Res. 30(20):4509–4518

Collier J, McAdams HH, Shapiro L (2007) A DNA methylation ratchet governs progression through a bacterial cell cycle. Proc Nat Acad Sci U S A 104(43):17111–17116. https://doi.org/10.1073/pnas.0708112104

Collins M, Myers RM (1987) Alterations in DNA helix stability due to base modifications can be evaluated using denaturing gradient gel electrophoresis. J Mol Biol. 198(4):737–744

Cummings DJ, Tait A, Goddard JM (1974) Methylated bases in DNA from Paramecium aurelia. Biochim Biophys Acta. 374(1):1–11

Degnen ST, Morris NR (1973) Deoxyribonucleic acid methylation and development in Caulobacter bacteroides. J Bacteriol. 116(1):48–53

Delk AS, Rabinowitz JC (1975) Biosynthesis of ribosylthymine in the transfer RNA of Streptococcus faecalis: a folate-dependent methylation not involving S-adenosylmethionine. Proc Nat Acad Sci U S A 72(2):528–530

Delk AS, Romeo JM, Nagle DP Jr, Rabinowitz JC (1976) Biosynthesis of ribothymidine in the transfer RNA of Streptococcus faecalis and Bacillus subtilis. A methylation of RNA involving 5,10-methylenetetrahydrofolate. J Biol Chem 251(23):7649–7656

Deng X, Chen K, Luo GZ, Weng X, Ji Q, Zhou T et al (2015) Widespread occurrence of N6-methyladenosine in bacterial mRNA. Nucleic Acids Res. 43(13):6557–6567. https://doi.org/10.1093/nar/gkv596

Diekmann S (1987) DNA methylation can enhance or induce DNA curvature. EMBO J 6(13):4213–4217

Dohno C, Shibata T, Nakatani K (2010) Discrimination of N6-methyl adenine in a specific DNA sequence. Chem Commun (Camb). 46(30):5530–5532. https://doi.org/10.1039/c0cc00172d

Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S et al (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature. 485(7397):201–206. https://doi.org/10.1038/nature11112

Douvlataniotis K, Bensberg M, Lentini A, Gylemo B, Nestor CE (2020) No evidence for DNA N (6)-methyladenine in mammals. Sci Adv 6(12):eaay3335. https://doi.org/10.1126/sciadv.aay3335

Drozdz M, Piekarowicz A, Bujnicki JM, Radlinska M (2012) Novel non-specific DNA adenine methyltransferases. Nucleic Acids Res. 40(5):2119–2130. https://doi.org/10.1093/nar/gkr1039

Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E et al (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell. 151(1):167–180. https://doi.org/10.1016/j.cell.2012.07.034

Du K, Zhang X, Zou Z, Li B, Gu S, Zhang S et al (2019) Epigenetically modified N(6)-methyladenine inhibits DNA replication by human DNA polymerase eta. DNA Repair (Amst) 78:81–90. https://doi.org/10.1016/j.dnarep.2019.03.015

Duncan T, Trewick SC, Koivisto P, Bates PA, Lindahl T, Sedgwick B (2002) Reversal of DNA alkylation damage by two human dioxygenases. Proc Nat Acad Sci U S A 99(26):16660–16665. https://doi.org/10.1073/pnas.262589799

Dunn DB, Smith JD (1955) Occurrence of a new base in the deoxyribonucleic acid of a strain of Bacterium coli. Nature. 175(4451):336–337

Dunn DB, Smith JD (1958) The occurrence of 6-methylaminopurine in deoxyribonucleic acids. Biochem J. 68(4):627–636

Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA et al (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res. 10(8):2709–2721

Ehrlich M, Gama-Sosa MA, Carreira LH, Ljungdahl LG, Kuo KC, Gehrke CW (1985) DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine. Nucleic Acids Res. 13(4):1399–1412

Ehrlich M, Wilson GG, Kuo KC, Gehrke CW (1987) N4-methylcytosine as a minor base in bacterial DNA. J Bacteriol. 169(3):939–943

Engel JD, von Hippel PH (1978a) D(M6ATP) as a probe of the fidelity of base incorporation into

polynucleotides by Escherichia coli DNA polymerase I. J Biol Chem 253(3):935–939

Engel JD, von Hippel PH (1978b) Effects of methylation on the stability of nucleic acid conformations. Studies at the polymer level. J Biol Chem 253(3):927–934

Falnes PO, Johansen RF, Seeberg E (2002) AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. Nature. 419(6903):178–182. https://doi.org/10.1038/nature01048

Fazakerley GV, Guy A, Teoule R, Quignard E, Guschlbauer W (1985) A proton 2D-NMR study of an oligodeoxyribonucleotide containing N6-methyladenine:d(GGm6ATATCC). Biochimie. 67(7-8):819–822

Fazakerley GV, Quignard E, Teoule R, Guy A, Guschlbauer W (1987) A two-dimensional 1H-NMR study of the dam methylase site: comparison between the hemimethylated GATC sequence, its unmethylated analogue and a hemimethylated CATG sequence. The sequence dependence of methylation upon base-pair lifetimes. Eur J Biochem. 167(3):397–404

Fazakerley GV, Gabarro-Arpa J, Lebret M, Guy A, Guschlbauer W (1989) The GTm6AC sequence is overwound and bent. Nucleic Acids Res. 17(7): 2541–2556

Fedeles BI, Singh V, Delaney JC, Li D, Essigmann JM (2015) The AlkB family of Fe(II)/alpha-Ketoglutarate-dependent Dioxygenases: repairing nucleic acid alkylation damage and beyond. J Biol Chem 290(34): 20734–20742. https://doi.org/10.1074/jbc.R115.656462

Fernandes SB, Grova N, Roth S, Duca RC, Godderis L, Guebels P et al (2021) N(6)-Methyladenine in Eukaryotic DNA: tissue distribution, early embryo development, and neuronal toxicity. Front Genet. 12: 657171. https://doi.org/10.3389/fgene.2021.657171

Fleissner E, Borek E (1962) A new enzyme of RNA synthesis: RNA methylase. Proc Nat Acad Sci U S A 48:1199–1203

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA et al (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods. 7(6):461–465. https://doi.org/10.1038/nmeth.1459

Frye M, Harada BT, Behm M, He C (2018) RNA modifications modulate gene expression during development. Science (New York, NY) 361(6409): 1346–1349. https://doi.org/10.1126/science.aau1646

Fu Y, Jia G, Pang X, Wang RN, Wang X, Li CJ et al (2013) FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. Nat Commun. 4:1798. https://doi.org/10.1038/ncomms2822

Fu Y, Luo GZ, Chen K, Deng X, Yu M, Han D et al (2015) N(6)-methyldeoxyadenosine marks active transcription start sites in chlamydomonas. Cell. 161(4): 879–892. https://doi.org/10.1016/j.cell.2015.04.010

Fujikawa N, Kurumizaka H, Nureki O, Tanaka Y, Yamazoe M, Hiraga S et al (2004) Structural and biochemical analyses of hemimethylated DNA binding by the SeqA protein. Nucleic Acids Res. 32(1):82–92. https://doi.org/10.1093/nar/gkh173

Fukui K (2010) DNA mismatch repair in eukaryotes and bacteria. J Nucleic Acids. 2010. https://doi.org/10.4061/2010/260512

Gama-Sosa MA, Midgett RM, Slagel VA, Githens S, Kuo KC, Gehrke CW et al (1983) Tissue-specific differences in DNA methylation in various mammals. Biochim Biophys Acta. 740(2):212–219

Geier GE, Modrich P (1979) Recognition sequence of the dam methylase of Escherichia coli K12 and mode of cleavage of Dpn I endonuclease. J Biol Chem 254(4): 1408–1413

Gerken T, Girard CA, Tung YC, Webby CJ, Saudek V, Hewitson KS et al (2007) The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. Science (New York, NY) 318(5855): 1469–1472. https://doi.org/10.1126/science.1151710

Glickman BW (1979) Spontaneous mutagenesis in Escherichia coli strains lacking 6-methyladenine residues in their DNA: an altered mutational spectrum in dam- mutants. Mutat Res. 61(2):153–162

Glickman B, van den Elsen P, Radman M (1978) Induced mutagenesis in dam- mutants of Escherichia coli: a role for 6-methyladenine residues in mutation avoidance. Mol Gen Genet. 163(3):307–312

Goedecke K, Pignot M, Goody RS, Scheidig AJ, Weinhold E (2001) Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. Nat Struct Biol. 8(2):121–125. https://doi.org/10.1038/84104

Goh YT, Koh CWQ, Sim DY, Roca X, Goh WSS (2020) METTL4 catalyzes m6Am methylation in U2 snRNA to regulate pre-mRNA splicing. Nucleic Acids Res. 48(16):9250–9261. https://doi.org/10.1093/nar/gkaa684

Gold M, Hurwitz J (1964) The enzymatic methylation of ribonucleic acid and deoxyribonucleic Acid. V. Purification and properties of the deoxyribonucleic acid-methylating activity of Escherichia Coli. J Biol Chem 239:3858–3865

Gold M, Hurwitz J, Anders M (1963) The enzymatic methylation of RNA and DNA. Biochem Biophys Res Commun. 11:107–114

Gommers-Ampt JH, Borst P (1995) Hypermodified bases in DNA. FASEB J. 9(11):1034–1042

Gommers-Ampt JH, Van Leeuwen F, de Beer AL, Vliegenthart JF, Dizdaroglu M, Kowalak JA et al (1993) beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan T. brucei. Cell. 75(6):1129–1136

Gorovsky MA, Hattman S, Pleger GL (1973) (6N)methyl adenine in the nuclear DNA of a eucaryote, Tetrahymena pyriformis. J Cell Biol. 56(3):697–701

Graham MW, Larkin PJ (1995) Adenine methylation at dam sites increases transient gene expression in plant cells. Transgenic Res. 4(5):324–331

Greer EL, Beese-Sims SE, Brookes E, Spadafora R, Zhu Y, Rothbart SB et al (2014) A histone methylation network regulates transgenerational epigenetic memory in C. elegans. Cell Reports. 7(1):113–126. https://doi.org/10.1016/j.celrep.2014.02.044

Greer EL, Blanco MA, Gu L, Sendinc E, Liu J, Aristizabal-Corrales D et al (2015) DNA Methylation on N(6)-Adenine in C. elegans. Cell. 161(4):868–878. https://doi.org/10.1016/j.cell.2015.04.005

Greer EL, Becker B, Latza C, Antebi A, Shi Y (2016) Mutation of C. elegans demethylase spr-5 extends transgenerational longevity. Cell Res. 26(2):229–238. https://doi.org/10.1038/cr.2015.148

Grosjean H (2009) Nucleic acids are not boring long polymers of only four types of nucleotides: A guided tour. In: Grosjean H (ed) DNA and RNA modi cation Enzymes: Structure, mechanism, function and evolution. CRC Press, pp 1–18

Gu L, Wang L, Chen H, Hong J, Shen Z, Dhall A et al (2020) CG14906 (mettl4) mediates m(6)A methylation of U2 snRNA in Drosophila. Cell Discov. 6:44. https://doi.org/10.1038/s41421-020-0178-7

Guarne A, Zhao Q, Ghirlando R, Yang W (2002) Insights into negative modulation of E. coli replication initiation from the structure of SeqA-hemimethylated DNA complex. Nat Struct Biol. 9(11):839–843. https://doi.org/10.1038/nsb857

Haag S, Sloan KE, Ranjan N, Warda AS, Kretschmer J, Blessing C et al (2016) NSUN3 and ABH1 modify the wobble position of mt-tRNAMet to expand codon recognition in mitochondrial translation. EMBO J 35(19): 2104–2119. https://doi.org/10.15252/embj.201694885

Hagerman KR, Hagerman PJ (1996) Helix rigidity of DNA: the meroduplex as an experimental paradigm. J Mol Biol. 260(2):207–223. https://doi.org/10.1006/jmbi.1996.0393

Hao Z, Wu T, Cui X, Zhu P, Tan C, Dou X et al (2020) N (6)-deoxyadenosine methylation in mammalian mitochondrial DNA. Mol Cell 78(3):382–395. e8. https://doi.org/10.1016/j.molcel.2020.02.018

Hassel M, Cornelius MG, Vom Brocke J, Schmeiser HH (2010) Total nucleotide analysis of Hydra DNA and RNA by MEKC with LIF detection and 32P-postlabeling. Electrophoresis. 31(2):299–302. https://doi.org/10.1002/elps.200900458

Hattman S, Kenny C, Berger L, Pratt K (1978) Comparative study of DNA methylation in three unicellular eucaryotes. J Bacteriol. 135(3):1156–1157

He PC, He C (2021) m(6) A RNA methylation: from mechanisms to therapeutic potential. EMBO J 40(3): e105977. https://doi.org/10.15252/embj.2020105977

He S, Zhang G, Wang J, Gao Y, Sun R, Cao Z et al (2019) 6mA-DNA-binding factor Jumu controls maternal-to-zygotic transition upstream of Zelda. Nat Commun. 10(1):2219. https://doi.org/10.1038/s41467-019-10202-3

Heindell HC, Liu A, Paddock GV, Studnicka GM, Salser WA (1978) The primary sequence of rabbit alpha-globin mRNA. Cell. 15(1):43–54

Hongay CF, Orr-Weaver TL (2011) Drosophila Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. Proc Nat Acad Sci U S A 108(36): 14855–14860. https://doi.org/10.1073/pnas.1111577108

Horton JR, Liebert K, Hattman S, Jeltsch A, Cheng X (2005) Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of dam methyltransferase. Cell. 121(3):349–361. https://doi.org/10.1016/j.cell.2005.02.021

Horton JR, Liebert K, Bekes M, Jeltsch A, Cheng X (2006) Structure and substrate recognition of the Escherichia coli DNA adenine methyltransferase. J Mol Biol. 358(2):559–570. https://doi.org/10.1016/j.jmb.2006.02.028

Hotchkiss RD (1948) The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem 175(1):315–332

Huang W, Xiong J, Yang Y, Liu SM, Yuan BF, Feng YQ (2015) Determination of DNA adenine methylation in genomes of mammals and plants by liquid chromatography/mass spectrometry. R Soc Chem Adv 5:64046–64054

Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, 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 SC, Collins LB, Swenberg JA et al (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science (New York, NY) 333(6047):1300–1303. https://doi.org/10.1126/science.1210597

Iyer LM, Abhiman S, Aravind L (2011) Natural history of eukaryotic DNA methylation systems. Prog Mol Biol Transl Sci. 101:25–104. https://doi.org/10.1016/B978-0-12-387685-0.00002-0

Jabbari K, Caccio S, Pais de Barros JP, Desgres J, Bernardi G (1997) Evolutionary changes in CpG and methylation levels in the genome of vertebrates. Gene. 205(1-2):109–118

Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature. 416(6880):556–560. https://doi.org/10.1038/nature731

Janulaitis A, Klimasauskas S, Petrusyte M, Butkus V (1983) Cytosine modification in DNA by BcnI methylase yields N4-methylcytosine. FEBS Lett. 161(1): 131–134

Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y et al (2011) N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol. 7(12):885–887. https://doi.org/10.1038/nchembio.687

Johnson TB, Coghill RD (1925) The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic

acid of the Tubercle bacillus. J Am Chem Soc 47: 2838–2844

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J et al (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. Current biology : CB. 17(4):379–384. https://doi.org/10.1016/j.cub.2007.01.009

Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 13(7):484–492. https://doi.org/10.1038/nrg3230

Kakutani T, Munakata K, Richards EJ, Hirochika H (1999) Meiotically and mitotically stable inheritance of DNA hypomethylation induced by ddm1 mutation of Arabidopsis thaliana. Genetics. 151(2):831–838

Kamat SS, Fan H, Sauder JM, Burley SK, Shoichet BK, Sali A et al (2011) Enzymatic deamination of the epigenetic base N-6-methyladenine. J Am Chem Soc. 133(7):2080–2083. https://doi.org/10.1021/ja110157u

Karrer KM, VanNuland TA (2002) Methylation of adenine in the nuclear DNA of Tetrahymena is internucleosomal and independent of histone H1. Nucleic Acids Res. 30(6):1364–1370

Katz DJ, Edwards TM, Reinke V, Kelly WG (2009) A C. elegans LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. Cell. 137(2):308–320

Koh CWQ, Goh YT, Toh JDW, Neo SP, Ng SB, Gunaratne J et al (2018) Single-nucleotide-resolution sequencing of human N6-methyldeoxyadenosine reveals strand-asymmetric clusters associated with SSBP1 on the mitochondrial genome. Nucleic Acids Res. 46(22):11659–11670. https://doi.org/10.1093/nar/gky1104

Kong Y, Cao L, Deikus G, Fan Y, Mead EA, Lai W et al (2022) Critical assessment of DNA adenine methylation in eukaryotes using quantitative deconvolution. Science (New York, NY.

Kornberg A, Zimmerman SB, Kornberg SR, Josse J (1959) Enzymatic synthesis of deoxyribonucleic acid. Influence of bacteriophage T2 on the synthetic pathway in host cells. Proc Nat Acad Sci U S A 45(6): 772–785

Kornberg SR, Zimmerman SB, Kornberg A (1961) Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected Escherichia coli. J Biol Chem 236:1487–1493

Kozdon JB, Melfi MD, Luong K, Clark TA, Boitano M, Wang S et al (2013) Global methylation state at base-pair resolution of the Caulobacter genome throughout the cell cycle. Proc Nat Acad Sci U S A 110(48): E4658–E4667. https://doi.org/10.1073/pnas.1319315110

Koziol MJ, Bradshaw CR, Allen GE, Costa AS, Frezza C, Gurdon JB (2016) Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. Nat Struct Mol Biol 23(1): 24–30. https://doi.org/10.1038/nsmb.3145

Krais AM, Cornelius MG, Schmeiser HH (2010) Genomic N(6)-methyladenine determination by MEKC with LIF. Electrophoresis. 31(21):3548–3551. https://doi.org/10.1002/elps.201000357

Kweon SM, Chen Y, Moon E, Kvederaviciute K, Klimasauskas S, Feldman DE (2019) An Adversarial DNA N(6)-Methyladenine-sensor network preserves polycomb silencing. Mol Cell. https://doi.org/10.1016/j.molcel.2019.03.018

Lahue RS, Su SS, Modrich P (1987) Requirement for d (GATC) sequences in Escherichia coli mutHLS mismatch correction. Proc Nat Acad Sci U S A 84(6): 1482–1486

Lentini A, Lagerwall C, Vikingsson S, Mjoseng HK, Douvlataniotis K, Vogt H et al (2018) A reassessment of DNA-immunoprecipitation-based genomic profiling. Nat Methods. 15(7):499–504. https://doi.org/10.1038/s41592-018-0038-7

Letunic I, Bork P (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 39(Web Server issue): W475-8. https://doi.org/10.1093/nar/gkr201

Li X, Zhao Q, Wei W, Lin Q, Magnan C, Emami MR et al (2019) The DNA modification N6-methyl-2'-deoxyadenosine (m6dA) drives activity-induced gene expression and is required for fear extinction. Nat Neurosci. 22(4):534–544. https://doi.org/10.1038/s41593-019-0339-x

Li Z, Zhao S, Nelakanti RV, Lin K, Wu TP, Alderman MH 3rd et al (2020) N(6)-methyladenine in DNA antagonizes SATB1 in early development. Nature. 583(7817):625–630. https://doi.org/10.1038/s41586-020-2500-9

Liang Z, Shen L, Cui X, Bao S, Geng Y, Yu G et al (2018) DNA N(6)-Adenine Methylation in Arabidopsis thaliana. Dev Cell. 45(3):406–416. e3. https://doi.org/10.1016/j.devcel.2018.03.012

Lichtsteiner S, Schibler U (1989) A glycosylated liver-specific transcription factor stimulates transcription of the albumin gene. Cell. 57(7):1179–1187

Lindahl T, Nyberg B (1974) Heat-induced deamination of cytosine residues in deoxyribonucleic acid. Biochemistry. 13(16):3405–3410

Linn S, Arber W (1968) Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form. Proc Nat Acad Sci U S A 59(4): 1300–1306

Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L et al (2014) A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol. 10(2):93–95. https://doi.org/10.1038/nchembio.1432

Liu F, Clark W, Luo G, Wang X, Fu Y, Wei J et al (2016a) ALKBH1-Mediated tRNA Demethylation Regulates Translation. Cell. 167(3):816–828. e16. https://doi.org/10.1016/j.cell.2016.09.038

Liu J, Zhu Y, Luo GZ, Wang X, Yue Y, Wang X et al (2016b) Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. Nat Commun. 7:13052. https://doi.org/10.1038/ncomms13052

Liu B, Liu X, Lai W, Wang H (2017) Metabolically Generated Stable Isotope-Labeled Deoxynucleoside Code for Tracing DNA N(6)-Methyladenine in Human Cells. Anal Chem. 89(11):6202–6209. https://doi.org/10.1021/acs.analchem.7b01152

Liu X, Lai W, Li Y, Chen S, Liu B, Zhang N et al (2021) N (6)-methyladenine is incorporated into mammalian genome by DNA polymerase. Cell Res. 31(1):94–97. https://doi.org/10.1038/s41422-020-0317-6

Lizarraga A, O'Brown ZK, Boulias K, Roach L, Greer EL, Johnson PJ et al (2020) Adenine DNA methylation, 3D genome organization, and gene expression in the parasite Trichomonas vaginalis. Proc Nat Acad Sci U S A 117(23):13033–13043. https://doi.org/10.1073/pnas.1917286117

Low DA, Weyand NJ, Mahan MJ (2001) Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. Infect Immun. 69(12): 7197–7204. https://doi.org/10.1128/IAI.69.12.7197-7204.2001

Lu M, Campbell JL, Boye E, Kleckner N (1994) SeqA: a negative modulator of replication initiation in E. coli. Cell. 77(3):413–426

Luo GZ, Blanco MA, Greer EL, He C, Shi Y (2015) DNA N-methyladenine: a new epigenetic mark in eukaryotes? Nat Rev 16(12):705–710. https://doi.org/10.1038/nrm4076

Luria SE, Human ML (1952) A nonhereditary, host-induced variation of bacterial viruses. J Bacteriol. 64(4):557–569

Lyko F, Ramsahoye BH, Jaenisch R (2000) DNA methylation in Drosophila melanogaster. Nature. 408(6812): 538–540. https://doi.org/10.1038/35046205

Ma C, Niu R, Huang T, Shao LW, Peng Y, Ding W et al (2019) N6-methyldeoxyadenine is a transgenerational epigenetic signal for mitochondrial stress adaptation. Nat Cell Biol. 21(3):319–327. https://doi.org/10.1038/s41556-018-0238-5

Macon JB, Wolfenden R (1968) 1-Methyladenosine. Dimroth rearrangement and reversible reduction. Biochemistry. 7(10):3453–3458

Mahdavi-Amiri Y, Chung Kim Chung K, Hili R (2020) Single-nucleotide resolution of N (6)-adenine methylation sites in DNA and RNA by nitrite sequencing. Chem Sci. 12(2):606–612. https://doi.org/10.1039/d0sc03509b

Malagnac F, Bartee L, Bender J (2002) An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. EMBO J 21(24): 6842–6852

Malygin EG, Lindstrom WM Jr, Schlagman SL, Hattman S, Reich NO (2000) Pre-steady state kinetics of bacteriophage T4 dam DNA-[N(6)-adenine] methyltransferase: interaction with native (GATC) or modified sites. Nucleic Acids Res. 28(21):4207–4211

Marinus MG (1976) Adenine methylation of Okazaki fragments in Escherichia coli. J Bacteriol. 128(3): 853–854

Marinus MG, Lobner-Olesen A (2014) DNA methylation. Ecosal Plus 6(1). https://doi.org/10.1128/ecosalplus.ESP-0003-2013

Marinus MG, Morris NR (1973) Isolation of deoxyribonucleic acid methylase mutants of Escherichia coli K-12. J Bacteriol. 114(3):1143–1150

Marinus MG, Morris NR (1974) Biological function for 6-methyladenine residues in the DNA of Escherichia coli K12. J Mol Biol. 85(2):309–322

Mason SF (1954) Purine Studies. Part II. The Ultra-violet absorption spectra of some mono- and poly-substituted purines. J Chem Soc:2071–2081

McClelland M (1984) Selection against dam methylation sites in the genomes of DNA of enterobacteriophages. J Mol Evol. 21(4):317–322

McIntyre ABR, Alexander N, Grigorev K, Bezdan D, Sichtig H, Chiu CY et al (2019) Single-molecule sequencing detection of N6-methyladenine in microbial reference materials. Nat Commun. 10(1):579. https://doi.org/10.1038/s41467-019-08289-9

Meselson M, Yuan R (1968) DNA restriction enzyme from E. coli. Nature. 217(5134):1110–1114

Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell. 149(7):1635–1646. https://doi.org/10.1016/j.cell.2012.05.003

Mills JB, Hagerman PJ (2004) Origin of the intrinsic rigidity of DNA. Nucleic Acids Res. 32(13): 4055–4059. https://doi.org/10.1093/nar/gkh740

Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, LaButti K et al (2017) Widespread adenine N6-methylation of active genes in fungi. Nat Genet 49(6):964–968. https://doi.org/10.1038/ng.3859

Montero LM, Filipski J, Gil P, Capel J, Martinez-Zapater JM, Salinas J (1992) The distribution of 5-methylcytosine in the nuclear genome of plants. Nucleic Acids Res. 20(12):3207–3210

Murchie AI, Lilley DM (1989) Base methylation and local DNA helix stability. Effect on the kinetics of cruciform extrusion. J Mol Biol. 205(3):593–602

Murray NE (2002) 2001 Fred Griffith review lecture. Immigration control of DNA in bacteria: self versus non-self. Microbiology. 148(Pt 1):3–20. https://doi.org/10.1099/00221287-148-1-3

Musheev MU, Baumgartner A, Krebs L, Niehrs C (2020) The origin of genomic N(6)-methyl-deoxyadenosine in mammalian cells. Nat Chem Biol. 16(6):630–634. https://doi.org/10.1038/s41589-020-0504-2

Nikolskaya II, Lopatina NG, Chaplygina NM, Debov SS (1976) The host specificity system in Escherichia coli SK. Mol Cell Biochem. 13(2):79–87

Nikolskaya II, Lopatina NG, Debov SS (1981) On heterogeneity of DNA methylases from Escherichia coli SK cells. Mol Cell Biochem. 35(1):3–10

Niu Y, Zhao X, Wu YS, Li MM, Wang XJ, Yang YG (2013) N6-methyl-adenosine (m6A) in RNA: an old modification with a novel epigenetic function. Genom

Proteom Bioinf 11(1):8–17. https://doi.org/10.1016/j.gpb.2012.12.002

Nordstrand LM, Svard J, Larsen E, Nilsen A, Ougland R, Furu K et al (2010) Mice lacking Alkbh1 display sex-ratio distortion and unilateral eye defects. PLoS One. 5(11):e13827. https://doi.org/10.1371/journal.pone.0013827

O'Brown ZK, Boulias K, Wang J, Wang SY, O'Brown NM, Hao Z et al (2019) Sources of artifact in measurements of 6mA and 4mC abundance in eukaryotic genomic DNA. BMC Genomics. 20(1):445. https://doi.org/10.1186/s12864-019-5754-6

Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 99(3):247–257

Pacini CE, Bradshaw CR, Garrett NJ, Koziol MJ (2019) Characteristics and homogeneity of N6-methylation in human genomes. Sci Rep. 9(1):5185. https://doi.org/10.1038/s41598-019-41601-7

Parfrey LW, Lahr DJ, Knoll AH, Katz LA (2011) Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc Nat Acad Sci U S A 108(33):13624–13629. https://doi.org/10.1073/pnas.1110633108

Peer E, Rechavi G, Dominissini D (2017) Epitranscriptomics: regulation of mRNA metabolism through modifications. Curr Opin Chem Biol. 41:93–98. https://doi.org/10.1016/j.cbpa.2017.10.008

Peng S, Padva A, LeBreton PR (1976) Ultraviolet photoelectron studies of biological purines: the valence electronic structure of adenine. Proc Nat Acad Sci U S A 73(9):2966–2968

Pogolotti AL Jr, Ono A, Subramaniam R, Santi DV (1988) On the mechanism of DNA-adenine methylase. J Biol Chem 263(16):7461–7464

Pomraning KR, Smith KM, Freitag M (2009) Genome-wide high throughput analysis of DNA methylation in eukaryotes. Methods. 47(3):142–150. https://doi.org/10.1016/j.ymeth.2008.09.022

Posfai G, Szybalski W (1988) A simple method for locating methylated bases in DNA using class-IIS restriction enzymes. Gene. 74(1):179–181

Pratt K, Hattman S (1983) Nucleosome phasing in Tetrahymena macronuclei. J Protozool. 30(3):592–598

Privat E, Sowers LC (1996) Photochemical deamination and demethylation of 5-methylcytosine. Chem Res Toxicol. 9(4):745–750. https://doi.org/10.1021/tx950182o

Proffitt JH, Davie JR, Swinton D, Hattman S (1984) 5-Methylcytosine is not detectable in Saccharomyces cerevisiae DNA. Mol Cell Biol. 4(5):985–988

Pukkila PJ, Peterson J, Herman G, Modrich P, Meselson M (1983) Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in Escherichia coli. Genetics. 104(4):571–582

Quignard E, Fazakerley GV, Teoule R, Guy A, Guschlbauer W (1985) Consequences of methylation on the amino group of adenine. A proton two-dimensional NMR study of d(GGATATCC) and d(GGm6ATATCC). Eur J Biochem. 152(1):99–105

Rae PM (1976) Hydroxymethyluracil in eukaryote DNA: a natural feature of the pyrrophyta (dinoflagellates). Science (New York, NY) 194(4269):1062–1064

Rae PM, Spear BB (1978) Macronuclear DNA of the hypotrichous ciliate Oxytricha fallax. Proc Nat Acad Sci U S A 75(10):4992–4996

Ratel D, Ravanat JL, Charles MP, Platet N, Breuillaud L, Lunardi J et al (2006) Undetectable levels of N6-methyl adenine in mouse DNA: Cloning and analysis of PRED28, a gene coding for a putative mammalian DNA adenine methyltransferase. FEBS Lett. 580(13):3179–3184. https://doi.org/10.1016/j.febslet.2006.04.074

Razin A, Razin S (1980) Methylated bases in mycoplasmal DNA. Nucleic Acids Res. 8(6):1383–1390

Reich NO, Mashhoon N (1991) Kinetic mechanism of the EcoRI DNA methyltransferase. Biochemistry. 30(11):2933–2939

Robbins-Manke JL, Zdraveski ZZ, Marinus M, Essigmann JM (2005) Analysis of global gene expression and double-strand-break formation in DNA adenine methyltransferase- and mismatch repair-deficient Escherichia coli. J Bacteriol. 187(20):7027–7037. https://doi.org/10.1128/JB.187.20.7027-7037.2005

Roberts D, Hoopes BC, McClure WR, Kleckner N (1985) IS10 transposition is regulated by DNA adenine methylation. Cell. 43(1):117–130

Rogers JC, Rogers SW (1995) Comparison of the effects of N6-methyldeoxyadenosine and N5-methyldeoxycytosine on transcription from nuclear gene promoters in barley. Plant J. 7(2):221–233

Rogers SD, Rogers ME, Saunders G, Holt G (1986) Isolation of mutants sensitive to 2-aminopurine and alkylating agents and evidence for the role of DNA methylation in Penicillium chrysogenum. Curr Genet. 10(7):557–560

Romanov GA, Vanyushin BF (1981) Methylation of reiterated sequences in mammalian DNAs. Effects of the tissue type, age, malignancy and hormonal induction. Biochim Biophys Acta. 653(2):204–218

Russell DW, Hirata RK (1989) The detection of extremely rare DNA modifications. Methylation in dam- and hsd- Escherichia coli strains. J Biol Chem 264(18):10787–10794

Sanchez-Romero MA, Olivenza DR, Gutierrez G, Casadesus J (2020) Contribution of DNA adenine methylation to gene expression heterogeneity in Salmonella enterica. Nucleic Acids Res. 48(21):11857–11867. https://doi.org/10.1093/nar/gkaa730

Saparbaev M, Laval J (1994) Excision of hypoxanthine from DNA containing dIMP residues by the Escherichia coli, yeast, rat, and human alkylpurine DNA glycosylases. Proc Nat Acad Sci U S A 91(13):5873–5877

Sarnacki SH, Castaneda Mdel R, Noto Llana M, Giacomodonato MN, Valvano MA, Cerquetti MC (2013) Dam methylation participates in the regulation

of PmrA/PmrB and RcsC/RcsD/RcsB two component regulatory systems in Salmonella enterica serovar Enteritidis. PLoS One. 8(2):e56474. https://doi.org/10.1371/journal.pone.0056474

Sater MR, Lamelas A, Wang G, Clark TA, Roltgen K, Mane S et al (2015) DNA Methylation Assessed by SMRT Sequencing Is Linked to Mutations in Neisseria meningitidis Isolates. PLoS One. 10(12):e0144612. https://doi.org/10.1371/journal.pone.0144612

Schiffers S, Ebert C, Rahimoff R, Kosmatchev O, Steinbacher J, Bohne AV et al (2017) Quantitative LC-MS provides no evidence for m(6) dA or m(4) dC in the genome of mouse embryonic stem cells and tissues. Angew Chem Int Ed Engl. 56(37): 11268–11271. https://doi.org/10.1002/anie.201700424

Sedgwick B, Bates PA, Paik J, Jacobs SC, Lindahl T (2007) Repair of alkylated DNA: recent advances. DNA Repair (Amst). 6(4):429–442. https://doi.org/10.1016/j.dnarep.2006.10.005

Shah K, Cao W, Ellison CE (2019) Adenine methylation in drosophila is associated with the tissue-specific expression of developmental and regulatory genes. G3 (Bethesda) 9(6):1893–1900. https://doi.org/10.1534/g3.119.400023

Shapiro R, Klein RS (1966) The deamination of cytidine and cytosine by acidic buffer solutions. Mutagenic implications. Biochemistry. 5(7):2358–2362

Shen L, Song CX, He C, Zhang Y (2014) Mechanism and function of oxidative reversal of DNA and RNA methylation. Annu Rev Biochem. 83:585–614. https://doi.org/10.1146/annurev-biochem-060713-035513

Slater S, Wold S, Lu M, Boye E, Skarstad K, Kleckner N (1995) E. coli SeqA protein binds oriC in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. Cell. 82(6):927–936

Smith JD, Arber W, Kuhnlein U (1972) Host specificity of DNA produced by Escherichia coli. XIV. The role of nucleotide methylation in in vivo B-specific modification. J Mol Biol. 63(1):1–8

Srivastava R, Gopinathan KP, Ramakrishnan T (1981) Deoxyribonucleic acid methylation in mycobacteria. J Bacteriol. 148(2):716–719

Stein R, Gruenbaum Y, Pollack Y, Razin A, Cedar H (1982) Clonal inheritance of the pattern of DNA methylation in mouse cells. Proc Nat Acad Sci U S A 79(1): 61–65

Stephens C, Reisenauer A, Wright R, Shapiro L (1996) A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. Proc Nat Acad Sci U S A 93(3):1210–1214

Sternglanz H, Bugg CE (1973) Conformation of N6-methyladenine, a base involved in DNA modification: restriction processes. Science (New York, NY) 182(4114):833–834

Su SS, Modrich P (1986) Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. Proc Nat Acad Sci U S A 83(14):5057–5061

Sugimoto K, Takeda S, Hirochika H (2003) Transcriptional activation mediated by binding of a plant GATA-type zinc finger protein AGP1 to the AG-motif (AGATCCAA) of the wound-inducible Myb gene NtMyb2. Plant J. 36(4):550–564

Sun Q, Huang S, Wang X, Zhu Y, Chen Z, Chen D (2015) N(6) -methyladenine functions as a potential epigenetic mark in eukaryotes. Bioessays. 37(11): 1155–1162. https://doi.org/10.1002/bies.201500076

Sundheim O, Talstad VA, Vagbo CB, Slupphaug G, Krokan HE (2008) AlkB demethylases flip out in different ways. DNA Repair (Amst). 7(11):1916–1923. https://doi.org/10.1016/j.dnarep.2008.07.015

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y et al (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science (New York, NY) 324(5929):930–935. https://doi.org/10.1126/science.1170116

Tawa R, Ueno S, Yamamoto K, Yamamoto Y, Sagisaka K, Katakura R et al (1992) Methylated cytosine level in human liver DNA does not decline in aging process. Mech Ageing Dev 62(3):255–261

Theil EC, Zamenhof S (1963) Studies on 6-Methylaminopurine (6-Methyladenine) in Bacterial Deoxyribonucleic Acid. J Biol Chem 238:3058–3064

Trewick SC, Henshaw TF, Hausinger RP, Lindahl T, Sedgwick B (2002) Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. Nature. 419(6903):174–178. https://doi.org/10.1038/nature00908

Tronche F, Rollier A, Bach I, Weiss MC, Yaniv M (1989) The rat albumin promoter: cooperation with upstream elements is required when binding of APF/HNF1 to the proximal element is partially impaired by mutation or bacterial methylation. Mol Cell Biol. 9(11):4759–4766

Tsuchiya H, Matsuda T, Harashima H, Kamiya H (2005) Cytokine induction by a bacterial DNA-specific modified base. Biochem Biophys Res Commun. 326(4):777–781. https://doi.org/10.1016/j.bbrc.2004.11.115

Unger G, Venner H (1966) Remarks on minor bases in spermatic desoxyribonucleic acid. Hoppe Seylers Z Physiol Chem. 344(4):280–283

Urbonavicius J, Skouloubris S, Myllykallio H, Grosjean H (2005) Identification of a novel gene encoding a flavin-dependent tRNA:m5U methyltransferase in bacteria--evolutionary implications. Nucleic Acids Res. 33(13): 3955–3964. https://doi.org/10.1093/nar/gki703

Urig S, Gowher H, Hermann A, Beck C, Fatemi M, Humeny A et al (2002) The Escherichia coli dam DNA methyltransferase modifies DNA in a highly processive reaction. J Mol Biol. 319(5):1085–1096. https://doi.org/10.1016/S0022-2836(02)00371-6

van den Born E, Omelchenko MV, Bekkelund A, Leihne V, Koonin EV, Dolja VV et al (2008) Viral AlkB proteins repair RNA damage by oxidative demethylation. Nucleic Acids Res. 36(17): 5451–5461. https://doi.org/10.1093/nar/gkn519

Van Etten JL, Schuster AM, Girton L, Burbank DE, Swinton D, Hattman S (1985) DNA methylation of viruses infecting a eukaryotic Chlorella-like green alga. Nucleic Acids Res. 13(10):3471–3478
Vanyushin BF, Belozersky AN, Kokurina NA, Kadirova DX (1968) 5-methylcytosine and 6-methylamino-purine in bacterial DNA. Nature. 218(5146): 1066–1067
Vanyushin BF, Tkacheva SG, Belozersky AN (1970) Rare bases in animal DNA. Nature. 225(5236):948–949
von Freiesleben U, Rasmussen KV, Schaechter M (1994) SeqA limits DnaA activity in replication from oriC in Escherichia coli. Mol Microbiol. 14(4):763–772
Wagner I, Capesius I (1981) Determination of 5-methylcytosine from plant DNA by high-performance liquid chromatography. Biochim Biophys Acta. 654(1):52–56
Wallecha A, Munster V, Correnti J, Chan T, van der Woude M (2002) Dam- and OxyR-dependent phase variation of agn43: essential elements and evidence for a new role of DNA methylation. J Bacteriol. 184(12): 3338–3347
Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D et al (2014) N6-methyladenosine-dependent regulation of messenger RNA stability. Nature. 505(7481): 117–120. https://doi.org/10.1038/nature12730
Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, Ma H et al (2015) N(6)-methyladenosine modulates messenger RNA translation efficiency. Cell. 161(6):1388–1399. https://doi.org/10.1016/j.cell.2015.05.014
Wang W, Xu L, Hu L, Chong J, He C, Wang D (2017) Epigenetic DNA modification N(6)-Methyladenine causes site-specific RNA polymerase II transcriptional pausing. J Am Chem Soc. 139(41):14436–14442. https://doi.org/10.1021/jacs.7b06381
Wang SY, Mao H, Shibuya H, Uzawa S, O'Brown ZK, Wesenberg S et al (2019) The demethylase NMAD-1 regulates DNA replication and repair in the Caenorhabditis elegans germline. PLoS Genet. 15(7): e1008252. https://doi.org/10.1371/journal.pgen.1008252
Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL et al (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet 37(8):853–862. https://doi.org/10.1038/ng1598
Wei YF, Carter KC, Wang RP, Shell BK (1996) Molecular cloning and functional analysis of a human cDNA encoding an Escherichia coli AlkB homolog, a protein involved in DNA alkylation damage repair. Nucleic Acids Res. 24(5):931–937
Willis DB, Granoff A (1980) Frog virus 3 DNA is heavily methylated at CpG sequences. Virology. 107(1): 250–257
Wion D, Casadesus J (2006) N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. Nat Rev Microbiol. 4(3):183–192. https://doi.org/10.1038/nrmicro1350
Wold S, Boye E, Slater S, Kleckner N, Skarstad K (1998) Effects of purified SeqA protein on oriC-dependent DNA replication in vitro. EMBO J 17(14): 4158–4165. https://doi.org/10.1093/emboj/17.14.4158
Woodcock CB, Yu D, Hajian T, Li J, Huang Y, Dai N et al (2019) Human MettL3-MettL14 complex is a sequence-specific DNA adenine methyltransferase active on single-strand and unpaired DNA in vitro. Cell Discov. 5:63. https://doi.org/10.1038/s41421-019-0136-4
Wu JC, Santi DV (1987) Kinetic and catalytic mechanism of HhaI methyltransferase. J Biol Chem 262(10): 4778–4786
Wu TP, Wang T, Seetin MG, Lai Y, Zhu S, Lin K et al (2016) DNA methylation on N-adenine in mammalian embryonic stem cells. Nature. https://doi.org/10.1038/nature17640
Wyatt GR (1950) Occurrence of 5-methylcytosine in nucleic acids. Nature. 166(4214):237–238
Wyatt GR, Cohen SS (1952) A new pyrimidine base from bacteriophage nucleic acids. Nature. 170(4338): 1072–1073
Xiao CL, Zhu S, He M, Chen D, Zhang Q, Chen Y et al (2018) N(6)-Methyladenine DNA modification in the human genome. Mol Cell 71(2):306–318. e7. https://doi.org/10.1016/j.molcel.2018.06.015
Xie Q, Wu TP, Gimple RC, Li Z, Prager BC, Wu Q et al (2018) N(6)-methyladenine DNA modification in glioblastoma. Cell. 175(5):1228–1243. e20. https://doi.org/10.1016/j.cell.2018.10.006
Xiong J, Ye TT, Ma CJ, Cheng QY, Yuan BF, Feng YQ (2019) N 6-Hydroxymethyladenine: a hydroxylation derivative of N6-methyladenine in genomic DNA of mammals. Nucleic Acids Res. 47(3):1268–1277. https://doi.org/10.1093/nar/gky1218
Yamaki H, Ohtsubo E, Nagai K, Maeda Y (1988) The oriC unwinding by dam methylation in Escherichia coli. Nucleic Acids Res. 16(11):5067–5073
Yang CG, Yi C, Duguid EM, Sullivan CT, Jian X, Rice PA et al (2008) Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. Nature. 452(7190):961–965. https://doi.org/10.1038/nature06889
Yao B, Cheng Y, Wang Z, Li Y, Chen L, Huang L et al (2017) DNA N6-methyladenine is dynamically regulated in the mouse brain following environmental stress. Nat Commun. 8(1):1122. https://doi.org/10.1038/s41467-017-01195-y
Yao B, Li Y, Wang Z, Chen L, Poidevin M, Zhang C et al (2018) Active N(6)-Methyladenine Demethylation by DMAD Regulates Gene Expression by Coordinating with Polycomb Protein in Neurons. Mol Cell 71(5): 848–857. e6. https://doi.org/10.1016/j.molcel.2018.07.005
Ye P, Luan Y, Chen K, Liu Y, Xiao C, Xie Z (2017) MethSMRT: an integrative database for DNA N6-methyladenine and N4-methylcytosine generated by single-molecular real-time sequencing. Nucleic

Acids Res. 45(D1):D85–DD9. https://doi.org/10.1093/nar/gkw950

Yoder JA, Soman NS, Verdine GL, Bestor TH (1997) DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. J Mol Biol. 270(3):385–395. https://doi.org/10.1006/jmbi.1997.1125

Yue Y, Liu J, He C (2015) RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. Genes Dev 29(13):1343–1355. https://doi.org/10.1101/gad.262766.115

Yuki H, Kawasaki H, Imayuki A, Yajima T (1979) Determination of 6-methyladenine in DNA by high-performance liquid chromatography. J Chromatogr. 168(2):489–494

Zaccara S, Ries RJ, Jaffrey SR (2019) Reading, writing and erasing mRNA methylation. Nature Rev 20(10): 608–624. https://doi.org/10.1038/s41580-019-0168-5

Zaleski P, Wojciechowski M, Piekarowicz A (2005) The role of Dam methylation in phase variation of Haemophilus influenzae genes involved in defence against phage infection. Microbiology. 151(Pt 10): 3361–3369. https://doi.org/10.1099/mic.0.28184-0

Zelinkova E, Paulicek M, Zelinka J (1990) Modification methylase M.Sau3239I from Streptomyces aureofaciens 3239. FEBS Lett. 271(1-2):147–148

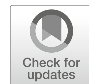
Zhang G, Huang H, Liu D, Cheng Y, Liu X, Zhang W et al (2015) N(6)-methyladenine DNA modification in Drosophila. Cell. 161(4):893–906. https://doi.org/10.1016/j.cell.2015.04.018

Zhang Q, Liang Z, Cui X, Ji C, Li Y, Zhang P et al (2018) N(6)-Methyladenine DNA methylation in Japonica and Indica rice genomes and its association with gene expression, plant development, and stress responses. Mol Plant. 11(12):1492–1508. https://doi.org/10.1016/j.molp.2018.11.005

Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, Li CJ et al (2013) ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell. 49(1):18–29. https://doi.org/10.1016/j.molcel.2012.10.015

Zhou J, Wan J, Gao X, Zhang X, Jaffrey SR, Qian SB (2015) Dynamic m(6)A mRNA methylation directs translational control of heat shock response. Nature. 526(7574):591–594. https://doi.org/10.1038/nature15377

Zhu S, Beaulaurier J, Deikus G, Wu T, Strahl M, Hao Z et al (2018) Mapping and characterizing N6-methyladenine in eukaryotic genomes using single molecule real-time sequencing. Genome Res. https://doi.org/10.1101/gr.231068.117



Pathways of DNA Demethylation

9

Wendy Dean

Abstract

The regulation of the genome relies on the overlying epigenome to instruct, define, and restrict the activities of cellular differentiation and growth integral to embryonic development, as well as defining the key activities of terminally differentiated cell types. These instructions are positioned as readers, writers, and erasers in their functional roles. Among the sizeable repertoire of epigenetic instructions, DNA methylation is perhaps the best understood process. In mammals, multiple cycles of reprogramming, the addition and removal of DNA methylation coupled with modulation of chromatin post-translational modifications (PMTs), constitute critical phases when the developing embryo must negotiate lineage specification and commitment events which serve to canalise development. During these reprogramming events the DNA methylation instruction is often removed, thereby allowing a change in developmental restriction, resulting in a return to a more plastic and pluripotent state. Thus, in germline reprogramming, DNA demethylation is essential in order to give rise to fully functional gametes which are inherited across generations and poised to restore totipotency. A similar return to a less differentiated state can also be achieved experimentally. DNA methylation constitutes one of the significant barriers to erroneous induced pluripotency, and loss of DNA methylation is a prerequisite for the generation of induced pluripotent stem cells (iPSCs). Taking fully differentiated cells, such as skin fibroblast cells or peripheral blood cells, and turning back the developmental clock by generating iPSCs constituted a technological breakthrough in 2006, offering unprecedented promise in precision regenerative medicine. In this chapter, I will explore mechanistic possibilities for DNA demethylation in the context of natural and experimentally induced epigenetic reprogramming. The balance of the maintenance of DNA methylation as a heritable mark together with its potential for timely removal is essential for lifelong health and may be key in our understanding of aging and the potential to limit or reverse that process.

Keywords

Active DNA demethylation · Passive DNA demethylation · Remodelling · Transcription-factor mediated reprogramming

W. Dean (✉)
Department of Cell Biology and Anatomy, University of Calgary, Calgary, AB, Canada
e-mail: wendy.dean@ucalgary.ca

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_9

Abbreviations

5caC	5-Carboxylcytosine
5fC	5-Formylcytosine
5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
A	Adenosine
AICDA	Activation-induced cytosine deaminase
AID	Activation-induced deaminase
APE1	Apurinic/apyrimidinic (AP) endonuclease 1
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
BER	Base excision repair
BS-seq	Bisulphite conversion coupled with next-generation sequencing
C	Cytosine
CGI	CpG islands
CHH	Asymmetric DNA methylation
CpA	Cytosine-adenosine dinucleotide
CpG	Cytosine-guanosine dinucleotide
CXXC	Zinc finger protein-binding domain to non-methylated CpG
DNA	Deoxyribonucleic acid
DNMT1	DNA (cytosine-5)-methyltransferase 1
DNMT1o	DNA (cytosine-5)-methyltransferase 1 oocyte form
DNMT1s	DNA (cytosine-5)-methyltransferase 1 somatic form
DNMT3a	DNA (cytosine-5)-methyltransferase 3a
DNMT3b	DNA (cytosine-5)-methyltransferase 3b
DNMT3L	DNA (cytosine-5)-methyltransferase 3-like
DNMTs	DNA methyltransferases
E13.5	Embryonic day 13.5
E6.5	Embryonic day 6.5
EGFP	Enhanced green fluorescent protein
ELP1	Elongator complex protein 1
ELP3	Elongator complex protein 3
ELP4	Elongator complex protein 4
ESCs	Embryonic stem cells
G	Guanosine
GV	Germinal vesicle
GVOs	Germinal vesicle oocytes
H3K9me2	Histone H3 lysine 9 dimethylation
H3K9me3	Histone H3 lysine 9 trimethylation
IAP	Intracisternal A particle
IF	Immunofluorescence
iPSCs	Induced pluripotent stem cells
KO	Knockout
MBD2	Methyl-CpG-binding protein 2
MBD4	Methyl-CpG-binding protein 4
NER	Nucleotide excision repair
NGS	Next-generation sequencing
PARP1	Poly-ADP-ribose polymerase 1
PGCs	Primordial germ cells
PMTs	Post-translational modifications
RNA	Ribonucleic acid
RNAi	RNA interference
RRBS	Reduced representation bisulphite sequencing
SAM	*S*-adenosyl-methionine
SCNT	Somatic cell nuclear transfer
siRNA	Small interfering RNA
SMUG	Single-strand selective monofunctional uracil DNA glycosylase 1
T	Thymine
TDG	Thymine DNA glycosylase
TET1/2/3	Ten-eleven-translocation methylcytosine dioxygenases 1, 2, and 3
U	Uracil
UNG	Uracil DNA glycosylase
XRCC1	X-ray repair complementing defective repair in Chinese hamster cells 1

9.1 DNA Methylation: One Building Block of the Epigenome

The genomic DNA sequence constitutes the blueprint of life. However, the generation of the 200 plus cell types which comprise the tissues and organs of the body requires another layer to support this diversification and interpret the underlying genetic code. This is the role of the epigenome.

One of the central pillars of the epigenome is the modification of cytosine by the addition of a methyl group at the C5 position of the pyrimidine base. This modification was discovered early in total hydrolysates of calf thymus DNA which identified a satellite to the cytosine spot on chromatograms. Thereafter, this was dubbed the fifth base of DNA (Hotchkiss 1948). This fifth base, 5-methylcytosine (5mC), was abundant and differed in its prevalence in a species- and tissue-specific manner (Ehrlich et al. 1982; Kothari and Shankar 1976). Deeper biochemical evaluation determined that methylcytosine was found in a CpG dinucleotide context. This symmetrical dinucleotide configuration has dominated our thinking related to the genomic nature of DNA methylation and the mechanism of its modulation, both gain and loss, over the last 45 years.

The machinery which adds DNA methylation to cytosines has been extensively studied and largely comprises a small group of related enzymatic activities, the DNA methyltransferases (DNMTs). This small cadre of DNA methylating proteins also includes non-enzymatic family members and essential co-factors along with the requisite *S*-adenosyl L-methionine methyl donor. Details of the DNA methyltransferases are investigated and outlined in other chapters of this book. Importantly, DNA methyltransferases are involved in diverse multi-functional partnerships via the key BAH and CXXC DNA binding domain of DNMT1, and the ADD and PWWP domains of the de novo DNMT3a/b/c enzymes, which bind chromatin and sequester specific histone modifications together with other proteins. Together with a wide selection of splice variants, principally among DNMT3b enzymes, these activities possess an expanding potential to create a wide division of labour for DNA methylation in the vertebrate genome. However, these mechanisms and activities are also at the centre of how DNA methylation may be lost.

In any discussion of DNA demethylation reprogramming, it is important to consider the role of DNA methylation. DNA methylation is regarded to have two non-exclusive functions. The first is the regulation of gene expression either in a developmentally timed or tissue-specific context. Here the most common occurrence involves DNA methylation as negatively correlated to gene expression. That is, DNA methylation turns off genes and in most textbooks you will read that DNA methylation is involved in tissue-specific gene expression (Holliday and Pugh 1975). This includes X chromosome inactivation, a ubiquitous requirement in female mammals, and genomic imprinting. Early experiments indicated that DNA methylation was also associated with the induction of changes in the state of cellular differentiation on exposure to experimental situations where 5-aza-cytidine, an inhibitor of DNA methylation, promoted the generation of muscle cells from non-muscle precursors through erosion of 5mC (Constantinides et al. 1977). This highlighted two epigenetic principles, which would dominate the field for many years: first, that transcription was negatively correlated with DNA methylation and, second, that the loss of DNA methylation resulted in the loss of heritable cell fate, frequently associated with tumourigenesis (Jaenisch and Bird 2003). In a related fashion, the second role for DNA methylation is in genome defence where the vast non-genic content of highly repeated sequences including retrotransposons is kept silenced or stabilised by the presence of DNA methylation (Bestor 1998). Importantly, this stabilisation includes the centromeric satellite repeats essential for chromosome integrity and segregation in mitosis (Lehnertz et al. 2003; Saksouk et al. 2014). The gene regulatory role of DNA methylation has led to the pervasive concept that it silences gene expression. In a developing embryo, this is congruous with the idea that sequential DNA methylation of genes that are no longer required is an essential element of cellular differentiation and lineage restriction as part of cellular commitment. These differential requirements add to the complexity of reprogramming and serve to explain the multifaceted mechanisms, both passive and active, which may be required during genomic resetting as part of reprogramming.

The link between DNA demethylation and development is long-standing. In a model system

of mammalian development, such as embryonal carcinoma cells, a multi-potent lineage-specific cell line isolated from teratocarcinomas, significant loss of DNA methylation (approx. 30%) on retinoic acid induction promoted cellular differentiation (Wei and Lee 1994). Germ cells, such as sperm, were found to be highly methylated, while extra-embryonic tissues including yolk sac and placenta were hypomethylated (Razin et al. 1984; Sanford et al. 1985). These observations exemplify the wide spectrum under which DNA methylation exerts its influence. Collectively, these early examples established that lineages undergo DNA demethylation during development (Razin et al. 1984). A loss of methylation had already been reported to be present in transformed cells and in tumours (Riggs and Jones 1983). These early examples formalised, very broadly, the concept of the two mechanistic possibilities for DNA demethylation. In the simplest of terms, DNA demethylation can be consigned to two clearly defined categories: active demethylation that requires a 'demethylase' and passive demethylation that engages the machinery around DNA replication and the associated activities of DNA maintenance methylation. These very clear distinctions formed the basis for understanding all examples of demethylation, at times creating impassioned schisms among the chief proponents (Ooi and Bestor 2008).

9.2 DNA Methylation Reprogramming: Setting the Epigenome Up for Success

In mammals, DNA methylation is reprogrammed at least twice across the life course. One phase takes place during the establishment of the germline, while the second is triggered immediately post-fertilisation as part of an enduring cycling that perpetuates genetic material across generations (Reik et al. 2001). In this respect, the idea of the immortality of the germline is formed, driven by the reiterative cycling of DNA methylation and demethylation. As such, much of the investigation of passive and active demethylation has focused around these cycles, especially in the mouse. However, these processes are not only integral to mouse development, but other mammals including humans rely on similar programs to achieve these same outcomes (Kurimoto and Saitou 2018).

Germline reprogramming, one of the key windows for DNA demethylation during development in mammals, serves the vital role of resetting the non-Mendelian imprinted genes. These genes are required, in part, for embryonic growth control and are expressed in a parent-of-origin manner. Hence, they must undergo erasure of the DNA methylation that marks the imprinted regulatory regions in each generation (Reik et al. 2001). Coupled with the erasure of genomic imprints, a number of germline-specific genes, such as *Dazl*, must undergo methylation erasure in order to be expressed, thus ensuring the complete development of gametes (Seisenberger et al. 2012). However, the temporal kinetics of the erasure is tightly regulated, such that methylation of germline-expressed genes is erased earlier (E10.5). This erasure supplies the terminal differentiation needed for gametes and precedes that of imprinted genes (from E11.5) which sets up future functional requirements and completes this reprogramming phase (Maatouk et al. 2006). These multiple trajectories may even require both active and passive mechanisms to operate simultaneously across the genome as well as at specific loci.

A second period of demethylation is initiated immediately on fertilisation. While quantitative and qualitative differences have been reported across Mammalia, the mechanisms are largely conserved (Dean et al. 2001; Young and Beaujean 2004). Owing to the extensive resource tools available to investigate mechanistic pathways, the mouse model has been most comprehensively studied and has revealed a number of activities involved in passive and active DNA demethylation. In recent years, technological advancements have been key in revealing a much richer and more intricate temporal change in DNA methylation during the preimplantation period of development. In this window, the fully mature gametes give way to the newly formed

zygote coupling chromatin remodelling with loss of DNA methylation from both the male and female genomes. This process culminates by setting up a hypomethylated landscape to coincide with the first lineage decisions that delimit the preimplantation stage of development. The early observation that the paternal contribution to the zygote, the remodelled sperm, and thereafter the male pronucleus underwent a rapid and extensive loss of methylation prior to replication inspired the field that an activity, a 'demethylase', might be present (Mayer et al. 2000; Oswald et al. 2000; Santos et al. 2002).

9.3 Active DNA Demethylation: The Hunt for the 'Demethylase'

The definition of active DNA demethylation is straightforward. It refers to the loss of DNA methylation in the absence of DNA replication. In cycling cells, this requires the proof that the methylation loss must occur outside of the 'S phase' or that it is faster than 50% loss per cell division. Importantly, for a direct demethylation event, a C-C bond would have to be broken. This direct one-step conversion of 5mC to cytosine is not energetically favourable and to date has not been documented in any organism. In order to fulfil this criterion, the mechanism must conform to something where it can be envisioned that the methyl group is directly 'snipped' away from the base, leaving behind cytosine. Another variation might include one where the base is stripped away leaving behind an abasic site, which then needs some type of DNA repair to restore the cytosine. This mechanism is very reminiscent of the plant family of activities, DEMETER and ROSI DNA glycosylases, which recognise the methylated base and remove it via the base excision repair (BER) pathway. In animals, no 5meC glycosylase has been found and the mechanisms requiring glycosylases only work when the 5meC base is first altered, such as by deamination, in advance of glycosylase removal (Drohat 2021). Enzymes which bind DNA and catalyse chemical reactions on it do so by distorting, bending, or kinking the B-form helices. Some of these enzymes, such as the DNMTs, achieve this by flipping the base out of the helix and modifying it, thus creating a methylated nucleotide. This led to the idea that the reverse reaction might occur in order for demethylation to take place. Base flipping is also used by glycosylases in order to gain access to the DNA bases (Cheng and Roberts 2001). These ideas supplied a variety of possibilities and gave clues to finding and confirming the existence of 'DNA demethylases' in animals.

The study of DNA methylation and its modulation has been ongoing for more than 45 years. Early studies using chromatography and nearest neighbour analysis could only have reported on total changes in 5-methylcytosine (Bird 1980). This was superseded by the use of CpG methylation-sensitive isoschizomers of restriction endonucleases that gave some sequence context but was severely limited by insufficient genomic information (Cedar et al. 1979). Breakthroughs came in the combination of differential methylation sensitivity and the polymerase chain reaction (Herman et al. 1995) coupled with base pair resolution afforded by bisulphite sequencing (BS-seq) (Clark et al. 1994; Frommer et al. 1992). These technological advancements were followed by refinements to sample sizes by orders of magnitude culminating now in the routine use of single-cell technologies (Denomme et al. 2012; Gravina et al. 2015; Gu et al. 2019; Niemoller et al. 2021). However, the field experienced a true paradigm shift with the mouse genome project in 2002 (Mouse Genome Sequencing Consortium et al. 2002), the human genome project in 2003 (International Human Genome Sequencing Consortium 2004), and later with the advent of next-generation sequencing in 2009 (Metzker 2010; Meyerson et al. 2010).

9.4 Direct DNA Demethylation

If DNA methylation is negatively correlated with gene expression, then the identification of a demethylase that regulates expression would serve as an integral part of mechanistic understanding of gene regulation. Beyond that, a demethylase might serve important roles

therapeutically in the targeted re-expression of key genes involved in developmental and cellular processes. With the advent of the spectrum of CRISPR/Cas9 technologies, tethering activities that afford DNA methylation removal and potentiate gene reactivation are now a viable option (Taghbalout et al. 2019; Xu et al. 2016). Arguably, among some of the most obvious targets for demethylation are the tumour suppressor genes frequently altered by inappropriate acquisition of DNA methylation found in precancerous and transformed cells of tumours (Herman et al. 1995). In nearly all tumours, cells have a signature where tumour suppressor genes are hypermethylated and repressed or silent, while the remainder of the genome is regarded as hypomethylated, genomically unstable, and subject to aneuploidy (Feinberg and Vogelstein 1983; Gama-Sosa et al. 1983). In this context, a concerted effort to identify and characterise a 'demethylase' had attracted much interest beyond the biological roles establishing the methylation landscape during development.

The publication of the discovery of an activity which could act as a 'demethylase' in 1999 was received with great interest and a degree of scepticism in the biomedical field and the burgeoning field of epigenetics (Bhattacharya et al. 1999). The existence of such an activity was highly controversial, as it had been regarded that the direct loss of the methyl group was chemically impossible. This new protein possessed the hallmark of such an activity with a methyl-CpG-binding domain and a reported demethylation activity on methylcytosine, in this context from artificially methylated plasmid sequence. Moreover, in the course of this enzymatic reaction the methyl group in the 5 position of the cytosine base was reported to be removed as methanol. This methyl binding protein had been cloned previously and identified as MBD2. Despite the flurry of interest in this protein and the demethylation process, independent verification of direct demethylation, and hence demethylase activity, has never been reported (Boeke et al. 2000; Ng et al. 1999; Wade et al. 1999). Aside from this controversy, MBD2 has been implicated in indirect demethylation function in a wide variety of autoimmune diseases in mice and humans including demethylation of a master regulatory gene Foxp3 in T-regulatory cell populations (Wang et al. 2013). The contribution of MBD2 to demethylation during the immediate post-fertilisation period prior to the first zygotic S phase seems less likely, as oocytes null for MBD2 undergo loss of DNA methylation from the male pronucleus in a kinetically similar manner to the control population (Santos et al. 2002).

9.5 Indirect Loss of DNA Methylation

The energetically unfavourable direct conversion of 5mC to C prompted investigation into alternative pathways that could otherwise achieve the de facto loss of DNA methylation, leaving behind the idea of 'demethylation via demethylase' activity. While direct 5-methylcytosine-dependent DNA glycosylases have not been found in animals, two dominant pathways have now emerged and have since been well studied and supported widely across multiple biological experimental systems. In both cases, the first step involves the modification of the 5mC base, either by deamination via the cytosine deamination family of activities or via singular or iterative oxidation steps directly targeted to the 5mC modification. In both instances, and indeed, in combination, the resolution of the further modified methylcytosine relies on DNA repair mechanisms which involve either base excision repair (BER) or nucleotide excision repair (NER) activities (Fig. 9.1).

9.5.1 Role of Cytosine Deamination in DNA Demethylation

One of the mechanisms which have shaped the mammalian genome is the process whereby the amine group (–NH_2) of the nucleotide base is removed. Hydrolytic deamination of cytosine is a prominent form of DNA damage and takes place at 60–500 residues per genome per day. This wide variation is a consequence of the

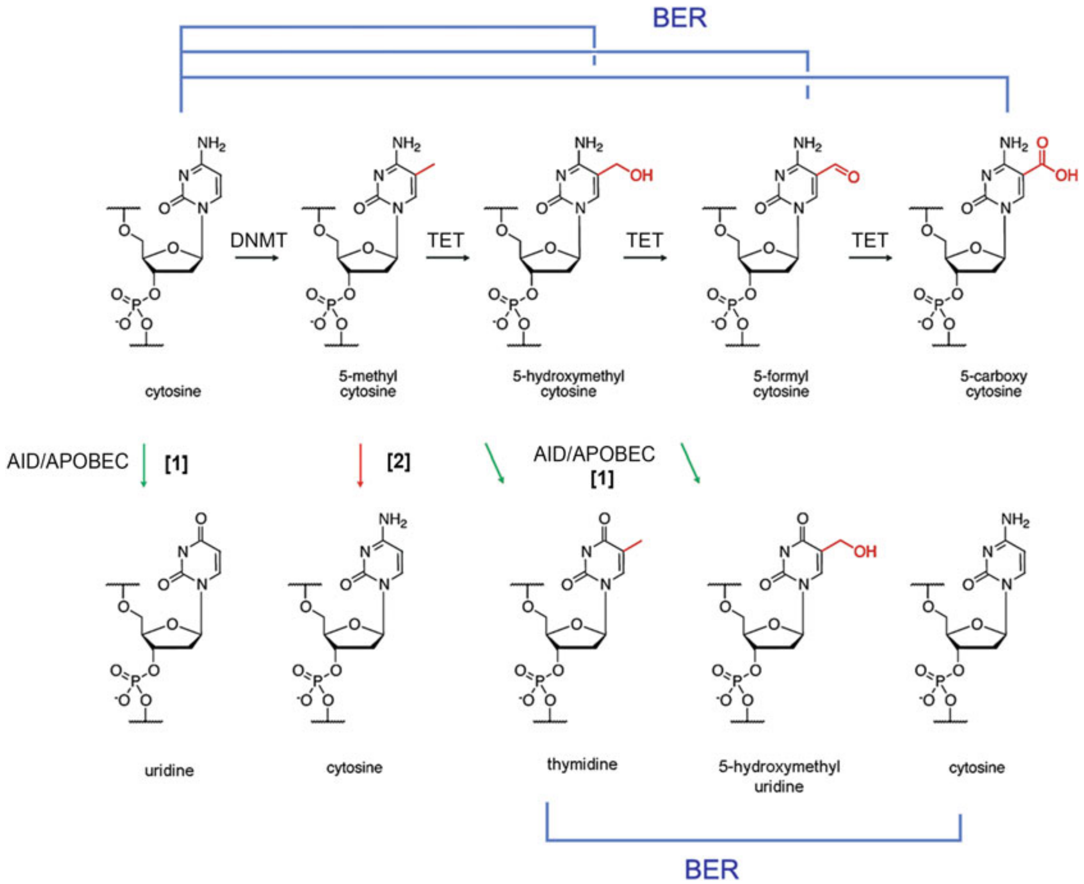


Fig. 9.1 Mechanisms of active demethylation. Cytosine is methylated by DNA methyltransferases (DNMTs) conferring genomic stability and serving as one layer of chromatin required for gene regulation. Much of what we have learned has been integral to the biology of pluripotent stem cells and in mammalian development. These processes are also accompanied by demethylation via a number of potential pathways. Cytosine may be altered by **[1]** AID/APOBEC-mediated deamination leaving uracil and 5-methylcytosine may likewise undergo deamination to thymidine. This pathway may be repaired to restore the cytosine residue by one of a number of base excision repair (BER) pathways. The 5mC residue may also be directly demethylated to cytosine although this is regarded as enzymatically unfavourable **[2]**. Another central mechanism to facilitate loss of DNA methylation occurs via TET-mediated oxidation of the methyl group to give 5-hydroxymethylcytosine (5hmC) that may in turn undergo further iterative oxidation to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). This reaction is Fe (II)/α-ketoglutarate-dependent and directly enhanced by vitamin C. In the absence of glycosylases that can detect this mismatch, 5fC may be further oxidised to 5-carboxylcytosine (5caC). BER mechanisms may aid in the repair of the cytosine base at each step of the process. The loss of methylation from cytosine may involve both pathways. In this case, 5mC is first oxidised to 5hmC and subsequently deaminated to 5-hydroxymethyluracil

different rates of deamination which occur with a more than 200-fold preference on single-stranded versus double-stranded DNA substrate (Lindahl 1993). This is particularly relevant for the pyrimidine base cytosine and especially the case for its modified version 5mC, as deamination leads to the genomic base substitution, if left unrepaired, known as the C to T transition. This propensity for alteration has had a significant impact on the mammalian genome and its unique sequence composition (Bird 1980).

Among the genomic signatures caused by cytosine deamination is the suppression of the expected frequency of CpG dinucleotides. The

genomic frequency of CpG sites in the human genome is closer to 1 in 24 rather than the predicted 1 in 16. This altered frequency occurs as a consequence, over evolutionary time, of the transition mutation of methylcytosine to thymidine owing to the hydrolytic deamination of the cytosine base where the fidelity of the repair process is apt to slip. Indeed, deamination forms the basis of the chemical modification used in bisulphite mutagenesis sequencing to discriminate between DNA methylation status, where cytosine is read as T and 5mC a C (Frommer et al. 1992). These two concepts led to the suggestion that an endogenous activity, either enzymatic or chemical such as through hydrolytic deamination, could serve as candidates capable of achieving functional loss of DNA methylation.

Whether spontaneous hydrolytic deamination alone would be competent to clear DNA methylation from the male pronucleus of the fertilised oocyte during the period of active demethylation is not known. However, by all accounts the rate of this chemical reaction seems far too slow to account for the observations arising from both immunofluorescence (IF) studies and that of contemporary molecular analyses of DNA methylation. In contrast, enzymatic deamination coupled with DNA repair processes offers a diverse and more feasible solution to explain the rate and magnitude of active loss of DNA methylation in the zygote. In mammals, deamination of cytosine and/or methylcytosine is achieved by a small family of highly related enzymatic activities. Cytosine deaminases are encoded by three enzyme families, APOBEC1, APOBEC3 (APOBEC2 is non-enzymatic (Conticello et al. 2005)), and AICDA. The activation-induced cytosine deaminase (AICDA or AID) takes its name from B-cell activation as part of the adaptive immune response, resulting in its functional role in somatic hypermutation and class switch recombination which permits antibody diversification (Longerich et al. 2006). The APOBEC family of activities are broadly involved in RNA editing and as such play a role in genome defence to restrict the movement of non-LTR and LTR retrotransposable elements, including long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) (Chiu and Greene 2008). While there is a family of deaminases, the implied function in active demethylation has fallen exclusively to AID.

9.5.2 Methylcytosine Oxidation-Based Demethylation Mechanisms

In the same way as deamination modifies 5mC to thymidine, the oxidation of the 5-methyl group can rapidly convert 5mC into 5-hydroxymethylcytosine (5hmC) without any further alteration to the double-stranded DNA molecule. Under favourable conditions this process may proceed through further iterative oxidative steps to form 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al. 2010; Ko et al. 2010). Should a decarboxylase be present or simply a repair cascade triggered, this pathway could mediate the conversion of 5mC to cytosine and the demethylation would be completed. The oxidative portion of this process is accomplished variously by the ten-eleven-translocation (TET) dioxygenase family, comprised of three active forms of the enzymes TET1-3. They all share a conserved function and use 5mC as a substrate in conjunction with 5-α-ketoglutarate and Fe^{2+} as co-factors (Kriaucionis and Heintz 2009; Tahiliani et al. 2009). As such, these enzymes are influenced by the metabolic state of the cell and the concentration of each oxidative intermediate is dependent on a number of other co-factors, including vitamins A and C, as well as the oxygen concentration of the tissue or cells (Blaschke et al. 2013; Chen et al. 2013). Despite the powerful potential of the TET family of activities to oxidise 5mC, mice null for *Tet1* or *Tet2* have little phenotypic effects either during embryogenesis or postnatally. Only upon combining deficiencies in *Tet1* and *Tet2* double null mutations semi-penetrant mid-gestational phenotypes can be observed. On combination of *Tet1/2/3* triple knockouts, the full impact of TET deletion becomes apparent. *Tet1/2/3* triple null embryos fail to undergo gastrulation likely due to hypermethylation of promoters in the LEFTY–

NODAL axis required for establishing early embryonic polarity (Dai et al. 2016). TET proteins may operate together with other 'loss of methylation' pathways creating a much more complex and diversified means of epigenetic regulation. Interestingly, these enzymes may serve wider biological roles beyond their canonical enzymatic function in keeping with their evolutionary secondment into a demethylation pathway.

Site-specific, feature-specific, or ultimately genome-wide loss of DNA methylation may require more than a single pathway to achieve a change in the epigenome. While some activities operate exclusively during the replicative portion of the cell cycle, others operate through a combination of DNA modifications leading to loss of methylation often by repair-coupled processes. This is the case with cytosine deamination and oxidative modification of 5mC (Fig. 9.1). The order of modifications may also change, permitting yet further complexity and specificity at target loci. This may widen and even accelerate the rate of repair or alteration with additional functional consequences. For example, deamination of 5mC leads to T which may be further altered or repaired, e.g., by thymidine DNA glycosylase (TDG). Alternatively, 5mC may be oxidised to 5hmC, which may be deaminated to give rise to 5-hydroxymethyluracil (5hmU) which is similarly repaired by TDG. In either case, loss of DNA methylation would be the outcome upon resolution of the repair process (Kohli and Zhang 2013).

9.6 Chromatin Remodelling, DNA Replication, and Repair: The Epigenetic Triumvirate

Whenever chromatin is remodelled, there is a likelihood of the occurrence of endogenous DNA damage. As such, the genome-wide remodelling during epigenetic reprogramming is particularly susceptible to DNA damage, especially in replicating cells where the replication fork can become a powerful nucleation site for activities focused on genomic fidelity including the DNA methyltransferases which maintain the epigenome. Both natural and experimental reprogramming, which activates new transcriptional requirements, may trigger the need for a DNA damage response. Collectively, these may result in functional demethylation and could become a source of cellular heterogeneity in development.

In the window defined in the immediate post-fertilisation period in the mouse, male and female gametes are remodelled in order to form a functional metaphase plate, the first of all the subsequent cellular divisions for life in mammals. This remodelling is essentially required in order to re-equilibrate the chromatin of the oocyte with that of the protamine-configured sperm nucleus (Brewer et al. 1999). In order to package sperm into a functional structure able to fertilise an ovulated oocyte, the genome must be confined to the dehydrated toroidal structure configured by protamines, something more akin to a virus than a cell (Braun 2001). This is the state of play when the largest cell, the oocyte, meets the smallest, the highly methylated sperm. Very shortly following fusion and sperm penetration, remodelling of the sperm genome ensues in the shared environment of the oocyte. Indeed, within one hour of fertilisation the protamine-encased chromatin has been replaced and the paternal genome becomes configured with the oocyte-derived histones (McLay and Clarke 2003). In this early remodelling phase significant DNA methylation is lost by paternal alleles (Dean et al. 2001; Santos et al. 2002). In contrast, the female pronucleus that forms on completion of meiosis is less subject to loss of DNA methylation (Santos et al. 2002). Quantification of the female compartment even under the most stringent conditions indicates a wide variance in DNA methylation and suggests that chromatin turnover also takes place on maternal alleles but is overall methylation neutral, losing and gaining DNA methylation in equal proportion (Santos et al. 2013).

Despite the introduction of a histone-based chromatin in the male pronucleus, the male and female components of the zygote remain epigenetically different (Probst et al. 2007; Santos et al.

2005). Post-translational modifications (PTMs) of the chromatin retain differences between the pronuclei beyond that of DNA methylation. The PTMs endow the female pronucleus with most of the modifications found in a somatic cell, while notably the male pronucleus has neither H3K9me2 nor H3K9me3 (Santos et al. 2005; van der Heijden et al. 2005). Both of these marks are critical to maintain the coupling of DNA methylation to these chromatin modifications, usually associated with transcriptional silencing, and are essential for reinforcing chromosome stability (Guenatri et al. 2004). The absence of H3K9me2 and H3K9me3 PTMs from the chromatin in the male pronucleus is thought to facilitate the loss of DNA methylation. In turn, this suggested that one or both of these marks may well be associated with protecting the maternal genome from extensive DNA demethylation (Nakamura et al. 2007; Szabo and Pfeifer 2012). Interestingly, the male pronucleus accumulates 5hmC licensed by cell cycle progression to S phase. This is a consequence of de novo DNA methylation which is immediately oxidised by TET in S phase. In contrast, the female pronucleus possesses significantly less of the 5hmC modification acquired during oogenesis (Amouroux et al. 2016; Santos et al. 2013). Interestingly, and in keeping with the growing appreciation of the complex interleaved dependency of DNA methylation and demethylation processes, maternal deletion of Dmmt3a/3b abrogates an increase in DNA methylation, but more importantly 5hmC in the male pronucleus. Thus, while the rapid paternal loss of 5mC is attributed to the presence of TET3, the process is dependent on a de novo DNA methyltransferase activity (Iqbal et al. 2011). However, whether the 5mC turnover is required for subsequent development and what mechanisms protect 5mC from demethylation in the female pronucleus is not fully known. The maternal effect protein, STELLA, is responsible to safeguard the oocyte methylome from overt DNA hypermethylation in a UHFR1-DNMT1-dependent manner (Li et al. 2018). Here, upon deletion of STELLA, UHRFI accumulates in the nucleus driving DNMT1 mislocalisation. Thus, STELLA functions to prevent aberrant hypermethylation de novo by DNMT1 in oocytes. Recently, DNMT1 has been reported to have in vitro and in vivo de novo methyltransferase function with specific retrotransposon targeting, a long suspected additional function which has been difficult to fully establish (Bestor and Ingram 1983; Haggerty et al. 2021).

Several groups have knocked out the TET3 catalytic activity with varying phenotypic results (Gu et al. 2011; Peat et al. 2014). All agree that the abrogation of the catalytic activity results in the loss of 5hmC immunofluorescence signal in the male pronucleus but a concomitant increase in 5mC was not always observed (Santos et al. 2013; Wossidlo et al. 2011). The absolute loss of TET3 is non-viable and homozygous null mice die shortly after birth. Maternal-specific oocyte deletion revealed a subtler and nonetheless variable effect. Maternal deletion of TET3 failed to have a significant effect on oocyte development and maturation and oocytes were fertilisable and pregnancies proceeded with only minor issues (Peat et al. 2014; Tsukada et al. 2015). Several reports highlighted that the failure to generate 5hmC per se is not responsible for the developmental defects, e.g. neonatal sub-lethality due to Tet3 haploinsufficiency in maternally deficient oocytes (Inoue et al. 2015; Tsukada et al. 2015). Interestingly, genome-wide DNA methylation analysis of the *Tet3* knockout revealed an unexpected role for this protein in protecting CpG islands from DNA methylation (Peat et al. 2014). This intriguing result reminds us that the machinery of DNA demethylation is complex, highly interleaved, and multi-dimensional. The protection of the female pronucleus gives us an insight into the overall mechanism of the regulation of the loss of DNA methylation and its role in setting up zygotic totipotency.

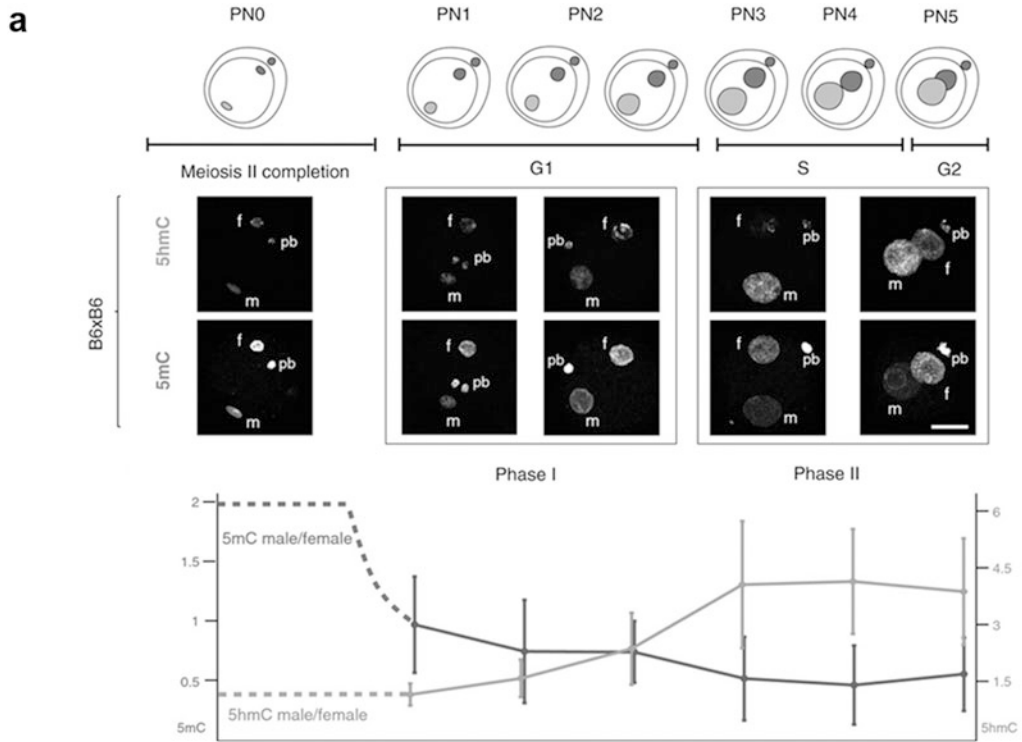
DNA repair has long been implicated in the process of active DNA demethylation especially in the mouse zygote. The mammalian oocyte is replete with many of the activities that are required across the myriad repair pathways supplying roles in maintaining the genome during replication and remodelling (Derijck et al. 2006, 2008; Zheng et al. 2005). In the special case of

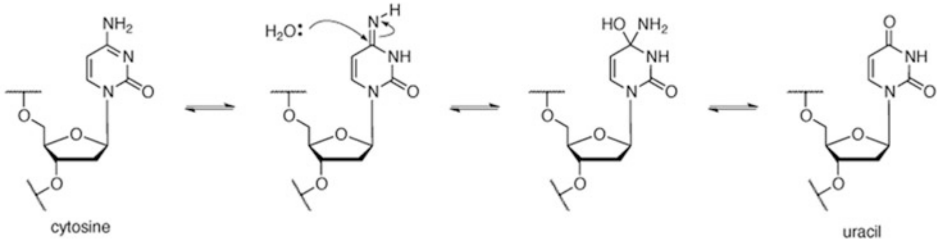
active CpG demethylation, more focus has been placed on activities that can read and repair the mismatched base pairs derived from C-G. It is now well established that both BER and NER pathways play a central role in active demethylation in the mouse zygote (Hajkova et al. 2010; Wossidlo et al. 2010). Inhibition of critical activities in BER, such as PARP1 [poly (ADP-ribose) polymerase 1] and APE1 (apurinic/apyrimidinic endonuclease), attenuated DNA demethylation with significantly higher levels of DNA methylation in the paternal pronucleus as judged by 5mC staining (Hajkova et al. 2010). Staining of XRCC1 (X-ray repair complementing defective repair in Chinese hamster cells 1) revealed high levels of bound protein exclusive to the male pronucleus, and hence ssDNA breaks were detectable from an early pronuclear stage (PN3), which coincides with the onset of DNA demethylation. Collectively, these pathways account for some of the early demethylation in the zygote; yet these studies failed to address the question of what initiates the pathways. DNMT3a/3b are abundant in oocytes and early embryos (Hirasawa and Sasaki 2009; Lucifero et al. 2007) and have been implicated in deamination roles in human cells (Kangaspeska et al. 2008; Metivier et al. 2008). However, whether they serve as the upstream initiator of the cascade has not been tested directly. Additional reports highlighted that de novo methylation is required for the replication-licensed oxidation of C-5 and generation of 5hmC in the male pronucleus (Amouroux et al. 2016). In contrast, conformational changes, such as those required during chromatin remodelling in the zygote, may be sufficient to act upstream of a DNA damage response and may be able to trigger repair mechanisms (McLay and Clarke 2003).

While it is well accepted that both deamination and TET-mediated oxidative processes are central to active DNA demethylation, it has taken some time to identify that they likely target C and 5mC, respectively. That is, DNA demethylation pathways use deamination targeting C, while TET-mediated demethylation employs oxidative processes to target 5mC. Consequently, these processes have an impact on different genomic regions. In either case, the outcome is still loss of DNA methylation. Here, the preferred target for deamination is cytosine and the resulting uracil becomes the repair-based lesion. Should methylated cytosines reside in proximity to the C-U mismatch, long patch repair loss of methylation would result from the replacement of 5mC with C (Santos et al. 2013).

Apart from deamination and oxidation-linked processes of DNA demethylation, the transcription elongation complex 3 (ELP3) and family members ELP1 and 4 have been identified in systematic screens and reported to act via an otherwise unknown mechanism to effect active demethylation. Using live cell imaging as a platform to capture the event, RNAi mediated depletion of ELPs in conjunction with expression of a CXXC-EGFP fusion protein reporter, able to bind unmethylated CpG-rich regions, accumulated in the male pronucleus suggesting a loss of methylation at this time (Okada et al. 2010). Why this should result in paternal-specific demethylation is not immediately obvious. Moreover, in light of reports suggesting that some turnover in DNA methylation occurs in the female compartment, the exclusive accumulation in the male might not be expected (Guo et al. 2014). Given that this phase of development is characterised by transcriptional silencing (Bouniol-Baly et al. 1999), the suggestion that ELPs play a role is somewhat counterintuitive and direct biochemical evidence is still lacking (Wu and Zhang 2014).

Several groups have investigated the kinetics of the loss of 5mC and tied it to the acquisition of 5hmC, but have failed to notice that this process took place in two clearly defined phases. Santos et al. used a genetic approach to test the possibility that both AID and TET3 could act in the ‘demethylation’ pathways (Santos et al. 2013). The dynamics and magnitude of the loss of DNA methylation were measured by semi-quantitative IF methods first in the wild-type (WT) zygotes (Fig. 9.2a) and then using a constitutive AID null mutant. In the AID null mice, the early phase of demethylation proceeded as in the WT, but by the post-replication stage PN5 a residual gain of methylation was evident. This suggested that an AID-dependent loss of DNA





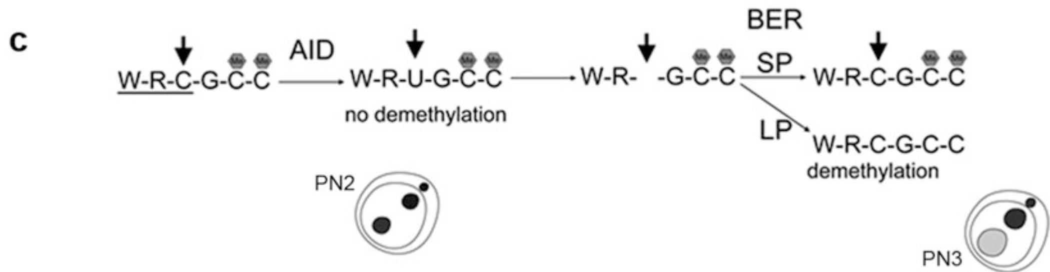


Fig. 9.2 Active demethylation during reprogramming in the zygote. Following fertilisation the maternal and paternal pronuclei are remodelled along two independent pathways. (**a**) Chromatin is reconfigured in stages across time points defined by the size and position of the respective pronuclei (PN_0 to PN_5). The paternal pronucleus is

methylation from the male pronucleus was possible in the first cell cycle in the mouse. But how can the temporal shift in the kinetics be explained? The answer came by way of an ingenious mechanism that fits well the temporal licensing of AID and the biochemical substrate preference data. AID ordinarily works during G1 phase of the cell cycle, a time consistent with an active demethylation process. Moreover, although AID can work on 5mC, its preferred substrate is C, as the enzymatic configuration does not work when the C5 position is derivatised (Nabel et al. 2012). Thus, it was reasoned that deamination of C to U (Fig. 9.2b), in close proximity to 5mC, would trigger DNA repair via a long patch BER mechanism and would in turn result in functional DNA demethylation (Fig. 9.2c). This mechanism would involve uracil DNA glycosylases (UNG and SMUG) rather than TDG, a result confirmed using a maternally deleted TDG null zygote (Santos et al. 2013). UNG and SMUG are abundant in oocytes, while TDG is not readily detectable. UNG null zygotes showed a reduced demethylation by IF (Santos et al. 2013) and hypermethylation of *Nanog* and *Line1* in the paternal alleles and of *Dnmt3b* and *Zbtb32* in the female (Xue et al. 2016).

The study also closely followed the kinetics of the specific acquisition of 5hmC as a function of the reciprocal loss of 5mC from the male pronucleus. Conditional deletion of *Tet3* from the oocyte, together with experiments inhibiting DNA replication with aphidicolin, clearly demonstrated that 5hmC was only present from S phase onwards and was coincident with, but not coupled to, DNA replication (Santos et al. 2013). Together, this ground-breaking study asserted that multiple and independent pathways were involved in paternal DNA demethylation in the zygote. These observations have been conclusively and independently confirmed in the follow-up studies using genetic and biochemical approaches together with IF, where both 5mC and 5hmC were measured using mass spectrometry (Amouroux et al. 2016; Okamoto et al. 2016). In support of these suggested mechanisms, two additional studies have also proposed that long patch repair may facilitate regional loss of 5mC (Franchini et al. 2014) adjacent to an AID binding site, i.e. WRC (W = A/T; R = A/G) or via a similar mechanism through repair of 5-hydroxymethyluracil (5hmU) arising from oxidative modification of T by TET1 (Pfaffeneder et al. 2014). One of the 11 glycosylases, Nei endonuclease VIII-like 1 and 3 (NEIL1) has DNA glycosylase/lyase activity towards mismatched uracil and thymine, in particular in U:C and T:C mismatches, and it specifically binds 5hmC, suggesting that it acts as a specific reader of 5hmC in advance of repair. As such, activities targeting both C and T via separate pathways may result in regional demethylation.

During the late fetal stages and early on in postnatal stages, oogenesis proceeds and DNA methylation is reacquired. This re-establishes the essential DNA methylation at imprinted loci needed for growth and development during the post-implantation phase. Some of this

Fig. 9.2 (continued) observed to lose methylation in two distinct phases. In the first phase, 5mC is rapidly lost with little or no change in 5hmC marking the end of G_1. The onset of the second phase at S phase of the cell cycle (PN_3 – PN_5) coincides with the appearance of 5hmC, a process that is TET3 dependent. Across both phases the maternal pronucleus retains DNA methylation and possesses very low levels of 5hmC and little overall change takes place. The maternal pronucleus resists loss of methylation in a Stella/PGC7-dependent manner. One mechanism proposed to achieve loss of DNA methylation in the paternal pronucleus is by deamination depicted in **b**. Here cytosine is modified through loss of the $-NH_2$ group to uracil triggering a BER cascade involving a short patch repair (SP) (**c**). The AID target sequence WRC marks the position of the deaminated cytosine in a place adjacent to 5mC residues. This mechanism envisions that in Phase I of demethylation the cytosine base is deaminated and the resection of the local 5mC only takes place in Phase II in a replication-activated process resulting in functional demethylation. Schematic depictions of fertilised oocytes indicate that initially both compartments have similar methylation levels that lead to paternal-specific loss of 5mC following short patch repair via BER

methylation is modified in a TET1-dependent manner, becoming 5hmC. This establishes, in part, the chromatin configuration for maternal alleles at fertilisation. Together with H3K9me2, the maternal chromatin is actively protected from post-fertilisation DNA demethylation by the presence of the protein STELLA/PGC7 that forms an inaccessible chromatin configuration maintained post-fertilisation (Pfaffeneder et al. 2014). While STELLA is found in both male and female pronuclei, the binding of the protein in maternal pronuclear chromatin confers protection against demethylation (Nakamura et al. 2007). Maternal deletion of Stella in the oocyte reinstates a substrate that is subject to TET3-dependent demethylation in both compartments and leads to lethality early on in development. Thus, DNA demethylation can be modulated through the configuration of the chromatin and may not simply be a function of presence or absence of the demethylating activity. Interestingly, there is little evidence that the loss of methylation from paternal alleles is required for development to proceed. The paternal-specific demethylation has been widely observed in mammals though often to varying degrees, yet an absolute demonstration of its requirement remains elusive. Experimental reconstitution of the zygote using round spermatids, which still possesses histones on paternal chromatin, does not lead to remodelling or demethylation and can give rise to a full-term viable mouse (Kishigami et al. 2006; Polanski et al. 2008).

Indeed, this raises the question of whether the extensive remodelling and dynamic loss of DNA methylation in the paternal compartment are rather a secondary consequence of remodelling. Across Mammalia many species do not apparently undergo paternal-specific active demethylation shortly after fertilisation (Barnetova et al. 2010; Beaujean et al. 2004a, b). However, experimental manipulation of oocytes from species of mammals which do not ordinarily undergo overt demethylation of the paternal pronucleus showed that paternal nuclei are subject to demethylation when they are artificially fertilised with sperm from other species (Beaujean et al. 2004b). Thus, in sheep oocytes, where there is little or no paternal-specific active demethylation, mouse sperm is extensively remodelled and appears to be demethylated as measured by IF. The reciprocal is also true; fertilisation of mouse oocytes with sheep sperm reveals extensive remodelling of the sperm nucleus including the specific loss of methylation from the paternal pronucleus (Barnetova et al. 2010; Beaujean et al. 2004b). These experimental examples typify the general trend that changes of cellular state such as development and cellular reprogramming often have an obligate loss of DNA methylation accompanying these transitions.

While many studies have addressed the question of the specific loss of paternal DNA methylation in the fertilised zygote, few have taken a quantitative approach. Using a small-scale mass spectrometry approach, Okamoto et al. (2016) were able to supply an absolute metric for 5mC and 5hmc levels without the ambiguity often confounded by bisulphite mutagenesis studies. Following active loss from the paternal pronucleus (~40%) and a small active loss from the female, the overall trajectory for 5C and 5hmC across preimplantation regresses with a best fit model that follows replication, suggesting that this loss of methylation occurs via passive DNA demethylation (Okamoto et al. 2016). The active loss of DNA methylation from the paternal pronucleus confirmed previous bisulphite analysis that had asserted this process in the maternal compartment as well (Guo et al. 2014). As such, maternal and paternal haploid contributions to the zygote are equalised in genomic 5mC content by 24 hours post-fertilisation (Okamoto et al. 2016). Finally, this study found no evidence for the reciprocal loss of 5mC and commensurate gain of 5hmC claimed in some studies (Okamoto et al. 2016; Wossidlo et al. 2011).

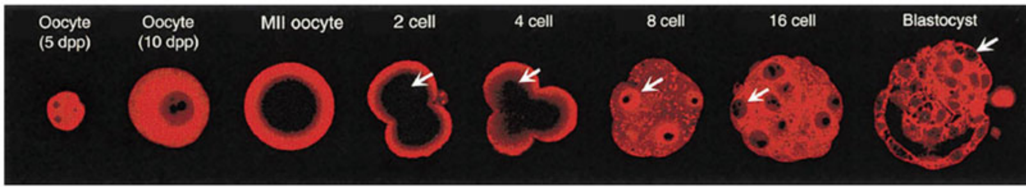
9.7 Replication-Coupled Loss of DNA Methylation: Passive Demethylation

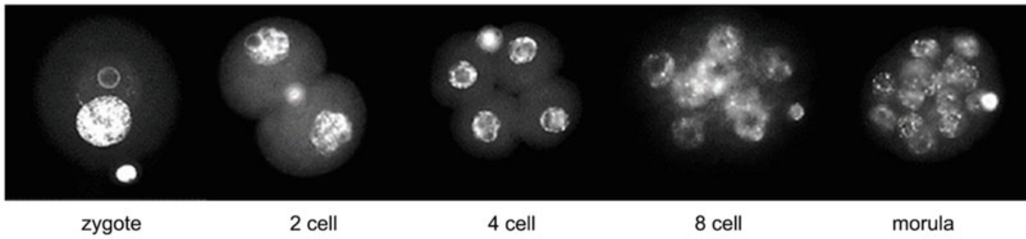
One of the best studied reprogramming events in development is the establishment and resetting of the germline in mammals. For the establishment

of the germline, precursor somatic cells of the epiblast must become reprogrammed, such that the somatic gene expression pattern is inhibited, giving way to the early expression of primordial germ cell markers. At this time, somatic epiblast cells are highly methylated with approximately 70% of CpGs modified. While the need for reprogramming of the germline has been known for several decades, the availability of only a few cells and the limitations of the available technology restricted what we could learn about this process (Monk et al. 1987). Early studies were confined to characterising maternally imprinted genes and highly repetitive sequences including centromeric satellites and repeat family sequences following their progression during germline maturation in mouse (Hajkova et al. 2002; Lane et al. 2003). The generality of this requisite reprogramming event could later be confirmed in other mammalian model systems (Hyldig et al. 2011). Early genome-wide analyses combined histology together with immunofluorescence detection using the 5mC antibody to give a semi-quantitative overview of early germ cells (Seki et al. 2005, 2007). Relying on an OCT4- or STELLA/PGC7-EGFP reporter to identify the germ cells, the low-resolution approach was sufficient to establish that at E7.5 the newly specified germ cells were highly methylated. However, over the next 24-h period this methylation was significantly reduced and thereafter no longer detectable up to E12.5. While inconclusive regarding a mechanism for the loss of methylation, these studies gave single-cell resolution and hinted at a variable level of methylation likely to arise from a highly repetitive family or families of genomic sequence.

Next-generation sequencing supplied the breakthrough needed for a deeper mechanistic understanding. This has been the impetus for the myriad studies which have described the detail of the reprogramming and programming phases of mammalian germline development. Whole genome-bisulphite sequencing was used to comprehensively profile DNA methylation of primordial germ cell (PGCs) from both wild-type and AID KO mice. The scope of this study was limited by the availability of sufficient sample DNA to generate libraries, and hence, E13.5 was elected for the PGC samples. Comparison of the DNA methylation status of WT with AID null gave the first high-resolution unbiased epigenome from PGCs (Popp et al. 2010). Interestingly, in the absence of AID, E13.5 female PGCs had three times more methylation (~ 21%) than the WT (~7% CpG methylation). Differences in male PGCs were less dramatic. The effect was genome-wide, except for CGI promoters, suggesting that AID had some role to play in the reprogramming machinery in germ cells. Despite this breakthrough, little or no resolution of the mechanism of DNA methylation loss was forthcoming although AID seemed to be involved.

One of the significant limitations of the Popp et al. (2010) study was the low genomic coverage. However, technological improvements to handle small sample sizes together with deep sequencing coverage were deployed quickly thereafter allowing for a systematic evaluation of the DNA methylation profile across multiple stages of PGCs. OCT4-EGFP-positive PGCs were isolated from selected stages from E9.5 through E13.5 during PGC development in both male and female embryos where appropriate. Genome-wide DNA methylation was analysed using BS-Seq and deep sequencing. High-depth coverage and bp resolution generated the first unbiased genome-wide survey of DNA methylation in PGCs across development. Moreover, this study had the potential to answer the question of the nature of demethylation mechanisms during germline reprogramming. Combining strand-specific bisulphite sequencing analysis and modelling of stage-specific data, the outcome was clear-cut. Germline reprogramming proceeded by a mechanism dominated by passive loss of DNA methylation through progressive replication up to the time of mitotic/meiotic arrest around E14 (Seisenberger et al. 2012) (Fig. 9.3c). These results confirmed an independent study using DNA methylation immunoprecipitation (meDIP) followed by a microarray (Guibert et al. 2012). This validation was particularly important owing to the ambiguity of calling 5mC and 5hmC by BS-seq. Evaluating the role of hydroxymethylation became important to





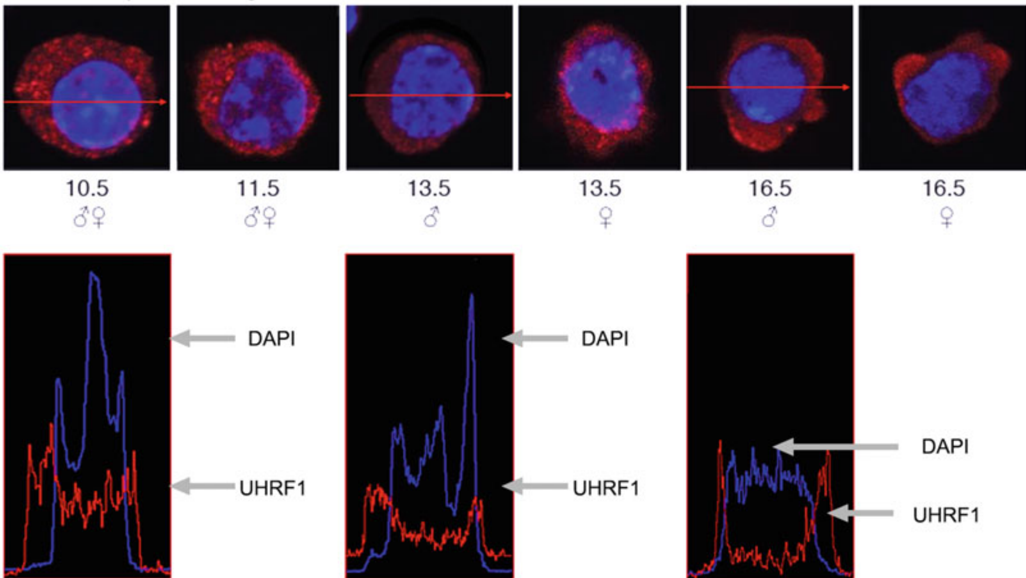


Fig. 9.3 Passive demethylation during preimplantation and germline reprogramming. Very early on the discovery that DNMT1 accumulated in the subcortical region of the oocyte and that of preimplantation stage embryos established the expectation that maintenance methylation was interrupted during preimplantation. (**a**) This panel illustrates this principle. DNMT1 staining is clearly abundant and excluded from the nucleus up until the 8-cell (8C) stage when it is found in the nucleus, thus establishing the conditions for passive demethylation. Beyond the 16-cell stage the protein is again found excluded from the nucleus, something which also included the blastocyst stage (**b**). Immunofluorescence detection of 5mC suggested that there was a diminishing quantity of DNA methylation over successive replication cycles across preimplantation stages in keeping with passive demethylation. Here this distribution of 5mC is reported from the zygote up to the morula stage. (**c**) A similar mechanism may operate in primordial germ cells to achieve passive demethylation. Staining primordial germ cells from E10.5 up to E16.5 indicated that UHRF1, an obligate chaperone for DNMT1 and an essential component of the maintenance machinery, is relegated to the cytoplasmic compartment. Despite DNMT1 residing in the nucleus, the cross-sectional tracing below identifies UHRF1 (arrow) depleted in the nucleus (marked by an arrow; stained with DAPI) and found in the cytoplasm. These conditions are requisite for passive demethylation and in keeping with the genome-wide loss of DNA methylation in the reprogramming of the germline. Panel **a** is reprinted from Cell *104*, Howell et al., (2001) Genomic

resolve the profiles observed during the passive loss of methylation. Primordial germ cells express high levels of TET1 and TET2 and these activities are licensed by DNA replication; however, the 5hmC mark which they generate from the 5mC is not maintained or read at CpG dinucleotides and hence may passively decline. Loss of 5mC via hydroxylation was thus found to contribute to asynchronous erasure of imprinted methylation in PGCs between E9.5 and E 10.5. However, thereafter, the genome average of the loss of DNA methylation obeyed kinetics in keeping with passive demethylation (Hackett et al. 2013). In contrast, in vitro generation of differentiated PGCs (iPGCs) where *Tet1* and *Tet2* were experimentally depleted or deficient indicated that iPGCs were unaffected by the absence of these activities and the lack of genomic 5hmC (Vincent et al. 2013). However, some loci were consequently hypermethylated, which necessitated an alternative loss of methylation mechanism to generate functional iPGCs. Indeed, germline erasure is sufficiently important that multiple and overlapping mechanisms of erasure have arisen to ensure the process and restrict transgenerational inheritance of epimutations (Hackett et al. 2012). In fact, the extent of the functional redundancy for loss of methylation pathways may even stretch to include multiple and overlapping mechanisms operating even at the same genomic locus (Ohno et al. 2013).

If we can measure the dynamic process of DNA demethylation at high resolution and derive a model that indicates that passive demethylation is the favoured mechanism, can we explain how this mechanism might come about? Kurimoto et al. (2008) isolated BLIMP1-positive PGC precursors and profiled RNA from them over the early stages of PGC establishment and commitment in order to attempt to understand the basis for passive demethylation. These cells were characterised by the downregulation of pathways associated with DNA methylation maintenance—DNMT1 and UHRF1, and BER pathways (including a role for TDG) in combination with loss of methylation by TET oxidation and AICDA deamination (Kurimoto et al. 2008). The selective staining of staged PGCs for DNMT1 and UHRF1 suggested that while DNMT1 remained in the nucleus, UHRF1 was cytoplasmic, a configuration in keeping with passive demethylation (Fig. 9.3c) (Seisenberger et al. 2012). Whether the exclusion of UHRF1 despite nuclear DNMT1 can account for the loss of 5mC in PGCs has not been tested. However, DNMT1 was inefficiently recruited in replicating gonadal germ cells, consistent with progressive loss of methylation during maturation (Ohno et al. 2013).

9.8 Resetting and Erasure of the Germline: A Barrier Against Transgenerational Inheritance

9.8.1 Demethylation During Preimplantation Development

Sanford et al. (1984) showed that sperm was highly methylated, while oocytes were less so and that following fertilisation DNA methylation declined progressively (Sanford et al. 1984). These methylation-sensitive Southern blots, although restricted in the genomic loci which could be interrogated, were accurately revealing locus-specific changes during development which included both hyper- and hypomethylation during this dynamic phase. In fact, Monk et al. (1987) confirmed and extend these observations using as few as 200 cells, showing dynamic and temporal specific regulation of DNA methylation and demethylation during early mouse development (Monk et al. 1987). However, molecular analysis, at this time, could not provide a genome-wide nuclear architectural perspective.

Fig. 9.3 (continued) Imprinting Disrupted by a Maternal Effect Mutation in the Dnmt1 Gene, pp. 829–838 with permission from Elsevier. Panel **b** is taken from Santos et al. (2013) Fig. 1 cited herein

The first glimpse of the dynamic genome-wide distribution of DNA methylation at the single cell level was reported not in mouse but in bovine. Investigating stage-specific chromosome spreads throughout the preimplantation development using a 5-methylcytidine antibody led to the conclusion of the absence of active demethylation and hence that the progressive loss of 5mC occurred by a passive mechanism (Bourc'his et al. 2001; Rougier et al. 1998). The visual compartmentalisation of DNMT1, the machinery of maintenance methylation, by IF had already predicted such an outcome (Fig. 9.3a). Antibodies raised against DNMT1 had already identified that the highly abundant protein was largely relegated to the cytoplasmic compartment of the zygote and throughout preimplantation and that DNMT1 appeared to be tethered to the subcortical region by an active process (Cardoso and Leonhardt 1999; Carlson et al. 1992). This striking and stark distribution offered a mechanism consistent with the observed conclusion that passive demethylation operated up to the blastocyst stage (Fig. 9.3b) (Carlson et al. 1992; Howell et al. 2001; Ratnam et al. 2002).

During preimplantation reprogramming the 'genome-wide' loss of methylation must be more complex and nuanced because some regions must retain their methylation. This is in contrast to the virtually full erasure and resetting observed during germline reprogramming. This specifically affects germline differentially methylated regions (DMRs), including the maternally imprinted genes, which during oogenesis are restored in a DNMT3a/DNMT3L-dependent manner (Kaneda et al. 2010). Transcriptional activity during this period stockpiles transcripts in order to maintain the oocyte's integrity as to supply an immediate store of protein for the completion of meiosis at fertilisation and the initiation of embryogenesis. This includes a number of maternal-specific transcripts that supply the oocyte with an extraordinary amount of protein, including DNMT1, both the somatic (DNMT1s) and the oocyte (DNMT1o) form, and UHRF1, an essential part of the DNA methylation maintenance machinery.

Advances in molecular profiling of DNA methylation have provided a more detailed understanding of the genome-wide methylation reprogramming during preimplantation in mouse and human. Reduced representation bisulphite sequencing (RRBS) is a technique that offers deeper sequencing of CpG-rich regions with the trade-off of lower genomic coverage. This technique covers 5–10% of the genome and favours CpG islands and the promoters of genes and hence is ideally suited for methylation profiling of imprinted genes. In contrast to earlier studies done by IF, RRBS studies in the mouse and human confirmed a rapid and global demethylation of paternal alleles post-fertilisation, but a much more limited loss of methylation to maternal alleles, such that at the blastocyst stage some DMRs were maintained (Guo et al. 2014; Smallwood et al. 2011; Smith et al. 2012). Applying a whole genome bisulphite sequencing (WGBS) approach revealed further details. Many of the 1600 DMRs inherited from the oocyte were only partially demethylated during preimplantation, indicative of mechanisms resistant to demethylation at play (Kobayashi et al. 2012). A similar study in human blastocysts found that maternal alleles were demethylated to a much lesser extent than in mouse, with notable locus and species differences attributed to both the de novo methyltransferase and demethylation machinery (Okae et al. 2014). Interestingly, in one of the earliest studies to probe the question of loss of methylation, Howlett and Reik (1991) found very little change in the methylation of L1 repeats after the zygote stage until the blastocyst stage following maternal chromosomes derived from parthenogenetic mouse embryos (Howlett and Reik 1991).

Passive DNA demethylation can occur when the methylation machinery is physically excluded or degraded from the substrate in each replication cycle, but other explanations are possible. Loss of methylation may occur when the ability to recognise the substrate is lost (Inoue et al. 2011; Inoue and Zhang 2011). This is the situation that arises in the presence of 5hmC and in the presence of asymmetric methylation of cytosine. Asymmetric DNA methylation has recently been recognised and new biological functions are being discovered especially in non-replicating cells in the

brain (Lister et al. 2009). Asymmetric DNA methylation has been characterised and mechanistically documented in plants, but functional roles in mammals have only just started to emerge. Non-canonical DNA methylation is found in the form of CHH (where H=A, C or T) and CpHpG, representing the non-symmetrical and symmetrical sequences, respectively. The most abundant non-symmetrical configuration for DNA methylation is CpA which is found in the male germline during the fetal maturation period, where DNMT3b is expressed at a time when imprinted methylation is first returned (Ichiyanagi et al. 2011). This mark is temporary, as the mitotic expansion of spermatocytes does not allow for the maintenance of the asymmetrical methylation and hence it disappears over time. In a similar manner, during the re-establishment of DNA methylation during oogenesis, the abundance of DNMT3a results in the acquisition of extensive CpA methylation. However, owing to the absence of replication in oogenesis, any accumulation of altered or aberrant bases ordinarily removed by replication does not happen until after ovulation and fertilisation has taken place. As such, the mammalian oocyte may accumulate a significant level of methylated CpAs, which remains until after fertilisation when methylation is progressively eroded away in each replication cycle (Shirane et al. 2013; Tomizawa et al. 2011). In the zygote and preimplantation embryo, DNA demethylation thus follows a path whereby the 5hmC of the male and the 5mCpA in the female are both subject to passive reduction irrespective of the cytoplasmic exclusion of the maintenance machinery, because neither 5hmC nor 5CpA can be maintained after DNA replication (Dean 2014; Okamoto et al. 2016).

9.9 Removing the Molecular Escapement Mechanism to Cell Fate and Aging by Modulation of DNA Methylation: How Cells Can Turn Back Time

The lessons learned from the activities used during the natural reprogramming cycles of development have given insight into the processes of cellular specialisation. In this context quantitative epigenetic marks, especially DNA methylation, and its modulation have been critical in establishing windows of cellular plasticity which represent developmental junctures for lineage establishment and their canalisation. Experimental reprogramming includes a wide spectrum of cellular transition states. These include (1) the transition of ‘primed’ but pluripotent embryonic stem cells (ESCs) to the naïve ground state possessing a wider cellular plasticity; (2) somatic cell nuclear transfer (SCNT)—returning a fully differentiated cell to totipotency; and (3) the generation of induced pluripotent stem cells from skin fibroblast or peripheral blood cells, the breakthrough technology at the very centre of precision medicine and regenerative medicine. Each of these changes in cellular plasticity are potentiated by the loss of DNA methylation via one mechanism or another.

The application of small molecule inhibitors, the so-called 2i, which facilitate the transitions to naïve pluripotency by uncoupling the ERK1/2 and the GSK signalling pathways, results in a remarkably rapid and quantitatively significant change in genomic DNA methylation (Lee et al. 2014). Genome-wide bisulphite sequencing profiling indicated that while most of the genomic features quickly lost methylation, major satellite repeats of the centromere, the IAP class of retrotransposons, and imprinted genes were ostensibly resistant to this change. Initial findings from the RNA-seq analyses noted a striking downregulation of the de novo methylation machinery, while the maintenance pathway, in particular DNMTs and UHRF1, was largely unchanged or upregulated as was the oxidative pathway mediated by TETs. This downregulation of the de novo methylation machinery was thought to be a function of the profound upregulation of *Prdm1*, an important negative transcriptional regulator of de novo methyltransferase genes. The upregulation of *Prdm14* was regarded as a satisfactory explanation accounting mechanistically for the observed rapid and widespread demethylation. In this transition, DNA methylation is reduced by 50%

(5mC/C) in 24 h (Ficz et al. 2013; Leitch et al. 2013; Yamaji et al. 2013). The conclusion that a wholly passive mechanism of DNA demethylation via reduction of the de novo DNA methylation machinery accounts for this rapid loss has been challenged as the two de novo methyltransferases *Dnmt3a/3b* would be unable to act this rapidly and to this extent. In fact, *Dnmt3a/3b* double knockout ESCs take upwards of 30 passages to become fully demethylated (Schubeler 2012). More congruous with the rapid decline of DNA methylation was the suggestion of the erosion in the fidelity of the maintenance methylation processes (Ficz et al. 2013). The best fit of the data suggested a combination of all three mechanisms, i.e. passive loss of DNA methylation, reduced activity of DNMT3a/3b, and TET-mediated oxidative pathways, albeit, with only a minor effect at most loci.

Somatic cell nuclear transfer (SCNT) has been likened to the reprogramming of the zygote to totipotency. Reduction of DNA methylation in the somatic donor nucleus results in improved reprogramming of the somatic DNA methylation landscape and the consequent enhancement of cloning efficiency (Samiec and Skrzyszowska 2018). DNA methylation reinforces somatic cellular memory of cell fate and hence presents the architectural configuration adopted by the nuclear donor. This nuclear configuration is completely foreign to the cytoplasm of the recipient oocyte with the dominating nuclear structure focused around the heterochromatic chromocentric organisation. This demonstrated inability of the fertilised oocyte to remodel somatic nuclear structure highlights that competency of remodelling activities, especially those of demethylation, is cell type- and chromatin type-specific. Depending on the epigenetic status of the donor nucleus, the same panoply of demethylation machinery components may not be effective, and reprogramming may be inefficient and developmentally flawed. Recent studies have explored the tightly regulated relationship of the TET enzymatic machinery to that of the de novo DNA methyltransferases, DNMT3a/3b. This overt partnership is reminiscent of the self-same conjunction of activities required in the generation of 5hmC in the paternal compartment of the zygote in the first cell cycle (Amouroux et al. 2016). TET enzymes are precisely recruited to the thousands of somatic enhancers that remain highly methylated in pluripotent cells in the presence of de novo methyltransferases. Thus, in changes of state, the potential for DNA demethylation is also influenced by the physical machinery and the need for recommissioning of DNA methylation reprogramming factors to enhancers in a pluripotency-dependent fashion (Charlton et al. 2020).

The generation of induced pluripotent cells has been one of the transformative discoveries in the post-genomic era. The expression of a small defined cocktail of early embryonic transcription factors, OCT4, SOX2, KLF4, and c-MYC (OSKM), demonstrated that the linear progression of development in the forward only direction could be interrupted and turned in a retrograde direction to revert differentiated cells types to an earlier pluripotent cellular state (Takahashi and Yamanaka 2006). This paradigm shift heralded the dawn of regenerative medicine which, in conjunction with CRISPR/Cas9 gene editing technology, renders possible the potential for personalised medicine for the first time (Jinek et al. 2012). Early on in this process, the question of whether or not this reprogramming was achieved by DNA demethylation mechanisms, and if so which ones, was mooted. Induced pluripotent reprogramming involves two clear phases: a transient global demethylation step which involves an AID-dependent downregulation of UHRF1, suggesting a passive loss of DNA methylation, and an AID-independent targeted demethylation process of enhancers and super-enhancers. In combination, these two phases lower the DNA methylation-mediated epigenetic barrier and potentiate transcription of the pluripotency network (Milagre et al. 2017). TET1 overexpression is able to replace OCT4 in the OSKM cocktail for reprogramming. Here this replacement is based solely on its enzymatic function (Gao et al. 2013) and TET1 targets enhancers and super-enhancers for demethylation which need to be activated to bring about cell-type change. During iPSC

reprogramming, demethylation-associated activities such as TET1 and TDG are essential, as mouse embryonic fibroblasts deficient in these factors fail to be reprogrammed owing to a failure in the mesenchymal-to-epithelial transitions (Hu et al. 2014). Multiple DNA demethylation pathways have been reported to operate during iPSC reprogramming in mouse and humans and the obligate expression of the methylation-demethylation activities, and reallocation of DNA methylation to somatic enhancers is a quintessential reprogramming step during the acquisition of pluripotency (Charlton et al. 2020).

Recent studies have queried the intersectional significance of TET1 activity during iPSC reprogramming. The ability of TET1 to demethylate enhancers coupled with its transcription factor functions positions this enzyme as essential for the demethylation of specific targets in both reprogramming phases, while also responding to transcriptional regulatory signals through OCT4 binding sites in its promoter to prime the pluripotency circuitry during reprogramming (Bartoccetti et al. 2020). These waves straddle the timing of first acquisition of pluripotency which progresses eventually into the 'naïve' state when other demethylation activities may engage to affect the loss of DNA methylation. These transitions are typified by the switching of stage-specific promoters. TET enzyme function has been found to be responsive to the context of flanking regions, over several orders of magnitude, and hence influences a spectrum of biological targets, contributing to genome-wide heterogeneity during reprogramming (Adam et al. 2022).

The ability of lineage-specific transcription factors to activate tissue-specific enhancers takes place via their capacity to sequester chromatin remodelling activities including those that demethylate DNA. This activation, whether direct or indirect, takes place through the transcription factor-dependent recruitment of TET proteins. Forced overexpression of tissue-specific transcription factors initiated by DNA demethylation of relevant enhancers leads ultimately to a stable DNA methylation-mediated cell fate identity. Interestingly, once the demethylation-dependent change to relevant enhancers has taken place, there is no longer a requirement for the continued presence of these enzymatic /transcription factors (Reizel et al. 2021). This stable balance is reinforced through the loss of DNA methylation by TET proteins together with the methylation reacquisition by DNMT3a/3b de novo methyltransferases (Charlton et al. 2020). The idea that transcription factors recruit the demethylation machinery to specific lineage enhancers as a requirement for reprogramming has been gaining mechanistic momentum (Charlton et al. 2020; Ginno et al. 2020; Sardina et al. 2018). Here, multiple demethylation pathway activities, including TET1, TDG, and the nucleotide repair activity XPC, function as critical co-activators of the stem cell-specific transcription factors OCT4 and SOX2. These repair activities are coupled with transcriptional coactivators to regulate transcription by cooperating with TDG to stimulate active demethylation across enhancers and regulatory regions of pluripotent stem cells (Ho et al. 2017). The interaction of XPC enhanced TDG turnover and led directly to improved somatic cell reprogramming and the quality and robustness of iPSCs.

Combining the requirements of somatic cell reprograming to iPSCs thus couples the multiple phases of DNA demethylation to achieve an objectively younger cellular state. Deletion of TET1 has been associated with telomere loss in differentiated cells, pointing to a definitive protective role of TET1 in cellular senescence and in the resetting of epigenetic age in iPSCs (Bartoccetti et al. 2020). Using the measure of the 'epigenetic clock' based on DNA methylation additions and subtractions, somatic cellular reprogramming does reset the de facto measure of aging profoundly. Based on the strictures of the 'epigenetic clock', reprogrammed iPSCs are estimated to be of an age equivalent to zero (Horvath 2013).

The importance of the many and varied mechanisms which effect DNA demethylation continues to broadly inform basic biological understanding. During development at least two reprogramming cycles take place, one in the germline and the other during preimplantation

development. These set up the programming cycles that establish the next generation. Here, demethylation mechanisms acting alone and in concert with DNA repair add and subtract instructive layers which direct chromatin and transcription factors to undertake the cell fate and lineage specifying decisions of early development. Many of these factors serve multiple functions and move seamlessly between enzymatic modification of 5mC and gene regulatory roles by binding to critical enhancer elements in a stage- and tissue-specific manner. These self-same activities operate to facilitate experimental reprogramming, driving cellular reprogramming to generate iPSCs. This is the foundational starting point for regenerative medicine, the underpinning technology transforming precision medicine into clinical care transformation. DNA demethylation and drugs/small molecule inhibitors which can affect it have been, and will continue to be, vital in the treatment of cancers as well as in the support and maintenance of healthy aging. In all of these facets, the machinery of DNA demethylation remains a centrally unifying theme.

References

Adam S, Bracker J, Klingel V, Osteresch B, Radde NE, Brockmeyer J, Bashtrykov P, Jeltsch A (2022) Flanking sequences influence the activity of TET1 and TET2 methylcytosine dioxygenases and affect genomic 5hmC patterns. Commun Biol 5:92. https://doi.org/10.1038/s42003-022-03033-4

Amouroux R, Nashun B, Shirane K, Nakagawa S, Hill PW, D'Souza Z, Nakayama M, Matsuda M, Turp A, Ndjetehe E et al (2016) De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nat Cell Biol 18:225–233. https://doi.org/10.1038/ncb3296

Barnetova I, Fulka H, Fulka J Jr (2010) Epigenetic characteristics of paternal chromatin in interspecies zygotes. J Reprod Dev 56:601–606. https://doi.org/10.1262/jrd.09-172a

Bartoccetti M, van der Veer BK, Luo X, Khoueiry R, She P, Bajaj M, Xu J, Janiszewski A, Thienpont B, Pasque V et al (2020) Regulatory dynamics of Tet1 and Oct4 resolve stages of global DNA Demethylation and transcriptomic changes in reprogramming. Cell Rep 30 (2150-2169):e2159. https://doi.org/10.1016/j.celrep.2020.01.065

Beaujean N, Hartshorne G, Cavilla J, Taylor J, Gardner J, Wilmut I, Meehan R, Young L (2004a) Non-conservation of mammalian preimplantation methylation dynamics. Curr Biol 14:R266–R267. https://doi.org/10.1016/j.cub.2004.03.019

Beaujean N, Taylor JE, McGarry M, Gardner JO, Wilmut I, Loi P, Ptak G, Galli C, Lazzari G, Bird A et al (2004b) The effect of interspecific oocytes on demethylation of sperm DNA. Proc Natl Acad Sci U S A 101:7636–7640. https://doi.org/10.1073/pnas.0400730101

Bestor TH (1998) The host defence function of genomic methylation patterns. Novartis Found Symp 214:187–195; discussion 195–189, 228–132. https://doi.org/10.1002/9780470515501.ch11.

Bestor TH, Ingram VM (1983) Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proc Natl Acad Sci U S A 80: 5559–5563. https://doi.org/10.1073/pnas.80.18.5559

Bhattacharya SK, Ramchandani S, Cervoni N, Szyf M (1999) A mammalian protein with specific demethylase activity for mCpG DNA. Nature 397: 579–583. https://doi.org/10.1038/17533

Bird AP (1980) DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res 8:1499–1504. https://doi.org/10.1093/nar/8.7.1499

Blaschke K, Ebata KT, Karimi MM, Zepeda-Martinez JA, Goyal P, Mahapatra S, Tam A, Laird DJ, Hirst M, Rao A et al (2013) Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature 500:222–226. https://doi.org/10.1038/nature12362

Boeke J, Ammerpohl O, Kegel S, Moehren U, Renkawitz R (2000) The minimal repression domain of MBD2b overlaps with the methyl-CpG-binding domain and binds directly to Sin3A. J Biol Chem 275:34963–34967. https://doi.org/10.1074/jbc.M005929200

Bouniol-Baly C, Hamraoui L, Guibert J, Beaujean N, Szollosi MS, Debey P (1999) Differential transcriptional activity associated with chromatin configuration in fully grown mouse germinal vesicle oocytes. Biol Reprod 60:580–587. https://doi.org/10.1095/biolreprod60.3.580

Bourc'his D, Le Bourhis D, Patin D, Niveleau A, Comizzoli P, Renard JP, Viegas-Pequignot E (2001) Delayed and incomplete reprogramming of chromosome methylation patterns in bovine cloned embryos. Curr Biol 11:1542–1546. https://doi.org/10.1016/s0960-9822(01)00480-8

Braun RE (2001) Packaging paternal chromosomes with protamine. Nat Genet 28:10–12. https://doi.org/10.1038/ng0501-10

Brewer LR, Corzett M, Balhorn R (1999) Protamine-induced condensation and decondensation of the same DNA molecule. Science 286:120–123. https://doi.org/10.1126/science.286.5437.120

Cardoso MC, Leonhardt H (1999) DNA methyltransferase is actively retained in the cytoplasm during early development. J Cell Biol 147:25–32. https://doi.org/10.1083/jcb.147.1.25

Carlson LL, Page AW, Bestor TH (1992) Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. Genes Dev 6:2536–2541. https://doi.org/10.1101/gad.6.12b.2536

Cedar H, Solage A, Glaser G, Razin A (1979) Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. Nucleic Acids Res 6:2125–2132. https://doi.org/10.1093/nar/6.6.2125

Charlton J, Jung EJ, Mattei AL, Bailly N, Liao J, Martin EJ, Giesselmann P, Brandl B, Stamenova EK, Muller FJ et al (2020) TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. Nat Genet 52:819–827. https://doi.org/10.1038/s41588-020-0639-9

Chen J, Guo L, Zhang L, Wu H, Yang J, Liu H, Wang X, Hu X, Gu T, Zhou Z et al (2013) Vitamin C modulates TET1 function during somatic cell reprogramming. Nat Genet 45:1504–1509. https://doi.org/10.1038/ng.2807

Cheng X, Roberts RJ (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping. Nucleic Acids Res 29:3784–3795. https://doi.org/10.1093/nar/29.18.3784

Chiu YL, Greene WC (2008) The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annu Rev Immunol 26:317–353. https://doi.org/10.1146/annurev.immunol.26.021607.090350

Clark SJ, Harrison J, Paul CL, Frommer M (1994) High sensitivity mapping of methylated cytosines. Nucleic Acids Res 22:2990–2997. https://doi.org/10.1093/nar/22.15.2990

Constantinides PG, Jones PA, Gevers W (1977) Functional striated muscle cells from non-myoblast precursors following 5-azacytidine treatment. Nature 267:364–366. https://doi.org/10.1038/267364a0

Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS (2005) Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol 22:367–377. https://doi.org/10.1093/molbev/msi026

Dai HQ, Wang BA, Yang L, Chen JJ, Zhu GC, Sun ML, Ge H, Wang R, Chapman DL, Tang F et al (2016) TET-mediated DNA demethylation controls gastrulation by regulating Lefty-Nodal signalling. Nature 538:528–532. https://doi.org/10.1038/nature20095

Dean W (2014) DNA methylation and demethylation: a pathway to gametogenesis and development. Mol Reprod Dev 81:113–125. https://doi.org/10.1002/mrd.22280

Dean W, Santos F, Stojkovic M, Zakhartchenko V, Walter J, Wolf E, Reik W (2001) Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. Proc Natl Acad Sci U S A 98:13734–13738. https://doi.org/10.1073/pnas.241522698

Denomme MM, Zhang L, Mann MR (2012) Single oocyte bisulfite mutagenesis. J Vis Exp. https://doi.org/10.3791/4046

Derijck AA, van der Heijden GW, Giele M, Philippens ME, van Bavel CC, de Boer P (2006) gammaH2AX signalling during sperm chromatin remodelling in the mouse zygote. DNA Repair (Amst) 5:959–971. https://doi.org/10.1016/j.dnarep.2006.05.043

Derijck A, van der Heijden G, Giele M, Philippens M, de Boer P (2008) DNA double-strand break repair in parental chromatin of mouse zygotes, the first cell cycle as an origin of de novo mutation. Hum Mol Genet 17:1922–1937. https://doi.org/10.1093/hmg/ddn090

Drohat AC (2021) DNA repair | DNA glycosylases. In: Encyclopedia of biological chemistry III (3rd ed), pp 281–288

Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, Gehrke C (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res 10:2709–2721. https://doi.org/10.1093/nar/10.8.2709

Feinberg AP, Vogelstein B (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 301:89–92. https://doi.org/10.1038/301089a0

Ficz G, Hore TA, Santos F, Lee HJ, Dean W, Arand J, Krueger F, Oxley D, Paul YL, Walter J et al (2013) FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. Cell Stem Cell 13:351–359. https://doi.org/10.1016/j.stem.2013.06.004

Franchini DM, Chan CF, Morgan H, Incorvaia E, Rangam G, Dean W, Santos F, Reik W, Petersen-Mahrt SK (2014) Processive DNA demethylation via DNA deaminase-induced lesion resolution. PLoS One 9:e97754. https://doi.org/10.1371/journal.pone.0097754

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A 89:1827–1831. https://doi.org/10.1073/pnas.89.5.1827

Gama-Sosa MA, Slagel VA, Trewyn RW, Oxenhandler R, Kuo KC, Gehrke CW, Ehrlich M (1983) The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res 11:6883–6894. https://doi.org/10.1093/nar/11.19.6883

Gao Y, Chen J, Li K, Wu T, Huang B, Liu W, Kou X, Zhang Y, Huang H, Jiang Y et al (2013) Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. Cell Stem Cell 12:453–469. https://doi.org/10.1016/j.stem.2013.02.005

Ginno PA, Gaidatzis D, Feldmann A, Hoerner L, Imanci D, Burger L, Zilbermann F, Peters A, Edenhofer F, Smallwood SA et al (2020) A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. Nat Commun 11:2680. https://doi.org/10.1038/s41467-020-16354-x

Gravina S, Ganapathi S, Vijg J (2015) Single-cell, locus-specific bisulfite sequencing (SLBS) for direct detection of epimutations in DNA methylation patterns. Nucleic Acids Res 43:e93. https://doi.org/10.1093/nar/gkv366

Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG et al (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 477:606–610. https://doi.org/10.1038/nature10443

Gu C, Liu S, Wu Q, Zhang L, Guo F (2019) Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes. Cell Res 29:110–123. https://doi.org/10.1038/s41422-018-0125-4

Guenatri M, Bailly D, Maison C, Almouzni G (2004) Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. J Cell Biol 166: 493–505. https://doi.org/10.1083/jcb.200403109

Guibert S, Forne T, Weber M (2012) Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Res 22:633–641. https://doi.org/10.1101/gr.130997.111

Guo F, Li X, Liang D, Li T, Zhu P, Guo H, Wu X, Wen L, Gu TP, Hu B et al (2014) Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. Cell Stem Cell 15:447–459. https://doi.org/10.1016/j.stem.2014.08.003

Hackett JA, Zylicz JJ, Surani MA (2012) Parallel mechanisms of epigenetic reprogramming in the germline. Trends Genet 28:164–174. https://doi.org/10.1016/j.tig.2012.01.005

Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, Surani MA (2013) Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science 339:448–452. https://doi.org/10.1126/science.1229277

Haggerty C, Kretzmer H, Riemenschneider C, Kumar AS, Mattei AL, Bailly N, Gottfreund J, Giesselmann P, Weigert R, Brandl B et al (2021) Dnmt1 has de novo activity targeted to transposable elements. Nat Struct Mol Biol 28:594–603. https://doi.org/10.1038/s41594-021-00603-8

Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani MA (2002) Epigenetic reprogramming in mouse primordial germ cells. Mech Dev 117:15–23. https://doi.org/10.1016/s0925-4773(02)00181-8

Hajkova P, Jeffries SJ, Lee C, Miller N, Jackson SP, Surani MA (2010) Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. Science 329:78–82. https://doi.org/10.1126/science.1187945

Herman JG, Merlo A, Mao L, Lapidus RG, Issa JP, Davidson NE, Sidransky D, Baylin SB (1995) Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. Cancer Res 55:4525–4530

Hirasawa R, Sasaki H (2009) Dynamic transition of Dnmt3b expression in mouse pre- and early post-implantation embryos. Gene Expr Patterns 9:27–30. https://doi.org/10.1016/j.gep.2008.09.002

Ho JJ, Cattoglio C, McSwiggen DT, Tjian R, Fong YW (2017) Regulation of DNA demethylation by the XPC DNA repair complex in somatic and pluripotent stem cells. Genes Dev 31:830–844. https://doi.org/10.1101/gad.295741.116

Holliday R, Pugh JE (1975) DNA modification mechanisms and gene activity during development. Science 187:226–232

Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol 14:R115. https://doi.org/10.1186/gb-2013-14-10-r115

Hotchkiss RD (1948) The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem 175:315–332

Howell CY, Bestor TH, Ding F, Latham KE, Mertineit C, Trasler JM, Chaillet JR (2001) Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell 104:829–838. https://doi.org/10.1016/s0092-8674(01)00280-x

Howlett SK, Reik W (1991) Methylation levels of maternal and paternal genomes during preimplantation development. Development 113:119–127. https://doi.org/10.1242/dev.113.1.119

Hu X, Zhang L, Mao SQ, Li Z, Chen J, Zhang RR, Wu HP, Gao J, Guo F, Liu W et al (2014) Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. Cell Stem Cell 14:512–522. https://doi.org/10.1016/j.stem.2014.01.001

Hyldig SM, Croxall N, Contreras DA, Thomsen PD, Alberio R (2011) Epigenetic reprogramming in the porcine germ line. BMC Dev Biol 11:11. https://doi.org/10.1186/1471-213X-11-11

Ichiyanagi K, Li Y, Watanabe T, Ichiyanagi T, Fukuda K, Kitayama J, Yamamoto Y, Kuramochi-Miyagawa S, Nakano T, Yabuta Y et al (2011) Locus- and domain-dependent control of DNA methylation at mouse B1 retrotransposons during male germ cell development. Genome Res 21:2058–2066. https://doi.org/10.1101/gr.123679.111

Inoue A, Zhang Y (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science 334:194. https://doi.org/10.1126/science.1212483

Inoue A, Shen L, Dai Q, He C, Zhang Y (2011) Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. Cell Res 21:1670–1676. https://doi.org/10.1038/cr.2011.189

Inoue A, Shen L, Matoba S, Zhang Y (2015) Haploinsufficiency, but not defective paternal 5mC oxidation, accounts for the developmental defects of maternal Tet3 knockouts. Cell Rep 10:463–470. https://doi.org/10.1016/j.celrep.2014.12.049

International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. Nature 431:931–945. https://doi.org/10.1038/nature03001

Iqbal K, Jin SG, Pfeifer GP, Szabo PE (2011) Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci U S A 108:3642–3647. https://doi.org/10.1073/pnas.1014033108

Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 466:1129–1133. https://doi.org/10.1038/nature09303

Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 33(Suppl):245–254. https://doi.org/10.1038/ng1089

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821. https://doi.org/10.1126/science.1225829

Kaneda M, Hirasawa R, Chiba H, Okano M, Li E, Sasaki H (2010) Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation. Genes Cells 15:169–179. https://doi.org/10.1111/j.1365-2443.2009.01374.x

Kangaspeska S, Stride B, Metivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, Benes V, Gannon F, Reid G (2008) Transient cyclical methylation of promoter DNA. Nature 452:112–115. https://doi.org/10.1038/nature06640

Kishigami S, Van Thuan N, Hikichi T, Ohta H, Wakayama S, Mizutani E, Wakayama T (2006) Epigenetic abnormalities of the mouse paternal zygotic genome associated with microinsemination of round spermatids. Dev Biol 289:195–205. https://doi.org/10.1016/j.ydbio.2005.10.026

Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, An J, Lamperti ED, Koh KP, Ganetzky R et al (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature 468:839–843. https://doi.org/10.1038/nature09586

Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y et al (2012) Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet 8:e1002440. https://doi.org/10.1371/journal.pgen.1002440

Kohli RM, Zhang Y (2013) TET enzymes, TDG and the dynamics of DNA demethylation. Nature 502:472–479. https://doi.org/10.1038/nature12750

Kothari RM, Shankar V (1976) 5-Methylcytosine content in the vertebrate deoxyribonucleic acids: species specificity. J Mol Evol 7:325–329. https://doi.org/10.1007/BF01743628

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

Kurimoto K, Saitou M (2018) Epigenome regulation during germ cell specification and development from pluripotent stem cells. Curr Opin Genet Dev 52:57–64. https://doi.org/10.1016/j.gde.2018.06.004

Kurimoto K, Yabuta Y, Ohinata Y, Shigeta M, Yamanaka K, Saitou M (2008) Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev 22:1617–1635. https://doi.org/10.1101/gad.1649908

Lane N, Dean W, Erhardt S, Hajkova P, Surani A, Walter J, Reik W (2003) Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis 35:88–93. https://doi.org/10.1002/gene.10168

Lee HJ, Hore TA, Reik W (2014) Reprogramming the methylome: erasing memory and creating diversity. Cell Stem Cell 14:710–719. https://doi.org/10.1016/j.stem.2014.05.008

Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol 13:1192–1200. https://doi.org/10.1016/s0960-9822(03)00432-9

Leitch HG, McEwen KR, Turp A, Encheva V, Carroll T, Grabole N, Mansfield W, Nashun B, Knezovich JG, Smith A et al (2013) Naive pluripotency is associated with global DNA hypomethylation. Nat Struct Mol Biol 20:311–316. https://doi.org/10.1038/nsmb.2510

Li Y, Zhang Z, Chen J, Liu W, Lai W, Liu B, Li X, Liu L, Xu S, Dong Q et al (2018) Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. Nature 564:136–140. https://doi.org/10.1038/s41586-018-0751-5

Lindahl T (1993) Instability and decay of the primary structure of DNA. Nature 362:709–715. https://doi.org/10.1038/362709a0

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:315–322. https://doi.org/10.1038/nature08514

Longerich S, Basu U, Alt F, Storb U (2006) AID in somatic hypermutation and class switch recombination. Curr Opin Immunol 18:164–174. https://doi.org/10.1016/j.coi.2006.01.008

Lucifero D, La Salle S, Bourc'his D, Martel J, Bestor TH, Trasler JM (2007) Coordinate regulation of DNA methyltransferase expression during oogenesis. BMC Dev Biol 7:36. https://doi.org/10.1186/1471-213X-7-36

Maatouk DM, Kellam LD, Mann MR, Lei H, Li E, Bartolomei MS, Resnick JL (2006) DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. Development 133:3411–3418. https://doi.org/10.1242/dev.02500

Mayer W, Niveleau A, Walter J, Fundele R, Haaf T (2000) Demethylation of the zygotic paternal genome. Nature 403:501–502. https://doi.org/10.1038/35000656

McLay DW, Clarke HJ (2003) Remodelling the paternal chromatin at fertilization in mammals. Reproduction 125:625–633. https://doi.org/10.1530/rep.0.1250625

Metivier R, Gallais R, Tiffoche C, Le Peron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G et al (2008) Cyclical DNA methylation of a transcriptionally active promoter. Nature 452:45–50. https://doi.org/10.1038/nature06544

Metzker ML (2010) Sequencing technologies—the next generation. Nat Rev Genet 11:31–46. https://doi.org/10.1038/nrg2626

Meyerson M, Gabriel S, Getz G (2010) Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet 11:685–696. https://doi.org/10.1038/nrg2841

Milagre I, Stubbs TM, King MR, Spindel J, Santos F, Krueger F, Bachman M, Segonds-Pichon A, Balasubramanian S, Andrews SR et al (2017) Gender differences in global but not targeted demethylation in iPSC reprogramming. Cell Rep 18:1079–1089. https://doi.org/10.1016/j.celrep.2017.01.008

Monk M, Boubelik M, Lehnert S (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development 99:371–382. https://doi.org/10.1242/dev.99.3.371

Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M et al (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562. https://doi.org/10.1038/nature01262

Nabel CS, Jia H, Ye Y, Shen L, Goldschmidt HL, Stivers JT, Zhang Y, Kohli RM (2012) AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nat Chem Biol 8:751–758. https://doi.org/10.1038/nchembio.1042

Nakamura T, Arai Y, Umehara H, Masuhara M, Kimura T, Taniguchi H, Sekimoto T, Ikawa M, Yoneda Y, Okabe M et al (2007) PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol 9:64–71. https://doi.org/10.1038/ncb1519

Ng HH, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, Tempst P, Reinberg D, Bird A (1999) MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat Genet 23:58–61. https://doi.org/10.1038/12659

Niemoller C, Wehrle J, Riba J, Claus R, Renz N, Rhein J, Bleul S, Stosch JM, Duyster J, Plass C et al (2021) Bisulfite-free epigenomics and genomics of single cells through methylation-sensitive restriction. Commun Biol 4:153. https://doi.org/10.1038/s42003-021-01661-w

Ohno R, Nakayama M, Naruse C, Okashita N, Takano O, Tachibana M, Asano M, Saitou M, Seki Y (2013) A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. Development 140:2892–2903. https://doi.org/10.1242/dev.093229

Okada Y, Yamagata K, Hong K, Wakayama T, Zhang Y (2010) A role for the elongator complex in zygotic paternal genome demethylation. Nature 463:554–558. https://doi.org/10.1038/nature08732

Okae H, Chiba H, Hiura H, Hamada H, Sato A, Utsunomiya T, Kikuchi H, Yoshida H, Tanaka A, Suyama M et al (2014) Genome-wide analysis of DNA methylation dynamics during early human development. PLoS Genet 10:e1004868. https://doi.org/10.1371/journal.pgen.1004868

Okamoto Y, Yoshida N, Suzuki T, Shimozawa N, Asami M, Matsuda T, Kojima N, Perry AC, Takada T (2016) DNA methylation dynamics in mouse preimplantation embryos revealed by mass spectrometry. Sci Rep 6:19134. https://doi.org/10.1038/srep19134

Ooi SK, Bestor TH (2008) The colorful history of active DNA demethylation. Cell 133:1145–1148. https://doi.org/10.1016/j.cell.2008.06.009

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. Curr Biol 10:475–478. https://doi.org/10.1016/s0960-9822(00)00448-6

Peat JR, Dean W, Clark SJ, Krueger F, Smallwood SA, Ficz G, Kim JK, Marioni JC, Hore TA, Reik W (2014) Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation. Cell Rep 9:1990–2000. https://doi.org/10.1016/j.celrep.2014.11.034

Pfaffeneder T, Spada F, Wagner M, Brandmayr C, Laube SK, Eisen D, Truss M, Steinbacher J, Hackner B, Kotljarova O et al (2014) Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. Nat Chem Biol 10:574–581. https://doi.org/10.1038/nchembio.1532

Polanski Z, Motosugi N, Tsurumi C, Hiiragi T, Hoffmann S (2008) Hypomethylation of paternal DNA in the late mouse zygote is not essential for development. Int J Dev Biol 52:295–298. https://doi.org/10.1387/ijdb.072347zp

Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 463:1101–1105. https://doi.org/10.1038/nature08829

Probst AV, Santos F, Reik W, Almouzni G, Dean W (2007) Structural differences in centromeric heterochromatin are spatially reconciled on fertilisation in the mouse zygote. Chromosoma 116:403–415. https://doi.org/10.1007/s00412-007-0106-8

Ratnam S, Mertineit C, Ding F, Howell CY, Clarke HJ, Bestor TH, Chaillet JR, Trasler JM (2002) Dynamics of Dnmt1 methyltransferase expression and intracellular localization during oogenesis and preimplantation development. Dev Biol 245:304–314. https://doi.org/10.1006/dbio.2002.0628

Razin A, Webb C, Szyf M, Yisraeli J, Rosenthal A, Naveh-Many T, Sciaky-Gallili N, Cedar H (1984) Variations in DNA methylation during mouse cell differentiation in vivo and in vitro. Proc Natl Acad Sci U S A 81:2275–2279. https://doi.org/10.1073/pnas.81.8.2275

Reik W, Dean W, Walter J (2001) Epigenetic reprogramming in mammalian development. Science 293:1089–1093. https://doi.org/10.1126/science.1063443

Reizel Y, Morgan A, Gao L, Schug J, Mukherjee S, Garcia MF, Donahue G, Baur JA, Zaret KS, Kaestner KH (2021) FoxA-dependent demethylation of DNA initiates epigenetic memory of cellular identity. Dev Cell 56(602-612):e604. https://doi.org/10.1016/j.devcel.2021.02.005

Riggs AD, Jones PA (1983) 5-methylcytosine, gene regulation, and cancer. Adv Cancer Res 40:1–30. https://doi.org/10.1016/s0065-230x(08)60678-8

Rougier N, Bourc'his D, Gomes DM, Niveleau A, Plachot M, Paldi A, Viegas-Pequignot E (1998) Chromosome methylation patterns during mammalian preimplantation development. Genes Dev 12:2108–2113. https://doi.org/10.1101/gad.12.14.2108

Saksouk N, Barth TK, Ziegler-Birling C, Olova N, Nowak A, Rey E, Mateos-Langerak J, Urbach S, Reik W, Torres-Padilla ME et al (2014) Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. Mol Cell 56:580–594. https://doi.org/10.1016/j.molcel.2014.10.001

Samiec M, Skrzyszowska M (2018) Intrinsic and extrinsic molecular determinants or modulators for epigenetic remodeling and reprogramming of somatic cell-derived genome in mammalian nuclear-transferred oocytes and resultant embryos. Pol J Vet Sci 21:217–227. https://doi.org/10.24425/119040

Sanford J, Forrester L, Chapman V, Chandley A, Hastie N (1984) Methylation patterns of repetitive DNA sequences in germ cells of Mus musculus. Nucleic Acids Res 12:2823–2836. https://doi.org/10.1093/nar/12.6.2823

Sanford JP, Chapman VM, Rossant J (1985) DNA methylation in extraembryonic lineages of mammals. Trends Genet 1:89–93

Santos F, Hendrich B, Reik W, Dean W (2002) Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev Biol 241:172–182. https://doi.org/10.1006/dbio.2001.0501

Santos F, Peters AH, Otte AP, Reik W, Dean W (2005) Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. Dev Biol 280:225–236. https://doi.org/10.1016/j.ydbio.2005.01.025

Santos F, Peat J, Burgess H, Rada C, Reik W, Dean W (2013) Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. Epigenetics Chromatin 6:39. https://doi.org/10.1186/1756-8935-6-39

Sardina JL, Collombet S, Tian TV, Gomez A, Di Stefano B, Berenguer C, Brumbaugh J, Stadhouders R, Segura-Morales C, Gut M et al (2018) Transcription factors drive Tet2-mediated enhancer demethylation to reprogram cell fate. Cell Stem Cell 23:727–741 e729. https://doi.org/10.1016/j.stem.2018.08.016

Schubeler D (2012) Molecular biology. Epigenetic islands in a genetic ocean. Science 338:756–757. https://doi.org/10.1126/science.1227243

Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W (2012) The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell 48:849–862. https://doi.org/10.1016/j.molcel.2012.11.001

Seki Y, Hayashi K, Itoh K, Mizugaki M, Saitou M, Matsui Y (2005) Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. Dev Biol 278:440–458. https://doi.org/10.1016/j.ydbio.2004.11.025

Seki Y, Yamaji M, Yabuta Y, Sano M, Shigeta M, Matsui Y, Saga Y, Tachibana M, Shinkai Y, Saitou M (2007) Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. Development 134:2627–2638. https://doi.org/10.1242/dev.005611

Shirane K, Toh H, Kobayashi H, Miura F, Chiba H, Ito T, Kono T, Sasaki H (2013) Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. PLoS Genet 9:e1003439. https://doi.org/10.1371/journal.pgen.1003439

Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G (2011) Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet 43:811–814. https://doi.org/10.1038/ng.864

Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature 484:339–344. https://doi.org/10.1038/nature10960

Szabo PE, Pfeifer GP (2012) H3K9me2 attracts PGC7 in the zygote to prevent Tet3-mediated oxidation of 5-methylcytosine. J Mol Cell Biol 4:427–429. https://doi.org/10.1093/jmcb/mjs038

Taghbalout A, Du M, Jillette N, Rosikiewicz W, Rath A, Heinen CD, Li S, Cheng AW (2019) Enhanced CRISPR-based DNA demethylation by Casilio-ME-mediated RNA-guided coupling of methylcytosine oxidation and DNA repair pathways. Nat Commun 10:4296. https://doi.org/10.1038/s41467-019-12339-7

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L et al (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324:930–935. https://doi.org/10.1126/science.1170116

Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126:663–676. https://doi.org/10.1016/j.cell.2006.07.024

Tomizawa S, Kobayashi H, Watanabe T, Andrews S, Hata K, Kelsey G, Sasaki H (2011) Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. Development 138:811–820. https://doi.org/10.1242/dev.061416

Tsukada Y, Akiyama T, Nakayama KI (2015) Maternal TET3 is dispensable for embryonic development but is required for neonatal growth. Sci Rep 5:15876. https://doi.org/10.1038/srep15876

van der Heijden GW, Dieker JW, Derijck AA, Muller S, Berden JH, Braat DD, van der Vlag J, de Boer P (2005) Asymmetry in histone H3 variants and lysine methylation between paternal and maternal chromatin of the early mouse zygote. Mech Dev 122:1008–1022. https://doi.org/10.1016/j.mod.2005.04.009

Vincent JJ, Huang Y, Chen PY, Feng S, Calvopina JH, Nee K, Lee SA, Le T, Yoon AJ, Faull K et al (2013) Stage-specific roles for tet1 and tet2 in DNA demethylation in primordial germ cells. Cell Stem Cell 12:470–478. https://doi.org/10.1016/j.stem.2013.01.016

Wade PA, Gegonne A, Jones PL, Ballestar E, Aubry F, Wolffe AP (1999) Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet 23:62–66. https://doi.org/10.1038/12664

Wang L, Liu Y, Han R, Beier UH, Thomas RM, Wells AD, Hancock WW (2013) Mbd2 promotes foxp3 demethylation and T-regulatory-cell function. Mol Cell Biol 33:4106–4115. https://doi.org/10.1128/MCB.00144-13

Wei LN, Lee CH (1994) Demethylation in the 5′-flanking region of mouse cellular retinoic acid binding protein-I gene is associated with its high level of expression in mouse embryos and facilitates its induction by retinoic acid in P19 embryonal carcinoma cells. Dev Dyn 201:1–10. https://doi.org/10.1002/aja.1002010102

Wossidlo M, Arand J, Sebastiano V, Lepikhov K, Boiani M, Reinhardt R, Scholer H, Walter J (2010) Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. EMBO J 29:1877–1888. https://doi.org/10.1038/emboj.2010.80

Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, Arand J, Nakano T, Reik W, Walter J (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun 2:241. https://doi.org/10.1038/ncomms1240

Wu H, Zhang Y (2014) Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell 156:45–68. https://doi.org/10.1016/j.cell.2013.12.019

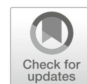
Xu X, Tao Y, Gao X, Zhang L, Li X, Zou W, Ruan K, Wang F, Xu GL, Hu R (2016) A CRISPR-based approach for targeted DNA demethylation. Cell Discov 2:16009. https://doi.org/10.1038/celldisc.2016.9

Xue JH, Xu GF, Gu TP, Chen GD, Han BB, Xu ZM, Bjoras M, Krokan HE, Xu GL, Du YR (2016) Uracil-DNA glycosylase UNG promotes Tet-mediated DNA demethylation. J Biol Chem 291:731–738. https://doi.org/10.1074/jbc.M115.693861

Yamaji M, Ueda J, Hayashi K, Ohta H, Yabuta Y, Kurimoto K, Nakato R, Yamada Y, Shirahige K, Saitou M (2013) PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. Cell Stem Cell 12:368–382. https://doi.org/10.1016/j.stem.2012.12.012

Young LE, Beaujean N (2004) DNA methylation in the preimplantation embryo: the differing stories of the mouse and sheep. Anim Reprod Sci 82-83:61–78. https://doi.org/10.1016/j.anireprosci.2004.05.020

Zheng P, Schramm RD, Latham KE (2005) Developmental regulation and in vitro culture effects on expression of DNA repair and cell cycle checkpoint control genes in rhesus monkey oocytes and embryos. Biol Reprod 72:1359–1369. https://doi.org/10.1095/biolreprod.104.039073



Structure and Function of TET Enzymes 10

Xiaotong Yin, Lulu Hu, and Yanhui Xu

Abstract

Mammalian DNA methylation mainly occurs at the carbon-C5 position of cytosine (5mC). TET enzymes were discovered to successively oxidize 5mC to 5-hydromethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC). Ten-eleven translocation (TET) enzymes and oxidized 5mC derivatives play important roles in various biological and pathological processes, including regulation of DNA demethylation, gene transcription, embryonic development, and oncogenesis. In this chapter, we will discuss the discovery of TET-mediated 5mC oxidation and the structure, function, and regulation of TET enzymes. We start with brief descriptions of the mechanisms of TET-mediated 5mC oxidation and TET-dependent DNA demethylation. We then discuss the TET-mediated epigenetic reprogramming in pluripotency maintenance and embryogenesis, as well as in tumorigenesis and neural system. We further describe the structural basis for substrate recognition and preference in TET-mediated 5mC oxidation. Finally, we summarize the chemical molecules and interacting proteins that regulate TET's activity.

X. Yin
Fudan University Shanghai Cancer Center, Institute of Biomedical Sciences, Shanghai Medical College of Fudan University, Shanghai, China
e-mail: xiaotongyin12@fudan.edu.cn

L. Hu (✉)
Fudan University Shanghai Cancer Center, Institute of Biomedical Sciences, Shanghai Medical College of Fudan University, Shanghai, China

Key Laboratory of Molecular Medicine, Ministry of Education, Department of Systems Biology for Medicine, School of Basic Medical Sciences, Shanghai Medical College of Fudan University, Shanghai, China
e-mail: luluhu@fudan.edu.cn

Y. Xu (✉)
Fudan University Shanghai Cancer Center, Institute of Biomedical Sciences, Shanghai Medical College of Fudan University, Shanghai, China

Key Laboratory of Molecular Medicine, Ministry of Education, Department of Systems Biology for Medicine, School of Basic Medical Sciences, Shanghai Medical College of Fudan University, Shanghai, China

State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China
e-mail: xuyh@fudan.edu.cn

Keywords

TET · Epigenetic modification · DNA demethylation · 5mC · 5hmC · 5fC · 5caC

Abbreviations

2HG	2-Hydroxyglutarate
5caC	5-Carboxylcytosine

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_10

5fC	5-Formylcytosine
5hmC	5-Hydromethylcytosine
5hmrC	5-Hydroxymethylcytidine
5mC	5-Methylcytosine
5mrC	5-Methylcytidine
6mA	N6-methyladenine
α-KG	α-Ketoglutarate
ABH2	AlkB homolog 2
AID	Activation-induced deaminase
AML	Acute myeloid leukemia
APOBEC	Apolipoprotein B mRNA-editing enzyme complex
BER	Base excision repair
CD	Catalytic domain
Chip-seq	Chromatin immunoprecipitation-sequencing
CMML	Chronic myelomonocytic leukemia
CpG	Cytosine-phosphate-guanine
CXXC	Cysteine-X-X-cysteine
Cys-C	Cys-rich C-terminal
Cys-N	Cys-rich N-terminal
Cys-rich	Cysteine rich
DMAD	DNA 6mA demethylase
DNMT	DNA methyltransferase
DSBH	Double-stranded β-helix
E11.5	Embryonic day 11.5
FH	Fumarate hydratase
HCF1	Host cell factor 1
HEK293	Human embryonic kidney 293
hmU	Hydroxymethyluracil
IDH	Isocitrate dehydrogenase
iPSCs	Induced pluripotent stem cells
JBP	J-binding protein
JmjC	Jumonji C
LC-MS	Liquid chromatography-mass spectrometry
MEFs	Mouse embryonic fibroblasts
mESCs	Mouse embryonic stem cells
MET	Mesenchymal-to-epithelial transition
NER	Nucleotide excision repair
NOG	*N*-oxalylglycine
OGT	O-linked β-*N*-acetylglucosamine transferase
OSKM	Oct4, Sox2, Klf4, and c-Myc
PGCs	Primordial germ cells
Pol II	RNA polymerase II
R-2HG	R-2-hydroxyglutarate
SAM	*S*-adenosyl methionine
SDH	Succinate dehydrogenase
SMUG1	Single-strand-selective monofunctional uracil DNA glycosylase 1
T7H	Thymine-7-hydroxylase
TAB-seq	Tet-assisted bisulfite sequencing
TCA	Tricarboxylic acid
TDG	Thymine-DNA glycosylase
TET	Ten-eleven translocation
TSKM	Tet1, Sox2, Klf4, and c-Myc
TSS	Transcription start site

10.1 Introduction

DNA methylation plays important roles in various biological processes through regulating gene expression, genome stability, genomic imprinting, and development (Bird 2002; Jaenisch and Bird 2003; Smith and Meissner 2013; Wu and Zhang 2017). Mammalian DNA methylation mainly occurs at the carbon-5 position of cytosine (5-methylcytosine, also known as 5mC) in CpG dinucleotide context (Bird 2002). Over the past several decades, mammalian DNA demethylation has been thought to be a passive process occurring through DNA replication-dependent dilution of the methylation mark in the absence of the maintenance DNA methyltransferase DNMT1. However, the passive dilution seems to be insufficient to account for the massive and cell division-independent DNA demethylation during specific stages of development. For example, genome-wide DNA demethylation in sperm-derived paternal pronuclei is detected shortly after fertilization before the completion of the first round of cell division (Mayer et al. 2000; Oswald et al. 2000). Proximal epiblast-derived primordial germ cells (PGCs) also undergo global DNA demethylation during their migration (Hajkova et al. 2002; Lee et al. 2002; Yamazaki et al. 2003). Considering that PGCs go through several cell cycles in the presence of DNMT1 before arriving at the genital ridge at E11.5, active DNA demethylation mechanisms are proposed to

be involved in this process (Wu and Zhang 2010). Additionally, DNA demethylation has been reported to take place in the promoter region of different genes in somatic cells upon stimulation by certain signals, indicating the existence of active DNA demethylation (Bruniquel and Schwartz 2003; Kangaspeska et al. 2008; Martinowich et al. 2003; Metivier et al. 2008).

Over the past few decades, a number of studies have reported the identification of different enzymes and pathways involved in mammalian DNA demethylation, including the enzymatic removal of the methyl groups, nucleotide excision repair (NER) pathway, and deamination followed by base excision repair (BER) pathway. However, none of the above findings has been essentially proven. Ten-eleven translocation (TET) proteins have been demonstrated to catalyze the successive oxidation of 5mC, 5-hydroxylmethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (He et al. 2011; Ito et al. 2011; Tahiliani et al. 2009). Further studies have indicated that TET and the oxidized 5mC derivatives play important roles in various biological and pathological processes, including the regulation of active DNA demethylation, gene transcription, embryonic development, and oncogenesis (Branco et al. 2011; Cimmino et al. 2011; Pastor et al. 2013; Rasmussen and Helin 2016; Tan and Shi 2012; Williams et al. 2011a; Wu and Zhang 2011, 2014; Xu and Walsh 2014). Phylogenetic analyses showed that members of the TET/JBP (J-binding protein) family are present in a vast variety of organisms from phages and fungus to mammals (Iyer et al. 2009; Iyer et al. 2013). The prokaryotic TET proteins may generate 5hmC or hydroxymethyluracil (hmU) as an epigenetic mark or to help the prokaryotes escape from the restriction-modification system of the host. In this chapter, we focus on the discovery of TET-mediated 5mC oxidation, summarize the structure, function, and regulation of TET enzymes, and briefly describe the subsequent steps in DNA demethylation.

10.2 Discovery of TET-Mediated 5mC Oxidation

10.2.1 TET-Mediated Iterative Oxidation of 5mC

Ten-eleven translocation 1 (TET1) was first identified as an MLL fusion partner in acute myeloid leukemia (AML) patients (Ono et al. 2002). Mammalian TET proteins have three family members, TET1, TET2, and TET3. TET proteins were considered as candidates in a search for 5mC-modifying enzymes in mammals (Tahiliani et al. 2009), because they show homology to the trypanosome proteins JBP1 and JBP2, which have been proposed to possess hydroxylase activity toward the C5-methyl group of thymine (Cliffe et al. 2009; Yu et al. 2007). Subsequently, TET1 was identified to catalyze the hydroxylation of 5mC to generate 5hmC in a manner dependent on α-ketoglutarate (α-KG) and Fe(II) (Tahiliani et al. 2009) (Fig. 10.1). This activity was observed for all three mouse Tet enzymes (Ito et al. 2010).

Later on, two groups independently demonstrated that TETs mediate iterative oxidation of 5mC to 5hmC, 5fC, and 5caC (He et al. 2011; Ito et al. 2011) (Fig. 10.1), reminiscing the thymine-7-hydroxylase (T7H)-catalyzed stepwise conversion of thymine to isoorotate (Liu et al. 1973; Neidigh et al. 2009; Smiley et al. 2005). The three 5mC oxidation derivatives counteract the DNMT1-mediated maintenance methylation, because CpG sites containing oxidized 5mC are not good substrates for DNMT1 (Hashimoto et al. 2012; Valinluck and Sowers 2007). Following studies have demonstrated that active DNA demethylation is dependent on TET3 in the early stage of embryo development (Gu et al. 2011; Guo et al. 2014a; Shen et al. 2014a). Thus, TET enzymes may facilitate DNA demethylation through passive dilution of the modified bases during replication and/or via active demethylation (Fig. 10.1).

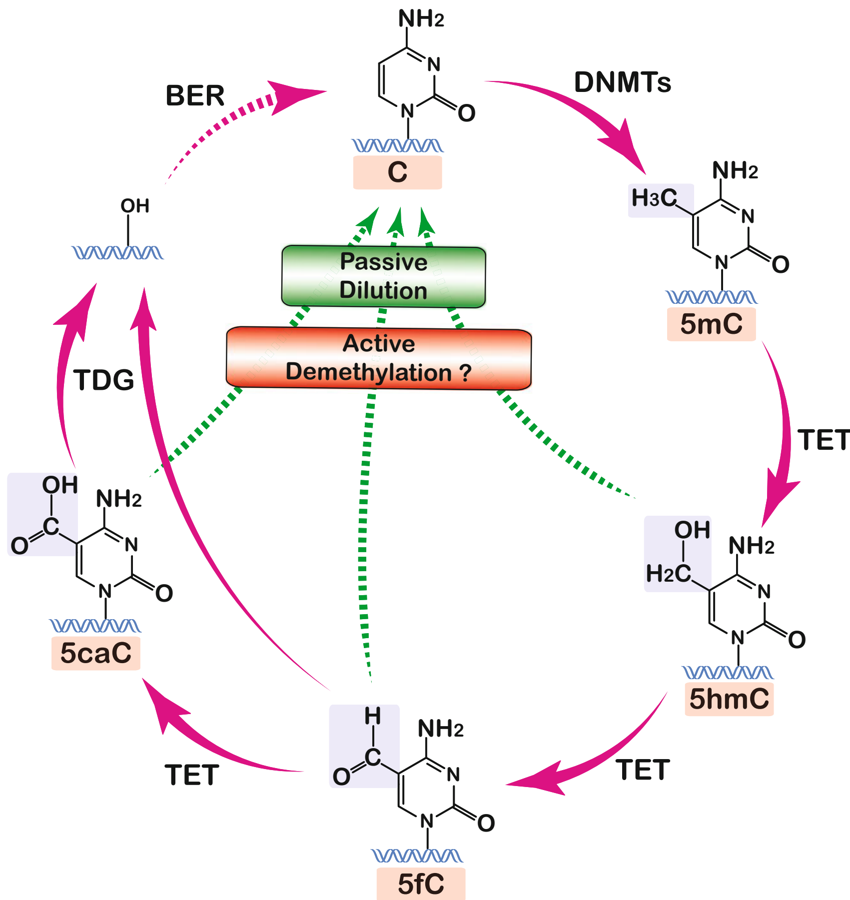


Fig. 10.1 Pathways for dynamic DNA methylation. Mammalian DNA methylation (5mC) is established by de novo DNA methyltransferases DNMT3A/3B, and the patterns of 5mC are maintained by DNMT1. TET enzymes successively oxidize 5mC to 5hmC, 5fC, and 5caC. The three 5mC derivatives may be involved in DNA demethylation through passive dilution and active demethylation. The latter has been observed in early embryo development, but the detailed mechanism remains enigmatic. 5fC and 5caC could be replaced by unmodified cytosine through the TDG-initiated BER pathway, which does not account for massive DNA demethylation in mouse zygotes but contributes to demethylation of specific loci in ESCs and neurons

10.2.2 TET-Dependent DNA Demethylation

An additional step in the active DNA demethylation should exist because 5mC and its oxidation derivatives 5hmC/5fC/5caC do not spontaneously convert to unmodified cytosine under physiological conditions. Interestingly, 5fC and 5caC can be recognized and excised by thymine-DNA glycosylase (TDG) coupled with the BER pathway (He et al. 2011; Maiti and Drohat 2011; Zhang et al. 2012) to complete the demethylation pathway (Fig. 10.1). However, TDG is not required for active demethylation during early-stage embryo development (Guo et al. 2014a). In the thymidine salvage pathway, the conversion of thymine to uracil is achieved by thymine hydroxylation to isoorotate, followed by decarboxylation catalyzed by the isoorotate decarboxylase (Neidigh et al. 2009; Smiley et al. 2005; Xu et al. 2013). Decarboxylation has been proposed to contribute to the conversion of 5caC to an unmodified cytosine. In support of this hypothesis, a 5caC-decarboxylation activity has been

observed in nuclear extracts of mouse embryonic stem cells (mESCs) (Schiesser et al. 2012) as well as of HEK293T cells (Feng et al. 2021a, b). Furthermore, mass spectrometry analysis demonstrated the direct decarboxylation in mammalian genomes by using metabolic labeling with 2′-fluorinated 5caC (Feng et al. 2021a). However, to what extent this decarboxylation activity contributes to the active DNA demethylation, and how this TDG-independent pathway couples with TET-mediated 5mC oxidation, remains largely unknown.

5mC derivatives have been proposed to serve as intermediates in DNA demethylation through alternative pathways. For example, cytidine deaminase AID (activation-induced deaminase)/APOBEC (apolipoprotein B mRNA-editing enzyme complex) family members have been proposed to deaminate 5hmC to generate 5hmU, which is further removed by SMUG1 (single-strand-selective monofunctional uracil DNA glycosylase 1) or TDG and ultimately replaced by cytosine via the BER pathway (Guo et al. 2011). However, controversial evidence for this hypothesis does exist. For example, 5hmU is not detectable in HEK293T cells concomitantly with the overexpression of AID/APOBEC. Purified AID/APOBEC possess decreased or undetectable deamination activity toward 5mC or 5hmC compared with the canonical substrate cytosine, because of steric effects of the 5-substituent groups (Nabel et al. 2012; Rangam et al. 2012). Intriguingly, DNMT3A/3B have been shown to directly convert 5hmC to cytosine in the absence of *S*-adenosyl-L-methionine (AdoMet) under certain reaction conditions (Chen et al. 2012; Liutkeviciute et al. 2009). Moreover, a decarboxylation activity toward 5caC mediated by bacterial and mammalian DNA methyltransferases, such as M.HhaI and DNMTs, was reported (Liutkeviciute et al. 2014). However, whether this reaction occurs in vivo remains elusive (Shen et al. 2014b), because AdoMet acts as a general methyl donor in various biological processes and it is very abundant in cells. One may speculate that it reflects an alternative activity of DNMTs when chromatin is locally exposed to certain extreme environment (van der Wijst et al. 2015).

Although 5mC primarily occurs at CpG dinucleotide sites, several studies discovered a significant amount of 5mC at non-CpG (CpH) sites in almost all kinds of human cells (Laurent et al. 2010; Lister et al. 2009, 2011; Ziller et al. 2011). Particularly, in some cell types, including embryonic stem cells (ESCs) and neuronal cells, as much as 1–6% of non-CpG sites are methylated (Guo et al. 2014b; Lister et al. 2013; Ziller et al. 2011). Experimental studies showed that TET2 could oxidize 5mCpH sites in double-stranded DNA in vitro (DeNizio et al. 2021; Dey et al. 2021). However, whether TETs participate in the demethylation of mCpH in vivo remains unclear.

10.2.3 Mechanisms and Processivity for TET-Mediated Oxidation Reaction

TET enzymes belong to the group of α-KG/Fe(II)-dependent dioxygenases (Iyer et al. 2009; Tahiliani et al. 2009) (Fig. 10.2a). Members of this enzyme family regulate secondary metabolisms in plants and microorganisms, biosynthesis of collagen, hypoxia response, and epigenetic modification in animals, through catalyzing a variety of oxidative reactions, such as hydroxylation, desaturation, epoxidation, epimerization, and oxidative halogenations (Aravind and Koonin 2001; Loenarz and Schofield 2008; McDonough et al. 2010). In mammals, the group includes the AlkB family of DNA/RNA demethylases, JmjC-containing histone lysine demethylases, TET enzymes, and other enzymes that will not be discussed hereafter. DNA repair enzymes in the AlkB family, including human ALKBH2, ALKBH3, and its *Escherichia coli* homolog AlkB, have been reported to oxidize 5mC to 5hmC, 5fC, and 5caC in vitro (Bian et al. 2019). However, whether they have a role in 5mC oxidation in vivo remains elusive.

Oxidation reactions mediated by the α-KG/Fe(II)-dependent dioxygenases can be divided into

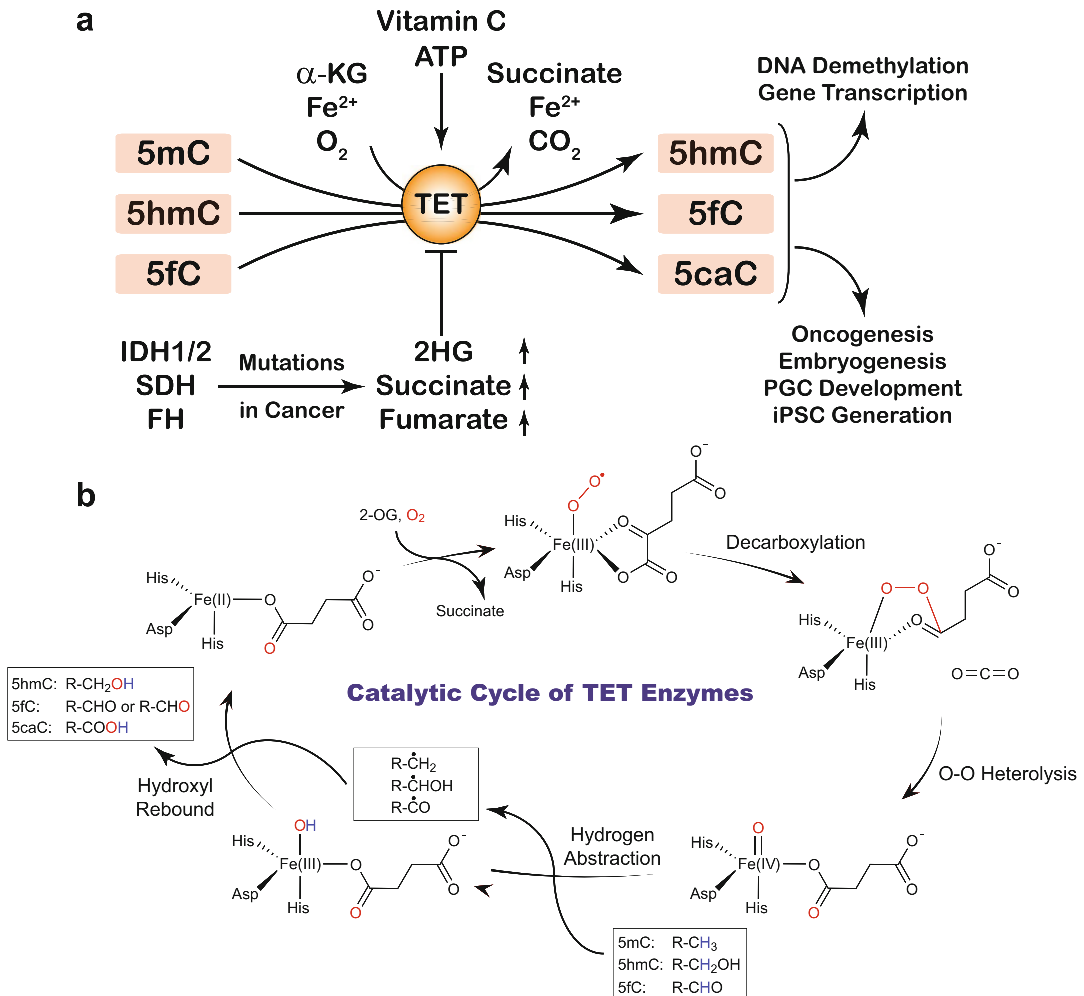


Fig. 10.2 TET-mediated oxidation of 5mC. (**a**) TET enzymes successively oxidize 5mC to 5hmC, 5fC, and 5caC. The oxidation depends on the presence of Fe(II), molecular oxygen, and α-KG as co-substrates. Vitamin C and ATP enhance the enzymatic activity of TET. TET enzymes and 5mC oxidation derivatives are involved in various molecular events (DNA demethylation and regulation of gene expression) and therefore play key roles in embryonic development and oncogenesis. Patient-derived mutations of enzymes in the tricarboxylic acid (TCA) cycle, including IDH1/2, SDH, and FH, lead to the accumulation of metabolites 2-HG, succinate, and fumarate. These α-KG analogs competitively inhibit the activities of TET enzymes and JmjC-containing histone demethylases and therefore may contribute to oncogenesis. (**b**) Schematic model for the oxidative reactions catalyzed by TET enzymes

two successive steps: dioxygen activation and substrate oxidation (Fig. 10.2b) (Shen et al. 2014b). In the first step, Fe(II) and α-KG coordinate to the "facial triad" composed of the conserved HxD/E. . .H (where x can be any residue) motif of the enzyme. Then, molecular dioxygen replaces the water molecule and binds to the Fe(II) in the catalytic center of the enzyme (Muller et al. 2004). One oxygen atom of the bound dioxygen inserts into the succinate derived from α-KG decarboxylation, and the other one couples with the iron to generate a high-valent Fe(IV)-oxo intermediate (Krebs et al. 2007; Seisenberger et al. 2012; Valegard et al. 2004).

The dioxygen activation procedure has been demonstrated to be accelerated in the presence of substrate within the catalytic cavity (Ryle et al. 1999), which induces environmental changes in Fe(II) coordination and then facilitates dioxygen binding (Muller et al. 2004). In the second step, the C-H bond of the substrate is cleaved by the Fe(IV)-oxo oxidizing radical, and the oxygen atom is transferred to the target carbon group through hydrogen abstraction (Hoffart et al. 2006; Price et al. 2003). Upon substrate oxidation, the iron returns to the Fe(II) state and thus completes one reaction cycle. TET enzymes oxidize 5mC through the same mechanism (Shen et al. 2014b) (Fig. 10.2b). In the first cycle of oxidation, 5mC is converted to 5hmC. In the next step, 5hmC is further oxidized to a germinal diol, which decomposes into 5fC, and then 5caC is generated in a third cycle. One α-KG is consumed for each cycle of the reaction (Fig. 10.2b).

As mentioned above, TET-generated oxidized methylcytosine derivatives have been shown to be crucial epigenetic markers. Characterizing TET's function in generating distinct 5mC derivatives helps to better understand the demethylation process and the establishment and maintenance of epigenetic landscape. Using isotope tracing, Crawford et al. reported that mouse Tet2 catalytic domain could generate 5fC/5caC de novo in an iterative manner, which means that 5mC is oxidized continuously in a single substrate–enzyme association event producing 5fC/5caC without the detachment of 5hmC intermediate (Crawford et al. 2016), supporting that TET could perform a chemical processive catalysis. On the contrary, Xu et al. reported the dominantly symmetric production of 5fC containing 5fC in both DNA strands during in vitro 5mC oxidation mediated by mouse Tet1 catalytic domain (Xu et al. 2014). By employing a chemical probe which could not only specifically target 5fC, but also discriminate symmetric 5fC sites from asymmetric ones, they observed the high abundance of symmetric 5fC sites in the in vitro oxidation reaction, indicating a non-processive/distributive function of TET. Consistently, Tamanaha et al. also detected a distributive mechanism of two mouse Tet catalytic domains and the full-length Tet1 from *Naegleria gruberi*, which exhibited favorable release of the intermediates 5hmC and 5fC bases at the end of each turnover (Tamanaha et al. 2016).

In vitro assays suggested that TET would dissociate from the DNA strand rather than slide along DNA in searching for the next targets (Tamanaha et al. 2016), while another study reported mobility of TET including sliding and interstrand transfer by a high-speed atomic-force microscope (AFM) (Xing et al. 2020), supporting the site-to-site processivity of TET-mediated oxidation reaction. These controversial outcomes might indicate that TET's processivity could be regulated by a broad range of factors composing the local chromatin environment, including the substrate strand configurations and flexibility, 5-oxidated-methyl cytosine site symmetricity and complementary strand composition (DeNizio et al. 2019; Xing et al. 2020), flanking sequences, the enzyme concentration, the interacting proteins, and whether full-length or truncated proteins are used.

10.2.4 Oxidation of 5mrC-RNA and 6mA-DNA

In addition to their activity toward 5mC in DNA, enzymes of the TET family have also been reported to possess oxidation activity on other substrates, including 5-methylcytidine (5mrC) in RNA and N^6-methyladenine (6mA) in DNA. Fu et al. found that TET enzymes convert 5mrC to 5-hydroxymethylcytidine (5hmrC) in RNA in vitro and induce the formation of 5hmrC in human cells (Fu et al. 2014; Xu et al. 2016). The presence of 5hmrC in RNA has been verified in mammalian cells using sensitive and accurate LC-MS/MS/MS approaches (Fu et al. 2014). By performing sequencing and data analyses, Lan et al. suggested that the RNA cytosine 5-hydroxymethylation converted by Tet has a crucial role as a marker of transcriptome flexibility which is correlated with the equilibrium of pluripotency maintenance and cell differentiation (Lan et al. 2020). Similar to oxidation on DNA, 5hmrC-RNA could be further oxidized,

generating 5fC and 5caC in mRNA in vivo (Huang et al. 2016; Zhang et al. 2016a). Catalysis of 5fC to 5caC conversion in RNA by Tet1 was also verified in vitro (Basanta-Sanchez et al. 2017). Tet-mediated oxidation of 5mrC in RNA may function through affecting the RNA interactome, rendering RNA chemically instable and thus influencing RNA metabolism and the transcriptome. A study reported TET2 promoting translation by mediating 5mC oxidation in tRNA, indicating the multilevel effects that TET could have on the central dogma (Shen et al. 2021a).

DNA 6mA is commonly found in bacterial genomes. Several studies have demonstrated the existence of this modified nucleotide in genomes of various eukaryotes, including *Chlamydomonas* (Fu et al. 2015), *C. elegans* (Greer et al. 2015), and *Drosophila* (Zhang et al. 2015a). While searching for a specific enzyme responsible for demethylation of 6mA, Zhang et al. found a *Drosophila* gene (CG2083) that possessed such activity and named the corresponding protein DMAD (DNA 6mA demethylase), which is a homolog of mammalian TET (Zhang et al. 2015a). Nuclear extracts from late-stage embryos showed considerable 6mA demethylation activity, whereas depletion of DMAD from the nuclear extracts using an anti-DMAD antibody or siRNA led to the loss of 6mA demethylase activity. Further study has suggested that DMAD removes 6mA primarily from transposon regions and is essential for development (Zhang et al. 2015a). In an in vitro assay, the catalytic domain (CD) of DMAD showed 5mC oxidation activity, albeit approximately 30-fold lower than that of Tet1-CD, suggesting dual substrate specificity. Intriguingly, structure-based sequence analysis has indicated that all residues critical for 5mC recognition are conserved in human TET2 and *Drosophila* DMAD (Hu et al. 2013). Because 6mA is considerably larger than 5mC, it would be of interest to investigate how 6mA is bound and specifically recognized by DMAD. Further studies delineate that TET proteins have considerable tolerance of alterations on substrates beyond 5-methylcytosine and DNA 6mA, such as novel synthetic N4-methyl substituents (Ghanty et al. 2020), 5-alkylcytosines (Ghanty et al. 2018; Kavoosi et al. 2019), and thymine (Pfaffeneder et al. 2014), suggesting a catalytic plasticity of TET oxygenases and possibly unforeseen roles of TET in various biological processes.

10.3 Function of TET Enzymes

10.3.1 Distribution of TET Enzymes and 5mC Oxidation Derivatives

The genomic levels of 5hmC are relatively high in neurons (15–40% of 5mC), and self-renewing and pluripotent stem cells, but are greatly reduced in cancer cells and along with differentiation (Globisch et al. 2010; Ito et al. 2011; Mellen et al. 2012; Pfaffeneder et al. 2011; Ruzov et al. 2011; Schutsky et al. 2018; Song et al. 2011; Szwagierczak et al. 2010; Tahiliani et al. 2009), suggesting a positive correlation between 5hmC and pluripotent states. Based on this finding 5hmC might serve as not only a DNA demethylation intermediate but also a relatively stable epigenetic mark. The 5mC derivatives may reduce DNA-binding affinity of methyl-CpG-binding proteins (Valinluck et al. 2004) or they could be recognized by other chromatin-associated factors for transcriptional regulation (Frauer et al. 2011; Yildirim et al. 2011; Zhou et al. 2014). 5hmC inhibits the activity of DNMT1, thus resulting in replication-dependent erasure of cytosine methylation (Hashimoto et al. 2012). Intriguingly, it has been reported that during zygotic reprogramming the initial loss of paternal 5mC and 5hmC formation is temporally separated, while 5hmC accumulation at late zygotic stages is dependent on de novo DNA methylation driven by zygotic Dnmt3a and Dnmt1 (Amouroux et al. 2016). The state-of-the-art new 5hmC sequencing technology suitable for a small amount of genomic DNA samples (genomic DNA from less than 1000 cells or even single cell) may provide further information for understanding the function of 5hmC.

Consistent with the enrichment of 5hmC in mESCs, relatively high expression levels of Tet1 and Tet2 (to a lesser extent Tet3) are also detected in mESCs and mouse iPSCs, whereas the protein

levels diminish during mESCs differentiation (Ficz et al. 2011; Ito et al. 2010; Mulholland et al. 2020; Szwagierczak et al. 2010; Wossidlo et al. 2011). As the stem cells begin to exit from pluripotency, DNMT3A/3B, with the opposing enzymatic activity to TETs, are re-expressed (Lee et al. 2014). Studies showed that co-existing TETs and DNMT3A/3B act competitively at somatic enhancers in human ESCs (Charlton et al. 2020), and similar competition between TET2 and DNMT3A was also found in hematopoietic stem cells (Zhang et al. 2016b). TET enzymes are expressed at different levels in adult human and mouse tissues (Ito et al. 2010). Tet1 is largely expressed in PGCs (Yamaguchi et al. 2012). Tet3 is highly expressed in oocytes and zygotes, whereas Tet1 or Tet2 shows undetectable or moderate expression (Wossidlo et al. 2011), indicating a critical role for Tet3 in epigenetic reprogramming during embryo development. The expression levels of TET enzymes are decreased in various cancers, which is also consistent with the low level of 5hmC in cancer cells (Haffner et al. 2011; Lian et al. 2012; Yang et al. 2013).

In accordance with other chromatin-modifying enzymes, TET proteins localize to the nucleus in cultured mammalian cells (Ito et al. 2010; Tahiliani et al. 2009). However, Tet3 is present in the male pronucleus in the zygotic stage, and translocates to the cytoplasm in the preimplantation stages (Gu et al. 2011). Besides, Aid may regulate the subcellular localization of Tet proteins. When co-transfected with Aid, the subcellular localization of Tet1/2/3 is altered from the nucleus to the cytoplasm, and this translocation is associated with Aid shuttling (Arioka et al. 2012).

A base resolution map of 5hmC in human and mouse ESCs determined by Tet-assisted bisulfite sequencing (TAB-seq) has revealed its widespread distribution in the genome and varied abundance across functional elements (Yu et al. 2012). Almost half (46.4%) of the 5hmCs reside in distal regulatory elements, in which p300-binding sites, enhancers, CTCF-binding sites, and DNaseI hypersensitive sites harbor more 5hmC than other genic regions. High levels of 5hmC concomitant with low levels of 5mC have been observed near the transcription factor binding sites but not within them. 5mC but not 5hmC was enriched in repetitive elements. Notably, TAB-seq analysis has indicated that 5hmC is abundant in regions with low CpG content (Yu et al. 2012), whereas antibody-based sequencing has indicated that 5hmC is enriched in CpG-rich transcription start sites (TSSs) (Williams et al. 2011b). This discrepancy may result from the different technological approaches used. A selective chemical labeling and enrichment method revealed that an age-dependent acquisition of 5hmC modification in specific gene bodies is linked to neurodegenerative disorders (Song et al. 2011).

Genome-wide analysis of 5fC localization in mESCs has revealed that 5fC is preferentially enriched at poised enhancers, suggesting its roles in epigenetic priming (Song et al. 2013). Base-resolution 5fC maps depicted by cyclization-enabled C-to-T transition of 5fC (fC-CET) showed limited overlap with 5hmC, with 5fC-marked regions being transcriptionally more active than 5hmC-marked ones (Xia et al. 2015). An additional study using DNA immunoprecipitation-coupled chemical modification-assisted bisulfite sequencing has indicated that 5fC and 5caC can be detected in enhancers, promoters, and intragenic regions but share limited overlap in mESCs genome (Lu et al. 2015). In addition, the 5mC oxidation activity negatively correlates with 5mC abundance and positively correlates with enhancer activity, suggesting its roles in regulating gene expression (Lu et al. 2015) by affecting the process of transcriptional elongation. In support of this hypothesis, *Tdg*-KO mESCs display an apparent retardation of RNA polymerase II (Pol II) elongation compared with the wild-type cells, possibly because of the increased level of 5caC. The crystal structure of RNA Pol II in complex with 5caC-DNA showed that an extra hydrogen bond formed between Pol II and 5caC may contribute to this retardation (Wang et al. 2015a). These studies suggest a functional impact of 5caC on transcriptional elongation and gene expression.

10.3.2 TET in ESCs and Cell Differentiation

Individual knockdown of *Tet1* or *Tet2* leads to a partial reduction of genomic 5hmC level, suggesting that both TET1 and TET2 are responsible for 5hmC maintenance in human and mouse ESCs (Dawlaty et al. 2013; Koh et al. 2011; Tahiliani et al. 2009). ChIP-seq analyses have revealed that most of the Tet1 binding sites co-localize with 5hmC in euchromatin regions in mESCs and strongly accumulate at hypomethylated CpG-rich promoters (Williams et al. 2011b; Wu et al. 2011). *Tet1* knockdown or knockout in mESCs results in increased DNA methylation levels, suggesting that Tet1 is required to maintain the hypomethylation state of many gene promoters (Wu et al. 2011). Tet1 regulates the expression of genes related to the maintenance of mESCs through promoting the transcription of pluripotency-associated factors and suppressing the expression of differentiation-associated factors (Dawlaty et al. 2011; Ficz et al. 2011; Williams et al. 2011b; Wu et al. 2011).

Although Tet1 plays a role in gene regulation, it is largely dispensable in the maintenance of pluripotency; however, it may regulate the cell lineage differentiation. *Tet1* knockdown and knockout mESCs present an unvaried propagation rate and morphology (Dawlaty et al. 2011; Ficz et al. 2011; Koh et al. 2011). The full pluripotency of *Tet1*$^{-/-}$ mESCs has also been verified by the tetraploid complementation assay (Dawlaty et al. 2011). *Tet1* knockdown or knockout mESCs generate teratomas that contain differentiated cells from all three germ layers but show skewed differentiation toward the trophectoderm (Dawlaty et al. 2011; Koh et al. 2011). Consistently, *Tet1* knockdown mESCs show increased mRNA levels of trophectoderm markers *Cdx2*, *Eomes*, and *Hand1* but decreased mRNA levels of *Pax6* and *Neurod1*, which are representatives of neuroectoderm (Koh et al. 2011). Similarly, TET1 deficiency also impairs differentiation of human ESCs to neuroectoderm (Li et al. 2020a). The *Tet1* and *Tet2* double-knockout mESCs retain pluripotency in the teratoma formation assay but generate abnormal chimeric embryos (Dawlaty et al. 2013). Although there exist possible defects in embryo development as indicated by the significant skewing, *Tet1*$^{-/-}$ mice are viable and fertile and grow normally, except for the smaller size of some of the pups (Dawlaty et al. 2011), and viable *Tet1* and *Tet2* double-knockout mice can also be generated, albeit with a low birthrate (Dawlaty et al. 2013). Besides its functions in ESCs differentiation, TET1 also plays an important role in adult stem cell differentiation. TET1 knockdown inhibits the odontogenic differentiation potential of human dental pulp cells (Rao et al. 2016). The overexpression of TET1 promotes myogenic differentiation and reduces the methylation of the myogenic differentiation-associated genes (Gao et al. 2020).

A series of studies have demonstrated how dynamic DNA methylation mediated by TETs and DNMT3A/3B might contribute to the cell lineage specification, particularly in the stage of cells exiting pluripotency and priming to differentiate. Compared to the abovementioned viable *Tet1* single- or *Tet1/2* double-knockout mice embryos, *Tet* triple-knockout mice embryos exhibit severe gastrulation defects with dysregulated Lefty-Nodal signaling pathway (Dai et al. 2016). Disruption of the *Dnmt3a* and *Dnmt3b* genes in *Tet*-null embryos restores the Lefty-Nodal signaling pathway and normal morphogenesis, suggesting that TET-mediated DNA demethylation modulates Lefty-Nodal signaling pathway through counteracting the de novo DNA methylation by DNMT3A/3B (Dai et al. 2016). This study implies that the dynamic DNA methylation regulated by TETs and DNMT3A/3B plays crucial roles in stem cells exiting pluripotency toward differentiation and cell developmental progression. Using metabolic labeling with stable isotopes and mass spectrometry, Spada et al. reported high rates of 5mC/5hmC/5fC turnover in cultured mouse pluripotent stem cells kept between naive and primed pluripotency (Spada et al. 2020). Acute knockout of *DNMT3a/3b* or/and *TETs* in human pluripotent stem cells or mouse ESCs combined with DNA

methylation analysis showed that DNA undergoes cyclic transitions between methylation states, and this methylation turnover is enriched at somatic enhancers (Charlton et al. 2020; Ginno et al. 2020). Allelic DNA methylation levels of two pluripotency super-enhancers, *Sox2* and *Mir290*, in ESCs, were tracked with allele-specific reporters and found dynamically switching (Song et al. 2019). Consistent with this active turnover of DNA methylation in ESCs, the methylation states were heterogeneously distributed from cell to cell (Bell et al. 2020; Rulands et al. 2018; Song et al. 2019). Analysis at genome scale by single-cell sequencing and biophysical modeling revealed genome-wide methylation oscillations with a periodicity of 2–3 h (Rulands et al. 2018). The DNA methylation oscillations occurred when both TETs and DNMT3A/3B were expressed and the oscillation amplitude was the greatest at a CpG density characteristic of enhancers (Rulands et al. 2018). By affecting the methylation states of the enhancers, the active turnover of DNA methylation may facilitate the cell lineage specification, through presenting a diversified epigenome for generating various cell lineages when a certain external stimulus emerges (Parry et al. 2021).

10.3.3 TETs Mediate Epigenetic Reprogramming in Early Embryogenesis and PGC Development

Extensive studies in mouse embryogenesis demonstrated two waves of epigenetic reprogramming occurring in pre-implantation and PGC development. Active demethylation of the paternal pronucleus occurs rapidly after fertilization (Mayer et al. 2000; Oswald et al. 2000). Several studies have provided evidence that TET-mediated 5mC oxidation participates in this demethylation process (Gu et al. 2011; Iqbal et al. 2011; Wossidlo et al. 2011). Immunofluorescence staining has indicated that 5mC of the maternal pronucleus remains constant through different pronucleus stages, whereas 5mC in paternal pronuclei is progressively lost and 5hmC shows up. Coincident with the appearance of 5hmC, Tet3 is enriched in oocytes and paternal pronuclei of zygotes (Gu et al. 2011; Iqbal et al. 2011; Wossidlo et al. 2011), suggesting its key role in DNA demethylation. Tet3-deficient zygotes show impaired conversion of 5mC to 5hmC in paternal genome and retardation of demethylation of paternal *Oct4* and *Nanog* genes, indicating that Tet3 plays an important role in epigenetic reprogramming (Gu et al. 2011). Tet3 also generates 5fC and 5caC in paternal pronuclei (Inoue et al. 2011), which may also contribute to DNA demethylation (Fig. 10.1).

It is generally believed that 5mC in maternal pronuclei does not undergo active demethylation (Gu et al. 2011; Inoue et al. 2011; Inoue and Zhang 2011; Iqbal et al. 2011; Xie et al. 2012). The 5mCs were considered to be protected by PGC7 (also known as Dppa3 or Stella) against oxidation by Tet (Iqbal et al. 2011; Nakamura et al. 2007; Wossidlo et al. 2011). However, advanced single-base resolution sequencing analyses of mouse oocytes and early embryos have indicated that both maternal and paternal pronuclei undergo active demethylation and replication-dependent dilution of methylation (Guo et al. 2014a; Shen et al. 2014a; Wang et al. 2014). Wang et al. have demonstrated the existence of 5hmC and 5fC in both male and female pronuclei and showed that a significant proportion of maternal genomes are subjected to active demethylation (Wang et al. 2014). Studies from another group have indicated that replication-dependent dilution contributes ~87% and ~75% to genome-wide demethylation in maternal and paternal pronuclei, respectively (Guo et al. 2014a). Nevertheless, active DNA demethylation does occur in a manner dependent on TET3, although the mechanism remains elusive (Fig. 10.1).

During the development of PGCs, massive genome-wide demethylation occurs which is required for the establishment of totipotency of the germ cells (Saitou et al. 2012). *Tet1* (but not *Tet2* or *Tet3*) is substantially expressed in mouse PGCs and is enriched between E10.5 and E11.5, suggesting its key role in regulating methylation states of imprinting genes during PGCs

development (Hackett et al. 2013; Yamaguchi et al. 2012, 2013). *Tet1* paternal knockout mice have been obtained by mating *Tet1*$^{-/-}$ male and wild-type female mice, but the mice exhibited partial fetal or postnatal defects as well as early embryo lethality (Yamaguchi et al. 2013). Intriguingly, a follow-up study indicated that during epigenetic reprogramming in gonadal PGCs Tet1 tends to protect the demethylated genome from new round of de novo methylation rather than directly demethylating the whole genome (Hill et al. 2018), which is reminiscent of the Tet3-driven hydroxylation in guarding against de novo DNA methylation during zygotic reprogramming (Amouroux et al. 2016).

Interestingly, 5mC and 5hmC diminish at different rates during PGCs development (Hackett et al. 2013), suggesting the occurrence of two waves of genomic demethylation. During the first wave, 5mC drops to approximately 30% until E9.5 during PGC migration (Seisenberger et al. 2012) owing to passive dilution (Kurimoto et al. 2008). During the second wave, active demethylation occurs in a manner dependent on TET1 and possibly TET2, leading to a rapid decrease in 5mC and a slight increase in 5hmC, followed by a decline resembling a dilution pattern (Hackett et al. 2013; Yamaguchi et al. 2013).

10.3.4 TET Enzymes in Somatic Cell Reprogramming

In induced pluripotent stem cells (iPSCs) induced from mouse embryonic fibroblasts (MEFs) by OSKM factors (Oct4, Sox2, Klf4, and c-Myc), Tet1 has a high expression level, whereas Tet2 expression is relatively moderate and Tet3 is undetectable (Gao et al. 2013). MEFs lacking all three Tet enzymes no longer produce reprogrammed colonies; this result is in contrast to MEFs deficient in only one or two of the Tet enzymes, indicating that Tet enzymes are redundant but essential for iPSCs generation (Hu et al. 2014). Furthermore, Tet1 overexpression can facilitate OSKM (Oct4, Sox2, Klf4, and c-Myc)-induced iPSC formation (Gao et al. 2013). The underlying mechanisms for Tet functions in iPSC generation have been studied from various aspects, as indicated below. Of note, reintroduction of biochemically engineered TET enzymes into Tet2-deficient mouse embryonic fibroblast revealed that TET-mediated oxidation to 5fC/5caC, but not 5hmC, promotes iPSC reprogramming efficiency. These data showed that 5fC/5caC is the major driver of DNA demethylation during iPSC reprogramming and DNA demethylation through 5fC/5caC has roles distinct from 5hmC in somatic reprogramming to pluripotency (Caldwell et al. 2021).

Nanog and *Esrrb*-included pluripotency loci require reactivation during somatic cell reprogramming. Tet2 is recruited to these pluripotency loci and may contribute to transcriptional induction in the early stage of reprogramming (Doege et al. 2012). Another study has shown the physical association and synergetic effect of Nanog and TET1 or TET2 in the enhancement of iPSCs generation. Nanog deficiency leads to a reduced recruitment of TET1 toward a subset of genomic loci shared by Nanog and TET1, suggesting that Nanog recruits TET1 to target genes for establishment of pluripotency and cell lineage specification (Costa et al. 2013). Intriguingly, somatic cell reprogramming can also be induced by TSKM factors (Tet1, Sox2, Klf4, and c-Myc) (Gao et al. 2013). In this system, Oct4 in OSKM factors (Oct4, Sox2, Klf4, and c-Myc) is replaced by Tet1, which facilitates iPSC generation by promoting the demethylation and reactivation of Oct4.

Another study has shown that the depletion of all three Tet enzymes prevents iPSC generation, because the mesenchymal-to-epithelial transition (MET) step is blocked. This barrier could be ascribed to the inactivation of a cluster of miRNAs (Hu et al. 2014). These miRNAs belong to the miR-200 family and are known to modulate the expression of transcription factors that inhibit the expression of epithelial markers. Tet and TDG are required for the demethylation and reactivation of miR-200, for the restoration of MET, and for the initiation of reprogramming process. Reintroducing only this cluster of miRNAs overcomes the barrier and completes the reprogramming, suggesting the nonessential

function of Tet enzymes in the subsequent processes.

10.3.5 TET Enzymes and Cancer

The *TET2* gene is more frequently mutated in hematopoietic malignancies than *TET1* or *TET3*. *TET2* mutations are frequently observed in chronic myelomonocytic leukemia (CMML), AML, and other myeloid malignancy patients (Abdel-Wahab et al. 2009; Langemeijer et al. 2009; Tefferi et al. 2009). Some of the patient-derived mutations, such as those involved in iron chelation (H1881, H1382, and D1384), α-KG interaction (R1896, R1261, and S1898), and DNA recognition (N1387, H1904, and Y1902), largely decrease or abolish TET2 activity (Hu et al. 2013; Ko et al. 2010). The prevalence of two catalytic inactive mutants (H1802 and R1817, corresponding to H1881 and R1896 in human TET2) strongly correlates with low genomic 5hmC levels in the bone marrow and blood of affected patients. Further evidence has shown that TET2 is critical for normal myelopoiesis (Ko et al. 2010). *Tet2* deficiency results in enhanced self-renewal and the abnormal proliferation of hematopoietic stem cells, resulting in splenomegaly, monocytosis, and extramedullary hematopoiesis in an animal model (Moran-Crusio et al. 2011). *Tet2*$^{+/-}$ mice are also predisposed to myeloid transformation (Moran-Crusio et al. 2011). *TET2* mutations have been identified in human lymphomas, and these loss-of-function mutants may perturb the early developmental state of hematopoietic stem cells, leading to myeloid and/or lymphoid malignancies (Quivoron et al. 2011). The host immune response plays a critical role in attenuating cancer progression. TET2 was reported to promote anti-tumor immunity by mediating IL-6/G-MDSCs/CD8+ T-cell immune response cascade and inhibiting PD-L1 gene expression in breast cancer (Li et al. 2020b; Shen et al. 2021b; Zhang et al. 2015b).

Melanoma is a highly malignant and aggressive type of cancer and displays global DNA hypomethylation and gene-specific hypermethylation at certain tumor suppressors (Hoon et al. 2004; Liu et al. 2008; Shen et al. 2007). 5hmC is considerably decreased in melanoma, compared with melanocytes and nevus cells, suggesting that the modification may be a unique feature of melanoma (Lian et al. 2012; Rodic et al. 2015). Moreover, all three TET proteins are downregulated in melanoma, which is consistent with the decreased 5hmC level (Lian et al. 2012). Low levels of 5hmC and downregulation of TET enzymes have also been found in other human cancer cells, including breast, liver, lung, pancreatic, colon, and prostate cancers (Haffner et al. 2011; Yang et al. 2013). As a stable epigenetic marker, the 5hmC modification in circulating cfDNA has been shown to function as biomarker for cancer diagnosis, including colorectal, gastric (Guler et al. 2020), and pancreatic cancers (Guler et al. 2020) and hepatocellular carcinoma (Cai et al. 2019).

10.3.6 TET Enzymes in Neural System

The presence of 5hmC in mammalian genomes was first discovered in Purkinje neurons and the brain (Kriaucionis and Heintz 2009). Later, relatively high levels of 5hmC have been observed in various adult brain regions (Munzel et al. 2010; Ruzov et al. 2011; Szwagierczak et al. 2010), and 5hmC was shown to mark postmitotic neural cells in the adult and developing vertebrate central nervous system (Diotel et al. 2017), suggesting important roles for TET enzymes and 5mC oxidation derivatives in neural systems. All TET proteins exhibit strong co-localization with the neuronal marker NeuN throughout the hippocampus, implicating a primary expression and distribution in neurons (Kaas et al. 2013; Li et al. 2014; Mi et al. 2015). Several lines of evidence suggest that TET1 is responsible for the basal level of neuronal 5hmC, and is related to crucial neuronal regulatory genes (Antunes et al. 2019). Overexpression of TET1 leads to increased conversion of 5mC to 5hmC in the central nervous system and TET1 is essential for the demethylation of fibroblast growth factor 1 (*Fgf1*) and brain-derived neurotrophic factor (*Bdnf*) promoters (Guo et al. 2011). *Tet1* knockout

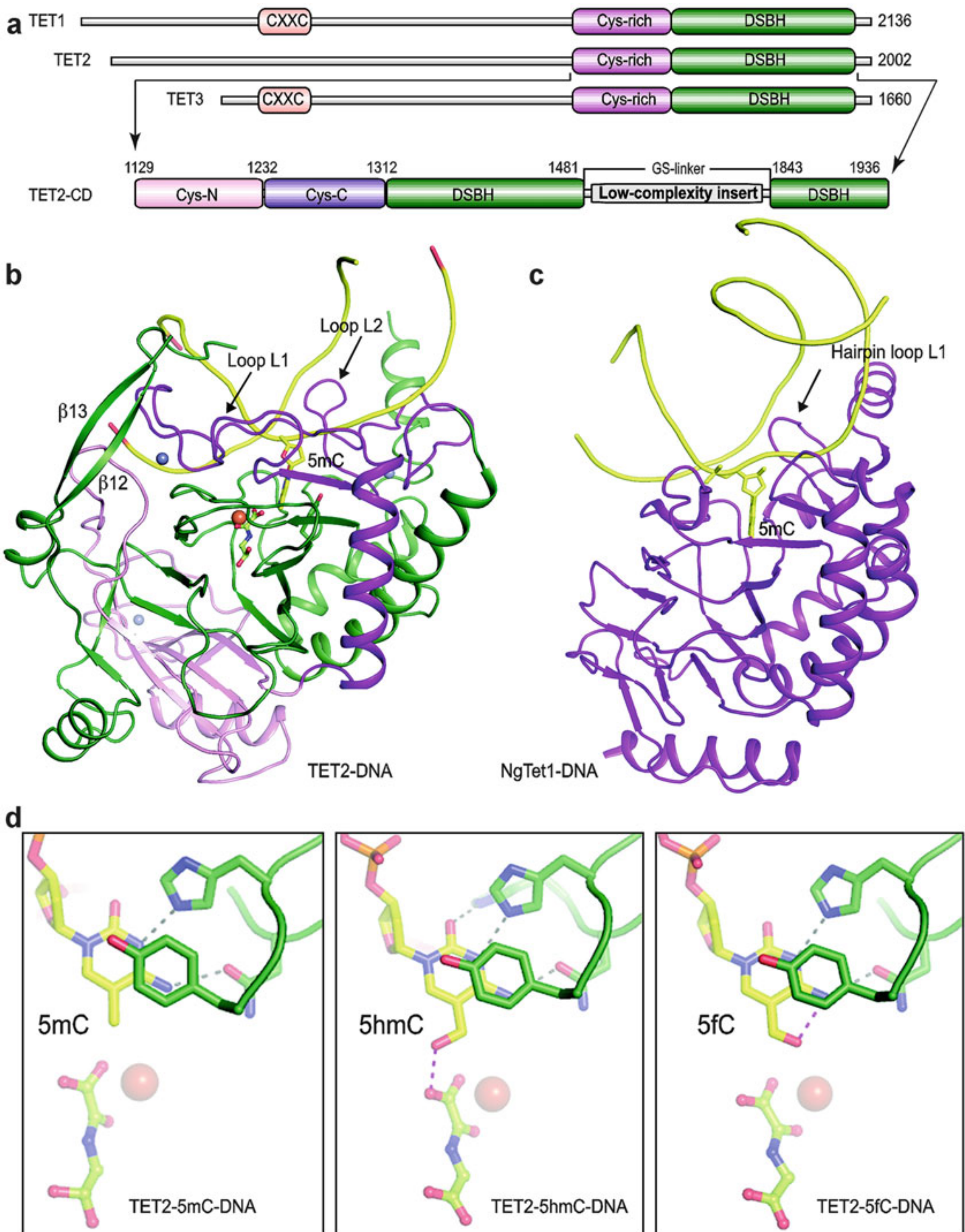


Fig. 10.3 Structures of TET enzymes in complex with DNA substrates (**a**) Color-coded domain architecture of human TET enzymes. All three TET family members share conserved Cys-rich and DSBH domains, which constitute the catalytic domain and are both essential for enzymatic activities of TET enzymes. TET1 and TET3 contain a CXXC domain, which recognizes CpGs and is essential for the function of TET enzymes in vivo. The TET2 construct used for structural study is indicated below. (**b**) Crystal structure of human TET2 in complex with methylated DNA. The color scheme is used as in **a**. The DNA is shown in ribbon representation and colored in *yellow*. NOG, the α-KG analog, is shown in stick representation. Fe(II) and zinc cations are shown as *red* and *gray* balls, respectively. (**c**) Ribbon representation of the crystal structure of NgTet1-DNA complex. The structure

mice display normal overall health and brain development (Gao et al. 2013; Rudenko et al. 2013). However, depletion of Tet1 in mice leads to a hypermethylation and downregulation of genes involved in progenitor cells proliferation and therefore impairs hippocampal neurogenesis, resulting in poor learning and memory (Gao et al. 2013). Another group reported that Tet1 ablation gives rise to the downregulation of genes involved in neuronal activity and results in impairment of memory extinction, synaptic plasticity, and hippocampal long-term depression (Rudenko et al. 2013). Like Tet1, Tet3 was found to regulate synaptic transmission and homeostatic plasticity via DNA demethylation (Yu et al. 2015). TET3 is necessary to maintain silencing of pluripotency genes and consequently neural stem cell identity, possibly through regulation of DNA methylation levels in neural precursor cells (Santiago et al. 2020). In addition, some reports showed that TETs-mediated DNA hydroxymethylation is involved in the axon regeneration, remyelination, and repair of spinal cord injury (Moyon et al. 2021; Weng et al. 2017; Yu et al. 2015).

10.4 Structure of TET Enzymes

10.4.1 Domain Structure of Human TET Enzymes

The three TET family members (TET1, TET2, and TET3) share a less conserved N-terminal region and highly conserved C-terminal catalytic domain with a fold characteristic for α-KG and Fe (II)-dependent dioxygenases (Fig. 10.3a). TET1 (also known as CXXC6) and TET3 (also known as CXXC10) contain an N-terminal zinc finger cysteine-X-X-cysteine (CXXC) domain, which was previously thought to recognize unmethylated CpGs (Long et al. 2013). Intriguingly, *TET2* does not encode a CXXC domain, but in the genome, it is located close to the *IDAX* gene, which encodes a CXXC domain similar to that of TET1 and TET3. IDAX directly interacts with TET2 and is enriched at unmethylated CpGs (Ko et al. 2013). Genome-wide analysis in mESCs indicates that Tet1 preferentially localizes to the TSS of unmethylated CpG-rich promoters and within genes (Williams et al. 2011b; Wu et al. 2011). Biochemical analysis indicates that the CXXC domain of Tet1 binds unmethylated CpG and methylated-CpG DNA (Xu et al. 2011b). A further structural study has indicated that the Tet3 CXXC domain prefers unmethylated cytosines within CpG or non-CpG DNA, and the CXXC domain is critical for Tet3 targeting (Xu et al. 2012). Thus, the CXXC domain of TET enzymes may recognize CpG-containing DNA and accommodate cytosine methylation, thereby providing flexibility for their genomic targeting. Therefore, it is of interest to investigate how the TET CXXC domains recognize CpG DNA when the cytosine is replaced by 5hmC, 5fC, and 5caC.

A previous study has predicted that TET enzymes contain a double-stranded β-helix (DSBH) fold, which is a characteristic domain of α-KG/Fe(II)-dependent dioxygenases and a cysteine-rich (Cys-rich) domain at the N-terminus of the DSBH (Iyer et al. 2009). Both the DSBH and Cys-rich domains are highly conserved among TET enzymes and across species. There is a much less conserved

Fig. 10.3 (continued) is shown in a similar orientation to that of TET2 in **b** for comparison. (**d**) Structural comparison of TET2-5mC-DNA (PDB: 4NM6, 2.02Å resolution), TET2-5hmC-DNA (PDB: 5DEU, 1.80 Å resolution), and TET2-5fC-DNA (PDB: 5D9Y, 1.97 Å resolution) complexes (Hu et al. 2013, 2015). The three complexes adopt similar overall structures (not shown). The close-up views of the TET2–DNA interactions show the different conformation of 5mC/5hmC/5fC within the catalytic cavity of TET2. Critical bases or residues are shown in stick representation. Hydrogen bonds are indicated as *dashed lines*. The nitrogen, oxygen, and phosphorous atoms are shown in *blue*, *red*, and *orange*, respectively. All the structural figures were modified from published literatures (Hashimoto et al. 2014; Hu et al. 2013, 2015)

low-complexity insert within the core DSBH domain. The deletion of the insert does not obviously affect the in vitro enzymatic activity of TET2 (Hu et al. 2013). However, the insert is present across the entire TET family enzymes, suggesting that it might be important for TET functions in vivo.

10.4.2 Crystal Structure of the TET2-5mC-DNA Complex

The function of TET enzymes has been extensively studied since the discovery of the TET-mediated 5mC oxidation in 2009 (Branco et al. 2011; Cimmino et al. 2011; Pastor et al. 2013; Tan and Shi 2012; Williams et al. 2011a; Wu and Zhang 2011, 2014; Xu and Walsh 2014). However, a few fundamental questions remain to be addressed. For example, how do TET enzymes specifically recognize their DNA substrate? How do TET enzymes successively oxidize 5mC to 5hmC, 5fC, and 5caC? Finally, how do patient-derived mutations of TET2 affect its enzymatic activity and contribute to oncogenesis? The three-dimensional structure of TET enzymes in complex with DNA substrate will provide valuable information to address these questions.

The main challenge for structural studies of TET enzymes is the difficulty in obtaining well-behaved proteins for crystallization. Hu et al. have mapped the minimal regions required for the enzymatic activity of TET2 (Hu et al. 2013). As shown in Fig. 10.3a, a human TET2 construct corresponding to residues 1129–1936 with an internal deletion (residues 1481–1843) maintains enzymatic activity and largely improves the solubility and yield of protein expression and purification. The TET2-5mC-DNA complex structure has been determined with this fragment at 2.02 Å resolution (Hu et al. 2013) (Fig. 10.3b). The α-KG analog *N*-oxalylglycine (NOG) was used to avoid the oxidation of 5mC in the crystals. Notably, the TET2 regions visible in the crystal structure are highly conserved in most TET enzymes, indicating that the structural features described below apply to other TET enzymes as well. The TET2 structure shows a compact fold of the catalytic domain in complex with the 5mC-DNA duplex. The central DSBH core is formed by two β-sheets and stabilized by other regions from both sides and on the bottom. The Cys-rich domain wraps around the DSBH core and is separated into Cys-rich N-terminal (Cys-N) and C-terminal (Cys-C) subdomains. Two β-strands (β12 and β13) should be connected by the large low-complexity insert, which has been removed for crystallization (Fig. 10.3b). This organization is consistent with the observation that the insert is not required for TET activity and suggests that in the full structure, the insert is pointing away from the central catalytic domain of TET2.

The overall structure of TET2 is further stabilized by three zinc cations. Notably, Zn2 and Zn3 are coordinated by residues from both Cys-rich and DSBH domains and thus bring flexible regions from the two domains together to facilitate the overall structure formation. Sequence analysis has indicated that all of the residues involved in zinc coordination are highly conserved, suggesting that this architecture is critical for TET enzymes. The Fe(II) ion and NOG cofactor are localized in the center of the DSBH core domain and are bound and stabilized by highly conserved residues. Mutation of these residues significantly decreases or abolishes the enzymatic activity of TET2 (Hu et al. 2013; Ko et al. 2010). As described above, most of the patient-derived somatic cancer mutations in TET2 occur at the residues for zinc and iron coordination or the α-KG interaction. However, there are quite several mutations occurring at residues that may not directly impair TET2 activity. For example, some of the mutations are at residues within the insert region or the N-terminus. Whether and how these mutations contribute to oncogenesis requires further investigation.

The methylated dsDNA is located above the DSBH core and is stabilized by two loops (L1 and L2) from the Cys-C subdomain (Fig. 10.3b). One methylcytosine (5mC) is flipped out of the DNA duplex and inserted into the catalytic cavity. As a replacement, a hydrophobic loop fills in this gap within the double-stranded DNA. A highly

conserved residue (Y1294) stabilizes the G:C base pair of the CpG dinucleotide through a base-stacking interaction. Therefore, TET2 may use a tipping mechanism to search for the modified CpG when sliding along the DNA (Tsai and Tainer 2013) in a mechanism similar to that observed for AlkB homolog 2 (ABH2) (Yang et al. 2009; Yang et al. 2008). The comparison between TET2-DNA and ABH2-DNA structures has been described previously (Hu et al. 2013; Tsai and Tainer 2013) and will not be discussed here.

Within the catalytic cavity, the 5mC base is specifically recognized by TET2 through a network of interactions, which allow the 5mC to adopt a specific orientation so that the methyl group faces toward the catalytic center for reaction (Hu et al. 2013). The catalytic cavity is large enough to accommodate 5mC and its derivatives for further oxidation. With the exception of the methyl-CpG dinucleotide, only the DNA phosphate groups are involved in the TET2-DNA contacts. Further biochemical analysis supports that TET2 has strong sequence preference toward the CpG dinucleotide (Hu et al. 2013).

10.4.3 Crystal Structure of the NgTet1-5mC-DNA Complex

TET enzymes are widely distributed across species, including the heterolobosean amoeboflagellate *Naegleria gruberi*. The genome of *Naegleria* encodes eight Tet-like dioxygenases (NgTet1-8). Sequence analysis has indicated that the NgTet enzymes have a DSBH core region and Fe(II)-chelating residues (HxD...H motif) but lack the Cys-rich region. Biochemical analyses have demonstrated that NgTet1 can successively oxidize 5mC to 5hmC, 5fC, and 5caC using DNA substrates in the XpG (X = 5mC, 5hmC, or 5fC) dinucleotide context (Hashimoto et al. 2014). Hashimoto et al. have determined the crystal structure of NgTet1 in complex with a 14-bp methylated DNA at 2.9 Å resolution (Hashimoto et al. 2014).

NgTet1 has a DSBH core formed by two β-sheets, with the eight-stranded β-sheet stabilized by five α-helices (Fig. 10.3c). The DNA binds to NgTet1 on the basic surface and the flipped 5mC inserts into the catalytic cavity. The overall structure and the pattern of DNA recognition are similar to that observed in the structure of TET2-DNA complex. The hairpin loop L1 of NgTet1 is equivalent to loop L2 of human TET2, which is important for DNA recognition. NgTet1 lacks the Cys-rich region. As a result, the enzyme lacks the equivalents of loop L1 and the regions involved in the coordination of the three zinc cations in human TET2. Moreover, no insert within the DSBH is present in any of the eight NgTet enzymes. TET enzymes may have gained these additional regions (loop L1 and the insert) for potential regulatory functions during evolution.

10.4.4 Structural Basis for Substrate Preference in TET-Mediated Oxidation

Previous studies showed that 5hmC is much (~10–100-fold) more abundant than 5fC/5caC (Globisch et al. 2010; Ito et al. 2011; Mellen et al. 2012; Pfaffeneder et al. 2011; Song et al. 2011; Tahiliani et al. 2009). The presence of TDG seems not to be predominately responsible for such different abundance of 5hmC and 5fC/5caC, because the depletion of TDG leads to an accumulation of 5fC and 5caC by 2–10-fold, but no apparent changes of the 5hmC and 5mC level in mouse ESCs (Shen et al. 2013). In vitro enzymatic analyses also show that TET enzymes, including human TET1/2, mouse Tet2, and *Naegleria* Tet-like protein, possess higher activity for DNA substrate containing 5mC than 5hmC/5fC-DNA (Hashimoto et al. 2014; Hu et al. 2015; Ito et al. 2011). These studies suggest that TET enzymes might play a major role in controlling the cellular level of 5mC oxidized derivatives.

We have previously determined the crystal structures of human TET2 in complex with 5hmC-DNA and 5fC-DNA (Hu et al. 2015)

(Fig. 10.3d). The structural analyses indicate that 5hmC or 5fC is specifically recognized by TET2 in a manner similar to that of 5mC in TET2-5mC-DNA structure (Hu et al. 2013). The cytosine portion of 5mC/5hmC/5fC adopts an almost identical conformation within the catalytic cavity in the three structures (Hu et al. 2013; Hu et al. 2015). The major difference between 5hmC and 5fC is that the hydroxyl group of 5hmC and carbonyl group of 5fC face toward opposite directions, because the hydroxymethyl group of 5hmC and the formyl group of 5fC form hydrogen bonds with 1-carboxylate of NOG and N4 exocyclic nitrogen of cytosine, respectively. The hydrogen bonds prevent the C-C bond between carbon 5 of cytosine and the methyl group (5hmC or 5fC) from free rotation. Therefore, the hydroxymethyl group of 5hmC and the formyl group of 5fC adopt restrained conformations within the catalytic cavity, whereas the methyl group of 5mC is not restrained, because no hydrogen bond is formed. Further biochemical analyses and molecular dynamic simulations suggest that such a restrained conformation may prevent the hydrogen(s) of 5hmC/5fC from adopting an orientation favorable for hydrogen abstraction during catalysis leading to low catalytic efficiency. The residues for catalytic cavity formation are highly conserved, which suggests that this process described a general mechanism for TET enzymes.

10.4.5 Crystal Structure of Algal TET Homologue CMD1 in Complex with VC and 5mC-DNA

C^5-glyceryl-methylcytosine (5gmC) is a novel DNA modification catalyzed by algal TET homolog CMD1 using vitamin C (VC) as co-substrate (Hu et al. 2013). CMD1 adds a glyceryl moiety to the C^5-methyl group of 5mC, generating a new DNA base modification named C^5-glyceryl-methylcytosine (5gmC). Li et al. report the structures of CMD1 in apo form and in complexes with VC or/and dsDNA (Li et al. 2021b). CMD1 contains the canonical DSBH fold of Fe(II)/2-OG-dependent dioxygenases. VC in the lactone form binds to the active site. Despite the low overall sequence identity, structural comparison reveals that the active site of CMD1 is similar to that of HsTET2 and NgTet1. However, the co-substrate binding pocket of CMD1 is significantly different from that of the TET proteins. The lack of a basic residue in stabilizing the carboxyl moiety of 2-OG and the potential steric hindrance might prevent the binding of 2-OG within the active site of CMD1, preventing the enzyme from "canonical" Tet activity.

10.5 Regulation of TET Enzymes

10.5.1 Inhibitors

As a co-substrate, α-KG directly binds to TET enzymes and is converted into succinate and carbon dioxide during each catalytic cycle. Succinate and α-KG bind to the catalytic cavity of TET enzymes in a similar manner. In various tumors, the pathological accumulation of natural metabolites (succinate and fumarate) and oncometabolites (2-hydroxyglutarate, 2HG) has been observed, which are structurally similar to α-KG and therefore lead to competitive inhibition of TET enzymes (Fig. 10.2a).

Isocitrate dehydrogenases (IDH) are important metabolic enzymes involved in the tricarboxylic acid (TCA) cycle through converting isocitrate to α-KG. The *IDH1/2* genes are frequently mutated in AML (Mardis et al. 2009), melanoma (Shibata et al. 2011), glioma (Parsons et al. 2008), and thyroid carcinomas (Hemerly et al. 2010). The *IDH1/2* mutations result in a gain of enzymatic activity for the production and accumulation of the oncometabolite R-2-hydroxyglutarate (R-2HG) (Figueroa et al. 2010). The resultant R-2HG functions as α-KG analog and competitive inhibitor of various α-KG/Fe(II)-dependent dioxygenases, including JmjC-containing histone demethylases and TET enzymes (Xu et al. 2011a; Ye et al. 2013). Loss-of-function mutations of two other key enzymes in the TCA cycle [fumarate hydratase (FH) and succinate dehydrogenase (SDH)] have also been observed in various

tumors. These mutations lead to the accumulation of their substrates (fumarate and succinate), which results in similar effects to IDH1/2 mutations (Xiao et al. 2012).

NOG (*N*-oxalylglycine) is an inactive analog of α-KG and binds to α-KG/Fe(II)-dependent dioxygenases in a manner similar to that of α-KG (Cloos et al. 2006; Hamada et al. 2009). However, it is unable to undergo decarboxylation and thus is commonly used as an inhibitor to block the enzymatic activity of these enzymes for in vitro biochemical and structural studies. However, NOG inhibits TET enzymes and all other α-KG/Fe(II)-dependent dioxygenases and is not an ideal inhibitor for in vivo studies. Inhibitors with high selectivity (only targeting TET enzymes) would advance the functional studies of TET enzymes under biological and pathological conditions, such as the early stage of embryo development.

TET proteins are direct substrates of calpains which are non-lysosomal cysteine proteases, and TETs are subject to calpain-mediated degradation. Calpain members have distinct functions. Specifically, calpain1 modulates TET1 and TET2 levels in mESCs, while calpain2 promotes TET3 turnover during neural differentiation (Wang and Zhang 2014).

10.5.2 Activators

Vitamin C (also known as L-ascorbic acid) is a dietary nutrient that is critical for mammals. It functions as an antioxidant and plays an important role in biosynthesis of collagen, catecholamine, and carnitine (Englard and Seifter 1986) and facilitates the generation of iPSCs (Esteban et al. 2010). Vitamin C has been reported to enhance the enzymatic activities of a number of α-KG/Fe(II)-dependent dioxygenases, including prolyl 4-hydroxylase (P4H) (Myllyla et al. 1978) and JmjC-containing histone demethylases (Wang et al. 2011). The vitamin C-induced enhancement of TET activity has been observed in vitro and under various physiological conditions, including mouse ESCs and MEFs, and shown to participate in cellular differentiation and lineage specification (Agathocleous et al. 2017; Blaschke et al. 2013; Chen et al. 2013a; Minor et al. 2013; Yin et al. 2013; Yue et al. 2016, 2021) (Fig. 10.2a). Therefore, vitamin C may contribute to epigenetic remodeling through regulating JmjC-containing histone demethylases and TET enzymes (Young et al. 2015). Deprivation of vitamin C is closely associated with decreased TET activity and oncogenesis, and has been considered as an epigenetic therapeutic (Agathocleous et al. 2017; Cimmino et al. 2017; Das et al. 2019, 2021). Vitamin C is generally believed to function as an antioxidant to prevent the oxidation of Fe(II). However, the activity enhancement of α-KG/Fe(II)-dependent dioxygenases does not occur when vitamin C is replaced by other reducing agents and antioxidants, such as DTT, glutathione, and L-cysteine (Blaschke et al. 2013; Yin et al. 2013), indicating a more specific effect. Some results show that vitamin C serves as a co-substrate of HIF PHD (hypoxia inducible factor prolyl hydroxylases, another member of non-heme iron α-ketoglutarate dioxygenases) that may compete for the binding site of αKG in the enzyme active center (Osipyants et al. 2018). Vitamin C appears to bind to the C-terminal catalytic domain of TET enzymes (Yin et al. 2013), but the underlying mechanism for the activity enhancement needs further investigation.

Adenosine triphosphate (ATP) has been found to enhance the in vitro activity of TET enzymes (He et al. 2011) (Fig. 10.2a). Although the cellular concentration of ATP is high enough for the enhancement, the physiological relevance of this finding remains unknown. Because the expression and activity of TET enzymes are impaired in various tumors, it is of interest to test whether vitamin C, ATP, or other TET activators yet to be discovered could be used for cancer treatment through enhancing the activity of TET enzymes.

The activity of TET proteins is also regulated by poly (ADP-ribosyl)ation (PARylation). Noncovalent binding of ADP-ribose polymers to TET1 catalytic domain decreases TET1 hydroxylase activity, while the covalent PARylation stimulates the TET1 enzyme (Ciccarone et al. 2015).

10.5.3 Interacting Proteins

O-linked β-*N*-acetylglucosamine transferase (OGT) is an enzyme that transfers the O-GlcNac moiety to the hydroxyl groups of threonine, serine, and cysteine residues of various protein substrates for specific regulations (Hanover et al. 2012; Maynard et al. 2016). The nutrient-responsive enzyme OGT directly interacts with the DSBH domain of TET enzymes, suggesting a direct link between metabolism and epigenomes (Chen et al. 2013b; Deplus et al. 2013; Vella et al. 2013). A following study further mapped this interaction to a short C-terminal region of TET1 and reported in vitro stimulation of TET activity after O-GlcNAcylation by OGT (Hrit et al. 2018). Genome-wide analyses have indicated that TET enzymes recruit OGT to CpG-rich promoters. The depletion of TET enzymes impairs the chromatin association of OGT and OGT-mediated O-GlcNAcylation of histone H2B (Chen et al. 2013b) and HCF1 (host cell factor 1) (Deplus et al. 2013). The OGT–TET interaction is necessary for TET1 function in embryonic development (Hrit et al. 2018) and is involved in cancer progression (Ciesielski et al. 2021; Li et al. 2021a).

Tet1 binds to and co-localizes with the SIN3A co-repressor complex. Tet1 contributes to the genomic targeting of SIN3A, while SIN3A does not affect Tet1 binding to the target genes (Williams et al. 2011b). Studies have indicated that WT1, encoded by Wilms' tumor gene *WT1*, interacts with and recruits TET2 to its target genes and regulates their gene expression (Rampal et al. 2014; Wang et al. 2015b). Interestingly, *WT1*, *TET2*, and *IDH1/2* are mutated in a mutually exclusive manner in AML, suggesting that these genes act in the same pathway for the suppression of oncogenesis. Moreover, Tet2 can recruit Hdac2 to interleukin-6 (IL-6) promoter and by this selectively mediate active repression of interleukin-6 (IL-6) transcription during inflammation resolution in innate myeloid cells via histone deacetylation (Zhang et al. 2015b). EBF1 interacts with TET2 at hypermethylated loci (Guilhamon et al. 2013), probably providing an explanation for the association of EBF1 and induction of CD79α promoter demethylation. As described above, IDAX (also known as CXXC4) interacts directly with the catalytic domain of TET2. IDAX recruits TET2 to DNA and promotes TET2 degradation by activating caspase (Ko et al. 2013). TET2 could be phosphorylated by JAK2 and phosphorylated TET2 interacts with KLF1 and is recruited to KLF1 binding motif, thus promoting the erythroid differentiation program (Jeong et al. 2019).

10.6 Concluding Remarks

It has now been well established that TET enzymes mediate 5mC oxidation in the genome and play important roles in DNA demethylation, gene transcription, embryonic development, and oncogenesis. What remains to be addressed is how the activity and genomic localization of TET enzymes are precisely determined and dynamically regulated, especially during developmental and pathological processes. What are the key factors that allow TET enzymes to exhibit basal activity to generate 5hmC or higher activity to generate 5fC/5caC in specific genomic regions? What are the specific signaling roles of 5hmC, 5fC, and 5caC? Are there specific readers to interpret these modifications? Moreover, specific TET inhibitors would provide valuable tools to study whether TET enzymes could be potential drug targets for therapeutic applications. Innovations of next-generation sequencing technologies for specific 5mC/5hmC sequencing from a small amount of genomic DNA could facilitate studies of the 5mC/5hmC dynamics during mammalian early embryogenesis.

Acknowledgments This work was supported by grants from the National Natural Science Foundation of China (31425008 and 91419301). We apologize that we could not cite many important papers due to space limitation.

References

Abdel-Wahab O, Mullally A, Hedvat C, Garcia-Manero G, Patel J, Wadleigh M, Malinge S, Yao J, Kilpivaara O,

Bhat R et al (2009) Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood 114:144–147

Agathocleous M, Meacham CE, Burgess RJ, Piskounova E, Zhao Z, Crane GM, Cowin BL, Bruner E, Murphy MM, Chen W et al (2017) Ascorbate regulates haematopoietic stem cell function and leukaemogenesis. Nature 549:476–481

Amouroux R, Nashun B, Shirane K, Nakagawa S, Hill PW, D'Souza Z, Nakayama M, Matsuda M, Turp A, Ndjetehe E et al (2016) De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nat Cell Biol 18:225–233

Antunes C, Sousa N, Pinto L, Marques CJ (2019) TET enzymes in neurophysiology and brain function. Neurosci Biobehav Rev 102:337–344

Aravind L, Koonin EV (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. Genome Biol 2:RESEARCH0007

Arioka Y, Watanabe A, Saito K, Yamada Y (2012) Activation-induced cytidine deaminase alters the subcellular localization of Tet family proteins. PLoS One 7:e45031

Basanta-Sanchez M, Wang R, Liu Z, Ye X, Li M, Shi X, Agris PF, Zhou Y, Huang Y, Sheng J (2017) TET1-Mediated Oxidation of 5-Formylcytosine (5fC) to 5-Carboxycytosine (5caC) in RNA. Chembiochem 18:72–76

Bell E, Curry EW, Megchelenbrink W, Jouneau L, Brochard V, Tomaz RA, Mau KHT, Atlasi Y, de Souza RA, Marks H et al (2020) Dynamic CpG methylation delineates subregions within super-enhancers selectively decommissioned at the exit from naive pluripotency. Nat Commun 11:1112

Bian K, Lenz SAP, Tang Q, Chen F, Qi R, Jost M, Drennan CL, Essigmann JM, Wetmore SD, Li D (2019) DNA repair enzymes ALKBH2, ALKBH3, and AlkB oxidize 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine in vitro. Nucleic Acids Res 47: 5522–5529

Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16:6–21

Blaschke K, Ebata KT, Karimi MM, Zepeda-Martinez JA, Goyal P, Mahapatra S, Tam A, Laird DJ, Hirst M, Rao A et al (2013) Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature 500:222–226

Branco MR, Ficz G, Reik W (2011) Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat Rev Genet 13:7–13

Bruniquel D, Schwartz RH (2003) Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. Nat Immunol 4:235–240

Cai J, Chen L, Zhang Z, Zhang X, Lu X, Liu W, Shi G, Ge Y, Gao P, Yang Y et al (2019) Genome-wide mapping of 5-hydroxymethylcytosines in circulating cell-free DNA as a non-invasive approach for early detection of hepatocellular carcinoma. Gut 68:2195–2205

Caldwell BA, Liu MY, Prasasya RD, Wang T, DeNizio JE, Leu NA, Amoh NYA, Krapp C, Lan Y, Shields EJ et al (2021) Functionally distinct roles for TET-oxidized 5-methylcytosine bases in somatic reprogramming to pluripotency. Mol Cell 81:859–869 e858

Charlton J, Jung EJ, Mattei AL, Bailly N, Liao J, Martin EJ, Giesselmann P, Brandl B, Stamenova EK, Muller FJ et al (2020) TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. Nat Genet 52:819-+

Chen CC, Wang KY, Shen CK (2012) The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. J Biol Chem 287:33116–33121

Chen J, Guo L, Zhang L, Wu H, Yang J, Liu H, Wang X, Hu X, Gu T, Zhou Z et al (2013a) Vitamin C modulates TET1 function during somatic cell reprogramming. Nat Genet 45:1504–1509

Chen Q, Chen Y, Bian C, Fujiki R, Yu X (2013b) TET2 promotes histone O-GlcNAcylation during gene transcription. Nature 493:561–564

Ciccarone F, Valentini E, Zampieri M, Caiafa P (2015) 5mC-hydroxylase activity is influenced by the PARylation of TET1 enzyme. Oncotarget 6:24333–24347

Ciesielski P, Jóźwiak P, Forma E, Krześlak A (2021) TET3- and OGT-dependent expression of genes involved in epithelial-mesenchymal transition in endometrial cancer. Int J Mol Sci 22

Cimmino L, Abdel-Wahab O, Levine RL, Aifantis I (2011) TET family proteins and their role in stem cell differentiation and transformation. Cell Stem Cell 9: 193–204

Cimmino L, Dolgalev I, Wang Y, Yoshimi A, Martin GH, Wang J, Ng V, Xia B, Witkowski MT, Mitchell-Flack M et al (2017) Restoration of TET2 function blocks aberrant self-renewal and leukemia progression. Cell 170:1079–1095 e1020

Cliffe LJ, Kieft R, Southern T, Birkeland SR, Marshall M, Sweeney K, Sabatini R (2009) JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. Nucleic Acids Res 37:1452–1462

Cloos PA, Christensen J, Agger K, Maiolica A, Rappsilber J, Antal T, Hansen KH, Helin K (2006) The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. Nature 442:307–311

Costa Y, Ding J, Theunissen TW, Faiola F, Hore TA, Shliaha PV, Fidalgo M, Saunders A, Lawrence M, Dietmann S et al (2013) NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature 495:370–374

Crawford DJ, Liu MY, Nabel CS, Cao XJ, Garcia BA, Kohli RM (2016) Tet2 catalyzes stepwise

5-methylcytosine oxidation by an iterative and de novo mechanism. J Am Chem Soc 138:730–733

Dai HQ, Wang BA, Yang L, Chen JJ, Zhu GC, Sun ML, Ge H, Wang R, Chapman DL, Tang F et al (2016) TET-mediated DNA demethylation controls gastrulation by regulating Lefty-Nodal signalling. Nature 538: 528–532

Das AB, Kakadia PM, Wojcik D, Pemberton L, Browett PJ, Bohlander SK, Vissers MCM (2019) Clinical remission following ascorbate treatment in a case of acute myeloid leukemia with mutations in TET2 and WT1. Blood Cancer J 9:82

Das AB, Smith-Diaz CC, Vissers MCM (2021) Emerging epigenetic therapeutics for myeloid leukemia: modulating demethylase activity with ascorbate. Haematologica 106:14–25

Dawlaty MM, Ganz K, Powell BE, Hu YC, Markoulaki S, Cheng AW, Gao Q, Kim J, Choi SW, Page DC et al (2011) Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell 9:166–175

Dawlaty MM, Breiling A, Le T, Raddatz G, Barrasa MI, Cheng AW, Gao Q, Powell BE, Li Z, Xu M et al (2013) Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev Cell 24:310–323

DeNizio JE, Liu MY, Leddin EM, Cisneros GA, Kohli RM (2019) Selectivity and promiscuity in TET-mediated oxidation of 5-methylcytosine in DNA and RNA. Biochemistry 58:411–421

DeNizio JE, Dow BJ, Serrano JC, Ghanty U, Drohat AC, Kohli RM (2021) TET-TDG active DNA demethylation at CpG and Non-CpG sites. J Mol Biol 433: 166877

Deplus R, Delatte B, Schwinn MK, Defrance M, Mendez J, Murphy N, Dawson MA, Volkmar M, Putmans P, Calonne E et al (2013) TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. EMBO J 32: 645–655

Dey AS, Bhattacharya C, Guan Y, Jha BK, Mukherji M (2021) Demethylation of non-CpG sites in DNA is initiated by TET2 5-methylcytosine dioxygenase. DNA 1:26–36

Diotel N, Merot Y, Coumailleau P, Gueguen MM, Serandour AA, Salbert G, Kah O (2017) 5-hydroxymethylcytosine marks postmitotic neural cells in the adult and developing vertebrate central nervous system. J Comp Neurol 525:478–497

Doege CA, Inoue K, Yamashita T, Rhee DB, Travis S, Fujita R, Guarnieri P, Bhagat G, Vanti WB, Shih A et al (2012) Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. Nature 488:652–655

Englard S, Seifter S (1986) The biochemical functions of ascorbic acid. Annu Rev Nutr 6:365–406

Esteban MA, Wang T, Qin B, Yang J, Qin D, Cai J, Li W, Weng Z, Chen J, Ni S et al (2010) Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. Cell Stem Cell 6:71–79

Feng Y, Chen JJ, Xie NB, Ding JH, You XJ, Tao WB, Zhang X, Yi C, Zhou X, Yuan BF et al (2021a) Direct decarboxylation of ten-eleven translocation-produced 5-carboxylcytosine in mammalian genomes forms a new mechanism for active DNA demethylation. Chem Sci 12:11322–11329

Feng Y, Xie N-B, Tao W-B, Ding J-H, You X-J, Ma C-J, Zhang X, Yi C, Zhou X, Yuan B-F et al (2021b) Transformation of 5-carboxylcytosine to cytosine through C–C bond cleavage in human cells constitutes a novel pathway for DNA demethylation. CCS Chem 3:994–1008

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature 473:398–402

Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF et al (2010) Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell 18:553–567

Frauer C, Hoffmann T, Bultmann S, Casa V, Cardoso MC, Antes I, Leonhardt H (2011) Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. PLoS One 6:e21306

Fu L, Guerrero CR, Zhong N, Amato NJ, Liu Y, Liu S, Cai Q, Ji D, Jin SG, Niedernhofer LJ et al (2014) Tet-mediated formation of 5-hydroxymethylcytosine in RNA. J Am Chem Soc 136:11582–11585

Fu Y, Luo GZ, Chen K, Deng X, Yu M, Han D, Hao Z, Liu J, Lu X, Dore LC et al (2015) N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. Cell 161:879–892

Gao Y, Chen J, Li K, Wu T, Huang B, Liu W, Kou X, Zhang Y, Huang H, Jiang Y et al (2013) Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. Cell Stem Cell 12:453–469

Gao L, Yang M, Wei Z, Gu M, Yang L, Bai C, Wu Y, Li G (2020) MSTN mutant promotes myogenic differentiation by increasing demethylase TET1 expression via the SMAD2/SMAD3 pathway. Int J Biol Sci 16:1324–1334

Ghanty U, DeNizio JE, Liu MY, Kohli RM (2018) Exploiting substrate promiscuity to develop activity-based probes for ten-eleven translocation family enzymes. J Am Chem Soc 140:17329–17332

Ghanty U, Wang T, Kohli RM (2020) Nucleobase modifiers identify TET enzymes as bifunctional DNA dioxygenases capable of direct N-demethylation. Angew Chem Int Ed Engl 59:11312–11315

Ginno PA, Gaidatzis D, Feldmann A, Hoerner L, Imanci D, Burger L, Zilbermann F, Peters A, Edenhofer F, Smallwood SA et al (2020) A genome-

scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. Nat Commun 11:2680

Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, Koch S, Bruckl T, Biel M, Carell T (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One 5:e15367

Greer EL, Blanco MA, Gu L, Sendinc E, Liu J, Aristizabal-Corrales D, Hsu CH, Aravind L, He C, Shi Y (2015) DNA methylation on N6-adenine in C. elegans. Cell 161:868–878

Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG et al (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 477:606–610

Guilhamon P, Eskandarpour M, Halai D, Wilson GA, Feber A, Teschendorff AE, Gomez V, Hergovich A, Tirabosco R, Fernanda Amary M et al (2013) Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. Nat Commun 4:2166

Guler GD, Ning Y, Ku CJ, Phillips T, McCarthy E, Ellison CK, Bergamaschi A, Collin F, Lloyd P, Scott A et al (2020) Detection of early stage pancreatic cancer using 5-hydroxymethylcytosine signatures in circulating cell free DNA. Nat Commun 11:5270

Guo JU, Su Y, Zhong C, Ming GL, Song H (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell 145: 423–434

Guo F, Li X, Liang D, Li T, Zhu P, Guo H, Wu X, Wen L, Gu TP, Hu B et al (2014a) Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. Cell Stem Cell 15:447–459

Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, Zhong C, Hu S, Le T, Fan G et al (2014b) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat Neurosci 17:215–222

Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, Surani MA (2013) Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science 339:448–452

Haffner MC, Chaux A, Meeker AK, Esopi DM, Gerber J, Pellakuru LG, Toubaji A, Argani P, Iacobuzio-Donahue C, Nelson WG et al (2011) Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. Oncotarget 2:627–637

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

Hamada S, Kim TD, Suzuki T, Itoh Y, Tsumoto H, Nakagawa H, Janknecht R, Miyata N (2009) Synthesis and activity of N-oxalylglycine and its derivatives as Jumonji C-domain-containing histone lysine demethylase inhibitors. Bioorg Med Chem Lett 19: 2852–2855

Hanover JA, Krause MW, Love DC (2012) Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation. Nat Rev Mol Cell Biol 13:312–321

Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM, Zhang X, Cheng X (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res 40:4841–4849

Hashimoto H, Pais JE, Zhang X, Saleh L, Fu ZQ, Dai N, Correa IR Jr, Zheng Y, Cheng X (2014) Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. Nature 506:391–395

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

Hemerly JP, Bastos AU, Cerutti JM (2010) Identification of several novel non-p.R132 IDH1 variants in thyroid carcinomas. Eur J Endocrinol 163:747–755

Hill PWS, Leitch HG, Requena CE, Sun Z, Amouroux R, Roman-Trufero M, Borkowska M, Terragni J, Vaisvila R, Linnett S et al (2018) Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. Nature 555:392–396

Hoffart LM, Barr EW, Guyer RB, Bollinger JM Jr, Krebs C (2006) Direct spectroscopic detection of a C-H-cleaving high-spin Fe(IV) complex in a prolyl-4-hydroxylase. Proc Natl Acad Sci U S A 103:14738–14743

Hoon DS, Spugnardi M, Kuo C, Huang SK, Morton DL, Taback B (2004) Profiling epigenetic inactivation of tumor suppressor genes in tumors and plasma from cutaneous melanoma patients. Oncogene 23:4014–4022

Hrit J, Goodrich L, Li C, Wang BA, Nie J, Cui X, Martin EA, Simental E, Fernandez J, Liu MY et al (2018) OGT binds a conserved C-terminal domain of TET1 to regulate TET1 activity and function in development. Elife 7

Hu L, Li Z, Cheng J, Rao Q, Gong W, Liu M, Shi YG, Zhu J, Wang P, Xu Y (2013) Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. Cell 155:1545–1555

Hu X, Zhang L, Mao SQ, Li Z, Chen J, Zhang RR, Wu HP, Gao J, Guo F, Liu W et al (2014) Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. Cell Stem Cell 14:512–522

Hu L, Lu J, Cheng J, Rao Q, Li Z, Hou H, Lou Z, Zhang L, Li W, Gong W et al (2015) Structural insight into substrate preference for TET-mediated oxidation. Nature 527:118–122

Huang W, Lan MD, Qi CB, Zheng SJ, Wei SZ, Yuan BF, Feng YQ (2016) Formation and determination of the oxidation products of 5-methylcytosine in RNA. Chem Sci 7:5495–5502

Inoue A, Zhang Y (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science 334:194

Inoue A, Shen L, Dai Q, He C, Zhang Y (2011) Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. Cell Res 21:1670–1676

Iqbal K, Jin SG, Pfeifer GP, Szabo PE (2011) Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci U S A 108: 3642–3647

Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 466:1129–1133

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333:1300–1303

Iyer LM, 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 8:1698–1710

Iyer LM, Zhang D, Burroughs AM, Aravind L (2013) Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. Nucleic Acids Res 41:7635–7655

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

Jeong JJ, Gu X, Nie J, Sundaravel S, Liu H, Kuo WL, Bhagat TD, Pradhan K, Cao J, Nischal S et al (2019) Cytokine-regulated phosphorylation and activation of TET2 by JAK2 in hematopoiesis. Cancer Discov 9: 778–795

Kaas GA, Zhong C, Eason DE, Ross DL, Vachhani RV, Ming GL, King JR, Song H, Sweatt JD (2013) TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation. Neuron 79:1086–1093

Kangaspeska S, Stride B, Metivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, Benes V, Gannon F, Reid G (2008) Transient cyclical methylation of promoter DNA. Nature 452:112–115

Kavoosi S, Sudhamalla B, Dey D, Shriver K, Arora S, Sappa S, Islam K (2019) Site- and degree-specific C-H oxidation on 5-methylcytosine homologues for probing active DNA demethylation. Chem Sci 10:10550–10555

Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, An J, Lamperti ED, Koh KP, Ganetzky R et al (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature 468:839–843

Ko M, An J, Bandukwala HS, Chavez L, Aijo T, Pastor WA, Segal MF, Li H, Koh KP, Lahdesmaki H et al (2013) Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature 497:122–126

Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G et al (2011) Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell 8:200–213

Krebs C, Galonic Fujimori D, Walsh CT, Bollinger JM Jr (2007) Non-heme Fe(IV)-oxo intermediates. Acc Chem Res 40:484–492

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

Kurimoto K, Yabuta Y, Ohinata Y, Shigeta M, Yamanaka K, Saitou M (2008) Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev 22:1617–1635

Lan J, Rajan N, Bizet M, Penning A, Singh NK, Guallar D, Calonne E, Li Greci A, Bonvin E, Deplus R et al (2020) Functional role of Tet-mediated RNA hydroxymethylcytosine in mouse ES cells and during differentiation. Nat Commun 11:4956

Langemeijer SM, Kuiper RP, Berends M, Knops R, Aslanyan MG, Massop M, Stevens-Linders E, van Hoogen P, van Kessel AG, Raymakers RA et al (2009) Acquired mutations in TET2 are common in myelodysplastic syndromes. Nat Genet 41:838–842

Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J et al (2010) Dynamic changes in the human methylome during differentiation. Genome Res 20:320–331

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

Lee HJ, Hore TA, Reik W (2014) Reprogramming the methylome: erasing memory and creating diversity. Cell Stem Cell 14:710–719

Li X, Wei W, Zhao QY, Widagdo J, Baker-Andresen D, Flavell CR, D'Alessio A, Zhang Y, Bredy TW (2014) Neocortical Tet3-mediated accumulation of 5-hydroxymethylcytosine promotes rapid behavioral adaptation. Proc Natl Acad Sci U S A 111:7120–7125

Li H, Hu Z, Jiang H, Pu J, Selli I, Qiu J, Zhang B, Feng J (2020a) TET1 deficiency impairs morphogen-free differentiation of human embryonic stem cells to neuroectoderm. Sci Rep 10:10343

Li S, Feng J, Wu F, Cai J, Zhang X, Wang H, Fetahu IS, Iwanicki I, Ma D, Hu T et al (2020b) TET2 promotes anti-tumor immunity by governing G-MDSCs and CD8(+) T-cell numbers. EMBO Rep 21:e49425

Li HJ, Wang Y, Li BX, Yang Y, Guan F, Pang XC, Li X (2021a) Roles of ten-eleven translocation family proteins and their O-linked beta-N-acetylglucosaminylated forms in cancer development. Oncol Lett 21:1

Li W, Zhang T, Sun M, Shi Y, Zhang XJ, Xu GL, Ding J (2021b) Molecular mechanism for vitamin C-derived

C(5)-glyceryl-methylcytosine DNA modification catalyzed by algal TET homologue CMD1. Nat Commun 12:744

Lian CG, Xu Y, Ceol C, Wu F, Larson A, Dresser K, Xu W, Tan L, Hu Y, Zhan Q et al (2012) Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell 150:1135–1146

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:315–322

Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, Antosiewicz-Bourget J, O'Malley R, Castanon R, Klugman S et al (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature 471:68–73

Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD et al (2013) Global epigenomic reconfiguration during mammalian brain development. Science 341: 1237905

Liu CK, Hsu CA, Abbott MT (1973) Catalysis of three sequential dioxygenase reactions by thymine 7-hydroxylase. Arch Biochem Biophys 159:180–187

Liu S, Ren S, Howell P, Fodstad O, Riker AI (2008) Identification of novel epigenetically modified genes in human melanoma via promoter methylation gene profiling. Pigment Cell Melanoma Res 21:545–558

Liutkeviciute Z, Lukinavicius G, Masevicius V, Daujotyte D, Klimasauskas S (2009) Cytosine-5-methyltransferases add aldehydes to DNA. Nat Chem Biol 5:400–402

Liutkeviciute Z, Kriukiene E, Licyte J, Rudyte M, Urbanaviciute G, Klimasauskas S (2014) Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. J Am Chem Soc 136:5884–5887

Loenarz C, Schofield CJ (2008) Expanding chemical biology of 2-oxoglutarate oxygenases. Nat Chem Biol 4: 152–156

Long HK, Blackledge NP, Klose RJ (2013) ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochem Soc Trans 41:727–740

Lu X, Han D, Zhao BS, Song CX, Zhang LS, Dore LC, He C (2015) Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. Cell Res 25:386–389

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

Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, McGrath SD et al (2009) Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med 361:1058–1066

Martinowich K, Hattori D, Wu H, Fouse S, He F, Hu Y, Fan G, Sun YE (2003) DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. Science 302:890–893

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

Maynard JC, Burlingame AL, Medzihradszky KF (2016) Cysteine S-linked N-acetylglucosamine (S-GlcNAcylation), a new post-translational modification in mammals. Mol Cell Proteomics 15:3405–3411

McDonough MA, Loenarz C, Chowdhury R, Clifton IJ, Schofield CJ (2010) Structural studies on human 2-oxoglutarate dependent oxygenases. Curr Opin Struct Biol 20:659–672

Mellen 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 151:1417–1430

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

Mi Y, Gao X, Dai J, Ma Y, Xu L, Jin W (2015) A novel function of TET2 in CNS: sustaining neuronal survival. Int J Mol Sci 16:21846–21857

Minor EA, Court BL, Young JI, Wang G (2013) Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. J Biol Chem 288:13669–13674

Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, Figueroa ME, Vasanthakumar A, Patel J, Zhao X et al (2011) Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell 20: 11–24

Moyon S, Frawley R, Marechal D, Huang D, Marshall-Phelps KLH, Kegel L, Bostrand SMK, Sadowski B, Jiang YH, Lyons DA et al (2021) TET1-mediated DNA hydroxymethylation regulates adult remyelination in mice. Nat Commun 12:3359

Mulholland CB, Traube FR, Ugur E, Parsa E, Eckl EM, Schonung M, Modic M, Bartoschek MD, Stolz P, Ryan J et al (2020) Distinct and stage-specific contributions of TET1 and TET2 to stepwise cytosine oxidation in the transition from naive to primed pluripotency. Sci Rep 10:12066

Muller I, Kahnert A, Pape T, Sheldrick GM, Meyer-Klaucke W, Dierks T, Kertesz M, Uson I (2004) Crystal structure of the alkylsulfatase AtsK: insights into the catalytic mechanism of the Fe(II) alpha-ketoglutarate-dependent dioxygenase superfamily. Biochemistry 43:3075–3088

Munzel M, Globisch D, Bruckl T, Wagner M, Welzmiller V, Michalakis S, Muller M, Biel M, Carell T (2010) Quantification of the sixth DNA base hydroxymethylcytosine in the brain. Angew Chem Int Ed Engl 49:5375–5377

Myllyla R, Kuutti-Savolainen ER, Kivirikko KI (1978) The role of ascorbate in the prolyl hydroxylase reaction. Biochem Biophys Res Commun 83:441–448

Nabel CS, Jia H, Ye Y, Shen L, Goldschmidt HL, Stivers JT, Zhang Y, Kohli RM (2012) AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nat Chem Biol 8:751–758

Nakamura T, Arai Y, Umehara H, Masuhara M, Kimura T, Taniguchi H, Sekimoto T, Ikawa M, Yoneda Y, Okabe M et al (2007) PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol 9:64–71

Neidigh JW, Darwanto A, Williams AA, Wall NR, Sowers LC (2009) Cloning and characterization of Rhodotorula glutinis thymine hydroxylase. Chem Res Toxicol 22:885–893

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

Osipyants AI, Poloznikov AA, Smirnova NA, Hushpulian DM, Khristichenko AY, Chubar TA, Zakhariants AA, Ahuja M, Gaisina IN, Thomas B et al (2018) L-ascorbic acid: A true substrate for HIF prolyl hydroxylase? Biochimie 147:46–54

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. Curr Biol 10:475–478

Parry A, Rulands S, Reik W (2021) Active turnover of DNA methylation during cell fate decisions. Nat Rev Genet 22:59–66

Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL et al (2008) An integrated genomic analysis of human glioblastoma multiforme. Science 321:1807–1812

Pastor WA, Aravind L, Rao A (2013) TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol 14:341–356

Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, Hagemeier C, Carell T (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew Chem Int Ed Engl 50:7008–7012

Pfaffeneder T, Spada F, Wagner M, Brandmayr C, Laube SK, Eisen D, Truss M, Steinbacher J, Hackner B, Kotljarova O et al (2014) Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. Nat Chem Biol 10:574–581

Price JC, Barr EW, Glass TE, Krebs C, Bollinger JM Jr (2003) Evidence for hydrogen abstraction from C1 of taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by taurine:alpha-ketoglutarate dioxygenase (TauD). J Am Chem Soc 125:13008–13009

Quivoron C, Couronne L, Della Valle V, Lopez CK, Plo I, Wagner-Ballon O, Do Cruzeiro M, Delhommeau F, Arnulf B, Stern MH et al (2011) TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell 20:25–38

Rampal R, Alkalin A, Madzo J, Vasanthakumar A, Pronier E, Patel J, Li Y, Ahn J, Abdel-Wahab O, Shih A et al (2014) DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. Cell Rep 9: 1841–1855

Rangam G, Schmitz KM, Cobb AJ, Petersen-Mahrt SK (2012) AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. PLoS One 7:e43279

Rao LJ, Yi BC, Li QM, Xu Q (2016) TET1 knockdown inhibits the odontogenic differentiation potential of human dental pulp cells. Int J Oral Sci 8:110–116

Rasmussen KD, Helin K (2016) Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev 30:733–750

Rodic N, Zampella J, Sharma R, Burns KH, Taube JM (2015) Diagnostic utility of 5-hydroxymethylcytosine immunohistochemistry in melanocytic proliferations. J Cutan Pathol 42:807–814

Rudenko A, Dawlaty MM, Seo J, Cheng AW, Meng J, Le T, Faull KF, Jaenisch R, Tsai LH (2013) Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. Neuron 79:1109–1122

Rulands S, Lee HJ, Clark SJ, Angermueller C, Smallwood SA, Krueger F, Mohammed H, Dean W, Nichols J, Rugg-Gunn P et al (2018) Genome-scale oscillations in DNA methylation during exit from pluripotency. Cell Syst 7:63–76 e12

Ruzov A, Tsenkina Y, Serio A, Dudnakova T, Fletcher J, Bai Y, Chebotareva T, Pells S, Hannoun Z, Sullivan G et al (2011) Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. Cell Res 21:1332–1342

Ryle MJ, Padmakumar R, Hausinger RP (1999) Stopped-flow kinetic analysis of Escherichia coli taurine/alpha-ketoglutarate dioxygenase: interactions with alpha-ketoglutarate, taurine, and oxygen. Biochemistry 38: 15278–15286

Saitou M, Kagiwada S, Kurimoto K (2012) Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. Development 139: 15–31

Santiago M, Antunes C, Guedes M, Iacovino M, Kyba M, Reik W, Sousa N, Pinto L, Branco MR, Marques CJ (2020) Tet3 regulates cellular identity and DNA methylation in neural progenitor cells. Cell Mol Life Sci 77: 2871–2883

Schiesser S, Hackner B, Pfaffeneder T, Muller M, Hagemeier C, Truss M, Carell T (2012) Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. Angew Chem Int Ed Engl 51:6516–6520

Schutsky EK, DeNizio JE, Hu P, Liu MY, Nabel CS, Fabyanic EB, Hwang Y, Bushman FD, Wu H, Kohli

RM (2018) Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. Nat Biotechnol. https://doi.org/10.1038/nbt.4204
Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W (2012) The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell 48:849–862
Shen L, Kondo Y, Guo Y, Zhang J, Zhang L, Ahmed S, Shu J, Chen X, Waterland RA, Issa JP (2007) Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS Genet 3:2023–2036
Shen L, Wu H, Diep D, Yamaguchi S, D'Alessio AC, Fung HL, Zhang K, Zhang Y (2013) Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. Cell 153:692–706
Shen L, Inoue A, He J, Liu Y, Lu F, Zhang Y (2014a) Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. Cell Stem Cell 15:459–471
Shen L, Song CX, He C, Zhang Y (2014b) Mechanism and function of oxidative reversal of DNA and RNA methylation. Annu Rev Biochem 83:585–614
Shen H, Ontiveros RJ, Owens MC, Liu MY, Ghanty U, Kohli RM, Liu KF (2021a) TET-mediated 5-methylcytosine oxidation in tRNA promotes translation. J Biol Chem 296:100087
Shen Y, Liu L, Wang M, Xu B, Lyu R, Shi YG, Tan L (2021b) TET2 Inhibits PD-L1 gene expression in breast cancer cells through histone deacetylation. Cancers (Basel):13
Shibata T, Kokubu A, Miyamoto M, Sasajima Y, Yamazaki N (2011) Mutant IDH1 confers an in vivo growth in a melanoma cell line with BRAF mutation. Am J Pathol 178:1395–1402
Smiley JA, Kundracik M, Landfried DA, Barnes VR Sr, Axhemi AA (2005) Genes of the thymidine salvage pathway: thymine-7-hydroxylase from a Rhodotorula glutinis cDNA library and iso-orotate decarboxylase from Neurospora crassa. Biochim Biophys Acta 1723:256–264
Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. Nat Rev Genet 14:204–220
Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X et al (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol 29:68–72
Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, Street C, Li Y, Poidevin M, Wu H et al (2013) Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell 153:678–691
Song Y, van den Berg PR, Markoulaki S, Soldner F, Dall'Agnese A, Henninger JE, Drotar J, Rosenau N, Cohen MA, Young RA et al (2019) Dynamic enhancer DNA methylation as basis for transcriptional and cellular heterogeneity of ESCs. Mol Cell 75:905–920 e906
Spada F, Schiffers S, Kirchner A, Zhang Y, Arista G, Kosmatchev O, Korytiakova E, Rahimoff R, Ebert C, Carell T (2020) Active turnover of genomic methylcytosine in pluripotent cells. Nat Chem Biol 16:1411–1419
Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res 38:e181
Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L et al (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324:930–935
Tamanaha E, Guan S, Marks K, Saleh L (2016) Distributive processing by the iron(II)/alpha-ketoglutarate-dependent catalytic domains of the TET enzymes is consistent with epigenetic roles for oxidized 5-methylcytosine bases. J Am Chem Soc 138:9345–9348
Tan L, Shi YG (2012) Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development 139:1895–1902
Tefferi A, Levine RL, Lim KH, Abdel-Wahab O, Lasho TL, Patel J, Finke CM, Mullally A, Li CY, Pardanani A et al (2009) Frequent TET2 mutations in systemic mastocytosis: clinical, KITD816V and FIP1L1-PDGFRA correlates. Leukemia 23:900–904
Tsai CL, Tainer JA (2013) Probing DNA by 2-OG-dependent dioxygenase. Cell 155:1448–1450
Valegard K, Terwisscha van Scheltinga AC, Dubus A, Ranghino G, Oster LM, Hajdu J, Andersson I (2004) The structural basis of cephalosporin formation in a mononuclear ferrous enzyme. Nat Struct Mol Biol 11:95–101
Valinluck V, Sowers LC (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res 67:946–950
Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, Sowers LC (2004) Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic Acids Res 32:4100–4108
van der Wijst MG, Venkiteswaran M, Chen H, Xu GL, Plosch T, Rots MG (2015) Local chromatin microenvironment determines DNMT activity: from DNA methyltransferase to DNA demethylase or DNA dehydroxymethylase. Epigenetics 10:671–676
Vella P, Scelfo A, Jammula S, Chiacchiera F, Williams K, Cuomo A, Roberto A, Christensen J, Bonaldi T, Helin K et al (2013) Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. Mol Cell 49:645–656
Wang Y, Zhang Y (2014) Regulation of TET protein stability by calpains. Cell Rep 6:278–284

Wang T, Chen K, Zeng X, Yang J, Wu Y, Shi X, Qin B, Zeng L, Esteban MA, Pan G et al (2011) The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. Cell Stem Cell 9:575–587

Wang L, Zhang J, Duan J, Gao X, Zhu W, Lu X, Yang L, Zhang J, Li G, Ci W et al (2014) Programming and inheritance of parental DNA methylomes in mammals. Cell 157:979–991

Wang L, Zhou Y, Xu L, Xiao R, Lu X, Chen L, Chong J, Li H, He C, Fu XD et al (2015a) Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. Nature 523:621–625

Wang Y, Xiao M, Chen X, Chen L, Xu Y, Lv L, Wang P, Yang H, Ma S, Lin H et al (2015b) WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. Mol Cell 57:662–673

Weng YL, An R, Cassin J, Joseph J, Mi R, Wang C, Zhong C, Jin SG, Pfeifer GP, Bellacosa A et al (2017) An intrinsic epigenetic barrier for functional axon regeneration. Neuron 94:337–346 e336

Williams K, Christensen J, Helin K (2011a) DNA methylation: TET proteins-guardians of CpG islands? EMBO Rep 13:28–35

Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, Helin K (2011b) TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature 473:343–348

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

Wu SC, Zhang Y (2010) Active DNA demethylation: many roads lead to Rome. Nat Rev Mol Cell Biol 11: 607–620

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

Wu H, Zhang Y (2014) Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell 156:45–68

Wu X, Zhang Y (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet 18:517–534

Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y (2011) Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature 473:389–393

Xia B, Han D, Lu X, Sun Z, Zhou A, Yin Q, Zeng H, Liu M, Jiang X, Xie W et al (2015) Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. Nat Methods 12:1047–1050

Xiao M, Yang H, Xu W, Ma S, Lin H, Zhu H, Liu L, Liu Y, Yang C, Xu Y et al (2012) Inhibition of alpha-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. Genes Dev 26:1326–1338

Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J, Dempster EL, Ren B (2012) Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell 148:816–831

Xing X, Sato S, Wong NK, Hidaka K, Sugiyama H, Endo M (2020) Direct observation and analysis of TET-mediated oxidation processes in a DNA origami nanochip. Nucleic Acids Res 48:4041–4051

Xu GL, Walsh CP (2014) Enzymatic DNA oxidation: mechanisms and biological significance. BMB Rep 47:609–618

Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim SH, Ito S, Yang C, Wang P, Xiao MT et al (2011a) Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. Cancer Cell 19:17–30

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

Xu Y, Xu C, Kato A, Tempel W, Abreu JG, Bian C, Hu Y, Hu D, Zhao B, Cerovina T et al (2012) Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. Cell 151:1200–1213

Xu S, Li W, Zhu J, Wang R, Li Z, Xu GL, Ding J (2013) Crystal structures of isoorotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase. Cell Res 23:1296–1309

Xu L, Chen Y-C, Chong J, Fin A, McCoy LS, Xu J, Zhang C, Wang D (2014) Pyrene-based quantitative detection of the 5-formylcytosine loci symmetry in the CpG Duplex content during TET-dependent demethylation. Angewandte Chemie 126:11405–11409

Xu Q, Wang K, Wang L, Zhu Y, Zhou G, Xie D, Yang Q (2016) IDH1/2 mutants inhibit TET-promoted oxidation of RNA 5mC to 5hmC. PLoS One 11:e0161261

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 492:443–447

Yamaguchi S, Shen L, Liu Y, Sendler D, Zhang Y (2013) Role of Tet1 in erasure of genomic imprinting. Nature 504:460–464

Yamazaki Y, Mann MR, Lee SS, Marh J, McCarrey JR, Yanagimachi R, Bartolomei MS (2003) Reprogramming of primordial germ cells begins before migration into the genital ridge, making these cells inadequate donors for reproductive cloning. Proc Natl Acad Sci U S A 100:12207–12212

Yang CG, Yi C, Duguid EM, Sullivan CT, Jian X, Rice PA, He C (2008) Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. Nature 452:961–965

Yang CG, Garcia K, He C (2009) Damage detection and base flipping in direct DNA alkylation repair. Chembiochem 10:417–423

Yang H, Liu Y, Bai F, Zhang JY, Ma SH, Liu J, Xu ZD, Zhu HG, Ling ZQ, Ye D et al (2013) Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. Oncogene 32:663–669

Ye D, Ma S, Xiong Y, Guan KL (2013) R-2-hydroxyglutarate as the key effector of IDH mutations promoting oncogenesis. Cancer Cell 23:274–276

Yildirim O, Li R, Hung JH, Chen PB, Dong X, Ee LS, Weng Z, Rando OJ, Fazzio TG (2011) Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell 147:1498–1510

Yin R, Mao SQ, Zhao B, Chong Z, Yang Y, Zhao C, Zhang D, Huang H, Gao J, Li Z et al (2013) Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. J Am Chem Soc 135:10396–10403

Young JI, Zuchner S, Wang G (2015) Regulation of the epigenome by vitamin C. Annu Rev Nutr 35:545–564

Yu Z, Genest PA, ter Riet B, Sweeney K, DiPaolo C, Kieft R, Christodoulou E, Perrakis A, Simmons JM, Hausinger RP et al (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. Nucleic Acids Res 35:2107–2115

Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B et al (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell 149:1368–1380

Yu H, Su Y, Shin J, Zhong C, Guo JU, Weng YL, Gao F, Geschwind DH, Coppola G, Ming GL et al (2015) Tet3 regulates synaptic transmission and homeostatic plasticity via DNA oxidation and repair. Nat Neurosci 18: 836–843

Yue X, Trifari S, Aijo T, Tsagaratou A, Pastor WA, Zepeda-Martinez JA, Lio CW, Li X, Huang Y, Vijayanand P et al (2016) Control of Foxp3 stability through modulation of TET activity. J Exp Med 213: 377–397

Yue X, Samaniego-Castruita D, Gonzalez-Avalos E, Li X, Barwick BG, Rao A (2021) Whole-genome analysis of TET dioxygenase function in regulatory T cells. EMBO Rep 22:e52716

Zhang L, Lu X, Lu J, Liang H, Dai Q, Xu GL, Luo C, Jiang H, He C (2012) Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nat Chem Biol 8:328–330

Zhang G, Huang H, Liu D, Cheng Y, Liu X, Zhang W, Yin R, Zhang D, Zhang P, Liu J et al (2015a) N6-methyladenine DNA modification in Drosophila. Cell 161:893–906

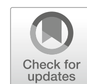
Zhang Q, Zhao K, Shen Q, Han Y, Gu Y, Li X, Zhao D, Liu Y, Wang C, Zhang X et al (2015b) Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. Nature 525:389–393

Zhang HY, Xiong J, Qi BL, Feng YQ, Yuan BF (2016a) The existence of 5-hydroxymethylcytosine and 5-formylcytosine in both DNA and RNA in mammals. Chem Commun (Camb) 52:737–740

Zhang X, Su J, Jeong M, Ko M, Huang Y, Park HJ, Guzman A, Lei Y, Huang YH, Rao A et al (2016b) DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. Nat Genet 48:1014–1023

Zhou T, Xiong J, Wang M, Yang N, Wong J, Zhu B, Xu RM (2014) Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. Mol Cell 54:879–886

Ziller MJ, Muller F, Liao J, Zhang Y, Gu H, Bock C, Boyle P, Epstein CB, Bernstein BE, Lengauer T et al (2011) Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. PLoS Genet 7:e1002389



11 Proteins That Read DNA Methylation

Ke Liu, Takashi Shimbo, Xiaosheng Song, Paul A. Wade, and Jinrong Min

Abstract

Covalent modification of DNA via deposition of a methyl group at the 5′ position on cytosine residues alters the chemical groups available for interaction in the major groove of DNA. This modification, thereby, alters the affinity and specificity of DNA-binding proteins; some of them favor interaction with methylated DNA, and others disfavor it. Molecular recognition of cytosine methylation by proteins often initiates sequential regulatory events that impact gene expression and chromatin structure. The known methyl-DNA-binding proteins have unique domains responsible for DNA methylation recognition: (1) the methyl-CpG-binding domain (MBD), (2) the SET- and RING finger-associated domain (SRA), and (3) some of TF families, such as the C2H2 zinc finger domain, basic helix-loop-helix (bHLH), basic leucine-zipper (bZIP), and homeodomain proteins. Structural analyses have revealed that each domain has a characteristic methylated DNA-binding pattern, and the difference in the recognition mechanisms renders the DNA methylation mark able to transmit complicated biological information. Recent genetic and genomic studies have revealed novel functions of methyl-DNA-binding proteins. These emerging data have also provided glimpses into how methyl-DNA-binding proteins possess unique features and, presumably, functions. In this chapter, we summarize structural and biochemical analyses elucidating the mechanisms for recognition of DNA methylation and correlate this information with emerging genomic and functional data.

Keywords

DNA methylation · MBD · Methylcytosine-binding proteins (MBPs) · Epigenetics

Abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
bHLH	Basic helix-loop-helix
bZIP	Basic leucine-zipper
CTCF	Multidomain CCCTC-binding factor

K. Liu · X. Song · J. Min (✉)
Hubei Key Laboratory of Genetic Regulation and Integrative Biology, School of Life Sciences, Central China Normal University, Wuhan, PR China
e-mail: keliu2015@mail.ccnu.edu.cn; minjinrong@ccnu.edu.cn

T. Shimbo · P. A. Wade
Epigenetics and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, Durham, NC, USA
e-mail: wadep2@niehs.nih.gov

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_11

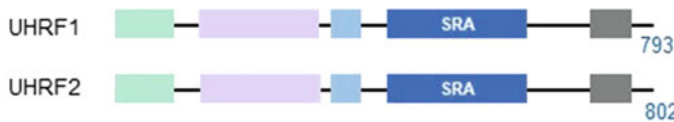
DNMT	DNA methyltransferase
GD	Glycosylase domain
MARs/SARs	Matrix/scaffold attachment regions
MBD	Methyl-CpG-binding domain
NSC	Neural stem cell
NuRD	Nucleosome remodeling deacetylase
PHD	Plant homeodomain
RING	Really interesting new gene domain
SRA	SET- and RING finger-associated domain
TF	Transcriptional factor
TRD	Transcriptional repression domain
TTD	Tandem Tudor domain
UBL	Ubiquitin-like domain
ZF	Zinc finger

11.1 Introduction

DNA methylation serves as a fundamental component of epigenetic regulation; dysregulation of DNA methylation impacts multiple biological processes, including tumorigenesis (Schubeler 2015). In mammals, most DNA methylation occurs in the context of the CpG dinucleotide. In general, 70–80% of the CpGs in mammalian genomes are methylated (Bird 2002). Nevertheless, cytosine methylation is also present at CpH (non-CpG methylation, H=A, T, or C) sites (Ramsahoye et al. 2000; Woodcock et al. 1987), which accounts for about 25% of the total cytosine methylation in both neurons and embryonic stem cells (*ESCs*). Similar to mCpG, mCpH mainly contributes to transcriptional repression and imprinting (Guo et al. 2014; Sanchez-Mut et al. 2016). mCpH is mainly located in a region with low CpG density, and is established and maintained by DNMT3A (Guo et al. 2014; Ramsahoye et al. 2000). Non-CpG methylation, although less abundant than CpG methylation, occurs in virtually all human tissues and is involved in the repression of development-related genes during stem cell differentiation (Schultz et al. 2015).

Although DNA methylation has historically been depicted as a relatively static modification, recent studies have revealed that the methyl group on cytosine can be further modified by oxidation; Fe (II)- and α-ketoglutarate-dependent oxidation mediated by ten-eleven translocation (TET) dioxygenases converts 5-methylcytosine (5mC) into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (Kohli and Zhang 2013; Kriaucionis and Tahiliani 2014; Tahiliani et al. 2009).

The "reader proteins," referred to as methylcytosine-binding proteins (MBPs), specifically recognize DNA methylation marks and initiate signaling pathways. MBPs often interact with other proteins and serve as hubs to recruit effector proteins to particular loci. It is the particular collection of effector proteins associated with each MBP and not the act of binding methylated CpG per se that typically elicits downstream transcriptional effects. The MBPs can be classified using structural information into three major families, each characterized by the presence of a critical recognition domain: the methyl-CpG-binding domain (MBD), the SET- and RING finger-associated domain (SRA), and some of TF families (Fig. 11.1). Although DNA methylation precludes the interaction of many TFs, such as MYC, CREB, NRF1, and members of the E2F family, with their specific DNA recognition sequences (Domcke et al. 2015; Tate and Bird 1993), some TFs, such as the extended homeodomain family, prefer methylated CpG sequences (Kribelbauer et al. 2017; Yin et al. 2017). Each MBP has unique features, including DNA-binding preferences, expression patterns, or protein–protein interaction partners, and has critical roles in various biological contexts. The domain architecture of each protein family is unique, and comparisons between these structures enable insights into the similarities and differences in recognition of methylcytosine, presenting opportunities for a single modification, CpG methylation, to nucleate different effectors (Fig. 11.1).



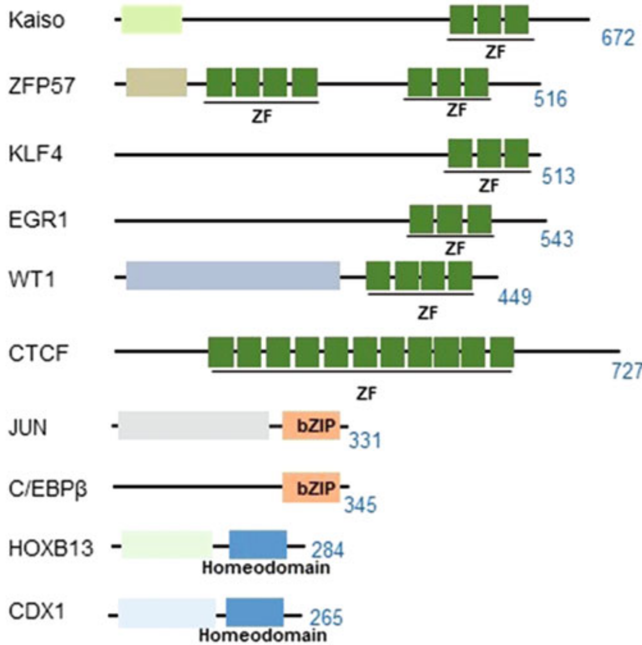


Fig. 11.1 Domain structures of methylated DNA-binding proteins. MeCP2 (NG_007107.2), MBD1 (NP_001191065.1), MBD2 (NP_003918.1), MBD3 (NP_001268382.1), MBD4 (NP_001263199.1), MBD5 (NP_060798.2), MBD6 (NP_060798.2), UHRF1 (NP_001041666.1), UHRF2 (NP_690856.1), Kaiso (NP_001171671.1), ZFP57 (Q9NU63-2), KLF4 (NP_001300981.1), EGR1 (NP_001955.1), WT1 (P19544.2), CTCF (XP_016878357.1), JUN (NP_002219.1), C/EBPβ (NP_005185.2), HOXB13 (NP_006352.2), CDX1 (NP_001795.2), CDX2 (NP_001256.3). *TRD* transcriptional repression domain, *CXXC* CXXC type zinc finger domain, *UBL* ubiquitin-like domain, *PHD* plant homeodomain, *SRA* SET- and RING finger-associated domain, *PWWP* Pro-Trp-Trp-Pro domain, *RING* really interesting new gene finger domain, *TTD* tandem tudor domain, *BTB* BR-C, ttk, and bab domain, *C2H2* C2H2 type zinc finger domain, *KRAB* Kruppel-associated box domain, *bZIP* basic leucine zipper domain, *WT1* Wilms' tumor 1

11.2 The Methyl-CpG-Binding Domain Family

11.2.1 MeCP2

MeCP2 was the first MBP to be purified biochemically and its cDNA cloned and sequenced (Lewis et al. 1992; Meehan et al. 1992; Meehan et al. 1989). The cDNA initially cloned by Bird and colleagues codes for a protein of 492 amino acids that contains an N-terminal MBD domain and a transcriptional repression domain in the C-terminal region. Surprisingly, MeCP2 was subsequently found to be homologous to a matrix attachment binding protein from chicken known as ARBP (attachment region binding protein), a protein identified by biochemical assays based on its binding to a sequence motif (5′-GGTGT-3′) found in matrix/scaffold attachment regions (MARs/SARs) (von Kries et al. 1991; Weitzel et al. 1997). The functional domain responsible for the binding of methylated CpG sites in MeCP2 was subsequently identified and termed the methyl-CpG-binding domain (MBD) that became the archetypal methyl-CpG-binding domain. Subsequent homology searches using the MBD from MeCP2 led to the identification of the remaining MBD family proteins (Hendrich and Bird 1998; Nan et al. 1993). MeCP2 thus represents the founding member of the MBD protein family.

Structure analysis revealed that MeCP2 recognizes the fully methylated CpG dinucleotide using a 5mC-Arg-Gua triad (Fig. 11.2a). Two arginine residues within the MBD (R111 and R133) each bind to a guanine with bidentate hydrogen bonds and to the 5mC with cation–π interactions, where cytosine methylation expands the aromatic ring structure and strengthens the cation–π interactions between the methylcytosine and the guanidinium group of the arginine residues (Ho et al. 2008; Lei et al. 2019; Liu et al. 2013, 2018; Zou et al. 2012). In addition, a tyrosine residue forms water-mediated hydrogen bonds with one of the two cytosine methyl groups. Although the SELEX experiments refined the model, stipulating that high-affinity interaction with methylated DNA was facilitated if the methylated CpG dinucleotide was flanked by A/T base pairs on each side (Ghosh et al. 2010; Klose et al. 2005), structural studies by different laboratories showed that the specific DNA recognition is largely confined to the mCpG dinucleotide, and no base-specific interaction was observed outside the mCpG dinucleotide (Fig. 11.2a) (Ho et al. 2008; Liu et al. 2018; Ohki et al. 2001; Otani et al. 2013). In addition, structural analysis and binding studies also confirmed that MeCP2 recognizes TpG DNA with a preference for GTG DNA (Fig. 11.2b) (Lei et al. 2019; Liu et al. 2018). The MeCP2 MBD binds to mCAC or CAC DNA by recognizing their complementary GTG trinucleotide, which explains why MeCP2 has a comparable binding affinity to both mCAC and hmCAC (hydroxymethylated CAC) DNAs (Fig. 11.2b) (Kinde et al. 2015; Lei et al. 2019). The above findings are also consistent with the original finding that chicken ARBP (or cMeCP2) binds to a conserved GGTGT DNA motif found in the MARs/SARs DNA regions that lacks any methylated CpG dinucleotides (Buhrmester et al. 1995; von Kries et al. 1991; Weitzel et al. 1997). The central GTG trinucleotide of the MARs DNA elements is crucial for its binding to cMeCP2, and mutating either of the guanine bases in this trinucleotide significantly reduced this binding (Buhrmester et al. 1995; Weitzel et al. 1997).

MeCP2 displays relatively high expression in neurons, where this level is approximate to that of histone octamers (Skene et al. 2010). MeCP2 plays important role in the normal chromatin architecture in neurons, and its mutations result in the neurodevelopmental disorder Rett syndrome (RTT) in humans (Amir et al. 1999; Chen et al. 2001; Guy et al. 2001). However, despite the important roles of MeCP2 mutations in development and disease, MeCP2 deletion in mice has only minimal impact on global gene regulation. Thus, detailed mechanistic insights into how disruption of MeCP2 causes developmental failure or Rett syndrome are currently lacking. MeCP2 has been found to bind to methylated DNA at CpG and CpA dinucleotides broadly throughout the genome (Gabel et al.

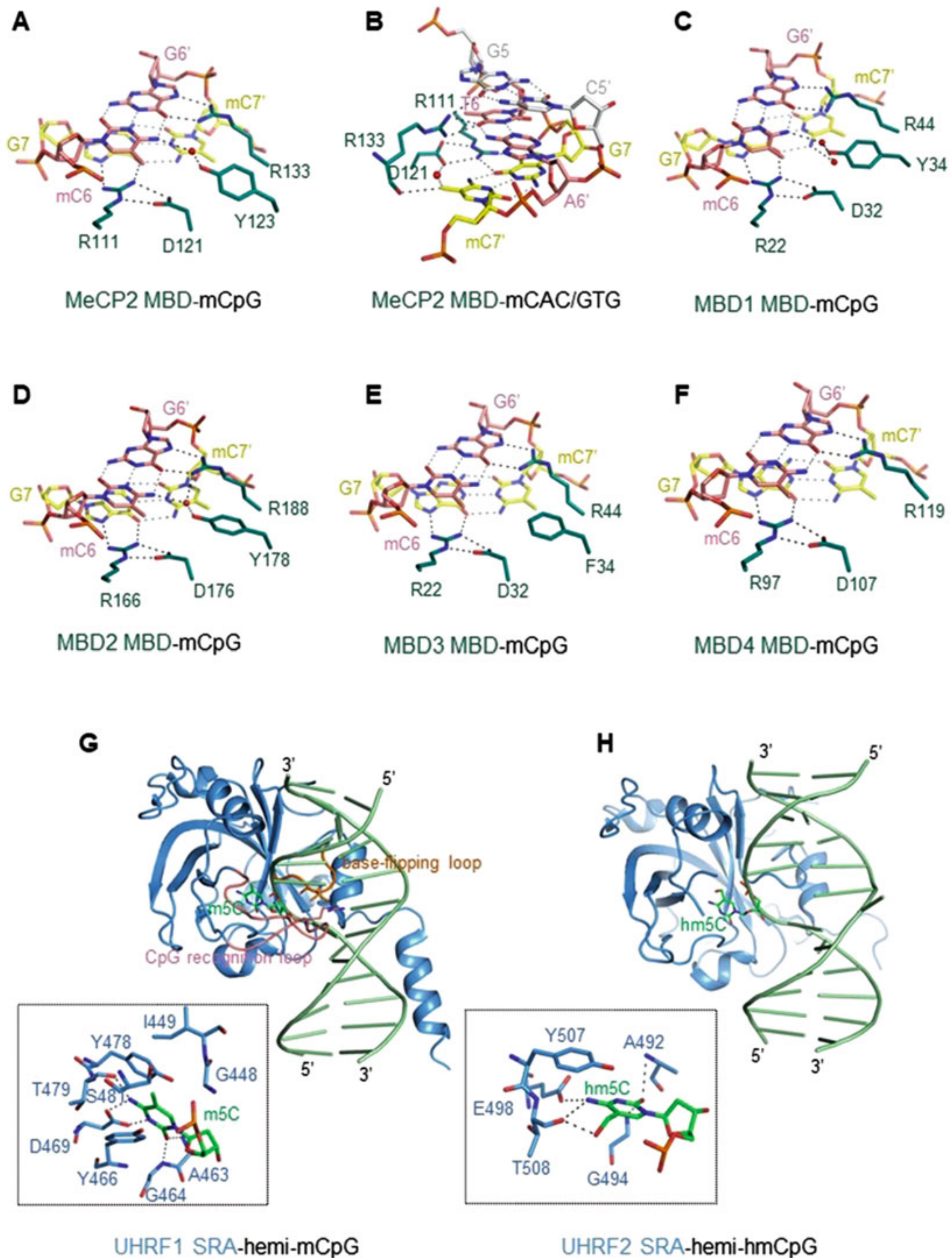


Fig. 11.2 Structures of human MBDs and SRA domains bound to methylated DNA. (**a**) and (**b**) Structures of MeCP2-MBD bound to mCpG (PDB: 6C1Y) and mCAC DNA (PDB: 6OGK), respectively. (**c**–**f**) Structures of MBDs from MBD1-4 in complex, respectively, with mCpG DNA (PDB: 6D1T, 6CNQ, 6CC8, and 4LG7). The protein residues are shown as green sticks, while nucleotides involved in base interactions are shown in sticks and colored in gray (G5-C5′), red (mC6-G6′ or T6-A6′), and yellow (G7-mC7′ or G7-C7′). The hydrogen bonds formed between residues and DNA bases are marked as black dashed lines; the DNA base pair interactions are shown as gray dashed lines. (**g**) Structure of UHRF1 SRA domain bound to the flipping-out 5mC of

2015; Kinde et al. 2016; Skene et al. 2010), which also complicates the connection between the MeCP2 binding and specific genes repression. In addition, studies in human neuronal SH-SY5Y cells (Yasui et al. 2007) or mouse hypothalamus (Chahrour et al. 2008) revealed that most genes bound by MeCP2 at their sparsely methylated promoters are actively expressed. In addition to the N-terminal MBD, the transcriptional repression domain of MeCP2 interacts with DNMT3A and inhibits its activity in vitro (Rajavelu et al. 2018). Moreover, recent structural revelation of the MeCP2 MBD bound to methylated and unmethylated CAC-containing DNA in a similar binding pattern might imply that MeCP2 also functions as a transcription activator and its binding to GTG DNA could provide disease implications in RTT syndrome (Lei et al. 2019; Liu et al. 2018).

11.2.2 MBD1

MBD1, like other MBD proteins, was initially discovered in homology screens using the MeCP2 MBD as a template. MBD1 is distinctive among the MBD proteins in that, in addition to the MBD, the protein has either two or three CXXC zinc finger domains resulting from alternative splicing (Fujita et al. 1999). Like MeCP2, MBD1 also contains a transcriptional repression domain (TRD) near its C-terminus (Fig. 11.1). The structures of MBD1 MBD in complex with methylated DNA revealed a conserved mCpG binding mode with that of MeCP2 (Fig. 11.2c) (Liu et al. 2018; Ohki et al. 2001). Although biochemical analysis previously indicated that nucleotides flanking the methylated CpG dinucleotide exert an influence on binding affinity of the MBD1 MBD (Clouaire et al. 2010), the complex structure of the MBD1 MBD with mCpG DNA does not support DNA base selectivity outside the mCpG dinucleotide (Liu et al. 2018).

In addition to the MBD, the CXXC domains of MBD1 add an additional DNA-binding interface. The CXXC domain selectively recognizes unmethylated DNA sites (Lee and Skalnik 2005; Liu and Min 2019; Xu et al. 2011, 2018). MBD1 isoform with all three CXXC domains can repress genes regardless of their promoter methylation status, while the MBD1 lacking the third CXXC domain (CXXC3) can only suppress gene expression when the promoter is methylated, suggesting that CXXC3 is essential for binding to unmethylated templates (Fujita et al. 2000; Jorgensen et al. 2004). Consistently, structure analysis and binding assays revealed that the CXXC3 domain of MBD1 specifically recognizes unmethylated CpG sites and cytosine methylation abrogates this binding; however, the first two CXXC domains of MBD1 (CXXC1 and CXXC2) do not bind CpG DNA (Jorgensen et al. 2004; Xu et al. 2018). This finding also explains why the increase of heterochromatin localization of TET1 and oxidation of 5mC require the CXXC3 domain-containing MBD1 (Zhang et al. 2017). In ESCs, biotin-tagged MBD1 is enriched at highly methylated regions, and this enrichment was lost after the depletion of DNA methylation. In addition, targeting of MBD1 to unmethylated DNA was observed only when the MBD was deleted, suggesting that the recruiting mechanism of MBD1 is dominated by the MBD–methyl-CpG interaction (Baubec et al. 2013).

Fig. 11.2 (continued) the hemi-methylated DNA (PDB: 3CLZ). The two DNA binding loops are colored in pink and orange, respectively. The interaction details of 5mC recognized by the unique binding pocket of UHRF1 are shown in the left illustration. (**h**) Structure of UHRF2 SRA domain in complex with flipping-out 5-hydroxymethylcytosine of DNA (PDB: 4PW6). The interaction details of 5hmC recognized by the unique binding pocket of UHRF2 are shown in the left illustration. The protein residues and modified cytosines are shown as sticks and colored in blue and green, respectively. The black dashed lines represent the hydrogen bonds between protein residues and base pairs

MBD1 is involved in neurodevelopment. MBD1 maintains the multipotency of neural stem cells (NSCs) by repressing neural cell differentiation-related genes (Jobe et al. 2017). Mutations or polymorphisms of MBD1 are associated with autism spectrum disorder (ASD), and also result in the accumulation of undifferentiated NSCs, impaired neurogenesis, and learning deficits in mice (Cukier et al. 2010; Jobe et al. 2017; Li et al. 2005).

11.2.3 MBD2

MBD2 is a member of the chromatin remodeling complex, nucleosome remodeling deacetylase (NuRD), which functions as a repressor connecting DNA methylation with histone deacetylation (Feng and Zhang 2001). Like MeCP2 and MBD1, MBD2 contains a TRD in addition to an MBD. MBD2 also has a glycine-arginine repeat and a coiled-coil domain (Fig. 11.1), which is essential for binding to the Mi-2/NuRD complex (Gnanapragasam et al. 2011). MBD2 was also reported to be an integral component of the MeCP1 complex, which was subsequently shown to have biochemical similarities to NuRD (Ng et al. 1999; Zhang et al. 1999). Single-molecule fluorescence techniques revealed that MBD2 spreads more quickly in the regions of CpG-rich sequences than that of CpG-free DNA, while MBD2 binding is static or with a slow exchange when interacting with DNA regions enriched for mCpG. It was thought that MBD2 facilitates the rapid movement and nucleosome remodeling of the NuRD complex in the CpG-rich regions, and this movement is limited when those loci are methylated, which further contributes to gene silencing (Pan et al. 2017).

The interaction of the MBD2 MBD with mCpG DNA was first defined for chicken MBD2 (which is >95% identical to human MBD2) by Williams and colleagues (Scarsdale et al. 2011). Subsequently, two human MBD2–mCpG complex structures were determined (Liu et al. 2018). Similar to MeCP2 and MBD1, base-specific contacts with the methylated CpG palindrome are mediated by a pair of arginine residues (R166 and R188), as well as a tyrosine residue (Y176) in MBD2 (Fig. 11.2d). Although MeCP2 and MBD2 MBDs exhibit a conserved binding mode, knock-in mice expressing a chimeric protein (MM2) by swapping the MBDs of MeCP2 and MBD2 exhibit severe phenotypic features that are largely similar to those seen in mouse models of RTT, suggesting that the conserved MBD of MBD2 could not functionally replace that of MeCP2 (Tillotson et al. 2021).

MBD2 is expressed in most somatic cells and is particularly abundant in ESCs in mice (Hendrich and Bird 1998). Despite this intriguing expression pattern, MBD2 knockout mice are viable and fertile (Hendrich et al. 2001). MBD2 has two predominant isoforms, MBD2a and MBD2c; MBD2c lacks the carboxyl-terminal region including the coiled-coil domain integral to the interaction with the NuRD complex (Hendrich and Bird 1998). Differential expression of MBD2a and MBD2c was shown in human pluripotent stem cells (hPSCs); MBD2c is dominant in hPSCs, while MBD2a is dominant in fibroblasts (Lu et al. 2014). Interestingly, MBD2a, but not MBD2c, can interact with the NuRD complex and promote differentiation, while MBD2c enhances reprogramming efficiency when overexpressed in fibroblasts.

11.2.4 MBD3

As predicted from its high sequence similarity with MBD2 (Fig. 11.1), MBD3 is also a member of the NuRD complex (Le Guezennec et al. 2006). MBD3 from mammals has been reported to lack the capacity for high-affinity interaction with methylated DNA in conventional biochemical assays (Fraga et al. 2003; Hendrich and Bird 1998). In contrast, the amphibian protein displays a strong preference for methylated substrates (Wade et al. 1999). Close inspection of the amino acid sequence of MBD3 MBDs from multiple species reveals that mammals differ from Amphibia, fish, reptiles, and birds at a critical position encoding the conserved tyrosine residue involved in specific contacts with the methylated

cytosine (changed to phenylalanine in mammals). Recently, crystal structures of human MBD3 MBD bound to mCpG DNA were determined (Liu et al. 2019). Similar to MBD2, the MBD3 MBD binds to mCpG via two conserved arginine fingers (Fig. 11.2e). By structural comparison to that of MBD2, the replacement of tyrosine by phenylalanine at F34 of MBD3 results in weaker mCpG DNA binding compared to MBD2, due to the loss of a solvent-mediated interaction of the hydroxyl group in tyrosine with the N4-amino group of the methylcytosine (Liu et al. 2019). The complex structure of the MBD3 MBD bound to a non-palindromic DNA also revealed that the MBDs recognize the mCpG DNA without orientation preference (Liu et al. 2019), consistent with the observations that the sequences flanking the mCpG dinucleotide do not influence the mCpG DNA binding significantly. Therefore, MBD3 is able to bind to methylated CpG DNA, albeit with a reduced binding affinity (Gunther et al. 2013; Liu et al. 2019).

MBD2 and MBD3 form mutually exclusive complexes (Le Guezennec et al. 2006). In contrast to the mild phenotypes of the MBD2 knockout mice, MBD3 deletion causes early embryonic lethality. Furthermore, MBD3-null ES cells can maintain stemness, even in the absence of leukemia inhibitory factor (LIF) (Hendrich et al. 2001; Kaji et al. 2006). These striking phenotypes suggest that MBD2 and MBD3 have nonredundant roles. Importantly, deletion of MBD3 can significantly enhance reprogramming efficiency, suggesting that MBD3 functions as a barrier to reprogramming (Luo et al. 2013; Rais et al. 2013). However, the field is not in complete agreement with the role of MBD3 in reprogramming, suggesting a possibility that the role of MBD3 is highly context-dependent (Dos Santos et al. 2014). Moreover, in uterine serous carcinoma patients, a small segment of chromosome 19 containing MBD3 is frequently deleted, suggesting a critical role of MBD3 in tumorigenesis or tumor progression (Zhao et al. 2013). Despite the accumulation of evidence on the critical function of MBD3 in biology, the detailed molecular mechanism of how MBD3, or MBD3/NuRD, regulates gene expression and chromatin structure is still unclear. Most cells express a splice variant of MBD3 that disrupts the canonical MBD sequence (Hendrich and Tweedie 2003). Genomic localization analyses revealed that MBD3 preferentially localizes at unmethylated CpG-rich regions, including CpG islands, while MBD2 distributes across the genome in a methylation-dependent manner (Baubec et al. 2013).

11.2.5 MBD4

MBD4 contains two functional domains, an N-terminal MBD and a C-terminal glycosylase domain (GD), separated by a long link of unknown function (Fig. 11.1). The presence of a glycosylase domain makes MBD4 a unique member of the MBD family. The MBD4 MBD has high affinity for methylated CpG-containing substrates. It has a similar affinity for the deamination product of that substrate, i.e., methylated CpG base-paired (mismatch) with TpG (Hendrich et al. 1999; Otani et al. 2013). Not surprisingly, MBD4 possesses enzymatic activity that can repair mCpG/TpG or mCpG/hmUpG double-stranded mismatches generated by spontaneous deamination of 5mC (Hendrich and Bird 1998; Hendrich et al. 1999). MBD4 knockout mice are viable and fertile with minor phenotypes, including a slight increase in C to T mutations at CpG sites (Millar et al. 2002; Wong et al. 2002). Although the deletion of MBD4 itself does not impact tumorigenesis, it increases tumor incidence in a susceptible genetic background (mutation in the adenomatous polyposis coli (APC) gene). In addition, mutations in MBD4 have been observed in human colorectal tumors with microsatellite instability (Riccio et al. 1999). Taken together, these observations suggest that MBD4 plays an important role in tumor progression by regulating DNA mismatch repair.

Similar to MeCP2 and other MBD family members, the crystal structures of the MBD4 MBD with different DNA sequences revealed that MBD4 recognizes mCpG or mismatched DNA by the 5mC (or T)-Arg-Gua triad interactions (Fig. 11.2f) (Liu et al. 2018; Otani

et al. 2013; Wu et al. 2021). While there are some minor alterations, the overall structure of the MBD4 MBD bound to a mismatched DNA is highly similar to that of the MBD4 MBD bound to symmetric methylated CpG DNA. In contrast to the MBD, the glycosylase domain of MBD4 binds to DNA containing a G:X mismatch in a very different manner (Hashimoto et al. 2012). The target nucleotide is flipped out from the DNA strand and an arginine residue from the MBD fills that space. The flipped base is associated with the active-site cleft. Importantly, the crystal structure of the full-length MBD4 containing both the MBD and GD is yet to be solved, so it is unclear to what level the two domains communicate. This question has been partially approached via analysis of solution structures; although MBD4 shows a slow exchange rate between different DNA molecules (intermolecular exchange), it has rapid exchange rate between the two binding sites on the same dsDNA molecule (intramolecular exchange) (Walavalkar et al. 2014). These data support a local hopping model in which the MBD of MBD4 rapidly scans multiple methylated CpG sites and supports the mismatch repair conducted by the GD (Walavalkar et al. 2014).

11.2.6 MBD5 and MBD6

MBD5 and MBD6 are characterized to be associated with neurodevelopmental disorders (Cukier et al. 2012; Talkowski et al. 2012). MBD5-null mice develop growth defects and preweaning lethality, exhibiting several phenotypic features seen in patients with 2q23.1 microdeletion (Du et al. 2012). Like other MBD proteins, MBD5 and MBD6 contain an MBD that is required for their heterochromatin localization (Laget et al. 2010). However, electrophoretic mobility shift assay (EMSA) experiments using purified MBDs from MBD5 and MBD6 indicated that these domains display no methylcytosine-binding capacity. This property may be explained by the loss of a characteristic loop structure and the conserved arginine fingers that contribute to DNA base-specific binding, which is critical for methylated mCpG DNA contacts in the MBDs (Hendrich and Tweedie 2003; Liu et al. 2018). Interestingly, although the incomplete MBDs of these proteins have lost the methylcytosine-binding affinity, they can interact with mammalian PR-DUB polycomb protein complex, which is known as a histone H2A deubiquitinase (Baymaz et al. 2014). These distinct differences from other MBD proteins may assign a specialized function to MBD5 and MBD6.

11.3 SET- and RING-Associated (SRA) Domain

11.3.1 UHRF1

UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1, also known as ICBP90 or Np95) contains a ubiquitin-like domain (UBL), tandem Tudor domain (TTD), plant homeodomain (PHD), SET- and RING-associated (SRA) domain, and really interesting new gene (RING) domain, which interdependently coordinate epigenetic functions of UHRF1 (Fig. 11.1). UHRF1 was originally identified as a potential regulator of topoisomerase IIα (Hopfner et al. 2000). UHRF1 is essential for the maintenance of proper DNA methylation levels by recruiting DNA methyltransferase 1 (DNMT1) to replication foci (Fang et al. 2016; Li et al. 2018). Deletion of UHRF1 in mice causes genome-wide DNA hypomethylation and results in embryonic lethality, presumably due to the dysfunction of DNMT1. UHRF1 was also observed to contribute to DNA damage repair by binding directly to interstrand crosslink (ICL)-containing DNA and facilitating the recruitment of other DNA repair factors (Liang et al. 2015; Mancini et al. 2021; Tian et al. 2015).

UHRF1 recognizes hemi-methylated DNA with its SRA domain employing the same base-flipping mechanism, which is commonly found in DNA methyltransferases (Arita et al. 2008; Avvakumov et al. 2008; Hashimoto et al. 2008; Song et al. 2012). The SRA domain uses the CpG recognition loop and the base-flipping loop approach the major groove and minor groove of

DNA, respectively. The flipped-out methylcytosine of the duplex DNA is stabilized in a binding pocket with van der Waals interactions, planar stacking contacts, and Watson–Crick polar hydrogen bonds (Fig. 11.2g). The UHRF1 SRA-DNA complex structure showed that cytosine methylation from the complementary strand interferes with the conformation of a conserved asparagine, e.g., human N489, which explains why UHRF1 SRA prefers binding to hemi-methylated DNA rather than symmetric mCpG DNA (Arita et al. 2008; Avvakumov et al. 2008; Hashimoto et al. 2008). The use of this base-flipping mechanism uniquely positions UHRF proteins in the MBP family; the SRA domain was the first domain that conducts base flipping without enzymatic activity (Song et al. 2012).

In addition to hemi-methylated DNA, UHRF1 also recognizes histone modifications, such as H3K9me3, unmodified H3K9, and H3R2 through its histone reader domains (Hu et al. 2011; Nady et al. 2011; Rajakumara et al. 2011). This interaction is allosterically regulated by phosphatidylinositol 5-phospate (PI5P), which alters the local structure around the Tudor and PHD domains (Gelato et al. 2014). Interestingly, it has been demonstrated that the interaction between UHRF1 and H3K9me3 is essential for the maintenance of DNA methylation (Rothbart et al. 2012). Furthermore, UHRF1 binding to hemi-methylated CpGs activates the ubiquitylation activity of the UHRF1 RING domain toward H3K18 and/or K23 adjacent to the H3 binding site of UHRF1, suggesting a role for UHRF1 as a molecular hub connecting DNA methylation and histone modifications (Harrison et al. 2016).

11.3.2 UHRF2

UHRF2 (also known as NIRF) has the same domain structure and high sequence similarity as its paralog UHRF1 (Fig. 11.1). They are the only two proteins with an SRA domain in humans (Mori et al. 2002). Based on their sequence similarity, UHRF2 and UHRF1 appear to share the same functions; UHRF2 also recognizes hemi-methylated DNA and interacts with DNMT1 (Zhang et al. 2011). However, there are critical differences between UHRF1 and UHRF2. Most importantly, unlike UHRF1, the SRA domain of UHRF2 creates a larger pocket to specifically bind to 5hmC by base flipping (Fig. 11.2h) (Zhou et al. 2014). In addition, UHRF2 and UHRF1 are differentially expressed; UHRF2 is downregulated in ESCs and gradually upregulated upon differentiation, whereas UHRF1 shows an opposite pattern (Pichler et al. 2011). Moreover, the introduction of UHRF2 into UHRF1-null ESCs cannot rescue the hypomethylation phenotype, suggesting a differential functionality of UHRF2, at least in ESCs (Zhang et al. 2011). The 5hmC binding can stimulate the E3 ligase activity of the UHRF2 RING domain to regulate the K33-linked polyubiquitination of the BER component XRCC1 (Liu et al. 2021). However, unlike UHRF1, UHRF2 could not ubiquitinate histones in the context of nucleosomes, implying that fragments outside the chromatin binding region also contribute to UHRF1 location in a productive conformation for nucleosomal histone ubiquitination (Vaughan et al. 2018). With these similarities and differences, it is yet unclear how UHRF1 and UHRF2 cooperatively (or distinctively) function in cells; analysis with UHRF2 knockout mice would provide more information. UHRF2 knockout mice exhibit frequent spontaneous seizures and abnormal electrical activities during adulthood. In addition, UHRF2 knockout mice only display a decreased 5mC level at certain genomic loci in brains. Therefore, UHRF2 might play a unique role differing from that of its paralog UHRF1 in the maintenance of 5mC levels (Liu et al. 2017).

11.4 Transcription Factors

In addition to the canonical MBD and SRA domains, an increasing number of transcriptional regulators have been identified to instruct downstream events depending on the recognition of different cytosine modification states, including

the members of C2H2 zinc finger (ZF) proteins, basic helix-loop-helix (bHLH), basic leucine-zipper (bZIP), and homeodomain transcriptional factor families.

11.4.1 Kaiso and ZBTB38

The C2H2 zinc finger is one of the most abundant DNA binding motifs. Each C2H2 zinc finger contains a ββα-fold core comprised of two β-strands packing against an α-helix, and this core is stabilized by a tetrahedrally coordinated Zn^{2+} with two cysteines and two histidine residues (Klug 2010). Being a member of the BTB/POZ (broad complex, Tramtrack, and bric-a-brac/poxvirus and zinc finger) family, Kaiso (also known as ZBTB33) contains three C2H2 zinc fingers and was originally identified as a binding partner of p120 catenin (Daniel and Reynolds 1999). In addition to p120 catenin, Kaiso also interacts with a repression complex, N-CoR (nuclear receptor corepressor), and suppresses the expression of MTA2, a member of the NuRD complex, in a methylation-dependent manner (Yoon et al. 2003). As the N-CoR complex contains histone deacetylases, the recruitment of the N-CoR complex mediated by Kaiso is proposed as a potential mechanism of DNA methylation-dependent gene repression. However, Kaiso is also known to associate with p53 and upregulate apoptosis-related genes, suggesting pleiotropic roles of Kaiso in different biological contexts (Koh et al. 2015; Koh et al. 2014). Moreover, Kasio showed both pro- and antitumorigenic activities, which also implies that Kaiso is a context-dependent regulator (Koh et al. 2014; Prokhortchouk et al. 2006; Soubry et al. 2010).

Kaiso preferentially binds to two consecutively methylated CpG dinucleotides or to a chemically similar, albeit unmethylated, TpG-containing sites, TCCTGCCA (also called the Kaiso binding sequence, KBS) (Daniel et al. 2002; Prokhortchouk et al. 2001). The crystal structures of Kaiso have been solved in complex with two different DNA templates: a methylated template, MeECad (promoter region of E-cadherin) containing two methylated CpG dinucleotides, and an unmethylated sequence TCCTGCCA (Buck-Koehntop et al. 2012; Daniel et al. 2002). The two structures are almost identical; Kaiso recognizes the methyl group, either of mCpG or TpG dinucleotides, using the 5mC-Arg-Gua triad structure (Fig. 11.3a) (Daniel et al. 2002). The first two zinc fingers hold the major groove of DNA, and the third zinc finger (together with the C-terminal extension) enables high-affinity binding (Fig. 11.3a) (Buck-Koehntop et al. 2012). A recent study suggests that E535 of Kaiso adopts different conformations to determine the distinct recognition of methylated and KBS motifs by Kaiso (Fig. 11.3a) (Nikolova et al. 2020). Recently, Kaiso was found to immunoprecipitate with the de novo DNA methyltransferases DNMT3A/3B, suggesting that Kaiso may recruit the DNA methyltransferases to modulate genome methylation apart from being a methyl-DNA-binding protein (Kaplun et al. 2021).

In addition to Kasio, ZBTB4 and ZBTB38 were also found to bind methylated DNA in vitro and in vivo. ZBTB4 and ZBTB38 specifically bind to the methylated allele of imprinting gene H19/Igf2 and become delocalized with loss of DNA methylation (Filion et al. 2006). ZBTB38 is involved in cellular proliferation, apoptosis, and genomic stability through modulating transcriptional activity (Miotto et al. 2014; Nishii et al. 2012; Oikawa et al. 2008; Pozner et al. 2018). ZBTB38 specifically binds to a DNA sequence ((A/G)TmCG(G/A)(mC/T)(G/A)) through its C-terminal ZF6-10 (Pozner et al. 2018). The crystal structure of ZBTB38 in complex with a DNA sequence of ATmCGGmCG revealed that ZF7 and ZF8 contribute to base-specific DNA interactions using a 5mC-Arg-Gua triad, while ZF6 and ZF9 mainly stabilize ZF7 and ZF8 to form base-specific interactions (Hudson et al. 2018).

11.4.2 CTCF

The multidomain CCCTC-binding factor (CTCF) is crucial for chromatin architecture organization

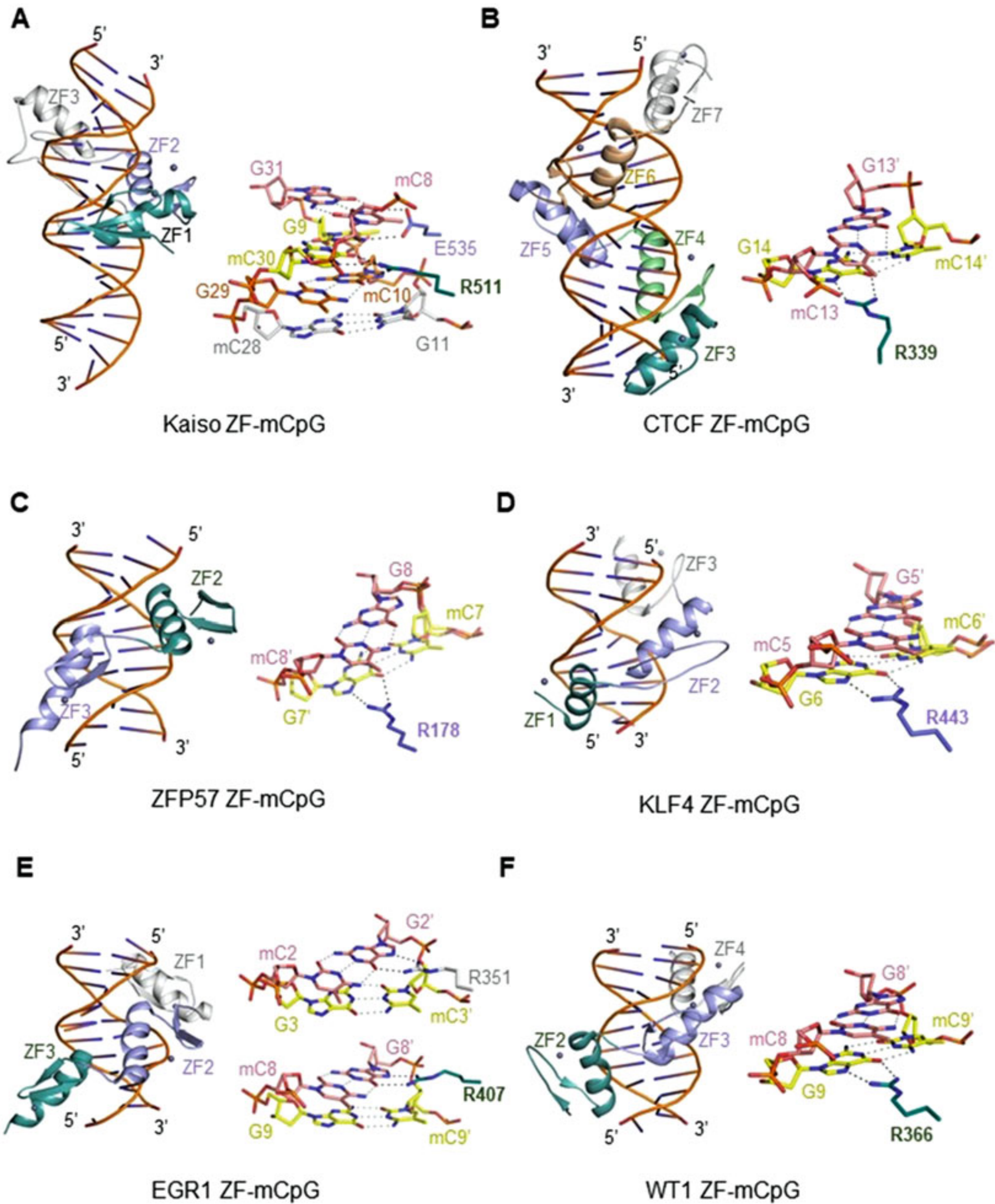


Fig. 11.3 Structural basis of C2H2 ZFs binding to methylated CpG DNA. (**a**) Structure of Kaiso ZF1-3 in complex with DNA containing two consecutively methylated CpG sites (PDB: 4F6N). (**b**) Structure of CTCF ZF3-7 in complex with methylated CpG DNA (PDB: 5T00). (**c**) Structure of ZFP57 ZF2-3 in complex with methylated CpG DNA (PDB: 4GZN). (**d**) Structure of KLF4 ZF1-3 in complex with methylated CpG DNA (PDB: 4M9E). (**e**) Structure of EGR1 ZF1-3 in complex with methylated CpG DNA (PDB: 4X9J). (**f**) Structure of WT1 ZF2-4 in complex with methylated CpG DNA (PDB: 4R2E). The zinc ions are shown as gray balls. The hydrogen bonds formed between protein residues and DNA bases are marked as black dashed lines, while the DNA base pair interactions are shown as gray dashed lines

and gene expression regulation. CTCF contains 11 tandem C2H2 zinc fingers and binds to a 15-base pair consensus sequence NCANNAG(G/A)NGGC(G/A)(C/G)(T/C) (N=A, C, G or T) that can be methylated on cytosines at positions 2 and 12 (C2 and C12) (Nakahashi et al. 2013; Rhee and Pugh 2011; Wang et al. 2012). The crystal structures of CTCF bound to unmethylated and methylated CpG DNA have been determined (Hashimoto et al. 2017), which showed that ZF3-7 recognize the major groove of DNA, and each ZF interacts with three adjacent DNA base pairs (also called the "triplet" element) (Fig. 11.3b) (Choo and Klug 1997). ZF8 and ZF9 span along the DNA phosphate backbone of the 15-base pair core sequence, and no contact is found for ZF1 and ZF10-11. In the complex structure with the methylated CpG DNA, ZF3 and ZF4 contribute to mCpG binding at position 12 using a 5mC-Arg-G triad (Fig. 11.3b). In contrast, methylation at position 2 cytosine forms a steric clash with D451 of CTCF (Hashimoto et al. 2017), explaining why methylation at C2 significantly abolishes DNA binding, whereas methylation at C12 increases DNA binding of CTCF. Thus, CTCF binds to methylated DNA and recognizes 5mC in a position-dependent manner (Hashimoto et al. 2017).

11.4.3 ZFP57

ZFP57, as a member of the KRAB-ZFP (Kruppel-associated box zinc finger) family, is a maternal-zygotic effect gene for maintaining DNA methylation memory in early mouse embryos and embryonic stem cells. ZFP57, together with its binding cofactor KAP1, binds to methylated hexanucleotides within CpG-rich sequences located in imprinting control regions (ICRs) and regulates imprinted genes (Li et al. 2008; Quenneville et al. 2011). In addition, ZFP57 and KAP1 coimmunoprecipitate with UHRF1 and DNA methyltransferases DNMT1, DNMT 3A, and DNMT 3B to maintain the DNA methylation at ICRs (Quenneville et al. 2011; Zuo et al. 2012). Loss-of-function mutations in human ZFP57 are associated with a global imprinting disorder (ID) and transient neonatal diabetes (TND) (Mackay et al. 2008).

ZFP57 preferentially binds to a methylated hexanucleotide sequence (TGCmCGC) using two classical C2H2 domains (Liu et al. 2012; Quenneville et al. 2011). Like MBD proteins and Kaiso, ZFP57 uses a 5mC-Arg-Gua triad to recognize the methylated cytosine (Fig. 11.3c). In addition to TGCmCGC, the ZFP57 DNA binding domain interacts with the oligonucleotide containing the sequence GGCmCGC in vivo and in vitro, albeit weaker than that of TGCmCGC sequence (Anvar et al. 2016).

11.4.4 KLF4

KLF4 is a member of the Kruppel-like protein family and contains three standard Kruppel-like zinc fingers. KLF4 is well known as one of the Yamanaka factors for reprogramming somatic cells to induce pluripotency (Takahashi and Yamanaka 2006). KLF4 recruits the histone H3K27me3 demethylase JMJD3 to reduce H3K27me3 levels at both enhancers and promoters of epithelial and pluripotency genes, facilitating reprogramming somatic cells to pluripotency (Huang et al. 2020). In addition, KLF4 also promotes somatic cell reprogramming by cooperating with transcription factors OCT4 and SOX2 (Chronis et al. 2017; Takahashi and Yamanaka 2006). OCT4 acts as a pioneer factor that opens heterochromatin and facilitates the binding of KLF4 (Chen et al. 2020). In addition to reprogramming, KLF4 is required for normal skin and colon development, and KLF4 knockout mice die soon after birth (Katz et al. 2002; Segre et al. 1999). The transcriptional activities of KLF4 have also been implicated in regulating genomic stability (El-Karim et al. 2013), cellular proliferation (Chen et al. 2003), DNA damage response, and apoptosis (Yoon et al. 2005).

KLF4 exhibits comparable binding activity to CpG, mCpG, or TpG containing DNA in vitro (Hashimoto et al. 2016; Liu et al. 2014). Interestingly, genome-wide studies showed that about half of the KLF4-binding sites in vivo are highly methylated (Hu et al. 2013). KLF4 binds to a

consensus sequence of GG(T/C)G with a preference for the methylated status (Sharma et al. 2021; Wan et al. 2017). Structural analysis revealed that the tandem C2H2 zinc fingers of KLF4 recognize the major groove of the DNA using a 5mC-Arg-Gua triad (Fig. 11.3d) (Liu et al. 2014; Schuetz et al. 2011). KLF4 progressively loses binding affinity as 5mC is oxidized into 5hmC, 5fC, and 5caC (Liu et al. 2014).

11.4.5 EGR1 and WT1

EGR1 (early growth response protein 1), also called ZIF-268, NGFI-A, KROX 24, or ZENK, is a member of the EGR protein family. EGR1 is involved in several biological processes, including cell proliferation, differentiation, inflammation, and apoptosis (Beckmann and Wilce 1997; Bozon et al. 2003; Duclot and Kabbaj 2017; Lee et al. 2004; Sanchez-Guerrero et al. 2013; Veyrac et al. 2014). EGR1 specifically recognizes and binds target genes using three C2H2 zinc fingers, which either promotes or inhibits the expression of target genes (Kim et al. 2011). The WT1 (Wilms' tumor 1) is a predisposition gene for Wilms' tumor, a pediatric kidney cancer (Call et al. 1990; Charlton and Pritchard-Jones 2016; Gessler et al. 1990). WT1 is involved in the regulation of BMP/pSMAD and FGF pathways in the early kidney anlagen (Motamedi et al. 2014), and also plays critical roles in tissue homeostasis, development, and disease (Hastie 2017). Although there are more than 35 potential mammalian WT1 isoforms generated by splicing or alternative translation start sites, all isoforms contain four C2H2 zinc fingers (Hastie 2017). Like EGR1, WT1 can either repress or activate specific target genes depending on its binding partners (Hastie 2017). EGR1 and WT1 have highly divergent cellular functions, but both TFs recognize the same consensus sequence, GCG (T/G)GGGCG (Hartwig et al. 2010; Pavletich and Pabo 1991; Rauscher et al. 1990; Stoll et al. 2007; Zandarashvili et al. 2015). Consistent with other MBPs, these two proteins interact with a methylated DNA template using a 5mC-Arg-Gua triad (Fig. 11.3e, f) (Hashimoto et al. 2014).

11.4.6 bZIP

Activator protein 1 (AP-1) is a transcription factor family with basic leucine zipper (bZIP), including c-Fos, c-Jun, CREB, C/EBP, and ATF (activating transcription factor), and is involved in multiple biological processes, such as development, metabolism, cell proliferation, and apoptosis (Angel and Karin 1991; Eferl and Wagner 2003; Hess et al. 2004; Karin et al. 1997). The AP-1 transcription factors usually form either heterodimers or homodimers depending on the sequences of their leucine zipper motifs (Miller 2009), and dimerization is considered a prerequisite for DNA binding (Landschulz et al. 1989). c-Fos and c-Jun form homo- or heterodimers to recognize three types of 7-bp 12-*O*-tetradecanoylphorbol-13-acetate (TPA)-response elements: TGAGTCA, mCGAGTCA, and TGAGmCCA (Bhende et al. 2004; Eferl and Wagner 2003; Gustems et al. 2014; Tulchinsky et al. 1996). C-Jun forms a clamp-like structure through its C-terminal leucine zipper, while its N-terminal basic region interacts with the major groove of the target DNA sequence (Fig. 11.4a). In contrast to the other MBPs, which use a 5mC-Arg-G triad to recognize mC or T, AP-1 binds to mCpG and TpG through van der Waals contacts between a conserved di-alanine (Ala265 and Ala266) and the methyl groups of 5mC or thymine (Fig. 11.4b) (Hong et al. 2017).

Another bZIP TF member, CCAAT/enhancer binding protein β (C/EBPβ), is associated with cytokine-mediated macrophage activation and rapid granulopoiesis (Hirai et al. 2006), and regulates gene expression at different developmental stages (Sun et al. 2017; Tsukada et al. 2011). C/EBPβ recognizes a DNA sequence TTGmCGCAA. Instead of the di-alanine, C/EBPβ bZIP has a unique Ala-Val dipeptide (A284-V285) that contributes to van der Waals interactions with the 5-position methyl group of methylated CpG and thymine, and the methylated CpG dinucleotide forms a classic 5mC-Arg-Gua triad with R289 (Yang et al. 2019). Oxidation of 5mC to 5hmC introduces a hydroxyl group and

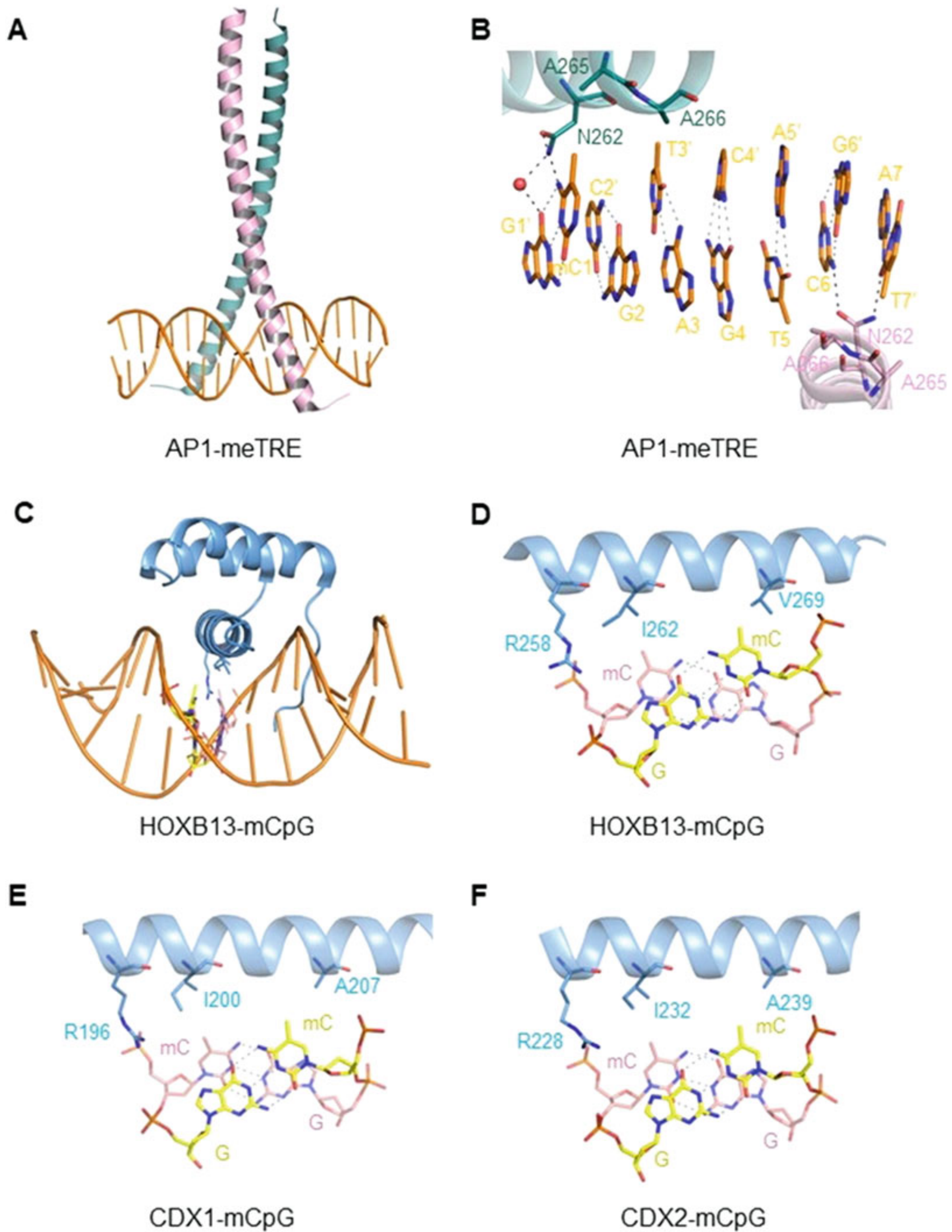


Fig. 11.4 Structures of other TFs binding to methylated DNA. (**a**) AP-1 bZIP domain forms dimer for binding to methylated cytosine DNA (PDB: 5T01). (**b**) AP-1 bZIP protein recognizes the DNA by the van der Waals contacts with the methyl groups of methylated cytosine and thymine. The protein residues are shown as green and pink sticks, respectively, while nucleotides are shown as sticks and colored orange. (**c**) Overall structure of HOXB13 homeodomain in complex with mCpG DNA (PDB: 5EF6). (**d**) HOXB13 homeodomain recognizes the DNA by the van der Waals contacts with the methyl groups of methylated cytosine. The protein residues are shown as blue sticks, while nucleotides are shown in sticks and colored in red and yellow, respectively. (**e**, **f**) CDX1 (PDB: 5LUX) and CDX2 (PDB: 5LTY) homeodomains recognize the DNA by the van der Waals contacts with the methyl groups of methylated cytosine, respectively. The protein residues and nucleotides are shown in the same way as in **d**. The hydrogen bonds formed between protein residues and DNA are marked as black dashed lines, while the DNA base pair interactions are shown as gray dashed lines

reduces its interaction with C/EBPβ (Yang et al. 2019).

11.4.7 Homeodomain Proteins

Homeodomain transcription factors bind to DNA and regulate the expression of morphogenesis-related target genes in eukaryotes (Burglin and Affolter 2016). Some homeodomain transcription factors have been reported to preferentially bind to methylated DNA (Yin et al. 2017). To illustrate the molecular basis for the methylated DNA binding preference, the structures of several homeodomain proteins, such as HOXB13, CDX1, and CDX2, bound to their cognate DNA sequences have been reported (Fig. 11.4c–f) (Yin et al. 2017). It was found that the homeodomain recognizes mCpG by direct hydrophobic interactions with the methyl groups of both 5mCs from the CpG dinucleotide. For example, the complex structure of HOXB13 bound to the DNA sequence CTmCGTAAA showed that its DNA binding domain consists of three α-helices, and the C-terminal α-helix (α3) lies in the major groove of the DNA and the N-terminal tail interacts with the minor groove (Fig. 11.4c) (Yin et al. 2017). Specifically, I262 has a direct hydrophobic interaction with the methyl group of the methylcytosine, while V269 forms another hydrophobic contact with the methyl group of methylcytosine from the complementary strand. Besides, the aliphatic chain of R258 interacts with I262 to enhance the hydrophobic environment (Fig. 11.4d) (Yin et al. 2017). This interaction is also conserved in the binding of CDX1 or CDX2 to GTmCGTAAAA (Fig. 11.4e, f) (Yin et al. 2017). In contrast to CDX1 and CDX2, homeodomain protein LHX4 displays a slightly weaker binding to TmCGTTA site and no significant binding to the unmethylated TCGTTA (Yin et al. 2017). The complex structure of LHX4 bound to the TAATA site showed that R127, V131, and A138 adopt a similar conformation to the corresponding residues in HOXB13, CDX1, and CDX2, which also engage in hydrophobic interaction with TmCGTTA (Yin et al. 2017).

11.5 Conclusion

In this chapter, we reviewed the research on proteins that recognize methylated DNA. Several themes have emerged from recent biochemical, structural, and genomic studies of MBPs. Among them, MBD and C2H2 ZFs DNA binding proteins use the same classic 5mC-Arg-Gua triad to recognize the methylated CpG and unmethylated TpG, while the bZIP proteins AP-1 and homeodomain proteins characterized in this chapter use van der Waals contacts to recognize the methyl groups of 5mC and thymine (Hong et al. 2017; Yin et al. 2017). The SRA proteins utilize specialized pockets to recognize 5mC or 5hmC. Although the POU and PAX proteins, NFAT proteins, SMAD proteins, and FOX proteins have also been reported to interact with mCpG sites (Yin et al. 2017), their methylcytosine recognition mechanism remains to be characterized. Thymine can replace mC for DNA binding due to their similar chemical structure; thus, the TpG recognition by POU and NFAT can also provide insights into their mC binding (Remenyi et al. 2001; Stroud et al. 2002).

Detailed analyses of structural and biochemical data have also indicated that the MBDs recognize DNA without sequence selectivity outside the methylated CpG sequence. However, it seems that individual MBD family members display some level of functional specificity as revealed by genomic mapping experiments in living cells, so how does a methylated CpG motif specify recruitment of a unique MBD has been puzzling, as a dinucleotide sequence like mCpG lacks the chemical information inherent in the longer binding sites typical for most transcription factors. The structural and biochemical work of some other MBPs reviewed here places this issue in an even wider context: MBD and most C2H2 ZF proteins utilize a common protein feature (arginine) to recognize chemical features of the methylated CpG dinucleotide. One possible explanation would be that additional functional domains of these MBPs recognize flanking sequence information, and the true consensus recognition sequence for these proteins is not a

simple methylated CpG, but the methylated CpG with its flanking sequence context that differs for each MBP. This feature endows the MBP proteins with considerable flexibility to respond to DNA methylation with different outputs at unique loci within the genome. Unraveling how cells utilize this surprising flexibility to resolve epigenetic regulation remains a principal challenge of current genetic and genomic experiments.

Acknowledgments This work was supported by funds from the National Natural Science Foundation of China (Number 31770834 to K.L) and the funds from Central China Normal University (Number 31101200001 to J.M. and CCNU19TS084 to K.L). This work was also supported by the Intramural Research Program of the National Institute of Environmental Health Sciences, NIH (ES101965 to P.A.W.). Due to the limited space, we apologize to those whose original work could not be cited in this chapter.

References

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet 23:185–188. https://doi.org/10.1038/13810

Angel P, Karin M (1991) The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation. Biochim Biophys Acta 1072:129–157. https://doi.org/10.1016/0304-419x(91)90011-9

Anvar Z, Cammisa M, Riso V, Baglivo I, Kukreja H, Sparago A, Girardot M, Lad S, De Feis I, Cerrato F, Angelini C, Feil R, Pedone PV, Grimaldi G, Riccio A (2016) ZFP57 recognizes multiple and closely spaced sequence motif variants to maintain repressive epigenetic marks in mouse embryonic stem cells. Nucleic Acids Res 44:1118–1132. https://doi.org/10.1093/nar/gkv1059

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature 455:818–821. https://doi.org/10.1038/nature07249

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, Arrowsmith CH, Dhe-Paganon S (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature 455:822–825. https://doi.org/10.1038/nature07273

Baubec T, Ivanek R, Lienert F, Schubeler D (2013) Methylation-dependent and -independent genomic targeting principles of the MBD protein family. Cell 153:480–492. https://doi.org/10.1016/j.cell.2013.03.011

Baymaz HI, Fournier A, Laget S, Ji Z, Jansen PW, Smits AH, Ferry L, Mensinga A, Poser I, Sharrocks A, Defossez PA, Vermeulen M (2014) MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. Proteomics 14:2179–2189. https://doi.org/10.1002/pmic.201400013

Beckmann AM, Wilce PA (1997) Egr transcription factors in the nervous system. Neurochem Int 31:477–510.; Discussion 517–6. https://doi.org/10.1016/s0197-0186(96)00136-2

Bhende PM, Seaman WT, Delecluse HJ, Kenney SC (2004) The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. Nat Genet 36:1099–1104. https://doi.org/10.1038/ng1424

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

Bozon B, Davis S, Laroche S (2003) A requirement for the immediate early gene zif268 in reconsolidation of recognition memory after retrieval. Neuron 40:695–701. https://doi.org/10.1016/s0896-6273(03)00674-3

Buck-Koehntop BA, Stanfield RL, Ekiert DC, Martinez-Yamout MA, Dyson HJ, Wilson IA, Wright PE (2012) Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. Proc Natl Acad Sci U S A 109:15229–15234. https://doi.org/10.1073/pnas.1213726109

Buhrmester H, von Kries JP, Stratling WH (1995) Nuclear matrix protein ARBP recognizes a novel DNA sequence motif with high affinity. Biochemistry 34: 4108–4117

Burglin TR, Affolter M (2016) Homeodomain proteins: an update. Chromosoma 125:497–521. https://doi.org/10.1007/s00412-015-0543-8

Call KM, Glaser T, Ito CY, Buckler AJ, Pelletier J, Haber DA, Rose EA, Kral A, Yeger H, Lewis WH et al (1990) Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. Cell 60:509–520. https://doi.org/10.1016/0092-8674(90)90601-a

Chahrour M, Jung SY, Shaw C, Zhou X, Wong ST, Qin J, Zoghbi HY (2008) MeCP2, a key contributor to neurological disease, activates and represses transcription. Science 320:1224–1229. https://doi.org/10.1126/science.1153252

Charlton J, Pritchard-Jones K (2016) WT1 mutation in childhood cancer. Methods Mol Biol 1467:1–14. https://doi.org/10.1007/978-1-4939-4023-3_1

Chen RZ, Akbarian S, Tudor M, Jaenisch R (2001) Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. Nat Genet 27:327–331. https://doi.org/10.1038/85906

Chen X, Whitney EM, Gao SY, Yang VW (2003) Transcriptional profiling of Kruppel-like factor 4 reveals a function in cell cycle regulation and epithelial

differentiation. J Mol Biol 326:665–677. https://doi.org/10.1016/s0022-2836(02)01449-3

Chen K, Long Q, Xing G, Wang T, Wu Y, Li L, Qi J, Zhou Y, Ma B, Scholer HR, Nie J, Pei D, Liu X (2020) Heterochromatin loosening by the Oct4 linker region facilitates Klf4 binding and iPSC reprogramming. EMBO J 39:e99165. https://doi.org/10.15252/embj.201899165

Choo Y, Klug A (1997) Physical basis of a protein-DNA recognition code. Curr Opin Struct Biol 7:117–125. https://doi.org/10.1016/s0959-440x(97)80015-2

Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, Ernst J, Plath K (2017) Cooperative binding of transcription factors orchestrates reprogramming. Cell 168: 442–459 e20. https://doi.org/10.1016/j.cell.2016.12.016

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

Cukier HN, Rabionet R, Konidari I, Rayner-Evans MY, Baltos ML, Wright HH, Abramson RK, Martin ER, Cuccaro ML, Pericak-Vance MA, Gilbert JR (2010) Novel variants identified in methyl-CpG-binding domain genes in autistic individuals. Neurogenetics 11:291–303. https://doi.org/10.1007/s10048-009-0228-7

Cukier HN, Lee JM, Ma D, Young JI, Mayo V, Butler BL, Ramsook SS, Rantus JA, Abrams AJ, Whitehead PL, Wright HH, Abramson RK, Haines JL, Cuccaro ML, Pericak-Vance MA, Gilbert JR (2012) The expanding role of MBD genes in autism: identification of a MECP2 duplication and novel alterations in MBD5, MBD6, and SETDB1. Autism Res 5:385–397. https://doi.org/10.1002/aur.1251

Daniel JM, Reynolds AB (1999) The catenin p120(ctn) interacts with Kaiso, a novel BTB/POZ domain zinc finger transcription factor. Mol Cell Biol 19:3614–3623. https://doi.org/10.1128/MCB.19.5.3614

Daniel JM, Spring CM, Crawford HC, Reynolds AB, Baig A (2002) The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. Nucleic Acids Res 30:2911–2919. https://doi.org/10.1093/nar/gkf398

Domcke S, Bardet AF, Adrian Ginno P, Hartl D, Burger L, Schubeler D (2015) Competition between DNA methylation and transcription factors determines binding of NRF1. Nature 528:575–579. https://doi.org/10.1038/nature16462

Dos Santos RL, Tosti L, Radzisheuskaya A, Caballero IM, Kaji K, Hendrich B, Silva JCR (2014) MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. Cell Stem Cell 15:392. https://doi.org/10.1016/j.stem.2014.08.005

Du Y, Liu B, Guo F, Xu G, Ding Y, Liu Y, Sun X, Xu G (2012) The essential role of Mbd5 in the regulation of somatic growth and glucose homeostasis in mice. PLoS One 7:e47358. https://doi.org/10.1371/journal.pone.0047358

Duclot F, Kabbaj M (2017) The role of early growth response 1 (EGR1) in brain plasticity and neuropsychiatric disorders. Front Behav Neurosci 11:35. https://doi.org/10.3389/fnbeh.2017.00035

Eferl R, Wagner EF (2003) AP-1: a double-edged sword in tumorigenesis. Nat Rev Cancer 3:859–868. https://doi.org/10.1038/nrc1209

El-Karim EA, Hagos EG, Ghaleb AM, Yu B, Yang VW (2013) Kruppel-like factor 4 regulates genetic stability in mouse embryonic fibroblasts. Mol Cancer 12:89. https://doi.org/10.1186/1476-4598-12-89

Fang J, Cheng J, Wang J, Zhang Q, Liu M, Gong R, Wang P, Zhang X, Feng Y, Lan W, Gong Z, Tang C, Wong J, Yang H, Cao C, Xu Y (2016) Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition. Nat Commun 7:11197. https://doi.org/10.1038/ncomms11197

Feng Q, Zhang Y (2001) The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. Genes Dev 15:827–832. https://doi.org/10.1101/gad.876201

Filion GJ, Zhenilo S, Salozhin S, Yamada D, Prokhortchouk E, Defossez PA (2006) A family of human zinc finger proteins that bind methylated DNA and repress transcription. Mol Cell Biol 26:169–181. https://doi.org/10.1128/MCB.26.1.169-181.2006

Fraga MF, Ballestar E, Montoya G, Taysavang P, Wade PA, Esteller M (2003) The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. Nucleic Acids Res 31:1765–1774. https://doi.org/10.1093/nar/gkg249

Fujita N, Takebayashi S, Okumura K, Kudo S, Chiba T, Saya H, Nakao M (1999) Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms. Mol Cell Biol 19: 6415–6426. https://doi.org/10.1128/MCB.19.9.6415

Fujita N, Shimotake N, Ohki I, Chiba T, Saya H, Shirakawa M, Nakao M (2000) Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1. Mol Cell Biol 20:5107–5118. https://doi.org/10.1128/MCB.20.14.5107-5118.2000

Gabel HW, Kinde B, Stroud H, Gilbert CS, Harmin DA, Kastan NR, Hemberg M, Ebert DH, Greenberg ME (2015) Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. Nature 522: 89–93. https://doi.org/10.1038/nature14319

Gelato KA, Tauber M, Ong MS, Winter S, Hiragami-Hamada K, Sindlinger J, Lemak A, Bultsma Y, Houliston S, Schwarzer D, Divecha N, Arrowsmith CH, Fischle W (2014) Accessibility of different histone H3-binding domains of UHRF1 is allosterically regulated by phosphatidylinositol 5-phosphate. Mol Cell 54:905–919. https://doi.org/10.1016/j.molcel.2014.04.004

Gessler M, Poustka A, Cavenee W, Neve RL, Orkin SH, Bruns GA (1990) Homozygous deletion in Wilms

tumours of a zinc-finger gene identified by chromosome jumping. Nature 343:774–778. https://doi.org/10.1038/343774a0

Ghosh RP, Horowitz-Scherer RA, Nikitina T, Shlyakhtenko LS, Woodcock CL (2010) MeCP2 binds cooperatively to its substrate and competes with histone H1 for chromatin binding sites. Mol Cell Biol 30:4656–4670. https://doi.org/10.1128/MCB.00379-10

Gnanapragasam MN, Scarsdale JN, Amaya ML, Webb HD, Desai MA, Walavalkar NM, Wang SZ, Zu Zhu S, Ginder GD, Williams DC Jr (2011) p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. Proc Natl Acad Sci U S A 108:7487–7492. https://doi.org/10.1073/pnas.1015341108

Gunther K, Rust M, Leers J, Boettger T, Scharfe M, Jarek M, Bartkuhn M, Renkawitz R (2013) Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences. Nucleic Acids Res 41:3010–3021. https://doi.org/10.1093/nar/gkt035

Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, Zhong C, Hu S, Le T, Fan G, Zhu H, Chang Q, Gao Y, Ming GL, Song H (2014) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat Neurosci 17:215–222. https://doi.org/10.1038/nn.3607

Gustems M, Woellmer A, Rothbauer U, Eck SH, Wieland T, Lutter D, Hammerschmidt W (2014) c-Jun/c-Fos heterodimers regulate cellular genes via a newly identified class of methylated DNA sequence motifs. Nucleic Acids Res 42:3059–3072. https://doi.org/10.1093/nar/gkt1323

Guy J, Hendrich B, Holmes M, Martin JE, Bird A (2001) A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat Genet 27:322–326. https://doi.org/10.1038/85899

Harrison JS, Cornett EM, Goldfarb D, DaRosa PA, Li ZM, Yan F, Dickson BM, Guo AH, Cantu DV, Kaustov L, Brown PJ, Arrowsmith CH, Erie DA, Major MB, Klevit RE, Krajewski K, Kuhlman B, Strahl BD, Rothbart SB (2016) Hemi-methylated DNA regulates DNA methylation inheritance through allosteric activation of H3 ubiquitylation by UHRF1. Elife 5. https://doi.org/10.7554/eLife.17101

Hartwig S, Ho J, Pandey P, Macisaac K, Taglienti M, Xiang M, Alterovitz G, Ramoni M, Fraenkel E, Kreidberg JA (2010) Genomic characterization of Wilms' tumor suppressor 1 targets in nephron progenitor cells during kidney development. Development 137:1189–1203. https://doi.org/10.1242/dev.045732

Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature 455:826–829. https://doi.org/10.1038/nature07280

Hashimoto H, Zhang X, Cheng X (2012) Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. Nucleic Acids Res 40:8276–8284. https://doi.org/10.1093/nar/gks628

Hashimoto H, Olanrewaju YO, Zheng Y, Wilson GG, Zhang X, Cheng X (2014) Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. Genes Dev 28:2304–2313. https://doi.org/10.1101/gad.250746.114

Hashimoto H, Wang D, Steves AN, Jin P, Blumenthal RM, Zhang X, Cheng X (2016) Distinctive Klf4 mutants determine preference for DNA methylation status. Nucleic Acids Res 44:10177–10185. https://doi.org/10.1093/nar/gkw774

Hashimoto H, Wang D, Horton JR, Zhang X, Corces VG, Cheng X (2017) Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. Mol Cell 66(711-720):e3. https://doi.org/10.1016/j.molcel.2017.05.004

Hastie ND (2017) Wilms' tumour 1 (WT1) in development, homeostasis and disease. Development 144:2862–2872. https://doi.org/10.1242/dev.153163

Hendrich B, Bird A (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol 18:6538–6547. https://doi.org/10.1128/MCB.18.11.6538

Hendrich B, Tweedie S (2003) The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends Genet 19:269–277. https://doi.org/10.1016/S0168-9525(03)00080-5

Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A (1999) The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature 401:301–304. https://doi.org/10.1038/45843

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

Hess J, Angel P, Schorpp-Kistner M (2004) AP-1 subunits: quarrel and harmony among siblings. J Cell Sci 117:5965–5973. https://doi.org/10.1242/jcs.01589

Hirai H, Zhang P, Dayaram T, Hetherington CJ, Mizuno S, Imanishi J, Akashi K, Tenen DG (2006) C/EBPbeta is required for 'emergency' granulopoiesis. Nat Immunol 7:732–739. https://doi.org/10.1038/ni1354

Ho KL, McNae IW, Schmiedeberg L, Klose RJ, Bird AP, Walkinshaw MD (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. Mol Cell 29:525–531. https://doi.org/10.1016/j.molcel.2007.12.028

Hong S, Wang D, Horton JR, Zhang X, Speck SH, Blumenthal RM, Cheng X (2017) Methyl-dependent and spatial-specific DNA recognition by the orthologous transcription factors human AP-1 and Epstein-Barr virus Zta. Nucleic Acids Res 45:2503–2515. https://doi.org/10.1093/nar/gkx057

Hopfner R, Mousli M, Jeltsch JM, Voulgaris A, Lutz Y, Marin C, Bellocq JP, Oudet P, Bronner C (2000)

ICBP90, a novel human CCAAT binding protein, involved in the regulation of topoisomerase IIalpha expression. Cancer Res 60:121–128
Hu L, Li Z, Wang P, Lin Y, Xu Y (2011) Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. Cell Res 21:1374–1378. https://doi.org/10.1038/cr.2011.124
Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, Shin J, Cox E, Rho HS, Woodard C, Xia S, Liu S, Lyu H, Ming GL, Wade H, Song H, Qian J, Zhu H (2013) DNA methylation presents distinct binding sites for human transcription factors. Elife 2:e00726. https://doi.org/10.7554/eLife.00726
Huang Y, Zhang H, Wang L, Tang C, Qin X, Wu X, Pan M, Tang Y, Yang Z, Babarinde IA, Lin R, Ji G, Lai Y, Xu X, Su J, Wen X, Satoh T, Ahmed T, Malik V, Ward C, Volpe G, Guo L, Chen J, Sun L, Li Y, Huang X, Bao X, Gao F, Liu B, Zheng H, Jauch R, Lai L, Pan G, Chen J, Testa G, Akira S, Hu J, Pei D, Hutchins AP, Esteban MA, Qin B (2020) JMJD3 acts in tandem with KLF4 to facilitate reprogramming to pluripotency. Nat Commun 11:5061. https://doi.org/10.1038/s41467-020-18900-z
Hudson NO, Whitby FG, Buck-Koehntop BA (2018) Structural insights into methylated DNA recognition by the C-terminal zinc fingers of the DNA reader protein ZBTB38. J Biol Chem 293:19835–19843. https://doi.org/10.1074/jbc.RA118.005147
Jobe EM, Gao Y, Eisinger BE, Mladucky JK, Giuliani CC, Kelnhofer LE, Zhao X (2017) Methyl-CpG-binding protein MBD1 regulates neuronal lineage commitment through maintaining adult neural stem cell identity. J Neurosci 37:523–536. https://doi.org/10.1523/JNEUROSCI.1075-16.2016
Jorgensen HF, Ben-Porath I, Bird AP (2004) Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. Mol Cell Biol 24:3387–3395. https://doi.org/10.1128/MCB.24.8.3387-3395.2004
Kaji K, Caballero IM, MacLeod R, Nichols J, Wilson VA, Hendrich B (2006) The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat Cell Biol 8:285–292. https://doi.org/10.1038/ncb1372
Kaplun D, Starshin A, Sharko F, Gainova K, Filonova G, Zhigalova N, Mazur A, Prokhortchouk E, Zhenilo S (2021) Kaiso regulates DNA methylation homeostasis. Int J Mol Sci 22. https://doi.org/10.3390/ijms22147587
Karin M, Liu Z, Zandi E (1997) AP-1 function and regulation. Curr Opin Cell Biol 9:240–246. https://doi.org/10.1016/s0955-0674(97)80068-3
Katz JP, Perreault N, Goldstein BG, Lee CS, Labosky PA, Yang VW, Kaestner KH (2002) The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon. Development 129:2619–2628
Kim JH, Jeong IY, Lim Y, Lee YH, Shin SY (2011) Estrogen receptor beta stimulates Egr-1 transcription via MEK1/Erk/Elk-1 cascade in C6 glioma cells. BMB Rep 44:452–457. https://doi.org/10.5483/BMBRep.2011.44.7.452
Kinde B, Gabel HW, Gilbert CS, Griffith EC, Greenberg ME (2015) Reading the unique DNA methylation landscape of the brain: non-CpG methylation, hydroxymethylation, and MeCP2. Proc Natl Acad Sci U S A 112:6800–6806. https://doi.org/10.1073/pnas.1411269112
Kinde B, Wu DY, Greenberg ME, Gabel HW (2016) DNA methylation in the gene body influences MeCP2-mediated gene repression. Proc Natl Acad Sci U S A 113:15114–15119. https://doi.org/10.1073/pnas.1618737114
Klose RJ, Sarraf SA, Schmiedeberg L, McDermott SM, Stancheva I, Bird AP (2005) DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. Mol Cell 19:667–678. https://doi.org/10.1016/j.molcel.2005.07.021
Klug A (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annu Rev Biochem 79:213–231. https://doi.org/10.1146/annurev-biochem-010909-095056
Koh DI, Han D, Ryu H, Choi WI, Jeon BN, Kim MK, Kim Y, Kim JY, Parry L, Clarke AR, Reynolds AB, Hur MW (2014) KAISO, a critical regulator of p53-mediated transcription of CDKN1A and apoptotic genes. Proc Natl Acad Sci U S A 111:15078–15083. https://doi.org/10.1073/pnas.1318780111
Koh DI, An H, Kim MY, Jeon BN, Choi SH, Hur SS, Hur MW (2015) Transcriptional activation of APAF1 by KAISO (ZBTB33) and p53 is attenuated by RelA/p65. Biochim Biophys Acta 1849:1170–1178. https://doi.org/10.1016/j.bbagrm.2015.07.008
Kohli RM, Zhang Y (2013) TET enzymes, TDG and the dynamics of DNA demethylation. Nature 502:472–479. https://doi.org/10.1038/nature12750
Kriaucionis S, Tahiliani M (2014) Expanding the epigenetic landscape: novel modifications of cytosine in genomic DNA. Cold Spring Harb Perspect Biol 6:a018630. https://doi.org/10.1101/cshperspect.a018630
Kribelbauer JF, Laptenko O, Chen S, Martini GD, Freed-Pastor WA, Prives C, Mann RS, Bussemaker HJ (2017) Quantitative analysis of the DNA methylation sensitivity of transcription factor complexes. Cell Rep 19:2383–2395. https://doi.org/10.1016/j.celrep.2017.05.069
Laget S, Joulie M, Le Masson F, Sasai N, Christians E, Pradhan S, Roberts RJ, Defossez PA (2010) The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. PLoS One 5:e11982. https://doi.org/10.1371/journal.pone.0011982
Landschulz WH, Johnson PF, McKnight SL (1989) The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science 243:1681–1688. https://doi.org/10.1126/science.2494700

Lee JH, Skalnik DG (2005) CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. J Biol Chem 280:41725–41731. https://doi.org/10.1074/jbc.M508312200

Lee JL, Everitt BJ, Thomas KL (2004) Independent cellular processes for hippocampal memory consolidation and reconsolidation. Science 304:839–843. https://doi.org/10.1126/science.1095760

Le Guezennec X, Vermeulen M, Brinkman AB, Hoeijmakers WA, Cohen A, Lasonder E, Stunnenberg HG (2006) MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. Mol Cell Biol 26:843–851. https://doi.org/10.1128/MCB.26.3.843-851.2006

Lei M, Tempel W, Chen S, Liu K, Min J (2019) Plasticity at the DNA recognition site of the MeCP2 mCG-binding domain. Biochim Biophys Acta Gene Regul Mech 1862:194409. https://doi.org/10.1016/j.bbagrm.2019.194409

Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell 69:905–914. https://doi.org/10.1016/0092-8674(92)90610-o

Li H, Yamagata T, Mori M, Yasuhara A, Momoi MY (2005) Mutation analysis of methyl-CpG binding protein family genes in autistic patients. Brain Dev 27:321–325. https://doi.org/10.1016/j.braindev.2004.08.003

Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, Ferguson-Smith AC (2008) A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Dev Cell 15:547–557. https://doi.org/10.1016/j.devcel.2008.08.014

Li T, Wang L, Du Y, Xie S, Yang X, Lian F, Zhou Z, Qian C (2018) Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation. Nucleic Acids Res 46:3218–3231. https://doi.org/10.1093/nar/gky104

Liang CC, Zhan B, Yoshikawa Y, Haas W, Gygi SP, Cohn MA (2015) UHRF1 is a sensor for DNA interstrand crosslinks and recruits FANCD2 to initiate the Fanconi anemia pathway. Cell Rep 10:1947–1956. https://doi.org/10.1016/j.celrep.2015.02.053

Liu K, Min J (2019) Structural basis for the recognition of non-methylated DNA by the CXXC domain. J Mol Biol. https://doi.org/10.1016/j.jmb.2019.09.025

Liu Y, Toh H, Sasaki H, Zhang X, Cheng X (2012) An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. Genes Dev 26:2374–2379. https://doi.org/10.1101/gad.202200.112

Liu Y, Zhang X, Blumenthal RM, Cheng X (2013) A common mode of recognition for methylated CpG. Trends Biochem Sci 38:177–183. https://doi.org/10.1016/j.tibs.2012.12.005

Liu Y, Olanrewaju YO, Zheng Y, Hashimoto H, Blumenthal RM, Zhang X, Cheng X (2014) Structural basis for Klf4 recognition of methylated DNA. Nucleic Acids Res 42:4859–4867. https://doi.org/10.1093/nar/gku134

Liu Y, Zhang B, Meng X, Korn MJ, Parent JM, Lu LY, Yu X (2017) UHRF2 regulates local 5-methylcytosine and suppresses spontaneous seizures. Epigenetics 12:551–560. https://doi.org/10.1080/15592294.2017.1314423

Liu K, Xu C, Lei M, Yang A, Loppnau P, Hughes TR, Min J (2018) Structural basis for the ability of MBD domains to bind methyl-CG and TG sites in DNA. J Biol Chem 293:7344–7354. https://doi.org/10.1074/jbc.RA118.001785

Liu K, Lei M, Wu Z, Gan B, Cheng H, Li Y, Min J (2019) Structural analyses reveal that MBD3 is a methylated CG binder. FEBS J 286:3240–3254. https://doi.org/10.1111/febs.14850

Liu X, Xu B, Yang J, He L, Zhang Z, Cheng X, Yu H, Liu X, Jin T, Peng Y, Huang Y, Xia L, Wang Y, Wu J, Wu X, Liu S, Shan L, Yang X, Sun L, Liang J, Zhang Y, Shang Y (2021) UHRF2 commissions the completion of DNA demethylation through allosteric activation by 5hmC and K33-linked ubiquitination of XRCC1. Mol Cell 81:2960–2974 e7. https://doi.org/10.1016/j.molcel.2021.05.022

Lu Y, Loh YH, Li H, Cesana M, Ficarro SB, Parikh JR, Salomonis N, Toh CX, Andreadis ST, Luckey CJ, Collins JJ, Daley GQ, Marto JA (2014) Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells. Cell Stem Cell 15:92–101. https://doi.org/10.1016/j.stem.2014.04.002

Luo M, Ling T, Xie W, Sun H, Zhou Y, Zhu Q, Shen M, Zong L, Lyu G, Zhao Y, Ye T, Gu J, Tao W, Lu Z, Grummt I (2013) NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. Stem Cells 31:1278–1286. https://doi.org/10.1002/stem.1374

Mackay DJ, Callaway JL, Marks SM, White HE, Acerini CL, Boonen SE, Dayanikli P, Firth HV, Goodship JA, Haemers AP, Hahnemann JM, Kordonouri O, Masoud AF, Oestergaard E, Storr J, Ellard S, Hattersley AT, Robinson DO, Temple IK (2008) Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nat Genet 40:949–951. https://doi.org/10.1038/ng.187

Mancini M, Magnani E, Macchi F, Bonapace IM (2021) The multi-functionality of UHRF1: epigenome maintenance and preservation of genome integrity. Nucleic Acids Res 49:6053–6068. https://doi.org/10.1093/nar/gkab293

Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell 58:499–507. https://doi.org/10.1016/0092-8674(89)90430-3

Meehan RR, Lewis JD, Bird AP (1992) Characterization of MeCP2, a vertebrate DNA binding protein with

affinity for methylated DNA. Nucleic Acids Res 20: 5085–5092. https://doi.org/10.1093/nar/20.19.5085

Millar CB, Guy J, Sansom OJ, Selfridge J, MacDougall E, Hendrich B, Keightley PD, Bishop SM, Clarke AR, Bird A (2002) Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. Science 297: 403–405. https://doi.org/10.1126/science.1073354

Miller M (2009) The importance of being flexible: the case of basic region leucine zipper transcriptional regulators. Curr Protein Pept Sci 10:244–269. https://doi.org/10.2174/138920309788452164

Miotto B, Chibi M, Xie P, Koundrioukoff S, Moolman-Smook H, Pugh D, Debatisse M, He F, Zhang L, Defossez PA (2014) The RBBP6/ZBTB38/MCM10 axis regulates DNA replication and common fragile site stability. Cell Rep 7:575–587. https://doi.org/10.1016/j.celrep.2014.03.030

Mori T, Li Y, Hata H, Ono K, Kochi H (2002) NIRF, a novel RING finger protein, is involved in cell-cycle regulation. Biochem Biophys Res Commun 296:530–536. https://doi.org/10.1016/s0006-291x(02)00890-2

Motamedi FJ, Badro DA, Clarkson M, Lecca MR, Bradford ST, Buske FA, Saar K, Hubner N, Brandli AW, Schedl A (2014) WT1 controls antagonistic FGF and BMP-pSMAD pathways in early renal progenitors. Nat Commun 5:4444. https://doi.org/10.1038/ncomms5444

Nady N, Lemak A, Walker JR, Avvakumov GV, Kareta MS, Achour M, Xue S, Duan S, Allali-Hassani A, Zuo X, Wang YX, Bronner C, Chedin F, Arrowsmith CH, Dhe-Paganon S (2011) Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. J Biol Chem 286:24300–24311. https://doi.org/10.1074/jbc.M111.234104

Nakahashi H, Kieffer Kwon KR, Resch W, Vian L, Dose M, Stavreva D, Hakim O, Pruett N, Nelson S, Yamane A, Qian J, Dubois W, Welsh S, Phair RD, Pugh BF, Lobanenkov V, Hager GL, Casellas R (2013) A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep 3:1678–1689. https://doi.org/10.1016/j.celrep.2013.04.024

Nan X, Meehan RR, Bird A (1993) Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res 21:4886–4892. https://doi.org/10.1093/nar/21.21.4886

Ng HH, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, Tempst P, Reinberg D, Bird A (1999) MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat Genet 23:58–61. https://doi.org/10.1038/12659

Nikolova EN, Stanfield RL, Dyson HJ, Wright PE (2020) A conformational switch in the zinc finger protein kaiso mediates differential readout of specific and methylated DNA sequences. Biochemistry 59:1909–1926. https://doi.org/10.1021/acs.biochem.0c00253

Nishii T, Oikawa Y, Ishida Y, Kawaichi M, Matsuda E (2012) CtBP-interacting BTB zinc finger protein (CIBZ) promotes proliferation and G1/S transition in embryonic stem cells via Nanog. J Biol Chem 287: 12417–12424. https://doi.org/10.1074/jbc.M111.333856

Ohki I, Shimotake N, Fujita N, Jee J, Ikegami T, Nakao M, Shirakawa M (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. Cell 105:487–497. https://doi.org/10.1016/s0092-8674(01)00324-5

Oikawa Y, Matsuda E, Nishii T, Ishida Y, Kawaichi M (2008) Down-regulation of CIBZ, a novel substrate of caspase-3, induces apoptosis. J Biol Chem 283:14242–14247. https://doi.org/10.1074/jbc.M802257200

Otani J, Arita K, Kato T, Kinoshita M, Kimura H, Suetake I, Tajima S, Ariyoshi M, Shirakawa M (2013) Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. J Biol Chem 288:6351–6362. https://doi.org/10.1074/jbc.M112.431098

Pan H, Bilinovich SM, Kaur P, Riehn R, Wang H, Williams DC Jr (2017) CpG and methylation-dependent DNA binding and dynamics of the methylcytosine binding domain 2 protein at the single-molecule level. Nucleic Acids Res 45:9164–9177. https://doi.org/10.1093/nar/gkx548

Pavletich NP, Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 252:809–817. https://doi.org/10.1126/science.2028256

Pichler G, Wolf P, Schmidt CS, Meilinger D, Schneider K, Frauer C, Fellinger K, Rottach A, Leonhardt H (2011) Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. J Cell Biochem 112:2585–2593. https://doi.org/10.1002/jcb.23185

Pozner A, Hudson NO, Trewhella J, Terooatea TW, Miller SA, Buck-Koehntop BA (2018) The C-terminal zinc fingers of ZBTB38 are novel selective readers of DNA methylation. J Mol Biol 430:258–271. https://doi.org/10.1016/j.jmb.2017.12.014

Prokhortchouk A, Hendrich B, Jorgensen 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 Dev 15:1613–1618. https://doi.org/10.1101/gad.198501

Prokhortchouk A, Sansom O, Selfridge J, Caballero IM, Salozhin S, Aithozhina D, Cerchietti L, Meng FG, Augenlicht LH, Mariadason JM, Hendrich B, Melnick A, Prokhortchouk E, Clarke A, Bird A (2006) Kaiso-deficient mice show resistance to intestinal cancer. Mol Cell Biol 26:199–208. https://doi.org/10.1128/MCB.26.1.199-208.2006

Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A, Trono D (2011) In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell 44:361–372. https://doi.org/10.1016/j.molcel.2011.08.032

Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, Mansour AA, Caspi I, Krupalnik V, Zerbib M, Maza I, Mor N, Baran D, Weinberger L, Jaitin DA, Lara-Astiaso D, Blecher-Gonen R, Shipony Z, Mukamel Z, Hagai T, Gilad S, Amann-Zalcenstein D, Tanay A, Amit I, Novershtern N, Hanna JH (2013) Deterministic direct reprogramming of somatic cells to pluripotency. Nature 502:65–70. https://doi.org/10.1038/nature12587

Rajakumara E, Wang Z, Ma H, Hu L, Chen H, Lin Y, Guo R, Wu F, Li H, Lan F, Shi YG, Xu Y, Patel DJ, Shi Y (2011) PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. Mol Cell 43:275–284. https://doi.org/10.1016/j.molcel.2011.07.006

Rajavelu A, Lungu C, Emperle M, Dukatz M, Brohm A, Broche J, Hanelt I, Parsa E, Schiffers S, Karnik R, Meissner A, Carell T, Rathert P, Jurkowska RZ, Jeltsch A (2018) Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2. Nucleic Acids Res 46:9044–9056. https://doi.org/10.1093/nar/gky715

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A 97:5237–5242. https://doi.org/10.1073/pnas.97.10.5237

Rauscher FJ 3rd, Morris JF, Tournay OE, Cook DM, Curran T (1990) Binding of the Wilms' tumor locus zinc finger protein to the EGR-1 consensus sequence. Science 250:1259–1262. https://doi.org/10.1126/science.2244209

Remenyi A, Tomilin A, Pohl E, Lins K, Philippsen A, Reinbold R, Scholer HR, Wilmanns M (2001) Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping. Mol Cell 8:569–580. https://doi.org/10.1016/s1097-2765(01)00336-7

Rhee HS, Pugh BF (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 147:1408–1419. https://doi.org/10.1016/j.cell.2011.11.013

Riccio A, Aaltonen LA, Godwin AK, Loukola A, Percesepe A, Salovaara R, Masciullo V, Genuardi M, Paravatou-Petsotas M, Bassi DE, Ruggeri BA, Klein-Szanto AJ, Testa JR, Neri G, Bellacosa A (1999) The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. Nat Genet 23:266–268. https://doi.org/10.1038/15443

Rothbart SB, Krajewski K, Nady N, Tempel W, Xue S, Badeaux AI, Barsyte-Lovejoy D, Martinez JY, Bedford MT, Fuchs SM, Arrowsmith CH, Strahl BD (2012) Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat Struct Mol Biol 19:1155–1160. https://doi.org/10.1038/nsmb.2391

Sanchez-Guerrero E, Jo SR, Chong BH, Khachigian LM (2013) EGFR and the complexity of receptor crosstalk in the cardiovascular system. Curr Mol Med 13:3–12

Sanchez-Mut JV, Heyn H, Vidal E, Moran S, Sayols S, Delgado-Morales R, Schultz MD, Ansoleaga B, Garcia-Esparcia P, Pons-Espinal M, de Lagran MM, Dopazo J, Rabano A, Avila J, Dierssen M, Lott I, Ferrer I, Ecker JR, Esteller M (2016) Human DNA methylomes of neurodegenerative diseases show common epigenomic patterns. Transl Psychiatry 6:e718. https://doi.org/10.1038/tp.2015.214

Scarsdale JN, Webb HD, Ginder GD, Williams DC Jr (2011) Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. Nucleic Acids Res 39:6741–6752. https://doi.org/10.1093/nar/gkr262

Schubeler D (2015) Function and information content of DNA methylation. Nature 517:321–326. https://doi.org/10.1038/nature14192

Schuetz A, Nana D, Rose C, Zocher G, Milanovic M, Koenigsmann J, Blasig R, Heinemann U, Carstanjen D (2011) The structure of the Klf4 DNA-binding domain links to self-renewal and macrophage differentiation. Cell Mol Life Sci 68:3121–3131. https://doi.org/10.1007/s00018-010-0618-x

Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, Rajagopal N, Nery JR, Urich MA, Chen H, Lin S, Lin Y, Jung I, Schmitt AD, Selvaraj S, Ren B, Sejnowski TJ, Wang W, Ecker JR (2015) Human body epigenome maps reveal noncanonical DNA methylation variation. Nature 523:212–216. https://doi.org/10.1038/nature14465

Segre JA, Bauer C, Fuchs E (1999) Klf4 is a transcription factor required for establishing the barrier function of the skin. Nat Genet 22:356–360. https://doi.org/10.1038/11926

Sharma R, Choi KJ, Quan MD, Sharma S, Sankaran B, Park H, LaGrone A, Kim JJ, MacKenzie KR, Ferreon ACM, Kim C, Ferreon JC (2021) Liquid condensation of reprogramming factor KLF4 with DNA provides a mechanism for chromatin organization. Nat Commun 12:5579. https://doi.org/10.1038/s41467-021-25761-7

Skene PJ, Illingworth RS, Webb S, Kerr AR, James KD, Turner DJ, Andrews R, Bird AP (2010) Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol Cell 37:457–468. https://doi.org/10.1016/j.molcel.2010.01.030

Song J, Teplova M, Ishibe-Murakami S, Patel DJ (2012) Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science 335:709–712. https://doi.org/10.1126/science.1214453

Soubry A, Staes K, Parthoens E, Noppen S, Stove C, Bogaert P, van Hengel J, van Roy F (2010) The transcriptional repressor Kaiso localizes at the mitotic spindle and is a constituent of the pericentriolar material. PLoS One 5:e9203. https://doi.org/10.1371/journal.pone.0009203

Stoll R, Lee BM, Debler EW, Laity JH, Wilson IA, Dyson HJ, Wright PE (2007) Structure of the Wilms tumor suppressor protein zinc finger domain bound to DNA. J

Mol Biol 372:1227–1245. https://doi.org/10.1016/j.jmb.2007.07.017

Stroud JC, Lopez-Rodriguez C, Rao A, Chen L (2002) Structure of a TonEBP-DNA complex reveals DNA encircled by a transcription factor. Nat Struct Biol 9: 90–94. https://doi.org/10.1038/nsb749

Sun C, Duan P, Luan C (2017) CEBP epigenetic dysregulation as a drug target for the treatment of hematologic and gynecologic malignancies. Curr Drug Targets 18:1142–1151. https://doi.org/10.2174/1389450117666161228160455

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324:930–935. https://doi.org/10.1126/science.1170116

Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126:663–676. https://doi.org/10.1016/j.cell.2006.07.024

Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, Ernst C, Hanscom C, Rossin E, Lindgren AM, Pereira S, Ruderfer D, Kirby A, Ripke S, Harris DJ, Lee JH, Ha K, Kim HG, Solomon BD, Gropman AL, Lucente D, Sims K, Ohsumi TK, Borowsky ML, Loranger S, Quade B, Lage K, Miles J, Wu BL, Shen Y, Neale B, Shaffer LG, Daly MJ, Morton CC, Gusella JF (2012) Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. Cell 149: 525–537. https://doi.org/10.1016/j.cell.2012.03.028

Tate PH, Bird AP (1993) Effects of DNA methylation on DNA-binding proteins and gene expression. Curr Opin Genet Dev 3:226–231. https://doi.org/10.1016/0959-437x(93)90027-m

Tian Y, Paramasivam M, Ghosal G, Chen D, Shen X, Huang Y, Akhter S, Legerski R, Chen J, Seidman MM, Qin J, Li L (2015) UHRF1 contributes to DNA damage repair as a lesion recognition factor and nuclease scaffold. Cell Rep 10:1957–1966. https://doi.org/10.1016/j.celrep.2015.03.038

Tillotson R, Cholewa-Waclaw J, Chhatbar K, Connelly JC, Kirschner SA, Webb S, Koerner MV, Selfridge J, Kelly DA, De Sousa D, Brown K, Lyst MJ, Kriaucionis S, Bird A (2021) Neuronal non-CG methylation is an essential target for MeCP2 function. Mol Cell 81(1260-1275):e12. https://doi.org/10.1016/j.molcel.2021.01.011

Tsukada J, Yoshida Y, Kominato Y, Auron PE (2011) The CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription factors is a multifaceted highly-regulated system for gene regulation. Cytokine 54:6–19. https://doi.org/10.1016/j.cyto.2010.12.019

Tulchinsky EM, Georgiev GP, Lukanidin EM (1996) Novel AP-1 binding site created by DNA-methylation. Oncogene 12:1737–1745

Vaughan RM, Dickson BM, Cornett EM, Harrison JS, Kuhlman B, Rothbart SB (2018) Comparative biochemical analysis of UHRF proteins reveals molecular mechanisms that uncouple UHRF2 from DNA methylation maintenance. Nucleic Acids Res 46:4405–4416. https://doi.org/10.1093/nar/gky151

Veyrac A, Besnard A, Caboche J, Davis S, Laroche S (2014) The transcription factor Zif268/Egr1, brain plasticity, and memory. Prog Mol Biol Transl Sci 122:89–129. https://doi.org/10.1016/B978-0-12-420170-5.00004-0

von Kries JP, Buhrmester H, Stratling WH (1991) A matrix/scaffold attachment region binding protein: identification, purification, and mode of binding. Cell 64:123–135. https://doi.org/10.1016/0092-8674(91)90214-j

Wade PA, Gegonne A, Jones PL, Ballestar E, Aubry F, Wolffe AP (1999) Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet 23:62–66. https://doi.org/10.1038/12664

Walavalkar NM, Cramer JM, Buchwald WA, Scarsdale JN, Williams DC Jr (2014) Solution structure and intramolecular exchange of methyl-cytosine binding domain protein 4 (MBD4) on DNA suggests a mechanism to scan for mCpG/TpG mismatches. Nucleic Acids Res 42:11218–11232. https://doi.org/10.1093/nar/gku782

Wan J, Su Y, Song Q, Tung B, Oyinlade O, Liu S, Ying M, Ming GL, 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 H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R, Thurman RE, Kaul R, Myers RM, Stamatoyannopoulos JA (2012) Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res 22:1680–1688. https://doi.org/10.1101/gr.136101.111

Weitzel JM, Buhrmester H, Stratling WH (1997) Chicken MAR-binding protein ARBP is homologous to rat methyl-CpG-binding protein MeCP2. Mol Cell Biol 17:5656–5666. https://doi.org/10.1128/MCB.17.9.5656

Wong E, Yang K, Kuraguchi M, Werling U, Avdievich E, Fan K, Fazzari M, Jin B, Brown AM, Lipkin M, Edelmann W (2002) Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. Proc Natl Acad Sci U S A 99:14937–14942. https://doi.org/10.1073/pnas.232579299

Woodcock DM, Crowther PJ, Diver WP (1987) The majority of methylated deoxycytidines in human DNA are not in the CpG dinucleotide. Biochem Biophys Res Commun 145:888–894. https://doi.org/10.1016/0006-291x(87)91048-5

Wu Z, Chen S, Zhou M, Jia L, Li Z, Zhang X, Min J, Liu K (2021) Family-wide characterization of methylated DNA binding ability of arabidopsis MBDs. J Mol

Biol 434:167404. https://doi.org/10.1016/j.jmb.2021.167404

Xu C, Bian C, Lam R, Dong A, Min J (2011) The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain. Nat Commun 2:227. https://doi.org/10.1038/ncomms1237

Xu C, Liu K, Lei M, Yang A, Li Y, Hughes TR, Min J (2018) DNA sequence recognition of human CXXC domains and their structural determinants. Structure 26:85–95 e3. https://doi.org/10.1016/j.str.2017.11.022

Yang J, Horton JR, Wang D, Ren R, Li J, Sun D, Huang Y, Zhang X, Blumenthal RM, Cheng X (2019) Structural basis for effects of CpA modifications on C/EBPbeta binding of DNA. Nucleic Acids Res 47:1774–1785. https://doi.org/10.1093/nar/gky1264

Yasui DH, Peddada S, Bieda MC, Vallero RO, Hogart A, Nagarajan RP, Thatcher KN, Farnham PJ, Lasalle JM (2007) Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. Proc Natl Acad Sci U S A 104:19416–19421. https://doi.org/10.1073/pnas.0707442104

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

Yoon HG, Chan DW, Reynolds AB, Qin J, Wong J (2003) N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. Mol Cell 12:723–734. https://doi.org/10.1016/j.molcel.2003.08.008

Yoon HS, Ghaleb AM, Nandan MO, Hisamuddin IM, Dalton WB, Yang VW (2005) Kruppel-like factor 4 prevents centrosome amplification following gamma-irradiation-induced DNA damage. Oncogene 24:4017–4025. https://doi.org/10.1038/sj.onc.1208576

Zandarashvili L, White MA, Esadze A, Iwahara J (2015) Structural impact of complete CpG methylation within target DNA on specific complex formation of the inducible transcription factor Egr-1. FEBS Lett 589:1748–1753. https://doi.org/10.1016/j.febslet.2015.05.022

Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D (1999) Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev 13:1924–1935. https://doi.org/10.1101/gad.13.15.1924

Zhang J, Gao Q, Li P, Liu X, Jia Y, Wu W, Li J, Dong S, Koseki H, Wong J (2011) S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. Cell Res 21:1723–1739. https://doi.org/10.1038/cr.2011.176

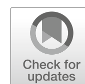
Zhang P, Rausch C, Hastert FD, Boneva B, Filatova A, Patil SJ, Nuber UA, Gao Y, Zhao X, Cardoso MC (2017) Methyl-CpG binding domain protein 1 regulates localization and activity of Tet1 in a CXXC3 domain-dependent manner. Nucleic Acids Res 45:7118–7136. https://doi.org/10.1093/nar/gkx281

Zhao S, Choi M, Overton JD, Bellone S, Roque DM, Cocco E, Guzzo F, English DP, Varughese J, Gasparrini S, Bortolomai I, Buza N, Hui P, Abu-Khalaf M, Ravaggi A, Bignotti E, Bandiera E, Romani C, Todeschini P, Tassi R, Zanotti L, Carrara L, Pecorelli S, Silasi DA, Ratner E, Azodi M, Schwartz PE, Rutherford TJ, Stiegler AL, Mane S, Boggon TJ, Schlessinger J, Lifton RP, Santin AD (2013) Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. Proc Natl Acad Sci U S A 110:2916–2921. https://doi.org/10.1073/pnas.1222577110

Zhou T, Xiong J, Wang M, Yang N, Wong J, Zhu B, Xu RM (2014) Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. Mol Cell 54:879–886. https://doi.org/10.1016/j.molcel.2014.04.003

Zou X, Ma W, Solov'yov IA, Chipot C, Schulten K (2012) Recognition of methylated DNA through methyl-CpG binding domain proteins. Nucleic Acids Res 40:2747–2758. https://doi.org/10.1093/nar/gkr1057

Zuo X, Sheng J, Lau HT, McDonald CM, Andrade M, Cullen DE, Bell FT, Iacovino M, Kyba M, Xu G, Li X (2012) Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. J Biol Chem 287:2107–2118. https://doi.org/10.1074/jbc.M111.322644



12 Recent Advances on DNA Base Flipping: A General Mechanism for Writing, Reading, and Erasing DNA Modifications

Ren Ren, John R. Horton, Samuel Hong, and Xiaodong Cheng

Abstract

The modification of DNA bases is a classic hallmark of epigenetics. Four forms of modified cytosine—5-methylcytosine, 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine—have been discovered in eukaryotic DNA. In addition to cytosine carbon-5 modifications, cytosine and adenine methylated in the exocyclic amine—N4-methylcytosine and N6-methyladenine—are other modified DNA bases discovered even earlier. Each modified base can be considered a distinct epigenetic signal with broader biological implications beyond simple chemical changes. Since 1994, several crystal structures of proteins and enzymes involved in writing, reading, and erasing modified bases have become available. Here, we present a structural synopsis of writers, readers, and erasers of the modified bases from prokaryotes and eukaryotes. Despite significant differences in structures and functions, they are remarkably similar regarding their engagement in flipping a target base/nucleotide within DNA for specific recognitions and/or reactions. We thus highlight base flipping as a common structural framework broadly applied by distinct classes of proteins and enzymes across phyla for epigenetic regulations of DNA.

Keywords

Epigenetic methylation · DNA base flipping · Reader, writer and eraser of DNA methyl marks · Methyltransferases · Demethylases

Abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5ghmC	Glucosylated 5-hydroxymethylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
αKG	α-ketoglutarate
AdoHcy	*S*-adenosyl-L-homocysteine (SAH)
AdoMet	*S*-adenosyl-L-methionine (SAM)
AlkB	*E. coli* Alkylated DNA repair protein AlkB
ALKBH5	Alkylated DNA repair protein AlkB homolog 5 in human
CMT2	Chromomethylase 2 (plant specific)

Ren Ren and John R. Horton contributed equally with all other contributors.

R. Ren · J. R. Horton · X. Cheng (✉)
Department of Epigenetics and Molecular Carcinogenesis, University of Texas MD Anderson Cancer Center, Houston, TX, USA
e-mail: RRen1@mdanderson.org; JRHorton@mdanderson.org; xcheng5@mdanderson.org

S. Hong
McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, USA
e-mail: samuel.hong@uth.tmc.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_12

CMT3	Chromomethylase 3 (plant specific)
DME	Demeter (plant)
DML3	Demeter-like protein 3 (plant)
DNMT1	Mammalian DNA methyltransferase 1
DNMT3A	Mammalian DNA methyltransferase 3A
DNMT3L	Mammalian DNA methyltransferase 3-like
DRM2	Domain rearranged methyltransferase 2 (plant)
FTO	Fat mass and obesity-associated protein
HhH	Helix-hairpin-helix
JBP	J-binding protein
MBD	Methyl-CpG binding domain
McrB	Modified cytosine restriction B
Met1	DNA methyltransferase 1 (plant)
MTase	Methyltransferase
N4mC	N4-methylcytosine
N6mA	N6-methyladenine
NOG	*N*-oxalylglycine
ROS1	Repressor of silencing 1 (plant specific)
SRA	SET- and RING-associated
TDG	Thymine DNA glycosylase
TET	Ten-eleven translocation
TRD	Target recognition domain
Uhrf1	Ubiquitin-like-containing PHD and RING finger domains protein 1
WH	Winged helix

12.1 Introduction

Chemical modifications of DNA bases have fundamental biological roles in virtually every living organism. In both prokaryotes and many eukaryotes, cytosine can be methylated at carbon-5 (C5) position by cytosine C5 methyltransferases (MTases) to generate 5-methylcytosine (5mC) (Kumar et al. 1994; Bestor 2000) (Fig. 12.1a). In higher eukaryotes, ten-eleven translocation (TET) 5mC dioxygenase enzymes utilize α-ketoglutarate (αKG) and Fe (II) to oxidize the methyl group of 5mC to generate 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) via discrete reactions (Kriaucionis and Heintz 2009; Tahiliani et al. 2009; Ito et al. 2011; He et al. 2011). In prokaryotes, 5mC and 5hmC can be introduced de novo into the genome during phage invasions, as both modified bases can be synthesized prior to incorporation into the phage genome during DNA synthesis (Warren 1980). After DNA synthesis, phage glucosyltransferases can modify 5hmC within the genome to generate glucosylated 5hmC (5ghmC) (Lehman and Pratt 1960; Kornberg et al. 1961; Lunt et al. 1964). Beyond cytosine C5 modifications, exocyclic amine groups of adenine and cytosine can be methylated in prokaryotes to generate N6-methyladenine (N6mA) and N4-methylcytosine (N4mC), respectively (Malone et al. 1995) (Fig. 12.1b, c). Crystal structures of DNA modification enzymes to date have consistently shown that the target nucleotide is flipped out of the double helix for reactions in a process called base flipping.

In addition to the modification writers, modified base readers have also been shown to flip the target base for recognition. Mammalian SET- and RING-associated (SRA) domains recognize 5mC within genome by base flipping (Arita et al. 2008; Avvakumov et al. 2008; Hashimoto et al. 2008), and have been characterized as non-enzymatic base flippers. Since the first discovery in eukaryotes, SRA has been rediscovered in prokaryotes, recognizing 5mC, 5hmC, and/or 5ghmC to coordinate restriction activity in a modification-dependent manner (Horton et al. 2014a–c). In addition to SRA, the bacterial modified cytosine restriction B enzyme (McrB) also flips 5mC for recognition but is structurally distinct from other known base flippers (Sukackaite et al. 2012). Structural homologs of McrB across different phyla may recognize modified bases in a similar way.

A brief survey of DNA base modifications in both prokaryotes and eukaryotes reveals that two major families of enzymes, methyltransferases, and dioxygenases are involved in writing DNA modifications in the four forms of modified cytosine: 5mC, 5hmC, 5fC, and 5caC. In plants, 5mC DNA glycosylase repressor of silencing 1 (ROS1) can excise 5mC and 5hmC (in vitro) (Gong et al.

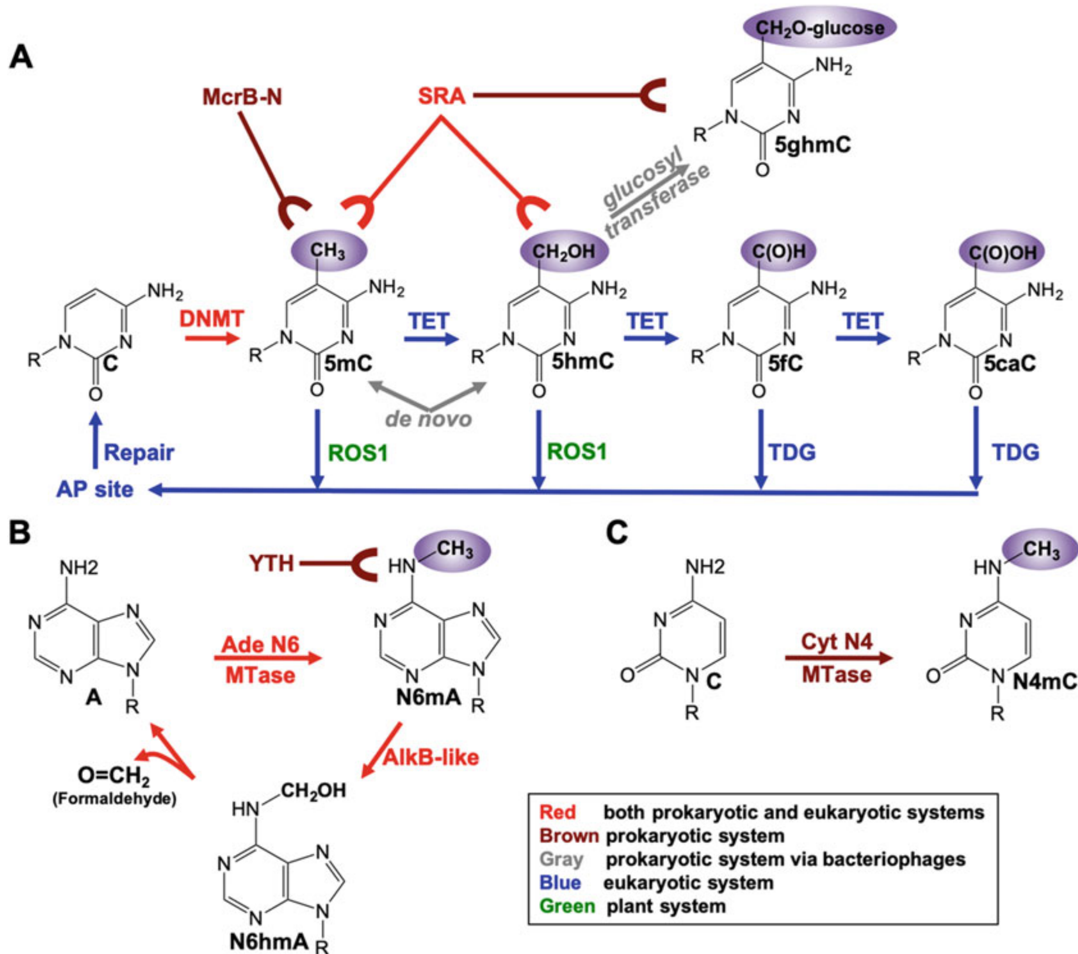


Fig. 12.1 Chemical modifications of nucleic acids. (**a**) DNA cytosine C5 modifications: enzymes and proteins involved in writing, reading, and erasing the modifications via base-flipping mechanisms. (**b**) Adenine N6 methylation in DNA or RNA: enzymes and proteins involved in writing, reading, and erasing adenine N6 methylation. (**c**) Cytosine N4 methylation

2002; Jang et al. 2014; Hong et al. 2014), and mammalian thymine DNA glycosylase (TDG) can excise 5fC and 5caC (He et al. 2011; Maiti and Drohat 2011; Zhang et al. 2012; Hashimoto et al. 2012a). These discoveries effectively link the base excision repair pathway, including AlkB homologs (see below), to DNA demethylation/demodification, by which epigenetic signals encoded in the modified cytosines can be reversed. DNA glycosylases represent the most structurally diverse family of enzymes that are involved in base flipping (also known as nucleotide flipping) (Brooks et al. 2013). Thus, base flipping is not restricted to writers and readers, but has been adopted by DNA glycosylases for erasing DNA modifications as well. Together, structural characterizations of writers, readers, and erasers of DNA base modifications in prokaryotes and eukaryotes effectively showcase base flipping as a general mechanism for regulating and translating fundamental epigenetic signals as well as for maintaining genome integrity (i.e., DNA damage repair).

12.2 Base Flipping for Methylation of DNA Bases

12.2.1 Bacterial DNMTs (HhaI, TaqI, Dam, CcrM, and CamA)

Biological methylation is widely engaged in various regulations, and it uses *S*-adenosyl-L-methionine (AdoMet or SAM) as a primary methyl donor. The methyl group of AdoMet is bound to a positively charged sulfur atom predisposed to a nucleophilic attack. During the methylation reaction, AdoMet loses the methyl group and becomes *S*-adenosyl-L-homocysteine (AdoHcy or SAH). A number of different families of MTases use AdoMet as cofactor, targeting diverse substrates ranging from small molecules to large macromolecules such as DNA, RNA, proteins, lipid, and polysaccharides. The atoms subjected to methylation also vary, including carbon (C), nitrogen (N), oxygen (O), sulfur (S), and several metals. AdoMet-dependent DNA MTases were first discovered in bacterial restriction-modification systems (Roberts et al. 2015). The known structures of AdoMet-dependent DNA MTases share a common "MTase fold" characterized by mixed seven-stranded β sheets (6↓ 7↑ 5↓ 4↓ 1↓ 2↓ 3↓) in which strand 7 is inserted between strands 5 and 6 antiparallel to the others (Cheng 1995; Schubert et al. 2003).

M.HhaI was the first DNA MTase to be structurally characterized (Cheng et al. 1993) (Fig. 12.2a). It contains an N-terminal MTase domain and a C-terminal target recognition domain (TRD). M.HhaI is a cytosine C5 MTase that methylates the first cytosine within 5′-GCGC-3′ recognition sequences and prevents R.HhaI restriction activity at this site (Roberts et al. 1976; Horton et al. 2020). Before the structure was available, the proposed mechanism predicted that catalytic Cys81 would make a nucleophilic attack on C6 of cytosine to form a covalent complex, followed by transferring the methyl group from AdoMet to cytosine C5 and releasing the covalent intermediate (Wu and Santi 1985, 1987). In 1994, the crystal structure of M. HhaI-DNA with AdoMet was solved as a trapped covalent enzyme-DNA intermediate using 5-fluorocytosine and directly supported the proposed mechanism, presenting the catalytic cysteine covalently linked to C6 and showing methylated C5 adjacent to AdoHcy (Klimasauskas et al. 1994). Yet, the most striking aspect of the structure was that both the MTase and the TRD of the enzyme work simultaneously to bind DNA and flip the target base into the active-site pocket. The mechanism of DNA base access by base flipping has since been described as the framework for other DNA MTases (Cheng and Roberts 2001).

After the first structure of M.HhaI-DNA was solved, many crystal structures of DNA MTase-DNA complexes have been solved. Besides cytosine C5 methylation, adenine exocyclic N6 methylation is also a critical modification in prokaryotic DNA (Fig. 12.2b–d) and in eukaryotic RNA (Anton and Roberts 2021; Wei and He 2021). The structure of the adenine N6 MTase M. TaqI in complex with DNA and a non-reactive AdoMet analog was solved in 2001 (Goedecke et al. 2001) (Fig. 12.2b). The enzyme methylates adenine within 5′-TCGA-3′ sequence and harbors a similar two-domain organization as M.HhaI, with the conserved N-terminal MTase domain, but a quite distinct C-terminal TRD. The ternary structure is remarkably reminiscent of M.HhaI, involving a flipped adenine in the active site, where the methyl group from the AdoMet analog is positioned near N6 of the flipped adenine. Instead of the catalytic cysteine residue as in M. HhaI, the asparagine 105 side chain and the following proline backbone oxygen make hydrogen bonds with the adenine N6 amine group, potentially modulating the direct transfer of the methyl group from AdoMet. A similar mode of interaction is also seen in the active site of the T4 phage DNA adenine MTase (T4 Dam) that flips adenine in 5′-GATC-3′ sequence, and an aspartate residue (Asp171) contacts adenine N6 (Horton et al. 2005).

Dam is an orphan MTase (Fig. 12.2c), a type of MTase that acts alone without associated cognate restriction endonuclease as part of R-M system (Roberts et al. 2015). Besides Dam, cell

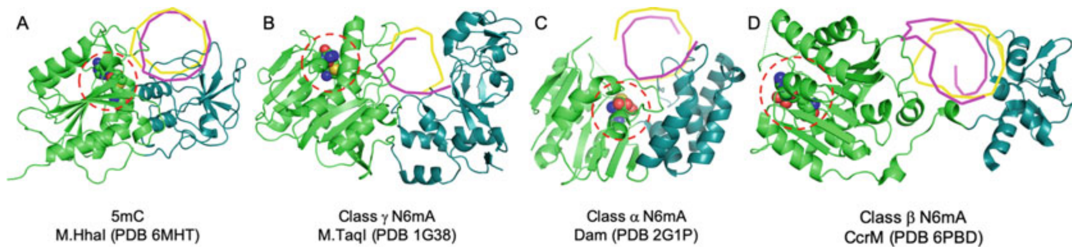


Fig. 12.2 Examples of bacterial DNA MTases. (**a**) M. HhaI, a 5mC MTase, (**b**) M.TaqI, a class γ MTase, (**c**) EcoDam, a class α MTase, and (**d**) CcrM, a class β MTase. The MTase domain (green) binds cofactor (in ball model), and the target recognition domain (TRD) is colored in dark blue. The DNA recognition strand containing the flipped target base is in magenta, and the complementary strand in yellow

cycle-regulated DNA MTase (CcrM) in *Caulobacter crescentus* (Zweiger et al. 1994) (Fig. 12.2d) and newly discovered *Clostridioides difficile* adenine MTase A (CamA) (Oliveira et al. 2020) are also orphan MTases, although they belong to different subgroups depending on the sequential order of conserved motifs (Malone et al. 1995; Woodcock et al. 2020a). The kinetics and structural studies revealed these orphan MTases shared similarities while functioning quite differently. Both Dam (α-group) and CcrM (β-group) are responsible for the post-replication maintenance of daughter strand adenine methylation of symmetrical GATC sequence or near-symmetric GAnTC (n=any base) (Messer and Noyer-Weidner 1988; Stephens et al. 1996). CamA (γ-group), predominantly presented in all *C. difficile* genomes, methylates an asymmetric 6-bp sequence: CAAAAA (underlining indicates the target A) (Oliveira et al. 2020; Zhou et al. 2021). Although the target adenine residue is flipped out during the catalysis for all three enzymes, each individual orphan MTase has its own distinct mechanism. Dam-bound DNA conformation has intact intrahelical paired bases (Horton et al. 2005), CcrM pulls the two DNA strands apart and creates a bubble comprising four unpaid bases for enzyme recognition (Horton et al. 2019), and CamA squeezes out the target adenine by base pair rearrangement (Zhou et al. 2021). In addition, Dam and CamA make base-specific contacts to both DNA strands, whereas CcrM only contacts the bases in the target strand. These unique features allow CcrM to methylate not only double-stranded but also single-stranded DNA (Reich et al. 2018; Konttinen et al. 2020), while both Dam and CamA are inactive on single-stranded DNA.

12.2.2 Mammalian DNMTs (DNMT1, DNMT3A/3L)

Structural features of classic prokaryotic cytosine C5 MTases are extensively shared by mammalian DNA MTases: DNMT1, DNMT3A, and DNMT3B. They are all cytosine C5 MTases containing the MTase domain with the catalytic cysteine and the TRD. DNMT1 is primarily implicated in methylation of the daughter strand during DNA replication to maintain the methylation pattern encoded in the mother strand by preferentially recognizing hemi-methylated DNA in CpG dinucleotide context (Li et al. 1992). On the other hand, DNMT3A and DNMT3B are considered de novo MTases that can methylate CpG sites as well as non-CpG sites (Okano et al. 1999; Ramsahoye et al. 2000; Gowher and Jeltsch 2001). Such differences in substrate specificities are partly due to the involvement of other domains outside the catalytic fragment. For example, the CXXC and BAH1 domains within DNMT1 hinder methylation of unmethylated CpG sites (Song et al. 2011), whereas DNMT3A and DNMT3B do not contain such domains and can readily methylate them.

Moreover, it appears there exist strong influences of flanking sequences on the CpG and non-CpG methylation activity of mammalian

DNA MTases and resulting patterns of methylation. Deep enzymology experiments, which utilize substrates with partly randomized sequences and then analyze the methylation levels by bisulfite conversion coupled with deep next-generation sequencing (NGS) readout, along with structural information, delineate an intricate interplay between flanking sequence, the enzyme-mediated base flipping, and the dynamic landscape of DNA methylation (Jeltsch et al. 2021). Flanking sequences of CpG directly influence catalysis by affecting base-flipping mechanisms which are accompanied by conformational changes in the MTase.

For DNMT1 (Fig. 12.3a), there are strong effects of the ± 2 flanking sequences of a CpG on its activity, with an about 100-fold difference in methylation rates of $N_{-2}N_{-1}CpGN_{+1}N_{+2}$ sites with the best and worst flanking sequences (Adam et al. 2020). Three crystal structures augment this study as a DNMT1 complex with DNA with a highly favorable flanking C_{-1} only showed flipping of the target base, while another complex with a less favorable A_{-1} showed the target base flipping together with rotation of the orphaned G. Finally, a structure with the least favorable G_{-1} displayed a rotation of the target base and invasion of the orphaned G into the −1 flank base pair followed by the formation of a GG non-canonical base pair and flipping of the C normally in the base pair with the G_{-1} residue (Song et al. 2011; Adam et al. 2020). Differences in positioning of the helix that follows the catalytic loop of DNMT1 could be observed with the different nucleotides at the N_{-1} position as well: with A and C at that position, the helix of DNMT1 predominantly adopted a kinked conformation, and with G, the helix mainly adopted a straight conformation moving the active-site loop away from the DNA. Differences of residues at the plus-side flank of the CpG appear to affect the minor groove width which may also influence the equilibrium between the two alternative conformations. Structural transition between these two states of active-site loop-helix is required for DNMT1 activity, suggesting that targeting this transition with small compounds could reduce DNA methylation in cancer cells (Ye et al. 2018). Excitingly, this has been shown to be the case with a new class of reversible DNMT1-selective inhibitors containing a dicyanopyridine moiety (Pappalardi et al. 2021) (Fig. 12.3c).

The flanking sequence preferences of DNMT3A and DNMT3B have also been investigated using the deep enzymology approach (Gao et al. 2020). Methylation levels were averaged for 4096 CpG sites with randomized ± 3 flanks, i.e., $N_{-3}N_{-2}N_{-1}CpGN_{+1}N_{+2}N_{+3}$ sites. The highly methylated sites of DNMT3A showed a preference for C at the +1 site, while DNMT3B preferred G/A at the +1 site and a G at the +2 site DNMT3B while both enzymes prefer T at the −2 site. In addition, the ratio of flanking sequence preferences of DNMT3A and DNMT3B for sites extended a more than 100-fold range, illustrating the noticeable divergence in flanking sequence preferences between the two enzymes.

12.2.3 Implications of DNA Methyltransferase Oligomers (DNMT3A/3L, DNMT3A/3B3, EcoP15I, CcrM, and MettL3-14)

The genomic targeting of DNMT3A and DNMT3B is further regulated by additional factors, including their N-terminal domains and DNMT3L. Besides being a catalytic domain, the MTase domain can participate in protein–protein interactions as exemplified by the DNMT3A MTase domain interacting with a naturally inactive MTase-like domain of DNMT3L, a scaffold protein that binds histone tail H3 to guide DNMT3A activities by forming a tetramer of 3L-3A-3A-3L (Jia et al. 2007; Ooi et al. 2007) (Fig. 12.4a). Moreover, DNMT3L can also form a similar linear assembly with DNMT3B (Lin et al. 2020) (Fig. 12.4b), and 3B-3B and 3B-3L interfaces share some conserved residues with those of the DNMT3A-3L complex.

As a key accessory protein of de novo DNA methylation, Dnmt3L predominantly exists in early embryos and embryonic stem cells and is silenced upon differentiation. Dnmt3b3, a catalytically inactive Dnmt3b isoform, which is

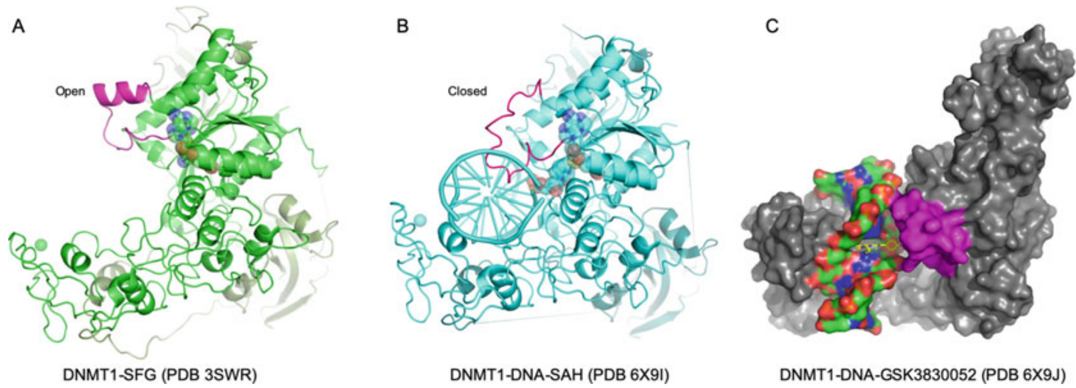


Fig. 12.3 Human DNMT1. (**a**) In the absence of DNA. (**b**) In the presence of DNA. (**c**) In the presence of DNA- and DNMT1-specicifc inhibitor

ubiquitously expressed in differentiated cells and contains the exact 63-residue deletion corresponding to that of Dnmt3L, carries out the regulatory role during late embryonic development and in somatic cells. Dnmt3b3 can positively regulate the catalytic activities of both Dnmt3a2 and Dnmt3b2 either in vitro or in vivo (Duymich et al. 2016; Zeng et al. 2020). The stimulatory effect of Dnmt3b3 is highly dependent on the direct interaction with active form of Dnmt3a or Dnmt3b proteins, and the optimal stimulation is reached at equal molar stoichiometry. The newly reported cryo-electron microscopy structure defined the architecture of a ternary complex of Dnmt3a2-Dnmt3b3 heterotetramer with a nucleosome core particle flanked by linker DNA (Xu et al. 2020) (Fig. 12.4c). This complex contains two monomers each of Dnmt3a2 and Dnmt3b3, forming a tetramer with 3b3-3a2-3a2-3b3, similar to the arrangement of 3L-3a-3a-3L complex. The 3b3-3a2-3a2-3b3 tetramer interacts asymmetrically with the nucleosome with one of the Dnmt3b3 molecules bound to the acidic patch of the nucleosome. The contact point orients both Dnmt3a2 catalytic domains and the second Dnmt3b3 to follow the path of the linker DNA and constrain the arrangement of Dnmt3a target recognition domain with linker DNA. This model suggested a crucial role of Dnmt3b3 in targeting the nucleosome core and driving the de novo methylation on a genome-wide scale.

Interestingly, a multi-subunit prokaryotic DNA N6mA methyltransferase, EcoP15I, contains a DNA MTase dimer in which one monomer is involved in target base flipping and the other in the recognition of DNA base context

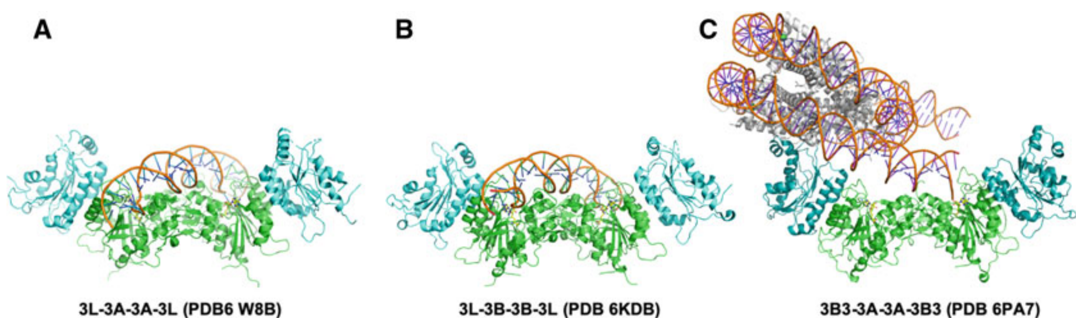


Fig. 12.4 Human DNMT3 family. (**a**) DNMT3A-3L tetramer in complex with DNA. (**b**) DNMT3B-3L tetramer in complex with DNA. (**c**) DNMT3A-3B3 tetramer in complex with nucleosome

(Gupta et al. 2015) (Fig. 12.5a). The "division of labor" might be a conserved feature among class-beta MTases, including M.EcoGII (Murray et al. 2018) and M.EcoP15I (Gupta et al. 2015) from *Escherichia coli*, *Caulobacter crescentus* cell cycle-regulated DNA methyltransferase (CcrM) (Horton et al. 2019) (Fig. 12.5b), the MTA1-MTA9 complex from the ciliate *Oxytricha* (Beh et al. 2019) (Fig. 12.5c), and the mammalian MettL3-MettL14 complex (Fig. 12.5d) (reviewed in Woodcock et al. 2020a). These MTases all generate N6-methyladenine in DNA, with some members having activity on single-stranded DNA as well as RNA. The beta class of MTases has a unique multimeric feature, forming either homo- or hetero-dimers, allowing the enzyme to use division of labor between two subunits in terms of substrate recognition and methylation. Thus, dimerization of two structurally comparable proteins for divergent functionalities may be a mechanism for fine-controlling DNA/RNA modifications.

12.2.4 Plant DNMTs

Plant DNA MTases show similar functionalities to their mammalian counterparts. Met1 is homologous to mammalian DNMT1 and is responsible for the maintenance CpG methylation, whereas domains rearranged methyltransferase 2 (DRM2) is involved in de novo DNA methylation (Law and Jacobsen 2010). DRM2 contains a rearranged MTase domain, such that its N-terminal half is equivalent to the C-terminal half of the conventional MTase fold and vice versa. A structural study of DRM2 family MTase domain has revealed that the rearranged domain still forms a classic MTase structure and functions as a homodimer (Zhong et al. 2014) analogous to the DNMT3A-3L heterodimer. In addition to Met1 and DRM2, plants also have plant-specific DNA MTases, such as CMT2 and CMT3 that are specifically involved in CNG methylation (Stroud et al. 2014; Lindroth et al. 2001; Zemach et al. 2013). The higher diversity of the MTase family within plants compared to the mammalian family suggests that DNA methylation may be more dynamically regulated in plants than in mammals.

12.3 Base Flipping in Oxidative Modifications of Methylated Bases

12.3.1 Eukaryotic TET Enzymes

The 5mC is by far the most widely studied modified base. Yet, if 5mC has been considered "the fifth" base of the genetic alphabet, 5hmC is increasingly being labeled as "the sixth" base and has garnered much attention. The existence of 5hmC in bacteriophage, modified from 2′-deoxycytidine before integration into the viral genome (Warren 1980), was first reported in the early 1950s (Wyatt and Cohen 1952, 1953). In 1993, a novel J base (β-D-glucosyl hydroxymethyluracil) was discovered in trypanosome, in which J-binding proteins (JBP1 and JBP2) are involved in oxidizing the C5 methyl group of thymine during J base synthesis by using αKG and Fe(II) as cofactors to generate 5-hydroxymethyluracil (Gommers-Ampt et al. 1993; Borst and Sabatini 2008). In 2009, mammalian JBP homolog TET enzymes were discovered to oxidize the methyl group of 5mC to generate 5hmC (Tahiliani et al. 2009). Further analysis revealed that TET enzymes could further oxidize 5hmC to 5fC and then to 5caC (Ito et al. 2011; He et al. 2011). Also, TET enzymes have been shown to convert thymine (5-methyluracil) to 5-hydroxymethyluracil by oxidizing the C5 methyl group of thymine (Pfaffeneder et al. 2011; Pais et al. 2015).

Eukaryotic JBP/TET homologs are present across many eukaryotic organisms including amoeboflagellate *Naegleria gruberi* (Iyer et al. 2013; Hashimoto et al. 2014a, 2015a). Crystal structures of *Naegleria gruberi* TET-like (NgTET) (Fig. 12.6a) and human TET2 (hTET2) (Fig. 12.6b) in complex with 5mC-, 5hmC, and 5fC-containing DNA have been characterized (Hashimoto et al. 2014a, 2015a; Hu et al. 2013, 2015). All TET structures show a flipped base positioned in the active-site pocket close to *N*-oxalylglycine (NOG)—an inactive αKG analog—and a divalent metal such as Fe(II) or Mn(II) for stalling the enzyme in the pre-reaction state. Some of the features of the flipped base recognition observed in

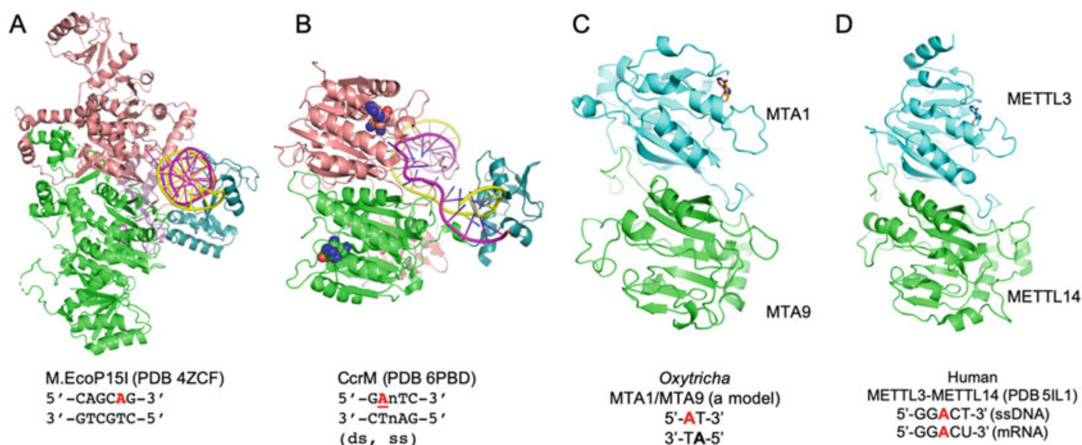


Fig. 12.5 Examples of dimeric class β MTases. (**a**) M.EcoP15I-DNA complex. (**b**) CcrM-DNA complex. (**c**) A model of Oxytricha MTA1-MTA9. (**d**) Human MettL3-MettL14 in the absence of substrate RNA or DNA

DNMT-DNA complex structures (Cheng and Roberts 2001) can also be seen in the structures of TET-DNA complexes. The flipped base in the active site of a TET enzyme in complex with DNA is stabilized by π stacking interactions involving an aromatic residue such as Phe295 in NgTET (Hashimoto et al. 2014a) and Tyr1902 in hTET2 (Hu et al. 2013). Also, polar residues such as Asn147, His297, and Asp234 in NgTET contact O2, N3, and N4, respectively, to guide substrate specificities (Hashimoto et al. 2014a), and the methyl or the hydroxymethyl group is oriented toward NOG and Fe(II)/Mn(II) (Hashimoto et al. 2015a; Hu et al. 2015). Often, active-site pockets for flipped bases not only contain residues for base recognition, but also specifically orient the base for distinct reactions depending on the type of enzymes. Base flipping is therefore a common mechanism applied by different classes of enzymes, such as AdoMet-dependent methyltransferases and αKG- and Fe(II)-dependent dioxygenases to recognize and stabilize the target base for specific reactions.

Like the mammalian DNMTs, flanking sequence preferences of TET1 and TET2 have been analyzed by a deep enzymology approach

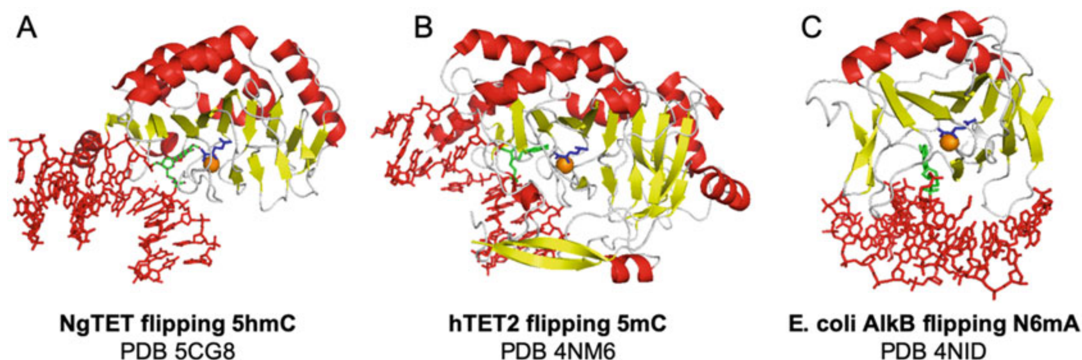


Fig. 12.6 Examples of Fe(II)- and αKG-dependent dioxygenases. (**a**) *Naegleria gruberi* Tet in complex with DNA. (**b**) Human Tet2 in complex with DNA. (**c**) *E. coli* AlkB in complex with DNA. The metal ion is shown as an orange ball in the active site, αKG is colored in blue, and the flipped nucleotide is in green

(Adam et al. 2022) and revealed that TET enzymes show up to 70-fold differences in oxidation rates of either 5mC or 5hmC in CpG target sites embedded in different $N_{-2}N_{-1}CpGN_{+1}N_{+2}$ flanking contexts. For TET1 and TET2 and both substrates, an A is strongly preferred at the −1 site and G is strongly disfavored. At the +1 site, C is generally disfavored, and at the +2 site, TET2 prefers T. TET1 prefers a TA dinucleotide at the −2/−1 site, particularly with 5mC substrates. Yet, a TG at the same place is disfavored especially by TET2. Moreover, a TT dinucleotide is preferred at the +1/+2 site, more so by TET2 and with 5hmC substrates. It was also found that sites with a high genomic 5hmC/5mC ratio are preferentially observed within an A_{-1} and T_{+1} context, while sites with a low genomic 5hmC/5mC ratio are associated with G_{-1} and C_{+1} flanks.

Inspection of structures of TET can help justify these findings. For instance, while only a water-mediated contact is formed with a TA base pair at the +1 site in a TET2 structure (Hu et al. 2015), in another complex with a CG base pair at the +1 site (Hu et al. 2013), a conserved arginine residue makes a direct hydrogen bond to G(+1′) in the minor groove. In turn, the CG(+1) base pair is shifted by 1.5 Å in the direction of the helix axis which may increase the stacking of the target C base in the helix, making base flipping of the target residue more difficult, and explaining the disfavor for C at the +1 flank position. Interestingly, CACGTG appears to be the best sequence which also is the canonical E-box motif, a well-known recognition site for many helix-loop-helix (bHLH) and basic zipper leucine domain (bZIP) transcription factors (Ravichandran et al. 2021). In the case of MAX, a binding partner of the oncogenic transcription factor MYC, MAX exhibits the greatest affinity for a 5caCpG containing E-box, and much reduced affinities for the corresponding 5mC, 5hmC, or 5fC forms (Wang et al. 2017). In the case of TCF4, which binds the E-box element in the context of CG_0-CA_1-CG_2-TG_3 (where the numbers indicate successive dinucleotides), modification of the central CG_2 has very little effect on TCF4 binding, the CA_1 modification has a negative influence on binding, while modification of the flanking CG_0, particularly carboxylation, has a strong positive impact on TCF4 binding to DNA (Yang et al. 2019).

12.3.2 AlkB and Homologs

Similar to TET enzymes, eukaryotic homologs of *E. coli* AlkB (Fig. 12.6c) such as FTO and ALKBH5 are also αKG- and Fe(II)-dependent dioxygenases that can oxidize the methyl group of N6mA within mRNA to yield demethylated adenine (Jia et al. 2011; Zheng et al. 2013; Zhu and Yi 2014). Another ALKBH family member, ALKBH1, appears to be a DNA N6mA demethylase in mammals with preference of N6mA nucleotide within a bubbled or bulged DNA with flanking duplex stems (Zhang et al. 2020). Interestingly, the human METTL3–METTL14 heterocomplex N6mA MTase (Liu et al. 2014a) has been shown by in vitro methylation assays to methylate ssDNA and unpaired DNA regions with flanking duplex DNA, as well as having activity on dsDNA containing cyclopyrimidine dimers, which are the major UV radiation-induced photoproducts (Woodcock et al. 2019; Yu et al. 2021; Qi et al. 2022).

TET-DNA complex structures are remarkably comparable to that of the AlkB-DNA complex, and both TET and AlkB enzymes are Fe(II)- and aKG-dependent dioxygenases using base flipping for reactions (Hashimoto et al. 2014a; Hu et al. 2013; McDonough et al. 2010). Common structural folds include two twisted β-sheets in the core where the active site is formed (Fig. 12.6). However, the two enzymes differ in an important mechanistic aspect. TET enzymes oxidize CH_3 attached to an inert carbon atom (cytosine or thymine C5). The resulting product (5hmC or 5hmU) is very stable and can undergo further oxidations in subsequent rounds of reactions to generate further oxidized products. On the other hand, FTO and ALKBH1/5 likely generate N6-hydroxymethyladenine intermediate in which the oxidized carbon is attached to a reactive nitrogen atom (adenine N6). This intermediate spontaneously releases the hydroxymethyl group as formaldehyde and decomposes to

adenine—the final "demethylated" product (Hashimoto et al. 2015b) (Fig. 12.1b). Therefore, AlkB and its homologs are demethylases, while TET enzymes should not be characterized as demethylases, but would rather be appropriately understood as "writers" that generate additional modifications on 5mC within genomes to alter epigenetic signals.

Several biochemical observations suggest that modified cytosines beyond 5mC may form distinct epigenetic signals. Many 5mCpG readers such as methyl-CpG binding domain (MBD) proteins have shown significantly reduced binding affinity toward 5hmC when compared to 5mC within CpG context (Hashimoto et al. 2012b, 2015b), whereas some proteins may preferentially bind 5hmC (Zhou et al. 2014). DNMT1 has a significantly reduced activity toward hemi-hydroxymethylated DNA substrate compared to hemi-methylated DNA (Hashimoto et al. 2012b), suggesting that methylation marks altered by TET enzymes can be lost in subsequent DNA replications. In addition, the RNA polymerase II transcription rate can be specifically reduced by 5fC and 5caC (Kellinger et al. 2012; Wang et al. 2015). These findings strongly point to the possibility that modifications beyond 5mC are distinct signals, and future work is needed to elucidate how the modified bases are differently implicated in larger biological contexts. Several 5caC reader proteins have been structurally characterized in the context of specific sequences (Wang et al. 2017; Yang et al. 2019; Hashimoto et al. 2014b).

12.4 Base Flipping in the Recognition of Modified Bases

12.4.1 Eukaryotic SRA Domains

The function of 5mC and N6mA in prokaryotes was classically understood in the context of restriction-modification systems, in which methylated bacterial DNA is protected from restriction digestion (Wilson and Murray 1991). Effects of DNA methylation are fundamentally determined by the way the methyl groups alter various protein–DNA interactions. In eukaryotes, genomic 5mC bases are considered widely involved in various regulatory processes to control gene expression, chromatin states, and genomic stability that are highly relevant in the human disease context (Robertson 2005). Such penetrating biological implications can be partly attributed to the large number of protein–DNA interactions that are potentially affected by DNA methylation in a direct manner. Evidence shows that several transcription factors are prevented from DNA binding when their binding site is methylated (Tate and Bird 1993), whereas several MBD family proteins are specific 5mCpG readers, as previously mentioned (Klose and Bird 2006). Furthermore, a few 5caC readers have been recently characterized (Yang et al. 2020). The interface between modified DNA and its biological effects can be further complicated by the involvement of histone modifications which DNA methylation is profoundly associated with (Cedar and Bergman 2009; Hashimoto et al. 2010).

The initial discovery of 5mC-binding proteins has raised the possibility of other readers involved in modified base recognitions. In 2007, another family of 5mC readers was discovered in plants and was termed SET- and RING-associated (SRA) domain as a part of VIM1 (Woo et al. 2007). A mammalian homolog to VIM1 is UHRF1, which can associate with DNMT1 during post-replicative maintenance of DNA methylation (Bostick et al. 2007; Sharif et al. 2007). In the following year, three crystal structures of the mammalian UHRF1 SRA domain in complex with 5mC-containing DNA were reported (Arita et al. 2008; Avvakumov et al. 2008; Hashimoto et al. 2008). The structures have revealed that SRA recognizes 5mC by base flipping, although it is not a DNA-modifying enzyme such as MTases or dioxygenases. SRA is also structurally distinct from other base flippers and is characterized by a twisted β-sheet fold resembling a half-moon shape (Fig. 12.7a). Remarkably, the 5mC-binding pocket of SRA features familiar modes of base recognitions exemplified by π stacking interactions, recognitions of the N3 and N4 by Asp474 side chain, and a van der Waals'

contact of the C5-methyl group of flipped 5mC by Ser486 Cβ.

Interestingly, the SRA of UHRF2 binds 5hmC with a slightly higher preference compared to 5mC, and the crystal structure of UHRF2 SRA in complex with 5hmC-containing DNA is available (Zhou et al. 2014). In the structure, 5hmC is flipped and stabilized, and the OH moiety of the hydroxymethyl group is contacted by the backbone carbonyl groups of Thr508 and Gly509 in the active-site pocket which is slightly larger in size compared to that of UHRF1 SRA. Therefore, eukaryotic SRA has been characterized as a base-flipping domain that recognizes both 5mC and 5hmC.

In addition, SRA domains have been rediscovered in prokaryotes in families of modification-dependent restriction enzymes that recognize modified bases and introduce a double-stranded break in some distances away. MspJI was among the first such enzymes to be reported, which recognizes hemi-modified 5mC or 5hmC by the N-terminal SRA-like domain and restricts the DNA by the C-terminal endonuclease domain (Cohen-Karni et al. 2011). The crystal structure of MspJI has been solved with substrate DNA, revealing an SRA-like structure in the N-terminal modification recognition domain that flips the target 5mC (Fig. 12.7b) (Horton et al. 2014b). Despite the lack of amino acid sequence conservation between eukaryotic UHRF1/2 SRA and MspJI SRA, all SRA domains feature a twisted β-sheet fold with a half-moon shape.

As more modification-dependent restriction enzymes have been identified, some of them are found with different specificities toward 5mC, 5hmC, and 5ghmC. AbaSI, unlike MspJI, has an N-terminal Vsr-like endonuclease domain and a C-terminal SRA-like domain (Horton et al. 2014a; Borgaro and Zhu 2013). Its SRA domain seems to preferentially recognize 5ghmC and 5hmC compared to 5mC, as the relative rate of cleavage of DNA containing the corresponding modification is 5ghmC:5hmC:5mC=8000:500:1 (Wang et al. 2011). Structural features within SRA domains that fine-tune such specificities await future characterizations.

12.4.2 EcMcrB-N Homologs as 5mC and N6mA Readers

Modification-dependent restriction enzymes also utilize yet another 5mC recognition domain. The N-terminus of *Escherichia coli* McrB (EcMcrB-N) recognizes 5mC next to adenine within 5′-ACCGGT-3′ sequences, and McrC associates with McrB to provide endonuclease activity (Stewart et al. 2000). The crystal structure of EcMcrB-N in complex with 5mC-containing DNA shows a flipped 5mC in the active site (Fig. 12.7c), revealing a novel fold distinct from all other known base flippers (Sukackaite et al. 2012). The active site displays familiar π stacking of the flipped 5mC via aromatic residues and van der Waals' contact of the C5-methyl group via the side chain of Leu68.

The history of the discovery of base flippers suggests a strong possibility of its structural homologs present in a wide spectrum of phyla. EcMcrB-N is poorly conserved among the wide array of McrBC homologs and other domains exist in other homologs for binding modified bases. Indeed, *T. gammatolerans* N-McrB is structurally distinct from the EcMcrB DNA-binding domain, adopting a YTH domain fold commonly found in eukaryotic proteins (Wu et al. 2017). It binds and flips methylated base out of the DNA (Hosford et al. 2020) (Fig. 12.7d). Similarly, it has been observed that, in addition to N6mA in RNA, the N6mA reader domain of human YTHDC1 binds N6mA in either a single-stranded DNA or a single-base gap between two canonical DNA helices (Yu et al. 2021; Woodcock et al. 2020b) (Fig. 12.7e).

12.4.3 5mC and N6mA Readers Use Non-Base-Flipping Recognition

While base flipping seems to be a major mechanism by which a modified DNA base can be recognized, it should be noted that many DNA-binding proteins recognize modified bases in a sequence-dependent manner without

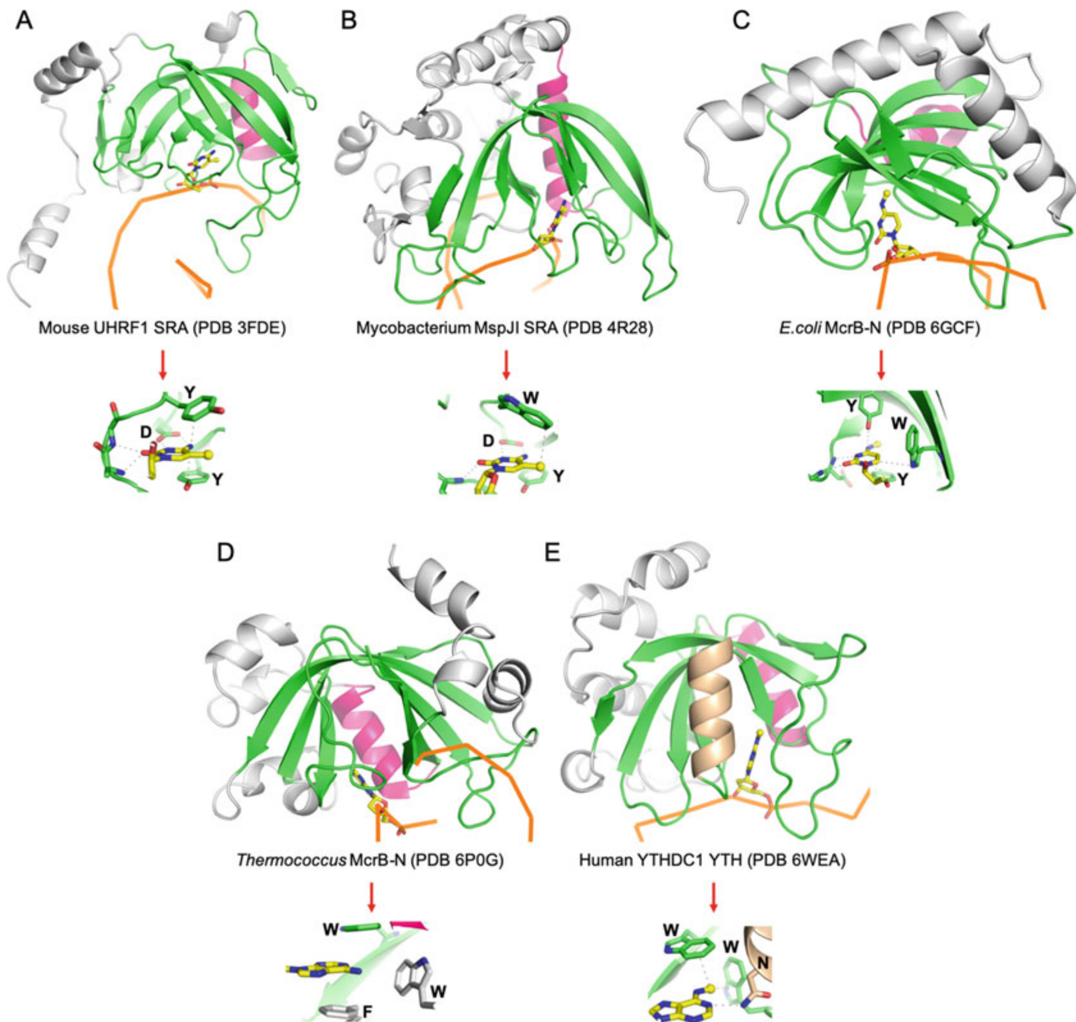


Fig. 12.7 Examples of reader domain proteins using base flipping. (**a**) Mouse UHRF1 SRA with 5mC, (**b**) MspJI SRA with 5mC, (**c**) *E. coli* McrB-N with N4mC, (**d**) *Thermococcus* McrB-N, and (**e**) human YTHDC1 YTH with N6mA. The conserved strands are in green and one conserved helix (colored in red) is behind the arch. The other helices are in gray. DNA strands are in ribbon with the flipped base in stick presentation. The modified bases are bound in a cage formed by 2–3 aromatic residues

involving base flipping. Along with previously mentioned MBD family proteins that recognize 5mC within the simple dinucleotide CpG sequence, mammalian DNA-binding proteins such as Kaiso (Buck-Koehntop et al. 2012), Zfp57 (Liu et al. 2012), Klf4 (Liu et al. 2014b), Egr1 (Hashimoto et al. 2014b; Zandarashvili et al. 2015), and AP-1 (Hong et al. 2017) bind 5mC within specific sequences via a common structural motif (Liu et al. 2013). In addition, transcription factors WT1 (Hashimoto et al. 2014b), MAX (Wang et al. 2017), Tcf3-Ascl1 heterodimer (Golla et al. 2014), and TCF4 (Yang et al. 2019) can specifically bind 5caC within their consensus sequences. In prokaryotes, DpnI harbors a C-terminal WH domain that recognizes the methyl group of N6mA within 5′-GATC-3′ sequence via Trp138 involving van der Waals' interactions (Mierzejewska et al. 2014). Therefore, DNA modifications may

regulate transcription-binding sites in much more dynamic and selective manners than previously understood.

12.5 Base Flipping in Removing Modified and Unmodified Bases

12.5.1 Mammalian Thymine DNA Glycosylase (TDG)

The discovery of TET-mediated modified cytosine bases has provided a fresh insight into a long-sought-after pathway of 5mC demethylation/demodification within mammalian genomes (Zhu 2009). In the base excision repair pathway, DNA glycosylases cleave the glycosidic bond between the ribose and the target base and represent the most structurally diverse family of base-flipping enzymes (Brooks et al. 2013). Initially, it was hypothesized that 5mC (and 5hmC) is removed by mammalian 5mC/5hmC DNA glycosylase activities (Vairapandi and Duker 1993; Cannon et al. 1988; Vairapandi et al. 2000; Vairapandi 2004). However, the glycosylase(s) involved was never identified. After the discovery of TET enzymes, mammalian TDG that generally removes uracil or thymine mismatched to guanine was surprisingly revealed to excise 5fC and 5caC to establish genome-wide DNA demethylation (He et al. 2011; Maiti and Drohat 2011; Hashimoto et al. 2012a). The crystal structure of the human TDG catalytic domain in complex with 5caC-containing DNA was solved (Fig. 12.8a), presenting the flipped base in the active site with an Asn140Ala mutation where the C5-carboxyl moiety of 5caC is specifically recognized by the side chain of Asn157 and the Tyr152 amide backbone (Zhang et al. 2012). In another crystal structure of TDG bound to DNA with a non-cleavable (2′-fluoroarabino) analog of 5-formyldeoxycytidine flipped into its active site (Pidugu et al. 2019), TDG provides a hydrogen bond (2.8 Å) from the Tyr152 backbone N–H to the 5fC formyl oxygen. A nucleophilic water molecule is bound by Asn140 and the backbone oxygen of Thr197, supporting an essential catalytic role for Asn140. The 1.6 Å high-resolution structures of TDG and its N140A mutant bound to DNA with 5caC flipped into the active site suggest that acid-catalyzed 5caC excision is facilitated by two water molecules and contact with Asn191, resulting in a protonated form of 5caC that would be ineffective for C, 5mC, or 5hmC (Pidugu et al. 2019). The discovery of TDG excising 5fC and 5caC has effectively linked the base excision repair pathway to DNA demethylation in mammalian systems.

12.5.2 Plant ROS1

In plants, paradoxically, bone fide 5mC DNA glycosylases were clearly demonstrated and identified in 2002 (Gong et al. 2002), approximately a decade before TET and TDG were implicated in DNA demethylation. In Arabidopsis, four closely related 5mC DNA glycosylases exist: ROS1, DME, DML2, and DML3 (Gehring et al. 2006; Morales-Ruiz et al. 2006; Ortega-Galisteo et al. 2008). They have a catalytic glycosylase domain homologous to *E. coli* endonuclease III (Fig. 12.8b), a helix-hairpin-helix (HhH) fold DNA glycosylase that harbors an iron-sulfur cluster-binding site (Ponferrada-Marin et al. 2009, 2010, 2011, 2012). Thus, plant ROS1 and mammalian TDG have mutually exclusive substrate specificities for 5mC, 5hmC, 5fC, and 5caC: the first two specific for ROS1 and the latter two for TDG (Hong et al. 2014; Hashimoto et al. 2012a) (as shown in Fig. 12.1a). One of the most surprising aspects of plant 5mC DNA glycosylases is that they excise the target base only when both the catalytic glycosylase domain and the C-terminal domain are present (Hong et al. 2014; Mok et al. 2010). The C-terminal domain of ROS1 is conserved only among plant 5mC DNA glycosylases and has been shown to strongly associate with the catalytic domain, suggesting that domain–domain interactions are important for target base recognition and excision (Hong et al. 2014).

While TDG and ROS1 have been clearly implicated in DNA demethylation pathways,

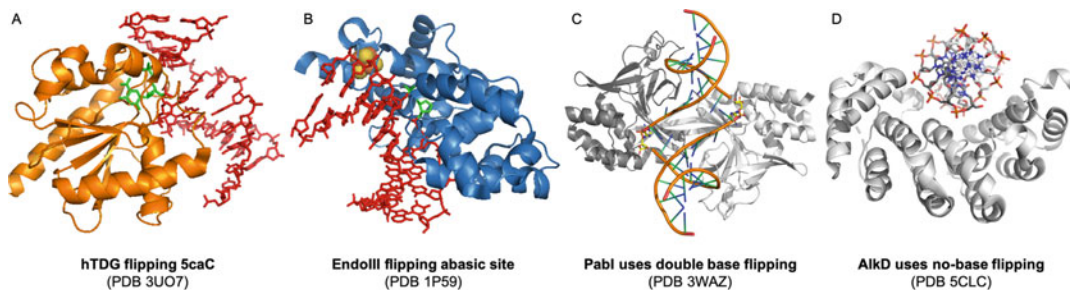


Fig. 12.8 Examples of DNA glycosylases. (**a**) Human TDG flipping 5caC opposite guanine. (**b**) *Geobacillus stearothermophilus* Endonuclease III in complex with DNA. Iron-sulfur cluster is colored in orange and yellow. (**c**) *Pyrococcus abyssi* PabI uses double base flipping. (**d**) *Bacillus cereus* AlkD uses no-base flipping

jury is still out on the possibility of other pathways to DNA demethylation. In addition to the previously mentioned mammalian 5mC DNA glycosylase activities, 5hmC DNA glycosylase activity was observed in a calf thymus extract (Cannon et al. 1988). A proteomic study has revealed that several mammalian DNA glycosylases such as NTH1, OGG1, NEIL1, and NEIL2 bind 5mC- and 5hmC-containing DNA in a modification-specific manner (Spruijt et al. 2013), though they by themselves do not have the glycosylase activity against 5mC or 5hmC (Hong et al. 2014).

The 5mC DNA glycosylase activity by ROS1 is interesting from a standpoint of historical characterization of DNA glycosylases as DNA damage repair enzymes. In a given genome, there can be many types of damaged bases, and their diversity is on par with many classes of DNA glycosylases that are structurally distinct (Brooks et al. 2013). On the other hand, 5mC in plants is not considered a damaged base and exists in substantial amounts in the *Arabidopsis* genome (Zhang et al. 2006). Thus, ROS1 must be regulated and specifically targeted to a certain genomic location to initiate DNA demethylation (Zheng et al. 2008; Qian et al. 2012). In addition to 5mC, ROS1 is comparably active on thymine mismatched to guanine and on some damaged pyrimidines, suggesting that ROS1 can be involved in both DNA demethylation and DNA damage repair (Ponferrada-Marin et al. 2009, 2010). Such dual functionality can be applied to TDG, which not only excises thymine or uracil mismatched to G during the process of DNA mismatch repair, but also excises 5fC and 5caC base-paired with guanine for DNA demodification in mammals.

12.5.3 Archaeon PabI Activity as Adenine DNA Glycosylase

Interestingly, the archaeal *Pyrococcus abyssi* PabI enzyme was initially thought to be a restriction endonuclease but has been re-characterized as a sequence-specific adenine DNA glycosylase (Miyazono et al. 2014, 2020). PabI is comparable to MutY family mismatch repair DNA glycosylases that excise target adenine mismatched to 8-oxoguanine (Fromme et al. 2004). However, PabI is remarkably distinct from MutY, because PabI excises adenine correctly base-paired to thymine in a targeted manner (Fig. 12.8c). It is therefore possible that DNA glycosylases have adapted to function in more processes than DNA damage repair by removing benign bases for various biological regulations.

12.6 Conclusions

First observed in 1994 in the crystal structure of M.HhaI with DNA, base flipping is now understood as a common mode of protein–DNA/RNA interactions adopted by structurally and functionally distinct classes of proteins across various

phyla. Base flipping is the only known mechanism for establishing DNA modifications in a targeted manner via DNA MTases and TET dioxygenases. What used to be considered a eukaryote-specific base-flipping SRA 5mC reader has later been shown to be a widely prevalent domain in prokaryotic systems for recognizing several modified bases including 5mC, 5hmC, and 5ghmC. In addition to SRA, more structurally diverse classes of modified base readers have been discovered in prokaryotes, such as the base-flipping McrB-N 5mC reader and the N6mA-recognizing WH domain of DpnI (using non-base-flipping mechanism). Also, DNA glycosylases are base flippers primarily characterized as DNA repair enzymes, though not all DNA glycosylases flip a base/nucleotide for base excision, as presented in the example of bacterial AlkD (Mullins et al. 2015) (Fig. 12.8d). Today, DNA demodification is considered a bone fide output of the base excision repair pathway through DNA glycosylases, such as mammalian TDG and plant ROS1 whose mechanism of action again involves base flipping. In an era in which DNA modifications are considered critical and increasingly complex epigenetic signals, this simple but elegant structural mechanism for protein–DNA interaction is preserved as a truly ubiquitous framework.

Acknowledgments We thank members of the Cheng laboratory during the last three decades of research on DNA base flipping. The work in the authors' laboratory is currently supported by grant from the National Institutes of Health (R35GM134744) and the Cancer Prevention and Research Institute of Texas (CPRIT RR160029). X.C. is a CPRIT Scholar in Cancer Research.

References

Adam S, Anteneh H, Hornisch M, Wagner V, Lu J, Radde NE, Bashtrykov P, Song J, Jeltsch A (2020) DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. Nat Commun 11:3723

Adam S, Bracker J, Klingel V, Osteresch B, Radde NE, Brockmeyer J, Bashtrykov P, Jeltsch A (2022) Flanking sequences influence the activity of TET1 and TET2 methylcytosine dioxygenases and affect genomic 5hmC patterns. Commun Biol 5:92

Anton BP, Roberts RJ (2021) Beyond restriction modification: epigenomic roles of DNA methylation in prokaryotes. Annu Rev Microbiol 75:129–149

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature 455:818–821

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, Arrowsmith CH, Dhe-Paganon S (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature 455:822–825

Beh LY, Debelouchina GT, Clay DM, Thompson RE, Lindblad KA, Hutton ER, Bracht JR, Sebra RP, Muir TW, Landweber LF (2019) Identification of a DNA N6-adenine methyltransferase complex and its impact on chromatin organization. Cell 177:1781–1796

Bestor TH (2000) The DNA methyltransferases of mammals. Hum Mol Genet 9:2395–2402

Borgaro JG, Zhu Z (2013) Characterization of the 5-hydroxymethylcytosine-specific DNA restriction endonucleases. Nucleic Acids Res 41:4198–4206

Borst P, Sabatini R (2008) Base J: discovery, biosynthesis, and possible functions. Annu Rev Microbiol 62:235–251

Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317:1760–1764

Brooks SC, Adhikary S, Rubinson EH, Eichman BF (2013) Recent advances in the structural mechanisms of DNA glycosylases. Biochim Biophys Acta 1834: 247–271

Buck-Koehntop BA, Stanfield RL, Ekiert DC, Martinez-Yamout MA, Dyson HJ, Wilson IA, Wright PE (2012) Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. Proc Natl Acad Sci U S A 109:15229–15234

Cannon SV, Cummings A, Teebor GW (1988) 5-Hydroxymethylcytosine DNA glycosylase activity in mammalian tissue. Biochem Biophys Res Commun 151:1173–1179

Cedar H, Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet 10:295–304

Cheng X (1995) DNA modification by methyltransferases. Curr Opin Struct Biol 5:4–10

Cheng X, Roberts RJ (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping. Nucleic Acids Res 29:3784–3795

Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ (1993) Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell 74:299–307

Cohen-Karni D, Xu D, Apone L, Fomenkov A, Sun Z, Davis PJ, Kinney SR, Yamada-Mabuchi M, Xu SY, Davis T et al (2011) The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. Proc Natl Acad Sci U S A 108:11040–11045

Duymich CE, Charlet J, Yang X, Jones PA, Liang G (2016) DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. Nat Commun 7:11453

Fromme JC, Banerjee A, Huang SJ, Verdine GL (2004) Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. Nature 427:652–656

Gao L, Emperle M, Guo Y, Grimm SA, Ren W, Adam S, Uryu H, Zhang ZM, Chen D, Yin J et al (2020) Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. Nat Commun 11:3355

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL (2006) DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell 124:495–506

Goedecke K, Pignot M, Goody RS, Scheidig AJ, Weinhold E (2001) Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. Nat Struct Biol 8:121–125

Golla JP, Zhao J, Mann IK, Sayeed SK, Mandal A, Rose RB, Vinson C (2014) Carboxylation of cytosine (5caC) in the CG dinucleotide in the E-box motif (CGCAG|GTG) increases binding of the Tcf3|Ascl1 helix-loop-helix heterodimer 10-fold. Biochem Biophys Res Commun 449:248–255

Gommers-Ampt JH, Van Leeuwen F, de Beer AL, Vliegenthart JF, Dizdaroglu M, Kowalak JA, Crain PF, Borst P (1993) beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan T. brucei. Cell 75:1129–1136

Gong Z, Morales-Ruiz T, Ariza RR, Roldan-Arjona T, David L, Zhu JK (2002) ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell 111:803–814

Gowher H, Jeltsch A (2001) Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. J Mol Biol 309:1201–1208

Gupta YK, Chan SH, Xu SY, Aggarwal AK (2015) Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. Nat Commun 6:7363

Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature 455: 826–829

Hashimoto H, Vertino PM, Cheng X (2010) Molecular coupling of DNA methylation and histone methylation. Epigenomics 2:657–669

Hashimoto H, Hong S, Bhagwat AS, Zhang X, Cheng X (2012a) Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. Nucleic Acids Res 40:10203–10214

Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM, Zhang X, Cheng X (2012b) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res 40:4841–4849

Hashimoto H, Pais JE, Zhang X, Saleh L, Fu ZQ, Dai N, Correa IR Jr, Zheng Y, Cheng X (2014a) Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. Nature 506:391–395

Hashimoto H, Olanrewaju YO, Zheng Y, Wilson GG, Zhang X, Cheng X (2014b) Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. Genes Dev 28:2304–2313

Hashimoto H, Pais JE, Dai N, Correa IR Jr, Zhang X, Zheng Y, Cheng X (2015a) Structure of Naegleria Tet-like dioxygenase (NgTet1) in complexes with a reaction intermediate 5-hydroxymethylcytosine DNA. Nucleic Acids Res 43:10713–10721

Hashimoto H, Zhang X, Vertino PM, Cheng X (2015b) The mechanisms of generation, recognition, and erasure of DNA 5-methylcytosine and thymine oxidations. J Biol Chem 290:20723–20733

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

Hong S, Hashimoto H, Kow YW, Zhang X, Cheng X (2014) The carboxy-terminal domain of ROS1 is essential for 5-methylcytosine DNA glycosylase activity. J Mol Biol 426:3703–3712

Hong S, Wang D, Horton JR, Zhang X, Speck SH, Blumenthal RM, Cheng X (2017) Methyl-dependent and spatial-specific DNA recognition by the orthologous transcription factors human AP-1 and Epstein-Barr virus Zta. Nucleic Acids Res 45:2503–2515

Horton JR, Liebert K, Hattman S, Jeltsch A, Cheng X (2005) Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of dam methyltransferase. Cell 121:349–361

Horton JR, Borgaro JG, Griggs RM, Quimby A, Guan S, Zhang X, Wilson GG, Zheng Y, Zhu Z, Cheng X (2014a) Structure of 5-hydroxymethylcytosine-specific restriction enzyme, AbaSI, in complex with DNA. Nucleic Acids Res 42:7947–7959

Horton JR, Wang H, Mabuchi MY, Zhang X, Roberts RJ, Zheng Y, Wilson GG, Cheng X (2014b) Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix. Nucleic Acids Res 42:12092–12101

Horton JR, Nugent RL, Li A, Mabuchi MY, Fomenkov A, Cohen-Karni D, Griggs RM, Zhang X, Wilson GG, Zheng Y et al (2014c) Structure and mutagenesis of the DNA modification-dependent restriction endonuclease AspBHI. Sci Rep 4:4246

Horton JR, Woodcock CB, Opot SB, Reich NO, Zhang X, Cheng X (2019) The cell cycle-regulated DNA adenine

methyltransferase CcrM opens a bubble at its DNA recognition site. Nat Commun 10:4600

Horton JR, Yang J, Zhang X, Petronzio T, Fomenkov A, Wilson GG, Roberts RJ, Cheng X (2020) Structure of HhaI endonuclease with cognate DNA at an atomic resolution of 1.0 A. Nucleic Acids Res 48:1466–1478

Hosford CJ, Bui AQ, Chappie JS (2020) The structure of the Thermococcus gammatolerans McrB N-terminal domain reveals a new mode of substrate recognition and specificity among McrB homologs. J Biol Chem 295:743–756

Hu L, Li Z, Cheng J, Rao Q, Gong W, Liu M, Shi YG, Zhu J, Wang P, Xu Y (2013) Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. Cell 155:1545–1555

Hu L, Lu J, Cheng J, Rao Q, Li Z, Hou H, Lou Z, Zhang L, Li W, Gong W et al (2015) Structural insight into substrate preference for TET-mediated oxidation. Nature 527:118–122

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333:1300–1303

Iyer LM, Zhang D, Maxwell Burroughs A, Aravind L (2013) Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. Nucleic Acids Res 41:7635–7655

Jang H, Shin H, Eichman BF, Huh JH (2014) Excision of 5-hydroxymethylcytosine by DEMETER family DNA glycosylases. Biochem Biophys Res Commun 446: 1067–1072

Jeltsch A, Adam S, Dukatz M, Emperle M, Bashtrykov P (2021) Deep enzymology studies on DNA methyltransferases reveal novel connections between flanking sequences and enzyme activity. J Mol Biol 433:167186

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449: 248–251

Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, Yi C, Lindahl T, Pan T, Yang YG et al (2011) N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol 7: 885–887

Kellinger MW, Song CX, Chong J, Lu XY, He C, Wang D (2012) 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. Nat Struct Mol Biol 19:831–833

Klimasauskas S, Kumar S, Roberts RJ, Cheng X (1994) HhaI methyltransferase flips its target base out of the DNA helix. Cell 76:357–369

Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci 31:89–97

Konttinen O, Carmody J, Pathuri S, Anderson K, Zhou X, Reich N (2020) Cell cycle regulated DNA methyltransferase: fluorescent tracking of a DNA strand-separation mechanism and identification of the responsible protein motif. Nucleic Acids Res 48: 11589–11601

Kornberg SR, Zimmerman SB, Kornberg A (1961) Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected Escherichia coli. J Biol Chem 236:1487–1493

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

Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, Roberts RJ, Wilson GG (1994) The DNA (cytosine-5) methyltransferases. Nucleic Acids Res 22:1–10

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11:204–220

Lehman IR, Pratt EA (1960) On the structure of the glucosylated hydroxymethylcytosine nucleotides of coliphages T2, T4, and T6. J Biol Chem 235:3254–3259

Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69:915–926

Lin CC, Chen YP, Yang WZ, Shen JCK, Yuan HS (2020) Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. Nucleic Acids Res 48:3949–3961

Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science 292:2077–2080

Liu Y, Toh H, Sasaki H, Zhang X, Cheng X (2012) An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. Genes Dev 26: 2374–2379

Liu Y, Zhang X, Blumenthal RM, Cheng X (2013) A common mode of recognition for methylated CpG. Trends Biochem Sci 38:177–183

Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, Jia G, Yu M, Lu Z, Deng X et al (2014a) A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol 10: 93–95

Liu Y, Olanrewaju YO, Zheng Y, Hashimoto H, Blumenthal RM, Zhang X, Cheng X (2014b) Structural basis for Klf4 recognition of methylated DNA. Nucleic Acids Res 42:4859–4867

Lunt MR, Siebke JC, Burton K (1964) Glucosylated nucleotide sequences from T2-bacteriophage deoxyribonucleic acid. Biochem J 92:27–36

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

Malone T, Blumenthal RM, Cheng X (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases,

and suggests a catalytic mechanism for these enzymes. J Mol Biol 253:618–632
McDonough MA, Loenarz C, Chowdhury R, Clifton IJ, Schofield CJ (2010) Structural studies on human 2-oxoglutarate dependent oxygenases. Curr Opin Struct Biol 20:659–672
Messer W, Noyer-Weidner M (1988) Timing and targeting: the biological functions of Dam methylation in E. coli. Cell 54:735–737
Mierzejewska K, Siwek W, Czapinska H, Kaus-Drobek M, Radlinska M, Skowronek K, Bujnicki JM, Dadlez M, Bochtler M (2014) Structural basis of the methylation specificity of R.DpnI. Nucleic Acids Res 42:8745–8754
Miyazono K, Furuta Y, Watanabe-Matsui M, Miyakawa T, Ito T, Kobayashi I, Tanokura M (2014) A sequence-specific DNA glycosylase mediates restriction-modification in Pyrococcus abyssi. Nat Commun 5:3178
Miyazono KI, Wang D, Ito T, Tanokura M (2020) Distortion of double-stranded DNA structure by the binding of the restriction DNA glycosylase R.PabI. Nucleic Acids Res 48:5106–5118
Mok YG, Uzawa R, Lee J, Weiner GM, Eichman BF, Fischer RL, Huh JH (2010) Domain structure of the DEMETER 5-methylcytosine DNA glycosylase. Proc Natl Acad Sci U S A 107:19225–19230
Morales-Ruiz T, Ortega-Galisteo AP, Ponferrada-Marin MI, Martinez-Macias MI, Ariza RR, Roldan-Arjona T (2006) DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. Proc Natl Acad Sci U S A 103:6853–6858
Mullins EA, Shi R, Parsons ZD, Yuen PK, David SS, Igarashi Y, Eichman BF (2015) The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions. Nature 527:254–258
Murray IA, Morgan RD, Luyten Y, Fomenkov A, Correa IR Jr, Dai N, Allaw MB, Zhang X, Cheng X, Roberts RJ (2018) The non-specific adenine DNA methyltransferase M.EcoGII. Nucleic Acids Res 46:840–848
Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99:247–257
Oliveira PH, Ribis JW, Garrett EM, Trzilova D, Kim A, Sekulovic O, Mead EA, Pak T, Zhu S, Deikus G et al (2020) Epigenomic characterization of Clostridioides difficile finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis. Nat Microbiol 5:166–180
Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin S-P, Allis CD et al (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448:714–717
Ortega-Galisteo AP, Morales-Ruiz T, Ariza RR, Roldan-Arjona T (2008) Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. Plant molecular biology 67:671–681
Pais JE, Dai N, Tamanaha E, Vaisvila R, Fomenkov AI, Bitinaite J, Sun Z, Guan S, Correa IR Jr, Noren CJ et al (2015) Biochemical characterization of a Naegleria TET-like oxygenase and its application in single molecule sequencing of 5-methylcytosine. Proc Natl Acad Sci U S A 112:4316–4321
Pappalardi MB, Keenan K, Cockerill M, Kellner WA, Stowell A, Sherk C, Wong K, Pathuri S, Briand J, Steidel M et al (2021) Discovery of a first-in-class reversible DNMT1-selective inhibitor with improved tolerability and efficacy in acute myeloid leukemia. Nat Cancer 2:1002–1017
Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, Hagemeier C, Carell T (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. Angewandte Chemie 50:7008–7012
Pidugu LS, Dai Q, Malik SS, Pozharski E, Drohat AC (2019) Excision of 5-carboxylcytosine by thymine DNA glycosylase. J Am Chem Soc 141:18851–18861
Ponferrada-Marin MI, Roldan-Arjona T, Ariza RR (2009) ROS1 5-methylcytosine DNA glycosylase is a slow-turnover catalyst that initiates DNA demethylation in a distributive fashion. Nucleic Acids Res 37:4264–4274
Ponferrada-Marin MI, Martinez-Macias MI, Morales-Ruiz T, Roldan-Arjona T, Ariza RR (2010) Methylation-independent DNA binding modulates specificity of Repressor of Silencing 1 (ROS1) and facilitates demethylation in long substrates. J Biol Chem 285:23032–23039
Ponferrada-Marin MI, Parrilla-Doblas JT, Roldan-Arjona-T, Ariza RR (2011) A discontinuous DNA glycosylase domain in a family of enzymes that excise 5-methylcytosine. Nucleic Acids Res 39:1473–1484
Ponferrada-Marin MI, Roldan-Arjona T, Ariza RR (2012) Demethylation initiated by ROS1 glycosylase involves random sliding along DNA. Nucleic Acids Res 40:11554–11562
Qi S, Mota J, Chan SH, Villarreal J, Dai N, Arya S, Hromas RA, Rao MK, Correa IR Jr, Gupta YK (2022) RNA binding to human METTL3-METTL14 restricts N(6)-deoxyadenosine methylation of DNA in vitro. eLife:11
Qian W, Miki D, Zhang H, Liu Y, Zhang X, Tang K, Kan Y, La H, Li X, Li S et al (2012) A histone acetyltransferase regulates active DNA demethylation in Arabidopsis. Science 336:1445–1448
Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A 97:5237–5242
Ravichandran M, Rafalski D, Ortega-Recalde O, Davies CI, Glanfield CR, Kotter A, Misztal K, Wang AH, Wojciechowski M, Rażew M et al (2021) Pronounced sequence specificity of the TET enzyme catalytic domain guides its cellular function. bioRxiv. https://doi.org/10.1101/2021.12.29.474486

Reich NO, Dang E, Kurnik M, Pathuri S, Woodcock CB (2018) The highly specific, cell cycle-regulated methyltransferase from Caulobacter crescentus relies on a novel DNA recognition mechanism. J Biol Chem 293:19038–19046

Roberts RJ, Myers PA, Morrison A, Murray K (1976) A specific endonuclease from Haemophilus haemolyticus. J Mol Biol 103:199–208

Roberts RJ, Vincze T, Posfai J, Macelis D (2015) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 43:D298–D299

Robertson KD (2005) DNA methylation and human disease. Nat Rev Genet 6:597–610

Schubert HL, Blumenthal RM, Cheng X (2003) Many paths to methyltransfer: a chronicle of convergence. Trends Biochem Sci 28:329–335

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

Song J, Rechkoblit O, Bestor TH, Patel DJ (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science 331:1036–1040

Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, Munzel M, Wagner M, Muller M, Khan F et al (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell 152:1146–1159

Stephens C, Reisenauer A, Wright R, Shapiro L (1996) A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. Proc Natl Acad Sci U S A 93: 1210–1214

Stewart FJ, Panne D, Bickle TA, Raleigh EA (2000) Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. J Mol Biol 298:611–622

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol 21:64–72

Sukackaite R, Grazulis S, Tamulaitis G, Siksnys V (2012) The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. Nucleic Acids Res 40:7552–7562

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L et al (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324:930–935

Tate PH, Bird AP (1993) Effects of DNA methylation on DNA-binding proteins and gene expression. Curr Opin Genet Dev 3:226–231

Vairapandi M (2004) Characterization of DNA demethylation in normal and cancerous cell lines and the regulatory role of cell cycle proteins in human DNA demethylase activity. J Cell Biochem 91:572–583

Vairapandi M, Duker NJ (1993) Enzymic removal of 5-methylcytosine from DNA by a human DNA-glycosylase. Nucleic Acids Res 21:5323–5327

Vairapandi M, Liebermann DA, Hoffman B, Duker NJ (2000) Human DNA-demethylating activity: a glycosylase associated with RNA and PCNA. J Cell Biochem 79:249–260

Wang H, Guan S, Quimby A, Cohen-Karni D, Pradhan S, Wilson G, Roberts RJ, Zhu Z, Zheng Y (2011) Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. Nucleic Acids Res 39:9294–9305

Wang L, Zhou Y, Xu L, Xiao R, Lu X, Chen L, Chong J, Li H, He C, Fu XD et al (2015) Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. Nature 523:621–625

Wang D, Hashimoto H, Zhang X, Barwick BG, Lonial S, Boise LH, Vertino PM, Cheng X (2017) MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. Nucleic Acids Res 45:2396–2407

Warren RA (1980) Modified bases in bacteriophage DNAs. Annu Rev Microbiol 34:137–158

Wei J, He C (2021) Chromatin and transcriptional regulation by reversible RNA methylation. Curr Opin Cell Biol 70:109–115

Wilson GG, Murray NE (1991) Restriction and modification systems. Annu Rev Genet 25:585–627

Woo HR, Pontes O, Pikaard CS, Richards EJ (2007) VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes Dev 21: 267–277

Woodcock CB, Yu D, Hajian T, Li J, Huang Y, Dai N, Correa IR Jr, Wu T, Vedadi M, Zhang X et al (2019) Human MettL3-MettL14 complex is a sequence-specific DNA adenine methyltransferase active on single-strand and unpaired DNA in vitro. Cell Discov 5:63

Woodcock CB, Horton JR, Zhang X, Blumenthal RM, Cheng X (2020a) Beta class amino methyltransferases from bacteria to humans: evolution and structural consequences. Nucleic Acids Res 48(18): 10034–10044

Woodcock CB, Horton JR, Zhou J, Bedford MT, Blumenthal RM, Zhang X, Cheng X (2020b) Biochemical and structural basis for YTH domain of human YTHDC1 binding to methylated adenine in DNA. Nucleic Acids Res 48:10329–10341

Wu JC, Santi DV (1985) On the mechanism and inhibition of DNA cytosine methyltransferases. Prog Clin Biol Res 198:119–129

Wu JC, Santi DV (1987) Kinetic and catalytic mechanism of HhaI methyltransferase. J Biol Chem 262:4778–4786

Wu B, Li L, Huang Y, Ma J, Min J (2017) Readers, writers and erasers of N(6)-methylated adenosine modification. Curr Opin Struct Biol 47:67–76

Wyatt GR, Cohen SS (1952) A new pyrimidine base from bacteriophage nucleic acids. Nature 170:1072–1073

Wyatt GR, Cohen SS (1953) The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. Biochem J 55:774–782

Xu TH, Liu M, Zhou XE, Liang G, Zhao G, Xu HE, Melcher K, Jones PA (2020) Structure of nucleosome-bound DNA methyltransferases DNMT3A and DNMT3B. Nature 586:151–155

Yang J, Horton JR, Li J, Huang Y, Zhang X, Blumenthal RM, Cheng X (2019) Structural basis for preferential binding of human TCF4 to DNA containing 5-carboxylcytosine. Nucleic Acids Res 47:8375–8387

Yang J, Zhang X, Blumenthal RM, Cheng X (2020) Detection of DNA modifications by sequence-specific transcription factors. J Mol Biol 432:1661–1673

Ye F, Kong X, Zhang H, Liu Y, Shao Z, Jin J, Cai Y, Zhang R, Li L, Zhang YW et al (2018) Biochemical studies and molecular dynamic simulations reveal the molecular basis of conformational changes in DNA methyltransferase-1. ACS Chem Biol 13:772–781

Yu D, Horton JR, Yang J, Hajian T, Vedadi M, Sagum CA, Bedford MT, Blumenthal RM, Zhang X, Cheng X (2021) Human MettL3-MettL14 RNA adenine methyltransferase complex is active on double-stranded DNA containing lesions. Nucleic Acids Res 49:11629–11642

Zandarashvili L, White MA, Esadze A, Iwahara J (2015) Structural impact of complete CpG methylation within target DNA on specific complex formation of the inducible transcription factor Egr-1. FEBS Lett 589: 1748–1753

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153:193–205

Zeng Y, Ren R, Kaur G, Hardikar S, Ying Z, Babcock L, Gupta E, Zhang X, Chen T, Cheng X (2020) The inactive Dnmt3b3 isoform preferentially enhances Dnmt3b-mediated DNA methylation. Genes Dev 34: 1546–1558

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE et al (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. Cell 126:1189–1201

Zhang L, Lu X, Lu J, Liang H, Dai Q, Xu GL, Luo C, Jiang H, He C (2012) Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nat Chem Biol 8:328–330

Zhang M, Yang S, Nelakanti R, Zhao W, Liu G, Li Z, Liu X, Wu T, Xiao A, Li H (2020) Mammalian ALKBH1 serves as an N(6)-mA demethylase of unpairing DNA. Cell Res 30:197–210

Zheng X, Pontes O, Zhu J, Miki D, Zhang F, Li WX, Iida K, Kapoor A, Pikaard CS, Zhu JK (2008) ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis. Nature 455:1259–1262

Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, Li CJ, Vagbo CB, Shi Y, Wang WL, Song SH et al (2013) ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell 49:18–29

Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, Chory J, Wohlschlegel JA, Patel DJ, Jacobsen SE (2014) Molecular mechanism of action of plant DRM de novo DNA methyltransferases. Cell 157:1050–1060

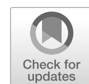
Zhou T, Xiong J, Wang M, Yang N, Wong J, Zhu B, Xu RM (2014) Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. Mol Cell 54:879–886

Zhou J, Horton JR, Blumenthal RM, Zhang X, Cheng X (2021) Clostridioides difficile specific DNA adenine methyltransferase CamA squeezes and flips adenine out of DNA helix. Nat Commun 12:3436

Zhu JK (2009) Active DNA demethylation mediated by DNA glycosylases. Annu Rev Genet 43:143–166

Zhu C, Yi C (2014) Switching demethylation activities between AlkB family RNA/DNA demethylases through exchange of active-site residues. Angewandte Chemie 53:3659–3662

Zweiger G, Marczynski G, Shapiro L (1994) A Caulobacter DNA methyltransferase that functions only in the predivisional cell. J Mol Biol 235:472–485



13 The Role of DNA Methylation and DNA Methyltransferases in Cancer

Daniel J. Weisenberger, Ranjani Lakshminarasimhan, and Gangning Liang

Abstract

The malignant transformation of normal cells is driven by both genetic and epigenetic changes. With the advent of next-generation sequencing and large-scale international consortia, it is now possible to profile the genomes and epigenomes of thousands of primary tumors from nearly every cancer type. These studies clearly demonstrate that the dynamic regulation of DNA methylation is a critical epigenetic mechanism of cancer initiation, maintenance, and progression. Proper control of DNA methylation is not only crucial for regulating gene transcription and tissue-specific cellular functions, but its broader consequences include maintaining the integrity of the genome and modulating the immune response. Here, we describe the aberrant DNA methylation changes in human cancers and how they contribute to the disease phenotypes. Aside from CpG island promoter DNA hypermethylation-based gene silencing, human cancers also display gene body DNA hypomethylation that is also associated with downregulated gene expression. In addition, the implementation of whole genome bisulfite sequencing (WGBS) has unveiled DNA hypomethylation of large blocks of the genome, known as partially methylated domains (PMDs), as well as cancer-specific DNA methylation aberrancies at enhancers and super-enhancers. Integrating WGBS and DNA methylation array data with mutation, copy number, and gene expression data has allowed for the identification of novel tumor suppressor genes and candidate driver genes of the disease state. Finally, we highlight potential clinical implications of these changes in the context of prognostic and diagnostic biomarkers, as well as therapeutic targets. Mounting evidence shows that DNA methylation data are effective and highly-sensitive disease classifiers, not only from analyses of the primary tumor but also from tumor-derived, cell free DNA (cfDNA) in blood of cancer patients. These findings highlight the power of DNA methylation aberrancies in providing efficacious biomarkers for clinical utility in improving patient diagnostics and their reversal using DNA methylation inhibitors in

D. J. Weisenberger
Department of Biochemistry and Molecular Medicine, University of Southern California, Norris Comprehensive Cancer Center, Los Angeles, CA, USA
e-mail: dan.weisenberger@med.usc.edu

R. Lakshminarasimhan
Department of Urology, University of Southern California, Norris Comprehensive Cancer Center, Los Angeles, CA, USA

Guardian Health, San Jose, CA, USA
e-mail: ranjanil@usc.edu

G. Liang (✉)
Department of Urology, University of Southern California, Norris Comprehensive Cancer Center, Los Angeles, CA, USA
e-mail: gliang@usc.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_13

cancer treatment may be key in surveillance, treatment, and quality of life for cancer patients.

Keywords

DNMTs · DNA methylation · Cancer epigenetics · Hypomethylation · Hypermethylation · cfDNA · Liquid biopsy · Biomarkers · DNA methylation inhibitors

Abbreviations

5-Aza-CR	5-azacytidine
5-Aza-CdR	5-Aza-2′deoxycytidine
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	The carbon-5 atom of cytosine
AML	Acute myeloid leukemia
ccRCC	Clear cell renal cell carcinoma
cfDNA	Cell free DNA
CGIs	CpG islands
CIMP	CpG island methylator phenotype
CMS	Consensus molecular subgroups
CpG	Cytosine-guanine dinucleotide
DNMT	DNA methyltransferases
DNMTi	DNA methyltransferase inhibitor
dsRNA	Double-stranded RNA
ENCODE	Encyclopedia of DNA Elements
EOC	Epithelial ovarian carcinoma
ERV	Endogenous retrovirus
FFPE	Paraffin embedded
GBM	Glioblastoma multiforme
H3K27M	Histone H3 lysine 27 to methionine
H3K27me3	Histone H3 lysine 27 trimethylation
Hypermutation	High global somatic mutation rates
ICR	Imprinting control region
IGF2	Insulin-like growth factor 2
LADs	Laminin-associated domains
LUAD	Lung adenocarcinomas
MAVS	Mitochondrial antiviral signaling
MDS	Myelodysplastic syndrome
MGMT	*O*-6-Methylguanine-DNA methyltransferase
MSI-H	High microsatellite instability
MSS	Microsatellite stable
NSCLC	Non-small cell lung cancer
pHGGs	Pediatric high-grade gliomas
TCGA	The Cancer Genome Atlas
TDG	Thymine DNA glycosylase
TET	Ten-eleven translocation
TMZ	Temozolomide
TSGs	Tumor suppressor genes
TSS	Transcription Start Site

13.1 Overview of Genetic and Epigenetic Alterations in Human Cancers

Classic hallmarks of cancer, as described by Hanahan and Weinberg, include maintenance of cell proliferation, evasion of growth suppression and cell death, promotion of angiogenesis, invasion, and metastasis (Hanahan and Weinberg 2011). Genetic and epigenetic alterations underlie these processes. Genetic changes contributing to tumorigenesis have been well studied and include DNA missense mutations, copy number variations, insertions, deletions, and recombination events. Complementary to genetic alterations, it is now generally accepted that oncogenic traits also accumulate through epigenetic disturbances (Baylin and Jones 2011; Sandoval and Esteller 2012), including changes of DNA methylation, histone tail modifications, nucleosome positioning, and noncoding RNAs. Gene promoter DNA methylation may be negatively related to gene activation status, while gene body DNA methylation is often positively related to gene expression, both of which are contributing mechanisms of gene regulation (Fig. 13.1a, b) (Liang and Weisenberger 2017;

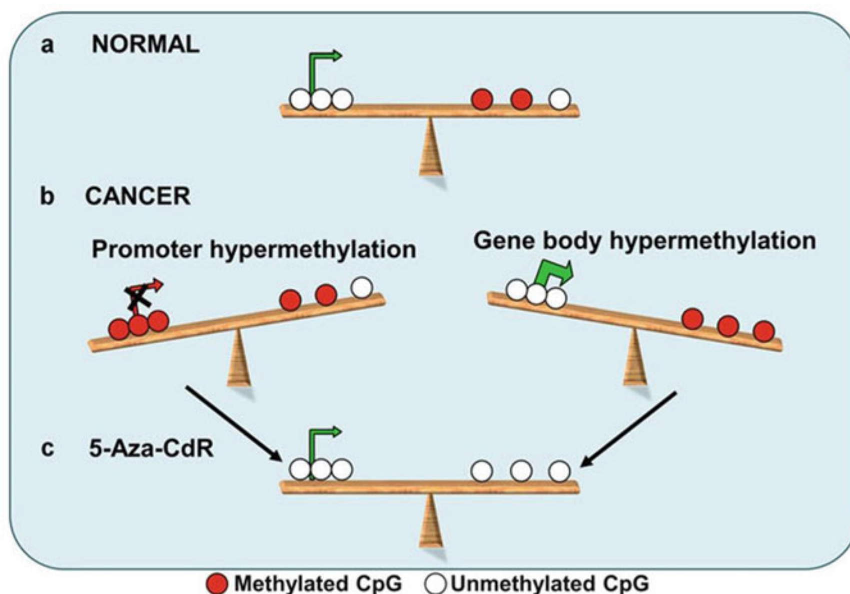


Fig. 13.1 DNA methylation equilibrium between the promoter and gene body modulates gene expression. In this diagram, methylated CpG sites are represented by red circles, unmethylated CpG sites are represented by white circles, and green arrows are indicative of active expression, while red arrow marks the absence of expression. (**a**) In normal mammalian tissues, genes that are actively transcribed have unmethylated promoters and some methylation in the gene body. (**b**) With the onset of cancer, however, promoter DNA hypermethylation can turn off the expression of genes, and gene body DNA hypermethylation can permit a more robust expression of some genes. (**c**) Treatment with DNA methyltransferase inhibitors such as 5-Aza-CdR can restore gene expression by removing aberrant methylation

Yang et al. 2014). Genetic and epigenetic alterations of DNA methylation, somatic mutations and insertions/deletions are separate mechanisms that can independently silence or activate one or both gene alleles, thereby coordinately impacting tumor suppressor genes and oncogenes across the human genome (Jones and Laird 1999). Interestingly, large-scale cancer genome DNA sequencing analyses have revealed somatic mutations in epigenetic modifier genes, indicating that genetic and epigenetic systems can also reinforce each other in a mutually dependent manner (You and Jones 2012).

13.2 DNA Methyltransferases

Mammalian DNA methylation primarily occurs as a covalent addition of a methyl group to the carbon-5 atom of cytosine (5mC) in a cytosine-guanine (CpG) dinucleotide sequence using *S*-adenosyl-L-methionine as a cofactor. This enzymatic reaction is catalyzed by the DNA methyltransferases (DNMTs) DNMT1, DNMT3A, and DNMT3B. It should be noted that DNMT2 has been identified as a tRNA methyltransferase, and *DNMT2* somatic mutations have been identified in several human cancer types (Elhardt et al. 2015; Goll et al. 2006). Finally, DNMT3L is a DNMT3-like protein that only shows expression in gametogenesis and does not contain a full catalytic domain that is required for enzymatic activity. DNMT3L deletion experiments showed that DNMT3L is important for maternal DNA methylation of imprinted regions (Bourc'his et al. 2001). DNMT3L has also been identified as an important interacting partner and stimulatory factor for DNMT3A and DNMT3B activity with preference for DNMT3A (Hata et al. 2002).

As the DNA replication fork performs leading and lagging strand synthesis of daughter strands, the newly synthesized strands are initially unmethylated. DNMT1 is positioned close to the DNA replication machinery and preferentially catalyzes the covalent addition of the methyl group onto the unmethylated strand of the newly synthesized hemi-methylated DNA molecule. While DNMT1 catalyzes most DNA methylation in a dividing cell, DNMT3A/3B strongly associate with nucleosomes to permit efficient propagation of DNA methylation by modification of the sites missed by DNMT1 (Liang et al. 2002; Okano et al. 1999; Rhee et al. 2000; Sharma et al. 2011). Rhee et al. (2002) subsequently showed that DNMT1 and DNMT3B cooperatively function to maintain DNA methylation profiles and retain cancer-specific gene silencing in colon cancer cells.

DNMT3A and DNMT3B show equal preference to hemi-methylated and unmethylated DNA molecules and are essential for the de novo creation of DNA methylation marks during embryonic development (Gujar et al. 2019; Okano et al. 1999). DNMT3A and DNMT3B are highly expressed in embryonic stem (ES) cells and, though downregulated, continue to be expressed in somatic cells (Sharma et al. 2011). Two DNMT3A isoforms, DNMT3A1 and DNMT3A2, have been identified (Chen et al. 2002). DNMT3A2 is expressed from an intronic promoter downstream of the *DNMT3A* promoter. It is the main form expressed in embryonic stem cells and is responsible for de novo DNA methylation activity (Chen et al. 2002). DNMT3B is expressed in more than 30 alternatively spliced variants (Duymich et al. 2016; Ostler et al. 2007; Wang et al. 2006), with DNMT3B1-3B8 and Δ3B1-Δ3B4 as the most characterized isoforms. These variants differ in which portions of the catalytic, N-terminal, and PWWP domains are removed due to alternative splicing. Although DNMT3B3 and DNMT3B4 are catalytically inactive, they can stimulate gene body DNA methylation, supporting their roles as potential scaffold proteins to signal DNA methylation by other catalytically active DNMTs, such as DNMT3A and DNMT3B (Duymich et al. 2016; Xu et al. 2020; Yang et al. 2014). In addition, the catalytically inactive murine Dnmt3b3 isoform also stimulates de novo DNA methylation by Dnmt3a2 and Dnmt3b2 (Zeng et al. 2020). Taken together, these studies demonstrated that DNMT3B3 without catalytic activity has a similar modulatory role as DNMT3L.

While DNMTs are responsible for catalyzing methyl transfer, the more recently identified ten-eleven translocation (TET) family of dioxygenases initiate DNA demethylation. These enzymes, through successive enzymatic reactions, oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) and 5-formylcytosine (5fC) to 5-carboxylcytosine (5caC) (Ko et al. 2010; Pastor et al. 2013; Pastor et al. 2011). The oxidization of 5mC contributes to the passive loss of DNA methylation over cell replication. In addition, the 5fC and 5caC oxidized intermediates can be restored to cytosine by base excision repair mediated by thymine DNA glycosylase (TDG) (Kohli and Zhang 2013). Together, DNMT and TET enzymes provide a toolkit for the dynamic regulation of DNA methylation and shaping the epigenomes of mammalian cells (Gujar et al. 2019).

13.3 Interplay Between DNA Methyltransferases and Histone Modifiers

The epigenome is shaped by not only DNA methylation, but also histone tail modifications, nucleosome occupancy and positioning, as well as the binding of transcriptional regulators. Histone modifiers catalyze the addition or removal of methyl, acetyl, phosphoryl, and other functional groups to specific amino acid residues of histone tail proteins. Subsequently, these modified histones work in concert with transcription factors and DNA methylation to promote or restrict gene expression. Unique histone modification profiles can be used to identify active or silenced promoters, enhancers, and areas of heterochromatin.

With these connections in mind, it is not surprising that DNA methylation and specific

histone modifications of the histone code are linked together to affect gene expression. For example, promoter DNA methylation can be associated with histone H3 lysine 27 trimethylation (H3K27me3) in gene silencing, while gene body DNA methylation is associated with H3K36me3 occupancy in facilitating gene expression. Interestingly, DNMT3B associates with H3K36me3 (Baubec et al. 2015), while DNMT3A associates with H3K36me2 marks and counteracts Polycomb-associated silencing (Shirane et al. 2020; Weinberg et al. 2019). H3K36me2 marks are placed by NSD1 and NSD2, while H3K36me3 marks are catalyzed by SETD2. H3K36me3 marks are focused on gene bodies, but H3K36me2 marks are diffusely organized in gene bodies and intergenic regions. In transcribed genes, H3K36me2 marks are mainly located through exon 1 downstream of the transcription start site, after which H3K36me3 marks are enriched. The associations of H3K36me2 and H3K36me3 marks with DNMT3A and DNMT3B, respectively, point to the high levels of organization and compartmentalization of DNA methylation profiles within transcribed regions in mammalian cells (Weinberg et al. 2019; Yang et al. 2014).

13.4 CpG Islands

Methylated cytosine residues are susceptible to spontaneous deamination resulting in the poorly repaired cytosine to thymine transitions. As a result, nearly a third of all disease-causing familial mutations and single-nucleotide polymorphisms are found in methylated CpG sites. Similarly, in somatic cells, CpG residues in gene body or coding regions habitually contribute to mutational hot spots, such as in the case of inactivating C to T transitions in the *TP53* tumor suppressor gene (Jones and Baylin 2002; Pfeifer 2000).

Another consequence of this phenomenon is that due to the mutational pressure there is a reduced representation of the CpG palindrome globally in the human genome, except in genomic regions designated as CpG islands (CGIs). CGIs were first defined by Gardiner-Garden and Frommer as ≥200-bp DNA regions with a C + G content of >50% and an observed/expected CpG content (observed/expected) >0.6 (Gardiner-Garden and Frommer 1987). However, a large proportion of *ALU* repetitive elements are included as CpG islands using these qualifiers, since these repeats are approximately 300 bp in length. To address this, Takai and Jones revised these criteria to ≥500 bp sequence length, ≥55% C + G content, and an observed/expected ≥0.65 (Takai and Jones 2002). These criteria exclude most ALU repetitive elements and capture CpG islands in promoter/5′ regions of genes.

While the majority of CpGs are methylated, CpGs located in CGIs remain overwhelmingly unmethylated (Meissner et al. 2008) and explain why they were refractory to evolutionary depletion. These islands are often, but not exclusively, found in the nearly half of all gene promoters (Meissner et al. 2008; Mikkelsen et al. 2007). Non-CGI promoters, on the other hand, are predominantly methylated and silent. The genes carrying non-CGI promoters are more likely to be tissue-specifically expressed; therefore, only a small subset of non-CGI promoters remains unmethylated and accessible for transcription factors in individual tissue types (Eckhardt et al. 2006).

13.5 DNA Methylation

13.5.1 Tissue-Specific DNA Methylation

Under normal physiological conditions, DNA methylation is vital to regulating gene expression in a tissue-specific context. Tissue-specific DNA methylation profiles can be used to identify cell of origin, cellular lineage, and developmental status. DNA methylation is also critical for facilitating selective allelic expression through parental genomic imprinting as well as the inactivation of one allele of the X-chromosome in females (Smith and Meissner 2013). X-chromosome inactivation is a developmentally necessary process by which the dosage of X-linked genes in females

is equalized to the dosage of X-linked genes in males (Pessia et al. 2012). In mammalian somatic cells, the choice of the X-chromosome to be inactivated is random. During development, however, the paternal X-chromosome is imprinted in mouse pre-implantation embryos. As the embryo continues to develop, the genes on the imprinted parental X-chromosome begin to reactivate at differing rates, and the parental X-chromosome is fully activated at the epiblast stage, after which time random X-chromosome inactivation occurs (Chen and Zhang 2020). Both random and imprinted X-chromosome inactivation are initiated and propagated by the increased expression of the noncoding RNA *XIST* on the X-chromosome that will be inactivated. This then triggers a cascade of events that ultimately result in RNA polymerase exclusion and the recruitment of repressive histone marks to the inactivating X-chromosome (Pontier and Gribnau 2011). Once this X-chromosome is silenced, DNA methylation across this allele is necessary to maintain the silenced state (Bestor et al. 2015).

The epigenetic phenomenon of genomic imprinting results in the unequal contribution of the chromosomes inherited from each parent to embryonic development. Imprinted genes are expressed in a parental-origin-specific manner rather than from both chromosomes. DNA methylation is a key mechanism by which allele-specific expression is established and maintained. For example, if the maternal allele is imprinted by DNA methylation, then it becomes silenced, and only the paternal allele is expressed (Ferguson-Smith 2011; Li et al. 1993). Allele-specific DNA methylation of the locus containing the *IGF2* and *H19* genes is a well-studied example of genomic imprinting impacting human development. IGF2 (insulin-like growth factor 2) is expressed from the paternal allele, while the *H19* noncoding RNA is transcribed from the maternal allele. *H19* expression results from allele-specific DNA methylation of an imprinting control region (ICR) located upstream of the *H19* transcription start site (Nordin et al. 2014). In turn, *IGF2* silencing of the maternal allele is mediated by CTCF binding to target sequences in the ICR (Takai et al. 2001). The binding of CTCF to unmethylated recognition sites serves as an insulating boundary to prevent IGF2 activation by distant enhancers (Kurukuti et al. 2006). CTCF also protects the maternal ICR allele from de novo DNA methylation (reviewed in Singh et al. 2012). Interestingly, aberrant DNA methylation of the *IGF2/H19* locus correlates with increased IGF2 expression (Ravenel et al. 2001; Kaneda and Feinberg 2005). Sustained IGF2 overexpression contributes to the development and progression of cancers such as colorectal and gastric, and the loss of imprinting at this locus is the most common alteration in Wilms' tumors (Bjornsson et al. 2007; Cui 2007; Li et al. 1993; Taniguchi et al. 1995; Wu et al. 1997).

13.5.2 DNA Methylation as a Function of Aging

In addition to tissue-specific DNA methylation profiles, DNA methylation is also dynamic with respect to aging (reviewed in Issa 1999; Jung and Pfeifer 2015; Unnikrishnan et al. 2019). Less pervasive than cancer-specific DNA methylation, age-related DNA methylation alterations occur at approximately 2% of all CpG sites (Unnikrishnan et al. 2019), but have similar consequences as DNA hypomethylation changes identified in human cancers, as they may provide impetus for gene expression alterations. Human cancers are generally thought as diseases related to human aging, because increased age is correlated with cancer risk (Issa 1999). Global DNA hypomethylation, as well as locus-specific DNA hypomethylation and DNA hypermethylation as a function of cellular aging has been identified. Indeed, Issa and colleagues first identified age-related DNA hypermethylation in the human *ESR1* CpG island (Issa et al. 1994). Following this discovery, additional age-related DNA hypermethylation changes were identified in human tissues. As technologies have advanced, it is now possible to identify age-associated DNA methylation profiles in mammalian cells and tissues using high-density, genome-scale DNA methylation microarrays. Based on the Illumina Infinium HumanMethylation27 (HM27), HM450

and EPIC BeadArrays, several panels of probes, termed Epigenetic Clocks, have been developed that accurately predict organismic age based on the DNA methylation profiles using three to 513 CpGs in the human genome (reviewed in Unnikrishnan et al. 2019). Epigenetic clocks are important for understanding how factors such as environmental exposures, chemical toxicity, emotional stress, and other stimuli impact cellular age and the risk of age-related diseases such as cancer.

13.6 Mutations of Epigenetic Modifier Genes in Human Cancers

Recent next-generation sequencing studies of cancer genomes have revealed frequent and recurrent mutations in a wide variety of epigenetic modulators, including mediators of DNA methylation, covalent histone modifiers, and genes encoding subunits of chromatin remodelers (Baylin and Jones 2011; Sandoval and Esteller 2012). Aberrant activities of these key epigenetic players result in deregulated gene expression and have been implicated in many malignancies, including numerous cancers (Hanahan and Weinberg 2011; Sharma et al. 2010).

Hematologic malignancies, most notably acute myeloid leukemia (AML), harbor high frequencies of *DNMT3A* somatic mutations and are associated with poor patient prognosis (Ley et al. 2010; Shah and Licht 2011; Stegelmann et al. 2011; Venugopal et al. 2021; Walter et al. 2011). Disease-associated DNMT3A mutations are thought to result in altered enzyme function. Specifically, the frequent R882 mutation results in poor DNA binding and CpG recognition that coincides with DNA hypomethylation at genes that contribute to tumorigenesis (reviewed in Venugopal et al. 2021).

Somatic mutations of the isocitrate dehydrogenase genes *IDH1* and *IDH2* are present in nearly 20% of AML patients. Heterozygous *IDH1* and *IDH2* mutations were first described in glioblastoma multiforme (GBM) (Parsons et al. 2008; Yan et al. 2009). *IDH1* heterozygous somatic mutations occur in nearly 15% of GBMs and *IDH1* mutation frequencies increase with lower glioma stage (Noushmehr et al. 2010). Wild-type IDH1 ($IDH1^{WT}$) functions in the citric acid cycle by converting isocitrate to alpha-ketoglutarate (α-KG), however, the mutant isoform ($IDH1^{mut}$) converts α-KG to 2-hydroxyglutarate (2-HG). Interestingly, 2-HG functions as an oncometabolite by inhibiting histone demethylases and the TET family of enzymes that catalyze hydroxyl-based DNA demethylation (reviewed in Weisenberger 2014).

Approximately 70% of pediatric high-grade gliomas (pHGGs) harbor histone H3 (*H3F3A*) mutations that result in the conversion of histone H3 lysine 27 to methionine (H3K27M) (Korshunov et al. 2015; Schwartzentruber et al. 2012). The K27M mutation inhibits the histone H3K27 trimethylase EZH2 and leads to reduced global levels of H3K27 trimethylation (H3K27me3), as well as extensive DNA hypomethylation and gene expression alterations (Bender et al. 2013; Chan et al. 2013; Cooper et al. 2018; Lewis et al. 2013). These findings highlight the interaction and coordination of histone modifications and DNA methylation in gene regulation, as well as the potential efficacy in targeting these aberrancies using novel cancer therapeutics.

Somatic mutations or downregulated expression of *SETD2* (the only known H3K36 trimethylase) are notable in clear cell renal cell carcinoma (ccRCC) (Pfister et al. 2015; Shoaib and Sorensen 2015; Turajlic et al. 2018a, b). *SETD2* mutations or downregulated *SETD2* expression occurs in 23% of ccRCCs and in 20% of papillary RCCs (Hsieh et al. 2017; Ricketts et al. 2018). *SETD2* mutations lead to decreased H3K36me3 occupancy and are highly associated with both cancer aggressiveness and dramatically reduced patient survival (Hakimi et al. 2013; Ho et al. 2016a, b). SETD2 has roles in DNA double-strand break (DSB) repair (Aymard et al. 2014; Carvalho et al. 2014; Li et al. 2013; Pai et al. 2014) and recruitment of splicing machinery (Bhattacharya et al. 2021; Bueno et al. 2016; de Almeida et al. 2011; Kolasinska-Zwierz et al. 2009; Luco et al. 2010;

Wilhelm et al. 2011; Yuan et al. 2017). *SETD2* mutations are especially relevant to DNA methylation as SETD2 and H3K36me3 marks recruit DNMT3B to catalyze gene body DNA methylation and prevent spurious intragenic transcriptional initiation of transposable elements (Baubec et al. 2015; Neri et al. 2017; Yang et al. 2014).

13.7 DNA Hypermethylation in Human Cancers

13.7.1 Tumor Stratification and DNA Methylation Marker Discovery Accelerated by International Consortia

Recent advancements in DNA sequencing and high-density microarray technologies have made it feasible to generate genome-wide genetic and epigenetic profiles for thousands of primary tumor and normal samples. Integrating molecular datasets allows for the construction of a more complete picture as to how tumor machinery constituents contribute to the initiation and progression of human cancers. Moreover, large sample collections and available patient information make it feasible to stratify tumors into subgroups that can be analyzed for the discovery of more personalized and efficacious treatment options. However, numerous bioinformatics and logistic challenges arise with such large datasets. To address these challenges, many research groups have assembled in multinational consortia, such as Encyclopedia of DNA Elements (ENCODE), the NIH Roadmap Epigenomics Mapping Consortium, and The Cancer Genome Atlas (TCGA). The ENCODE project has surveyed a large number of cell lines to extrapolate functional and regulatory elements of the genome, the NIH Roadmap Initiative has focused its resources on interrogating various tissue types to identify tissue-specific regulation of the epigenome, while TCGA has comprehensively generated and integrated molecular data from 11,000 tumor samples across 30 cancer types (Chadwick 2012; ENCODE Project Consortium 2004, 2012; ENCODE Project Consortium et al. 2007; The Cancer Genome Atlas Research Network 2013; Tomczak et al. 2015). Each TCGA specimen used aliquots of nucleic acid analytes from the same tissue specimen to perform whole exome sequencing for mutation detection, RNA sequencing for transcriptome characterization, single-nucleotide polymorphism (SNP) arrays to determine somatic copy number variations, and Illumina Infinium BeadArray analysis of DNA methylation (Tomczak et al. 2015) in order to integrate the molecular data profiles for each tumor tissue (Cooper et al. 2018). Along with the molecular information, the TCGA database contains clinical features including tumor grade, stage, prognosis, as well as gender, age, and race/ethnicity. A large number of integrated molecular analyses have been performed on tumors within and between tumor types to identify tumors with somatic mutations and copy number aberrancies, effects of DNA methylation on gene regulation, pathway analyses based on gene expression and integrated tumor subgroups in order to gain insights into the molecular mechanisms specific to these subgroups and identify subgroup-specific therapeutic targets (Weisenberger 2014).

13.7.2 Promoter DNA Hypermethylation

Broad epigenomic changes accompany cancer initiation and progression. It has been known for decades that cancer cells display a global loss of CpG methylation, including regions with low density of CpG sites, repeat elements, retrotransposons, and nuclear lamina-associated domains (LADs). This phenomenon occurs juxtaposed with concomitant locus-specific DNA hypermethylation at CpG islands and CpG island shores (reviewed in Weisenberger and Liang 2015). Gene silencing via promoter DNA hypermethylation serves as a secondary mechanism for the inactivation of tumor suppressor genes (TSGs) in addition to genetic changes such as somatic mutations and deletions (Ehrlich and Lacey 2013; Irizarry et al. 2009; Jones and Baylin 2007; Jones and Laird 1999; Shen and

Laird 2013; You and Jones 2012) (Fig. 13.1b). Silencing of cell cycle regulators and DNA repair genes through DNA hypermethylation has been reported in multiple cancer types and is often mutually exclusive with the genetic inactivation of the gene consistent with Knudsen's two-hit hypothesis (Alvarez-Nunez et al. 2006; Chiang et al. 2006; Jones and Laird 1999; Sakai et al. 1991). For example, sporadic breast and ovarian cancers harbor *BRCA1* loss due to inactivating somatic mutations as well as downregulated expression of the WT allele by promoter DNA hypermethylation. Similarly, the *VHL* tumor suppressor gene is inactivated by promoter DNA hypermethylation and somatic mutations in human ccRCCs (Chiang et al. 2006; Esteller et al. 2000; Herman et al. 1994; The Cancer Genome Atlas Research Network 2013).

O-6-Methylguanine-DNA methyltransferase (MGMT), a DNA repair enzyme responsible for repairing O6-Methylguanine-DNA adducts, is frequently inactivated by promoter DNA hypermethylation in multiple cancer types, most notably those derived from brain (gliomas) and colorectal tissues. Promoter DNA hypermethylation-based *MGMT* silencing results in increased susceptibility to genetic mutations in essential genes such as *TP53* and *KRAS*. Interestingly, *MGMT* silencing sensitizes the cancer cell to the chemotherapeutic agent temozolomide (TMZ). Clinical studies in glioblastoma multiforme (GBM) suggest that TMZ treatment is most beneficial for patients whose tumors display *MGMT* promoter DNA hypermethylation (Donson et al. 2007; Silber et al. 2012; Zarnett et al. 2015) and is ineffective in tumors in which MGMT is expressed, since the MGMT enzyme repairs the DNA damage incurred by TMZ administration.

TCGA DNA methylation data have been particularly informative with respect to identifying promoter DNA hypermethylation mediated gene silencing and associations with gene mutations. Analyses of DNA methylation data from 489 TCGA high-grade serous ovarian tumors have unveiled promoter DNA methylation events in 168 genes, including *BRCA1*. *BRCA1* inactivation due to promoter DNA hypermethylation and somatic mutations are mutually exclusive. While high-grade serous ovarian patients carrying somatic *BRCA1* mutations show improved overall survival than patients with the WT gene, patients with *BRCA1* epigenetic silencing do not carry this survival advantage, suggesting that patient outcome is dependent on the mechanism of *BRCA1* inactivation (The Cancer Genome Atlas Research Network 2011). Similarly in lung squamous cell carcinomas (LUSC), the *CDKN2A*$^{(INK4A/p16)}$ tumor suppressor is inactivated by exon 1a promoter DNA hypermethylation in 21% of cases, by mutations in 18% of cases and by deletion in 29% of cases (The Cancer Genome Atlas Research Network 2012b), indicating that this tumor suppressor gene is altered by several unique mechanisms in lung squamous tumors. Lung adenocarcinomas (LUAD) harboring *CDKN2A*$^{(INK4A/p16)}$ DNA hypermethylation also display overall low copy number alterations, and *CDKN2A*$^{(INK4A/p16)}$ somatic mutations are enriched in a subgroup of LUAD tumors with extensive DNA hypermethylation and *SETD2* mutations (The Cancer Genome Atlas Research Network 2014b). Since SETD2 catalyzes the placement of H3K36me3 marks that are intrinsically intertwined with gene body DNA methylation, this mutation suggests that a subset of LUAD tumors have multiple levels of epigenetic dysregulation.

High *SETD2* mutation frequencies were also identified in TCGA clear cell renal cell carcinomas (ccRCCs) (The Cancer Genome Atlas Research Network 2013) and were associated with non-promoter DNA hypomethylation. Interestingly, a comprehensive examination of ccRCC DNA methylation profiles showed that overall promoter DNA hypermethylation frequency positively correlated with tumor stage and grade. TCGA also identified *VHL* somatic mutations in approximately 50% of ccRCCs and *VHL* epigenetic silencing in 7% of tumors in a mutually exclusive fashion. *UQCRH* silencing due to promoter DNA hypermethylation was observed in 36% of TCGA ccRCC tumors.

MIR21 overexpression due to promoter DNA hypomethylation correlates with advanced tumor

stage and poor ccRCC patient outcome (The Cancer Genome Atlas Research Network 2013). Importantly, MIR21 helps drive a glycolytic shift similar to the Warburg Effect. The Warburg Effect is a phenomenon in which cancer cells generate adenosine 5′-triphosphate (ATP) energy equivalents by aerobic glycolysis rather than through mitochondrial oxidative phosphorylation that is common in normal somatic cells (Vander Heiden et al. 2009). Although aerobic glycolysis is a less efficient mechanism of ATP production as compared to oxidative phosphorylation, cancer cells show large increases in glycolysis and lactate under normoxic conditions (reviewed in Gaal 2021). MIR21 is induced by high glucose levels and inhibits the PTEN tumor suppressor that ultimately results in PI(3)K pathway enhancement (The Cancer Genome Atlas Research Network 2013).

13.7.3 CpG Island Methylator Phenotypes (CIMPs) Stratify Tumor Subclasses

In 1999, Toyota et al. identified a subset of colorectal cancers (CRCs) that showed cancer-specific DNA hypermethylation of a unique subset of gene regions, most notably the concordant promoter DNA hypermethylation of *CDKN2A*$^{(INK4A/p16)}$, *THBS1,* and *MLH1*. This phenomenon, referred to as CpG island methylator phenotype (CIMP), represents approximately 15% of colon cancers. Using a genome-wide approach, Weisenberger and colleagues characterized CIMP in 295 primary CRC samples, correlated CIMP to clinical features and developed a five-gene CIMP-specific DNA methylation marker panel (Weisenberger et al. 2006). Interestingly, CIMP is strongly associated with the *BRAF(V600E)* somatic mutation, *MLH1* silencing via promoter DNA hypermethylation, high microsatellite instability (MSI-H), high global somatic mutation rates (hypermutation), diploid genome copy number, *TP53* somatic mutations, tumor location in the proximal region of the colon, female gender, advanced age, and family history of disease. Alternatively, non-CIMP CRCs are highlighted by *KRAS* somatic mutations, non-hypermutated genotype and extensive copy number alterations (The Cancer Genome Atlas Research Network 2012a; Hinoue et al. 2012; Weisenberger 2014; Weisenberger et al. 2006, 2015). Ogino and colleagues first described a CIMP-low (CIMP-L) subgroup as having an attenuated CIMP phenotype and association with *KRAS* mutations (Ogino et al. 2006) and Shen et al. identified the CIMP2 subgroup as displaying CIMP-associated DNA methylation with *KRAS* mutation enrichment (Shen et al. 2007). The classical CIMP subgroup was subsequently termed as CIMP-high (CIMP-H) to reflect the extensive DNA methylation in these tumors compared to CIMP-L tumors. Two-dimensional unsupervised clustering of CRC DNA methylation data from Hinoue et al. and TCGA identified four CRC tumor subgroups: CIMP-H, CIMP-L, and two non-CIMP subgroups (The Cancer Genome Atlas Research Network 2012a; Hinoue et al. 2012). Currently, CRCs are classified by four consensus molecular subgroups (CMS), termed CMS1-4 (Guinney et al. 2015). CMS1 tumors display CIMP-H features with MSI-H, *BRAF(V600E)*, DNA hypermutation, and poor patient outcome; CMS2 tumors are non-CIMP; CMS3 tumors are CIMP-L, *KRAS*-mutated, microsatellite stable (MSS) and show low copy number variation; CMS4 tumors are non-CIMP and are associated with poor patient survival (reviewed in Weisenberger et al. 2018).

Early mechanistic insights into CIMP generation in cancer came from investigating promoter-associated DNA hypermethylation in gliomas. Using TCGA data, Noushmehr et al. comprehensively characterized DNA methylation of GBM tumors and identified a glioma CIMP subgroup (G-CIMP) that defines a subset of gliomas (Noushmehr et al. 2010). With respect to clinical features, G-CIMP patients are younger in age and have significantly increased survival compared to non-G-CIMP patients. Interestingly, G-CIMP tumors are tightly associated with a high frequency of *IDH1* somatic mutations (Brennan et al. 2013; Noushmehr et al. 2010). IDH1^{WT} converts isocitrate to alpha-ketoglutarate (α-KG), however, IDH1^{mut} further converts

α-KG to 2-hydroxyglutarate (2-HG), which subsequently inhibits the TET family of enzymes that catalyze hydroxyl-based DNA demethylation (reviewed in Weisenberger 2014). Thus, production of 2-HG results in the accumulation of DNA methylation along with aberrant histone methylation (Dang et al. 2009). Furthermore, the *IDH1* mutation alone is sufficient to establish G-CIMP and constitutes an early event that is likely driving the tumorigenesis (Hill et al. 2014; Turcan et al. 2012). It should also be noted that AML tumors also harbor *IDH1*, *IDH2,* and *TET* somatic mutations. *IDH1* and *IDH2* mutations are mutually exclusive, while *TET2* mutations are mutually exclusive with all *IDH* mutations, suggesting redundant activity of their gene products. TCGA and others have shown that AML tumors with *IDH* or *TET* mutations show substantial DNA hypermethylation (The Cancer Genome Atlas Research Network et al. 2013b; Figueroa et al. 2010; Shih et al. 2012).

To date, several reports have described CIMPs in many additional cancers including gastric, breast, bladder, melanoma, prostate, hepatocellular, and endometrial cancers. Interestingly, two CIMP subgroups are present in gastric cancer: one CIMP group is akin to colorectal CIMP-H with MSI-H, DNA hypermutation and *MLH1* epigenetic silencing, while the EBV-CIMP group shows more extensive DNA hypermethylation than seen in other CIMP subgroups of human cancers and is highly correlated with Epstein-Barr virus (EBV) infection (EBV-CIMP) (The Cancer Genome Atlas Research Network 2014a). These findings suggest that EBV-CIMP tumors may have unique molecular mechanisms of tumorigenesis as compared to other CIMP tumor types. Finally, Fang et al. (2011) found that breast CIMP tumors (B-CIMP) were associated with positive estrogen receptor (*ESR1*) and progesterone receptor (*PGR*) status. Moreover, B-CIMP patients had a lower risk of metastasis and improved clinical outcome than non-CIMP patients. Recognizing and understanding how CIMPs are associated with patient clinical co-variates and their potential mechanism for their genesis and propagation may have important consequences for cancer patient monitoring and treatment.

13.7.4 DNA Hypermethylation of Noncoding RNAs

Aside from the functions of noncoding RNAs as epigenetic regulators (Matzke and Mosher 2014; Wei et al. 2017), DNA methylation also plays an important role in the regulation of noncoding RNAs (ncRNAs) including microRNA (miRNA), small nucleolar RNA (snoRNA), vault RNA (vtRNA), and long noncoding RNA (lncRNA). These elements are critical regulators of cellular processes including proliferation, differentiation, and development (Esteller 2011). DNA hypermethylation of miR-127 results in its downregulated expression and subsequent upregulation of the *BCL6* proto-oncogene to drive tumorigenesis (Ehrlich 2009; Kulis et al. 2012; Saito et al. 2006). Similarly, miR-101 is also downregulated in bladder cancer via DNA hypermethylation. miR-101 functions in repressing EZH2, the catalytic component of PRC2 responsible for introducing H3K27me3, highlighting its role in global epigenomic regulation (Friedman et al. 2009). Moreover, microRNA miR-124a silencing due to DNA hypermethylation in acute lymphoid leukemia (ALL) activates the CDK6-RB1 oncogene pathway, contributing to poor patient survival (Agirre et al. 2009).

It has also been observed that the CpG islands upstream of the snoRNAs *SNORD123*, *U70C*, and *ACA59B* endure cancer-specific DNA hypermethylation resulting in their transcriptional silencing in ALL. Gastric cancer and acute myeloid leukemia (AML) patients with DNA hypermethylation of n*c866*, also known as *vtRNA2-1*, show poor survival (Lee et al. 2014; Treppendahl et al. 2012). In vitro knockdown of *nc866* in human gastric cell lines induces known oncogenes, and overexpression of the ncRNA reduces cellular proliferation (Lee et al. 2014). In myelodysplastic syndrome (MDS), both *vtRNA1-2* and *vtRNA1-3* are silenced by DNA hypermethylation and are associated with

decreased survival in lower-risk MDS patients (Helbo et al. 2015). Finally, epigenetic silencing of the *MORT* lncRNA via DNA hypermethylation is significant for immortalizing human mammary epithelial cells. Deficient *MORT* expression is common in most cancers and can be reactivated by treatment of the DNA methylation inhibitor 5-aza-2′-deoxycytidine (5-Aza-CdR), implicating its role in immortalization during oncogenesis (Vrba et al. 2015). These findings and several others make it clear that aberrant DNA hypermethylation of ncRNAs is associated with tumor suppression, a fundamental feature of cancer, and has a vital role in the disease initiation and progression.

13.7.5 DNA Hypomethylation

13.7.5.1 Repetitive Element DNA Hypomethylation

Although DNA hypomethylation was the first DNA methylation change discovered in cancer, the implication of this dysregulation in tumorigenesis has often been overlooked. Feinberg and Vogelstein, as well as Gama-Sosa et al., identified a global decrease in 5mC content across numerous cancer types in the 1980s (Feinberg and Vogelstein 1983; Gama-Sosa et al. 1983). DNA hypomethylation is an early event in tumorigenesis, occurs frequently in benign hyperplasia, becomes more prominent with tumor progression, and metastatic lesions possess greater DNA hypomethylation than primary tumors (Li et al. 2014b). Finally, non-CGI promoter DNA hypomethylation is much less frequent than promoter CGI DNA hypermethylation but may result in oncogene and proto-oncogene upregulation (Feinberg and Vogelstein 1983; Soes et al. 2014).

Most DNA hypomethylation occurs in intergenic and intragenic regions. These genomic areas are replete with repetitive and transposable elements. DNA methylation suppresses these elements and their DNA hypomethylation, most evident in human cancers, contributes to their activation and ectopic expression. Repetitive elements comprise nearly 50% of the human genome and consist of interspersed repeats derived from transposable elements (Deininger et al. 2003; Jordan et al. 2003; Prak and Kazazian 2000; Weiner 2002), including short interspersed nucleotide elements (SINEs), long interspersed nucleotide elements (LINEs), as well as simple and complex DNA sequence tandem repeats. *ALU* repetitive elements are short (~300 bp) DNA sequences and represent the most abundant SINE in the human genome, with one million copies per haploid genome (Weiner 2002). LINEs are more abundant with over 500,000 copies and 20% of the human genome. These elements have inserted themselves and transposed in eukaryotic germlines in waves during evolution. They have the potential to modify gene control in the host organism (Zhou et al. 2020), however, only a small percentage are active in normal somatic cells (reviewed in Deininger et al. 2003; Ehrlich 2002).

LINE-1 and *ALU* DNA hypomethylation is recognized as indicators of tumor progression and prognosis in several cancer types including prostate, melanoma, bladder, renal, and ovarian cancers (Andreotti et al. 2014; Ecsedi et al. 2013; Karami et al. 2015; Saied et al. 2012; Su et al. 2014; Yegnasubramanian et al. 2008). Indeed, Li and colleagues (Li et al. 2014b) showed in a meta-analysis of repetitive element DNA methylation data from 146 studies that global DNA methylation of *ALU*, *LINE-1,* and satellite repeats is strongly associated with poor cancer patient outcome. In addition, *LINE-1* DNA hypomethylation can greatly affect gene expression, not only of the *LINE-1* repetitive element, but also the gene region in which the *LINE-1* repeat resides. In this regard, *LINE-1* DNA hypomethylation can unveil alternative promoters to activate oncogenes (Jang et al. 2019; Wolff et al. 2010). For example, *LINE-1* DNA hypomethylation stimulates the adoption of a permissive chromatin architecture of an alternative gene promoter of the *MET* oncogene, thereby activating its expression (Wolff et al. 2010).

The *ALU* and *LINE-1* classes of repetitive elements can cause genomic instability if they copy and insert themselves into genic regions of the genome (Konkel and Batzer 2010). They also

have been shown to confuse the homologous recombination repair machinery into non-allelic recombination and cause replication fork stalling during DNA replication, thus leading to insertions, deletions, and other rearrangements (reviewed in Konkel and Batzer 2010). Repetitive element DNA hypomethylation can potentially further exacerbate this phenomenon by not only creating an abundance of genetic aberrancies in cancer cells, but also promoting the accrual of double-stranded DNAs (dsDNAs) and double-stranded RNAs (dsRNAs) that stimulate a host defense response that ultimately degrades repetitive element transcripts (reviewed in Ishak et al. 2018). This is the basis for the therapeutic strategy to promote viral mimicry, in which activation of endogenous retroviruses (ERVs) via drug-induced DNA demethylation can be exploited for cancer therapies (Jones et al. 2019). However, repetitive element induction in cancer cells may also result in reduced therapeutic efficacy after treatment with DNA methylation inhibitors (Ishak et al. 2018; Shen et al. 2021), thus creating a conundrum of immune evasion.

The mechanisms of repetitive element silencing in cancer cells have proven to be evasive. Since repetitive element silencing is generally associated with heterochromatin and the H3K9me3 repressive mark, Shen et al. (2021) recently screened for H3K9me3-associated mediators of repetitive element silencing in human cancer cells. Interestingly, they identified FBXO44 as an essential mediator of H3K9me3-based repetitive element silencing in which FBXO44 recruits a multitude of chromatin remodelers and modifiers known to be associated with heterochromatin. In addition, FBXO44 also interacts with the DNA replication machinery to silence repetitive elements to reduce genome instability immediately after DNA replication. Hence, FBXO44 inhibition results in immune response activation and genomic instability in the form of DNA strand breaks. Moreover, FBXO44 inhibition results in reduced tumor growth, cellular proliferation, and cell survival, as well as a heightened immune response and increased sensitivity to PD-L1 targeted immune therapy. These results highlight not only the critical role of repetitive element silencing, but also the approach for characterizing mediators of repetitive element silencing that can be therapeutically targeted for cancer treatments.

13.7.5.2 Partially Methylated Domains (PMDs)

The development of whole genome bisulfite sequencing (WBGS) technologies to interrogate DNA methylation profiles at single-base resolution has unveiled complexities of the methylome and associations with other epigenetic modifications and gene expression. Lister and colleagues first presented WGBS-based analyses of the human H1 embryonic stem (ES) cells and IMR90 fetal lung fibroblast methylomes (Lister et al. 2009). While IMR90 cells showed that DNA methylation was almost exclusively located in CpG dinucleotides, the H1 ES cells showed substantial levels of non-CpG methylation, namely at CHG or CHH contexts in which H nucleotides are A, C, or T. While non-CpG DNA methylation was previously shown in ES cells (Ramsahoye et al. 2000), this WGBS analysis showed it is more pervasive and widespread than previously determined.

Pairwise comparisons between H1 and IMR90 DNA methylomes also revealed substantial DNA hypomethylation in IMR90 cells. Indeed, large, continuous regions of >150 kb displayed mean DNA methylation levels below 70% and were termed Partially Methylated Domains (PMDs). PMDs were identified on every autosomal chromosome and also correlated with gene body DNA hypomethylation on the inactive X-chromosome in female IMR90 cells (Lister et al. 2009).

WGBS analyses of a primary human colon tumor and its paired adjacent normal tissue showed that PMDs >100 kb in length occupied a substantial portion of the colon cancer genome, and small regions of focal DNA hypermethylation were identified within these hypomethylated PMDs. PMDs are associated with heterochromatic, gene poor regions of the genome that normally are under-expressed (Hon et al. 2012; Hovestadt et al. 2014; Salhab et al. 2018). PMDs also strongly overlapped with nuclear lamina-associated domains (LADs) and

late-replicating areas of the human genome (Berman et al. 2011), as was also shown for multiple tumor types (Du et al. 2019; Hama et al. 2018; Hansen et al. 2011). PMDs have subsequently been identified across multiple cancer and tissue types and genes contained within large regions of DNA hypomethylation were also shown to be enriched for matrix remodeling (Hansen et al. 2011). A meta-analysis of over 170 WGBS datasets generated from a wide array of human tissue types revealed that PMDs comprise 50–75% of the human genome, but only 25% of the genome is shared by PMDs across all surveyed tissue types (Salhab et al. 2018).

Accumulation of PMD DNA hypomethylation correlates with increased enrichment of heterochromatin marks and cellular proliferation (Salhab et al. 2018). Short PMDs are linked to early-mid S phase replicating genes, as well as H3K9me3 and H3K27me3 marked chromatin, whereas intermediate length PMDs are associated with mid-late S phase replication and lower H3K27me3 occupancy than short PMDs. Finally, late S-G2 replicating genes have extensive PMD regions that are coupled with H3K9me3 marks bordered by H3K27me3 chromatin (Salhab et al. 2018)

WGBS data from 39 TCGA primary tumors and eight matched normal-adjacent tissues across multiple cancer/cell types was integrated with 343 human and 206 mouse WGBS datasets (Zhou et al. 2018) and revealed that PMD DNA hypomethylation can be predicted by DNA sequence context. Specifically, low CpG density, CpGs that are flanked by an A or T (W) on both sides (WCGW) and CpGs in the WCGW context without neighboring CpGs (solo-WCGW) were the most prone to DNA hypomethylation. Even though solo-WCGW sites are an extreme sequence context, these represent 13% of the human genome. Solo-WCGW containing PMDs range from 100 kb to 5 Mb in length, correlate with chronological age, and are mostly shared across cell type and developmental lineage.

PMD DNA hypomethylation is also associated with increased somatic mutation and copy number frequencies and the extent of tumor proliferation, suggesting a link between PMD DNA hypomethylation and cumulative mitotic cell division, specifically its role as a mitotic clock. Solo-WCGW DNA methylation was correlated with replication timing and to a lesser extent, H3K36me3 occupied chromatin. Indeed, less than 15% of all solo-WCGWs overlapping H3K36me3-enriched regions were methylated, while the remaining solo-WCGWs displayed lower overall DNA methylation with a strong link to replication timing (Zhou et al. 2018). However, Decato and colleagues (Decato et al. 2020) identified a panel of late-replicating genes that retain H3K36me3 occupancy but are not DNA hypomethylated, suggestive of a higher level of granularity between PMD DNA hypomethylation and replication timing.

13.7.6 Whole Genome Bisulfite Sequencing (WGBS) of Cancer Genomes

The discovery of PMDs in human cancers highlights the intricate complexities of cancer-specific DNA methylation alterations in not only gene expression but also replication timing and cellular proliferation. WGBS-based profiling of primary human cancers has been reported for multiple tumor types and integrating WGBS data with gene expression, copy number variation, somatic mutation, and chromatin modifications provide complete windows as to how DNA methylation regulates gene expression and is associated with DNA sequence alterations. WGBS and whole genome sequencing (WGS) data sets for esophageal squamous cell carcinomas and adenocarcinomas (Lin et al. 2018) show that these two cancer types display disparate driver genes, suggesting that these tumor types follow unique tumorigenic programs. The transcription factor *ZPF36L2* was also discovered as a novel significantly mutated gene in esophageal carcinomas and, interestingly, is associated with a super-enhancer. ZPF36L2 is silenced by DNA hypermethylation of its super-enhancer in esophageal carcinomas.

WGBS analysis of 100 primary metastatic castration resistant prostate tumors and 10 benign

prostate tissues showed that DNA methylation variation between tumors was primarily found in gene bodies, transcription factor-binding sites, enhancers, and heterochromatic regions (Zhao et al. 2020). In addition, integrating WGBS DNA methylation data with WGS data showed that differential DNA methylation as a function of disease progression occurs at regulatory regions and somatic mutation hotspots. 20% of tumors displayed a CIMP-like DNA methylation signature that is associated with *BRAF*, *DNMT3B*, *IDH1,* and *TET2* somatic mutations (Zhao et al. 2020). Incorporating WGBS DNA methylation data with RNA sequencing expression data highlighted associations of DNA methylation with expression of *AR*, *MYC,* and *ERG* driver genes.

A similar approach was performed for 33 primary liver tumors in which the integration of DNA methylation and gene expression data resulted in the identification of 611 gene regions that are differentially methylated and differentially expressed in liver tumors (Huang et al. 2021). These gene regions were enriched in activated DNA repair and cell cycle pathways and downregulated in metabolic pathways. The majority of these genes are also more effective prognostic biomarkers over those identified by DNA methylation and gene expression alone.

13.7.7 Gene Body DNA Methylation

For decades, much of the research efforts in cancer epigenetics had been concentrated on the regulation of DNA methylation at gene promoters. Due to advances in next-generation sequencing and high-density microarray technologies, DNA methylation profiling can now be performed in a genome-wide context. It is now increasingly evident that DNA methylation at non-promoter intragenic and intergenic regions is also dynamically regulated and contributes to physiological changes as well as to the development of disease states. Unlike promoters, where DNA methylation contributes to a "closed" chromatin architecture resulting in gene repression, DNA methylation in transcribed regions of genes (gene bodies) is often positively correlated with gene expression (Liu et al. 2018; Yang et al. 2014). Indeed, transcribed regions display H3K36me3 marks, and DNMT3B binds to H3K36me3 marks via its PWWP domain (Baubec et al. 2015) to facilitate gene body DNA methylation (Baubec et al. 2015; Duymich et al. 2016; Yang et al. 2014). Gene body regions are generally CpG poor, and therefore, most gene bodies are methylated in normal somatic cells and are enriched within repetitive elements (Brocks et al. 2017; Zhou et al. 2020).

Gene body DNA methylation can also add to transcription efficiency by regulating the usage of alternate transcription start sites. For example, DNA methylation profiling of primary glioblastoma tumors purports a role for gene body DNA hypomethylation in stimulating the transcription from alternate promoters, thereby resulting in an increased expression of alternative transcripts and expression of oncogenic protein isoforms (Nagarajan et al. 2014). In addition, gene body DNA hypomethylation can reveal distal regulatory elements (enhancers) that are muted in a tissue-specific manner (Yang et al. 2014). Moreover, DNA methylation profiles of normal B cell and chronic myeloid leukemia revealed widespread gene body DNA hypomethylation targeting enhancer sites (Kulis et al. 2012).

13.7.8 Enhancer DNA Methylation

Together with promoters and gene bodies, enhancers play a significant role in regulating the expression and activity of their target genes. Enhancers serve as platforms for transcription factors (TFs) that bind DNA through sequence recognition. Remarkably, one enhancer can regulate the activity of multiple promoters and vice versa, each promoter can be regulated by several enhances (Bulger and Groudine 2011). The presence of multiple TFs is usually necessary for enhancer activation. Additionally, functional enhancers are decorated with active histone marks including H3K4 monomethylation (H3K4me1) and H3K27 acetylation (H3K27ac). Through long-range interactions such as

chromatin "looping," these distal elements deliver the accessory proteins to promoters and stimulate robust transcription.

Although DNA methylation is inversely correlated with the presence of active histone marks, such as those that delineate active enhancers, expression-related DNA methylation of enhancers has also been observed. Enhancer DNA methylation inversely correlated with gene expression in a manner similar to promoters and is often a superior predictor of expression levels over promoter DNA methylation (Aran and Hellman 2013). Furthermore, enhancers can regulate gene expression in a cell-type-specific manner even when the gene promoter is continually unmethylated (Aran and Hellman 2013).

TF recognition sequences and other DNA-binding elements are mostly situated in unmethylated DNA regions. DNA methylation can thwart the association of TFs to DNA, as was shown for bHLH, bZIP, and ETS TF classes (Yin et al. 2017). Additionally, the presence of TFs can promote DNA hypomethylation by preventing DNMTs from accessing DNA (Calo and Wysocka 2013). However, some TFs also show preference in binding to methylated recognition sequences (Yin et al. 2017). TF involved in development, most notably homeodomain, NFAT and POU TFs and the OCT4 pluripotency factor, bind to methylated DNA. Thus, the subtle modulation of DNA methylation at enhancers can greatly affect gene expression of multiple target genes.

In cancer, DNA hypomethylation of intergenic and intragenic enhancers can reveal TF binding motifs and induce downstream expression changes (Aran and Hellman 2013; Kulis et al. 2012). Alternatively, enhancer DNA hypermethylation can result in enhancer decommission, resulting in losses of both active histone marks and TF binding. Consequently, these alterations can modulate gene transcription independent of promoter DNA methylation fluctuations (Kulis et al. 2012). DNA methylation also plays an unexpected dual role at enhancer regions by being both anti-correlated focally at transcription factor-binding sites and positively correlated globally with the active H3K27ac mark to ensure structural enhancer integrity (Charlet et al. 2016).

Xiong et al. (2019) used WGBS to identify differentially methylated enhancers in human hepatocellular carcinomas (HCCs) and unveiled global enhancer DNA hypomethylation. Interestingly, DNA hypomethylation of the CCAAT/enhancer-binding protein beta (C/EBPβ) enhancer correlates with C/EBPβ overexpression and poor patient outcome. C/EBPβ is a transcription factor that is also involved in cellular growth control and differentiation (Nerlov 2007). In addition, C/EBPβ is thought to function as a leukemic driver in AML as it interacts with other transcription factors to stimulate tumor proliferation (Abdel Ghani et al. 2022). In HCC, CEPBB overexpression due to enhancer DNA hypomethylation stimulates C/EPPB enhancer RNA expression by occupancy of the active H3K27ac mark and recruitment of RNA polymerase II and bromodomain proteins (Xiong et al. 2019).

Similarly, Heyn et al. (2016) identified cancer-associated DNA methylation alterations in super-enhancer regions of the genome. Originally described by Whyte et al. (2013), super-enhancers are large clusters of enhancers that are bound by master transcription factors and drive developmental programming and cell identity. Super-enhancers are marked by H3K27ac occupancy and are associated with oncogenic drivers in human cancers, most notably encompassing the *MYC* gene locus (Hnisz et al. 2013; Loven et al. 2013). Cancer-specific promoter DNA hypermethylation occurs extensively at polycomb repressive complex 2 (PRC2) targets (Widschwendter et al. 2007), suggestive of a molecular cross-talk between stem cell development and cancer genesis. In normal primary tissues, super-enhancers generally display more extensive DNA hypomethylation than of conventional enhancers, however, comparisons between super-enhancers show extensive differential DNA methylation (Heyn et al. 2016). Primary tumors show both super-enhancer DNA hypomethylation and DNA hypermethylation that correlated with target gene activation and inactivation, respectively.

13.8 Liquid Biopsy Measurements of Cancer-Specific DNA Methylation

Cancer-specific DNA methylation data have been abundantly described for virtually every tumor type. Thus, DNA methylation markers are convenient prognostic and diagnostic tools (Laird 2003; Levenson 2010). DNA methylation is a stable chemical modification, and the methylation status of loci of interest can be readily obtained from primary fresh/frozen tissues as well as highly processed tissues, including formalin-fixed, paraffin embedded (FFPE) tissue sections. In addition to primary tumor specimens, liquid biopsies of cell free DNA (cfDNA) derived from the tumor can be analyzed in blood plasma/serum, urine sediment, cerebral spinal fluid, and other biological fluids (Taback and Hoon 2004). Tumors are inherently unstable and shed tumor material into the blood stream due to tumor necrosis and apoptosis. The presence of tumor-derived cfDNA in blood from early and late-stage cancer patients has been well characterized (Bettegowda et al. 2014; Lo et al. 1999; Sun et al. 2015). Tumor-derived cfDNA levels in the blood stream increase with tumor stage. cfDNAs in blood or urine represent non-invasive means of sampling tumor material and can be used in place of primary tissues since these are derived from the primary tumor (reviewed in Angeles et al. 2021) and can also be exploited for determining tumor cell type (Barefoot et al. 2021). Coupled with technological advancements in analyzing ultra-low sample amounts using DNA methylation microarrays or whole genome bisulfite sequencing, epigenomic profiling of cfDNAs has promise in cancer detection, prognosis, and diagnostics (Angeles et al. 2021; van der Pol and Mouliere 2019).

Cancer-specific detection of *SEPT9* DNA methylation in blood plasma is an important example of a liquid biopsy-based epigenomic profile driving CRC patient diagnostics. *SEPT9* DNA methylation occurs in nearly all colorectal tumors and adenomas, but not in normal colonic mucosae. Accordingly, the initial cfDNA *SEPT9* DNA methylation assays showed high CRC detection sensitivity (72%) and specificity (86%) in plasma samples (deVos et al. 2009). Overall, *SEPT9* DNA methylation assays show pronounced improved detection sensitivity and specificity as compared to carcino-embryonic antigen (CEA) and fecal occult blood and tests (Weisenberger et al. 2018).

Next-generation sequencing and DNA methylation array-based data of cfDNA samples have been successfully implemented as a means of biomarker discovery. Cai et al. developed the ColonAiQ assay (Cai et al. 2021), a six-gene DNA methylation classifier that stratifies colorectal tumors and adenomas as an early detection system, after curating a large discovery dataset (Chen et al. 2020). ColonAiQ is effective in monitoring disease progression in colorectal cancer patients after surgical resection. Similarly, Pulverer et al. describe independent cfDNA methylation classifiers for the early detection of colorectal adenomas and to monitor response to neo-adjuvant therapy (Pulverer et al. 2021). Jin et al. developed a 10-gene classifier that showed improved ability over SEPT9 to detect patients with stage I and II disease, the presence of pre-malignant polyps, and the ability to predict disease recurrence earlier than conventional radiologic imaging (Jin et al. 2021).

Liu et al. (2020) performed WGBS on blood plasma cfDNA samples from 1493 patients of over 50 cancer types and 1135 apparently healthy individuals to identify correlations of DNA methylation with tumor stage. Cancer-specific DNA methylation profiles identified in plasma showed exceptional specificity (99%). Cancer detection sensitivity increased with tumor stage with sensitivities of 18% for stage I patients, 43% in stage II disease, 81% in stage III cases, and 93% in stage IV patients, highlighting the increased cfDNA concentration in blood as a function of tumor stage. A follow-up analysis of 2129 cfDNA samples from cancer patients also predicted patient cancer status, cell of origin, disease mortality, tumor stage, and patient age (Chen et al. 2021).

WGBS analysis of cfDNAs isolated from blood plasma of women with metastatic breast

cancer, disease-free breast cancer survivors, and apparently healthy women unveiled a panel of differential DNA methylation markers that determines the presence of metastatic breast cancer versus disease-free survivors or healthy women (Legendre et al. 2015). In addition, a prediction model of early-stage breast cancer was developed based on cfDNA methylation data from plasma of women with benign disease (Liu et al. 2021). Combining the DNA methylation classifier with diagnostic ultrasound and mammogram imaging improved cancer detection sensitivity and reduced false-negative calls (Liu et al. 2021). Finally, targeted bisulfite sequencing of a panel of classifier candidate genes showed significant *EGFR* and *PPM1E* DNA hypermethylation in cfDNA samples as compared to healthy controls (Li et al. 2016), further implicating the utility of cfDNA methylation as a diagnostic tool.

13.9 DNA Methylation as a Therapeutic Target

Epigenetic aberrations in cancers including differential DNA methylation can be used to distinguish tumor subtypes, indicate treatment responsiveness, predict clinical outcomes, and determine therapeutic strategies. With respect to clinical outcome, epigenetic profiles can reveal molecular pathways most vulnerable to chemotherapeutic agents, and DNA methylation changes can often serve as a barometer for treatment efficacy (Jones et al. 2016; Weisenberger et al. 2018). Unlike genetic modifications, epigenetic changes, most notably DNA methylation and histone modifications, are both somatically heritable and reversible. Thus, DNA methylation changes affected through pharmacological intervention can have long-lasting impact. To this end, DNA methyltransferase inhibitors (DNMTis) have been employed in preclinical and clinical settings with the goal of reversing aberrant DNA methylation to more normal-like profiles (Fig. 13.1c) (Juo et al. 2015; Yamazaki and Issa 2013).

The first generation DNMTis are the cytidine analogs 5-azacytidine (5-Aza-CR) and 5-Aza-2′deoxycytidine (5-Aza-CdR), whose clinical labels are Vidaza and Decitabine, respectively. 5-Aza-CR, an RNA analog, is converted to a nucleoside triphosphate and is incorporated into RNA and DNA. Alternatively, since 5-Aza-CdR is a DNA analog, it is only incorporated into genomic DNA during replication. After incorporation into genomic DNA, the Aza-analogs are recognized as natural cytosine nucleotide substrate by the DNMT enzymes. Normally, DNMTs initiate the methylation transfer reaction by first forming a covalent bond at the C-6 position of the targeted cytosine reside that enables the complex to abstract the methyl group from *S*-adenosyl-L-methionine and transfer it to the C-5 cytosine position. Interestingly, when the DNMT attempts to methylate the Aza-cytosine moiety, the Aza-DNMT tertiary complex is stable, establishing a large DNA adduct structure that triggers proteolytic degradation of the bound DNMT, thereby contributing to the loss of available DNMT enzyme in the cell and DNA methylation marks over subsequent cell divisions (Christman 2002; Stresemann and Lyko 2008). Both 5-Aza-CR and 5-Aza-CdR are currently FDA-approved to treat high-risk myelodysplastic syndrome (MDS) patients and have resulted in successful clinical outcomes (Issa 2013; Navada et al. 2014; Plimack et al. 2007).

The second-generation Aza-cytidine analog, guadecitabine (S110), consists of 5-Aza-CdR linked to a deoxyguanosine residue via a phosphodiester bond (5-Aza-CdR-dG) showing improved stability and activity relative to 5Aza-CdR (Chuang et al. 2010; Yoo et al. 2007). S110 treatment of hepatocellular carcinoma cells also significantly synergized with oxaliplatin and resulted in greater cytotoxicity (Kuang et al. 2015). Unfortunately, the Aza-derived analogs are chemically unstable, suffer from poor pharmacokinetics and their effectiveness is dependent on their transport and uptake into target cells (Stresemann and Lyko 2008) which limits their efficacy in the clinical setting. Recently, a reversible, a non-nucleoside, selective DNMT1

inhibitor, GSK3685032, was described (Pappalardi et al. 2021). GSK3685032 treatment of acute myeloid leukemia (AML) tumor cells showed improved tolerability and more extensive DNA demethylation compared to the Aza-cytidine analogs, resulting in increased anti-tumor activity and survival. Ultimately, GSK3685032 may provide increased clinical efficacy compared to Aza-cytidine analogs.

Numerous tumor suppressor gene promoters have been identified to be activated after DNMTi treatment including *CDKN2A*$^{(INK4A/p16)}$, *MYOD1*, *RASSF1,* and *TIMP3* (Christman 2002; Toyota et al. 2001). The chromatin remodeler protein *CHD5* considered as tumor suppressor in many cancer types is frequently silenced through multiple epigenetic mechanisms including promoter DNA hypermethylation (Fujita et al. 2008; Gorringe et al. 2008; Wang et al. 2011). Colorectal cancer cells treated with 5-Aza-CdR displayed partially restored CHD5 protein expression (Fatemi et al. 2014). In AML cells, DNMTis can initiate apoptosis in a p53-independent manner. Here, 5-Aza-CdR administration demethylates the gene promoter of *TP73*, a member of the p53 family of transcription factors. TP73 expression induces p21 protein expression, which in turn renders the cell more sensitive to chemotherapeutics and mediates drug cytotoxicity (Schmelz et al. 2005).

Improved understanding of intergenic DNA methylation has shown that in addition to promoter DNA methylation, gene body DNA methylation also serves as a therapeutic target for demethylating chemotherapeutic agents such as 5-AzaCdR (Fig. 13.1b, c). Genome-wide DNA methylation levels were assayed at various time points after 5-AzaCdR treatment of the HCT116 human colorectal cell line. Gene body DNA demethylation correlated with loss of gene expression and the rate of DNA re-methylation after drug withdrawal determined the strength of gene re-expression. Evaluating DNMTi treatments of HCT116 derivative lines lacking various DNA methyltransferases showed that gene body DNA re-methylation was dependent on DNMT3B. Moreover, clustering the genomic regions into groups according to the rates of DNA re-methylation indicated that rapidly remethylating genes are enriched for oncogenes such as c-MYC targets and metabolic pathway genes. This unveils a potential mechanism of action for DNMTis in mitigating the effect of deregulated c-MYC signaling (Kasinathan and Henikoff 2014; Yang et al. 2014). Moreover, gene body DNA methylation is a therapeutic DNMTi target for lowering oncogene gene expression and reversing the cellular machinery to a more normal-like state. Low DNMTi doses are sufficient for long-lasting loss of tumorigenicity and self-renewal with minimal cytotoxic effect, indicating that supplementary to acute tumor suppressor gene re-expression or oncogene downregulation, DNMTis target aberrant DNA methylation via unique mechanism(s) of action (Licht 2015; Oki et al. 2007; Tsai et al. 2012).

DNA demethylating agents may also mediate therapeutic response by rendering the cell more visible to the immune system (Chiappinelli et al. 2015; Jones et al. 2019; Roulois et al. 2015). There is substantial evidence of tumor antigen DNA demethylation and activation after 5-Aza-CdR treatment of tumor cells (De Smet et al. 1996; Shiohama et al. 2014). Indeed, human epithelial ovarian carcinoma (EOC) cells treated with S110 displayed increased cancer-testis antigen expression, thereby, enhancing the recognition of EOC cells by antigen-specific CD8+ T-cells. This was accompanied by restricted tumor growth and improved survival in a xenograft setting (Srivastava et al. 2015). In clinical trials involving patients with non-small cell lung cancer (NSCLC), it has been observed that a subset of patients showed a robust response to immune checkpoint blockade therapy after 5-Aza-CdR treatment, suggesting that DNMT inhibition sensitizes tumor cells to immune checkpoint inhibition (Wrangle et al. 2013).

5-Aza-CdR also stimulates the immune system by triggering antiviral immune response and permitting the expression of endogenous retroviruses (ERVs) that were originally silenced by DNA hypermethylation (Fig. 13.2). Moreover, interferon genes and genes involved in antigen presentation accounted for most genes commonly upregulated in solid tumor cell lines upon Aza

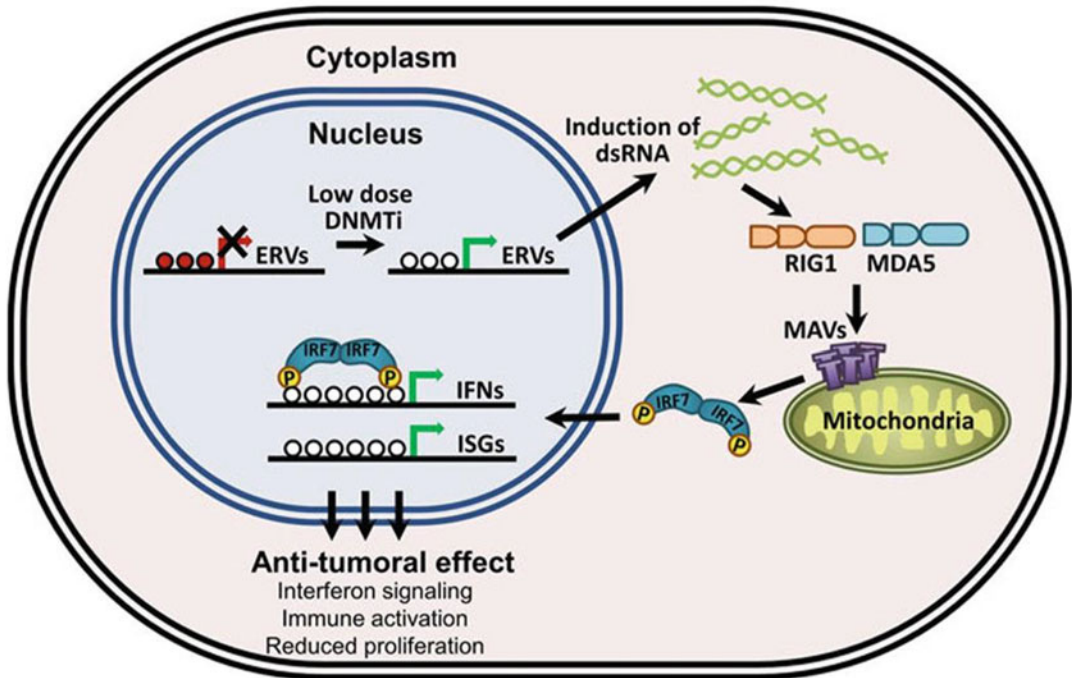


Fig. 13.2 DNMTis exert anti-tumoral effect by eliciting immune response in cancer cells. Treatment with DNA methyltransferase inhibitors induces transcription of endogenous retroviral (ERV) elements. Double-stranded RNAs (dsRNAs) are recognized by viral recognition proteins such as RIG1 and MDA5, which in turn interact with the mitochondrial antiviral signaling (MAVS) proteins. MAVS-mediated IRF7 activation leads to the translocation of IRF7 from the cytoplasm to the nucleus, where it initiates transcription of interferons (IFNs) and interferon-stimulated genes (ISGs), which then contribute to reduced proliferation (modified from Chiappinelli et al. 2015; Licht 2015; Roulois et al. 2015)

treatment (Li et al. 2014a). Gene expression profiling after transient low dose 5-Aza-CdR treatments of colorectal cell lines demonstrated that most of the late occurring expression changes (24 days past initial exposure) were interferon-responsive genes. These genes showed little DNA methylation changes at their promoters or coding regions, and in fact, many even displayed low pre-treatment DNA methylation levels. Thus, gene expression changes after 5-Aza-CdR treatment were independent of DNA methylation inhibition.

The mechanistic explanation for how DNMTis activate unmethylated gene promoters of immune response genes is based on double-stranded RNA (dsRNA), endogenous retrovirus (ERV), and/or *ALU* activation after DNMTi treatment of cancer cells (Mehdipour et al. 2020; Roulois et al. 2015). dsRNAs are recognized by RIG1 and MDA5, cytosolic pattern recognition receptors whose primary role is to recognize viral RNAs. This recognition initiates a signaling cascade dependent on the mitochondrial antiviral signaling (MAVS) adaptor molecule, leading to an activation of downstream targets, such as IRF7, culminating in an anti-tumor response. This mechanism was also shown using human ovarian cancer cell lines, in which 5-Aza-CdR treatment activates interferon signaling mediated by IRF7. Furthermore, the strength of interferon response to 5-Aza-CdR was reflected in the immune checkpoint sensitivity (Chiappinelli et al. 2015). Drug-induced DNA demethylation of repetitive elements and ERVs, together with dsRNA activation, elicits a strong antiviral response. This consequently results in anti-tumor effects including interferon induction,

reduced cell proliferation, and loss of self-renewal capacity upon treatment (Fig. 13.2). In addition, DNMTis can target not only cancer cells but also reactivate exhausted T-cells (Loo Yau et al. 2021). These mechanistic aspects provide important evidence for not only the use of DNMTis in cancer treatment, but their combination with immune checkpoint inhibitors as an efficacious treatment scheme (Jones et al. 2019).

13.10 Concluding Remarks

DNA methylation is a complex epigenetic process crucial to regulating gene expression in normal and tumor cells. Methylation of CpGs at the promoters of genes attenuates their expression, while gene body methylation levels positively correlate with expression. By modulating gene expression, DNA methylation alters signaling pathways that affect cellular processes such as cell cycle, DNA repair, cell growth, and proliferation. Dysregulation of DNA methylation can, therefore, lead to inappropriate silencing of tumor suppressors or expression of oncogenes, thus contributing to the development of disease states, most notably in human cancers. However, unlike genetic changes, DNA methylation alterations can be potentially reversed by DNA methylation inhibitors with therapeutic effects of tumor suppressor gene reactivation, oncogene downregulation, and immune response stimulation. Genome-wide screens can be efficiently used to identify genes that are influenced by the pathways being affected by aberrant DNA methylation. Furthermore, with improved access to next-generation sequencing, large-scale multinational consortia led research has resulted in a wealth of genomic and epigenomic data and the identification of actionable therapeutic targets. Integrating molecular data profiles with patient clinical co-variates enables putative therapeutic epigenetic target discovery and validation, as well as stratifying tumors into clinically relevant subgroups according to molecular features, thereby, facilitating more effective and personalized therapeutic strategies.

Acknowledgments The work in the Liang laboratory has been supported in part by the generous contribution of George and Vicky Joseph.

References

Abdel Ghani L, Yusenko MV, Frank D, Moorthy R, Widen JC, Dorner W, Khandanpour C, Harki DA, Klempnauer KH (2022) A synthetic covalent ligand of the C/EBPbeta transactivation domain inhibits acute myeloid leukemia cells. Cancer Lett 530:170–180. https://doi.org/10.1016/j.canlet.2022.01.024

Agirre X, Vilas-Zornoza A, Jimenez-Velasco A, Martin-Subero JI, Cordeu L, Garate L, San Jose-Eneriz E, Abizanda G, Rodriguez-Otero P, Fortes P et al (2009) Epigenetic silencing of the tumor suppressor microRNA Hsa-miR-124a regulates CDK6 expression and confers a poor prognosis in acute lymphoblastic leukemia. Cancer Res 69:4443–4453. https://doi.org/10.1158/0008-5472.CAN-08-4025

Alvarez-Nunez F, Bussaglia E, Mauricio D, Ybarra J, Vilar M, Lerma E, de Leiva A, Matias-Guiu X, Thyroid Neoplasia Study Group (2006) PTEN promoter methylation in sporadic thyroid carcinomas. Thyroid 16:17–23. https://doi.org/10.1089/thy.2006.16.17

Andreotti G, Karami S, Pfeiffer RM, Hurwitz L, Liao LM, Weinstein SJ, Albanes D, Virtamo J, Silverman DT, Rothman N, Moore LE (2014) LINE1 methylation levels associated with increased bladder cancer risk in pre-diagnostic blood DNA among US (PLCO) and European (ATBC) cohort study participants. Epigenetics 9:404–415. https://doi.org/10.4161/epi.27386

Angeles AK, Janke F, Bauer S, Christopoulos P, Riediger AL, Sultmann H (2021) Liquid biopsies beyond mutation calling: genomic and epigenomic features of cell-free DNA in cancer. Cancers (Basel) 13. https://doi.org/10.3390/cancers13225615

Aran D, Hellman A (2013) DNA methylation of transcriptional enhancers and cancer predisposition. Cell 154:11–13. https://doi.org/10.1016/j.cell.2013.06.018

Aymard F, Bugler B, Schmidt CK, Guillou E, Caron P, Briois S, Iacovoni JS, Daburon V, Miller KM, Jackson SP, Legube G (2014) Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. Nat Struct Mol Biol 21:366–374. https://doi.org/10.1038/nsmb.2796

Barefoot ME, Loyfer N, Kiliti AJ, McDeed AP, Kaplan T, Wellstein A (2021) Detection of cell types contributing to cancer from circulating, cell-free methylated DNA. Front Genet 12:671057. https://doi.org/10.3389/fgene.2021.671057

Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, Akalin A, Schubeler D (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520:243–247. https://doi.org/10.1038/nature14176

Baylin SB, Jones PA (2011) A decade of exploring the cancer epigenome—biological and translational implications. Nat Rev Cancer 11:726–734. https://doi.org/10.1038/nrc3130

Bender S, Tang Y, Lindroth AM, Hovestadt V, Jones DT, Kool M, Zapatka M, Northcott PA, Sturm D, Wang W et al (2013) Reduced H3K27me3 and DNA hypomethylation are major drivers of gene expression in K27M mutant pediatric high-grade gliomas. Cancer Cell 24:660–672. https://doi.org/10.1016/j.ccr.2013.10.006

Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, Noushmehr H, Lange CP, van Dijk CM, Tollenaar RA et al (2011) Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. Nat Genet 44:40–46. https://doi.org/10.1038/ng.969

Bestor TH, Edwards JR, Boulard M (2015) Notes on the role of dynamic DNA methylation in mammalian development. Proc Natl Acad Sci USA 112:6796–6799. https://doi.org/10.1073/pnas.1415301111

Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, Bartlett BR, Wang H, Luber B, Alani RM et al (2014) Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med 6:224ra224. https://doi.org/10.1126/scitranslmed.3007094

Bhattacharya S, Levy MJ, Zhang N, Li H, Florens L, Washburn MP, Workman JL (2021) The methyltransferase SETD2 couples transcription and splicing by engaging mRNA processing factors through its SHI domain. Nat Commun 12:1443. https://doi.org/10.1038/s41467-021-21663-w

Bjornsson HT, Brown LJ, Fallin MD, Rongione MA, Bibikova M, Wickham E, Fan JB, Feinberg AP (2007) Epigenetic specificity of loss of imprinting of the IGF2 gene in Wilms tumors. J Natl Cancer Inst 99:1270–1273. https://doi.org/10.1093/jnci/djm069

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints. Science 294:2536–2539. https://doi.org/10.1126/science.1065848

Brennan CW, Verhaak RG, McKenna A, Campos B, Noushmehr H, Salama SR, Zheng S, Chakravarty D, Sanborn JZ, Berman SH et al (2013) The somatic genomic landscape of glioblastoma. Cell 155:462–477. https://doi.org/10.1016/j.cell.2013.09.034

Brocks D, Schmidt CR, Daskalakis M, Jang HS, Shah NM, Li D, Li J, Zhang B, Hou Y, Laudato S et al (2017) DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. Nat Genet 49:1052–1060. https://doi.org/10.1038/ng.3889

Bueno R, Stawiski EW, Goldstein LD, Durinck S, De Rienzo A, Modrusan Z, Gnad F, Nguyen TT, Jaiswal BS, Chirieac LR et al (2016) Comprehensive genomic analysis of malignant pleural mesothelioma identifies recurrent mutations, gene fusions and splicing alterations. Nat Genet 48:407–416. https://doi.org/10.1038/ng.3520

Bulger M, Groudine M (2011) Functional and mechanistic diversity of distal transcription enhancers. Cell 144:327–339. https://doi.org/10.1016/j.cell.2011.01.024

Cai G, Cai M, Feng Z, Liu R, Liang L, Zhou P, ColonAi QG, Zhu B, Mo S, Wang H et al (2021) A multilocus blood-based assay targeting circulating tumor DNA methylation enables early detection and early relapse prediction of colorectal cancer. Gastroenterology 161:2053–2056 e2052. https://doi.org/10.1053/j.gastro.2021.08.054

Calo E, Wysocka J (2013) Modification of enhancer chromatin: what, how, and why? Mol Cell 49:825–837. https://doi.org/10.1016/j.molcel.2013.01.038

Carvalho S, Vitor AC, Sridhara SC, Martins FB, Raposo AC, Desterro JM, Ferreira J, de Almeida SF (2014) SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. Elife 3:e02482. https://doi.org/10.7554/eLife.02482

Chadwick LH (2012) The NIH Roadmap Epigenomics Program data resource. Epigenomics 4:317–324. https://doi.org/10.2217/epi.12.18

Chan KM, Fang D, Gan H, Hashizume R, Yu C, Schroeder M, Gupta N, Mueller S, James CD, Jenkins R et al (2013) The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. Genes Dev 27:985–990. https://doi.org/10.1101/gad.217778.113

Charlet J, Duymich CE, Lay FD, Mundbjerg K, Dalsgaard Sorensen K, Liang G, Jones PA (2016) Bivalent regions of cytosine methylation and H3K27 acetylation suggest an active role for DNA methylation at enhancers. Mol Cell 62:422–431. https://doi.org/10.1016/j.molcel.2016.03.033

Chen Z, Zhang Y (2020) Maternal H3K27me3-dependent autosomal and X chromosome imprinting. Nat Rev Genet 21:555–571. https://doi.org/10.1038/s41576-020-0245-9

Chen T, Ueda Y, Xie S, Li E (2002) A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. J Biol Chem 277:38746–38754. https://doi.org/10.1074/jbc.M205312200

Chen X, Gole J, Gore A, He Q, Lu M, Min J, Yuan Z, Yang X, Jiang Y, Zhang T et al (2020) Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. Nat Commun 11:3475. https://doi.org/10.1038/s41467-020-17316-z

Chen X, Dong Z, Hubbell E, Kurtzman KN, Oxnard GR, Venn O, Melton C, Clarke CA, Shaknovich R, Ma T et al (2021) Prognostic significance of blood-based multi-cancer detection in plasma cell-free DNA. Clin Cancer Res 27:4221–4229. https://doi.org/10.1158/1078-0432.CCR-21-0417

Chiang JW, Karlan BY, Cass L, Baldwin RL (2006) BRCA1 promoter methylation predicts adverse

ovarian cancer prognosis. Gynecol Oncol 101:403–410. https://doi.org/10.1016/j.ygyno.2005.10.034

Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, Hein A, Rote NS, Cope LM, Snyder A et al (2015) Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. Cell 162:974–986. https://doi.org/10.1016/j.cell.2015.07.011

Christman JK (2002) 5-Azacytidine and 5-aza-2-′-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene 21:5483–5495. https://doi.org/10.1038/sj.onc.1205699.

Chuang JC, Warner SL, Vollmer D, Vankayalapati H, Redkar S, Bearss DJ, Qiu X, Yoo CB, Jones PA (2010) S110, a 5-Aza-2′-deoxycytidine-containing dinucleotide, is an effective DNA methylation inhibitor in vivo and can reduce tumor growth. Mol Cancer Ther 9:1443–1450. https://doi.org/10.1158/1535-7163.MCT-09-1048

Cooper LA, Demicco EG, Saltz JH, Powell RT, Rao A, Lazar AJ (2018) PanCancer insights from The Cancer Genome Atlas: the pathologist's perspective. J Pathol 244:512–524. https://doi.org/10.1002/path.5028

Cui H (2007) Loss of imprinting of IGF2 as an epigenetic marker for the risk of human cancer. Dis Markers 23: 105–112. https://doi.org/10.1155/2007/363464

Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, Fantin VR, Jang HG, Jin S, Keenan MC et al (2009) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature 462:739–744. https://doi.org/10.1038/nature08617

de Almeida SF, Grosso AR, Koch F, Fenouil R, Carvalho S, Andrade J, Levezinho H, Gut M, Eick D, Gut I et al (2011) Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. Nat Struct Mol Biol 18:977–983. https://doi.org/10.1038/nsmb.2123

Decato BE, Qu J, Ji X, Wagenblast E, Knott SRV, Hannon GJ, Smith AD (2020) Characterization of universal features of partially methylated domains across tissues and species. Epigenetics Chromatin 13:39. https://doi.org/10.1186/s13072-020-00363-7

Deininger PL, Moran JV, Batzer MA, Kazazian HH Jr (2003) Mobile elements and mammalian genome evolution. Curr Opin Genet Dev 13:651–658. https://doi.org/10.1016/j.gde.2003.10.013

De Smet C, De Backer O, Faraoni I, Lurquin C, Brasseur F, Boon T (1996) The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. Proc Natl Acad Sci U S A 93:7149–7153. https://doi.org/10.1073/pnas.93.14.7149

deVos T, Tetzner R, Model F, Weiss G, Schuster M, Distler J, Steiger KV, Grutzmann R, Pilarsky C, Habermann JK et al (2009) Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer. Clin Chem 55:1337–1346. https://doi.org/10.1373/clinchem.2008.115808

Donson AM, Addo-Yobo SO, Handler MH, Gore L, Foreman NK (2007) MGMT promoter methylation correlates with survival benefit and sensitivity to temozolomide in pediatric glioblastoma. Pediatr Blood Cancer 48:403–407. https://doi.org/10.1002/pbc.20803

Du Q, Bert SA, Armstrong NJ, Caldon CE, Song JZ, Nair SS, Gould CM, Luu PL, Peters T, Khoury A et al (2019) Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. Nat Commun 10:416. https://doi.org/10.1038/s41467-019-08302-1

Duymich CE, Charlet J, Yang X, Jones PA, Liang G (2016) DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. Nat Commun 7:11453. https://doi.org/10.1038/ncomms11453

Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA et al (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 38:1378–1385. https://doi.org/10.1038/ng1909

Ecsedi SI, Hernandez-Vargas H, Lima SC, Herceg Z, Adany R, Balazs M (2013) Transposable hypomethylation is associated with metastatic capacity of primary melanomas. Int J Clin Exp Pathol 6:2943–2948

Ehrlich M (2002) DNA methylation in cancer: too much, but also too little. Oncogene 21:5400–5413. https://doi.org/10.1038/sj.onc.1205651

Ehrlich M (2009) DNA hypomethylation in cancer cells. Epigenomics 1:239–259. https://doi.org/10.2217/epi.09.33

Ehrlich M, Lacey M (2013) DNA hypomethylation and hemimethylation in cancer. Adv Exp Med Biol 754: 31–56. https://doi.org/10.1007/978-1-4419-9967-2_2

Elhardt W, Shanmugam R, Jurkowski TP, Jeltsch A (2015) Somatic cancer mutations in the DNMT2 tRNA methyltransferase alter its catalytic properties. Biochimie 112:66–72. https://doi.org/10.1016/j.biochi.2015.02.022

ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306:636–640. https://doi.org/10.1126/science.1105136

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74. https://doi.org/10.1038/nature11247

ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET et al (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447:799–816. https://doi.org/10.1038/nature05874

Esteller M (2011) Non-coding RNAs in human disease. Nat Rev Genet 12:861–874. https://doi.org/10.1038/nrg3074

Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, Bussaglia E, Prat J, Harkes IC, Repasky EA et al (2000) Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. J Natl Cancer Inst 92:564–569. https://doi.org/10.1093/jnci/92.7.564

Fang F, Turcan S, Rimner A, Kaufman A, Giri D, Morris LG, Shen R, Seshan V, Mo Q, Heguy A et al (2011) Breast cancer methylomes establish an epigenomic foundation for metastasis. Sci Transl Med 3:75ra25. https://doi.org/10.1126/scitranslmed.3001875

Fatemi M, Paul TA, Brodeur GM, Shokrani B, Brim H, Ashktorab H (2014) Epigenetic silencing of CHD5, a novel tumor-suppressor gene, occurs in early colorectal cancer stages. Cancer 120:172–180. https://doi.org/10.1002/cncr.28316

Feinberg AP, Vogelstein B (1983) Hypomethylation of ras oncogenes in primary human cancers. Biochem Biophys Res Commun 111:47–54. https://doi.org/10.1016/s0006-291x(83)80115-6

Ferguson-Smith AC (2011) Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet 12:565–575. https://doi.org/10.1038/nrg3032

Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF et al (2010) Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell 18:553–567. https://doi.org/10.1016/j.ccr.2010.11.015

Friedman JM, Liang G, Liu CC, Wolff EM, Tsai YC, Ye W, Zhou X, Jones PA (2009) The putative tumor suppressor microRNA-101 modulates the cancer epigenome by repressing the polycomb group protein EZH2. Cancer Res 69:2623–2629. https://doi.org/10.1158/0008-5472.CAN-08-3114

Fujita T, Igarashi J, Okawa ER, Gotoh T, Manne J, Kolla V, Kim J, Zhao H, Pawel BR, London WB et al (2008) CHD5, a tumor suppressor gene deleted from 1p36.31 in neuroblastomas. J Natl Cancer Inst 100:940–949. https://doi.org/10.1093/jnci/djn176

Gaal Z (2021) MicroRNAs and metabolism: revisiting the Warburg effect with emphasis on epigenetic background and clinical applications. Biomolecules 11. https://doi.org/10.3390/biom11101531

Gama-Sosa MA, Slagel VA, Trewyn RW, Oxenhandler R, Kuo KC, Gehrke CW, Ehrlich M (1983) The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res 11:6883–6894. https://doi.org/10.1093/nar/11.19.6883

Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. J Mol Biol 196:261–282. https://doi.org/10.1016/0022-2836(87)90689-9

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golic KG, Jacobsen SE, Bestor TH (2006) Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science 311:395–398. https://doi.org/10.1126/science.1120976

Gorringe KL, Choong DY, Williams LH, Ramakrishna M, Sridhar A, Qiu W, Bearfoot JL, Campbell IG (2008) Mutation and methylation analysis of the chromodomain-helicase-DNA binding 5 gene in ovarian cancer. Neoplasia 10:1253–1258. https://doi.org/10.1593/neo.08718

Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P et al (2015) The consensus molecular subtypes of colorectal cancer. Nat Med 21:1350–1356. https://doi.org/10.1038/nm.3967

Gujar H, Weisenberger DJ, Liang G (2019) The roles of human DNA methyltransferases and their isoforms in shaping the epigenome. Genes (Basel) 10. https://doi.org/10.3390/genes10020172

Hakimi AA, Chen YB, Wren J, Gonen M, Abdel-Wahab-O, Heguy A, Liu H, Takeda S, Tickoo SK, Reuter VE et al (2013) Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. Eur Urol 63:848–854. https://doi.org/10.1016/j.eururo.2012.09.005

Hama N, Totoki Y, Miura F, Tatsuno K, Saito-Adachi M, Nakamura H, Arai Y, Hosoda F, Urushidate T, Ohashi S et al (2018) Epigenetic landscape influences the liver cancer genome architecture. Nat Commun 9:1643. https://doi.org/10.1038/s41467-018-03999-y

Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144:646–674. https://doi.org/10.1016/j.cell.2011.02.013

Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep D et al (2011) Increased methylation variation in epigenetic domains across cancer types. Nat Genet 43:768–775. https://doi.org/10.1038/ng.865

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

Helbo AS, Treppendahl M, Aslan D, Dimopoulos K, Nandrup-Bus C, Holm MS, Andersen MK, Liang G, Kristensen LS, Gronbaek K (2015) Hypermethylation of the VTRNA1-3 promoter is associated with poor outcome in lower risk myelodysplastic syndrome patients. Genes (Basel) 6:977–990. https://doi.org/10.3390/genes6040977

Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, Liu S, Samid D, Duan DS, Gnarra JR, Linehan WM et al (1994) Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. Proc Natl Acad Sci U S A 91:9700–9704. https://doi.org/10.1073/pnas.91.21.9700

Heyn H, Vidal E, Ferreira HJ, Vizoso M, Sayols S, Gomez A, Moran S, Boque-Sastre R, Guil S, Martinez-Cardus A et al (2016) Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. Genome Biol 17:11. https://doi.org/10.1186/s13059-016-0879-2

Hill VK, Shinawi T, Ricketts CJ, Krex D, Schackert G, Bauer J, Wei W, Cruickshank G, Maher ER, Latif F

(2014) Stability of the CpG island methylator phenotype during glioma progression and identification of methylated loci in secondary glioblastomas. BMC Cancer 14:506. https://doi.org/10.1186/1471-2407-14-506

Hinoue T, Weisenberger DJ, Lange CP, Shen H, Byun HM, Van Den Berg D, Malik S, Pan F, Noushmehr H, van Dijk CM et al (2012) Genome-scale analysis of aberrant DNA methylation in colorectal cancer. Genome Res 22:271–282. https://doi.org/10.1101/gr.117523.110

Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA (2013) Super-enhancers in the control of cell identity and disease. Cell 155:934–947. https://doi.org/10.1016/j.cell.2013.09.053

Ho TH, Kapur P, Joseph RW, Serie DJ, Eckel-Passow JE, Tong P, Wang J, Castle EP, Stanton ML, Cheville JC et al (2016a) Loss of histone H3 lysine 36 trimethylation is associated with an increased risk of renal cell carcinoma-specific death. Mod Pathol 29:34–42. https://doi.org/10.1038/modpathol.2015.123

Ho TH, Park IY, Zhao H, Tong P, Champion MD, Yan H, Monzon FA, Hoang A, Tamboli P, Parker AS et al (2016b) High-resolution profiling of histone h3 lysine 36 trimethylation in metastatic renal cell carcinoma. Oncogene 35:1565–1574. https://doi.org/10.1038/onc.2015.221

Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, Valsesia A, Ye Z, Kuan S, Edsall LE et al (2012) Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res 22:246–258. https://doi.org/10.1101/gr.125872.111

Hovestadt V, Jones DT, Picelli S, Wang W, Kool M, Northcott PA, Sultan M, Stachurski K, Ryzhova M, Warnatz HJ et al (2014) Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. Nature 510:537–541. https://doi.org/10.1038/nature13268

Hsieh JJ, Chen D, Wang PI, Marker M, Redzematovic A, Chen YB, Selcuklu SD, Weinhold N, Bouvier N, Huberman KH et al (2017) Genomic biomarkers of a randomized trial comparing first-line everolimus and sunitinib in patients with metastatic renal cell carcinoma. Eur Urol 71:405–414. https://doi.org/10.1016/j.eururo.2016.10.007

Huang P, Xu M, Han H, Zhao X, Li MD, Yang Z (2021) Integrative analysis of epigenome and transcriptome data reveals aberrantly methylated promoters and enhancers in hepatocellular carcinoma. Front Oncol 11:769390. https://doi.org/10.3389/fonc.2021.769390

Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M et al (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet 41:178–186. https://doi.org/10.1038/ng.298

Ishak CA, Classon M, De Carvalho DD (2018) Deregulation of retroelements as an emerging therapeutic opportunity in cancer. Trends Cancer 4:583–597. https://doi.org/10.1016/j.trecan.2018.05.008

Issa JP (1999) Aging, DNA methylation and cancer. Crit Rev Oncol Hematol 32:31–43. https://doi.org/10.1016/s1040-8428(99)00019-0

Issa JP (2013) The myelodysplastic syndrome as a prototypical epigenetic disease. Blood 121:3811–3817. https://doi.org/10.1182/blood-2013-02-451757

Issa JP, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, Baylin SB (1994) Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. Nat Genet 7:536–540. https://doi.org/10.1038/ng0894-536

Jang HS, Shah NM, Du AY, Dailey ZZ, Pehrsson EC, Godoy PM, Zhang D, Li D, Xing X, Kim S et al (2019) Transposable elements drive widespread expression of oncogenes in human cancers. Nat Genet 51:611–617. https://doi.org/10.1038/s41588-019-0373-3

Jin S, Zhu D, Shao F, Chen S, Guo Y, Li K, Wang Y, Ding R, Gao L, Ma W et al (2021) Efficient detection and post-surgical monitoring of colon cancer with a multi-marker DNA methylation liquid biopsy. Proc Natl Acad Sci U S A 118. https://doi.org/10.1073/pnas.2017421118

Jones PA, Baylin SB (2002) The fundamental role of epigenetic events in cancer. Nat Rev Genet 3:415–428. https://doi.org/10.1038/nrg816

Jones PA, Baylin SB (2007) The epigenomics of cancer. Cell 128:683–692. https://doi.org/10.1016/j.cell.2007.01.029

Jones PA, Laird PW (1999) Cancer epigenetics comes of age. Nat Genet 21:163–167. https://doi.org/10.1038/5947

Jones PA, Issa JP, Baylin S (2016) Targeting the cancer epigenome for therapy. Nat Rev Genet 17:630–641. https://doi.org/10.1038/nrg.2016.93

Jones PA, Ohtani H, Chakravarthy A, De Carvalho DD (2019) Epigenetic therapy in immune-oncology. Nat Rev Cancer 19:151–161. https://doi.org/10.1038/s41568-019-0109-9

Jordan IK, Rogozin IB, Glazko GV, Koonin EV (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet 19:68–72. https://doi.org/10.1016/s0168-9525(02)00006-9

Jung M, Pfeifer GP (2015) Aging and DNA methylation. BMC Biol 13:7. https://doi.org/10.1186/s12915-015-0118-4

Juo YY, Gong XJ, Mishra A, Cui X, Baylin SB, Azad NS, Ahuja N (2015) Epigenetic therapy for solid tumors: from bench science to clinical trials. Epigenomics 7:215–235. https://doi.org/10.2217/epi.14.73

Karami S, Andreotti G, Liao LM, Pfeiffer RM, Weinstein SJ, Purdue MP, Hofmann JN, Albanes D, Mannisto S, Moore LE (2015) LINE1 methylation levels in pre-diagnostic leukocyte DNA and future renal cell

carcinoma risk. Epigenetics 10:282–292. https://doi.org/10.1080/15592294.2015.1006505

Kasinathan S, Henikoff S (2014) 5-Aza-CdR delivers a gene body blow. Cancer Cell 26:449–451. https://doi.org/10.1016/j.ccell.2014.09.004

Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, An J, Lamperti ED, Koh KP, Ganetzky R et al (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature 468:839–843. https://doi.org/10.1038/nature09586

Kohli RM, Zhang Y (2013) TET enzymes, TDG and the dynamics of DNA demethylation. Nature 502:472–479. https://doi.org/10.1038/nature12750

Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet 41:376–381. https://doi.org/10.1038/ng.322

Konkel MK, Batzer MA (2010) A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. Semin Cancer Biol 20:211–221. https://doi.org/10.1016/j.semcancer.2010.03.001

Korshunov A, Jakobiec FA, Eberhart CG, Hovestadt V, Capper D, Jones DT, Sturm D, Stagner AM, Edward DP, Eagle RC et al (2015) Comparative integrated molecular analysis of intraocular medulloepitheliomas and central nervous system embryonal tumors with multilayered rosettes confirms that they are distinct nosologic entities. Neuropathology 35:538–544. https://doi.org/10.1111/neup.12227

Kuang Y, El-Khoueiry A, Taverna P, Ljungman M, Neamati N (2015) Guadecitabine (SGI-110) priming sensitizes hepatocellular carcinoma cells to oxaliplatin. Mol Oncol 9:1799–1814. https://doi.org/10.1016/j.molonc.2015.06.002

Kulis M, Heath S, Bibikova M, Queiros AC, Navarro A, Clot G, Martinez-Trillos A, Castellano G, Brun-Heath-I, Pinyol M et al (2012) Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat Genet 44:1236–1242. https://doi.org/10.1038/ng.2443

Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc Natl Acad Sci U S A 103:10684–10689. https://doi.org/10.1073/pnas.0600326103

Laird PW (2003) The power and the promise of DNA methylation markers. Nat Rev Cancer 3:253–266. https://doi.org/10.1038/nrc1045

Lee KS, Park JL, Lee K, Richardson LE, Johnson BH, Lee HS, Lee JS, Kim SB, Kwon OH, Song KS et al (2014) nc886, a non-coding RNA of anti-proliferative role, is suppressed by CpG DNA methylation in human gastric cancer. Oncotarget 5:3944–3955. https://doi.org/10.18632/oncotarget.2047

Legendre C, Gooden GC, Johnson K, Martinez RA, Liang WS, Salhia B (2015) Whole-genome bisulfite sequencing of cell-free DNA identifies signature associated with metastatic breast cancer. Clin Epigenetics 7:100. https://doi.org/10.1186/s13148-015-0135-8

Levenson VV (2010) DNA methylation as a universal biomarker. Expert Rev Mol Diagn 10:481–488. https://doi.org/10.1586/erm.10.17

Lewis PW, Muller MM, Koletsky MS, Cordero F, Lin S, Banaszynski LA, Garcia BA, Muir TW, Becher OJ, Allis CD (2013) Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science 340:857–861. https://doi.org/10.1126/science.1232245

Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J et al (2010) DNMT3A mutations in acute myeloid leukemia. N Engl J Med 363:2424–2433. https://doi.org/10.1056/NEJMoa1005143

Li E, Beard C, Jaenisch R (1993) Role for DNA methylation in genomic imprinting. Nature 366:362–365. https://doi.org/10.1038/366362a0

Li F, Mao G, Tong D, Huang J, Gu L, Yang W, Li GM (2013) The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha. Cell 153:590–600. https://doi.org/10.1016/j.cell.2013.03.025

Li H, Chiappinelli KB, Guzzetta AA, Easwaran H, Yen RW, Vatapalli R, Topper MJ, Luo J, Connolly RM, Azad NS et al (2014a) Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers. Oncotarget 5:587–598. https://doi.org/10.18632/oncotarget.1782

Li J, Huang Q, Zeng F, Li W, He Z, Chen W, Zhu W, Zhang B (2014b) The prognostic value of global DNA hypomethylation in cancer: a meta-analysis. PLoS One 9:e106290. https://doi.org/10.1371/journal.pone.0106290

Li Z, Guo X, Tang L, Peng L, Chen M, Luo X, Wang S, Xiao Z, Deng Z, Dai L et al (2016) Methylation analysis of plasma cell-free DNA for breast cancer early detection using bisulfite next-generation sequencing. Tumour Biol 37:13111–13119. https://doi.org/10.1007/s13277-016-5190-z

Liang G, Weisenberger DJ (2017) DNA methylation aberrancies as a guide for surveillance and treatment of human cancers. Epigenetics 12:416–432. https://doi.org/10.1080/15592294.2017.1311434

Liang G, Chan MF, Tomigahara Y, Tsai YC, Gonzales FA, Li E, Laird PW, Jones PA (2002) Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Mol Cell Biol 22:480–491. https://doi.org/10.1128/MCB.22.2.480-491.2002

Licht JD (2015) DNA methylation inhibitors in cancer therapy: the immunity dimension. Cell 162:938–939. https://doi.org/10.1016/j.cell.2015.08.005

Lin DC, Dinh HQ, Xie JJ, Mayakonda A, Silva TC, Jiang YY, Ding LW, He JZ, Xu XE, Hao JJ et al (2018) Identification of distinct mutational patterns and new

driver genes in oesophageal squamous cell carcinomas and adenocarcinomas. Gut 67:1769–1779. https://doi.org/10.1136/gutjnl-2017-314607

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:315–322. https://doi.org/10.1038/nature08514

Liu M, Zhang L, Li H, Hinoue T, Zhou W, Ohtani H, El-Khoueiry A, Daniels J, O'Connell C, Dorff TB et al (2018) Integrative epigenetic analysis reveals therapeutic targets to the DNA methyltransferase inhibitor guadecitabine (SGI-110) in hepatocellular carcinoma. Hepatology 68:1412–1428. https://doi.org/10.1002/hep.30091

Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, CCGA Consortium (2020) Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann Oncol 31:745–759. https://doi.org/10.1016/j.annonc.2020.02.011

Liu J, Zhao H, Huang Y, Xu S, Zhou Y, Zhang W, Li J, Ming Y, Wang X, Zhao S et al (2021) Genome-wide cell-free DNA methylation analyses improve accuracy of non-invasive diagnostic imaging for early-stage breast cancer. Mol Cancer 20:36. https://doi.org/10.1186/s12943-021-01330-w

Lo YM, Chan LY, Chan AT, Leung SF, Lo KW, Zhang J, Lee JC, Hjelm NM, Johnson PJ, Huang DP (1999) Quantitative and temporal correlation between circulating cell-free Epstein-Barr virus DNA and tumor recurrence in nasopharyngeal carcinoma. Cancer Res 59:5452–5455

Loo Yau H, Bell E, Ettayebi I, de Almeida FC, Boukhaled GM, Shen SY, Allard D, Morancho B, Marhon SA, Ishak CA et al (2021) DNA hypomethylating agents increase activation and cytolytic activity of CD8(+) T cells. Mol Cell 81:1469–1483 e1468. https://doi.org/10.1016/j.molcel.2021.01.038

Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, Bradner JE, Lee TI, Young RA (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell 153:320–334. https://doi.org/10.1016/j.cell.2013.03.036

Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T (2010) Regulation of alternative splicing by histone modifications. Science 327:996–1000. https://doi.org/10.1126/science.1184208

Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet 15:394–408. https://doi.org/10.1038/nrg3683

Mehdipour P, Marhon SA, Ettayebi I, Chakravarthy A, Hosseini A, Wang Y, de Castro FA, Loo Yau H, Ishak C, Abelson S et al (2020) Epigenetic therapy induces transcription of inverted SINEs and ADAR1 dependency. Nature 588:169–173. https://doi.org/10.1038/s41586-020-2844-1

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB et al (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454:766–770. https://doi.org/10.1038/nature07107

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP et al (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448:553–560. https://doi.org/10.1038/nature06008

Nagarajan RP, Zhang B, Bell RJ, Johnson BE, Olshen AB, Sundaram V, Li D, Graham AE, Diaz A, Fouse SD et al (2014) Recurrent epimutations activate gene body promoters in primary glioblastoma. Genome Res 24:761–774. https://doi.org/10.1101/gr.164707.113

Navada SC, Steinmann J, Lubbert M, Silverman LR (2014) Clinical development of demethylating agents in hematology. J Clin Invest 124:40–46. https://doi.org/10.1172/JCI69739

Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, Maldotti M, Anselmi F, Oliviero S (2017) Intragenic DNA methylation prevents spurious transcription initiation. Nature 543:72–77. https://doi.org/10.1038/nature21373

Nerlov C (2007) The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control. Trends Cell Biol 17:318–324. https://doi.org/10.1016/j.tcb.2007.07.004

Nordin M, Bergman D, Halje M, Engstrom W, Ward A (2014) Epigenetic regulation of the Igf2/H19 gene cluster. Cell Prolif 47:189–199. https://doi.org/10.1111/cpr.12106

Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, Pan F, Pelloski CE, Sulman EP, Bhat KP et al (2010) Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell 17:510–522. https://doi.org/10.1016/j.ccr.2010.03.017

Ogino S, Kawasaki T, Kirkner GJ, Loda M, Fuchs CS (2006) CpG island methylator phenotype-low (CIMP-low) in colorectal cancer: possible associations with male sex and KRAS mutations. J Mol Diagn 8:582–588. https://doi.org/10.2353/jmoldx.2006.060082

Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99:247–257. https://doi.org/10.1016/s0092-8674(00)81656-6

Oki Y, Aoki E, Issa JP (2007) Decitabine—bedside to bench. Crit Rev Oncol Hematol 61:140–152. https://doi.org/10.1016/j.critrevonc.2006.07.010

Ostler KR, Davis EM, Payne SL, Gosalia BB, Exposito-Cespedes J, Le Beau MM, Godley LA (2007) Cancer cells express aberrant DNMT3B transcripts encoding truncated proteins. Oncogene 26:5553–5563. https://doi.org/10.1038/sj.onc.1210351

Pai CC, Deegan RS, Subramanian L, Gal C, Sarkar S, Blaikley EJ, Walker C, Hulme L, Bernhard E, Codlin S et al (2014) A histone H3K36 chromatin switch coordinates DNA double-strand break repair pathway

choice. Nat Commun 5:4091. https://doi.org/10.1038/ncomms5091

Pappalardi MB, Keenan K, Cockerill M, Kellner WA, Stowell A, Sherk C, Wong K, Pathuri S, Briand J, Steidel M et al (2021) Discovery of a first-in-class reversible DNMT1-selective inhibitor with improved tolerability and efficacy in acute myeloid leukemia. Nat Cancer 2:1002–1017

Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL et al (2008) An integrated genomic analysis of human glioblastoma multiforme. Science 321:1807–1812. https://doi.org/10.1126/science.1164382

Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P et al (2011) Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature 473:394–397. https://doi.org/10.1038/nature10102

Pastor WA, Aravind L, Rao A (2013) TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol 14:341–356. https://doi.org/10.1038/nrm3589

Pessia E, Makino T, Bailly-Bechet M, McLysaght A, Marais GA (2012) Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. Proc Natl Acad Sci U S A 109:5346–5351. https://doi.org/10.1073/pnas.1116763109

Pfeifer GP (2000) p53 mutational spectra and the role of methylated CpG sequences. Mutat Res 450:155–166. https://doi.org/10.1016/s0027-5107(00)00022-1

Pfister SX, Markkanen E, Jiang Y, Sarkar S, Woodcock M, Orlando G, Mavrommati I, Pai CC, Zalmas LP, Drobnitzky N et al (2015) Inhibiting WEE1 selectively kills histone H3K36me3-deficient cancers by dNTP starvation. Cancer Cell 28:557–568. https://doi.org/10.1016/j.ccell.2015.09.015

Plimack ER, Kantarjian HM, Issa JP (2007) Decitabine and its role in the treatment of hematopoietic malignancies. Leuk Lymphoma 48:1472–1481. https://doi.org/10.1080/10428190701471981

Pontier DB, Gribnau J (2011) Xist regulation and function explored. Hum Genet 130:223–236. https://doi.org/10.1007/s00439-011-1008-7

Prak ET, Kazazian HH Jr (2000) Mobile elements and the human genome. Nat Rev Genet 1:134–144. https://doi.org/10.1038/35038572

Pulverer W, Kruusmaa K, Schonthaler S, Huber J, Bitenc M, Bachleitner-Hofmann T, Bhangu JS, Oehler R, Egger G, Weinhausel A (2021) Multiplexed DNA methylation analysis in colorectal cancer using liquid biopsy and its diagnostic and predictive value. Curr Issues Mol Biol 43:1419–1435. https://doi.org/10.3390/cimb43030100

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A 97:5237–5242. https://doi.org/10.1073/pnas.97.10.5237

Rhee I, Jair KW, Yen RW, Lengauer C, Herman JG, Kinzler KW, Vogelstein B, Baylin SB, Schuebel KE (2000) CpG methylation is maintained in human cancer cells lacking DNMT1. Nature 404:1003–1007. https://doi.org/10.1038/35010000

Rhee I, Bachman KE, Park BH, Jair KW, Yen RW, Schuebel KE, Cui H, Feinberg AP, Lengauer C, Kinzler KW et al (2002) DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature 416:552–556. https://doi.org/10.1038/416552a

Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, Reznik E, Bowlby R, Gibb EA, Akbani R, Beroukhim R et al (2018) The cancer genome atlas comprehensive molecular characterization of renal cell carcinoma. Cell Rep 23:313–326 e315. https://doi.org/10.1016/j.celrep.2018.03.075

Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, Han H, Liang G, Jones PA, Pugh TJ et al (2015) DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. Cell 162:961–973. https://doi.org/10.1016/j.cell.2015.07.056

Saied MH, Marzec J, Khalid S, Smith P, Down TA, Rakyan VK, Molloy G, Raghavan M, Debernardi S, Young BD (2012) Genome wide analysis of acute myeloid leukemia reveal leukemia specific methylome and subtype specific hypomethylation of repeats. PLoS One 7:e33213. https://doi.org/10.1371/journal.pone.0033213

Saito Y, Liang G, Egger G, Friedman JM, Chuang JC, Coetzee GA, Jones PA (2006) Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. Cancer Cell 9:435–443. https://doi.org/10.1016/j.ccr.2006.04.020

Sakai T, Toguchida J, Ohtani N, Yandell DW, Rapaport JM, Dryja TP (1991) Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. Am J Hum Genet 48:880–888

Salhab A, Nordstrom K, Gasparoni G, Kattler K, Ebert P, Ramirez F, Arrigoni L, Muller F, Polansky JK, Cadenas C et al (2018) A comprehensive analysis of 195 DNA methylomes reveals shared and cell-specific features of partially methylated domains. Genome Biol 19:150. https://doi.org/10.1186/s13059-018-1510-5

Sandoval J, Esteller M (2012) Cancer epigenomics: beyond genomics. Curr Opin Genet Dev 22:50–55. https://doi.org/10.1016/j.gde.2012.02.008

Schmelz K, Wagner M, Dorken B, Tamm I (2005) 5-Aza-2′-deoxycytidine induces p21WAF expression by demethylation of p73 leading to p53-independent apoptosis in myeloid leukemia. Int J Cancer 114:683–695. https://doi.org/10.1002/ijc.20797

Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, Sturm D, Fontebasso AM, Quang DA, Tonjes M et al (2012) Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric

glioblastoma. Nature 482:226–231. https://doi.org/10.1038/nature10833
Shah MY, Licht JD (2011) DNMT3A mutations in acute myeloid leukemia. Nat Genet 43:289–290. https://doi.org/10.1038/ng0411-289
Sharma S, Kelly TK, Jones PA (2010) Epigenetics in cancer. Carcinogenesis 31:27–36. https://doi.org/10.1093/carcin/bgp220
Sharma S, De Carvalho DD, Jeong S, Jones PA, Liang G (2011) Nucleosomes containing methylated DNA stabilize DNA methyltransferases 3A/3B and ensure faithful epigenetic inheritance. PLoS Genet 7: e1001286. https://doi.org/10.1371/journal.pgen.1001286
Shen H, Laird PW (2013) Interplay between the cancer genome and epigenome. Cell 153:38–55. https://doi.org/10.1016/j.cell.2013.03.008
Shen L, Toyota M, Kondo Y, Lin E, Zhang L, Guo Y, Hernandez NS, Chen X, Ahmed S, Konishi K et al (2007) Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. Proc Natl Acad Sci U S A 104:18654–18659. https://doi.org/10.1073/pnas.0704652104
Shen JZ, Qiu Z, Wu Q, Finlay D, Garcia G, Sun D, Rantala J, Barshop W, Hope JL, Gimple RC et al (2021) FBXO44 promotes DNA replication-coupled repetitive element silencing in cancer cells. Cell 184: 352–369 e323. https://doi.org/10.1016/j.cell.2020.11.042
Shih AH, Abdel-Wahab O, Patel JP, Levine RL (2012) The role of mutations in epigenetic regulators in myeloid malignancies. Nat Rev Cancer 12:599–612. https://doi.org/10.1038/nrc3343
Shiohama Y, Ohtake J, Ohkuri T, Noguchi D, Togashi Y, Kitamura H, Nishimura T (2014) Identification of a meiosis-specific protein, MEIOB, as a novel cancer/testis antigen and its augmented expression in demethylated cancer cells. Immunol Lett 158:175–182. https://doi.org/10.1016/j.imlet.2014.01.004
Shirane K, Miura F, Ito T, Lorincz MC (2020) NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing. Nat Genet 52:1088–1098. https://doi.org/10.1038/s41588-020-0689-z
Shoaib M, Sorensen CS (2015) Epigenetic deficiencies and replicative stress: driving cancer cells to an early grave. Cancer Cell 28:545–547. https://doi.org/10.1016/j.ccell.2015.10.009
Silber JR, Bobola MS, Blank A, Chamberlain MC (2012) O(6)-methylguanine-DNA methyltransferase in glioma therapy: promise and problems. Biochim Biophys Acta 1826:71–82. https://doi.org/10.1016/j.bbcan.2011.12.004
Singh P, Lee DH, Szabo PE (2012) More than insulator: multiple roles of CTCF at the H19-Igf2 imprinted domain. Front Genet 3:214. https://doi.org/10.3389/fgene.2012.00214
Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. Nat Rev Genet 14:204–220. https://doi.org/10.1038/nrg3354
Soes S, Daugaard IL, Sorensen BS, Carus A, Mattheisen M, Alsner J, Overgaard J, Hager H, Hansen LL, Kristensen LS (2014) Hypomethylation and increased expression of the putative oncogene ELMO3 are associated with lung cancer development and metastases formation. Oncoscience 1:367–374. https://doi.org/10.18632/oncoscience.42
Srivastava P, Paluch BE, Matsuzaki J, James SR, Collamat-Lai G, Taverna P, Karpf AR, Griffiths EA (2015) Immunomodulatory action of the DNA methyltransferase inhibitor SGI-110 in epithelial ovarian cancer cells and xenografts. Epigenetics 10:237–246. https://doi.org/10.1080/15592294.2015.1017198
Stegelmann F, Bullinger L, Schlenk RF, Paschka P, Griesshammer M, Blersch C, Kuhn S, Schauer S, Dohner H, Dohner K (2011) DNMT3A mutations in myeloproliferative neoplasms. Leukemia 25:1217–1219. https://doi.org/10.1038/leu.2011.77
Stresemann C, Lyko F (2008) Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. Int J Cancer 123:8–13. https://doi.org/10.1002/ijc.23607
Su SF, de Castro Abreu AL, Chihara Y, Tsai Y, Andreu-Vieyra C, Daneshmand S, Skinner EC, Jones PA, Siegmund KD, Liang G (2014) A panel of three markers hyper- and hypomethylated in urine sediments accurately predicts bladder cancer recurrence. Clin Cancer Res 20:1978–1989. https://doi.org/10.1158/1078-0432.CCR-13-2637
Sun K, Jiang P, Chan KC, Wong J, Cheng YK, Liang RH, Chan WK, Ma ES, Chan SL, Cheng SH et al (2015) Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. Proc Natl Acad Sci U S A 112:E5503–E5512. https://doi.org/10.1073/pnas.1508736112
Taback B, Hoon DS (2004) Circulating nucleic acids in plasma and serum: past, present and future. Curr Opin Mol Ther 6:273–278
Takai D, Jones PA (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A 99:3740–3745. https://doi.org/10.1073/pnas.052410099
Takai D, Gonzales FA, Tsai YC, Thayer MJ, Jones PA (2001) Large scale mapping of methylcytosines in CTCF-binding sites in the human H19 promoter and aberrant hypomethylation in human bladder cancer. Hum Mol Genet 10:2619–2626. https://doi.org/10.1093/hmg/10.23.2619
Taniguchi T, Sullivan MJ, Ogawa O, Reeve AE (1995) Epigenetic changes encompassing the IGF2/H19 locus associated with relaxation of IGF2 imprinting and silencing of H19 in Wilms tumor. Proc Natl Acad Sci U S A 92:2159–2163. https://doi.org/10.1073/pnas.92.6.2159

The Cancer Genome Atlas Research Network (2011) Integrated genomic analyses of ovarian carcinoma. Nature 474:609–615. https://doi.org/10.1038/nature10166

The Cancer Genome Atlas Research Network (2012a) Comprehensive molecular characterization of human colon and rectal cancer. Nature 487:330–337. https://doi.org/10.1038/nature11252

The Cancer Genome Atlas Research Network (2012b) Comprehensive genomic characterization of squamous cell lung cancers. Nature 489:519–525. https://doi.org/10.1038/nature11404

The Cancer Genome Atlas Research Network (2013) Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature 499:43–49. https://doi.org/10.1038/nature12222

The Cancer Genome Atlas Research Network (2014a) Comprehensive molecular characterization of gastric adenocarcinoma. Nature 513:202–209. https://doi.org/10.1038/nature13480

The Cancer Genome Atlas Research Network (2014b) Comprehensive molecular profiling of lung adenocarcinoma. Nature 511:543–550. https://doi.org/10.1038/nature13385

The Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM (2013a) The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 45:1113–1120. https://doi.org/10.1038/ng.2764

The Cancer Genome Atlas Research Network, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A, Hoadley K, Triche TJ Jr, Laird PW et al (2013b) Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 368:2059–2074. https://doi.org/10.1056/NEJMoa1301689

Tomczak K, Czerwinska P, Wiznerowicz M (2015) The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn) 19: A68–A77. https://doi.org/10.5114/wo.2014.47136

Toyota M, Kopecky KJ, Toyota MO, Jair KW, Willman CL, Issa JP (2001) Methylation profiling in acute myeloid leukemia. Blood 97:2823–2829. https://doi.org/10.1182/blood.v97.9.2823

Treppendahl MB, Qiu X, Sogaard A, Yang X, Nandrup-Bus C, Hother C, Andersen MK, Kjeldsen L, Mollgard L, Hellstrom-Lindberg E et al (2012) Allelic methylation levels of the noncoding VTRNA2-1 located on chromosome 5q31.1 predict outcome in AML. Blood 119:206–216. https://doi.org/10.1182/blood-2011-06-362541

Tsai HC, Li H, Van Neste L, Cai Y, Robert C, Rassool FV, Shin JJ, Harbom KM, Beaty R, Pappou E et al (2012) Transient low doses of DNA-demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells. Cancer Cell 21:430–446. https://doi.org/10.1016/j.ccr.2011.12.029

Turajlic S, Xu H, Litchfield K, Rowan A, Chambers T, Lopez JI, Nicol D, O'Brien T, Larkin J, Horswell S et al (2018a) Tracking cancer evolution reveals constrained routes to metastases: TRACERx renal. Cell 173:581–594 e512. https://doi.org/10.1016/j.cell.2018.03.057

Turajlic S, Xu H, Litchfield K, Rowan A, Horswell S, Chambers T, O'Brien T, Lopez JI, Watkins TBK, Nicol D et al (2018b) Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. Cell 173:595–610 e511. https://doi.org/10.1016/j.cell.2018.03.043

Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, Yilmaz E, Campos C, Fabius AW, Lu C, Ward PS et al (2012) IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. Nature 483:479–483. https://doi.org/10.1038/nature10866

Unnikrishnan A, Freeman WM, Jackson J, Wren JD, Porter H, Richardson A (2019) The role of DNA methylation in epigenetics of aging. Pharmacol Ther 195: 172–185. https://doi.org/10.1016/j.pharmthera.2018.11.001

Vander Heiden MG, Cantley LC, Thompson CB (2009) Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science 324:1029–1033. https://doi.org/10.1126/science.1160809

van der Pol Y, Mouliere F (2019) Toward the early detection of cancer by decoding the epigenetic and environmental fingerprints of cell-free DNA. Cancer Cell 36: 350–368. https://doi.org/10.1016/j.ccell.2019.09.003

Venugopal K, Feng Y, Shabashvili D, Guryanova OA (2021) Alterations to DNMT3A in hematologic malignancies. Cancer Res 81:254–263. https://doi.org/10.1158/0008-5472.CAN-20-3033

Vrba L, Garbe JC, Stampfer MR, Futscher BW (2015) A lincRNA connected to cell mortality and epigenetically-silenced in most common human cancers. Epigenetics 10:1074–1083. https://doi.org/10.1080/15592294.2015.1106673

Walter MJ, Ding L, Shen D, Shao J, Grillot M, McLellan M, Fulton R, Schmidt H, Kalicki-Veizer J, O'Laughlin M et al (2011) Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. Leukemia 25:1153–1158. https://doi.org/10.1038/leu.2011.44

Wang J, Walsh G, Liu DD, Lee JJ, Mao L (2006) Expression of Delta DNMT3B variants and its association with promoter methylation of p16 and RASSF1A in primary non-small cell lung cancer. Cancer Res 66: 8361–8366. https://doi.org/10.1158/0008-5472.CAN-06-2031

Wang J, Chen H, Fu S, Xu ZM, Sun KL, Fu WN (2011) The involvement of CHD5 hypermethylation in laryngeal squamous cell carcinoma. Oral Oncol 47:601–608. https://doi.org/10.1016/j.oraloncology.2011.05.003

Wei JW, Huang K, Yang C, Kang CS (2017) Non-coding RNAs as regulators in epigenetics (Review). Oncol Rep 37:3–9. https://doi.org/10.3892/or.2016.5236

Weinberg DN, Papillon-Cavanagh S, Chen H, Yue Y, Chen X, Rajagopalan KN, Horth C, McGuire JT, Xu X, Nikbakht H et al (2019) The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. Nature 573: 281–286. https://doi.org/10.1038/s41586-019-1534-3

Weiner AM (2002) SINEs and LINEs: the art of biting the hand that feeds you. Curr Opin Cell Biol 14:343–350. https://doi.org/10.1016/s0955-0674(02)00338-1

Weisenberger DJ (2014) Characterizing DNA methylation alterations from The Cancer Genome Atlas. J Clin Invest 124:17–23. https://doi.org/10.1172/JCI69740

Weisenberger DJ, Liang G (2015) Contributions of DNA methylation aberrancies in shaping the cancer epigenome. Transl Cancer Res 4:219–234

Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D et al (2006) CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat Genet 38:787–793. https://doi.org/10.1038/ng1834

Weisenberger DJ, Levine AJ, Long TI, Buchanan DD, Walters R, Clendenning M, Rosty C, Joshi AD, Stern MC, LeMarchand L et al (2015) Association of the colorectal CpG island methylator phenotype with molecular features, risk factors, and family history. Cancer Epidemiol Biomarkers Prev 24:512–519. https://doi.org/10.1158/1055-9965.EPI-14-1161

Weisenberger DJ, Liang G, Lenz HJ (2018) DNA methylation aberrancies delineate clinically distinct subsets of colorectal cancer and provide novel targets for epigenetic therapies. Oncogene 37:566–577. https://doi.org/10.1038/onc.2017.374

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153:307–319. https://doi.org/10.1016/j.cell.2013.03.035

Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I, Laird PW (2007) Epigenetic stem cell signature in cancer. Nat Genet 39:157–158. https://doi.org/10.1038/ng1941

Wilhelm BT, Marguerat S, Aligianni S, Codlin S, Watt S, Bahler J (2011) Differential patterns of intronic and exonic DNA regions with respect to RNA polymerase II occupancy, nucleosome density and H3K36me3 marking in fission yeast. Genome Biol 12:R82. https://doi.org/10.1186/gb-2011-12-8-r82

Wolff EM, Byun HM, Han HF, Sharma S, Nichols PW, Siegmund KD, Yang AS, Jones PA, Liang G (2010) Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. PLoS Genet 6:e1000917. https://doi.org/10.1371/journal.pgen.1000917

Wrangle J, Wang W, Koch A, Easwaran H, Mohammad HP, Vendetti F, Vancriekinge W, Demeyer T, Du Z, Parsana P et al (2013) Alterations of immune response of non-small cell lung cancer with azacytidine. Oncotarget 4:2067–2079. https://doi.org/10.18632/oncotarget.1542

Wu MS, Wang HP, Lin CC, Sheu JC, Shun CT, Lee WJ, Lin JT (1997) Loss of imprinting and overexpression of IGF2 gene in gastric adenocarcinoma. Cancer Lett 120:9–14. https://doi.org/10.1016/s0304-3835(97)00279-6

Xiong L, Wu F, Wu Q, Xu L, Cheung OK, Kang W, Mok MT, Szeto LLM, Lun CY, Lung RW et al (2019) Aberrant enhancer hypomethylation contributes to hepatic carcinogenesis through global transcriptional reprogramming. Nat Commun 10:335. https://doi.org/10.1038/s41467-018-08245-z

Xu TH, Liu M, Zhou XE, Liang G, Zhao G, Xu HE, Melcher K, Jones PA (2020) Structure of nucleosome-bound DNA methyltransferases DNMT3A and DNMT3B. Nature 586:151–155. https://doi.org/10.1038/s41586-020-2747-1

Yamazaki J, Issa JP (2013) Epigenetic aspects of MDS and its molecular targeted therapy. Int J Hematol 97:175–182. https://doi.org/10.1007/s12185-012-1197-4

Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, Kos I, Batinic-Haberle I, Jones S, Riggins GJ et al (2009) IDH1 and IDH2 mutations in gliomas. N Engl J Med 360:765–773. https://doi.org/10.1056/NEJMoa0808710

Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G (2014) Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer Cell 26:577–590. https://doi.org/10.1016/j.ccr.2014.07.028

Yegnasubramanian S, Haffner MC, Zhang Y, Gurel B, Cornish TC, Wu Z, Irizarry RA, Morgan J, Hicks J, DeWeese TL et al (2008) DNA hypomethylation arises later in prostate cancer progression than CpG island hypermethylation and contributes to metastatic tumor heterogeneity. Cancer Res 68:8954–8967. https://doi.org/10.1158/0008-5472.CAN-07-6088

Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F et al (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 356. https://doi.org/10.1126/science.aaj2239

Yoo CB, Jeong S, Egger G, Liang G, Phiasivongsa P, Tang C, Redkar S, Jones PA (2007) Delivery of 5-aza-2′-deoxycytidine to cells using oligodeoxynucleotides. Cancer Res 67:6400–6408. https://doi.org/10.1158/0008-5472.CAN-07-0251

You JS, Jones PA (2012) Cancer genetics and epigenetics: two sides of the same coin? Cancer Cell 22:9–20. https://doi.org/10.1016/j.ccr.2012.06.008

Yuan H, Li N, Fu D, Ren J, Hui J, Peng J, Liu Y, Qiu T, Jiang M, Pan Q et al (2017) Histone methyltransferase SETD2 modulates alternative splicing to inhibit intestinal tumorigenesis. J Clin Invest 127:3375–3391. https://doi.org/10.1172/JCI94292

Zarnett OJ, Sahgal A, Gosio J, Perry J, Berger MS, Chang S, Das S (2015) Treatment of elderly patients

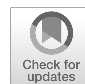
with glioblastoma: a systematic evidence-based analysis. JAMA Neurol 72:589–596. https://doi.org/10.1001/jamaneurol.2014.3739

Zeng Y, Ren R, Kaur G, Hardikar S, Ying Z, Babcock L, Gupta E, Zhang X, Chen T, Cheng X (2020) The inactive Dnmt3b3 isoform preferentially enhances Dnmt3b-mediated DNA methylation. Genes Dev 34: 1546–1558. https://doi.org/10.1101/gad.341925.120

Zhao SG, Chen WS, Li H, Foye A, Zhang M, Sjostrom M, Aggarwal R, Playdle D, Liao A, Alumkal JJ et al (2020) The DNA methylation landscape of advanced prostate cancer. Nat Genet 52:778–789. https://doi.org/10.1038/s41588-020-0648-8

Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, Shen H, Laird PW, Berman BP (2018) DNA methylation loss in late-replicating domains is linked to mitotic cell division. Nat Genet 50:591–602. https://doi.org/10.1038/s41588-018-0073-4

Zhou W, Liang G, Molloy PL, Jones PA (2020) DNA methylation enables transposable element-driven genome expansion. Proc Natl Acad Sci U.S.A. 117: 19359–19366. https://doi.org/10.1073/pnas.1921719117



DNA Methyltransferases and DNA Damage

14

Peter Sarkies

Abstract

Ever since the discovery of depletion of CG sites in mammalian genomes it has been clear that cytosine DNA methyltransferases (DNMTs) are linked to the rate at which mutations accumulate in DNA. Research in the intervening decades has shown that DNMTs influence mutation rates through the indirect consequences of methylation on the mechanism of mutation and the mechanisms for DNA repair. Additionally, recent studies have shown that DNA methyltransferases have the potential to directly introduce damage into DNA. Here, I will discuss both aspects of the connection between DNMTs and DNA damage, evaluating the potential consequences for evolution across species and in human diseases such as cancer where cellular evolution plays a key role.

Keywords

Epigenetics · DNA methylation · DNA damage · DNA repair · Evolution · Mutation · Cancer

P. Sarkies (✉)
Department of Biochemistry, University of Oxford, London, UK
e-mail: Peter.sarkies@bioch.ox.ac.uk

14.1 Brief Summary of DNA Methyltransferases

The mechanisms of eukaryotic DNA methyltransferases have been reviewed extensively, including elsewhere in this volume. To summarise here, DNMTs are an ancient family of enzymes, whose origins predate those of eukaryotes (Jurkowski and Jeltsch 2011; Ponger and Li 2005). The superfamily of DNA methyltransferases can be divided into those with specificity for position 6 on adenine residues and those with specificity for cytosine, either at position 5 or 4 (Cheng 1995). In addition to methylation on cytosine 5, methylation of cytosine 4 is common in bacteria (Seong et al. 2021) and has recently been documented in a basal plant, although this is likely to have evolved uniquely in this lineage through horizontal gene transfer (Walker et al. 2021). Methylation of adenine at the sixth position is also widespread in prokaryotes (Seong et al. 2021) and is present in some eukaryotes (Beh et al. 2019; Greer et al. 2015; Mondo et al. 2017), although the extent to which this occurs is currently under debate (Bochtler and Fernandes 2021; O'Brown et al. 2019).

Within eukaryotes, the most phylogenetically widely distributed cytosine methyltransferases are homologous to the mammalian enzymes DNMT1 and DNMT3 (Feng et al. 2010; Ponger and Li 2005; Zemach et al. 2010). Additionally, the DNMT5 enzyme, although not found in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_14

animals, can be inferred to have been present in the last eukaryotic common ancestor (de Mendoza et al. 2020; Huff and Zilberman 2014; Ponger and Li 2005). These enzymes predominantly act to methylate cytosine at the fifth position (Jeltsch 2006). They preferentially act on cytosines within CG dinucleotides, known as CpG dinucleotides to avoid confusion with general CG content (Jeltsch 2006), though there is evidence that DNMT3 enzymes in vertebrates also methylate cytosine in non-CpG contexts, particularly in embryonic and neuronal tissue (de Mendoza et al. 2021; He and Ecker 2015).

Classically, DNMT1 possesses "maintenance" methyltransferase function, whereby it acts on hemimethylated DNA after the passage of the replication fork to maintain DNA methylation in the CG context (Holliday 2006). DNMT3 enzymes are known as "de novo" methyltransferases as they have the ability to act on unmethylated DNA to introduce methylation (Edwards et al. 2017). These distinctions are not absolute, with DNMT1 having some de novo activity and DNMT3 contributing to faithful maintenance of methylation patterns through DNA replication. Indeed, the ability of DNMT1 to act on unmethylated CpG sequences is likely to explain the fact that many animals have only homologues of DNMT1 despite possessing robust levels of DNA methylation genome-wide (Bewick et al. 2017; Lewis et al. 2020). Although characterisation of DNMT5 is less advanced, in Cryptococcus neoformans it has been shown to possess robust CpG maintenance activity and little propensity for de novo methylation (Catania et al. 2020).

In many lineages, multiple paralogues of particular DNMTs have evolved. For example, mammals contain two DNMT3 enzymes, DNMT3A and DNMT3B, as well as a third protein, DNMT3L, which is catalytically inactive but has an essential role in regulating DNMT3A and B (Greenberg and Bourc'his 2019). Mouse and Rat also contain a third DNMT3 paralogue, DNMT3C, which is important in germline development (Barau et al. 2016).

Beyond CpG sequences, the sequence preferences of DNMTs vary widely across evolution. In mammals and plants, DNA methylation is found at both transposable elements and within transcribed genes (Lister et al. 2008, 2009). Additionally, in mammals, promoter methylation is important for controlling gene expression, with methylation of promoters with high CG content being associated with stable silencing (Edwards et al. 2017). Whilst gene body methylation seems to be conserved across animals, transposable element (TE) and promoter methylation seem to be much more labile, indicating that these can evolve independently in different lineages (Bewick et al. 2017; de Mendoza et al. 2019, 2020; Feng et al. 2010; Lewis et al. 2020; Rošić et al. 2018; Zemach et al. 2010). As discussed below, the propensity of DNA methylation to damage DNA might be an important factor in the evolution of diverse DNA methylation patterns.

14.2 DNA Methylation and Regulation of DNA Repair

14.2.1 Adenine Methylation and DNA Repair in Bacteria

The best understood influence of DNA methylation on DNA repair is in the regulation of mismatch repair in bacteria. Mismatch repair is a mechanism whereby aberrant bases incorporated during DNA replication can be detected and removed. The first step involves recognition of mispairs through direct sensing of the distorted base by the MutS protein (Lamers et al. 2000). MutS recruits MutL to the lesion and together they migrate along the DNA until they reach MutH (Acharya et al. 2003). MutH is an endonuclease which specifically cuts the newly synthesised strand containing the mispair. A helicase, UvrD, displaces the newly synthesised strand (Dao and Modrich 1998), thus offering the opportunity for the replication machinery to attempt to replicate the lesion again (Pluciennik et al. 2009), which will be highly likely to correct the error. MutH is able to specifically cut the newly synthesised strand due to DNA methylation (Welsh et al. 1987). The Dam DNA

methyltransferase methylates adenine residues within GATC sequences (Adhikari and Curtis 2016). MutH binds specifically to a hemimethylated form at GATC sequences containing methylation in one DNA strand (Welsh et al. 1987). Since Dam is a highly processive enzyme but is not present at replication forks, hemimethylated GATC sequences will only be present near replication forks. Hence, the unmethylated strand will be the newly synthesised DNA (Cooper et al. 1993; Lahue et al. 1987).

The mismatch repair pathway in *E. coli* is critically dependent on the Dam-mediated discrimination of the parental and daughter DNA strands. Mutations that inactivate Dam lead to increased mutation rate (Marinus 2010). Conversely, mutations that lead to increased Dam activity also increase mutation rate, presumably because the time that DNA remains hemimethylated after replication is reduced, thus reducing the time available for repair (Herman and Modrich 1981; Pukkila et al. 1983).

Although the mechanism of methyl-guided mismatch repair is fairly well understood, some intriguing questions remain over its evolution. In order for the system to work effectively, GATC sequences must be relatively frequent through the genome. In *E. coli* they are around 240 bp apart on average (Putnam 2021). The extent to which this places selection pressure on bacteria with the methylation guided mismatch repair system to retain GATC sequences is unclear. A second question arises from the fact that many bacteria possess mismatch repair systems that do not involve MutH and therefore do not rely on DNA methylation to identify the newly synthesised strand (Putnam 2016). In these species it seems that MutL has endonuclease activity (Lenhart et al. 2016). Thus, the methyl-guided repair would appear to be a more recent evolutionary innovation that occurred in gamma-proteobacteria. Despite the elegance of the strand discrimination mechanism, therefore, it seems that DNA methylation is not essential for the DNA repair process when taking a broad view across the phylogenetic tree. The evolutionary factors that promote acquisition and retention of the methyl-guided mismatch repair system are still unclear (Putnam 2016).

The mismatch repair pathway is conserved in eukaryotes, including homologues of MutS and MutL. However, consistent with the idea that methyl-guided mismatch repair is evolutionarily more recent than mismatch repair itself, methyl-based mismatch repair does not seem to be present in eukaryotes (Putnam 2021). The mechanism of strand discrimination instead is connected to the intrinsic nature of DNA replication. In vitro, mismatch repair is more efficient on strands with a break in the phosphate backbone (known as a nick) (Pluciennik et al. 2010). This has been hypothesised to enable it to distinguish newly synthesised DNA on the lagging strand, which is synthesised discontinuously (Kunkel and Erie 2015). Consistent with this, mismatch repair appears more active on the lagging strand in vivo (Andrianova et al. 2017).

14.2.2 DNA Methylation and Recombination

DNA recombination is vital for the repair of double strand breaks (DSB) in DNA. DSBs can arise due to exposure to ionising radiation and certain carcinogenic agents known as radiomimetics (Mehta and Haber 2014). The most frequent source of spontaneous DSBs, however, is DNA replication. Replication forks that encounter lesions in the DNA, such as nicks, single strand gaps or chemically modified bases that cannot be accommodated by polymerases can "collapse", resulting in the release of the DNA from the replication fork and a DSB (Mehta and Haber 2014). Accurate repair of DSBs can be achieved by homologous recombination whereby the undamaged strand of the sister chromatid is used as a template for the repair of the lesion (Scully et al. 2019). Additionally, programmed DSBs arise in meiosis, triggering exchange of genetic material between homologous pairs of chromosomes at points along the sequence known as crossovers (Mehta and Haber 2014).

Clear evidence for a role for DNA methylation in meiotic recombination comes from plants. In

Arabidopsis, DNA methylation can take place at cytosine in either CpG, CHG or CHH contexts, where H stands for A or T (Law and Jacobsen 2010). DNA methylation in these three contexts is anticorrelated with crossover formation (Yelina et al. 2012). Most strikingly, the centromeres in Arabidopsis which carry dense methylation are strongly depleted for crossovers (Naish et al. 2021). Direct evidence for the ability of DNA methylation to suppress crossover formation was provided by artificial induction of methylation at a specific crossover hotspot using small (18–30 nt) RNAs with complementary sequences to the region. This resulted in strongly reduced crossover formation at that region (Yelina et al. 2015).

How does DNA methylation restrain crossover formation in Arabidopsis? The enzyme Spo-11 is responsible for inducing meiotic DSBs across eukaryotes (Cole et al. 2010). Spo-11 induces several DSBs on each chromosome but only one of these becomes a recombination region where a crossover between homologous chromosomes happens (Pazhayam et al. 2021). DNA methylation does not change the locations of the DSBs, changing instead the probability that any particular DSB becomes a crossover (Choi et al. 2018). Exactly how this happens is not clear. Mapping of crossover hotspots in Arabidopsis has implicated H3K4 trimethylation and nucleosome depletion as important factors promoting crossover formation (Choi et al. 2013). DNA methylation may block crossover formation by blocking H3K4me3 and increasing nucleosome density.

Evidence that DNA methylation may be important in controlling recombination in mammals comes from the observation that mouse mutants lacking DNMT3L, which are deficient in de novo methylation, show failure of meiosis during spermatogenesis (Bourc'his and Bestor 2004). This corresponds to a loss of methylation at transposable element sequences and increased TE expression, leading to the idea that loss of DNA methylation might prevent meiosis indirectly through TE-induced DNA damage (Bourc'his and Bestor 2004). However, no increase in TE copy number occurs. Instead, it is thought that failure of meiosis occurs because of non-homologous recombination between TEs with different sequences (Zamudio et al. 2015). Crucially, the transcriptional activation that occurs in the absence of DNA methylation results in the acquisition of histone modifications, such as H3K4me3, that promote recombination (Zamudio et al. 2015). Thus, the failure of meiosis might be due to a loss of the ability of DNA methylation to restrain recombination. If so, the ability of DNA methylation to restrain recombination through blocking H3K4me3 seems to be present in both mammals and plants. It is interesting that this correlates, in both cases, to DNA methylation of transposable elements, which is not conserved in most animals (de Mendoza et al. 2020). Recent evolution of TE methylation has occurred in a few animal lineages (de Mendoza et al. 2019; Lewis et al. 2020) and it would be interesting to investigate whether this also corresponds to the acquisition of DNA methylation-based repression of recombination in these species.

14.3 Direct Effects of DNA Methylation on DNA Damage

14.3.1 Effect of DNA Methylation of Mutation Rate via Cytosine Deamination

The first understanding that DNA methylation might influence the rate of mutation was the observation that mammalian genomes were strongly depleted of CpG sequences (Bird 1980). This depletion can be quantified as a metric comparing the proportion of CpG dinucleotides found to the expected proportion, given the separate proportions of C and G, which is written as the mathematical formula:

$$R = (f_{\mathrm{CpG}})/(f_C \times f_G)$$

where f_{CpG}, f_{C} and f_{G} are the respective proportions of CpG, C and G within the sequence.

The value is calculated over a fixed window, such as 1 kb, over the genome to enable local variation to be identified (Gardiner-Garden and

Frommer 1987). In mammals, this metric tends to show very low values of R, at around 0.1 on average. Crucially, however, sites that show higher R values, typically around 0.6, correspond to regions that are often unmethylated. These sequences are known as “CpG islands” (Illingworth and Bird 2009). Together this indicates that methylation tends to increase the probability that a cytosine will become mutated. The first studies to identify this proposed a plausible mechanism, whereby cytosine deaminates to uracil, methylated cytosine deaminates to Thymine. Uracil, which is not a normal component of DNA, was proposed to be more easily recognised and excised (Gardiner-Garden and Frommer 1987). This will result, over time, in a depletion of CpGs, except in regions that are less frequently methylated in the germline, where CpG content will remain high.

In mammals, because the majority of the genome is methylated, the CpG content as a whole is low. Even CpG islands have lower CpG contents than many invertebrate genomes, and most genes have low CpG content corresponding to high and consistent methylation levels (Suzuki and Bird 2008). However, in some species where methylation is more sparsely distributed across the genomes, a bimodal distribution of genic CpG content is found. High CpG content corresponds to genes with low levels of methylation, whereas genes with high methylation have low CpG content. Thus, methylation is “concentrated” at a subset of genes. This has even been used, with some accuracy, to infer methylated genes, characterised by low CpG content, without assessing methylation directly (Keller et al. 2016).

It is important to note that bimodal distribution of genic CpG content is not found in all species with low methylation levels. In some cases this is because methylation itself is uniformly distributed, and no subset of methylated genes is particularly highly methylated (Lewis et al. 2020). However, in many species with high methylation in a subset of genes, bimodality in CpG content within genes is not observed (Bewick et al. 2017; Lewis et al. 2020). Bimodal CpG content across genes in some insects is therefore likely to be influenced by selective forces that are driving overall CG content upwards across the genome, making the depletion of CpGs in methylated genes appear particularly marked. Analysis of polymorphic regions in worldwide populations has confirmed this effect in the honey bee *Apis mellifera* (Wallberg et al. 2015). Indeed, even in mammals, the maintenance of CpG islands is a combination of relatively low methylation and specific selective forces that drive high CG content within those regions (Cohen et al. 2011). Nevertheless, the clear patterns of CpG content across species provide strong evidence for the mutagenic properties of methylated cytosines over evolutionary time.

Comparison of DNA methylation levels over evolutionary time does not easily enable a quantitative assessment of the difference in actual mutation rates at methylated CpG compared to unmethylated cytosine, because selective forces are also active so that any apparent mutational pattern observed in comparison within or between species is shaped by a combination of selection and mutation. However, DNA mutation rates can be assessed over short timescales in human cancers, where the majority of mutations are neutral (Cannataro and Townsend 2018). Here, the most common form of mutation is C→T and the rate of this transition is ~5 times greater at methylated CGs than unmethylated (Alexandrov et al. 2013). In order for mutations to translate into evolutionary differences, they have to occur in germline cells or in the zygote. Recent work employing “trios” (parents and offspring) to track de novo mutation rates have confirmed that DNA methylation increases the rate of C→T mutations in de novo mutations in humans by a factor of ~5 (Francioli et al. 2015; Rahbari et al. 2016). Together, these studies provide a solid body of evidence for the promotion of mutations by cytosine-5-methylation.

What is the mechanism for the promotion of C→T mutation by methylation of cytosine? The key to this is the well-known mechanism of cytosine deamination, which occurs spontaneously in solution and converts cytosine to uracil. This is the most frequent spontaneous base mutation that occurs in DNA (Lindahl 1996). The initial rate of

deamination is around 2 times faster for 5mC than for unmethylated cytosine (Shen et al. 1994). However, this is probably not sufficient to explain the 5 times greater rate of mutation. The excess mutagenicity is likely to arise from downstream processing of the products of deamination (De Bont and van Larebeke 2004).

Deamination converts cytosine to uracil, resulting in a U-G mismatch in the DNA. 5mC deamination results in a T-G mismatch instead. The canonical suggestion has been that the UG mismatch is easier to detect and repair because uracil is "foreign" and so more easily recognised and excised (Bird 1980). Uracil in DNA is recognised by the enzyme UNG (uracil DNA glycosylase). UNG excises uracil from the DNA to leave an abasic site, which is then a substrate for the base excision repair (BER) pathway (Visnes et al. 2009). BER responds to an abasic site by cutting the phosphate backbone of the DNA and initiating new DNA synthesis. This can be subdivided into two mechanisms. In "short patch" BER an enzyme known as AP lyase cuts either side of the abasic site, and only this nucleotide is then replaced by DNA synthesis (Robertson et al. 2009). Contrastingly, long-patch BER initiates DNA synthesis at the abasic site but then extends several nucleotides beyond (Robertson et al. 2009). Factors that determine whether short or long patch is used are unclear. The overall effect is to replace the original uracil with a newly synthesised cytosine.

The argument concerning "foreign" uracil recognition is not quite so simple however. The T-G mismatch that results from 5mC deamination is a substrate for a family of specific glycosylases, called Thymine DNA glycosylases or TDG (Bellacosa and Drohat 2015). Mammals possess two such enzymes, TDG and MBD4 (Sjolund et al. 2013). TDG enzymes coevolve with DNA methyltransferases (Rošić et al. 2018), thus the majority of species that possess DNA methylation also have the capacity to repair T-G mismatches directly. Increased mutagenicity of 5mC therefore predicts that UNG should recognise uracil in DNA with greater efficiency than TDG recognises T-G mismatches. Biochemical evidence, alongside structural modelling, does support this hypothesis, suggesting that TDG has relatively inefficient activity (Maiti et al. 2012; Tarantino et al. 2018). Nevertheless, this prediction remains to be tested in a cellular context.

One alternative possible mechanism for the mutagenicity of 5mC has recently been proposed on the basis of data from cancer genomes (Tomkova et al. 2018). In this model, 5mC increases the probability that the replicative polymerase pol epsilon inserts an erroneous residue (A rather than G) opposite 5mC, thus increasing the rate of C→T mutations directly (Fig. 14.1b) (Tomkova et al. 2018). Together with the difference in deamination rate of 5mC this may be sufficient to explain the increased rate of 5mC mutation in cancer (Tomkova and Schuster-Böckler 2018). In order to test whether this mechanism is responsible for the mutagenic effect of 5mC over evolutionary time it will be important to investigate whether it also occurs in germline development.

The mutagenic properties of methylated cytosine are related to the phenomenon of Repeat Induced Point mutation (RIP) (Gladyshev 2017). RIP occurs in Ascomycete fungi and is characterised by an extremely high rate of C→T mutations in transposable element sequences (Amselem et al. 2015; van Wyk et al. 2021). A cytosine DNA methyltransferase RID1 is essential for RIP, leading to the proposal that methylation of cytosine is required for the C→T mutation to occur (Aramayo and Selker 2013). However, this proposal is still under debate. The catalytic domain of RID1 is highly divergent, so an alternative proposal is that RID1 catalyses deamination directly (Freitag et al. 2002; Gladyshev 2017). However, recent work showed that an alternative DNA methyltransferase, DIM2, which is much more similar to mammalian DNMT1, can promote RIP as well (Gladyshev and Kleckner 2017). Moreover, a highly related process known as MIP (meiotic induced point mutation) has been shown to use a methyltransferase MASC1 that is a fully competent cytosine methyltransferase in vitro. The rate of mutations in RIP and MIP is far too high for spontaneous deamination to be responsible for

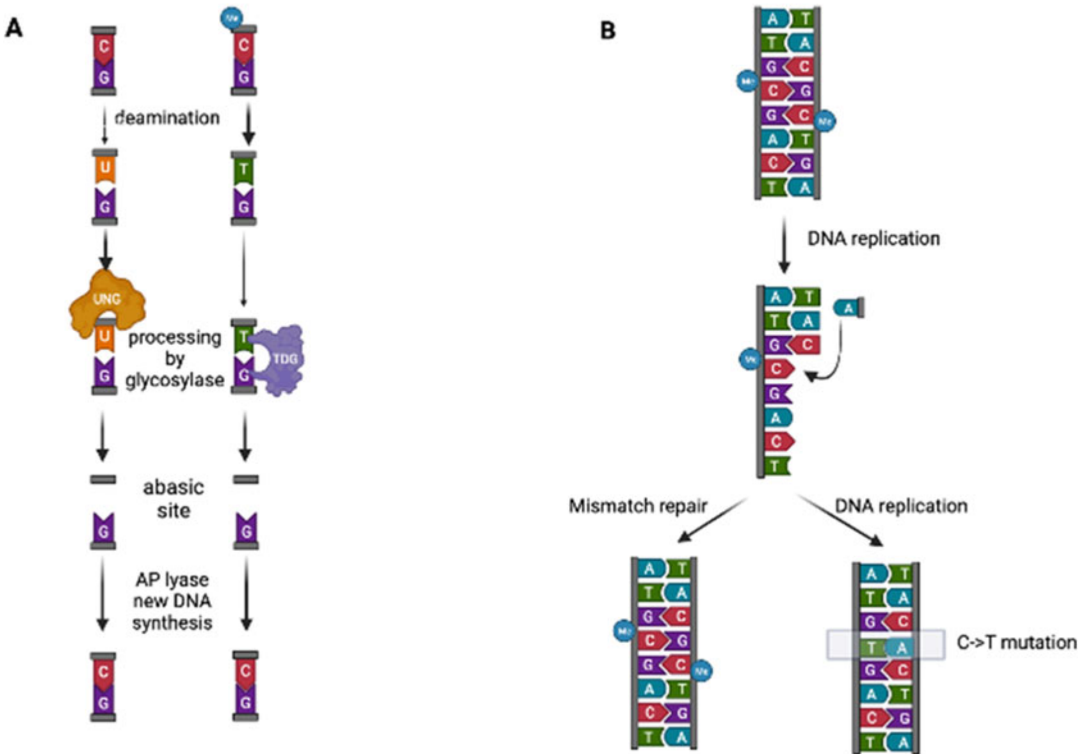


Fig. 14.1 Models for how cytosine DNA methylation promotes mutagenesis. (**a**) Cytosine deamination occurs more readily at methylated cytosine, and recognition of the resulting T-G mismatch is slower than the U-G mismatch resulting from deamination of unmodified cytosine. Overall, this is likely to increase the chance of a mutation occurring at methylated cytosine. (**b**) DNA replication is more likely to incorporate adenine erroneously opposite methylated cytosine than unmethylated cytosine. If this is not corrected by mismatch repair, a C→T mutation may result. *UNG* uracil deglycosylase, *TDG* thymine deglycosylase. Figure made with BioRender©

subsequent mutation (Gladyshev 2017), thus a deaminase activity, still unknown, is presumed to be required downstream of the cytosine methylation (Freitag et al. 2002).

14.3.2 DNA Alkylation Damage Induction

Recently, another way in which DNA methylation is linked to DNA damage was discovered. Key to the identification of this mechanism was an evolutionary study that showed coevolution between DNA methyltransferases and a specific DNA repair enzyme called ALKB2 (Rošić et al. 2018). ALKB2 tends to be present in species with DNA methylation and absent in those without (Rošić et al. 2018). Moreover, in arthropods, the presence of ALKB2 is associated with quantitatively higher DNA methylation levels (Lewis et al. 2020). ALKB2 specifically repairs 3mC and 1mA in DNA (Ougland et al. 2015), leading to the hypothesis that DNMTs might introduce these lesions as an “off-target” effect. Consistent with this, 3mC is associated with DNMT activity in vivo and in vitro (Dukatz et al. 2019; Rošić et al. 2018). Modelling of the catalytic site of DNMT3A suggested that 3mC induction might be an intrinsic property of DNMTs due to occasional rotation of the target base during the base flipping process (Dukatz et al. 2019).

3mC is a highly toxic lesion, which interferes with base pairing and causes replication forks to stall (Sedgwick 2004). Consequently, the

induction of 3mC by DNMTs may be a fitness cost associated with DNMT activity which might be important in driving the frequent loss of DNMTs in different eukaryotic lineages. Loss of ALKB2 would make this particularly likely, but a low-level burden of 3mC even in the presence of ALKB2 might still be sufficiently detrimental to promote loss of DNMTs in some lineages. 3mC induction would also be predicted to lead to mutations (Sedgwick 2004) but how this influences the mutagenic properties of DNMTs in cancer or across species is still unclear.

14.3.3 DNA Demethylation and DNA Damage

DNA methylation marks are sometimes removed from the genome. This process is notable in the genome-wide DNA demethylation processes that happen at key stages during mammalian development as part of a general reprogramming of epigenetic states (Nashun et al. 2015). Similar whole-scale demethylation also occurs during early development in plants (Zhang et al. 2018). Demethylation also happens on a finer scale at specific loci to ensure cell type specific gene expression programs are enacted (Luo et al. 2018). DNA demethylation can be classified as either passive, whereby maintenance methylation is reduced leading to dilution of DNA methylation through multiple rounds of cell division, or active, whereby enzymes specifically remove methylation from the genome on a rapid timescale (Edwards et al. 2017; Luo et al. 2018). Active DNA demethylation has been linked to the induction and repair of DNA damage, thus indirectly linking the sites of activity of DNA methyltransferases to the sites of DNA damage.

In plants, active DNA demethylation pathways have evolved by co-opting the base excision repair pathway discussed above (Roldán-Arjona et al. 2019). In Arabidopsis four enzymes, DME, DME2, DME3 and ROS1, are members of a specialised subfamily of glycosylases which act on methylated cytosine, excising the base to leave an abasic site (Choi et al. 2002; Gong et al. 2002; Ortega-Galisteo et al. 2008). The abasic site then triggers the base excision repair pathway, which results in the incorporation of an unmethylated cytosine (Roldán-Arjona et al. 2019). This mechanism occurs in all sequence contexts in which methylation is found in Arabidopsis (CAA, CHH and CG) (Penterman et al. 2007).

This remarkable co-option of the DNA repair machinery to enact DNA demethylation poses some important questions. First, the introduction of an abasic site appears highly dangerous as, if the continuation of the BER pathway does not correct it, it would lead to a mutation or even formation of a toxic double strand break. Second, in the case of CG sequences, the methylation must be removed from both strands, raising the question of how this is coordinated with the requirement for base excision repair, which would not be able to act on both strands simultaneously (Robertson et al. 2009). Interestingly, despite these apparent risks, demethylation is not restricted to developmental transitions as it appears that ROS1 has a constitutive role in counteracting DNA methylation at several regions across the genome (Tang et al. 2016). Recently, analysis of mutations in plant genomes suggested that there was an increased propensity for mutations to occur in the region surrounding methylated cytosines, which was not observed in mammals (Kusmartsev et al. 2020). DNA damage resulting from demethylation activity might be responsible for this by inducing long-patch base excision repair or through collision between the replication fork and an unrepaired abasic site (Kusmartsev et al. 2020).

DNA glycosylases that act on methylated cytosine seem to have evolved specifically in plants and are not present in animal lineages. Active demethylation in animals can be performed by the Ten-Eleven-Translocase (TET) family of enzymes, which are iron(II) and 2-oxoglutarate dependent oxidases (Luo et al. 2018; Nashun et al. 2015). Nevertheless, base excision repair might be involved in active demethylation downstream of TET proteins. TETs act on cytosine producing 5-hydroxymethylcytosine, followed by further TET mediated oxidation to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). Both

5fmC and 5caC have been shown to be potential substrates for base excision repair by triggering recognition by TDG (Pidugu et al. 2016, 2019). This process may not be required for demethylation because neither 5hmC, 5fC nor 5caC can be propagated through DNA replication. Hence, dilution through cell division would be sufficient to remove these modifications (Luo et al. 2018).

Despite the ability of TET proteins to mediate demethylation, they do not seem to be responsible for all stages of active DNA demethylation in mammalian development (Amouroux et al. 2016). The mechanisms underlying demethylation in both germline and zygotic development are still unclear (Luo et al. 2018). Interestingly however TDG is essential for early development (Cortellino et al. 2011) suggesting that there may yet be an essential role for DNA damage and repair in demethylation to be elucidated.

14.4 Conclusion

As a regulatory mechanism DNA methylation has a role in protecting the genome against DNA damage, both in bacteria, where the mismatch repair pathway relies on methylation cues for correct repair, and in eukaryotes where DNA methylation suppresses ectopic recombination as well as reducing crossover formation during meiosis. However, DNA methylation itself has the propensity to induce DNA damage, through increasing the rate of cytosine deamination, perturbing its repair and through the propensity of DNMT activity to introduce toxic alkylation damage in the form of 3mC into DNA. In addition, the process of DNA demethylation has been linked to DNA damage formation. Therefore, DNA methylation carries a risk, which may explain why it has so frequently been lost across evolution. Understanding the selective forces that promote retention of DNA methylation despite this clear disadvantage is a fascinating challenge for the future research to address.

Acknowledgements Thanks to members of my laboratory for their helpful discussions on this topic. Work in the Sarkies laboratory is funded by the Medical Research Council, The John Fell Fund, the EPA Fund and World Wide Cancer Research.

References

Acharya S, Foster PL, Brooks P, Fishel R (2003) The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair. Mol Cell 12(1):233–246. https://doi.org/10.1016/S1097-2765(03)00219-3

Adhikari S, Curtis PD (2016) DNA methyltransferases and epigenetic regulation in bacteria. FEMS Microbiol Rev 40:575–591. https://doi.org/10.1093/femsre/fuw023

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV et al (2013) Signatures of mutational processes in human cancer. Nature 500(7463):415–421. https://doi.org/10.1038/nature12477

Amouroux R, Nashun B, Shirane K, Nakagawa S, Hill PWS, D'Souza Z et al (2016) De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nat Cell Biol 18(2):225–233. https://doi.org/10.1038/ncb3296

Amselem J, Lebrun MH, Quesneville H (2015) Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. BMC Genomics 16(1). https://doi.org/10.1186/s12864-015-1347-1

Andrianova MA, Bazykin GA, Nikolaev SI, Seplyarskiy VB (2017) Human mismatch repair system balances mutation rates between strands by removing more mismatches from the lagging strand. Genome Res. https://doi.org/10.1101/gr.219915.116

Aramayo R, Selker EU (2013) Neurospora crassa, a model system for epigenetics research. Cold Spring Harb Perspect Biol 5(10). https://doi.org/10.1101/cshperspect.a017921

Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Hérault Y et al (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354(6314):909–912. https://doi.org/10.1126/science.aah5143

Beh LY, Debelouchina GT, Clay DM, Thompson RE, Lindblad KA, Hutton ER et al (2019) Identification of a DNA N6-adenine methyltransferase complex and its impact on chromatin organization. Cell 177(7):1781–1796.e25. https://doi.org/10.1016/j.cell.2019.04.028

Bellacosa A, Drohat AC (2015) Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites. DNA Repair 32:33–42. https://doi.org/10.1016/j.dnarep.2015.04.011

Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ (2017) Evolution of DNA methylation across insects. Mol Biol Evol 34(3):654–665. https://doi.org/10.1093/molbev/msw264

Bird AP (1980) DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res 8(7): 1499–1504. https://doi.org/10.1093/nar/8.7.1499

Bochtler M, Fernandes H (2021) DNA adenine methylation in eukaryotes: enzymatic mark or a form of DNA damage? Bioessays 43. https://doi.org/10.1002/bies.202000243

Bourc'his D, Bestor TH (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature 431:96–99. https://doi.org/10.1038/nature02886

Cannataro VL, Townsend JP (2018) Neutral theory and the somatic evolution of cancer. Mol Biol Evol 35(6): 1308–1315. https://doi.org/10.1093/molbev/msy079

Catania S, Dumesic PA, Pimentel H, Nasif A, Stoddard CI, Burke JE et al (2020) Evolutionary persistence of DNA methylation for millions of years after ancient loss of a de novo methyltransferase. Cell 180(2):263–277.e20. https://doi.org/10.1016/j.cell.2019.12.012

Cheng X (1995) Structure and function of DNA methyl transferases. Annu Rev Biophys Biomol Struct 24: 293–318. https://doi.org/10.1146/annurev.bb.24.060195.001453

Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB et al (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. Cell 110(1):33–42. https://doi.org/10.1016/S0092-8674(02)00807-3

Choi K, Zhao X, Kelly KA, Venn O, Higgins JD, Yelina NE et al (2013) Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. Nat Genet 45(11):1327–1338. https://doi.org/10.1038/ng.2766

Choi K, Zhao X, Tock AJ, Lambing C, Underwood CJ, Hardcastle TJ et al (2018) Nucleosomes and DNA methylation shape meiotic DSB frequency in Arabidopsis thaliana transposons and gene regulatory regions. Genome Res 28(4):532–546. https://doi.org/10.1101/gr.225599.117

Cohen NM, Kenigsberg E, Tanay A (2011) Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. Cell 145(5): 773–786. https://doi.org/10.1016/j.cell.2011.04.024

Cole F, Keeney S, Jasin M (2010) Evolutionary conservation of meiotic DSB proteins: more than just Spo11. Genes Dev 24:1201–1207. https://doi.org/10.1101/gad.1944710

Cooper DL, Lahue RS, Modrich P (1993) Methyl-directed mismatch repair is bidirectional. J Biol Chem 268(16): 11823–11829. https://doi.org/10.1016/s0021-9258(19)50274-5

Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A et al (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell 146(1):67–79. https://doi.org/10.1016/j.cell.2011.06.020

Dao V, Modrich P (1998) Mismatch-, MutS-, MutL-, and helicase II-dependent unwinding from the single-strand break of an incised heteroduplex. J Biol Chem 273(15):9202–9207. https://doi.org/10.1074/jbc.273.15.9202

De Bont R, van Larebeke N (2004) Endogenous DNA damage in humans: a review of quantitative data. Mutagenesis 19:169–185. https://doi.org/10.1093/mutage/geh025

de Mendoza A, Hatleberg WL, Pang K, Leininger S, Bogdanovic O, Pflueger J et al (2019) Convergent evolution of a vertebrate-like methylome in a marine sponge. Nat Ecol Evol 3(10):1464–1473. https://doi.org/10.1038/s41559-019-0983-2

de Mendoza A, Lister R, Bogdanovic O (2020) Evolution of DNA methylome diversity in eukaryotes. J Mol Biol 432:1687–1705. https://doi.org/10.1016/j.jmb.2019.11.003

de Mendoza A, Poppe D, Buckberry S, Pflueger J, Albertin CB, Daish T et al (2021) The emergence of the brain non-CpG methylation system in vertebrates. Nat Ecol Evol 5(3):369–378. https://doi.org/10.1038/s41559-020-01371-2

Dukatz M, Requena CE, Emperle M, Hajkova P, Sarkies P, Jeltsch A (2019) Mechanistic insights into cytosine-N3 methylation by DNA methyltransferase DNMT3A. J Mol Biol. https://doi.org/10.1016/j.jmb.2019.06.015

Edwards JR, Yarychkivska O, Boulard M, Bestor TH (2017) DNA methylation and DNA methyltransferases. Epigenetics Chromatin 10(1): 1–10. https://doi.org/10.1186/s13072-017-0130-8

Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M, Goll MG et al (2010) Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci USA 107(19):8689–8694. https://doi.org/10.1073/pnas.1002720107

Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, Renkens I et al (2015) Genome-wide patterns and properties of de novo mutations in humans. Nat Genet 47(7):822–826. https://doi.org/10.1038/ng.3292

Freitag M, Williams RL, Kothe GO, Selker EU (2002) A cytosine methyltransferase homologue is essential for repeat-induced point mutation in Neurospora crassa. Proc Natl Acad Sci U S A 99(13):8802–8807. https://doi.org/10.1073/pnas.132212899

Gardiner-Garden M, Frommer M (1987) CpG Islands in vertebrate genomes. J Mol Biol 196(2):261–282. https://doi.org/10.1016/0022-2836(87)90689-9

Gladyshev E (2017) Repeat-induced point mutation and other genome defense mechanisms in fungi. Microbiol Spectr 5(4). https://doi.org/10.1128/microbiolspec.funk-0042-2017

Gladyshev E, Kleckner N (2017) DNA sequence homology induces cytosine-to-thymine mutation by a heterochromatin-related pathway in Neurospora. Nat Genet 49(6):887–894. https://doi.org/10.1038/ng.3857

Gong Z, Morales-Ruiz T, Ariza RR, Roldán-Arjona T, David L, Zhu JK (2002) ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a

DNA glycosylase/lyase. Cell 111(6):803–814. https://doi.org/10.1016/S0092-8674(02)01133-9

Greenberg MVC, Bourc'his D (2019) The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol 20:590–607. https://doi.org/10.1038/s41580-019-0159-6

Greer EL, Blanco MA, Gu L, Sendinc E, Liu J, Aristizábal-Corrales D et al (2015) DNA methylation on N6-adenine in C. elegans. Cell 161(4):868–878. https://doi.org/10.1016/j.cell.2015.04.005

He Y, Ecker JR (2015) Non-CG methylation in the human genome. Annu Rev Genomics Hum Genet. https://doi.org/10.1146/annurev-genom-090413-025437

Herman GE, Modrich P (1981) Escherichia coli K-12 clones that overproduce dam methylase are hypermutable. J Bacteriol 145(1):644–646. https://doi.org/10.1128/jb.145.1.644-646.1981

Holliday R (2006) Epigenetics: a historical overview. Epigenetics 1:76–80. https://doi.org/10.4161/epi.1.2.2762

Huff JT, Zilberman D (2014) Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. Cell 156(6):1286–1297. https://doi.org/10.1016/j.cell.2014.01.029

Illingworth RS, Bird AP (2009) CpG islands—"A rough guide". FEBS Lett 583:1713–1720. https://doi.org/10.1016/j.febslet.2009.04.012

Jeltsch A (2006) Molecular enzymology of mammalian DNA methyltransferases. Curr Top Microbiol Immunol 301:203–225. https://doi.org/10.1007/3-540-31390-7_7

Jurkowski TP, Jeltsch A (2011) On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. PLoS One 6(11). https://doi.org/10.1371/journal.pone.0028104

Keller TE, Han P, Yi SV (2016) Evolutionary transition of promoter and gene body DNA methylation across invertebrate-vertebrate boundary. Mol Biol Evol 33(4):1019–1028. https://doi.org/10.1093/molbev/msv345

Kunkel TA, Erie DA (2015) Eukaryotic mismatch repair in relation to DNA replication. Annu Rev Genet 49:291–313. https://doi.org/10.1146/annurev-genet-112414-054722

Kusmartsev V, Drozdz M, Schuster-Böckler B, Warnecke T (2020) Cytosine methylation affects the mutability of neighboring nucleotides in germline and soma. Genetics 214(4):809–823. https://doi.org/10.1534/GENETICS.120.303028

Lahue RS, Su SS, Modrich P (1987) Requierement for d (GATC) sequences in Escherichia coli mutHLS mismatch correction. Proc Natl Acad Sci U S A 84(6):1482–1486. https://doi.org/10.1073/pnas.84.6.1482

Lamers M, Perrakis A, Enzlin J, Winterwerp H, de Wind N, Sixma TK (2000) Crystal structure of the E. coli DNA mismatch repair protein MutS in complex with a GT mismatch. Acta Crystallogr A Found Crystallogr 56(s1):s90. https://doi.org/10.1107/s0108767300022546

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11:204–220. https://doi.org/10.1038/nrg2719

Lenhart JS, Pillon MC, Guarné A, Biteen JS, Simmons LA (2016) Mismatch repair in Gram-positive bacteria. Res Microbiol 167:4–12. https://doi.org/10.1016/j.resmic.2015.08.006

Lewis SH, Ross L, Bain SA, Pahita E, Smith SA, Cordaux R et al (2020) Widespread conservation and lineage-specific diversification of genome-wide DNA methylation patterns across arthropods. PLoS Genet 16(6). https://doi.org/10.1371/journal.pgen.1008864

Lindahl T (1996) The Croonian Lecture, 1996: endogenous damage to DNA. Philos Trans R Soc Lond Ser B Biol Sci 351(1347):1529–1538. https://doi.org/10.1098/rstb.1996.0139

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in arabidopsis. Cell 133(3):523–536. https://doi.org/10.1016/j.cell.2008.03.029

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462(7271):315–322. https://doi.org/10.1038/nature08514

Luo C, Hajkova P, Ecker JR (2018) Dynamic DNA methylation: in the right place at the right time. Science 361:1336–1340. https://doi.org/10.1126/science.aat6806

Maiti A, Noon MS, MacKerell AD, Pozharski E, Drohat AC (2012) Lesion processing by a repair enzyme is severely curtailed by residues needed to prevent aberrant activity on undamaged DNA. Proc Natl Acad Sci U S A 109(21):8091–8096. https://doi.org/10.1073/pnas.1201010109

Marinus MG (2010) DNA methylation and mutator genes in Escherichia coli K-12. Mutat Res Rev Mutat Res 705(2):71–76. https://doi.org/10.1016/j.mrrev.2010.05.001

Mehta A, Haber JE (2014) Sources of DNA double-strand breaks and models of recombinational DNA repair. Cold Spring Harb Perspect Biol 6(9). https://doi.org/10.1101/cshperspect.a016428

Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, LaButti K et al (2017) Widespread adenine N6-methylation of active genes in fungi. Nat Genet 49(6):964–968. https://doi.org/10.1038/ng.3859

Naish M, Alonge M, Wlodzimierz P, Tock AJ, Abramson BW, Schmücker A et al (2021) The genetic and epigenetic landscape of the Arabidopsis centromeres. Science 374(6569). https://doi.org/10.1126/science.abi7489

Nashun B, Hill PWS, Hajkova P (2015) Reprogramming of cell fate: epigenetic memory and the erasure of memories past. EMBO J 34(10):1296–1308. https://doi.org/10.15252/embj.201490649

O'Brown ZK, Boulias K, Wang J, Wang SY, O'Brown NM, Hao Z et al (2019) Sources of artifact in

measurements of 6mA and 4mC abundance in eukaryotic genomic DNA. BMC Genomics 20(1). https://doi.org/10.1186/s12864-019-5754-6

Ortega-Galisteo AP, Morales-Ruiz T, Ariza RR, Roldán-Arjona T (2008) Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. Plant Mol Biol 67(6):671–681. https://doi.org/10.1007/s11103-008-9346-0

Ougland R, Rognes T, Klungland A, Larsen E (2015) Non-homologous functions of the AlkB homologs. J Mol Cell Biol 7:494–504. https://doi.org/10.1093/jmcb/mjv029

Pazhayam NM, Turcotte CA, Sekelsky J (2021) Meiotic crossover patterning. Front Cell Dev Biol 9. https://doi.org/10.3389/fcell.2021.681123

Penterman J, Zilberman D, Jin HH, Ballinger T, Henikoff S, Fischer RL (2007) DNA demethylation in the Arabidopsis genome. Proc Natl Acad Sci U S A 104(16):6752–6757. https://doi.org/10.1073/pnas.0701861104

Pidugu LS, Flowers JW, Coey CT, Pozharski E, Greenberg MM, Drohat AC (2016) Structural basis for excision of 5-formylcytosine by thymine DNA glycosylase. Biochemistry 55(45):6205–6208. https://doi.org/10.1021/acs.biochem.6b00982

Pidugu LS, Dai Q, Malik SS, Pozharski E, Drohat AC (2019) Excision of 5-carboxylcytosine by thymine DNA glycosylase. J Am Chem Soc 141(47):18851–18861. https://doi.org/10.1021/jacs.9b10376

Pluciennik A, Burdett V, Lukianova O, O'Donnell M, Modrich P (2009) Involvement of the β clamp in methyl-directed mismatch repair in vitro. J Biol Chem 284(47):32782–32791. https://doi.org/10.1074/jbc.M109.054528

Pluciennik A, Dzantiev L, Iyer RR, Constantin N, Kadyrov FA, Modrich P (2010) PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. Proc Natl Acad Sci U S A. https://doi.org/10.1073/pnas.1010662107

Ponger L, Li W-H (2005) Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. Mol Biol Evol 22(4):1119–1128. https://doi.org/10.1093/molbev/msi098

Pukkila PJ, Peterson J, Herman G, Modrich P, Meselson M (1983) Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in Escherichia coli. Genetics 104(4):571–582. https://doi.org/10.1093/genetics/104.4.571

Putnam CD (2016) Evolution of the methyl directed mismatch repair system in Escherichia coli. DNA Repair 38:32–41. https://doi.org/10.1016/j.dnarep.2015.11.016

Putnam CD (2021) Strand discrimination in DNA mismatch repair. DNA Repair 105. https://doi.org/10.1016/j.dnarep.2021.103161

Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S et al (2016) Timing, rates and spectra of human germline mutation. Nat Genet 48(2):126–133. https://doi.org/10.1038/ng.3469

Robertson AB, Klungland A, Rognes T, Leiros I (2009) Base excision repair: the long and short of it. Cell Mol Life Sci 66:981–993. https://doi.org/10.1007/s00018-009-8736-z

Roldán-Arjona T, Ariza RR, Córdoba-Cañero D (2019) DNA base excision repair in plants: an unfolding story with familiar and novel characters. Front Plant Sci 10. https://doi.org/10.3389/fpls.2019.01055

Rošić S, Amouroux R, Requena CE, Gomes A, Emperle M, Beltran T et al (2018) Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. Nat Genet 50(3):452–459. https://doi.org/10.1038/s41588-018-0061-8

Scully R, Panday A, Elango R, Willis NA (2019) DNA double-strand break repair-pathway choice in somatic mammalian cells. Nat Rev Mol Cell Biol 20:698–714. https://doi.org/10.1038/s41580-019-0152-0

Sedgwick B (2004) Repairing DNA-methylation damage. Nat Rev Mol Cell Biol 5:148–157. https://doi.org/10.1038/nrm1312

Seong HJ, Han SW, Sul WJ (2021) Prokaryotic DNA methylation and its functional roles. J Microbiol 59:242–248. https://doi.org/10.1007/s12275-021-0674-y

Shen JC, Rideout WM, Jones PA (1994) The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. Nucleic Acids Res 22(6):972–976. https://doi.org/10.1093/nar/22.6.972

Sjolund AB, Senejani AG, Sweasy JB (2013) MBD4 and TDG: multifaceted DNA glycosylases with ever expanding biological roles. Mutat Res 743–744:12–25. https://doi.org/10.1016/j.mrfmmm.2012.11.001

Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9(6):465–476. https://doi.org/10.1038/nrg2341

Tang K, Lang Z, Zhang H, Zhu JK (2016) The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. Nat Plants 2. https://doi.org/10.1038/nplants.2016.169

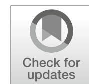
Tarantino ME, Dow BJ, Drohat AC, Delaney S (2018) Nucleosomes and the three glycosylases: high, medium, and low levels of excision by the uracil DNA glycosylase superfamily. DNA Repair 72:56–63. https://doi.org/10.1016/j.dnarep.2018.09.008

Tomkova M, Schuster-Böckler B (2018) DNA modifications: naturally more error prone? Trends Genet 34:627–638. https://doi.org/10.1016/j.tig.2018.04.005

Tomkova M, McClellan M, Kriaucionis S, Schuster-Böckler B (2018) DNA replication and associated repair pathways are involved in the mutagenesis of methylated cytosine. DNA Repair 62:1–7. https://doi.org/10.1016/j.dnarep.2017.11.005

van Wyk S, Wingfield BD, De Vos L, van der Merwe NA, Steenkamp ET (2021) Genome-wide analyses of repeat-induced point mutations in the Ascomycota.

Front Microbiol 11. https://doi.org/10.3389/fmicb.2020.622368
Visnes T, Doseth B, Pettersen HS, Hagen L, Sousa MML, Akbari M et al (2009) uracil in DNA and its processing by different DNA glycosylases. Philos Trans R Soc Lond B Biol Sci 364(1517):563–568. https://doi.org/10.1098/rstb.2008.0186
Walker J, Zhang J, Liu Y, Vickers M, Dolan L, Nakajima K, Feng X (2021) Extensive N4 cytosine methylation is essential for Marchantia sperm function. BioRxiv. Retrieved from https://www.biorxiv.org/content/10.1101/2021.02.12.428880v1
Wallberg A, Glémin S, Webster MT (2015) Extreme recombination frequencies shape genome variation and evolution in the honeybee, Apis mellifera. PLoS Genet 11(4). https://doi.org/10.1371/journal.pgen.1005189
Welsh KM, Lu AL, Clark S, Modrich P (1987) Isolation and characterization of the Escherichia coli mutH gene product. J Biol Chem 262(32):15624–15629. https://doi.org/10.1016/s0021-9258(18)47772-1
Yelina NE, Choi K, Chelysheva L, Macaulay M, de Snoo B, Wijnker E et al (2012) Epigenetic remodeling of meiotic crossover frequency in Arabidopsis thaliana DNA methyltransferase mutants. PLoS Genet 8(8). https://doi.org/10.1371/journal.pgen.1002844
Yelina NE, Lambing C, Hardcastle TJ, Zhao X, Santos B, Henderson IR (2015) DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in Arabidopsis. Genes Dev 29(20):2183–2202. https://doi.org/10.1101/gad.270876.115
Zamudio N, Barau J, Teissandier A, Walter M, Borsos M, Servant N, Bourc'his D (2015) DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. Genes Dev 29(12):1256–1270. https://doi.org/10.1101/gad.257840.114
Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328(5980):916–919. https://doi.org/10.1126/science.1186366
Zhang H, Lang Z, Zhu JK (2018) Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol 19:489–506. https://doi.org/10.1038/s41580-018-0016-z



Role of DNMTs in the Brain

15

Can Bora Yildiz and Geraldine Zimmer-Bensch

Abstract

DNA methyltransferases (DNMTs) are widely expressed in the brain, dictating the transcriptional activity of genes through various epigenetic mechanisms. Functional irregularities, alterations in the activity, and aberrant expression levels of DNMTs have been linked to various neurodevelopmental abnormalities, neuropsychiatric disorders, neurodegenerative diseases, and brain cancer. A continuously increasing number of studies address the roles DNMTs have in the brain, to reach a better understanding of their involvement in disease-related pathophysiologies, which in turn is required to dissect their applicability as potential therapeutic targets. This chapter provides an overview of DNMT function in the developing and the adult brain, putting a spotlight on their role in orchestrating diverse aspects of brain development, memory, and aging, followed by a discussion of associated neurodevelopmental and neurodegenerative disorders, and the implications in brain cancer.

C. B. Yildiz · G. Zimmer-Bensch (✉)
Functional Epigenetics in the Animal Model, Institute for Biology II, RWTH Aachen University, Aachen, Germany

Research Training Group 2416 MultiSenses-MultiScales, Institute for Biology II, RWTH Aachen University, Aachen, Germany
e-mail: can.yildiz@rwth-aachen.de; zimmer@bio2.rwth-aachen.de

Keywords

DNA methylation · Brain development · Synapse · Memory · Cell-intrinsic engram mechanisms · Aging · Neurodegeneration · Glioma · Cancer · lncRNAs

List of Abbreviations

5-AZA	5-azacytidine
5-hmC	5-hydroxymethylcytosine
5-mC	5-methylcystosinee
Aβ	Amyloid-beta protein
AD	Alzheimer's disease
aIPC	Apical intermediate progenitor cells
APP	Amyloid precursor protein
ARC	Activity regulated cytoskeletal-associated protein
BAX	Bcl-2-associated X protein
BCL2	B-cell lymphoma 2
BCL2L2	Bcl-2-like protein 2
BDNF	Brain-derived neurotrophic factor
bIPC	Basal intermediate progenitor cells
BIRC5	Baculoviral inhibitor of apoptosis repeat-containing 5; Survivin
bRGC	Basal radial glial cells
cAMP	Cyclic adenosine monophosphate
CASP8	Caspase-8
CDK5	Cyclin-dependent kinase 5
CDKN2A	Cyclin-dependent kinase inhibitor 2A

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_15

CGE	Caudal ganglionic eminence
CNS	Central nervous system
CpG	Cytosine-phosphate-guanine
CREB	cAMP response element-binding protein
CRISPR	Clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
CTIP2	B-cell lymphoma/leukemia 11B
dCGE	Dorsal caudal ganglionic eminence
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
E-LTP	Early long-term potentiation
EZH2	Enhancer of zeste homolog 2
GABA	Gamma aminobutyric acid
GBM	Glioblastoma
GFAP	Glial fibrillary acidic protein
GluN2A	*N*-methyl D-aspartate receptor subtype 2A
GSK3B	Glycogen synthase kinase 3 beta
H3K27me3	Trimethylation of lysine 27 on histone H3
H3K4me3	Trimethylation of lysine 4 on histone H3
H3K9me3	Trimethylation of lysine 9 on histone H3
HD	Huntington's disease
HIC1	Hypermethylated in cancer 1 protein
HTT	Huntingtin protein
ICF	Immunodeficiency, centromere region instability, facial anomalies syndrome
IDH	Isocitrate dehydrogenase
IPC	Intermediate progenitor cells
iPSC	Induced pluripotent stem cell
iSVZ	Inner subventricular zone
JAK	Janus kinase
L-LTP	Late long-term potentiation
lncRNA	Long non-coding RNA
LTD	Long-term depression
LTP	Long-term potentiation
MAPT	Microtubule-associated protein TAU
MBD	Methyl-binding domain
MeCP2	Methyl-CpG binding protein 2
MGE	Medial ganglionic eminence
MGMT	O^6-alkylguanine DNA alkyltransferase
mHTT	Mutant Huntingtin protein
miRNA	Micro RNA
mRNA	Messenger RNA
MZ	Marginal zone
NDD	Neurodegenerative diseases
NF2	Neurofibromin 2; Merlin
NFIA	Nuclear factor 1 A-type
NMDA	*N*-methyl-D-aspartic acid
NMDAR	*N*-methyl-D-aspartic acid receptor
NPY	Neuropeptide Y
oSVZ	Outer subventricular zone
p21	Cyclin-dependent kinase inhibitor 1
PAK6	p21 activated kinase 6
PDGFRA	Platelet-derived growth factor receptor A
PET	Positron emission tomography
piRNA	Piwi-interacting RNA
POa	Pre-optic area
PR C2	Polycomb repressive complex 2
PSEN	Presenilin
PTCH1	Protein patched homolog 1
PV	Parvalbumin
REST	RE1-silencing transcription factor
RG108	*N*-Phthalyl-L-tryptophan
RGC	Radial glial cell
RNA	Ribonucleic acid
SCZ	Schizophrenia
SFRP	Secreted frizzled-related protein
SHH	Sonic-Hedgehog
sncRNA	Small non-coding RNA
SST	Somatostatin
STAT	Signal transducer and activator of transcription
SVZ	Sub-ventricular zone
TBRS	Tatton-Brown-Rahman syndrome
TET	Ten-eleven translocation enzyme
TF	Transcription factor
TMS1	PYD and CARD domain containing, transcript variant 1
TMZ	Temozolomide
VIP	Vasointestinal peptide
VZ	Ventricular zone
Wnt	Wingless and Int-1
ZIF268	Zinc finger protein 268; Early growth response protein 1

15.1 Introduction

DNMTs are widely expressed in the developing, adult, and aged brain, suggesting implications in neuronal differentiation, maturation, and function (Guo et al. 2011; Simmons et al. 2013; Fasolino et al. 2017). Moreover, DNMT functionality and expression are altered in neurons of the aged brain, and in the context of neuropsychiatric and neurodegenerative diseases (Linde and Zimmer-Bensch 2020; Zimmer-Bensch 2020; Zimmer-Bensch and Zempel 2021), for which they are proposed as putative therapeutic targets. In addition to their canonical function of catalyzing DNA methylation, DNMTs can act on gene expression through crosstalk with histone modifications (Du et al. 2015; Symmank et al. 2018, 2020), hence displaying a diverse mechanistic spectrum. Moreover, the different DNMTs exhibit a brain region and cell-type-specific expression and seem to fulfill partly redundant (Feng et al. 2010) but also distinct functions in the brain (Morris and Monteggia 2014; Morris et al. 2016). While DNMT3A seems crucial for learning (Morris and Monteggia 2014), DNMT1 appears to be involved in anxiety (Morris et al. 2016), where the subcellular mechanisms remain unknown. Similar to the particular implications of the different DNMTs in orchestrating brain development and function (Zimmer-Bensch 2019b; Reichard and Zimmer-Bensch 2021), DNMTs contribute to certain diseases (Klein et al. 2011; Ding et al. 2018) and are themselves distinctively affected in brain cancer such as glioma (Rajendran et al. 2011). To exploit DNMTs therapeutically, we need to dissect their precise functional implications in the developing, aging, and diseased brain, which is discussed in this chapter with the focus on mammals.

15.2 The Mammalian Brain

The human brain is extraordinary in many regards, being recognized as the crown of evolution (Pascual-Leone et al. 2005; Hofman 2014). Still, we are far away from understanding in detail how this fascinating organ evolved, how the human brain works, and is established during ontogenesis. Despite the extraordinary features that account for the elaborated cognitive ability of the human brain, there are essential common principles in the architecture, the development and the function of the mammalian brain. For this, rodent and primate models are frequently used in neuroscientific research to extend our understanding of human brain evolution, function, development, and related diseases.

The mammalian brain is anatomically divided into three major parts: the hindbrain (including the cerebellum and the brain stem), the midbrain, and the forebrain (including the diencephalon and the cerebrum) (Fig. 15.1a). The brain stem, incorporating the pons and the medulla (Fig. 15.1b), is processing involuntary activities such as vomiting and breathing, while the cerebellum coordinates muscular movements and, in concert with the midbrain, it monitors posture.

The thalamus and the hypothalamus are major parts of the diencephalon (Fig. 15.1b). While the thalamus is a relay station for incoming sensory information routing these to the appropriate higher centers, the hypothalamus regulates heartbeat, body temperature, and fluid balance, in addition to appetite and body weight control (Sherman and Guillery 2006; Saper and Lowell 2014).

By far the largest region of the mammalian brain is the telencephalon (cerebrum) (Fig. 15.1a, b), composed of the superficial gray matter (cerebral cortex) and the white matter (axonal tracts). The telencephalon is distinguished vertically into left and right hemispheres. The two hemispheres communicate with each other through a large axonal tract, the corpus callosum. The cerebral cortex, the most evolved structure of the human brain holding its higher cognitive function, is strongly folded in gyri and sulci in humans and other primates (Hilgetag and Barbas 2005). In the cortex sensory data are processed, and motor impulses are generated that initiate, reinforce, or inhibit the entire spectrum of muscle and gland activity. The cerebral cortex is further involved in learning and memory and the control of affection (Thompson 1986; Jin

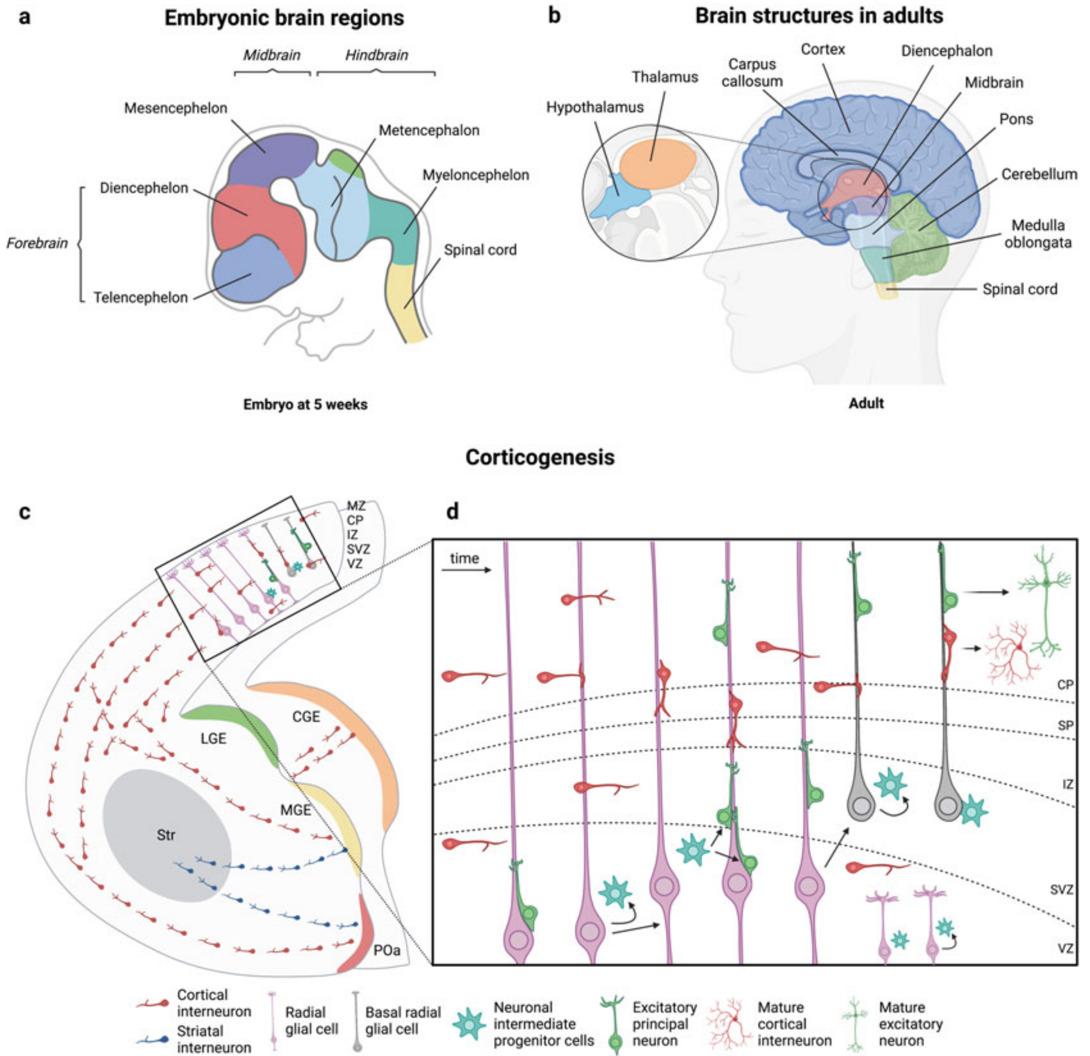


Fig. 15.1 Human brain anatomy and developmental principles of the mammalian brain. (**a**) Anatomical regions of the developing human brain. (**b**) Anatomy of the adult human brain. (**c**) Graphic depiction of a coronally sectioned hemisphere illustrating the sites of origin and the migratory streams of cortical inhibitory interneurons in the basal telencephalon (POA, MGE, and CGE), as well as of excitatory neurons in the dorsal telencephalon. (**d**) Schematic illustration of proliferation, differentiation, and migration of cortical precursor cells. Radial glial cells (RCGs) increase in number by symmetric division, and asymmetrically divide into basal radial glial cells (bRGCs) or neuronal intermediate progenitor cells (nIPCs). The latter further divide symmetrically to give rise to young excitatory principal neurons, which migrate into the cortical plate (CP). Inhibitory interneurons invade the developing neocortex along two migratory streams, the marginal zone (MZ), and the subplate (SP)/subventricular zone (SVZ), before they switch to radial migration to enter the cortical plate. *CGE* caudal ganglionic eminence, *CP* cortical plate, *IZ* intermediate zone, *LGE* lateral ganglionic eminence, *MGE* medial ganglionic eminence, *MZ* marginal zone, *POa* pre-optic area, *SP* subplate, *Str* striatum, *SVZ* subventricular zone, *VZ* ventricular zone

and Maren 2015). The cortex is composed of the six-layered neocortex, and the way smaller, three- or four layered allocortex. The allocortex consists of the paleocortex, the archicortex, and the periallocortex. The hippocampus and dentate gyrus are the main parts of the archicortex,

being functionally relevant for learning and memory.

Excitatory glutamatergic neurons and inhibitory local GABAergic interneurons represent the major neuronal subtypes of the cerebral cortex. While most glutamatergic cortical neurons display a long axon projecting either sub-cortically to other cortical areas or contralaterally to the other hemisphere, spiny stellate cells of layer IV are local excitatory interneurons receiving input from the thalamus (Shepherd 2004; Costa and Müller 2015). The glutamatergic neurons of the different cortical layers differ in their morphology, molecular features, and connectivity, establishing the neuronal circuits as basic modules of cortical information processing (Bayer and Altman 1991; Lodato et al. 2011; Greig et al. 2013). These circuits are shaped by the inhibitory action of the GABAergic cortical interneurons, which represent a highly diverse group of neurons differing in their electrophysiological features, morphology, targeting, and molecular properties (Nery et al. 2002; Fishell 2008; Gelman et al. 2009; Miyoshi et al. 2015; Wamsley and Fishell 2017; Lim et al. 2018; Zimmer-Bensch 2018) .

15.2.1 Developmental Principles of the Cerebral Cortex as the Seat of Higher Cognitive Functions

Prerequisite for correct cortical functionality is the proper establishment of the mammalian neocortex during embryonic and postnatal development. Processes such as progenitor proliferation and differentiation, cellular migration, morphological maturation, and the establishment of synaptic contacts, as well as programmed cell death, have to be highly controlled to form the circuits of billions of morphologically and functionally distinct neurons (Jones 2009; Huang and Paul 2018; Sultan and Shi 2018; Subramanian et al. 2020). Disturbances of these developmental processes cause a variety of neurodevelopmental disorders (Reichard and Zimmer-Bensch 2021).

The proportionally larger population of excitatory principal neurons (70–85% of the neuronal cells in the cortex) originates from progenitors of the dorsal telencephalon located in the zone delineating the lateral ventricles (ventricular zone; VZ) called radial glia cells (RGCs) (Fig. 15.1c) (Kriegstein and Alvarez-Buylla 2009; Sun and Hevner 2014). In addition to symmetric proliferative division to increase the pool of progenitors, RGCs' asymmetric division leads to the generation of neurons ("direct neurogenesis") (Kriegstein and Alvarez-Buylla 2009; Sun and Hevner 2014). Moreover, asymmetric division can result in the formation of intermediate progenitors, which translocate to the SVZ, where they generate neurons by symmetric divisions ("indirect neurogenesis") (Kriegstein and Alvarez-Buylla 2009; Agirman et al. 2017; Borrell 2019) (Fig. 15.1c). In humans, the developing cortex contains an inner (iSVZ) and an outer SVZ (oSVZ), hosting basal intermediate progenitor cells (bIPCs) and basal RGCs (bRGCs) (Sun and Hevner 2014). In contrast to humans and mammals with a high rate of gyrification such as ferrets, bRGCs are less prominent in lissencephalic species such as the mouse (Penisson et al. 2019; Subramanian et al. 2020). For this, bRGCs have been linked to cerebral gyrification. IPCs are also present in the VZ in humans and mice, being named apical IPCs (aIPCs) (García-Moreno et al. 2012; Sun and Hevner 2014).

Upon becoming post-mitotic, excitatory neurons migrate along the scaffold of radial glial cell processes spanning the whole cortical wall, into the cortical plate and settle in their target layer, establishing apical dendrites and axons (Zimmer-Bensch 2019a) (Fig. 15.1c). The inhibitory GABA-expressing interneurons originate in particular domains of the basal telencephalon (Lim et al. 2018; Mukhtar and Taylor 2018; Subramanian et al. 2020). The medial ganglionic eminence (MGE) generates parvalbumin (PV)-positive basket and chandelier cells, and somatostatin (SST)-expressing Martinotti and multipolar interneurons. The pre-optic area (POa) gives rise to neuropeptide Y (NPY)-, reelin-, SST-, and

CTIP2-expressing interneurons. Further reelin-expressing interneurons emerge in the caudal ganglionic eminence (CGE) alongside vasointestinal-peptide- (VIP)/calretinin-positive bipolar cells and VIP-/cholecystokinin-expressing basket cells (Gelman et al. 2011; Zimmer-Bensch 2019b). Thereby, the majority of interneurons is born in the MGE and the dorsal part of the CGE (dCGE) (Gelman et al. 2009; Marín et al. 2010; Faux et al. 2012; Lim et al. 2018; Sultan and Shi 2018; Zimmer-Bensch 2018). In humans and monkeys, some GABAergic interneurons appear to be generated in parts of the dorsal telencephalon at developmental later stages, proposing an evolutionary strategy of primate corticogenesis (Petanjek et al. 2009; Krienen et al. 2020).

Cortical inhibitory interneurons perform glial cell-independent long-range migration through the basal telencephalon toward the cortex, following defined routes depending on their site of origin, being guided by diverse sets of spatially and temporally expressed chemoattractive and repellent signaling molecules (Marín et al. 2003; Zimmer et al. 2007, 2010, 2011; Petanjek et al. 2009; Marín et al. 2010; Rudolph et al. 2010; Friocourt and Parnavelas 2011; Faux et al. 2012; Guo and Anton 2014; Symmank et al. 2019). Upon reaching the cortex, interneurons spread tangentially over the cortical areas along the SVZ/intermediate zone and the marginal zone (MZ) (Fig. 15.1c, d) (Tanaka and Nakajima 2012; Guo and Anton 2014), before they switch to radial migration invading the cortical layers that begin to be formed by the excitatory neurons at this embryonic stage (López-Bendito et al. 2004; Hatanaka et al. 2016).

Not all neurons being born successfully integrate into cortical circuits. It seems an evolutionarily conserved strategy to overproduce cortical neurons that are then being fine-tuned in their numbers by controlled cell death (Wong and Marín 2019). In matters of GABAergic interneurons about half of their embryonic population is reduced within there early postnatal period in mice (Yamaguchi and Miura 2015). At the same stage, improperly or disconnected pyramidal cells are also eliminated (Raff 1992). Apart from post-migratory regulation of the neuronal survival, regulatory mechanisms for survival regulation during neuronal migration have been described (Symmank et al. 2018).

15.3 DNMT Expression in the Brain

DNA methylation is catalyzed by the DNA methyltransferases DNMT1, DNMT3A, and DNMT3B in the mammalian brain. In line with the well-known function of DNMT1 in maintaining DNA methylation in dividing neural progenitor cells, and its reported functions in post-mitotic and mature neurons (Pensold et al. 2017, 2020), DNMT1 expression is remarkably high in the embryonic as well as adult nervous system (Goto et al. 1994; Inano et al. 2000; Fan et al. 2001; Kadriu et al. 2012). In the developing brain, DNMT1 is expressed in neuronal progenitors (Feng et al. 2007; Noguchi et al. 2015) and oligodendrocyte progenitor cells (OPCs) (Moyon et al. 2016), as well as in newly generated post-mitotic neurons. DNMT1 expression is maintained until adulthood, with prominent expression in GABAergic interneurons of the cerebral cortex (Kadriu et al. 2012; Pensold et al. 2020), as well as in excitatory cortical neurons (Hutnick et al. 2009; Feng et al. 2010), hippocampal neurons (Noguchi et al. 2015), and cerebellar neurons (Fan et al. 2001). Similar to DNMT1, DNMT3A can be detected in the developing, postnatal, and adult central nervous system (CNS) (Watanabe et al. 2002; Feng et al. 2005). DNMT3A was detected in progenitors of the cerebral cortex, in post-mitotic and adult cortical neurons, as well as in post-mitotic cerebellar cells. Similar findings were reported for the olfactory epithelia, which revealed expression of DNMT3A in maturing olfactory receptor neurons (MacDonald et al. 2005). In contrast to its neuronal expression, GFAP-positive astrocytes seem to only have a weak or no expression of DNMT3A. However, strong expression of DNMT3A was detected in postnatal cerebellar oligodendrocytes (Feng et al. 2005). Different from DNMT3A, DNMT3B expression is mainly restricted to neuronal precursor during early neurogenesis (Feng

et al. 2005). These stage- and cell-type-specific patterns of DNMT expression are suggestive of important roles in brain development, adult functionality, and associated diseases.

15.4 DNMT Function in the Developing Brain: Neurogenesis

Neuronal circuit formation depends on the correct generation of its neuronal constituents. Neurons derive from neuronal stem cells, which become progressively restricted to give first rise to the different neuronal subtypes (neurogenesis) and afterwards to glia cells (gliogenesis). Moreover, the sequential generation of the excitatory neurons destined for the distinct layers of the cerebral cortex, with deep layer neurons being born prior to the upper layer neurons, relies on progressive fate restriction (Martynoga et al. 2012). Apart from such temporal confinement, a spatial determination becomes evident early in embryonic development (Kiecker and Lumsden 2005). A prominent example is the distinct site of origin of inhibitory and excitatory neurons of the cerebral cortex in the ventral and dorsal telencephalon, respectively (Martynoga et al. 2012; Hu et al. 2017). Further, discrete spatial domains in the ventral telencephalon are suggested to give rise to different cortical interneuron subtypes (Hu et al. 2017).

Subtype-specific transcriptional programs orchestrate cell fate determination of both excitatory principal cortical neurons and inhibitory interneurons, directing subsequent developmental steps such as migration and morphological differentiation (Franco and Müller 2013; Hu et al. 2017). Thereby, a close connection between the stage- and subtype-specific transcriptional programs and the epigenetic machinery including DNMTs is proposed by an ever-increasing body of evidence. Indeed, as mentioned above, DNMTs are found widely expressed in neuronal precursors of the CNS (Feng et al. 2005). DNMT1 is suggested to be implicated in driving the differentiation into neurons by inhibiting the astroglial cell fate through DNA methylation of astroglia-associated genes during the neurogenic period. *Dnmt1* deficiency in progenitor cells of the spinal cord was reported to trigger precocious astroglial differentiation and hypomethylation of genes related to the gliogenic JAK/STAT pathway (Fan et al. 2005). Similarly, in the dentate gyrus, *Dnmt1* deficiency drives the differentiation of neuronal stem cells into astrocytes (Murao et al. 2016). The role of DNA methylation as an intrinsic driver of astrocyte differentiation in the embryonic brain has already been shown by Takizawa et al. (2001). The promoter sites of *Gfap* (glial fibrillary acidic protein) and *s100β* become demethylated at later stages of corticogenesis, promoting the generation of astrocytes from cortical progenitors. Demethylation of *Gfap* has been found to depend on the binding of NFIA (Nuclear Factor I/A), which is activated downstream of Notch and JAK/STAT signaling, leading to the dissociation of DNMT1 (Namihira et al. 2009), which then results in reduced methylation levels.

DNA demethylation further involves the oxidation of 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC) and subsequent oxidized forms by ten-eleven translocation (TET) methylcytosine dioxygenases, enabling cells to edit methylation patterns and thus maintain epigenomic flexibility during embryogenesis (Kohli and Zhang 2013). In line with this, TET1 function was reported to mediate the onset of neurogenesis by favoring the expression of neuronal genes (Kim et al. 2016).

In support of the functional implications of DNMTs and TETs in neurogenesis, dynamic temporal alterations of DNA methylation signatures have been detected during the sequential generation of neuronal subtypes (Lister et al. 2013; Lister and Mukamel 2015; Mo et al. 2015; Sharma et al. 2016). The use of epigenome editing approaches has already provided support for an instructive role of DNA methylation in neuronal differentiation driving subtype-specific developmental programs (Baumann et al. 2019).

15.5 DNMT Function in the Developing Brain: Post-mitotic Neuronal Maturation

Newly generated post-mitotic neurons usually migrate out of their proliferative zones to respective target regions, finally adopting subtype-specific morphological features mediated by axonal and dendritic growth as well as the formation of synapses, which underlie their specific connectivity and firing patterns. Intrinsically programmed cell death represents another crucial aspect of post-mitotic neuronal maturation, removing unconnected neurons and ultimately determining the final neuron numbers (Southwell et al. 2012). Subtype-specific establishment of DNA methylation signatures during neuronal and glial maturation has been reported by numerous studies (Lister et al. 2013; Lister and Mukamel 2015; Mo et al. 2015; Sharma et al. 2016), implying an important function of DNMTs in setting up the maturation-related DNA methylation patterns.

It was already shown that *Dnmt1* deletion in nestin expressing progenitor cells of the CNS is associated with increased rates of cell death in postnatal animals (Fan et al. 2001). In addition to this, morphological maturation was found to be impaired upon *Dnmt1* deletion in excitatory forebrain neurons (Hutnick et al. 2009), and after deletion of both *Dnmt1* and *Dnmt3A* (Feng et al. 2010). Together, these findings indicate that DNMTs regulate important aspects of postnatal neuronal development such as cell survival and morphological maturation.

Similar to morphological maturation, e.g., dendritic/axonal elaboration, neuronal migration critically relies on cytoskeleton remodeling. In addition to morphological refinement of the excitatory neurons of the cerebral cortex (Hutnick et al. 2009; Feng et al. 2010), DNMT1 function was found to regulate the migration of cortical inhibitory interneurons generated in the pre-optic area (POa) by acting on cytoskeletal organization, thereby promoting their polarized migratory morphology. Moreover, *Dnmt1* deficient interneurons showed increased rates of cell death. One of the involved target genes repressed by DNMT1 is *Pak6* (Pensold et al. 2017). PAK6 belongs to the p21-activated kinases known to drive neurite complexity in excitatory cortical neurons (Civiero et al. 2015) and is implicated in cell survival regulation (Kumar et al. 2017). Hence, the increased *Pak6* expression detected in *Dnmt1*-deficient POA-derived interneurons seems to account for their abnormal multipolar morphology and their impaired survival (Pensold et al. 2017). Another cell survival-associated gene, repressed by DNMT1 in migrating cortical interneurons, is *Lhx1* (Symmank et al. 2020). This homeobox transcription factor drives cell the expression of death associated genes and its downregulation promotes neuronal survival (Symmank et al. 2019; Symmank et al. 2020).

Of note, despite increased expression levels, DNA methylation signatures of the *Pak6* and the *Lhx1* gene locus were not changed in *Dnmt1*-deficient embryonic interneurons (Pensold et al. 2017; Symmank et al. 2018, 2020), implying DNMT1 to have activities beyond locus-specific DNA methylation, which may account for the transcriptional regulation of *Pak6* and *Lhx1*. Indeed, DNMT1 is known to affect histone modifications in neuronal and non-neuronal cells by transcriptional regulation of associated genes as well as through interactions with key enzymes at protein level (Du et al. 2015) (Fig. 15.2). Interactions between DNMTs and histone modifying enzymes have been reported to influence the catalytic activity of their binding partners and the recruitment to protein complexes (Viré et al. 2006; Smallwood et al. 2007; Clements et al. 2012). DNMT1 has been described to interact with EZH2 in non-neuronal cells (Viré et al. 2006; Ning et al. 2015; Purkait et al. 2016). EZH2 represents the core enzyme of the polycomb repressor complex 2 (PRC2) catalyzing repressive trimethylations on lysine 27 at the N-terminal amino acid tail of histone 3 (H3K27me3) (Margueron and Reinberg 2011). In addition to such putatively non-canonical functions via interactions with histone modifying proteins, DNMT1 has been reported to affect

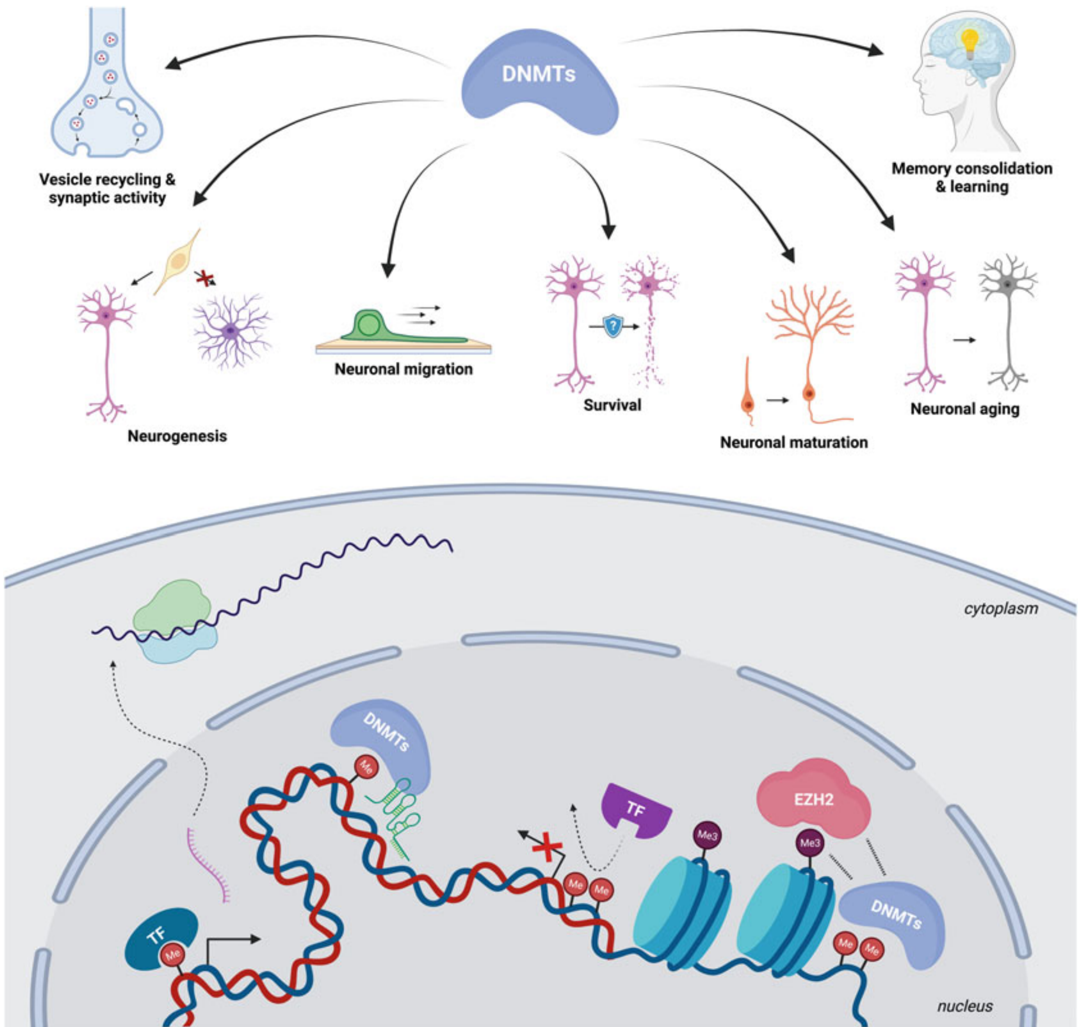


Fig. 15.2 The functional spectrum of DNMTs in the brain. Upper panel: DNMTs are involved in the regulation of neuronal development (neurogenesis, differentiation, neuronal migration, and maturation as well as survival), but also neuronal functionality by modulating synaptic function, learning and memory, in addition to neuronal aging. Lower panel: At molecular level, DNMTs act on gene expression through different mechanisms. The methylation of DNA segments via DNMTs can facilitate or hinder transcription factor (TF) binding and thereby modulate the expression of a gene. The recruitment of DNMTs may occur via long non-coding RNAs acting as adapters as well as through cross-talking with histone modifying complexes and histone marks. *DNMTs* DNA methyltransferases, *EZH2* enhancer of zeste homolog 2, *Me* methylation, *Me3* trimethylation, *TF* transcription factor

H3K27me3 levels by regulating the expression of *Ezh2* (So et al. 2011; Purkait et al. 2016).

In migrating cortical interneurons, the interaction of DNMT1 with EZH2 at protein level seems implicated in the establishment of H3K27me3 marks that represses the transcription of *Pak6* (Symmank et al. 2018), which is essentially involved in maintaining the migratory morphology and promoting the survival of migrating interneurons (Pensold et al. 2017). In support of this, inhibition of EZH2 causes similar defects in neuronal complexity as *Dnmt1* deletion, which

were rescued by *Pak6* depletion (Symmank et al. 2018). LHX1, another regulator of post-mitotic cortical interneuron development (Symmank et al. 2018), is likewise indirectly transcriptionally controlled by DNMT1 by interfering with histone acetylation and deacetylation through transcriptional control of genes coding for relevant enzymes (Symmank et al. 2020). Hence, DNMT1 regulates the migration and survival of post-mitotic cortical interneurons through distinct mechanisms (Fig. 15.2).

The post-mitotic development of other neuronal subtypes such as retinal ganglion cells, motor neurons, and dentate gyrus neurons has also been reported to involve DNMT1-mediated survival regulation (Chestnut et al. 2011; Rhee et al. 2012; Noguchi et al. 2016).

In sum, DNMTs regulate important aspects of post-mitotic neuronal development, such as migration, morphological maturation, neuronal survival, and cell death through canonical as well as non-canonical mechanisms (Fig. 15.2).

15.6 Role of DNMTs in Brain Function, Learning, and Memory

15.6.1 Functional Implications of DNMTs in Learning and Memory

Communication of nerve cells in neuronal circuits via their synaptic connections is considered the basis of brain functionality, learning, and memory. DNMTs and DNA methylation seem to be critically implicated in all these processes (Fig. 15.2). Downregulation of DNMT1 in excitatory neurons of the cerebral cortex differentiated from human induced pluripotent stem cells (iPSCs) reduced the proportion of active neurons as revealed by calcium imaging studies (Bachmann et al. 2021), indicating that DNMT1 function promotes neuronal activity of excitatory neurons. In contrast, in the inhibitory interneurons of the cerebral cortex, DNMT1-mediated DNA methylation was shown to reduce synaptic transmission by repressing endocytosis-related genes and endocytosis-dependent vesicle recycling (Pensold et al. 2020). Hence, by affecting the neuronal activity in excitatory versus inhibitory cortical neurons differently, DNMT1 could balance the net excitation of cortical networks. This process is critical for proper cortical functionality that is shown to be disturbed in diverse neurodevelopmental and neuropsychiatric diseases (Linde and Zimmer-Bensch 2020; Reichard and Zimmer-Bensch 2021).

It was already described that alterations in neuronal activity can induce global changes in the DNA methylation landscape (Guo et al. 2014). As DNA methylation modulates synapse- and plasticity-related gene expression that can mediate memory formation, DNMT function and DNA methylation could act on neuronal plasticity as well as metaplasticity, which is discussed in more detail below.

The concept of experience- and activity-dependent synaptic changes has long been accepted as the fundamental mechanism of learning and memory retention and is nowadays dubbed the synaptic plasticity and memory (SPM) hypothesis (Abraham et al. 2019). Several observations have made synaptic plasticity a leading candidate cellular mechanism for memory formation and storage. As numerous forms of learning have been shown to induce synaptic plasticity in learning-relevant brain regions, diverse forms of reversal learning were shown to trigger a reversal of synaptic plasticity, complementary to what has been induced by the initial learning paradigm (Abraham et al. 2019). The SPM hypothesis involves activity-dependent long-lasting changes in synaptic efficacy such as long-term potentiation (LTP) and long-term depression (LTD) (Bliss and Collingridge 1993; Roberts and Glanzman 2003). Recent studies with modern imaging methods capable of real-time monitoring of changes in synaptic spine morphology accompanied by well-established electrophysiological measures to detect functional synaptic changes have opened a new door for a better understanding of pre- and post-synaptic LTP and LTD (Abraham et al. 2019).

LTP is the most intensively investigated form of synaptic plasticity, captured by the Hebbian

phrase: "cells that fire together wire together" (Lowel and Singer 1992). It has been shown to depend on DNMTs and DNA methylation in addition to other chromatin modifications (Levenson et al. 2006; Miller and Sweatt 2007; Muñoz et al. 2016). Joint firing of the pre- and post-synaptic cell generates LTP, a strengthening of the synapses, while asynchronous firing generates the opposite, LTD. LTP and LTD are induced by different mechanisms, involving ionotropic and metabotropic receptor activation by neurotransmitters, such as the amino acid glutamate acting mainly as excitatory neurotransmitter in the mammalian nervous system. Hallmarks of LTP involve input specificity and associativity, which can be achieved by the activation of the *N*-methyl-D-aspartate (NMDA) subtype of glutamate receptor (Abraham et al. 2019). The NMDA receptor is an ion channel, which requires glutamate binding (specificity) and coincident depolarization achieved by multiple co-active synapses (association) for opening and channel unblocking, hence functioning as "coincidence detector" (Bliss and Collingridge 1993). These channels are also permeable for calcium ions, well-known second messenger and initiators of signaling cascades triggering LTP (Bear and Abraham 1996; Cummings et al. 1996).

LTP persistence is prerequisite for long-term memory. LTP can be categorized in early LTP (E-LTP) and late-LTP (L-LTP), or alternatively, in LTP1, LTP2, and LTP3 (Racine et al. 1983; Morrell 1991). While E-LTP or LTP1, which lasts only a few hours in maximum, occurs independently of de novo protein synthesis, L-LTP involves protein synthesis. L-LTP can be further subdivided into LTP that is transcription-independent (LTP2) or dependent (LTP3) (Racine et al. 1983). Transcription-independent (LTP2) can be achieved by the local protein translation machinery present in dendritic-synaptic compartments using existent local mRNA species. Transcriptional activation involved in LTP3 is mediated by transcription factors such as cAMP response element-binding protein (CREB), serum response factor, and nuclear factor kappa B, which in turn trigger the expression of downstream-induced transcription factors like ZIF268, c-FOS, JUNB. Manipulations of diverse epigenetic writers and erasers, including DNMTs and TET enzymes, have been shown to affect the different forms of LTP as well as the expression of key genes such as CREB (Kandel 2012; Rajasethupathy et al. 2012), and thus, memory formation.

Indeed, variable forms of learning and memory formation and/or consolidation involve DNMT function and DNA methylation (Day and Sweatt 2010). *Dnmt1* and *Dnmt3A* deletion in excitatory forebrain neurons has been shown to affect learning and memory in the hippocampus (Feng et al. 2010). Inhibition of DNA methyltransferases or genetic deletion of *Dnmt3A* potently hampers LTP (Levenson et al. 2006; Morris et al. 2014). *Dnmt3A* knockout mice display deficits in associative and episodic memory tasks and synaptic alterations, indicating that DNMT3A function in post-mitotic neurons is crucial for normal memory formation. Furthermore, associative learning tasks impact *Dnmt3A* expression, underlining the implication of DNMT3A in learning-related processes (Morris et al. 2014). Moreover, synaptic plasticity and fear memory consolidation in the lateral amygdala in addition to hippocampal structures seem to depend on DNA methylation and DNMT activity (Monsey et al. 2011). Of note, it was also found that E-LTP is dependent on DNA methylation, as the DNA methylation inhibitor 5-aza-2-deoxycytidine (5-AZA) impaired hippocampal long-term potentiation (LTP) induced by a twenty-minute theta burst stimulation. In contrast, 5-AZA treatment 2 hours after stimulation had no effect on transcription-dependent LTP in the applied experimental setup (Muñoz et al. 2016). This indicates that early alterations in DNA methylation are sufficient to impair LTP. The role of DNMTs in the induction of synaptic plasticity was already reported by Levenson et al. (2006). This study showed that inhibiting DNMT changed the DNA methylation signatures within promoters of *Reln* (reelin) and *Bdnf* (brain-derived neurotrophic factor), two factors essential for synaptic plasticity induction in the adult hippocampus (Levenson et al. 2006).

15.6.2 DNMTs as Potential Mediators of Cell-Intrinsic Mechanisms for Memory Consolidation and Maintenance

It is still under debate whether the stability of such synaptic changes is prerequisite for memory maintenance. Highly varying degrees of spine turnover were detected by high resolution imaging in the neocortex (Bhatt et al. 2009), while experimentally triggered LTP can be long-lasting (Abraham et al. 2002). Alternative views involve the idea that there may be multiple synaptic weight distributions being capable of properly coupling inputs with outputs. Learning new information would result in an update of synaptic weights to enable the incorporation of new information, while at the same time retaining the old ones. In artificial network models with highly coupled layers of cells, such changes of synaptic weights have been revealed to be necessary (Abraham and Robins 2005). Another model based on experimental data implies that learning establishes specific connectivity patterns between cells of a memory circuit, which is named engram. These new connections, rather than the potentiation of existing synapses, are suggested to support memory storage. According to this view, the LTP of existing synapses rather serves to recall the memory (Tonegawa et al. 2015). Still, both models somehow support the synaptic plasticity and memory hypothesis, with synapses representing critical units of memory storage. A strong argument being discussed against this general hypothesis is the observation that synaptic molecules in the adult brain are not stable with half-lives of only 2–5 days (Cohen et al. 2013), although it should be mentioned that individual molecules might not need to last for the duration of a memory (Lisman 1985). Moreover, the successful transfer of memory from a trained to an untrained animal via RNA injection tremendously challenged the synaptic plasticity hypothesis of memory storage (Bédécarrats et al. 2018). Of note, synaptic transmission is not the only way of communication between neurons. Non-synaptic flow of information between neuronal somata in the form of non-coding and protein coding RNA as well as proteins can be achieved through neuronal activity triggered release of miRNA-containing exosomes (Chivet et al. 2014; Goldie et al. 2014; Higa et al. 2014) or channeling nanotubes (Ariazi et al. 2017), and it has been described to mediate learning-related epigenetic alterations in neurons (Abraham et al. 2019). An important example is the activity regulated cytoskeletal-associated protein ARC, an immediate early gene product and a vital regulator of synaptic activity. Upon neural activity, ARC is released alongside *Arc* mRNA via exosomes that are taken up by neighboring neurons where the transferred mRNA is translated locally (Ashley et al. 2018; Pastuzyn et al. 2018).

An alternative model involves the cell-intrinsic storage of memory enabled by thermodynamically stable molecules, which is supported by numerous studies that are discussed elsewhere in more detail (Abraham et al. 2019). As firstly suggested by Holliday (1999), DNA methylation represents an attractive mechanism for cell-intrinsic engram storage, which brings DNMT function into play. Besides relative stability, DNA methylation comes with the advantages of compactness and energy efficiency. Apart from that, this epigenetic signature is capable of storing a vast amount of information, due to the extraordinary numbers of methylation sites in the whole genome (Holliday 1999). The finding of active DNA demethylation involving TET-mediated oxidation of 5-mC allows cytosines in neurons to function as on-off switches, hence providing principally a “binary code” (Abraham et al. 2019).

Since the hypothesis raised by Holliday in 1999, studies in mammals and invertebrates confirmed the functional implication of DNA methylation and DNMTs in diverse learning paradigms (Day and Sweatt 2010; Biergans et al. 2015; Pearce et al. 2017). It was shown that contextual fear conditioning triggers global genome-wide changes in DNA methylation seen after an hour persisting for at least 24 h (Mizuno et al. 2012), but not after 4 weeks (Halder et al.

2016). These changes in DNA methylation, mostly detected in neurons that correlated with the spatio-temporal location of memory, were specific for genes and *cis*-regulatory sites and were reported to be dynamic or stable (Halder et al. 2016; Duke et al. 2017). In line with these studies, inhibition of DNMTs has been shown to block remote memory in rats (Miller et al. 2010). Moreover, DNMT inhibition impairs memory formation and consolidation and even eliminates well-consolidated long-term memory in Aplysia (Pearce et al. 2017). DNMTs and DNA methylation further seem to be implicated in the RNA injection-mediated memory transfer described previously. Blocking DNMTs by RG108 immediately after RNA injection successfully impeded the behavioral enhancement and, hence, memory transfer (Bédécarrats et al. 2018). The learning-triggered changes in DNA methylation were reported to regulate the transcription or splicing of plasticity-, neuronal transmission- and function-related genes in different learning paradigms (Halder et al. 2016; Duke et al. 2017). Hence, DNMTs and DNA methylation might play a key role in long-term memory being stored either as RNA-induced epigenetic alterations and/or synaptic plasticity.

Both models, the synaptic plasticity and the cell-intrinsic model of memory formation, do not necessarily exclude each other and rather could be integrated, with epigenetic writers such as DNMTs as key mediators. A prerequisite for that would be a synapse-nucleus communication to provide an explanation for how nuclear changes can account for input (or synapse) specificity, shown to be evident in LTP, as well as other forms of learning-related long-term synaptic plasticity (Martin et al. 1997; Schuman 1997). Hence, a challenging question in this context is how the synaptic information is translated into discrete changes of DNA methylation and how the signaling mechanisms from the nucleus back to specific synapses occur?

A recent study showed that protein levels of DNMT3A1 are intimately linked to the activation of *N*-methyl-D-aspartate receptors (NMDAR) containing the GluN2A subunit. Synaptic NMDARs were found to promote the degradation of this DNMT in a neddylation-dependent process. Interference with neddylation leads to reduced degradation of DNMT3A1, which causes changes in promoter methylation of activity-driven genes and deficits in synaptic plasticity and memory formation. Hence, plasticity-relevant signals from GluN2A-containing NMDARs seem to orchestrate activity-dependent DNA methylation implication in memory formation (Bayraktar et al. 2020).

Moreover, non-coding RNAs provide possible alternative mechanisms to control DNA methylation. Similar to mRNAs, which undergo local translation at dendrites/synapses (Donnelly et al. 2010), also non-coding RNAs have been reported to shuttle from the neuron's nucleus to dendritic compartments (Qureshi and Mehler 2012). Shuttling back from the cytosol to the nucleus has already been shown for piRNAs, a class of small-non-coding RNAs, which in addition to cytosolic functions influence transcription in the nucleus, e.g., by recruiting DNMTs and through this targeting DNA methylation (Liu et al. 2019). Of note, piRNAs are known to be key for establishing stable long-term changes in neurons in memory persistence by mediating the methylation of a conserved CpG island in the *Creb2* promoter in a serotonin-dependent fashion (Rajasethupathy et al. 2012). CREB, a critical plasticity-related protein, acts as crucial inhibitory constraint of memory in Aplysia (Bartsch et al. 1995). Hence, cytosine methylation of the *Creb2* promoter triggered by the Piwi-piRNA complex containing a DNMT could provide a mechanistic link for how transient external stimuli culminate in long-lasting alterations in the expression of genes implicated in long-term memory storage in neurons.

DNMTs have been further reported to interact with long non-coding RNAs (lncRNAs), defined as non-coding transcripts longer than 200 nucleotides (Hung and Chang 2010), which were suggested to either recruit DNMTs to specific genomic loci or prevent their binding (Rinn and Chang 2012; Merry et al. 2015; Zhao et al. 2016; Somasundaram et al. 2018; Zimmer-Bensch 2019a), similar to piRNAs. The expression of lncRNAs has been described to be

modulated by altered neuronal activity (Barry et al. 2017), and they can be shuttled from the nucleus to the cytoplasm (Bridges et al. 2021). Thereby, their subcellular location determines their function (Carlevaro-Fita and Johnson 2019). LncRNAs can be precursors for miRNAs, representing a class of small non-coding RNAs which modulate translation in the cytoplasm (Leung 2015). Hence, lncRNAs could influence the translation of synapse-related gene expression, "transferring" information from the nucleus into the cytoplasm, and potentially in discrete microcompartments of a neuron "hitting" local translation. The ability of lncRNAs to localize to diverse yet specific subcellular locations has further been described (Bridges et al. 2021), where they regulate synapse stability (Wang et al. 2021), synaptic activity (Raveendra et al. 2018; Keihani et al. 2019), or structural plasticity of dendritic spines in an activity-dependent manner (Grinman et al. 2021). However, in contrast to piRNAs, whether and how shuttling of lncRNAs back from the cytoplasm to nucleus occurs, is still largely unknown.

15.7 DNMTs in Neurodevelopmental and Neuropsychiatric Diseases

After having discussed the crucial functions of DNMTs in neurodevelopment and brain function, it is not surprising that mutations or defective expression of DNMTs are implicated in a broad spectrum of neurodevelopmental and neuropsychiatric diseases, being the cause or mediator of the underlying pathophysiology. Biallelic missense mutations in *DNMT3B* are causative for the immunodeficiency-centromeric instability-facial anomalies (ICF) syndrome, a rare autosomal recessive disorder presenting with cognitive and intellectual disability (Miniou et al. 1997; Kondo et al. 2000; Jin et al. 2008). *DNMT3B* mutations are associated with DNA hypomethylation of genes relevant for the immune system, but also for neurogenesis, neuronal differentiation, and migration in the affected patients (Jin et al. 2008). In line with this, another study found that conditional *Dnmt3B* deletion in the hippocampus impairs recognition memory and revealed differential expression of K+ channel subunits in mice (Kong et al. 2020). Mutations in *DNMT3A* that lead to different variants of this methyltransferase (Lane et al. 2020) were reported for the Tatton-Brown-Rahman syndrome (TBRS), a rare neurodevelopmental congenital anomaly syndrome characterized by macrocephaly and characteristic facial features (Yokoi et al. 2020).

Patients suffering from schizophrenia (SCZ) show a significant upregulation of *DNMT1* expression in *postmortem* GABAergic interneurons (Veldic et al. 2005), which is suggested to alter the expression of genes relevant for GABAergic transmission (Linde and Zimmer-Bensch 2020). These observations are in line with the finding of disturbed interneuron functionality as crucial hallmark of schizophrenia (Nakazawa et al. 2012), and DNMT1 was found to modulate GABAergic transmission of cortical interneurons by regulating endocytosis-dependent vesicle replenishment through DNA methylation-dependent transcriptional control of associated genes (Pensold et al. 2020). *DNMT1* overexpression in SCZ patient brains is proposed to cause hypermethylation of *RELN*, coding for Reelin that is a key player in cortical development (Kirkbride et al. 2012). In addition to abnormal functionality in mature neurons, defects during interneuron development are suggested to contribute to the manifestation of neuropsychiatric diseases such as schizophrenia (Linde and Zimmer-Bensch 2020). In agreement with this, altering the expression levels of *Dnmts* in embryonic cortical interneurons in mice elicited SZ-like phenotypes in offspring (Matrisciano et al. 2013). In line with the reported relevance of DNMT1 for interneuron migration (Pensold et al. 2017), dysregulated expression levels during development could contribute to the manifestation of SCZ. Furthermore, increased activity of DNMTs as well as DNA hypermethylation has been suggested to be implicated in the development of epilepsy in humans as well as in rodent models (Jesus-Ribeiro et al. 2021).

15.8 DNMTs in Neuronal Aging

Brain aging is a dynamic process characterized by structural, neurochemical, and physiological alterations that altogether cause memory decline, cognitive impairments, and behavioral changes (Rozycka and Liguz-Lecznar 2017; Zimmer-Bensch 2019b). Cognitive aging predominantly manifests itself as attention and memory deficits involving the function of the hippocampus and frontal brain areas, such as the prefrontal cortex, with the working memory function being affected the most in aged individuals (Nolde et al. 1998; Davidson and Glisky 2002; Glisky 2007). Neuronal circuits formed within the hippocampus and the synaptic connections with other brain regions are widely considered to constitute the basis for its function in learning and memory, indicating that age-associated perturbances in these regions increase susceptibility to learning deficits later in life (Eichenbaum et al. 1992; Glisky 2007). Thereby, healthy adult neurogenesis in the hippocampus is proposed to be essential for higher cognitive functions (Bekinschtein et al. 2010).

In the mature CNS, neuronal plasticity and long-term memory are modulated by DNA methylation through DNMT activity in the hippocampus (Levenson et al. 2006). The neuronal methylome changes dramatically upon neuronal activity, in association with synaptic plasticity genes gaining or losing DNA methylation (Guo et al. 2011). Additionally, adult neurogenesis is defined as a pivotal process in the generation of neurons in adulthood, thus directly affecting learning and memory functions (Ming and Song 2011). It has been shown that hypomethylation in the brain during aging is responsible for a decline in adult neurogenesis (Liu et al. 2009), which is in line with the reported decline in the expression of DNMTs in the brain upon aging (Oliveira et al. 2012).

The aging process hits different brain regions and neuronal cell types distinctively. In addition to reduced excitability and plasticity (Clark and Taylor 2011), an increased vulnerability of inhibitory interneurons and GABAergic synapses (Rozycka and Liguz-Lecznar 2017) has been reported for particular regions of aged brains (Shetty and Turner 1998; Stanley and Shetty 2004; Cheng and Lin 2013). Besides functional and structural changes of GABAergic synapses, several studies have reported reduced numbers of cortical interneuron subtypes across different species and brain regions upon aging (Zimmer-Bensch 2019b). Features of cortical inhibitory defects involve loss of synaptic contacts, decreased neurotransmitter release, and reduced post-synaptic responsiveness to neurotransmitters. Due to the critical function of GABAergic interneurons in cortical information processing, the age-related structural and functional defects are strongly suggested to be implicated in the age-associated cognitive decline (Rozycka and Liguz-Lecznar 2017).

DNMT1 has been described to be implicated in the age-associated loss of cortical interneurons (Hahn et al. 2020) (Fig. 15.2). Conditional deletion of *Dnmt1* in parvalbumin-expressing cortical interneurons ameliorates their age-related decline, which is accompanied by improved senso-motoric performances of aged mice (Hahn et al. 2020). However, DNMT1-dependent regulation of cell death- and survival-associated genes seems to play a rather subordinate role, whereas the DNMT1-dependent regulation of proteostasis-related gene expression might be important (Hahn et al. 2020).

15.9 DNMTs in Neurodegeneration

Neurodegenerative diseases (NDDs) encompass a wide variety of disorders characterized by functional perturbances in neurons accompanied by neuron loss and tissue degeneration in the peripheral or central nervous system (Vila and Przedborski 2003). Aspects of cellular homeostasis underlying NDDs range from dysfunctional mitochondria and compromised proteostasis to altered gene expression and abnormal transcriptional regulation, with epigenetics gaining significant attention over the years due to its involvement in these processes (Lovrečić et al. 2013). Furthermore, modern research has

frequently highlighted the role of epigenetics in brain development where dynamic epigenetic signatures, such as histone modifications and DNA methylation, drive and coordinate important processes such as neuronal differentiation and cell survival (Zimmer-Bensch, 2018). Interestingly, the dynamicity of epigenetic signatures was revealed to carry on into the adult brain with implications in memory acquisition and consolidation as well as age-related loss of neural cells (Sweatt 2016; Hahn et al. 2020). Unsurprisingly, these findings have propelled epigenetic mechanisms and dysregulations to the forefront in investigations of NDDs for better diagnostic agents and therapeutics.

15.9.1 Alzheimer's Disease and Tauopathies

The vast majority of NDD patients are affected by Alzheimer's disease (AD) (Selkoe and Lansbury Jr 1999; Zimmer-Bensch and Zempel 2021). Symptomatically, AD manifests itself initially as cognitive deficits, such as memory loss, confusion, and poor judgment, with a high risk of developing into a full-blown dementia where it accounts for 60–80% of all cases (Korolev et al. 2016; Fishman 2017). The patient demise mainly results from concomitant lack of adequate nutrition and severe loss of body weight, but also from typical diseases that affect bedridden patients, such as pneumonia (Korolev et al. 2016). The pathophysiology of AD is characterized by the extracellular accumulation of plaques made up of the Amyloid-beta (Aβ) protein, and the intracellular aggregation of the microtubule-associated protein TAU, encoded by the *MAPT* gene (Zimmer-Bensch and Zempel 2021).

Albeit heavily debated, several studies agree that abnormal levels of Aβ protein, cleaved out of the amyloid precursor protein (APP) primarily by the PSEN1/PSEN2 complex, as well as the extracellular deposition of Aβ are the main culprits behind the development and progression of AD (Selkoe and Hardy 2016; Gulisano et al. 2018). Indeed, mutations in the genes *APP*, *PSEN1*, and *PSEN2* were found to be causative for patients with autosomal dominant inheritable forms of early-onset AD (Zimmer-Bensch and Zempel 2021).

Interestingly, in the absence or suppression of TAU protein, mouse and cell culture models for AD failed to show a significant effect upon exposure to Aβ or its overproduction, hinting toward a possible role of the TAU protein as a mediator in neurodegeneration (Roberson et al. 2007; Zempel et al. 2013). Furthermore, the accumulation and aggregation of TAU protein was found to correlate better with the AD-associated loss of synapses and cognitive impairment, with PET-imaging technologies being able to predict structural brain deterioration in full-blown AD (La Joie et al. 2020; Biel et al. 2021).

As the importance of the TAU protein in neurodegeneration and neuronal dysfunction became more evident, many studies went beyond AD and started to investigate the heterogenous group of TAU protein-related NDDs called tauopathies that are characterized by the neural and/or glial deposition of TAU. Histopathological hallmarks of tauopathies include but are not limited to the hyperphosphorylation of the TAU protein and the formation of neurofibrillary tangles (Zimmer-Bensch and Zempel 2021). Tauopathies clinically manifest themselves as cognitive deficits, motor neuron disease, and movement disorders in diverse combinations or in an isolated manner (Murley et al. 2020) and can be classified into primary and secondary tauopathies depending on whether TAU is instrumental to the pathology or appears secondary alongside other cerebral pathologies (Zimmer-Bensch and Zempel 2021). Despite growing appeals, genetic and signaling-based aberrations, such as familial mutations in the *MAPT* gene, fail to illustrate a mechanistic basis for the emergence and progression of both sporadic and genetic forms of AD and tauopathies. Epigenetics may provide a further piece of the puzzle and contribute to a better understanding of environmental triggers that are involved in AD and tauopathies.

Abnormal gene expression, loss of chromatin structure, and genomic instability are considered to be hallmarks of both aging and complex diseases such as AD (López-otín et al. 2013;

Spiegel et al. 2014). These changes in cellular homeostasis are deeply associated with epigenetic mechanisms that can respond to environmental cues (Grant et al. 2002; Rowbotham et al. 2015), such as DNA methylation catalyzed by DNMTs (Greenberg 2020). Prior studies have pointed out that altered expression levels of DNMTs (Cui and Xu 2018) are associated with changes in synaptic plasticity, memory, and learning (Levenson et al. 2006; Morris and Monteggia 2014), further emphasizing the role of DNMTs in aging and AD-related symptoms. In particular, the aging-associated decrease in *Dnmt3A2* expression is implicated in cognitive impairment, as the symptoms were alleviated upon a rescue of the *Dnmt3A2* expression levels in mice (Oliveira et al. 2012).

As mentioned previously, the expression of *DNMT*s decreases upon aging, accompanied by a global hypomethylation and local hypermethylation in aging brains of various species (Johnson et al. 2012; Hahn et al. 2020). Such changes are presumed to contribute to transcriptional alterations seen in AD and tauopathies (McKinney et al. 2019; Salameh et al. 2020). Hence, DNA methylation could pose as a mechanism in the transcriptional regulation of AD-associated genes. Indeed, previous studies have revealed an age-related hypomethylation in the promoter region of *APP*, *PSEN1*, and *PSEN2* (Fig. 15.3), which were linked to the extracellular deposition of Aβ in the aged brain (Tohgi et al. 1999a, b).

Similarly, an age-related decrease in *MAPT* expression was evidenced alongside alterations in the methylation levels of its promoter region, emphasizing the role of DNA methylation in tauopathies (Tohgi et al. 1999b). Furthermore, aberrant methylation levels in the promoter regions of genes involved in TAU phosphorylation, such as *GSK3B* (Nicolia et al. 2017) and *Cdk5* (Li et al. 2015), as well as their increased expression, were shown to play a crucial role in tauopathies and AD (Yu et al. 2019). In addition to DNMTs, the hyperphosphorylation of TAU could be further influenced by TET-dependent, active DNA demethylation (Zimmer-Bensch and Zempel 2021). Indeed, this becomes evident for BDNF, a key player in synaptic plasticity and synaptogenesis in the hippocampus (Song et al. 2015). *BDNF*, whose transcriptional accessibility is regulated partly via TET1 (Ambigapathy et al. 2015), is implicated in TAU hyperphosphorylation (Tanila 2017), indicating that a TET-mediated demethylation of *BDNF* could influence the phosphorylation of the TAU protein.

Overall, these findings underline the role of DNMTs as well as TETs in AD and tauopathies (Fig. 15.3). Although a substantial number of the affected genes are proposed to be downstream effectors of Aβ pathology, the majority are suggested to be upstream of TAU pathology which seems to be the driving force behind the cognitive dysfunction seen in AD and tauopathy patients (Zimmer-Bensch and Zempel 2021). In the future, the modulation of DNMT activity to restore its healthy function or locus-specific gene editing methods to re-establish DNA methylation patterns could open new doors for targeted epigenetic therapies against AD and tauopathies.

15.9.2 Huntington's Disease

The interest in the role of epigenetics in neurodegenerative diseases was further fueled by recent developments in Huntington's disease (HD) research. HD is a neurodegenerative disease predominantly caused by an inherited expansion mutation in the Huntingtin protein (HTT), leading to N-terminal polyQ repeats and a subsequent misfolding of HTT (Zimmer-Bensch 2020). This mutant form of HTT (mHTT) is prone to aggregations and forms intracellular inclusion bodies, ultimately leading to severe atrophy in the dorsal striatum accompanied by an abnormal increase in astrocytes as well as a loss of striatal and cortical neurons (Hedreen et al. 1991; DiFiglia et al. 1997; Lee et al. 2013).

For decades, researchers have been trying to decipher the exact functions of healthy and mutated HTT where prominent progress has been made on the epigenetics front. Multiple studies have suggested that HTT can interact with transcription factors and histone modifying

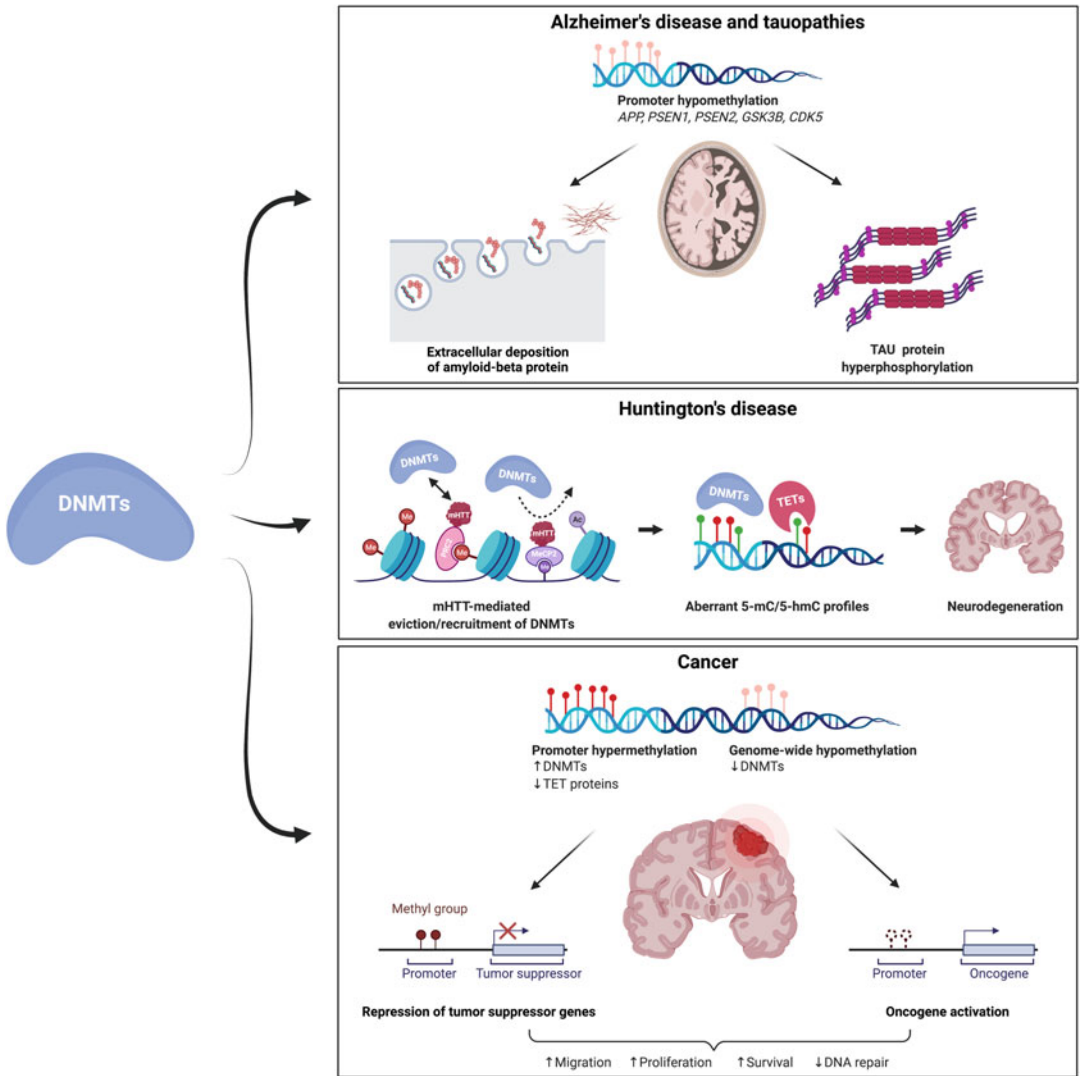


Fig. 15.3 The implications of DNMTs in neurodegenerative diseases and cancers of the brain. Upper panel: Altered methylation level in Alzheimer's disease and other tauopathies. Promoter hypomethylation of amyloid precursor protein gene (*APP*) and *PSEN1/2*, encoding for proteins involved in APP-cleavage, leads to increased extracellular deposition of amyloid-beta protein (Aβ), regarded as a hallmark for the development and progression of Alzheimer's disease (AD). Genes encoding for proteins involved in the hyperphosphorylation of TAU protein such as *GSK3B* and *CDK5* are similarly hypomethylated in their promoter regions in AD and related tauopathies. Middle panel: In Huntington's disease (HD), the recruitment or eviction of DNMTs mediated by the mutant Huntingtin protein (mHTT) and its interactions with epigenetic writers and readers, such as PRC2 and MeCP2, is proposed to lead to aberrant 5-mC/5-hmC profiles of genes involved in neuronal development, function, and survival, leading to significant cerebral atrophy. Lower panel: In the context of cancer, differential global activity of DNMTs, as well as reduced TET activity in promoter regions, results in promoter hypermethylation and genome-wide hypomethylation, leading to the repression of tumor suppressor genes and the activation of oncogenes. Resulting aberrant expression profiles facilitate tumorigenesis and metastasis by enhanced migration, invasion, proliferation, and survival while decreased DNA repair activity accelerates the mutation rate

enzymes, e.g., REST, PRC2, and MeCP2, thus highlighting HTTs ability to interact with key epigenetic players (Seong et al. 2009; Buckley et al. 2010). In the case of mHTT, disruptions of these interactions could potentially contribute to the abnormal transcriptional regulation which is a hallmark of HD.

In addition to HTT's interaction with histone modifying complexes and transcription factors, the academic community has extensively explored changes in DNA methylation levels and signatures, such as 5-mC and 5-hmC patterns, where aberrations were reported in HD patients and transgenic mice (Fig. 15.3) (Ng et al. 2013; Villar-Menéndez et al. 2013; Wood 2013). The oxidation of 5-mC to 5-hmC is a vital step in active DNA demethylation and crucial for the dynamic DNA methylation-dependent transcriptional regulation (Xu and Wong 2015). Abnormal 5-mC/5-hmC profiles were seen for genes implicated in neuronal development, function, and survival (Wang et al. 2013) and could thus be attributed to the decline in neuronal function and neuronal death present in HD. Yet, little research has been conducted to investigate the mechanistic behind such changes in DNA methylation patterns in full molecular detail, whether they are direct consequences due to the altered interactome of mHTT, or whether these occur indirectly, e.g., as adaptive response to impaired cellular physiology.

Beside the ability of HTT to directly interact with histone modifying enzymes, abnormal, mHTT-mediated deposition and/or removal of histone marks can lead to the eviction or recruitment of DNMTs (Zimmer-Bensch 2020) (Fig. 15.3). Prior studies suggest the enrichment of histone marks to correlate with DNA methylation, inversely or positively, depending on whether they are associated with the decondensed euchromatin or the tightly packed heterochromatin. The euchromatin-associated H3K4me3 has an inverse correlation with DNA methylation levels and was shown to preclude DNMT3-dependent DNA methylation (Rose and Klose 2014). In contrast, the heterochromatin-associated H3K9me3 and H3K27me3 were found to have a positive correlation with DNA methylation where both were shown to directly recruit DNMTs (Lehnertz et al. 2003; Viré et al. 2006; Hashimoto et al. 2009; Liu et al. 2018). Similarly, the binding of DNA demethylases such as TETs could be influenced, adding another layer to this mechanistic conundrum.

Based on observations made in prior studies, another scenario could be that altered expression of DNMTs leads to the abnormal DNA methylation signatures seen in HD (Thomas 2016). The expression of *Dnmt1* was found to be decreased in a HD cell model (Tobin and Signer 2000), and a reduced striatal and cortical *Dnmt1* expression was documented in transgenic HD mice. Despite the assumption that reduced *Dnmt1* expression correlates with lower levels of DNA methylation, mHTT-expressing neurons showed increased methylation levels in promoter regions of key HD-relevant genes such as *Bdnf* (Pan et al. 2016). These counter-intuitive findings were attributed to reduced expression levels of DNA demethylases leading to elevated methylation levels in HD patients and model systems (Thomas 2016). It was proposed that this increase in methylation levels could lead to a diminished expression of *Dnmt*s by means of a feedback-regulation. Indeed, the reduced expression of *Dnmt*s may be favorable in the context of HD, as both the inhibition of DNMTs and a knockdown of *Dnmt1* and *Dnmt3A* were shown to decrease mHTT-associated neurotoxicity in primary striatal and cortical neurons (Pan et al. 2016).

Altogether, these findings add another level of complexity to the pathophysiology of NDDs, while further emphasizing the hierarchical dilemma of epigenetics, i.e., the imperative distinction between the cause and the consequence of the observed changes in DNA methylation.

15.10 Role of DNMTs in Brain Cancer

Cancer is a group of diseases, arising from abnormal gene expression programs that shift the balance of oncogenic and tumor suppressive mechanisms. Making up 2% and 23% of all primary tumors in adults and children, respectively, brain cancer is one of the most malignant forms of cancer with more than a quarter of all pediatric cases resulting in the patient's demise (Marie and

Shinjo 2011). Gliomas are the most frequent and malignant type of brain tumor in adults, with glioblastoma (GBM) having a median overall survival of 14.6 months. In turn, medulloblastoma, which starts in the cerebellum, accounts for the most cases of malignant, pediatric brain tumors (Marie and Shinjo 2011; Bartlett et al. 2013). Prior studies on glioblastoma (Parsons et al. 2008) and medulloblastoma (Parsons et al. 2011) demonstrated the presence of mutations implicated in their initiation and development. Yet, in the case of medulloblastoma, the tumors tend to have a low mutational burden as pediatric patients are incapable of acquiring spontaneous mutations due to their young age (Blaeschke et al. 2019). Indeed, cumulative evidence points to the involvement of other mechanisms beside somatic mutations in the formation of brain tumors, such as structural and numerical aberrations in chromosomes (Larsen 2010) and epigenetic alterations (Sharma et al. 2010).

Over the past decade, epigenetics has taken a center stage in cancer research due to its crucial role in the maintenance and regulation of gene expression programs. Many studies have shown how exposure to environmental agents, genetic alterations, or even aging can perturb the epigenetic machinery, creating a permissive environment for cancer to develop and progress (Easwaran and Baylin 2019). As one of the key epigenetic mechanisms, aberrant DNA methylation signatures were implicated in various types of cancer, with diverse gene-bodies, intergenic regions, and repetitive elements being hypomethylated (Ehrlich and Lacey 2013) (Fig. 15.3). The global hypomethylation seen in cancer has been implicated in the dysregulation of the genome, leading to genomic instability, oncogenic activation, loss of genomic imprinting, and the reactivation of transposable elements, ultimately resulting in increased mutational rates and tumorigenesis (Chen et al. 1998; Eden et al. 2003; Gaudet et al. 2003; Holm et al. 2005; Hur et al. 2014). Interestingly, a local hypermethylation accompanies this global hypomethylation, potentially indicating a differential activity of epigenetic modifiers, such as DNMTs, depending on the genomic region (Easwaran and Baylin 2019).

15.10.1 Promoter Methylation

One of the proposed mechanisms for the abnormal activation of oncogenes and the inactivation of tumor suppressor genes is the methylation of cytosines in promoter regions of genes, with the gain/presence of methylation being associated with gene silencing, and the loss/absence being associated with transcriptional activation (Ehrlich and Lacey 2013) (Fig. 15.3). DNA methylation in the promoter region prevents the binding of key transcription factors (TFs) and by this directly inhibits gene expression (Moore et al. 2013). Additionally, DNA methylation can also influence post-translational modifications of histones through methyl-binding proteins (MBD). MBDs act as adaptors for histone modifying enzymes that can change the chromatin state and regulate the accessibility of the methylated site (Ng et al. 2000). In addition, DNMTs themselves were found to crosstalk with histone modifying enzymes in a methylation-independent fashion (Symmank et al. 2018), further complexifying DNMT-mediated transcriptional regulation.

In the context of glioma, mutations in isocitrate dehydrogenase genes *IDH1* and *IDH2* lead to the production and accumulation of 2-hydroxyglutarate instead of α-ketoglutarate, the subsequent inhibition of α-ketoglutarate-dependent enzymes such as TET2, and ultimately to DNA hypermethylation (Dang et al. 2009; Figueroa et al. 2010; Scourzic et al. 2015). Furthermore, the epigenetic silencing of the DNA repair gene *MGMT* via promoter hypermethylation represents another prominent epigenetic alteration in glioma with the promoter methylation status of the *MGMT* gene becoming a predictive biomarker in neuro-oncology. The hypermethylation of the promoter, resulting in the absence of MGMT, was found to be beneficial in the treatment of glioma via temozolomide (TMZ), as active MGMT can repair O6-methylguanine, a toxic DNA lesion caused by TMZ, and diminish the effects of the treatment (Hegi et al. 2005).

In medulloblastoma, the promoters of tumor suppressor genes *CASP8*, *HIC1*, and *CDKN2A* (Lindsey et al. 2004; Sexton-Oates et al. 2015)

and the DNA repair gene *MGMT* (von Bueren et al. 2012) were found to be methylated. In line with the role of Sonic-Hedgehog (SHH) and Wnt signaling pathways in the activation of tumor formation in medulloblastoma (Cambruzzi 2018), *PTCH1* and the *SFRP* family of proteins that are involved in the negative regulation of SHH and Wnt signaling, respectively, were found to be silenced via promoter methylation (Pritchard and Olson 2008; Kongkham et al. 2010).

15.10.2 Methylation of Distal Regulatory Elements

Distal regulatory elements in the genome are able to modulate the transcription of distinct genes through structuring the chromatin organization. Due to their enrichment for TF binding sites, changes in methylation levels and/or patterns in enhancer regions might interfere with the binding of TFs (Easwaran and Baylin 2019), either by changes in methylation itself or by secondary changes in histone modifications that result in an altered chromatin structure. Indeed, this was shown to be the case in gliomas with mutated *IDH*, where the hypermethylation of the binding site for the transcriptional regulator CTCF leads to the eviction of the CTCF/cohesion complex, resulting in an interaction between the enhancer and the oncogene *PDGFRA*. Notably, this interaction was not found in healthy individuals or glioma cases without the IDH mutation (Flavahan et al. 2016).

15.10.3 Implications of Altered DNMT Expression and Targeting in Brain Cancer and Therapy Resistance

As DNMTs catalyze DNA methylation, alteration in their expression could mediate the aberrant methylation signatures in cancer. Indeed, significant overexpression of *DNMT1* and *DNMT3B* was shown in gliomas where the promoter regions of *DNMT1* and *DNMT3B* had a differential histone code with distinct marks for euchromatin, compared to normal tissue that are predominantly enriched with repressive histone marks (Rajendran et al. 2011) (Fig. 15.3). In human glioma biopsies, the expression levels of *DNMT1* and *DNMT3B* were shown to coregulate the methylation status of the apoptosis-related *BIRC5*, *TMS1,* and *CASP8*, but not of other apoptosis-related genes such as *BCL2*, *BCL2L2,* and *BAX*, indicating that DNMTs could orchestrate the emergence of the apoptosis evasion phenotype in glioma by mediating the regulation of distinct apoptosis-associated genes (Hervouet et al. 2010).

A similar overexpression of *DNMT*s was observed for medulloblastoma patients, with the overexpression of *DNMT3B* being the most common (Pócza et al. 2016). Yet, no correlation was found between the expression levels of *DNMT*s and the age of onset, histological subtype, or overall survival in medulloblastoma (Pócza et al. 2016).

Apart from transcriptional dysregulation of DNMTs, their targeting to specific genomic loci could cause the alterations in DNA methylation signatures seen in the different types of brain, mediating pathophysiological processes and/or therapy resistance. Emerging evidence proposes lncRNAs to orchestrate the recruitment of epigenetic writers such as DNMTs or histone modifying complexes to specific genomic sites (Wang et al. 2015; Jain et al. 2016; Xiong et al. 2018). Modern research has enabled genome-wide studies of tumor samples, which have identified a great number of lncRNAs implicated in various types of cancer (Bhan et al. 2017). Even though lncRNAs were shown to directly interact with DNMTs as well (Wang et al. 2015), relatively little is known about this interaction in the context of cancer. Recent studies have begun to slowly close the gap in literature, with the lncRNA–DNMT interactions having been demonstrated in renal, breast, and thyroid cancer (Wu et al. 2018; Song et al. 2019; Zhao and Hu 2019). In GBM, an overexpression of the lncRNA *HOTAIRM1* was found to promote

tumor growth and upregulate the expression of the oncogene *HOXA1* by evicting DNMTs, G9A, and EZH2, leading to the demethylation of H3K9 and H3K27 and a reduction in DNA methylation levels (Li et al. 2018). In TMZ-resistant glioma, the lncRNA *SNHG12* was shown to be upregulated due to a loss of DNA methylation in its promoter region, with clinical studies evidencing poor survival of GBM patients in presence of an *SNHG12* overexpression (Lu et al. 2020).

15.10.4 Crosstalk of DNMTs and miRNA-Mediated Translational Control

In addition to the transcriptional level within the nucleus, lncRNAs but also small non-coding RNAs (sncRNA) such as microRNAs (miRNAs) can modulate post-transcriptional events in the cytoplasm (Wei et al. 2017). miRNAs can regulate the expression of target genes on a post-transcriptional level by binding to complementary sequences in mRNA molecules and silencing them (Bartel 2009). The expression of many miRNAs is increased or decreased in brain cancer leading to dysregulations in cellular pathways involved in proliferation, apoptosis, cell survival, and metastasis (Li et al. 2013; Haltom et al. 2020). The expression of miRNAs can be modulated via DNMT-mediated DNA methylation, underlining the crosstalk between the two epigenetic regulatory mechanisms (Chuang and Jones 2007). Indeed, this was evidenced by Zhou et al. (2015), as *DNMT1* expression was shown to be downregulated in a TMZ-resistant GBM cell line compared to the control, leading to a hypomethylation of the miR-20a promoter and an increase in its expression. The overexpression of *DNMT1* was shown to suppress miR-20a expression and restore sensitivity to TMZ, highlighting the crucial role of DNMT1 in the development of chemoresistance in glioma. Conversely, miRNAs are able to influence the expression of *DNMT*s. In a previous study, miR-152-3p was shown to directly target *DNMT1* and lower its expression (Sun et al. 2017). Due to the downregulation of miR-152-3p in GBM tissue and glioma cells, the expression of *DNMT1* was found to be increased, leading to a hypermethylation of the tumor suppressor gene *NF2* and its subsequent downregulation. Both the overexpression of miR-152-3p and the knockdown of *DNMT1* were shown to result in a rescue of *NF2* expression, increased apoptosis, and reduced invasive activity (Sun et al. 2017).

In summary, abnormal DNA methylation signatures seen in brain cancer could be attributed to alterations in the recruitment and activity of DNMTs in distinct genomic regions, rather than a global loss or gain of their activity, as the global hypomethylation is accompanied by a concomitant hypermethylation of specific genomic loci. Indeed, preclinical studies, where DNMT activity was inhibited in in vivo and in vitro models of glioma, have shown efficacy (Rajendran et al. 2011), but have not been translated into successful therapies so far (Stewart et al. 2009). Beyond changes in the expression levels or activity of DNMTs, understanding how DNMTs target or avoid distinct loci in the genome, which results in the global hypomethylation and the local hypermethylation seen in cancer, remains the most challenging problem to this date, and solving it might be the key in discovering epigenetic biomarkers or therapies for cancer.

15.11 Conclusions

DNMTs are widely and distinctively expressed in different cell types of the brain, being implicated in orchestrating brain development, functionality, and age-related processes. Their dysregulated expression and function have been proposed to be implicated in a wide range of diseases, including neurodevelopmental and neuropsychiatric disorders, neurodegenerative diseases as well as brain cancer. However, to approach their full-blown potential as therapeutic targets, we need to dissect their interactome, mechanisms of transcriptional regulation, context-specific targeting to specific genomic sites, and regulation of their activity. DNMTs have been shown to crosstalk with histone modifying and miRNA-mediated

mechanisms, and to bind specific lncRNAs. How are these specific interactions and the recruitment to distinct genomic sites achieved? What role do post-translational modifications of the different DNMTs play in this diverse spectrum of interactions? These questions have to be addressed in cell-type specific contexts. The enormous progress that has been achieved in sequencing-based technologies, allowing single cell resolution even at multi-omics level, might provide an answer to these challenging questions in the near future. Furthermore, we need to combine multi-omics approaches with functional readouts, reaching a higher integrational level of analyses. CRISPR-Cas mediated epigenomic and genomic editing in combination with iPSC approaches might allow the development of targeted and personalized therapeutics, even for so-far incurable diseases of the brain.

Acknowledgment All figures were generated with https://BioRender.com with a respective publication license provided by the Biology department of the RWTH Aachen University.

Funding This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—368482240/GRK2416DFG; DFG ZI1224/8-1; DFG ZI1224/13-1.

References

Abraham WC, Robins A (2005) Memory retention—the synaptic stability versus plasticity dilemma. Trends Neurosci 28(2):73–78

Abraham WC, Logan B, Greenwood JM, Dragunow M (2002) Induction and experience-dependent consolidation of stable long-term potentiation lasting months in the hippocampus. J Neurosci 22(21):9626–9634

Abraham WC, Jones OD, Glanzman DL (2019) Is plasticity of synapses the mechanism of long-term memory storage? NPJ Sci Learn 4(1):1–10

Agirman G, Broix L, Nguyen L (2017) Cerebral cortex development: an outside-in perspective. FEBS Lett 591(24):3978–3992

Ambigapathy G, Zheng Z, Keifer J (2015) Regulation of BDNF chromatin status and promoter accessibility in a neural correlate of associative learning. Epigenetics 10(10):981–993

Ariazi J, Benowitz A, De Biasi V, Den Boer ML, Cherqui S, Cui H, Douillet N, Eugenin EA, Favre D, Goodman S (2017) Tunneling nanotubes and gap junctions—their role in long-range intercellular communication during development, health, and disease conditions. Front Mol Neurosci 10:333

Ashley J, Cordy B, Lucia D, Fradkin LG, Budnik V, Thomson T (2018) Retrovirus-like Gag protein Arc1 binds RNA and traffics across synaptic boutons. Cell 172(1-2):262–274. e211

Bachmann S, Linde J, Bell M, Spehr M, Zempel H, Zimmer-Bensch G (2021) DNA methyltransferase 1 (DNMT1) shapes neuronal activity of human iPSC-derived glutamatergic cortical neurons. Int J Mol Sci 22(4):2034

Barry G, Briggs JA, Hwang DW, Nayler SP, Fortuna PR, Jonkhout N, Dachet F, Maag JL, Mestdagh P, Singh EM (2017) The long non-coding RNA NEAT1 is responsive to neuronal activity and is associated with hyperexcitability states. Sci Rep 7(1):1–11

Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136(2):215–233

Bartlett F, Kortmann R, Saran F (2013) Medulloblastoma. Clin Oncol 25(1):36–45

Bartsch D, Ghirardi M, Skehel PA, Karl KA, Herder SP, Chen M, Bailey CH, Kandel ER (1995) Aplysia CREB2 represses long-term facilitation: relief of repression converts transient facilitation into long-term functional and structural change. Cell 83(6): 979–992

Baumann V, Wiesbeck M, Breunig CT, Braun JM, Köferle A, Ninkovic J, Götz M, Stricker SH (2019) Targeted removal of epigenetic barriers during transcriptional reprogramming. Nat Commun 10(1):1–12

Bayer SA, Altman J (1991) Neocortical development, vol 1. Raven, New York

Bayraktar G, Yuanxiang P, Confettura AD, Gomes GM, Raza SA, Stork O, Tajima S, Suetake I, Karpova A, Yildirim F (2020) Synaptic control of DNA methylation involves activity-dependent degradation of DNMT3A1 in the nucleus. Neuropsychopharmacology 45(12):2120–2130

Bear MF, Abraham WC (1996) Long-term depression in hippocampus. Annu Rev Neurosci 19(1):437–462

Bédécarrats A, Chen S, Pearce K, Cai D, Glanzman DL (2018) RNA from trained Aplysia can induce an epigenetic engram for long-term sensitization in untrained Aplysia. eneuro 5(3)

Bekinschtein P, Katche C, Slipczuk L, Gonzalez C, Dorman G, Cammarota M, Izquierdo I, Medina JH (2010) Persistence of long-term memory storage: new insights into its molecular signatures in the hippocampus and related structures. Neurotox Res 18(3): 377–385

Bhan A, Soleimani M, Mandal SS (2017) Long noncoding RNA and cancer: a new paradigm. Cancer Res 77(15): 3965–3981

Bhatt DH, Zhang S, Gan W-B (2009) Dendritic spine dynamics. Annu Rev Physiol 71:261–282

Biel D, Brendel M, Rubinski A, Buerger K, Janowitz D, Dichgans M, Franzmeier N (2021) Tau-PET and in vivo Braak-staging as prognostic markers of future cognitive decline in cognitively normal to demented individuals. Alzheimers Res Ther 13(1):137

Biergans SD, Galizia CG, Reinhard J, Claudianos C (2015) Dnmts and Tet target memory-associated

genes after appetitive olfactory training in honey bees. Sci Rep 5(1):1–15

Blaeschke F, Paul MC, Schuhmann MU, Rabsteyn A, Schroeder C, Casadei N, Matthes J, Mohr C, Lotfi R, Wagner B (2019) Low mutational load in pediatric medulloblastoma still translates into neoantigens as targets for specific T-cell immunotherapy. Cytotherapy 21(9):973–986

Bliss TV, Collingridge GL (1993) A synaptic model of memory: long-term potentiation in the hippocampus. Nature 361(6407):31–39

Borrell V (2019) Recent advances in understanding neocortical development. F1000Research 8

Bridges MC, Daulagala AC, Kourtidis A (2021) LNCcation: lncRNA localization and function. J Cell Biol 220(2):e202009045

Buckley NJ, Johnson R, Zuccato C, Bithell A, Cattaneo E (2010) The role of REST in transcriptional and epigenetic dysregulation in Huntington's disease. Neurobiol Dis 39(1):28–39

Cambruzzi E (2018) Medulloblastoma, WNT-activated/SHH-activated: clinical impact of molecular analysis and histogenetic evaluation. Childs Nerv Syst 34(5): 809–815

Carlevaro-Fita J, Johnson R (2019) Global positioning system: understanding long noncoding RNAs through subcellular localization. Mol Cell 73(5):869–883

Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R (1998) DNA hypomethylation leads to elevated mutation rates. Nature 395(6697):89–93

Cheng C-H, Lin Y-Y (2013) Aging-related decline in somatosensory inhibition of the human cerebral cortex. Exp Brain Res 226(1):145–152

Chestnut BA, Chang Q, Price A, Lesuisse C, Wong M, Martin LJ (2011) Epigenetic regulation of motor neuron cell death through DNA methylation. J Neurosci 31(46):16619–16636

Chivet M, Javalet C, Laulagnier K, Blot B, Hemming FJ, Sadoul R (2014) Exosomes secreted by cortical neurons upon glutamatergic synapse activation specifically interact with neurons. J Extracell Vesicles 3(1): 24722

Chuang JC, Jones PA (2007) Epigenetics and microRNAs. Pediatr Res 61(7):24–29

Civiero L, Cirnaru MD, Beilina A, Rodella U, Russo I, Belluzzi E, Lobbestael E, Reyniers L, Hondhamuni G, Lewis PA (2015) Leucine-rich repeat kinase 2 interacts with p21-activated kinase 6 to control neurite complexity in mammalian brain. J Neurochem 135(6): 1242–1256

Clark BC, Taylor JL (2011) Age-related changes in motor cortical properties and voluntary activation of skeletal muscle. Curr Aging Sci 4(3):192–199

Clements EG, Mohammad HP, Leadem BR, Easwaran H, Cai Y, Van Neste L, Baylin SB (2012) DNMT1 modulates gene expression without its catalytic activity partially through its interactions with histone-modifying enzymes. Nucleic Acids Res 40(10): 4334–4346

Cohen LD, Zuchman R, Sorokina O, Müller A, Dieterich DC, Armstrong JD, Ziv T, Ziv NE (2013) Metabolic turnover of synaptic proteins: kinetics, interdependencies and implications for synaptic maintenance. PLoS One 8(5):e63191

Costa MR, Müller U (2015) Specification of excitatory neurons in the developing cerebral cortex: progenitor diversity and environmental influences. Front Cell Neurosci 8:449

Cui D, Xu X (2018) DNA methyltransferases, DNA methylation, and age-associated cognitive function. Int J Mol Sci 19(5):1315

Cummings JA, Mulkey RM, Nicoll RA, Malenka RC (1996) Ca2+ signaling requirements for long-term depression in the hippocampus. Neuron 16(4):825–833

Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, Fantin VR, Jang HG, Jin S, Keenan MC (2009) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature 462(7274):739–744

Davidson PSR, Glisky EL (2002) Neuropsychological correlates of recollection and familiarity in normal aging. Cogn Affect Behav Neurosci 2(2):174–186

Day JJ, Sweatt JD (2010) DNA methylation and memory formation. Nat Neurosci 13(11):1319–1323

DiFiglia M, Sapp E, Chase KO, Davies SW, Bates GP, Vonsattel JP, Aronin N (1997) Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. Science 277(5334): 1990–1993

Ding E, Liu J, Guo H, Shen H, Zhang H, Gong W, Song H, Zhu B (2018) DNMT1 and DNMT3A haplotypes associated with noise-induced hearing loss in Chinese workers. Sci Rep 8(1):1–10

Donnelly CJ, Fainzilber M, Twiss JL (2010) Subcellular communication through RNA transport and localized protein synthesis. Traffic 11(12):1498–1505

Du J, Johnson LM, Jacobsen SE, Patel DJ (2015) DNA methylation pathways and their crosstalk with histone methylation. Nat Rev Mol Cell Biol 16(9): 519–532

Duke CG, Kennedy AJ, Gavin CF, Day JJ, Sweatt JD (2017) Experience-dependent epigenomic reorganization in the hippocampus. Learn Mem 24(7):278–288

Easwaran H, Baylin SB (2019) Origin and mechanisms of DNA methylation dynamics in cancers. In: The DNA, RNA, and histone methylomes. Springer, pp 27–52

Eden A, Gaudet F, Waghmare A, Jaenisch R (2003) Chromosomal instability and tumors promoted by DNA hypomethylation. Science 300(5618):455–455

Ehrlich M, Lacey M (2013) DNA hypomethylation and hemimethylation in cancer. Epigenetic Alterations Oncogenesis 31–56

Eichenbaum H, Otto T, Cohen NJ (1992) The hippocampus—what does it do? Behav Neural Biol 57(1):2–36

Fan G, Beard C, Chen RZ, Csankovszki G, Sun Y, Siniaia M, Biniszkiewicz D, Bates B, Lee PP, Kühn R (2001) DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. J Neurosci 21(3):788–797

Fan G, Martinowich K, Chin MH, He F, Fouse SD, Hutnick L, Hattori D, Ge W, Shen Y, Wu H (2005) DNA methylation controls the timing of astrogliogenesis through regulation of JAK-STAT signaling. Development 132(15):3345–3356

Fasolino M, Liu S, Wang Y, Zhou Z (2017) Distinct cellular and molecular environments support aging-related DNA methylation changes in the substantia nigra. Epigenomics 9(1):21–31

Faux C, Rakic S, Andrews W, Britto JM (2012) Neurons on the move: migration and lamination of cortical interneurons. Neurosignals 20(3):168–189

Feng J, Chang H, Li E, Fan G (2005) Dynamic expression of de novo DNA methyltransferases Dnmt3a and Dnmt3b in the central nervous system. J Neurosci Res 79(6):734–746

Feng J, Fouse S, Fan G (2007) Epigenetic regulation of neural gene expression and neuronal function. Pediatr Res 61(7):58–63

Feng J, Zhou Y, Campbell SL, Le T, Li E, Sweatt JD, Silva AJ, Fan G (2010) Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. Nat Neurosci 13(4):423–430

Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF (2010) Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell 18(6):553–567

Fishell G (2008) Perspectives on the developmental origins of cortical interneuron diversity. In: Cortical development: genes and genetic abnormalities, pp 21–34

Fishman E (2017) Risk of developing dementia at older ages in the United States. Demography 54(5): 1897–1919

Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suvà ML, Bernstein BE (2016) Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature 529(7584):110–114

Franco SJ, Müller U (2013) Shaping our minds: stem and progenitor cell diversity in the mammalian neocortex. Neuron 77(1):19–34

Friocourt GM, Parnavelas JG (2011) Identification of Arx targets unveils new candidates for controlling cortical interneuron migration and differentiation. Front Cell Neurosci 5:28

García-Moreno F, Vasistha NA, Trevia N, Bourne JA, Molnar Z (2012) Compartmentalization of cerebral cortical germinal zones in a lissencephalic primate and gyrencephalic rodent. Cereb Cortex 22(2): 482–492

Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, Leonhardt H, Jaenisch R (2003) Induction of tumors in mice by genomic hypomethylation. Science 300(5618):489–492

Gelman DM, Martini FJ, Nóbrega-Pereira S, Pierani A, Kessaris N, Marín O (2009) The embryonic preoptic area is a novel source of cortical GABAergic interneurons. J Neurosci 29(29):9380–9389

Gelman D, Griveau A, Dehorter N, Teissier A, Varela C, Pla R, Pierani A, Marín O (2011) A wide diversity of cortical GABAergic interneurons derives from the embryonic preoptic area. J Neurosci 31(46): 16570–16580

Glisky EL (2007) Changes in cognitive function in human aging. Brain Aging 3–20

Goldie BJ, Dun MD, Lin M, Smith ND, Verrills NM, Dayas CV, Cairns MJ (2014) Activity-associated miRNA are packaged in Map1b-enriched exosomes released from depolarized neurons. Nucleic Acids Res 42(14):9195–9208

Goto K, Numata M, Komura J-I, Ono T, Bestor TH, Kondo H (1994) Expression of DNA methyltransferase gene in mature and immature neurons as well as proliferating cells in mice. Differentiation 56(1-2): 39–44

Grant WB, Campbell A, Itzhaki RF, Savory J (2002) The significance of environmental factors in the etiology of Alzheimer's disease. J Alzheimers Dis 4(3):179–189

Greenberg MVC (2020) Get Out and Stay Out: New Insights Into DNA Methylation Reprogramming in Mammals. Front Cell Dev Biol 8:1721

Greig LC, Woodworth MB, Galazo MJ, Padmanabhan H, Macklis JD (2013) Molecular logic of neocortical projection neuron specification, development and diversity. Nat Rev Neurosci 14(11):755–769

Grinman E, Nakahata Y, Avchalumov Y, Espadas I, Swarnkar S, Yasuda R, Puthanveettil SV (2021) Activity-regulated synaptic targeting of lncRNA ADEPTR mediates structural plasticity by localizing Sptn1 and AnkB in dendrites. Sci Adv 7(16):eabf0605

Gulisano W, Maugeri D, Baltrons MA, Fà M, Amato A, Palmeri A, D'Adamio L, Grassi C, Devanand D, Honig LS (2018) Role of amyloid-β and tau proteins in Alzheimer's disease: confuting the amyloid cascade. J Alzheimers Dis 64(s1):S611–S631

Guo J, Anton E (2014) Decision making during interneuron migration in the developing cerebral cortex. Trends Cell Biol 24(6):342–351

Guo JU, Ma DK, Mo H, Ball MP, Jang M-H, Bonaguidi MA, Balazer JA, Eaves HL, Xie B, Ford E (2011) Neuronal activity modifies the DNA methylation landscape in the adult brain. Nat Neurosci 14(10): 1345–1351

Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, Yan J, Ren X, Lin S, Li J (2014) The DNA methylation landscape of human early embryos. Nature 511(7511):606–610

Hahn A, Pensold D, Bayer C, Tittelmeier J, González-Bermúdez L, Marx-Blümel L, Linde J, Groß J, Salinas-Riester G, Lingner T et al (2020) DNA methyltransferase 1 (DNMT1) function is implicated in the age-related loss of cortical interneurons. Front Cell Dev Biol 8:639

Halder R, Hennion M, Vidal RO, Shomroni O, Rahman R-U, Rajput A, Centeno TP, Van Bebber F, Capece V, Vizcaino JCG (2016) DNA methylation changes in

plasticity genes accompany the formation and maintenance of memory. Nat Neurosci 19(1):102–110
Haltom AR, Toll SA, Cheng D, Maegawa S, Gopalakrishnan V, Khatua S (2020) Medulloblastoma epigenetics and the path to clinical innovation. J Neurooncol 1–12
Hashimoto H, Horton JR, Zhang X, Cheng X (2009) UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. Epigenetics 4(1):8–14. https://doi.org/10.4161/epi.4.1.7370
Hatanaka Y, Zhu Y, Torigoe M, Kita Y, Murakami F (2016) From migration to settlement: the pathways, migration modes and dynamics of neurons in the developing brain. Proc Jpn Acad Ser B 92(1):1–19
Hedreen JC, Peyser CE, Folstein SE, Ross CA (1991) Neuronal loss in layers V and VI of cerebral cortex in Huntington's disease. Neurosci Lett 133(2):257–261
Hegi ME, Diserens A-C, Gorlia T, Hamou M-F, De Tribolet N, Weller M, Kros JM, Hainfellner JA, Mason W, Mariani L (2005) MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med 352(10):997–1003
Hervouet E, Vallette F, Cartron P (2010) Impact of the DNA methyltransferases expression on the methylation status of apoptosis-associated genes in glioblastoma multiforme. Cell Death Dis 1(1):e8–e8
Higa GSV, de Sousa E, Walter LT, Kinjo ER, Resende RR, Kihara AH (2014) MicroRNAs in neuronal communication. Mol Neurobiol 49(3):1309–1326
Hilgetag CC, Barbas H (2005) Developmental mechanics of the primate cerebral cortex. Anat Embryol 210(5-6):411
Hofman MA (2014) Evolution of the human brain: when bigger is better. Front Neuroanat 8:15
Holliday R (1999) Is there an epigenetic component in long-term memory? J Theor Biol 200(3):339–341
Holm TM, Jackson-Grusby L, Brambrink T, Yamada Y, Rideout WM III, Jaenisch R (2005) Global loss of imprinting leads to widespread tumorigenesis in adult mice. Cancer Cell 8(4):275–285
Hu X-L, Chen G, Zhang S, Zheng J, Wu J, Bai Q-R, Wang Y, Li J, Wang H, Feng H (2017) Persistent expression of VCAM1 in radial glial cells is required for the embryonic origin of postnatal neural stem cells. Neuron 95(2):309-325. e306
Huang ZJ, Paul A (2018) Diversity of GABAergic interneurons and diversification of communication modules in cortical networks. bioRxiv. https://doi.org/10.1101/490797
Hung T, Chang HY (2010) Long noncoding RNA in genome regulation: prospects and mechanisms. RNA Biol 7(5):582–585
Hur K, Cejas P, Feliu J, Moreno-Rubio J, Burgos E, Boland CR, Goel A (2014) Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. Gut 63(4):635–646
Hutnick LK, Golshani P, Namihira M, Xue Z, Matynia A, Yang XW, Silva AJ, Schweizer FE, Fan G (2009) DNA hypomethylation restricted to the murine forebrain induces cortical degeneration and impairs postnatal neuronal maturation. Hum Mol Genet 18(15):2875–2888
Inano K, Suetake I, Ueda T, Miyake Y, Nakamura M, Okada M, Tajima S (2000) Maintenance-type DNA methyltransferase is highly expressed in post-mitotic neurons and localized in the cytoplasmic compartment. J Biochem 128(2):315–321
Jain AK, Xi Y, McCarthy R, Allton K, Akdemir KC, Patel LR, Aronow B, Lin C, Li W, Yang L (2016) LncPRESS1 is a p53-regulated LncRNA that safeguards pluripotency by disrupting SIRT6-mediated de-acetylation of histone H3K56. Mol Cell 64(5):967–981
Jesus-Ribeiro J, Pires LM, Melo JD, Ribeiro IP, Rebelo O, Sales F, Freire A, Melo JB (2021) Genomic and epigenetic advances in focal cortical dysplasia types I and II: a scoping review. Front Neurosci 14:1412
Jin J, Maren S (2015) Prefrontal-hippocampal interactions in memory and emotion. Front Syst Neurosci 9:170
Jin B, Tao Q, Peng J, Soo HM, Wu W, Ying J, Fields CR, Delmas AL, Liu X, Qiu J (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. Hum Mol Genet 17(5):690–709
Johnson AA, Akman K, Calimport SR, Wuttke D, Stolzing A, De Magalhaes JP (2012) The role of DNA methylation in aging, rejuvenation, and age-related disease. Rejuvenation Res 15(5):483–494
Jones EG (2009) The origins of cortical interneurons: mouse versus monkey and human. Cereb Cortex 19(9):1953–1956
Kadriu B, Guidotti A, Chen Y, Grayson DR (2012) DNA methyltransferases1 (DNMT1) and 3a (DNMT3a) colocalize with GAD67-positive neurons in the GAD67-GFP mouse brain. J Comp Neurol 520(9):1951–1964
Kandel ER (2012) The molecular biology of memory: cAMP, PKA, CRE, CREB-1, CREB-2, and CPEB. Mol Brain 5(1):1–12
Keihani S, Kluever V, Mandad S, Bansal V, Rahman R, Fritsch E, Gomes LC, Gärtner A, Kügler S, Urlaub H (2019) The long noncoding RNA neuroLNC regulates presynaptic activity by interacting with the neurodegeneration-associated protein TDP-43. Sci Adv 5(12):eaay2670
Kiecker C, Lumsden A (2005) Compartments and their boundaries in vertebrate brain development. Nat Rev Neurosci 6(7):553–564
Kim H, Jang WY, Kang M-C, Jeong J, Choi M, Sung Y, Park S, Kwon W, Jang S, Kim MO (2016) TET1 contributes to neurogenesis onset time during fetal brain development in mice. Biochem Biophys Res Commun 471(4):437–443

Kirkbride JB, Susser E, Kundakovic M, Kresovich JK, Davey Smith G, Relton CL (2012) Prenatal nutrition, epigenetics and schizophrenia risk: can we test causal effects? Epigenomics 4(3):303–315
Klein CJ, Botuyan M-V, Wu Y, Ward CJ, Nicholson GA, Hammans S, Hojo K, Yamanishi H, Karpf AR, Wallace DC (2011) Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. Nat Genet 43(6):595–600
Kohli RM, Zhang Y (2013) TET enzymes, TDG and the dynamics of DNA demethylation. Nature 502(7472): 472–479
Kondo T, Bobek MP, Kuick R, Lamb B, Zhu X, Narayan A, Bourc'his D, Viegas-Péquignot E, Ehrlich M, Hanash SM (2000) Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats D4Z4 and NBL2. Hum Mol Genet 9(4):597–604
Kong Q, Yu M, Zhang M, Wei C, Gu H, Yu S, Sun W, Li N, Zhou Y (2020) Conditional Dnmt3b deletion in hippocampal dCA1 impairs recognition memory. Mol Brain 13(1):1–4
Kongkham P, Northcott P, Croul S, Smith C, Taylor M, Rutka J (2010) The SFRP family of WNT inhibitors function as novel tumor suppressor genes epigenetically silenced in medulloblastoma. Oncogene 29(20): 3017–3024
Korolev IO, Symonds LL, Bozoki AC, Herholz K (2016) Alzheimer's Disease Neuroimaging Initiative. Predicting progression from mild cognitive impairment to alzheimer's dementia using clinical, MRI, and plasma biomarkers via probabilistic pattern classification. PLoS One 11:e0138866
Kriegstein A, Alvarez-Buylla A (2009) The glial nature of embryonic and adult neural stem cells. Annu Rev Neurosci 32:149–184
Krienen FM, Goldman M, Zhang Q, Del Rosario RC, Florio M, Machold R, Saunders A, Levandowski K, Zaniewski H, Schuman B (2020) Innovations present in the primate interneuron repertoire. Nature 586(7828):262–269
Kumar R, Sanawar R, Li X, Li F (2017) Structure, biochemistry, and biology of PAK kinases. Gene 605: 20–31
La Joie R, Visani AV, Baker SL, Brown JA, Bourakova V, Cha J, Chaudhary K, Edwards L, Iaccarino L, Janabi M (2020) Prospective longitudinal atrophy in Alzheimer's disease correlates with the intensity and topography of baseline tau-PET. Sci Transl Med 12(524)
Lane C, Tatton-Brown K, Freeth M (2020) Tatton-Brown-Rahman syndrome: cognitive and behavioural phenotypes. Dev Med Child Neurol 62(8):993–998
Larsen C (2010) Genetic and molecular abnormalities of glioblastomas (GBM). Bull Cancer 97(11):1389–1407
Lee J, Hwang YJ, Kim KY, Kowall NW, Ryu H (2013) Epigenetic mechanisms of neurodegeneration in Huntington's disease. Neurotherapeutics 10(4): 664–676
Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol 13 (14):1192–1200. https://doi.org/10.1016/S0960-9822(03)00432-9
Leung AK (2015) The whereabouts of microRNA actions: cytoplasm and beyond. Trends Cell Biol 25(10): 601–610
Levenson JM, Roth TL, Lubin FD, Miller CA, Huang I-C, Desai P, Malone LM, Sweatt JD (2006) Evidence that DNA (cytosine-5) methyltransferase regulates synaptic plasticity in the hippocampus. J Biol Chem 281(23): 15763–15773
Li M, Li J, Liu L, Li W, Yang Y, Yuan J (2013) MicroRNA in human glioma. Cancers 5(4):1306–1331
Li L, Zhang C, Zi X, Tu Q, Guo K (2015) Epigenetic modulation of Cdk5 contributes to memory deficiency induced by amyloid fibrils. Exp Brain Res 233(1): 165–173
Li Q, Dong C, Cui J, Wang Y, Hong X (2018) Over-expressed lncRNA HOTAIRM1 promotes tumor growth and invasion through up-regulating HOXA1 and sequestering G9a/EZH2/Dnmts away from the HOXA1 gene in glioblastoma multiforme. J Exp Clin Cancer Res 37(1):1–15
Lim L, Mi D, Llorca A, Marín O (2018) Development and functional diversification of cortical interneurons. Neuron 100(2):294–313
Linde J, Zimmer-Bensch G (2020) DNA methylation-dependent dysregulation of GABAergic interneuron functionality in neuropsychiatric diseases. Front Neurosci 14:982
Lindsey JC, Lusher ME, Anderton JA, Bailey S, Gilbertson RJ, Pearson AD, Ellison DW, Clifford SC (2004) Identification of tumour-specific epigenetic events in medulloblastoma development by hypermethylation profiling. Carcinogenesis 25(5): 661–668
Lisman JE (1985) A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase. Proc Natl Acad Sci USA 82(9):3055–3057
Lister R, Mukamel EA (2015) Turning over DNA methylation in the mind. Front Neurosci 9:252
Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD (2013) Global epigenomic reconfiguration during mammalian brain development. Science 341(6146)
Liu L, van Groen T, Kadish I, Tollefsbol TO (2009) DNA methylation impacts on learning and memory in aging. Neurobiol Aging 30(4):549–560
Liu X, Li C, Zhang R, Xiao W, Niu X, Ye X, Li Z, Guo Y, Tan J, Li Y (2018) The EZH2-H3K27me3-DNMT1 complex orchestrates epigenetic silencing of the wwc1 gene, a Hippo/YAP pathway upstream effector, in breast cancer epithelial cells. Cell Signal 51:243–256. https://doi.org/10.1016/j.cellsig.2018.08.011

Liu Y, Dou M, Song X, Dong Y, Liu S, Liu H, Tao J, Li W, Yin X, Xu W (2019) The emerging role of the piRNA/piwi complex in cancer. Mol Cancer 18(1): 1–15

Lodato S, Rouaux C, Quast KB, Jantrachotechatchawan C, Studer M, Hensch TK, Arlotta P (2011) Excitatory projection neuron subtypes control the distribution of local inhibitory interneurons in the cerebral cortex. Neuron 69(4):763–779

López-Bendito G, Sturgess K, Erdélyi F, Szabó G, Molnár Z, Paulsen O (2004) Preferential origin and layer destination of GAD65-GFP cortical interneurons. Cereb Cortex 14(10):1122–1133

López-otín C, Blasco M, Partridge L, Serrano M, Kroemer G (2013) The hallmarks of aging longevity. Cell 153: 1194–1217

Lovrečić L, Maver A, Zadel M, Peterlin B (2013) The role of epigenetics in neurodegenerative diseases. Neurodegener Dis IntechOpen 345–365

Lowel S, Singer W (1992) Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. Science 255(5041):209–212

Lu C, Wei Y, Wang X, Zhang Z, Yin J, Li W, Chen L, Lyu X, Shi Z, Yan W (2020) DNA-methylation-mediated activating of lncRNA SNHG12 promotes temozolomide resistance in glioblastoma. Mol Cancer 19(1):1–19

MacDonald JL, Gin CS, Roskams AJ (2005) Stage-specific induction of DNA methyltransferases in olfactory receptor neuron development. Dev Biol 288(2): 461–473

Margueron R, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. Nature 469(7330):343–349

Marie SKN, Shinjo SMO (2011) Metabolism and brain cancer. Clinics 66:33–43

Marín O, Plump AS, Flames N, Sánchez-Camacho C, Tessier-Lavigne M, Rubenstein JL (2003) Directional guidance of interneuron migration to the cerebral cortex relies on subcortical Slit1/2-independent repulsion and cortical attraction. Development 130(9): 1889–1901

Marín O, Valiente M, Ge X, Tsai L-H (2010) Guiding neuronal cell migrations. Cold Spring Harb Perspect Biol 2(2):a001834

Martin KC, Casadio A, Zhu H, Yaping E, Rose JC, Chen M, Bailey CH, Kandel ER (1997) Synapse-specific, long-term facilitation of aplysia sensory to motor synapses: a function for local protein synthesis in memory storage. Cell 91(7):927–938

Martynoga B, Drechsel D, Guillemot F (2012) Molecular control of neurogenesis: a view from the mammalian cerebral cortex. Cold Spring Harb Perspect Biol 4(10): a008359

Matrisciano F, Tueting P, Dalal I, Kadriu B, Grayson DR, Davis JM, Nicoletti F, Guidotti A (2013) Epigenetic modifications of GABAergic interneurons are associated with the schizophrenia-like phenotype induced by prenatal stress in mice. Neuropharmacology 68:184–194

McKinney BC, Lin C-W, Rahman T, Oh H, Lewis DA, Tseng G, Sibille E (2019) DNA methylation in the human frontal cortex reveals a putative mechanism for age-by-disease interactions. Transl Psychiatry 9(1):1–10

Merry CR, Forrest ME, Sabers JN, Beard L, Gao X-H, Hatzoglou M, Jackson MW, Wang Z, Markowitz SD, Khalil AM (2015) DNMT1-associated long non-coding RNAs regulate global gene expression and DNA methylation in colon cancer. Hum Mol Genet 24(21):6240–6253

Miller CA, Sweatt JD (2007) Covalent modification of DNA regulates memory formation. Neuron 53(6): 857–869

Miller CA, Gavin CF, White JA, Parrish RR, Honasoge A, Yancey CR, Rivera IM, Rubio MD, Rumbaugh G, Sweatt JD (2010) Cortical DNA methylation maintains remote memory. Nat Neurosci 13(6):664–666

Ming GL, Song H (2011) Adult neurogenesis in the mammalian brain: significant answers and significant questions. Neuron 70(4):687–702

Miniou P, Jeanpierre M, Bourc'his D, Barbosa ACC, Blanquet V, Viegas-Péquignot E (1997) α-Satellite DNA methylation in normal individuals and in ICF patients: heterogeneous methylation of constitutive heterochromatin in adult and fetal tissues. Hum Genet 99(6):738–745

Miyoshi G, Young A, Petros T, Karayannis T, Chang MM, Lavado A, Iwano T, Nakajima M, Taniguchi H, Huang ZJ (2015) Prox1 regulates the subtype-specific development of caudal ganglionic eminence-derived GABAergic cortical interneurons. J Neurosci 35(37): 12869–12889

Mizuno K, Dempster E, Mill J, Giese K (2012) Long-lasting regulation of hippocampal Bdnf gene transcription after contextual fear conditioning. Genes Brain Behav 11(6):651–659

Mo A, Mukamel EA, Davis FP, Luo C, Henry GL, Picard S, Urich MA, Nery JR, Sejnowski TJ, Lister R (2015) Epigenomic signatures of neuronal diversity in the mammalian brain. Neuron 86(6):1369–1384

Monsey MS, Ota KT, Akingbade IF, Hong ES, Schafe GE (2011) Epigenetic alterations are critical for fear memory consolidation and synaptic plasticity in the lateral amygdala. PLoS One 6(5):e19958

Moore LD, Le T, Fan G (2013) DNA methylation and its basic function. Neuropsychopharmacology 38(1): 23–38

Morrell F (1991) Introduction: the legacy of Graham Goddard. In: Kindling and synaptic plasticity. Birkhauser, Boston

Morris MJ, Monteggia LM (2014) Role of DNA methylation and the DNA methyltransferases in learning and memory. Dialogues Clin Neurosci 16(3):359

Morris MJ, Adachi M, Na ES, Monteggia LM (2014) Selective role for DNMT3a in learning and memory. Neurobiol Learn Mem 115:30–37

Morris MJ, Na ES, Autry AE, Monteggia LM (2016) Impact of DNMT1 and DNMT3a forebrain knockout

on depressive-and anxiety like behavior in mice. Neurobiol Learn Mem 135:139–145
Moyon S, Huynh JL, Dutta D, Zhang F, Ma D, Yoo S, Lawrence R, Wegner M, John GR, Emery B (2016) Functional characterization of DNA methylation in the oligodendrocyte lineage. Cell Rep 15(4):748–760
Mukhtar T, Taylor V (2018) Untangling cortical complexity during development. J Exp Neurosci 12: 1179069518759332
Muñoz P, Estay C, Díaz P, Elgueta C, Ardiles ÁO, Lizana PA (2016) Inhibition of DNA methylation impairs synaptic plasticity during an early time window in rats. Neural Plast. https://doi.org/10.1155/2016/4783836
Murao N, Noguchi H, Nakashima K (2016) Epigenetic regulation of neural stem cell property from embryo to adult. Neuroepigenetics 5:1–10
Murley AG, Coyle-Gilchrist I, Rouse MA, Jones PS, Li W, Wiggins J, Lansdall C, Rodríguez PV, Wilcox A, Tsvetanov KA (2020) Redefining the multidimensional clinical phenotypes of frontotemporal lobar degeneration syndromes. Brain 143(5):1555–1571
Nakazawa K, Zsiros V, Jiang Z, Nakao K, Kolata S, Zhang S, Belforte JE (2012) GABAergic interneuron origin of schizophrenia pathophysiology. Neuropharmacology 62(3):1574–1583
Namihira M, Kohyama J, Semi K, Sanosaka T, Deneen B, Taga T, Nakashima K (2009) Committed neuronal precursors confer astrocytic potential on residual neural precursor cells. Dev Cell 16(2):245–255
Nery S, Fishell G, Corbin JG (2002) The caudal ganglionic eminence is a source of distinct cortical and subcortical cell populations. Nat Neurosci 5(12):1279–1287
Ng H-H, Jeppesen P, Bird A (2000) Active repression of methylated genes by the chromosomal protein MBD1. Mol Cell Biol 20(4):1394–1406
Ng CW, Yildirim F, Yap YS, Dalin S, Matthews BJ, Velez PJ, Labadorf A, Housman DE, Fraenkel E (2013) Extensive changes in DNA methylation are associated with expression of mutant huntingtin. Proc Natl Acad Sci U S A 110(6):2354–2359
Nicolia V, Ciraci V, Cavallaro RA, Ferrer I, Scarpa S, Fuso A (2017) GSK3β 5′-flanking DNA methylation and expression in Alzheimer's disease patients. Curr Alzheimer Res 14(7):753–759
Ning X, Shi Z, Liu X, Zhang A, Han L, Jiang K, Kang C, Zhang Q (2015) DNMT1 and EZH2 mediated methylation silences the microRNA-200b/a/429 gene and promotes tumor progression. Cancer Lett 359(2): 198–205
Noguchi H, Kimura A, Murao N, Matsuda T, Namihira M, Nakashima K (2015) Expression of DNMT1 in neural stem/precursor cells is critical for survival of newly generated neurons in the adult hippocampus. Neurosci Res 95:1–11
Noguchi H, Murao N, Kimura A, Matsuda T, Namihira M, Nakashima K (2016) DNA methyltransferase 1 is indispensable for development of the hippocampal dentate gyrus. J Neurosci 36(22):6050–6068
Nolde SF, Johnson MK, D'Esposito M (1998) Left prefrontal activation during episodic remembering: An event-related fMRI study. Neuroreport 9(15): 3509–3514
Oliveira AM, Hemstedt TJ, Bading H (2012) Rescue of aging-associated decline in Dnmt3a2 expression restores cognitive abilities. Nat Neurosci 15(8): 1111–1113
Pan Y, Daito T, Sasaki Y, Chung YH, Xing X, Pondugula S, Swamidass SJ, Wang T, Kim AH, Yano H (2016) Inhibition of DNA methyltransferases blocks mutant huntingtin-induced neurotoxicity. Sci Rep 6:31022
Parsons DW, Jones S, Zhang X, Lin JCH, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu I-M, Gallia GL, Olivi A, Mclendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA Jr, Hartigan J, Smith DR, Strausberg RL, Marie SKN, Shinjo SMO, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW (2008) An integrated genomic analysis of human glioblastoma multiforme. Science 321(5897):1807–1812. https://doi.org/10.1126/science.1164382
Parsons DW, Li M, Zhang X, Jones S, Leary RJ, Lin JC-H, Boca SM, Carter H, Samayoa J, Bettegowda C (2011) The genetic landscape of the childhood cancer medulloblastoma. Science 331(6016):435–439
Pascual-Leone A, Amedi A, Fregni F, Merabet LB (2005) The plastic human brain cortex. Annu Rev Neurosci 28:377–401
Pastuzyn ED, Day CE, Kearns RB, Kyrke-Smith M, Taibi AV, McCormick J, Yoder N, Belnap DM, Erlendsson S, Morado DR (2018) The neuronal gene arc encodes a repurposed retrotransposon gag protein that mediates intercellular RNA transfer. Cell 172(1-2): 275-288. e218
Pearce K, Cai D, Roberts AC, Glanzman DL (2017) Role of protein synthesis and DNA methylation in the consolidation and maintenance of long-term memory in Aplysia. Elife 6:e18299
Penisson M, Ladewig J, Belvindrah R, Francis F (2019) Genes and mechanisms involved in the generation and amplification of basal radial glial cells. Front Cell Neurosci 13:381
Pensold D, Symmank J, Hahn A, Lingner T, Salinas-Riester G, Downie BR, Ludewig F, Rotzsch A, Haag N, Andreas N (2017) The DNA methyltransferase 1 (DNMT1) controls the shape and dynamics of migrating POA-derived interneurons fated for the murine cerebral cortex. Cereb Cortex 27(12): 5696–5714
Pensold D, Reichard J, Van Loo KM, Ciganok N, Hahn A, Bayer C, Liebmann L, Groß J, Tittelmeier J, Lingner T (2020) DNA methylation-mediated modulation of endocytosis as potential mechanism for synaptic function regulation in murine inhibitory cortical interneurons. Cereb Cortex 30(7):3921–3937

Petanjek Z, Berger B, Esclapez M (2009) Origins of cortical GABAergic neurons in the cynomolgus monkey. Cereb Cortex 19(2):249–262
Pócza T, Krenács T, Turányi E, Csáthy J, Jakab Z, Hauser P (2016) High expression of DNA methyltransferases in primary human medulloblastoma. Folia Neuropathol 54(2):105–113
Pritchard JI, Olson JM (2008) Methylation of PTCH1, the Patched-1 gene, in a panel of primary medulloblastomas. Cancer Genet Cytogenet 180(1): 47–50
Purkait S, Sharma V, Kumar A, Pathak P, Mallick S, Jha P, Sharma MC, Suri V, Julka PK, Suri A (2016) Expression of DNA methyltransferases 1 and 3B correlates with EZH2 and this 3-marker epigenetic signature predicts outcome in glioblastomas. Exp Mol Pathol 100(2):312–320
Qureshi IA, Mehler MF (2012) Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. Nat Rev Neurosci 13(8): 528–541
Racine R, Milgram N, Hafner S (1983) Long-term potentiation phenomena in the rat limbic forebrain. Brain Res 260(2):217–231
Raff MC (1992) Social controls on cell survival and cell death. Nature 356(6368):397–400
Rajasethupathy P, Antonov I, Sheridan R, Frey S, Sander C, Tuschl T, Kandel ER (2012) A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. Cell 149(3): 693–707
Rajendran G, Shanmuganandam K, Bendre A, Mujumdar D, Goel A, Shiras A (2011) Epigenetic regulation of DNA methyltransferases: DNMT1 and DNMT3B in gliomas. J Neurooncol 104(2):483–494
Raveendra BL, Swarnkar S, Avchalumov Y, Liu X-A, Grinman E, Badal K, Reich A, Pascal BD, Puthanveettil SV (2018) Long noncoding RNA GM12371 acts as a transcriptional regulator of synapse function. Proc Natl Acad Sci UUUSA 115(43): E10197–E10205
Reichard J, Zimmer-Bensch G (2021) The epigenome in neurodevelopmental disorders. Front Neurosci 1415
Rhee K, Yu J, Zhao C, Fan G, Yang X (2012) Dnmt1-dependent DNA methylation is essential for photoreceptor terminal differentiation and retinal neuron survival. Cell Death Dis 3(11):e427–e427
Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. Annu Rev Biochem 81:145–166
Roberson ED, Scearce-Levie K, Palop JJ, Yan F, Cheng IH, Wu T, Gerstein H, Yu G-Q, Mucke L (2007) Reducing endogenous tau ameliorates amyloid ß-induced deficits in an Alzheimer's disease mouse model. Science 316(5825):750–754
Roberts AC, Glanzman DL (2003) Learning in Aplysia: looking at synaptic plasticity from both sides. Trends Neurosci 26(12):662–670
Rose NR, Klose RJ (2014) Understanding the relationship between DNA methylation and histone lysine methylation. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms 1839(12):1362–1372
Rowbotham DA, Marshall EA, Vucic EA, Kennett JY, Lam WL, Martinez VD (2015) Epigenetic changes in aging and Age-related disease. J Aging Sci 3:130
Rozycka A, Liguz-Lecznar M (2017) The space where aging acts: focus on the GABA ergic synapse. Aging Cell 16(4):634–643
Rudolph J, Zimmer G, Steinecke A, Barchmann S, Bolz J (2010) Ephrins guide migrating cortical interneurons in the basal telencephalon. Cell Adh Migr 4(3):400–408
Rudolph J, Gerstmann K, Zimmer G, Steinecke A, Döding A, Bolz J (2014) A dual role of EphB1/ephrin-B3 reverse signaling on migrating striatal and cortical neurons originating in the preoptic area: should I stay or go away? Front Cell Neurosci 8:185
Salameh Y, Bejaoui Y, El Hajj N (2020) DNA methylation biomarkers in aging and age-related diseases. Front Genet 11:171
Saper CB, Lowell BB (2014) The hypothalamus. Curr Biol 24(23):R1111–R1116
Schuman EM (1997) Synapse specificity and long-term information storage. Neuron 18(3):339–342
Scourzic L, Mouly E, Bernard OA (2015) TET proteins and the control of cytosine demethylation in cancer. Genome Med 7(1):1–16
Selkoe DJ, Hardy J (2016) The amyloid hypothesis of Alzheimer's disease at 25 years. EMBO Mol Med 8(6):595–608
Selkoe DJ, Lansbury PJ Jr (1999) Alzheimer's disease is the most common neurodegenerative disorder. Basic Neurochem Mol Cell Med Aspect 6:101–102
Seong IS, Woda JM, Song J-J, Lloret A, Abeyrathne PD, Woo CJ, Gregory G, Lee J-M, Wheeler VC, Walz T, Kingston RE, Gusella JF, Conlon RA, MacDonald ME (2009) Huntingtin facilitates polycomb repressive complex 2. Hum Mol Genet 19(4):573–583
Sexton-Oates A, MacGregor D, Dodgshun A, Saffery R (2015) The potential for epigenetic analysis of paediatric CNS tumours to improve diagnosis, treatment and prognosis. Ann Oncol 26(7):1314–1324
Sharma S, Kelly TK, Jones PA (2010) Epigenetics in cancer. Carcinogenesis 31(1):27–36
Sharma A, Klein SL, Barboza L, Lodhi N, Toth M (2016) Principles governing DNA methylation during neuronal lineage and subtype specification. J Neurosci 36(5): 1711–1722
Shepherd GM (2004) The synaptic organization of the brain. Oxford University Press, New York
Sherman SM, Guillery RW (2006) Exploring the thalamus and its role in cortical function. MIT Press, Cambridge, MA
Shetty AK, Turner DA (1998) Hippocampal interneurons expressing glutamic acid decarboxylase and calcium-binding proteins decrease with aging in Fischer 344 rats. J Comp Neurol 394(2):252–269
Simmons RK, Stringfellow SA, Glover ME, Wagle AA, Clinton SM (2013) DNA methylation markers in the postnatal developing rat brain. Brain Res 1533:26–36

Smallwood A, Estève P-O, Pradhan S, Carey M (2007) Functional cooperation between HP1 and DNMT1 mediates gene silencing. Genes Dev 21(10): 1169–1178

So A-Y, Jung J-W, Lee S, Kim H-S, Kang K-S (2011) DNA methyltransferase controls stem cell aging by regulating BMI1 and EZH2 through microRNAs. PLoS One 6(5):e19503

Somasundaram S, Forrest ME, Moinova H, Cohen A, Varadan V, LaFramboise T, Markowitz S, Khalil AM (2018) The DNMT1-associated lincRNA DACOR1 reprograms genome-wide DNA methylation in colon cancer. Clin Epigenetics 10(1):1–15

Song J-H, Yu J-T, Tan L (2015) Brain-derived neurotrophic factor in Alzheimer's disease: risk, mechanisms, and therapy. Mol Neurobiol 52(3): 1477–1493

Song C, Xiong Y, Liao W, Meng L, Yang S (2019) Long noncoding RNA ATB participates in the development of renal cell carcinoma by downregulating p53 via binding to DNMT1. J Cell Physiol 234(8): 12910–12917

Southwell DG, Paredes MF, Galvao RP, Jones DL, Froemke RC, Sebe JY, Alfaro-Cervello C, Tang Y, Garcia-Verdugo JM, Rubenstein JL (2012) Intrinsically determined cell death of developing cortical interneurons. Nature 491(7422):109–113

Spiegel AM, Sewal AS, Rapp PR (2014) Epigenetic contributions to cognitive aging: disentangling mindspan and lifespan. Learn Mem 21(10):569–574

Stanley DP, Shetty AK (2004) Aging in the rat hippocampus is associated with widespread reductions in the number of glutamate decarboxylase-67 positive interneurons but not interneuron degeneration. J Neurochem 89(1):204–216

Stewart DJ, Issa J-P, Kurzrock R, Nunez MI, Jelinek J, Hong D, Oki Y, Guo Z, Gupta S, Wistuba II (2009) Decitabine effect on tumor global DNA methylation and other parameters in a phase I trial in refractory solid tumors and lymphomas. Clin Cancer Res 15(11):3881–3888

Subramanian L, Calcagnotto ME, Paredes MF (2020) Cortical malformations: lessons in human brain development. Front Cell Neurosci 13:576

Sultan KT, Shi SH (2018) Generation of diverse cortical inhibitory interneurons. Wiley Interdiscip Rev Dev Biol 7(2):e306

Sun T, Hevner RF (2014) Growth and folding of the mammalian cerebral cortex: from molecules to malformations. Nat Rev Neurosci 15(4):217–232

Sun J, Tian X, Zhang J, Huang Y, Lin X, Chen L, Zhang S (2017) Regulation of human glioma cell apoptosis and invasion by miR-152-3p through targeting DNMT1 and regulating NF2. J Exp Clin Cancer Res 36(1):1–13

Sweatt JD (2016) Dynamic DNA methylation controls glutamate receptor trafficking and synaptic scaling. J Neurochem 137(3):312–330

Symmank J, Bayer C, Schmidt C, Hahn A, Pensold D, Zimmer-Bensch G (2018) DNMT1 modulates interneuron morphology by regulating Pak6 expression through crosstalk with histone modifications. Epigenetics 13(5):536–556

Symmank J, Gölling V, Gerstmann K, Zimmer G (2019) The transcription factor LHX1 regulates the survival and directed migration of POA-derived cortical interneurons. Cereb Cortex 29(4):1644–1658

Symmank J, Bayer C, Reichard J, Pensold D, Zimmer-Bensch G (2020) Neuronal Lhx1 expression is regulated by DNMT1-dependent modulation of histone marks. Epigenetics 15(11):1259–1274

Takizawa T, Nakashima K, Namihira M, Ochiai W, Uemura A, Yanagisawa M, Fujita N, Nakao M, Taga T (2001) DNA methylation is a critical cell-intrinsic determinant of astrocyte differentiation in the fetal brain. Dev Cell 1(6):749–758

Tanaka DH, Nakajima K (2012) GABA ergic interneuron migration and the evolution of the neocortex. Dev Growth Differ 54(3):366–372

Tanila H (2017) The role of BDNF in Alzheimer's disease. Neurobiol Dis 97(Pt B):114–118

Thomas EA (2016) DNA methylation in Huntington's disease: Implications for transgenerational effects. Neurosci Lett 625:34–39

Thompson RF (1986) The neurobiology of learning and memory. Science 233(4767):941–947

Tobin AJ, Signer ER (2000) Huntington's disease: the challenge for cell biologists. Trends Cell Biol 10(12): 531–536

Tohgi H, Utsugisawa K, Nagane Y, Yoshimura M, Genda Y, Ukitsu M (1999a) Reduction with age in methylcytosine in the promoter region− 224∼− 101 of the amyloid precursor protein gene in autopsy human cortex. Mol Brain Res 70(2):288–292

Tohgi H, Utsugisawa K, Nagane Y, Yoshimura M, Ukitsu M, Genda Y (1999b) The methylation status of cytosines in a τ gene promoter region alters with age to downregulate transcriptional activity in human cerebral cortex. Neurosci Lett 275(2):89–92

Tonegawa S, Pignatelli M, Roy DS, Ryan TJ (2015) Memory engram storage and retrieval. Curr Opin Neurobiol 35:101–109

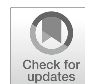
Veldic M, Guidotti A, Maloku E, Davis JM, Costa E (2005) In psychosis, cortical interneurons overexpress DNA-methyltransferase 1. Proc Natl Acad Sci U S A 102(6):2152–2157

Vila M, Przedborski S (2003) Targeting programmed cell death in neurodegenerative diseases. Nat Rev Neurosci 4(5):365–375

Villar-Menéndez I, Blanch M, Tyebji S, Pereira-Veiga T, Albasanz JL, Martín M, Ferrer I, Pérez-Navarro E, Barrachina M (2013) Increased 5-methylcytosine and decreased 5-hydroxymethylcytosine levels are associated with reduced striatal A2AR levels in Huntington's disease. Neuromol Med 15(2):295–309

Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, Morey L, Van Eynde A, Bernard D, Vanderwinden JM, Bollen M (2006) The Polycomb group protein EZH2 directly controls DNA

methylation. Nature 439(7078):871–874. https://doi.org/10.1016/j.bbagrm.2014.02.007
von Bueren AO, Bacolod MD, Hagel C, Heinimann K, Fedier A, Kordes U, Pietsch T, Koster J, Grotzer MA, Friedman HS (2012) Mismatch repair deficiency: a temozolomide resistance factor in medulloblastoma cell lines that is uncommon in primary medulloblastoma tumours. Br J Cancer 107(8):1399–1408
Wamsley B, Fishell G (2017) Genetic and activity-dependent mechanisms underlying interneuron diversity. Nat Rev Neurosci 18(5):299–309
Wang F, Yang Y, Lin X, Wang JQ, Wu YS, Xie W, Wang D, Zhu S, Liao YQ, Sun Q, Yang YG, Luo HR, Guo C, Han C, Tang TS (2013) Genome-wide loss of 5-hmC is a novel epigenetic feature of Huntington's disease. Hum Mol Genet 22(18):3641–3653
Wang L, Zhao Y, Bao X, Zhu X, Kwok YK-Y, Sun K, Chen X, Huang Y, Jauch R, Esteban MA (2015) LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration. Cell Res 25(3):335–350
Wang F, Wang Q, Liu B, Mei L, Ma S, Wang S, Wang R, Zhang Y, Niu C, Xiong Z (2021) The long noncoding RNA Synage regulates synapse stability and neuronal function in the cerebellum. Cell Death Differ 1–17
Watanabe D, Suetake I, Tada T, Tajima S (2002) Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. Mech Dev 118(1-2):187–190
Wei J-W, Huang K, Yang C, Kang C-S (2017) Non-coding RNAs as regulators in epigenetics. Oncol Rep 37(1):3–9
Wong FK, Marín O (2019) Developmental cell death in the cerebral cortex. Annu Rev Cell Dev Biol 35:523–542
Wood H (2013) Neurodegenerative disease: altered DNA methylation and RNA splicing could be key mechanisms in Huntington disease. Nat Rev Neurol 9(3):119
Wu J, Shuang Z, Zhao J, Tang H, Liu P, Zhang L, Xie X, Xiao X (2018) Linc00152 promotes tumorigenesis by regulating DNMTs in triple-negative breast cancer. Biomed Pharmacother 97:1275–1281
Xiong Y, Kuang W, Lu S, Guo H, Wu M, Ye M, Wu L (2018) Long noncoding RNA HOXB 13-AS 1 regulates HOXB 13 gene methylation by interacting with EZH 2 in glioma. Cancer Med 7(9):4718–4728
Xu G-L, Wong J (2015) Oxidative DNA demethylation mediated by Tet enzymes. Natl Sci Rev 2(3):318–328
Yamaguchi Y, Miura M (2015) Programmed cell death in neurodevelopment. Dev Cell 32(4):478–490
Yokoi T, Enomoto Y, Naruto T, Kurosawa K, Higurashi N (2020) Tatton-Brown-Rahman syndrome with a novel DNMT3A mutation presented severe intellectual disability and autism spectrum disorder. Human genome variation 7(1):1–3
Yu C-C, Jiang T, Yang A-F, Du Y-J, Wu M, Kong L-H (2019) Epigenetic modulation on tau phosphorylation in Alzheimer's disease. Neural Plast. https://doi.org/10.1155/2019/6856327
Zempel H, Luedtke J, Kumar Y, Biernat J, Dawson H, Mandelkow E, Mandelkow EM (2013) Amyloid-β oligomers induce synaptic damage via Tau-dependent microtubule severing by TTLL6 and spastin. EMBO J 32(22):2920–2937
Zhao X, Hu X (2019) Downregulated long noncoding RNA LINC00313 inhibits the epithelial–mesenchymal transition, invasion, and migration of thyroid cancer cells through inhibiting the methylation of ALX4. J Cell Physiol 234(11):20992–21004
Zhao Y, Sun H, Wang H (2016) Long noncoding RNAs in DNA methylation: new players stepping into the old game. Cell Biosci 6(1):1–6
Zhou D, Wan Y, Xie D, Wang Y, Wei J, Yan Q, Lu P, Mo L, Xie J, Yang S (2015) DNMT1 mediates chemosensitivity by reducing methylation of miRNA-20a promoter in glioma cells. Exp Mol Med 47(9):e182
Zimmer G, Kästner B, Weth F, Bolz J (2007) Multiple effects of ephrin-A5 on cortical neurons are mediated by SRC family kinases. J Neurosci 27(21):5643–5653
Zimmer G, Schanuel SM, Bürger S, Weth F, Steinecke A, Bolz J, Lent R (2010) Chondroitin sulfate acts in concert with semaphorin 3A to guide tangential migration of cortical interneurons in the ventral telencephalon. Cereb Cortex 20(10):2411–2422
Zimmer G, Rudolph J, Landmann J, Gerstmann K, Steinecke A, Gampe C, Bolz J (2011) Bidirectional ephrinB3/EphA4 signaling mediates the segregation of medial ganglionic eminence-and preoptic area-derived interneurons in the deep and superficial migratory stream. J Neurosci 31(50):18364–18380
Zimmer-Bensch G (2018) Diverse facets of cortical interneuron migration regulation—implications of neuronal activity and epigenetics. Brain Res 1700:160–169
Zimmer-Bensch G (2019a) Emerging roles of long non-coding RNAs as drivers of brain evolution. Cells 8(11):1399
Zimmer-Bensch G (2019b) Functional implications of dynamic DNA methylation for the developing, aging and diseased brain. In: The DNA, RNA, and histone methylomes. Springer, pp 141–163
Zimmer-Bensch G (2020) Epigenomic remodeling in Huntington's disease—master or servant? Epigenomes 4(3):15
Zimmer-Bensch G, Zempel H (2021) DNA methylation in genetic and sporadic forms of neurodegeneration: lessons from Alzheimer's, related tauopathies and genetic tauopathies. Cells 10(11):3064



16 Current and Emerging Technologies for the Analysis of the Genome-Wide and Locus-Specific DNA Methylation Patterns

Jörg Tost

Abstract

DNA methylation is the most studied epigenetic modification, and altered DNA methylation patterns have been identified in cancer and more recently also in many other complex diseases. Furthermore, DNA methylation is influenced by a variety of environmental factors, and the analysis of DNA methylation patterns might allow deciphering previous exposures. A number of techniques to study DNA methylation either genome-wide or at specific loci have been devised using a limited number of principles for differentiating the methylation state: (1) methylation-sensitive/dependent restriction enzymes, (2) antibody or methyl-binding protein-based enrichment, or (3) chemical or enzymatic conversion, (4) direct sequence readout. Second-generation sequencing has largely replaced microarrays as a readout platform and is also becoming more popular for locus-specific DNA methylation analysis. In this chapter, the currently used methods for both genome-wide and locus-specific analysis of 5-methylcytosine as well as its oxidative derivatives such as 5-hydroxymethylcytosine are reviewed in detail and advantages and limitations of each approach are discussed. Furthermore, emerging technologies avoiding PCR amplification and allowing a direct read-out of DNA methylation are summarized, together with novel applications, such as the detection of DNA methylation in single cells or in circulating cell-free DNA.

Keywords

DNA methylation · NGS · MethylC-Seq · Bisulphite · Nanopore · Cell-free DNA · Enzymatic conversion · Single cell analysis · Methylcytosine · Hydroxymethylcytosine

List of Abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
Aba-seq	AbaSI-coupled sequencing
Ccf DNA	Circulating cell-free DNA
ChIP	Chromatin immunoprecipitation
CMS	Cytosinemethylenesulfonate
COBRA	Combined bisulfite restriction analysis
COLD	Coamplification at lower denaturation temperature
ddPCR	Digital droplet PCR

J. Tost (✉)
Laboratory for Epigenetics and Environment, Centre National de Recherche en Génomique Humaine, CEA-Institut de Biologie François Jacob, Université Paris-Saclay, Evry, France
e-mail: tost@cnrgh.fr

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_16

DREAM	Digital restriction enzyme analysis of methylation
FFPE	Formalin-fixed paraffin embedded
GLIB	Glucosylation, periodate oxidation, biotinylation
hMeSEAL	5hmC selective chemical labeling
M	Million
MALDI-MS	Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
MBD	Methyl-binding domain
MeDIP	Methylated DNA immunoprecipitation
MIRA	Methylated-CpG island recovery assay
MRE	Methylation-sensitive restriction enzyme
MS	Methylation-sensitive
MSCC	Methylation-sensitive cut counting
MS-FLAG	Methylation-sensitive fluorescent amplicon generation
MS-HRM	Methylation-sensitive high-resolution melting analysis
MS-MLPA	Methylation-specific multiplexed ligation probe amplification
MSP	Methylation-specific PCR
MS-SNuPE	Methylation-sensitive single-nucleotide primer extension
NGS	Next-/second-generation sequencing
OxBS	Oxidative bisulfite
PBAT	Post-adaptor bisulfite tagging
PBMC	Peripheral blood mononuclear cells
QAMA	Quantitative analysis of methylated alleles
RRBS	Reduced representation bisulfite sequencing
RRHP	Reduced representation 5-hydroxymethylcytosine profiling
SBS	Sequencing by synthesis
SCAN	Single chromatin molecule analysis at the nanoscale
SMART-MSP	Sensitive melting analysis after real-time methylation-specific PCR
SMRT	Single molecule real time
TAB-seq	TET-assisted bisulfite sequencing
TET	Ten-eleven translocation (enzyme)

16.1 Introduction

Epigenetic phenomena are mediated by a variety of molecular mechanisms including post-translational histone modifications, histone variants, ATP-dependent chromatin remodeling, small and other non-coding RNAs, and DNA methylation (Tost 2008a). These diverse molecular mechanisms are all closely intertwined and stabilize each other to ensure the faithful propagation of epigenetic states over time and especially through cell division. Epigenetics and the analysis of epigenetic modifications have come to a central stage for many developmental and biomedical questions and the advent of second-/next-generation sequencing (NGS) has revolutionized the way of interrogating the epigenome.

In this chapter, I will provide a comprehensive overview of the technologies for the genome-wide and locus-specific analysis of DNA methylation, the best-studied epigenetic modification. As the description of all methods used for the different applications of DNA methylation analysis is beyond the scope of this chapter, I will concentrate on the most widely used methods, powerful emerging concepts as well as approaches that are or have the potential to be used in the clinics.

16.1.1 DNA Methylation

In mammals, DNA methylation is the most prevalent DNA modification and is almost entirely found on the fifth position of the pyrimidine ring of cytosines in the context of CpG dinucleotides (Bird 2002). 5-methylcytosine (5mC) accounts for ~1% of all bases whereby the majority (75%) of CpG dinucleotides throughout mammalian genomes is methylated. Methylation of cytosines in the context of CpHpG or CpHpH sequences (where H is A, C, or T) has been detected in embryonic stem cells, brain, and plants, but is rarely found in other somatic mammalian/human tissues. DNA methylation can have profound effects on gene expression, is crucial for proper development, and is implicated in

disease processes, particularly tumorigenesis (Baylin and Jones 2011), but more recently also connections to other complex diseases, including autoimmune, allergic, and inflammatory diseases, neurodegenerative, psychiatric, and metabolic disorders have been discovered (Rydbirk et al. 2020; Kabesch and Tost 2020; Fraszczyk et al. 2022; de la Calle-Fabregat et al. 2022; McCartney et al. 2022; Miceli-Richard et al. 2016; Hop et al. 2022; Fogel et al. 2017; Marshall et al. 2020). Online repositories list hundreds of thousands of associations between DNA methylation and diseases or phenotypic traits (Xiong et al. 2022; Huang et al. 2021; Battram et al. 2022). DNA methylation might act as memory of both internal and external environmental influences to which the cells of an organism have been exposed. The variety of epigenetic modifications in mammalian DNA has further increased with the discovery of 5-hydroxymethylcytosine (5hmC) and its oxidative derivatives, 5-formyl (5fC) and 5-carboxylcytosine (5caC). These modifications are formed from 5-methylcytosines by a catalytic oxidation mediated by the Ten-eleven translocation (TET) proteins and appear during DNA demethylation (Kriaucionis and Heintz 2009; Tahiliani et al. 2009). 5hmC is, however, not only an intermediate of the DNA demethylation process, but has also distinct regulatory functions. 5hmC levels change dynamically during embryonic development and are important for cell lineage specification (Wu et al. 2018; Lopez et al. 2017). 5hmC is depleted from active promoters, which normally show an overall low level of methylation, but enriched in distant regulatory elements such as enhancers and transcribed regions/gene bodies (Skvortsova et al. 2017). 5hmC has been detected in nearly all mouse embryonic tissues, but with the exception of brain tissue and bone marrow only very low levels (0.2–0.05% of all cytosines) are found in adult tissue (Ruzov et al. 2011; Globisch et al. 2010; Wagner et al. 2015).

As a covalent DNA-based modification, which is technically relatively easy to investigate, DNA methylation has been intensively studied since the 1980s and is the best-studied epigenetic mark. For the analysis of DNA methylation patterns sensitive and quantitative methods are required to detect even subtle changes in the degree of methylation, as biological samples often represent a heterogeneous mixture of different cells, for example tumor and non-tumor cells from tissue biopsies or DNA extracted from peripheral blood mononuclear cells (PBMCs). Realizing the importance of epigenetic changes in development and disease, a large number of technologies for the study of DNA methylation have been developed in the last years. No single method has emerged as the "gold" standard technique unifying cost-efficiency, throughput, quantitative accuracy and sensitivity, possibility for whole-genome analysis, and precise investigation of individual CpG positions. Therefore, the choice of the method mainly depends on the nature of the fundamental or biomedical research problem and the required answer to be provided by the analysis. A major advantage of DNA methylation analysis is that it can be carried out on DNA isolated from nearly all biological tissues or body fluids as DNA methylation marks seem to be stable even under prolonged storage conditions. It has, for example, been shown that DNA from the blood spots of Guthrie cards can be reliably extracted for up to 20 years with a quality suitable for genome-wide DNA methylation analyses (Ghantous et al. 2014). Archived specimens fixed in a variety of fixatives are also suitable for DNA methylation analysis, but to a different degree. Despite suffering from large damage and modification of the nucleic acids (Srinivasan et al. 2002), DNA from *F*ormalin-*F*ixed *P*araffin-*E*mbedded (FFPE) samples can be used for locus-specific analyses of DNA methylation patterns using, for example, pyrosequencing (Leong et al. 2013) or after a ligation and amplification-based repair step for genome-wide analysis using the Infinium BeadChips (Thirlwell et al. 2010). However, alternative fixatives such as PaxGene based preservation proved superior in terms of accuracy and reproducibility compared to FFPE specimens (Andersen et al. 2014).

Although DNA methylation is technically easier to handle and the requirements for pre-analytical sample processing are less stringent compared to the analysis of RNA and protein

modifications, several challenges are nonetheless associated with the analysis of DNA methylation patterns in tissues. Biological specimens are composed of a large number of different cell types, each associated with its own DNA methylome and appropriate control of the cellular composition needs to be carried out for complex material using either careful pathological examination, complementary molecular methods, e.g., for assessing the tumor percentage (Van Loo et al. 2010) or statistical methods, e.g., for blood cell composition (Houseman et al. 2012). These points are, of course, also valid for the analysis of RNA or protein-based biomarkers. Heterogeneity at the level of DNA methylation, which is transformed into polymorphisms of the target sequence through bisulfite conversion, might also be problematic for some of the technologies used for the assessment of DNA methylation levels in a candidate region, when primers fail to amplify, e.g., partially methylated molecules (Mikeska et al. 2010; Alnaes et al. 2015). Furthermore, the exact location for DNA methylation analysis might not have been fully optimized to identify positions with the largest DNA methylation difference (Koch et al. 2018). Another yet unanswered question is under which conditions and to which extent accessible tissues such as blood, urine, or saliva can be used as surrogates for the target organ, if the primary disease organ is not available for analysis. All these reasons contribute to the current situation that despite a large amount of literature and studies performed, currently few DNA methylation-based biomarkers are used in clinical diagnostic assays (Beltran-Garcia et al. 2019; Taryma-Lesniak et al. 2020).

16.2 Principles of DNA Methylation Detection

Several generations of genome-wide methods for DNA methylation analysis have been developed adapting to different analytical supports with increasing levels of resolution and coverage. Except for the highly popular epigenotyping arrays, NGS-based technologies have largely replaced microarrays as the readout platform for DNA methylation analysis and the latter are therefore not covered in this review. Historical methods for DNA methylation analysis have been reviewed in Tost (2008b) and Harrison and Parle-McDermott (2011).

The main approaches for the discrimination of methylation are based on four main principles:

1. **Enzymatic methods** including notably methylation-sensitive restriction endonucleases, i.e., enzymes that are blocked by methylated cytosines in their recognition sequence (Bird and Southern 1978) are widely used for the analysis of methylation patterns in combination with their methylation-insensitive isoschizomers. Although methods based on methylation-sensitive restriction enzymes are simple and cost-effective and might be able to distinguish between methylcytosine and its oxidative derivatives, they are hampered by the limitation to specific restriction sites, as only CpG sites found within the recognition sequences of restriction enzymes can be analyzed (Fazzari and Greally 2004). Information complementary to that obtained by methylation-sensitive restriction digests can be obtained by the methylation-dependent restriction enzymes such as McrBC (Stewart et al. 2000).
2. **Antibody and protein-based enrichment:** The methylated fraction of a genome can be enriched by precipitation with a bead-immobilized antibody specific for 5-methylcytosine following a protocol similar to chromatin immunoprecipitation (ChIP) and analyzed on microarrays (Weber et al. 2005) or by sequencing (Down et al. 2008; Feber et al. 2011). Similar results can be obtained by affinity purification of methylated DNA with methyl-binding domain (MBD) proteins such as MeCP2 (Brinkman et al. 2010) or MBD2 in combination with MBD3L1 (Rauch and Pfeifer 2005).
3. **Chemical or enzymatic conversion**: The most widely used approach consists of the chemical modification of genomic DNA with sodium bisulfite. This chemical reaction

induces hydrolytic deamination of non-methylated cytosines to uracils, while methylated cytosines are resistant to the conversion under the chosen reaction conditions (Frommer et al. 1992; Shapiro et al. 1974). This method translates the methylation signal into a sequence difference. After performing PCR, the methylation status at a given position is manifested in the ratio of C (former methylated cytosine) to T (former non-methylated cytosine) and can be analyzed as a virtual C/T polymorphism spanning the entire allele frequency spectrum from 0-100% in the bisulfite-treated DNA. A number of commercial kits have been developed which allow the user-friendly conversion of genomic DNA from various sources. Bisulfite conversion can also be directly applied to suspended cell pellets with similar technical performance (Verhoef et al. 2021). Nonetheless, the chemical treatment degrades a significant amount of the input DNA leading to the loss of a substantial amount of the starting material (Holmes et al. 2014). Furthermore, dependent on DNA quantity, quality, and conversion protocol incomplete or overconversion of DNA can lead to specific artifacts (Genereux et al. 2008). Specific quality control assays have been developed, which allow for the accurate quantification of total, amplifiable converted and unconverted DNA after bisulfite treatment (Campan et al. 2009). Furthermore, the difference in GC content for different molecules induced by the bisulfite conversion and depending on their former DNA methylation status might influence in a DNA polymerase dependent manner their amplification efficiency and optimal annealing temperature. Therefore, the preferable amplification of unmethylated or methylated DNA is a common complication for methods including PCR amplification of bisulfite-treated DNA (Warnecke et al. 1997; Grunau et al. 2001). Several methods have been devised for the potential correction of the amplification bias, including single-molecule PCR (Chhibber and Schroeder 2008), addition of betaine to the PCR (Voss et al. 1998), inclusion of CpGs in the sequence of the amplification primers (Wojdacz and Hansen 2006), and/or variation of the annealing temperature (Shen et al. 2007). Nonetheless all these methods cannot be generally applied, but have to be carefully tested using standards with a known degree of DNA methylation at the specific locus of interest. Another important problem might arise from clonal amplification during the PCR amplification following bisulfite conversion leading to an ostensible faultless readout but which is not representing the true distribution of DNA methylation in the original sample (Zhang and Jeltsch 2010). Technical replicates of the bisulfite conversion or at least the PCR amplification or the addition of molecule-specific barcodes during the first strand synthesis can help to detect such problems (Zhang and Jeltsch 2010; Miner et al. 2004). It should also be noted that standard bisulfite conversion protocols cannot discriminate between 5-hydroxymethylcytosine and 5-methylcytosine, which are converted with a similar efficiency. However, as in most tissues the relative abundance of 5hmC is much lower than the abundance of 5mC, bisulfite conversion remains a valid approach for many research questions. Specialized protocols such as oxidative bisulfite sequencing (Booth et al. 2012) or TAB-seq (Yu et al. 2012) are required to avoid any potential influence of concomitant hydroxymethylation on accurate quantification of DNA methylation. Methods for the specific analysis of 5-hydroxymethylation are detailed in Sect. 16.9 of this chapter.

To avoid the strong degradation induced by the bisulfite conversion, alternative protocols using an enzymatic conversion strategy including EM-seq and TAPS-seq have been developed (Vaisvila et al. 2021; Liu et al. 2019).

4. The **direct readout** of DNA methylation patterns can be achieved using nanopore or zeptowell sequencing approaches (Song et al. 2012; Flusberg et al. 2010; Clarke et al. 2009).

For Nanopore sequencing, a single-stranded DNA molecule is passed through a pore and the different physical properties of the bases in the DNA lead to changes in the electrical current as a short sequence stretch (k-mer) pass the pore. As the change in current is DNA sequence specific, it allows reconstitution of the DNA sequence. As methylated cytosines within a k-mer induce a different change compared to unmodified cytosines, Nanopore sequencing can be used for the direct readout of the DNA methylation patterns (Simpson et al. 2017).

Depending on the requirements for resolution, coverage, quantification, and throughput, these four main assay principles have been combined with PCR, microarray, or sequencing-based readout technologies. Nowadays, NGS methods have largely replaced microarray-based readouts, because they allow a truly genome-wide discovery of differentially methylated CpGs and/or regions, they require in most cases a lower starting amount of DNA and limit DNA amplification during the analysis, thereby reducing potential amplification biases, and typically provide a more quantitative readout.

16.3 Global Methylation Content of a Sample

Methods for the analysis of global DNA methylation levels in a sample determine the overall 5-methylcytosine content or changes affecting the entire epigenome, respectively. They do, however, not provide any information about the static or dynamic patterns of DNA methylation in the genome. 5-methylcytosine can be differentiated from its unmethylated counterpart by the different mass or polarity of the two bases (cytosine and 5-methylcytosine), which can be used for chromatographic or mass spectrometric separation (Wiebers 1976; Eick et al. 1983; Fraga et al. 2002; Berdasco et al. 2009). The 5-methylcytosine content is measured after hydrolysis to mononucleosides, and comparison to an internal standard enables quantification. Particularly since the discovery of the oxidative derivatives of 5-methylcytosine, the mass spectrometry detection has become very popular, which is also due to its exquisite sensitivity and the low amounts of required starting material (Globisch et al. 2010; Le et al. 2011). Less accurate quantification of the global methylation content can also be made with a variety of ELISA-based assays (Kremer et al. 2012). A number of commercial kits are available, but normally they are much less sensitive than mass spectrometric or chromatographic methods enabling only the detection of large changes in the DNA methylation content.

Bacterial methyltransferases, e.g., *Sss*I, transfer a methyl group from the universal methyl donor *S*-adenosyl-L-methionine to unmethylated CpG positions. The methyl acceptor assay makes use of these enzymes and analyzes the amount of incorporated radioactively labeled methyl groups into a sample (Bestor and Ingram 1983). The measured amount of radioactive label of a given sample correlates thus inversely with the degree of its DNA methylation prior to labeling. Similarly, the cytosine extension assay combines methylation-sensitive restriction digestion and single-nucleotide extension with radio- or fluorescently labeled dCTP complementary to the guanine 5′ overhang created by the digestion (Pogribny et al. 1999; Bönsch et al. 2004). The pyrosequencing-based Luminometric Methylation Assay (LUMA) is based on the differential digestion of a sample with a methylation-sensitive endonuclease or its methylation-insensitive isoschizomer and the successive dispensation of four nucleotides complementary to the overhang created by the endonucleases (Karimi et al. 2006). The pyrosequencing-based analysis of repetitive elements such as Alu and LINE1 elements has also been widely used as surrogate for the global DNA methylation level (Yang et al. 2004; Choi et al. 2009) and is often used as a complement to BeadChip-based epigenome-wide association studies (EWAS) (Rousseaux et al. 2020; Jedynak et al. 2022). Methods analyzing the total amount of 5-methylcytosine in a sample are used to analyze and follow global DNA methylation changes

induced by demethylating pharmaceutical agents in patients with hematological malignancies at various time points of treatment (Mund et al. 2005; Liu et al. 2007; Kantarjian et al. 2006), to investigate the efficacy of novel demethylating agents (Balch et al. 2005), to detect and predict the outcome of various human cancers (Hur et al. 2014; Inamura et al. 2014) as well as to detect and quantify the effect of environmental exposure on the DNA methylome (Bollati et al. 2007; Marques-Rocha et al. 2016).

In situ hybridization methods with antibodies directed against 5-methylcytosine or its oxidative derivatives allow the measurement of the methylation content and its potentially cell-type specific distribution. They can be used to visualize the modified bases in the cell nuclear context (Miller et al. 1974; Rougier et al. 1998; Salvaing et al. 2015). As only clustered methylated CpGs prevalent in, for example, repeat elements can be recognized at the chromosomal level, methylation patterns at relatively small loci such as CpG islands contribute little to the overall staining profile.

16.4 Whole Methylome Analyses

Although potentially confounded by the presence of 5-hydroxymethylcytosine, whole-genome bisulfite sequencing or MethylC-seq can be considered as the current gold standard for the genome-wide identification of differentially methylated regions at single-nucleotide resolution. This technology is currently used in a number of international large-scale projects to map the methylomes of various human tissues and cell types (Adams et al. 2012; Roadmap Epigenomics et al. 2015; Schultz et al. 2015; Bujold et al. 2016). It overcomes the limitations of cloning and Sanger sequencing, a low-throughput method limited to a small number of loci of interest, in which the quantitative resolution was limited by the number of clones analyzed (in most studies < 20). Furthermore, whole-genome bisulfite sequencing avoids problems with the primer design that often introduces multiple biases (Grunau et al. 2001; Warnecke et al. 2002). However, this unprecedented quantitative and spatial resolution that is currently transforming DNA methylation analysis comes at a high cost, requires substantial sequencing depth to obtain a proper and even coverage and necessitates profound bioinformatic expertise and resources. Although low-coverage bisulfite sequencing can yield some information about global DNA methylation alterations, it does not yield reliable locus-specific information (Popp et al. 2010). For the identification of differentially methylated regions, a sequencing coverage of 5–10× is recommended, while for association analysis based on single CpG, a coverage of at least 20× is required (Ziller et al. 2015), while the IHEC consortium even recommends a coverage of at least 30×. Beyond this level of coverage, resources should be rather dedicated to sequence more biological replicates rather than obtain a higher coverage—at least for samples with high cellular homogeneity. The Roadmap consortium therefore suggested the sequencing of two biological replicated with a combined coverage of 30×. However, even at 30× coverage, only half of differentially methylated CpGs can be detected as assessed by saturation analysis (Libertini et al. 2016).

The most widely used whole-genome bisulfite sequencing protocol (Fig. 16.1) consists of the fragmentation of genomic DNA, adapter ligation, bisulfite conversion, and limited amplification using adapter-specific PCR primers. While initially several micrograms of DNA were required to perform whole-genome bisulfite sequencing, the replacement of electrophoresis and gel extraction steps by magnetic beads has enabled to create libraries suitable for sequencing from ng of input material (Urich et al. 2015). Libraries have been reported to be constructed from even less input material, but require in most cases a high number of PCR cycles (up to 25 cycles (Kobayashi et al. 2012)), inducing potentially a large bias for the estimation of the DNA methylation levels and a much more substantial sequencing effort to obtain a homogeneous and sufficient coverage. To assess the bisulfite conversion efficiency, DNA of the bacteriophage λ, which contains only unmethylated

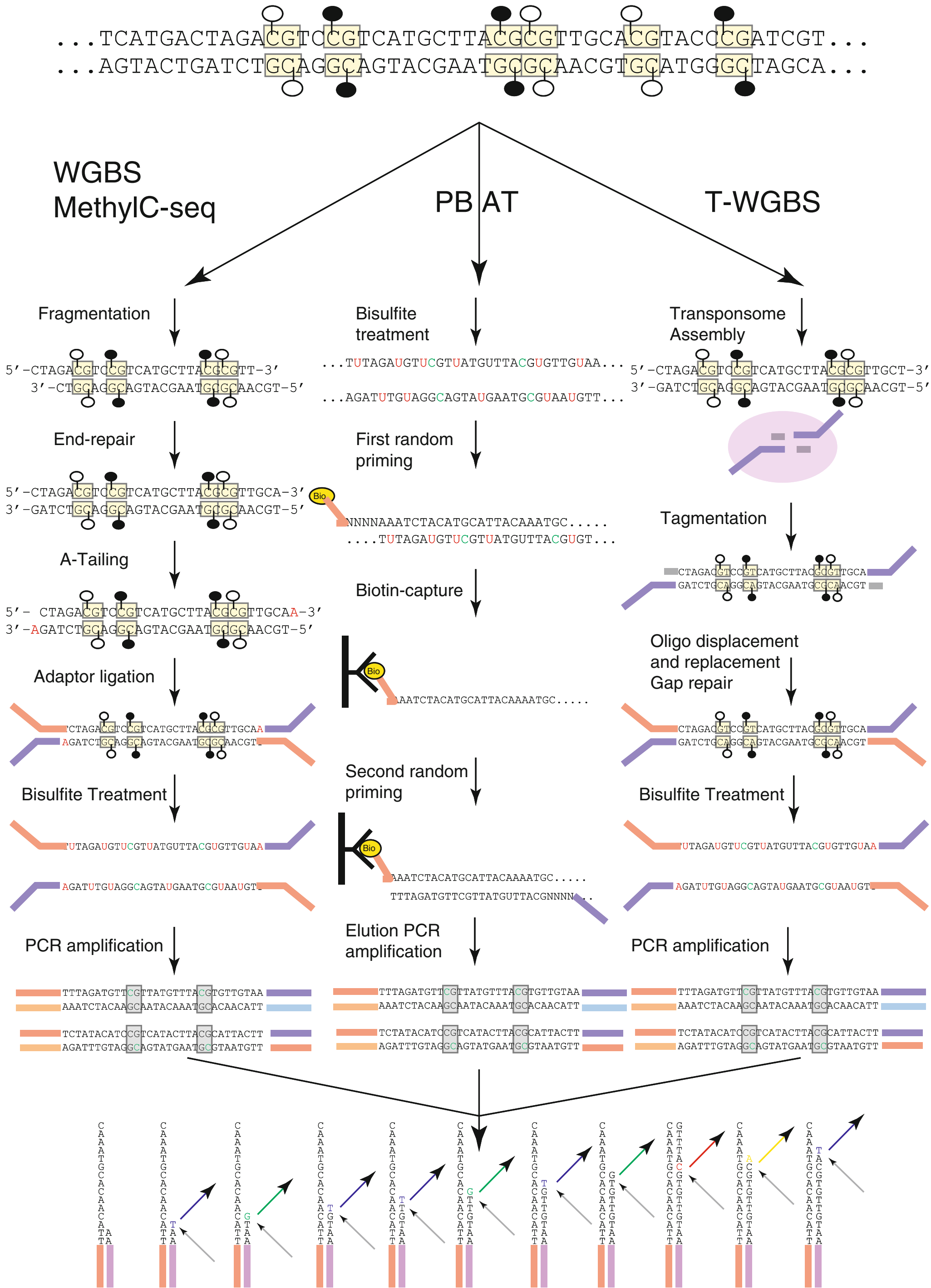


Fig. 16.1 Schematic outline of the three most commonly used approaches for whole-genome bisulfite sequencing (from left to right): MethylC-seq, PBAT, and tagmentation-based WGBS. Details of the procedures are given in the text

cytosines, is spiked in the reaction. Mapping the reads against the bisulfite-converted genome of the phage and counting any remaining cytosines allows to identify problems of bisulfite conversion and estimate the conversion rates. A large number of programs have been developed to perform the quality control, pre-processing steps (such as adaptor, barcode, and quality score trimming), mapping the reads to a bisulfite-converted reference genome, scoring of DNA methylation levels (count statistics), and identification of differentially methylated CpG positions (DMCs) and regions (DMRs) (Adusumalli et al. 2015). This method has been widely used for the methylome analysis of a large number of organisms and plants, as well as human tissues, and ~90–95% of the cytosines present in the genome are routinely covered at a 30× sequencing coverage (Lister et al. 2009; Lister et al. 2008, 2011; Li et al. 2010a; Chalhoub et al. 2014; Lyko et al. 2010; Guo et al. 2014).

As the Watson and Crick strands of the DNA are no longer complementary after bisulfite conversion, after PCR amplification and the synthesis of the DNA complementary to either the Watson or Crick strand a bisulfite-converted genome does contain four distinct strands. As DNA methylation is mostly symmetrical, the MethylC-seq protocol employs directional sequencing library construction and will analyze only two of the four strands. A variation of the MethylC-seq protocol creating non-directional libraries allows capturing all four strands of a bisulfite-treated genomic DNA using an alternative sequencing adaptor strategy (BS-seq (Cokus et al. 2008)). While allowing a more comprehensive mapping of cytosines, this strategy requires a large amount of input DNA and comes at the cost of a more complex bioinformatics analysis. Therefore, this approach has been rarely used (Popp et al. 2010). Similarly, whole-genome pre-amplification of bisulfite-treated DNA has been proposed as an alternative to obtain sufficient material for sequencing (Kobayashi et al. 2012), however, the accuracy and reproducibility of the whole-genome amplification are still under debate, as it is prone to bias, especially if low amounts of input DNA are used (Bundo et al. 2012). Furthermore, there is a polymerase dependent bias toward highly methylated DNA for WGBS, which might skew the DNA methylation measurement (Ji et al. 2014; Olova et al. 2018). To avoid the amplification bias, an amplification-free protocol for WGBS based on two rounds of adaptor ligations has been devised and shown an improvement in GC bias, but has so far been little used (McInroy et al. 2016).

A variation of the protocol, called tagmentation-based WGBS (T-WGBS) has been developed. Tagmentation is based on a hyperactive variant of the prokaryotic Tn5 transposase that randomly fragments DNA and tags ends with the sequencing adaptors that can be subsequently used for amplification. While initially used for low-input genome sequencing (Adey et al. 2010), the approach has been adapted to the analysis of genome-wide DNA methylation analysis (Tn5mC-seq or T-WGBS (Adey and Shendure 2012; Wang et al. 2013; Weichenhan et al. 2018)) (Fig. 16.1), single-cell RNA sequencing (Brouilette et al. 2012), chromatin accessibility (ATAC-seq (Buenrostro et al. 2013)), transcription factor binding sites and histone modifications (Schmidl et al. 2015; Kaya-Okur et al. 2019). As the tagmentation requires double-stranded DNA as a substrate for the transposition reaction, it is performed prior to the bisulfite conversion. Transposase complexes are loaded with methylated oligonucleotide (except for the 19 bp transposase recognition sequence) to retain sequence identity after the bisulfite treatment and enable the use of the standard amplification primers. A second complementary methylated adaptor is added replacing the transposase recognition sequence and ligated using gap repair. Double-stranded DNA fragments are bisulfite converted and subsequently PCR amplified to append the outer, flow cell compatible primers. The protocol can be carried out with very little input (down to 10 ng), as the tagmentation step removes the need for the multiple steps of the conventional protocol for library preparation (DNA shearing, 3′ end repair, adenylation, and adapter ligation). The presence of unmethylated nucleotides during the gap repair step serves as internal control for bisulfite

conversion efficiency and abolishes the need for DNA spike-ins. On the other hand, the incorporation of methylated cytosines during this step improves sequencing performance. It enables sequencing to be initiated in sequences with a GC content comparable to unconverted genomic DNA before sequencing the bisulfite-treated stretches of DNA, optimizing the cluster passing filter in the initial sequencing cycles and addressing a historical challenge with Illumina short-read sequencing (Lu et al. 2015). Further optimizations to the protocol include the additional use of custom oligos to amplify a single of the two complementary strands abolishing the need of specific pre-methylated adaptors as well as the use of spike-ins with a high GC content for base equilibration during sequencing (Suzuki et al. 2018). The tagmentation reaction has been shown to be little affected by the GC content (Adey and Shendure 2012). Up to 96% of CpGs can be covered with this approach and ~70% of reads align to the genome, a number slightly lower compared to the standard MethylC-Seq protocol. However, the coverage has been shown to be slightly more uniform compared to the standard MethylC-seq (Adey and Shendure 2012). Overall, MethylC-seq and T-WGBS do yield highly similar results in terms of methylation levels and coverage of the genome and there seems to be no sequence bias for the insertion of the transposase (Wang et al. 2013). Of note, T-WGBS has been found unsuitable for the analysis of DNA extracted from FFPE tissue (Wang et al. 2013).

A potential drawback of both the MethylC-seq and the tagmentation-based protocol is that the adaptors are ligated to the DNA fragments prior to the bisulfite conversion. One problem related to this is that the treatment of DNA with sodium bisulfite leads to a substantial degradation of DNA and reduces significantly the amount of amplifiable DNA through the induction of double strand breaks between the adaptors. Therefore, protocols performing the adaptor tagging after bisulfite treatment (PBAT: *P*ost-*A*daptor *B*isulfite *T*agging) have been devised and shown to enable efficient library construction from as little as 125 pg of DNA (Miura et al. 2012) and even single cells (Smallwood et al. 2014). The protocol can be performed without amplifications as originally devised (Miura et al. 2012), and combined with amplification, which allows to sequence the methylomes of rare cell populations, such as primordial germ cells or zygotes (Kobayashi et al. 2012; Peat et al. 2014; Kobayashi et al. 2013; Shirane et al. 2013). Adaptor tagging is performed with two rounds of random primer extension using oligonucleotides with a random tetramer sequences at the 3′-end of the amplification primers containing the Illumina adaptor sequences. Primers are biotinylated to allow capture of the biotinylated fragments after the first strand synthesis on magnetic beads. When starting from ~100 ng of DNA, this method allows for the routine PCR-free construction of libraries for methylome-wide sequencing (Miura et al. 2012), thereby avoiding the problem of high PCR duplicate rates that is frequently occurring in PBAT protocols due to the preferential binding of the random amplification primers when the library has insufficient complexity and diversity. Variations of the PBAT protocol, which are partly used in commercial kits, include the use of random hexamer priming, followed by extension and 3′ end tagging, the ligation of proprietary adaptor constructs, or the *Sp*linted *L*igation *A*daptor *T*agging (SPLAT) protocol, in which adaptors with random hexamers are consecutively ligated to the 3′ and 5′ ends of the bisulfite-treated template prior to PCR amplification (Raine et al. 2017).

A number of commercial kits for WGBS are currently available. The performance of these commercial products has been benchmarked in several studies, revealing large differences in a variety of quality metrics such as sequencing quality score, coverage of CpGs and genome-wide uniformity, library fragment size, GC bias and duplicate reads (Olova et al. 2018; Raine et al. 2017; Zhou et al. 2019; Nair et al. 2018; Daviaud et al. 2018; Morrison et al. 2021; Foox et al. 2021). Amplification-free PBAT has shown to yield global DNA methylation levels close to mass spectrometry-based measurements and outperformed classic WGBS protocols, which tend to overestimate DNA methylation levels for some genomic regions with high methylation

levels (Olova et al. 2018; Zhou et al. 2019). However, other benchmark studies experienced suboptimal coverage and sequencing quality metric due to the little amount of material recovered from the amplification-free PBAT protocol (Morrison et al. 2021). Evaluation of the different steps in both the classic MethylC-seq and the PBAT protocol showed that the bisulfite conversion step is the major cause for biased DNA methylation level estimation, which was increased through PCR amplification (Olova et al. 2018).

Alternative approaches based on enzymatic and chemical conversion methods such as EM-seq or TAPS-seq have been developed (Vaisvila et al. 2021; Liu et al. 2019). Enzymatic methylation conversion (EM) is a recent technology, which yields data similar to bisulfite conversion, but using an enzymatic method instead of the chemical conversion (Vaisvila et al. 2021). Methylated cytosines are sequentially, enzymatically oxidized to hydroxy-, formyl- and eventually carboxymethylated cytosines using TET2. Subsequently, unprotected, i.e. formerly unmethylated cytosines, are deaminated to uracils using APOBEC proteins. Following PCR, unmethylated cytosines are represented as thymines, while methylated cytosines lose their protective modifications and will be represented as cytosines. Enzymatic conversion is supposed to allow for lower amounts of input material as it fragments less the DNA compared to bisulfite conversion and avoid some of the technical biases observed with the chemical conversion. It was recently shown to outperform chemical conversion-based protocols yielding notably larger library insert sizes and higher library complexity for both standard and low-input library preparation (Morrison et al. 2021; Foox et al. 2021; Han et al. 2021), improved conversion efficiency (Sun et al. 2021), and the lowest number of reads to obtain a 20 x coverage (Foox et al. 2021). However, although promising, the technology has not yet been widely used and potential problems and technical biases still need to be comprehensively assessed.

TAPS-seq uses the TET1 dioxygenase to oxidize both 5mC and 5hmC to 5caC, and pyridine borane reduces 5caC to dihydrouracil, which is subsequently read as thymine (Liu et al. 2019). Both EM-seq and TAPS-seq have recently been shown to enable long read sequencing and phasing of DNA methylation patterns over kilobases of DNA (Sun et al. 2021; Liu et al. 2020a).

16.5 Genome-Wide Methylation Analyses Using NGS

Whole methylome analyses covering each CpG in the genome at single base resolution remain complex and resource intensive when aiming for a reasonable coverage of at least 20–30× on a given CpG site and are, therefore, not yet feasible in large cohorts. Furthermore, it has been shown that more than half of all reads do not contain even a single CpG dinucleotide and are thus without any information for DNA methylation analysis, making suboptimal use of the sequencing capacities of current sequencers (Ziller et al. 2013). Furthermore, many CpGs will not show variable DNA methylation under any condition. Therefore, several approaches have been developed to concentrate on the “potentially informative” fraction of the genome (Table 16.1). These approaches make use of either sequence features such as the CpG density or use antibody, protein, or chemical labeling-based methods to enrich the methylated or unmethylated fraction of the genome. However, it should be kept in mind that all of them will only analyze part of the CpGs present in a genome and none does provide a comprehensive analysis of the methylome.

16.5.1 Bisulfite-Based Methods

*R*educed *R*epresentation *B*isulfite *S*equencing (RRBS (Meissner et al. 2005, 2008; Gu et al. 2011)) is currently the most popular alternative to WGBS, as it requires significantly less sequencing and CpG-rich regions that are enriched by restriction enzyme digestion are relatively well-covered (Meissner et al. 2008; Bock et al. 2010; Harris et al. 2010). RRBS makes use of a methylation-insensitive restriction

Table 16.1 Key characteristics of the most widely used methods for the genome-wide analysis of DNA methylation

Technology	The technology behind	*M*ethylated/ *U*nmethylated fraction	Single-nucleotide resolution (yes/no)	DNA input in µg	Required number of reads per sample in M	Coverage of the human genome (%)	Coverage of CpGs in the human genome	Comments	References
WGBS/ MethylC-seq	Adaptor ligation before or after bisulfite treatment of genomic DNA prior to sequencing	U+M	Yes	0.1–5	500	95	29 M	Comprehensive coverage of all cytosines in a genome, several variations of library preparation exist, cost-intensive, but will with decreasing sequencing costs become the gold standard for the discovery technologies, no distinction between 5mC and 5hmC	Urich et al. (2015), Wang et al. (2013), Miura et al. (2012)
RRBS	Restriction digestion allows for size selection of small fragments, used for library construction, bisulfite conversion and sequencing	U+M	Yes	0.01–1	20	4	1.6 M	85% of CpG islands are covered, cost-effective sequencing approach, low coverage in CpG poor regions, no distinction between 5mC and 5hmC	Meissner et al. (2005), Gu et al. (2011)
COBRA-seq	Library construction, bisulfite conversion, restriction enzyme digestion, enrichment of cut fragments with biotin labels, amplification and sequencing	M	Yes	0.1–1	10–15	<1	1–4 M	Allows for analysis of DNA methylation in other sequence context than CpG and might be an interesting method in organisms with very little methylation, relative quantification only, less dependent on methylation density compared to MeDIP/MBD-seq	Varinli et al. (2015)
Agilent SureSelect^{XT} methylation capture	Libraries are prepared from genomic DNA fragments, hybridized to the capture probes, eluted, bisulfite converted, and amplified	U+M	Yes	3	50	2.8	3.7 M	Off-the-shelf product, captures all known promoters and upstream regions, uses the procedure used for exome sequencing with an additional bisulfite	Borno et al. (2012), Walker et al. (2015)

	before being sequenced by a NGS							conversion step, relatively expensive, custom design possible	
SeqCapEpi CpGiant enrichment	Libraries are prepared from genomic DNA fragments, bisulfite converted, pre-amplified, hybridized to the capture probes, eluted, and amplified before being sequenced by NGS	U+M	Yes	1	40–50	2.7	5.5 M	Less input compared to the SureSelect capture as capture is performed on bisulfite-treated libraries after pre-amplification, relatively expensive, custom design possible	Walker et al. (2015), Li et al. (2015a)
MBD-seq/ MethylCap-seq	Enrichment of methylated DNA using a Methyl-binding domain protein before or after a library construction followed by sequencing	M	No	0.2–1	30–40	95	17 M	Also suitable for FFPE, amenable to high throughput and analysis of large cohorts, absolute quantification of DNA methylation difficult, specific for 5mC	Brinkman et al. (2010), Serre et al. (2010)
MeDIP-seq	Library construction, enrichment of methylated DNA using an antibody against 5mC followed by sequencing	M	No	0.3–0.5	50–60	95	17 M	Captures and allows to analyze also the transposons of the human genome, specific for 5mC, absolute quantification of DNA methylation difficult, lower signal to noise compared to MBD- and MIRA-seq, better coverage in CpG poor regions	Weber et al. (2005), Taiwo et al. (2012)
MIRA-seq	Enrichment of methylated DNA using a Methyl-binding domain protein before or after a library construction followed by sequencing	M	No	0.1–1	40–50	95	17 M	Higher specificity compared to MBD-seq, absolute quantification of DNA methylation difficult, specific for 5mC	Jung et al. (2015)
MRE-seq	Digestion of genomic DNA in parallel with several methylation-sensitive restriction enzymes, size	U	Yes, at restriction site	1.5–2.5	30	4	1.7 M	Analyzes preferentially unmethylated CpGs in high density CpG regions, in combination with MeDIP	Li et al. (2015b), Maunakea et al. (2010)

(continued)

Table 16.1 (continued)

Technology	The technology behind	*M*ethylated/ *U*nmethylated fraction	Single-nucleotide resolution (yes/no)	DNA input in µg	Required number of reads per sample in M	Coverage of the human genome (%)	Coverage of CpGs in the human genome	Comments	References
	selection, library construction, sequencing							comprehensive genome coverage is achieved	
DREAM-seq	Sequential digestion with the SmaI/XmaI isoschizomer pair, creates fragments with methylation-specific ends, adaptor ligation and sequencing	U+M	Yes, at restriction site	1–2	25	1.8	50,000	Analyzes preferentially unmethylated CpGs in high density CpG regions (e.g., CpG islands), low number of CpGs covered	Jelinek et al. (2012)
Infinium BeadChips	Genotyping of bisulfite-converted DNA	U+M	Yes	0.5	Microarray	2–3	480,000/ 840,000	Absolute quantification, cost-effective solution, amenable to the analysis of large cohorts, high data quality, standard bioinformatic pipelines available, limited coverage	Moran et al. (2016), Sandoval et al. (2011)

Abbreviations: *5mC* 5-methylcytosine, *5hmC* 5-hydroxymethylcytosine, *WGBS* whole-genome bisulfite sequencing, *RRBS* reduced representation bisulfite sequencing, *COBRA* combined bisulfite restriction analysis, *MBD* methyl-binding domain, *MeDIP* methylated DNA immunoprecipitation, *MIRA* methylated-CpG island recovery assay, *MRE* methylation-sensitive restriction enzyme sequencing, *DREAM* digital restriction enzyme analysis of methylation

endonuclease with a CG-rich recognition sequence, such as MspI, which cuts between the two Cs in the target sequence CCGG, which is frequently found in CpG islands and promoter regions. After a size selection step, the generated DNA fragments representing approximately 1% of the entire genome are subsequently used for a standard library construction using methylated adaptors followed by bisulfite conversion. RRBS requires about 30 times less sequencing reads than WGBS and interrogates approximately 2–4 M CpGs covering 80% of CpG islands and 60% of promoters, but only ~10% of enhancers, which are often located in CpG poor regions (Smith et al. 2009). Furthermore, the method lends itself to automation and can be used to analyze large cohorts (Smith et al. 2009; Boyle et al. 2012; Klughammer et al. 2018). Typically, more than 70% of the expected MspI fragments are detected by sequencing (Kacmarczyk et al. 2018). However, this corresponds to only ~10% of the 28 million (M) CpGs in a human genome, mainly in regions of high CpG density, such as CpG islands. Increased coverage of CpGs can be obtained by selecting two population of MspI fragments of different sizes, which are subsequently bisulfite converted and PCR amplified in parallel prior to pooling of the two fractions and sequencing (Akalin et al. 2012; Garrett-Bakelman et al. 2015). An alternative approach to target regions and genomic features of interest consists of using combination of several different restriction enzymes optimized for the desired fragment size and target regions (Lee et al. 2014; Kirschner et al. 2016; Martin-Herranz et al. 2017; Tanas et al. 2017). The number of covered CpGs will thereby depend on the number of fragments created by the restriction enzyme(s) and the selection of the fragments with a certain length. Notably, RRBS can be performed from minute amounts of DNA and especially for non-human samples, for which no Infinium Methylation BeadChips exist, RRBS is probably the most cost-efficient method to obtain comprehensive high-resolution and quantitative methylome data, with as little as 20M sequencing reads approaching saturation (Bock et al. 2010). However, non-uniform coverage of CpGs across samples might be an issue and many CpG poor regions might be missed. Furthermore, the possibility to investigate regions of particular interest will depend on the presence of nearby restriction sites.

An alternative approach for large-scale methylome sequencing uses long probes to capture DNA for regions of interest, reducing sequencing costs by approximately ~90% compared to standard WGBS. This capture can be performed using either oligonucleotide microarrays (Hodges et al. 2009) or solution-based hybridization (Lee et al. 2011) and capture can be performed on the bisulfite-converted fragments (Hodges et al. 2009) or prior to bisulfite conversion (Lee et al. 2011). Solution-based hybrid methods present clear advantages in terms of flexibility and specificity. They make use of biotinylated RNA baits to capture regions of interest (Lee et al. 2011), similarly as in standard exome sequencing for the detection of genetic variations (Gnirke et al. 2009). If the capture is performed prior to conversion, this method requires an increased amount of input DNA, with little DNA being recovered after hybridization, necessitating a high number of PCR cycles after bisulfite treatment. Only one of the two DNA strands is generally captured during hybridization, which allows maximizing the number of regions than can be captured simultaneously. Furthermore, the harsh conditions of the bisulfite treatment can further reduce the complexity of the captured fragments. If capture is performed after bisulfite conversion, the design of probes complementary to all possible alleles generated by the bisulfite conversion is required to avoid the preferential capture of molecules with a distinct DNA methylation pattern. While capturing pre-amplified bisulfite-converted DNA on a microarray (Hodges et al. 2009) can partially circumvent the problem of the required starting material, this approach leads nonetheless to a reduced percentage of "on-target" sequences compared to the capture after conversion-based methods (~10% vs. 80%).

Capture probes can be custom designed and predesigned products enabling the capture of promoters, and CpG islands are nowadays commercially available. Agilent's SureSelect™

human Methyl-seq covers ~3% of the genome (84 Mb), corresponding to ~1/7 of all sequencing accessible CpG sites (3.7 M CpG sites (Borno et al. 2012)). The Tru Seq EPIC panel from Illumina captures ~107 Mb of DNA including 3.3 M CpGs, while Arbor Biosciences focuses on smaller panels of ultraconserved sequences in different evolutionary classes. The same technology also allows customized capture panels targeting regions of interest (Ivanov et al. 2013; Morselli et al. 2021), which are currently available from several vendors. Although the required amount of starting material has been significantly reduced since the method was originally devised (from 20 to 30 μg (Lee et al. 2011) to 0.5–3 μg (Kacmarczyk et al. 2018; Walker et al. 2015)), it remains one of the major drawbacks of these methods. This requirement is mainly due to the fact that the capture is generally carried out after adapter ligation and prior to the bisulfite conversion proscribing an amplification step. Furthermore, the high number of PCR cycles that need to be performed following the release of the captured fragments potentially distorts the DNA methylation values. To avoid a too high number of PCR amplification cycles after bisulfite treatment of the released DNA fragments, DNA from several hybridizations can be pooled prior to bisulfite conversion (Lee et al. 2011). However, recent advances show the feasibility of a post-bisulfite adaptor tagging (PBAT) protocol, reducing significantly the amount of required material (Miura and Ito 2015). The design of custom capture arrays targeting a high number of specific regions of interest makes this approach competitive to RRBS, which costs substantially less, but might also be less suited if specific regions of interest are not easily accessible by a (combination of) restriction digests.

The SeqCap Epi CpGiant from Roche/ NimbleGen, which has now been discontinued, allowed for a reduced input of DNA (250 ng–1 μg) compared to the SureSelect MethylSeq system and was performed on a whole-genome bisulfite sequencing (WGBS) library by using capture probes complementary to methylated, unmethylated, and partially methylated targets after bisulfite conversion (Walker et al. 2015; Allum et al. 2015; Li et al. 2015a). Up to 5.5M CpGs in ~80 Mb of the human genome could be interrogated simultaneously at single-nucleotide resolution. A sophisticated design in combination with the use of long probes allowed for a very efficient capture and focused on only regions of interest at much increased coverage compared to WGBS (Allum et al. 2015). The probe design did allow pre-amplifying the sequencing library prior to capture and thereby starting from smaller amounts of input material compared to other capture approaches. In contrast to all other capture approaches, the CpG Giant captured CpGs on both strands allowing to assess strand methylation symmetry.

In a direct comparison the Illumina, Roche, and Agilent off-the-shelf capture products, which do not target the same CpGs, showed a similar performance capturing well the targeted regions with few off-target regions being covered and methylation values for CpGs measured by different protocols correlated well with each other (Kacmarczyk et al. 2018). Due to the design of the capture arrays, enriched coverage in CpG island surrounding regions and promoters devoid of CpG islands is achieved compared to RRBS and the capture approaches cover more targeted regions at higher coverage, allowing a more precise quantification. Capture approaches require increased amount of starting material, but are less sensible to the quality of the DNA (Kacmarczyk et al. 2018). The main advantage of the capture approaches remains the possibility to design custom panels targeting regions of interest or focus on specific genomic features for organisms, for which no BeadChips are currently available.

Several technologies have been developed that allow the rapid generation of a large number of amplification products simultaneously. Padlock probes are horseshoe shaped oligonucleotides with their both ends complementary to a bisulfite-converted sequence surrounding a CpG of interest (Deng et al. 2009; Ball et al. 2009). The region complementary to the target is filled in by a polymerase and the padlock probe is circularized, which protects it against an exonuclease digestion used to remove linear DNA. The

target regions are PCR amplified using primers with molecular barcodes and sequencing platform-specific adaptors annealing to the common backbone of the padlocks, and subsequently sequenced. This approach allows analyzing tens to hundreds of thousands of CpGs (Deng et al. 2009; Ball et al. 2009; Diep et al. 2012). As the primer design and set-up of padlock probes requires significant upfront investment, as well as expertise in the design and some optimization, these technologies are most suited for the analysis of DNA methylation patterns of a large number of loci in either large cohorts or in a routine setting with a fixed panel of target regions.

16.5.2 Affinity and Antibody-Based Enrichment Methods

In contrast to the above-described sequencing assays, several methods have been developed, which yield enrichment values for the methylation states over genomic regions, but do not provide quantitative values of DNA methylation levels at individual CpGs. Genome-wide DNA methylation profiles with such a region-specific resolution can be obtained using methylated DNA immunoprecipitation (MeDIP (Weber et al. 2005)) or enrichment with methylated-CpG binding proteins (Brinkman et al. 2010; Serre et al. 2010), which will yield alterations at the level of regions, but not single nucleotides. Several proteins are commonly used for MBD sequencing approaches, including MBD2 and MECP2, and the inherent principle of affinity purification of methylated DNA is known under several names: MBD isolated genomic sequencing (Serre et al. 2010), MCip (Gebhard et al. 2006), MBD-seq (Li et al. 2010b), or MethylCap-seq (Brinkman et al. 2010), or MIRA-seq (Jung et al. 2015) (Table 16.1). Although MBDs might display a certain target specificity and MBD columns enrich methylated DNA significantly, they do not fully purify methylated sequences (Selker et al. 2003). When single base resolution is not required, these technologies provide a compromise for DNA methylation studies in large cohorts at a reasonable cost, which the methods covering all CpGs at single-nucleotide resolution are not yet able to deliver. Furthermore, these approaches are specific for the analysis of 5-methylcytosine and therefore allow distinguishing 5-methylcytosine from 5-hydroxymethylcytosine (Williams et al. 2011; Wu et al. 2011). The latter can be analyzed using specific antibodies against 5-hydroxymethylcytosine, as described in the section on methods for the analysis of 5-hydroxymethylcytosine (Sect. 16.9). The shift to NGS as a readout platform has significantly increased the accuracy, as the microarray-based protocols required a whole-genome amplification step, which led to a substantial bias in CpG-rich regions (Robinson et al. 2010). Nonetheless, it should not be forgotten that standard amplifications using, e.g., the Illumina adaptors are tailored to a specific GC content and regions with very high or low GC content might still be subject to amplification bias.

One of the major advantages of MeDIP-seq is the simplicity of the protocol and its resemblance with widely used ChIP-seq protocols, as well as the possibility for automation (Taiwo et al. 2012). Briefly, genomic DNA is fragmented using ultrasound or acoustic shearing and (unmethylated) sequencing-platform specific adaptors are ligated to the fragmented DNA. This step has to be performed prior to the immunoprecipitation as the immunoprecipitation yields single-stranded DNA, which is incompatible with standard library preparation protocols (Taiwo et al. 2012). Methylated DNA is immunoprecipitated with high affinity and specificity with an antibody against 5-methylcytosine, and most commercial suppliers provide the same monoclonal antibody. Immunoprecipitated fragments are released from the beads, PCR amplified and sequenced. Improved and automated protocols allow for the standardized high-throughput analysis of samples with little starting material (Taiwo et al. 2012). The number of sequencing reads in MeDIP-seq depends on both the DNA methylation level and the CpG content of the sequence. Computational algorithms are therefore required to account for this bias in amplification and to convert signal intensities into a methylation percentage (Down

et al. 2008; Pelizzola et al. 2008; Huang et al. 2012; Xiao et al. 2015). As most CpGs in a mammalian genome are methylated, MeDIP-seq allows to cover approximately 70% of all CpGs (Taiwo et al. 2012).

In combination with the complementary *M*ethylation-sensitive *R*estriction *E*nzyme sequencing (MRE-seq) approach, which makes use of methylation-sensitive restriction enzymes to map unmethylated cytosines within restriction recognition sites at single-nucleotide resolution (see also Sect. 16.5.3), a comprehensive coverage of ~80% of the CpGs in the human genome can be obtained (Li et al. 2015b).

MeDIP-seq does not provide data with single-nucleotide resolution, which needs to be obtained by locus-specific analyses such as (pyro)-sequencing or mass spectrometric analysis of the regions of interest and thus requires extensive follow-up studies. However, partial data on the DNA methylation patterns at single-nucleotide resolution can be obtained by bisulfite conversion of the immunoprecipitated methylated DNA fragments. However, as the fragments are enriched for completely methylated molecules (Sengenes et al. 2010), information of variable DNA methylation patterns will be lost and the observed profiles might not be representative of the methylation states of the original sample.

In an alternative approach, the methylated fraction of a genome is isolated by affinity purification of methylated DNA with MBD proteins. Of the different MBD proteins, MBD2b has the highest affinity for methylated DNA (Fraga et al. 2003). For the affinity-based enrichment with MBD proteins, genomic DNA is sonicated prior to capture with the respective MBD protein coupled to a solid support, such as streptavidin beads. Following the capture reaction, the bound methylated DNA can be eluted as a single fraction or in several fractions using increasing concentrations of salt in the elution buffer, which enables to target fractions with a specific CpG density, because fragments with a high density of CpGs are eluting last. As double-stranded DNA is recovered from the elution, the library preparation can be performed after affinity purification. MBD-seq methods can also be applied to fragmented DNA, such as DNA extracted from FFPE tissues. It is amenable to very high throughput through automation, and the analysis of very large cohorts including 1500 samples (750 schizophrenia patients and 750 controls) has been reported (Aberg et al. 2012). Information at single-nucleotide resolution can be obtained by bisulfite conversion of the captured DNA and subsequent massively parallel sequencing (Brinkman et al. 2012).

In the *M*ethylated-CpG *I*sland *R*ecovery *A*ssay (MIRA), a glutathione S-transferase (GST)-tagged full-length MBD2b is used to bind sonicated methylated DNA fragments and the affinity to methylated-CpG dinucleotides is further enhanced in a dose-dependent manner by the addition of the MBD3-like-1 protein (Rauch and Pfeifer 2005). The combined effect significantly improves the sensitivity of the assay and a single methylated-CpG dinucleotide allows for capture of the corresponding DNA molecule. Ligation of oligonucleotide linkers to enzymatically digested DNA prior to affinity chromatography permits efficient amplification of eluted fractions and subsequent analysis of input DNA and MIRA-enriched amplification products by NGS (Jung et al. 2015; Choi et al. 2010). Of note, as the capture is performed on double-stranded DNA, library preparation can be performed both before or after the enrichment step. There are a number of commercial kits for MBD sequencing on the market which have been benchmarked, highlighting some differences in performance between the different kits (De Meyer et al. 2013; Aberg et al. 2015). As for MeDIP-seq, CpG density and GC content of the fragmented DNA are major biases affecting the efficiency of the affinity purification, as well as the subsequent sequencing (Robinson et al. 2010). Therefore, computational approaches have been developed for background estimation, to correct for CpG coverage and CpG density in reads (Lan et al. 2011; Riebler et al. 2014).

Instead of using antibody or protein affinity purification, methylated cytosines can also be chemically modified after oxidation with Tet enzymes, followed by the attachment of a biotin group (TAmC-seq (Zhang et al. 2013)). After

protection of 5hmC with a glucose moiety (as described in Sect. 16.9 on the analysis of 5-hydroxymethylation (Song et al. 2011)), methylated cytosines are oxidized to hydroxymethylated cytosines using a catalytically active fragment of the Tet enzyme, modified with a glucose molecule with an azide group, which by "click"-chemistry can be used for the labeling with biotin or other chemical tags. This label can subsequently be used to enrich selectively for sequences with methylated cytosines. Although this method has been demonstrated to have less density bias and allow a more even coverage, it has been little used probably due to its multi-step protocol and the fact that some of the reagents might not be readily available in all laboratories.

Both MBD-seq and MeDIP-seq are well suited for the identification of differentially methylated regions, and when correctly accounted for CpG density, they can distinguish between methylated and unmethylated regions with a precision similar to RRBS (Bock et al. 2010; Harris et al. 2010), Infinium BeadChips (Harris et al. 2010; Clark et al. 2012), or the mass spectrometry-based EPITYPER assay (Nair et al. 2011). The quantitative accuracy is nonetheless reduced for regions with intermediate DNA methylation levels such as CpG island shores (Harris et al. 2010; Irizarry et al. 2008) and despite the advantage of having a higher coverage of the genome compared to RRBS, substantially more sequencing is required (40–60M reads for MBD/MeDIP vs 20M reads for RRBS) (Bock et al. 2010). Both technologies suffer from false-positive signals in repetitive CpG-rich regions, in which minor methylation differences are amplified through the enrichment, while copy number variation do not generally seem to influence the DNA methylation profile (Bock et al. 2010; Robinson et al. 2010). There are also some differences in the target distribution between MeDIP and MBD (Nair et al. 2011). While MBD-seq enriches preferentially CpG islands and regions with high CpG density, MeDIP does also enrich for methylated CpGs in regions with low CpG density and in general the signal over baseline ratio is lower for MeDIP (Bock et al. 2010; Robinson et al. 2010; Nair et al. 2011). In practice this means that MeDIP will require a significantly higher number of reads to capture all methylated cytosines compared to MBD-seq and might identify less DMRs at the same read depth. Nonetheless, MeDIP was found to display the best coverage for the whole genome and especially for gene body regions (Walker et al. 2015). Furthermore, while MeDIP can detect methylated cytosines in any sequence context, the MBD-based approaches will only detect CpG methylation. If the target organism might contain methylated cytosines in other sequence contexts, such as in plants, MeDIP will allow for a more comprehensive DNA methylation profiling. Of note, the CpG density of the DNA fragments eluted for sequencing using the MBD-seq approaches can be modulated by using different salt concentrations for elution (Serre et al. 2010). The absence of methylation is difficult to assess using these enrichment methods as a lack of reads in a given region can be due to the absence of methylation, but could also be due to technical problems, such as lack of amplification or just chance.

Complementary information on unmethylated cytosines and regions of the genome can be obtained using an enrichment technology that has been termed mTAG-seq (Kriukiene et al. 2013). Unmodified cytosines are chemically modified using an engineered M.SssI methyltransferase and synthetic AdoMet analogs containing activated amine or azide groups , followed by chemoselective tagging of the AdoMet analog with a covalent biotin molecule, which is subsequently used for enrichment and sequencing.

16.5.3 Sequencing Approaches Using Methylation-Sensitive/Dependent Restriction Enzymes

Methylation-sensitive restriction endonucleases, i.e., enzymes that are blocked by methylated cytosines in their recognition sequence (Bird and Southern 1978), have been widely used in the past for the analysis of methylation patterns in

combination with their methylation-insensitive isoschizomers. There are about 50 methylation-sensitive restriction enzymes known, but few of them are available in combination with a methylation-insensitive isochizomer. One of the most commonly used pairs of enzymes is HpaII/MspI; both recognize and cleave the four base palindrome C^CGG in double-stranded DNA, but while MspI cleaves the DNA independent of the methylation status of the inner CpG, HpaII is unable to cleave when the second cytosine is methylated (C^^{me}CGG). Although methods based on methylation-sensitive restriction enzymes are simple and relatively cost-effective as they do not require any special instrumentation, they are hampered by the limitation to specific restriction sites, as only CpG sites found within these sequences can be analyzed. For example, only ~4% of CpG sites in non-repetitive sequences are located in HpaII recognition sites and only 0.03% can be cleaved by NotI (Fazzari and Greally 2004). The fraction of the genome interrogated is often arbitrary and not associated with a specific functionality. In addition, methods using these enzymes might be prone to false-positive results due to incomplete cleavage and some sequences are intrinsically resistant to digestion if not appropriately controlled. For example, non-CpG methylation on cytosines or DNA adducts in the vicinity of the cleavage site might influence the restriction capacity of an enzyme. Digestions are therefore difficult to perform on material extracted from formaldehyde-fixed paraffin-embedded (FFPE) samples. Although in many cases restriction enzyme-based approaches have been replaced by more quantitative sequencing methods allowing the more rapid identification of altered DNA methylation levels at higher spatial resolution, a few protocols are still in use and have been successfully transferred to NGS instruments, which permit a more quantitative analysis of the isolated (methylated or unmethylated) fraction of the genome (Table 16.1).

*M*ethylation-sensitive *R*estriction *E*nzyme sequencing (MRE-seq) identifies unmethylated-CpG sites at single CpG site resolution by sequencing size-selected fragments from parallel DNA digestions with a number of methyl-sensitive restriction enzymes (e.g., HpaII, Hin6I and AciI) and covers ~ 1.7M of the 29M CpG sites in the human genome (Maunakea et al. 2010). After restriction digestion and size selection, Illumina adaptors are ligated to the DNA fragments, PCR amplified and sequenced using short-read (50 bp) single-end sequencing, thereby identifying unmethylated CpG sites within the restriction sites with single base resolution. Interrogation of the methylated fraction of the same sample by MeDIP-seq, described in more detail in the previous section, yields complementary information and the combined use of MRE-seq and MeDIP-seq allows for genome-wide DNA methylation analysis at high coverage and resolution, while limiting the biases of each technology (Li et al. 2015b). The combined approach analyzes up to 22M of the 29M CpGs of the human genome and the quantitative accuracy of the methylation levels obtained by MeDIP in regions with low or intermediate DNA methylation is improved (Harris et al. 2010). Similarly, the methylated fraction of the genome can also be interrogated using methylation-dependent restriction enzymes of the mrr-like family (FspEI, MspJI, LpnPI, AspBHI, etc.) used in the MethylRAD assay, which creates small fragments of 32 nucleotides around a methylated cytosine which are subsequently sequenced (Wang et al. 2015a). A very similar protocol to MRE-seq, termed Methyl-MAPS, isolates both the methylated and unmethylated fraction of the genome by using the methylation-dependent restriction endocuclease McrBC and a combination of several restriction enzymes, respectively, prior to library preparation and paired-end sequencing (Edwards et al. 2010).

Digital Restriction Enzyme Analysis of Methylation (DREAM) is based on sequential cuts of the genomic DNA with a pair of restriction enzymes (SmaI and XmaI) at CCCGGG target sites (Jelinek et al. 2012, 2018). Unmethylated sites are first digested with SmaI. This enzyme cuts the sites in the middle at CCC^GGG, leaving behind blunt-ended fragments. CpG methylation completely blocks SmaI; therefore, only unmethylated sites are cleaved. In the next step,

the remaining methylated sites are digested with XmaI, which is not blocked by CpG methylation. It cuts the recognition site sideways at C^CCGGG forming 5′-CCGG overhangs. The sequential cuts thus create distinct methylation-specific signatures at the ends of restriction fragments: 5′-GGG for unmethylated CpG sites and 5′-CCGGG for methylated sites. The DNA fragments resulting from the digestions are ligated to barcoded NGS adapters, libraries with distinct barcodes are pooled and sequenced using a paired-end protocol. The sequencing reads are aligned to the genome and mapped to unique CCCGGG target sites. Methylation at individual CpG sites is calculated as the digital counting of sequencing reads with the methylated signature to the total number of reads mapping to the site. Quantitative accuracy can be improved using spike-ins with defined levels of DNA methylation at the enzyme recognition sites (Jelinek et al. 2018). Sequencing of 25 million reads per sample typically yields accurate determination of DNA methylation levels at 50,000 unique CpG sites with high coverage (Jelinek et al. 2012), but results of good quality have been reported with as little as eight million reads (Kitazawa et al. 2021). Due to its cost effectiveness, high reproducibility, and good correlation with other read-out technologies such as the 450K array and RRBS, the technology continues to be used (Kitazawa et al. 2021; Yamazaki et al. 2021).

For the analysis of gene-specific methylation patterns or individual CpG positions, methods using methylation-sensitive endonucleases have largely been replaced by PCR-based methods following treatment of genomic DNA with sodium bisulfite, although as mentioned briefly in Sect. 16.6, they are still used and do provide some advantages when combined with high-throughput qPCR and multiplex amplification systems.

16.5.4 Epigenotyping Arrays

With whole-genome bisulfite sequencing being not yet affordable at a large scale, and given the low resolution of antibody and methyl-binding protein enrichment of methylated DNA, epigenotyping technologies have emerged as an alternative tool for the identification of differentially methylated regions and DNA methylation-based biomarkers. Epigenotyping technologies, such as the Infinium Human Methylation 450K or EPIC BeadChip (Illumina Inc., CA, USA) generate a methylation state-specific "pseudo-SNP" through bisulfite conversion, thereby translating differences in DNA methylation patterns into sequence differences that can be analyzed using quantitative genotyping methods (Bibikova et al. 2009, 2011). The 450K BeadChip has dramatically expanded the genome coverage compared to previous generations of the BeadChip, analyzing more than 480,000 CpG sites covering 99% of all RefSeq genes with an average of 17 probes per gene. This array is not only focused on CpG islands, but probes are distributed over various functional elements that are more prone to alter their DNA methylation status in response to environmental conditions or in cancer, such as CpG island shores and shelfs (Irizarry et al. 2009). The 450K BeadChip was replaced by the Illumina MethylationEPIC BeadChip in 2016. On this BeadChip ~ 400,000 CpGs were added to the content of the Illumina HM450array, focusing essentially on enhancer regions identified in the ENCODE and FANTOM5 projects, thereby extending significantly the information content of the BeadChip (Moran et al. 2016). However, due to the cell-type specificity of distal regulatory elements, the total number of elements covered for any cell type remains relatively low. Nonetheless, the addition of the intergenic gene regulatory regions, which do often display intermediate and variable DNA methylation levels and might contain other cytosine modifications than methylation, makes this array also suitable for the analysis of hydroxymethylcytosine.

These arrays have been widely used for large-scale high-throughput studies as they employ highly standardized protocols that can be integrated with a large degree of automation into existing genotyping pipelines The analysis of the results is relatively straightforward compared to the required correction for CpG density or the cost- and time-intensive bioinformatic calculations when compared to sequencing-

based DNA methylation analysis, but requires specific normalization protocols due to the combination of two different assay chemistries on the 450K and EPIC BeadChip that display a different dynamic behavior (Aryee et al. 2014; Morris et al. 2014; Touleimat and Tost 2012; Tian et al. 2017; Xu et al. 2021a; Muller et al. 2019; Zhou et al. 2018; Morris and Beck 2015): The Infinium I probes (InfI), which represent ~1/3 of the 450K array and less than 5% on the EPIC array, convey information about the methylation state in the type of the bead (InfI). Two different bead codes are used to interrogate allele—specifically the base following an unmethylated or methylated cytosine (T or C after bisulfite conversion) at the same genomic location. If the 3′ end of the probe hybridizes correctly, the probe is extended with the bases following the potential methylation variable position. The extended base is thus the same for methylated or unmethylated alleles, which means that the fluorescent signal does not carry any information on the methylation status. The Infinium II (InfII) probes are attached to a single type of beads and the methylation information is obtained through dual channel single-nucleotide primer extension with labeled dideoxynucleotides on the methylation variable position of a CpG. These probes thus take less physical space on the microarray. The use of the two chemistries is a compromise to ensure the coverage of a large number of CpGs throughout the entire genome and also in CpG island associated promoter regions. While the InfII probes require a single probe to interrogate a CpG positions increasing thereby the number of CpGs potentially analyzed on the BeadArray, they can only tolerate up to three CpG positions in the genomic sequence complementary to the 50-mer probe, which are covered using degenerate bases (e.g., R (A/G)) complementary to the potentially methylated position after bisulfite treatment. The InfI probes tolerate more variable positions, but it should be noted that their design assumes an identical methylation pattern (methylated or unmethylated) of CpGs present in the 50-mer probe sequence. Therefore, InfI probes are more adequate to analyze CpG positions in regions of high CpG density such as CpG islands. InfI/InfII are therefore not equally distributed among functional or CpG island-based categories present on the array.

Although data from the BeadChips is relatively easy to generate, a number of technical issues, such as batch effects, or underlying SNPs might confound the analyses if not appropriately controlled for. Furthermore, the low sequence complexity of bisulfite-converted DNA could induce cross-hybridization events and a number of potentially problematic probes have been reported (Chen et al. 2013; Pidsley et al. 2016). However, when properly handled these epigenotyping arrays show a high reproducibility between technical and biological replicates and display high correlation to WGBS and RRBS (Bock et al. 2010; Pidsley et al. 2016) as well as locus-specific quantitative assays such as Pyrosequencing (Roessler et al. 2012) or MethyLight (Campan et al. 2011). The technical variability of the Infinium arrays is about two- to three-fold lower as for WGBS and it has been estimated that to achieve a comparable precision as provided by the Infinium arrays, sequencing coverage of at least 100x would be required (Zhou et al. 2019), which at least for WGBS would be cost-prohibitive. Nonetheless, the sensitivity is reduced for extreme methylation values, with increased DNA methylation levels for lowly methylated regions, and lower methylation levels for highly methylated regions compared to other quantitative locus-specific DNA methylation technologies (BLUEPRINT Consortium 2016).

The BeadChips can be applied to a variety of biological specimens including fresh-frozen samples, as well as formalin-fixed paraffin-embedded (FFPE) samples or PAXgene conserved samples. While DNA extracted from fresh-frozen and FACS or MACS sorted cells can be directly used on the array, FFPE samples, which are the most common form of tissue preservation in pathological archives, are not suitable for the direct use in the procedure due to the extensive cross-linking, fragmentation and generation of apurinic/apyrimidinic sites, which all impede enzymatic processing steps such as whole-genome amplification, which is an essential step in the BeadChip protocol. A restoration method for FFPE DNA uses a ligation-based

approach to obtain DNA fragments of sufficient size (Thirlwell et al. 2010). However, there is still a controversy about the concordance between differentially methylated loci detected in fresh-frozen tissue and DNA restored from FFPE tissue (Jasmine et al. 2012). While this approach has been found to yield relatively good results that lead to similar biological findings to those obtained from fresh-frozen samples (Moran et al. 2014; Dumenil et al. 2014), it was also pointed out that there are substantial differences, which will prohibit a combined analysis of fresh frozen and FFPE in the same study. PAXgene preserved samples provide a new source for the analysis of a wide range of biomolecules and have shown to provide superior results compared to restored FFPE samples (Andersen et al. 2014).

As all bisulfite-based analysis techniques, epigenotyping arrays are not able to differentiate between cytosine methylation and hydroxymethylation in a standard protocol. Specialized protocols based on oxidative bisulfite conversion or TET-assisted bisulfite analysis allow for the assessment of hydroxymethylation also on the BeadChip platform (Skvortsova et al. 2017; Stewart et al. 2015; Nazor et al. 2014).

The human Infinium arrays have also been evaluated and used for the analysis of DNA methylation in great apes and non-human primates, where due to their short phylogenetic distance to human 40–70% of the available probes can be used with high confidence (Hernando-Herraez et al. 2013; Ong et al. 2014; Pichon et al. 2021). Recently, an Infinium array specific for DNA methylation analysis of 285K CpG sites in mice was released, which will allow to accelerate DNA methylation profiling studies in different mouse strains and models of human disease (Garcia-Prieto et al. 2022). Similarly, a custom array analyzing 36K CpGs in highly conserved sequences common to a large number of mammalian species was recently devised and evaluated in multiple species (Arneson et al. 2022). It makes use of the tolerance of the Illumina probes to a certain number of mismatches in the complementary genomic sequence. While initially used to tolerate deviations from the expected co-methylation patterns, it also allows for maintaining efficient annealing in the presence of species-specific sequence variations. It is however not clear if this array will be available as an off-the-shelf product.

Despite their limited coverage of less than 3% of the 28 million CpG sites of the human genome, the 450K and EPIC array provide currently a good compromise between coverage, throughput, cost, resolution, and accuracy, permitting genome-wide epigenome analysis by epigenotyping. Therefore, this approach has been rapidly adapted by the community for epigenome-wide association studies for the analysis of a large variety of diseases and phenotypes. Indeed, the hundreds of thousands of associations between DNA methylation and diseases or phenotypic traits present in online repositories have been established based on the different generations of the Infinium arrays (Xiong et al. 2022; Huang et al. 2021; Battram et al. 2022).

16.6 Locus-Specific DNA Methylation Analysis

While all the above-described technologies are well suited for the identification of differentially methylated genes, most have inherent biases as described in the respective paragraphs, are still too costly, and/or do not provide the required analytical sensitivity and specificity for detailed locus-specific analyses. Large DNA methylation data sets from clinical samples are now available and can be mined for clinical associations in public databases or through the TCGA/ICGC project portals. However, identified DNA methylation-based biomarkers need to be validated and replicated using locus-specific methods for DNA methylation analysis. Furthermore, technologies for the potential use in a clinical setting have to be cost-effective, sensitive, and specific. It would be a non-negligible advantage if the method can also be applied to DNA extracted from formalin-fixed paraffin-embedded (FFPE) clinical specimens. Furthermore, it would be preferable if the analysis can be performed at a high-throughput scale and in a relatively short time span. In general, closed-tube assay formats

should be preferred to avoid cross contamination and thereby false-positive results. Following the interest in DNA methylation for the various clinical applications, many technologies have been developed for the quantitative analysis of DNA methylation patterns or levels at specific loci, mostly relying on the conversion of genomic DNA with sodium bisulfite (Table 16.2). While a large number of methods have been devised, relatively few methods are currently in use and are amenable to the throughput that is nowadays required. In this chapter, only the most commonly used methods including amplicon bisulfite sequencing, pyrosequencing, mass spectrometry, and real-time methylation-specific PCRs are described in detail. In a multi-laboratory and multi-technology benchmark study on the performance of locus-specific DNA methylation methods, results obtained through amplicon bisulfite sequencing, pyrosequencing, or mass spectrometry showed an excellent degree of agreement to each other as well as standards with a known degree of DNA methylation, demonstrating that assays for the analysis of locus-specific DNA methylation patterns are sufficiently mature to be used for routine analysis of DNA methylation-based biomarkers (BLUEPRINT Consortium 2016). When single-nucleotide resolution is not required, other methods such as *M*ethylation-*S*ensitive *H*igh-*R*esolution *M*elting analysis (MS-HRM (Wojdacz et al. 2008a)) or *M*ethylation-*S*pecific *M*ultiplexed *L*igation *P*robe *A*mplification (MS-MLPA (Nygren et al. 2005)) might be useful, for example, in case of screening for differentially methylated regions or if a method not relying on bisulfite treatment is required for validation (for MS-MLPA). MS-HRM (also termed *M*ethylation-*S*pecific *M*elting *C*urve Analysis, MS-MCA) makes use of the melting profiles of PCR products for a target locus originating from methylated and unmethylated variants of the same template, which differ after bisulfite treatment in their GC content. Therefore, the methylation status of an unknown sample can be determined by comparing the melting profile of the sample to calibration standards. A gradual increase of the temperature leads to a step-wise dissociation of the double strand in domains of the PCR product in function of their GC content differing between methylated and unmethylated molecules after bisulfite treatment. The application of this simple and cost-efficient technology, devised a long time ago (Worm et al. 2001), was previously limited by the toxicity of the intercalating agent *SYBR Green I* for DNA polymerases, which prohibited working at the required saturating concentrations. Advances in fluorescence detection technology, new algorithms for data calculation, and the use of novel dyes permitted the development of this high-resolution melting analysis (HRM (Wojdacz and Dobrovic 2007)), allowing for the rapid scanning of a large number of genes for the presence of differential DNA methylation. Careful primer design (e.g., including or not including cytosine residues) allows to fine tune the discriminative window to a specific range of DNA methylation levels, facilitating discrimination of samples with small differences (Wojdacz et al. 2009). MS-HRM can also be used on FFPE samples without a major loss of accuracy in the DNA methylation assessment (Daugaard et al. 2015). MS-HRM has been applied to the detection of aberrant methylation profiles in imprinting disorders (White et al. 2007; Alders et al. 2009; Wojdacz et al. 2008b), cancer (Balic et al. 2009; Gupta et al. 2014) and in epidemiological studies analyzing environmental exposure (Li et al. 2015c, 2016). Of note, this technology does not provide DNA methylation profiles at single-nucleotide resolution and yields only semi-quantitative results. Moreover, correlation to results obtained by other methods not relying on specific DNA methylation patterns was moderate (BLUEPRINT Consortium 2016), which will complicate its potential implementation in clinical diagnosis. Notably, the MS-HRM/MCA based assays failed to detect differential DNA methylation at a substantial proportion of loci supporting a use for qualitative purpose only (BLUEPRINT Consortium 2016). Of note, optimized protocols have been devised for the subsequent pyrosequencing of the amplification products,

Table 16.2 Technologies for the analysis of locus-specific DNA methylation patterns

Technology	The technology behind	Advantages	Inconveniences	Throughput	Limit of Detection or quantitative resolution	Suitability for cell-free DNA methylation analysis	References
Sanger Bisulfite sequencing	Bisulfite conversion, PCR amplification and in most cases cloning of PCR products. Sanger sequencing of clones and/or bulk amplification product	Standard equipment, single-nucleotide resolution, clonal methylation patterns (if cloned fragments are analyzed)	Quantitative resolution is usually low, high workload and cost-intensive when combined with cloning	+/++	~ 10–15%	–	Zhang et al. (2009)
Amplicon BS-Seq	Regions of interest are PCR amplified from bisulfite-converted DNA and libraries are constructed and sequenced by NGS (MiSeq/ IonTorrent)	Clonal fragments, high coverage, multiple amplicons and individuals can be analyzed in parallel, can be combined with microfluidics and/or droplets for PCR product preparation	Analysis requires some bioinformatic expertise, relatively expensive	+++	~ 1–5%	++	Korbie et al. (2015), Masser et al. (2013), Paliwal et al. (2013)
Pyrosequencing	Bisulfite conversion, PCR amplification and sequencing-by-synthesis method. Incorporation of a complementary nucleotide releases a phosphate group that will be involved in the ATP-driven oxidation of luciferin. The light emitted from this reaction is monitored	Highly quantitative, single CpG resolution	High workload and feasible only for a limited number of target regions	++	~ 5%	+/–	Tost and Gut (2007)
MassCLEAVE/ EpiTYPER	T7-promoter-tagged bisulfite specific PCR followed by ssRNA transcription and base specific cleavage by RNase A. The mixture is analyzed quantitatively by MALDI-TOF-MS	Allele-specific methylation can be addressed, amenable to high throughput and identification of heterogeneous DNA methylation patterns	Not all CpG positions can be analyzed, requires expensive instrumentation	+++	~ 5%	+	Ehrich et al. (2005), Coolen et al. (2007)

(continued)

Table 16.2 (continued)

Technology	The technology behind	Advantages	Inconveniences	Throughput	Limit of Detection or quantitative resolution	Suitability for cell-free DNA methylation analysis	References
MS-HRM	High-resolution melting curve analysis following quantitative real-time PCR RT-PCR yields quantitative methylation levels. Inclusion of a few CpG sites in the 5′ end of MIP primers allows for the amplification of unmethylated as well as methylated DNA	Rapid screening for the presence of differential DNA methylation	A limited number of CpG sites between the primers are necessary to obtain highly distinguishable melting curves but lead to biased amplification. Heterogeneous DNA methylation patterns are difficult to resolve	++	~5–20% (depending on primer design)	++	Wojdacz et al. (2008a)
MS-MLPA	MLPA primers containing an additionally restriction site are hybridized to the target of interest. Methylation-sensitive endonucleases will digest the probes hybridized to unmethylated targets. Thus, only methylated targets are subsequently amplified (Nygren et al. 2005)	There is opportunity for multiplexing. MS-MLPA resolves the copy number variations simultaneously with the methylation status. Does not rely on bisulfite conversion	Difficult primer design and assay optimization, depends on the presence of methylation-sensitive restriction enzymes	+/–	~ 10%	+/–	Nygren et al. (2005)
MS-qPCR	Digestion of DNA with methylation-sensitive restriction enzymes and subsequent amplification of undigested target DNA monitored quantitatively in real-time	Simple assay design, no special instrumentation, amenable to high throughput using microfluidic devices	Depends on the presence of recognition sites for methylation-sensitive restriction enzymes, Incomplete digestion gives false-positive results	+++	~ 5%	++	Oakes et al. (2006), Wielscher et al. (2015)
MS-SNuPE/ SNaPshot	The region of interest is amplified and labeled ddNTPs are used to extend a primer that terminates at the CpG site. The products are	Multiplexing assays enable investigation of several targets of interest	Only specific CpG positions are analyzed	–	~ 5%	–	Gonzalgo and Jones (1997), Kaminsky et al. (2005)

	visualized on a gel or by capillary electrophoresis (Gonzalgo and Jones 1997; Kaminsky et al. 2005)						
COBRA	Restriction enzymes are used to digest PCR products amplified from bisulfite-treated DNA, which contain a sequence difference in the enzymatic recognition site in function of the methylation status prior to the bisulfite conversion	Easy to implement and possibility for quantification	Gel based readout, targets may be missed due to inhibition of enzymatic activity	–	–10%	–	Xiong and Laird (1997), Brena et al. (2006)
MSP-PCR	At least one CpG site at or near the 3′ of the amplification primers ensures amplification of methylated DNA only. By including non-CpG cytosines in the primer sequence and having stringent annealing temperatures, amplification of incomplete converted- and non-target DNA is limited	MSP is a cost-effective approach, detects one methylated molecule in background of 1000 molecules	Qualitative readout, high rate of false positives and negatives	++	1%	++	Herman et al. (1996)
qMSP	Same as MSP, but amplification is followed in real-time	Sensitive technology, more cost-efficient than MethyLight while better control over false-positive and -negative results compared to MSP	Heterogeneous DNA methylation might lead to amplification failure and thus false negative results	+++	1%	++	Nikolaidis et al. (2012)
SMART-MSP	Same as qMSP followed by a melting curve analysis allowing for the exclusion of some false-positive results	Cost-effective, no probe required, and yields a low-rate of false positives	Heterogeneous DNA methylation might lead to amplification failure and thus false negative results or difficult to interpret melting curves	+++	1%	+++	Kristensen et al. (2008)

(continued)

Table 16.2 (continued)

Technology	The technology behind	Advantages	Inconveniences	Throughput	Limit of Detection or quantitative resolution	Suitability for cell-free DNA methylation analysis	References
MethyLight	Cleavage of a fluorescent hydrolysis probe designed to hybridize between MSP primers, principle based on the TaqMan principle. Detects one methylated molecule in background of 10,000 molecules	Highly sensitive; a reference assay (*Alu* or *β-Actin*) permits the quantification of the methylation level. Further improved sensitivity in a digital	Heterogeneous DNA methylation might lead to amplification failure and thus false negative results	+++	~0.1%	+++	Eads et al. (2000), Yu et al. (2015)
HeavyMethyl	Blockers, designed to bind unmethylated DNA only, ensure the amplification of methylated DNA with methylation independent amplification primers. A probe hybridizing to CpG sites containing a fluorophore and a quencher is designed to detect the amplification of methylated molecules	Highly sensitive; principle used in FDA approved DNA methylation assays	Sophisticated design requiring multiple probes for each assay	+++	~0.1	+++	Cottrell et al. (2004)

Abbreviations: *BS* bisulfite, *MS* methylation-sensitive, *MS-HRM* methylation-sensitive high-resolution melting, *MS-MLPA* methylation-specific multiplexed ligation probe amplification, *MS-SNuPE* methylation-sensitive single-nucleotide primer extension, *COBRA* combined bisulfite restriction analysis, *MSP* methylation-specific PCR, *SMART-MSP*, sensitive melting analysis after real-time methylation-specific PCR

which can then yield detailed information on the methylation status of the CpGs included in the amplification product (Candiloro et al. 2011).

MS-MLPA has been widely used for the diagnosis of imprinting disorders. While initially devised for the analysis of copy number alterations, this technique has been useful for the parallel analysis of up to 40 loci permitting a comprehensive analysis for all possible variations of DNA methylation aberrations in imprinting disorders (Dikow et al. 2007; Priolo et al. 2008; Henkhaus et al. 2012), the combined analysis of genetic and epigenetic alterations in imprinting disorders (Scott et al. 2008), as well as tumor analysis (Serizawa et al. 2010; Homig-Holzel and Savola 2012). Two oligonucleotides with universal primer binding sites are annealed to a target region or ligated in case of complete target complementarity. A methylation-sensitive enzyme is added to the ligation reaction digesting unmethylated templates and reducing the amount of ligated product. A semi-quantitative readout is then performed using capillary electrophoresis allowing the detection of methylation differences of 10% or more compared to the standards. While this technology does not rely on bisulfite conversion, it might yield false-positive results if the digestion is not complete and limits the applicability of the method to targets with restriction enzyme recognition sites.

PCR amplification following methylation-sensitive restriction digestion is an alternative strategy that requires substantially less DNA and no prior bisulfite conversion treatment, which is well suited as a rapid screening tool for differential methylation (Singer-Sam et al. 1990). Multiple targets can be simultaneously analyzed by locus-specific multiplex PCR following methylation-sensitive restriction digestion of genomic DNA (Melnikov et al. 2005). Additional information on the methylation status of a target region can be achieved by digesting the DNA with either methylation-sensitive restriction enzymes or methylation-dependent enzymes such as McrBC, thus allowing to distinguish complete methylation, partial methylation, or absence of methylation in the sequence (Yamada et al. 2004). Quantification can be improved by monitoring the increase in fluorescence by quantitative real-time PCR with intercalating dyes (Bastian et al. 2005; Oakes et al. 2006). Combined with microfluidic preparation of the PCR products, this method allows the analysis of a large number of target sequences from a very limited amount of starting DNA (Wielscher et al. 2015). To minimize false-positive results due to incomplete digestion, DNA is overdigested using a combination of several restriction enzymes and two or better three restriction sites have to be present within the target sequence.

16.6.1 Amplicon Bisulfite Sequencing

Established methods, such as pyrosequencing (Tost and Gut 2007) or mass spectrometry-based DNA methylation analysis (Ehrich et al. 2005), which are described in more details below, allow for the quantitative analysis of DNA methylation in a region of interest and are well suited for the analysis of a limited number of regions in large number of samples, but they are difficult to upscale if a large number of potential candidate regions identified in genome-wide analyses have to be verified. Amplicon bisulfite sequencing makes use of benchtop sequencers (e.g., Illumina's MiSeq) and allows generating high levels of coverage (e.g., 100's–1000's ×) that yield precise measurements of the quantitative levels of cytosine methylation. Due to the digital counting at methylation positions, amplicon bisulfite sequencing (together with padlock-based protocols) reported more frequently methylation values of either 0 or 100% in a benchmark study on the performance of locus-specific DNA methylation methods (BLUEPRINT Consortium 2016). Amplicon bisulfite sequencing using these instruments has become a widely used approach for the validation of genomic regions following methylome analyses and for answering hypothesis driven research questions. In addition, due to the sequencing of clonal clusters generated in the sequencing machine, these methods provide co-methylation patterns on individual molecules within the limits of the length of the reads (up to 600 base pairs in paired-end modus on the

MiSeq). With a current output of ~50M reads for the MiSeq, yielding between 3.8 and 15 GB of sequence depending on the used sequencing kit, several tens to hundreds of target regions can be analyzed simultaneously depending on the desired coverage and number of samples analyzed in parallel. In a large benchmarking study, amplicon bisulfite sequencing showed together with pyrosequencing the best overall performance (BLUEPRINT Consortium 2016). Their short run time, relatively low running costs and wide availability make them a valuable alternative for targeted DNA methylation analysis.

In general, PCR amplification products are prepared from bisulfite-treated DNA using a two-round amplification protocol with a first pair of target region-specific primers that contain tag sequences to label the created amplicons with sequences compatible for subsequent PCR amplification with the full-length Illumina p5 and p7 adaptor sequences. Molecular barcodes and full-length adaptor sequences are added in a second round of amplification after pooling all amplicons from a sample. If sample quantity is limited, the first amplification can also be performed as multiplex PCR, but this requires some more optimization for the multiplex set-up (Korbie et al. 2015). Another strategy consists of using conventional amplification primers and the molecular barcode and adaptors complementary to the sequences immobilized on the flow cell are subsequently added in a standard library preparation protocol after pooling of the PCR products from the first round of amplification (Jenkins et al. 2014).

An alternative protocol, termed Bisulfite Amplicon Sequencing (BSAS), makes use of the hyperactive Tn5 transposase (Nextera/Illumina) for random insertion of the sequencing primers in the amplification product after a PCR-based targeted amplification using conventional primers (Masser et al. 2013). Amplification of multiple amplification products in parallel can also be performed using microfluidic tools such as the Fluidigm access array, amplifying simultaneously 48 target regions in 48 samples starting from as little as 50 ng of DNA (Paliwal et al. 2013). The Access Array uses a two-step amplification-procedure where universal forward and reverse adaptor sequences are added to the 5′-ends of the gene-specific amplification primers. Sample barcodes and platform-specific sequencing primers are added in a second round of amplification. However, this approach is unsuitable for DNA extracted from FFPE samples (Korbie et al. 2015). Bisulfite-patch PCR is another approach enabling multiplex amplification by using a restriction enzyme digestion to anneal exonuclease-resistant patch oligonucleotides and universal primers complementary to the created overhang, while unselected fragments are eliminated by an exonuclease digestion (Varley and Mitra 2010). Following bisulfite conversion, fragments are amplified using sequencing platform-specific universal primers. Bisulfite-patch PCR has been shown to analyze up to 94 simultaneously amplified sequences with little off-target sequences, however, the requirement for a specific restriction site imposes serious limitations on the targets that can be analyzed in parallel. Nonetheless, multiple restriction enzymes can be used to select different sets of target sequences (Varley and Mitra 2010).

As for the genome-wide approaches for DNA methylation analysis, post-bisulfite adaptor tagging (PBAT) can also be applied to amplicon bisulfite sequencing approaches, allowing to start from as little as 10 ng of DNA and to reduce the number of amplification cycles compared to the standard library preparation protocol (Miura and Ito 2015). In general, bisulfite sequencing using next-generation sequencers with their digital readout enables a more accurate quantification of DNA methylation levels, as they show a reduced error of quantitation and lower standard deviations compared to the conventional (analog) sequencing approaches (5% vs. 5–20% (Masser et al. 2013)). Sequencing depth of ~1000× is sufficient for a precise measurement of the DNA methylation levels and increasing sequencing depth does not improve the accuracy further (Masser et al. 2013). However, accuracy is already very high (>99%) if only regions with a reasonable amount of coverage (i.e. > 50×) are used for base calling and quantitative

determination. Up to 96 samples with different regions of interest can be currently analyzed in parallel using conventional multiplexing strategies such as (dual) indexing. The introduction of longer (8 base) indices as well as dual indexing has further increased multiplexing capabilities. Of note, the MiSeq does already exist in a version certified for diagnostics, which potentially allows the analysis of DNA methylation-based biomarkers in a clinical setting. Specialized bioinformatics pipelines coming with the instruments or freely available pipelines such as Bismark (Krueger and Andrews 2011) or BiQ-Analyzer (Lutsik et al. 2011) enable convenient and standardized analysis of the sequencing results including the demultiplexation of individual samples, alignment to target regions, and estimation of the DNA methylation degree, allowing the analysis of DNA methylation in target regions without great bioinformatic expertise.

Ion Torrent's PGM sequencer has also been used for locus-specific DNA methylation analysis (Nones et al. 2014) including to sequence the methylation patterns of candidate genes in circulating cell-free DNA (Vaca-Paniagua et al. 2015).

16.6.2 Pyrosequencing

Pyrosequencing® (Ronaghi et al. 1998; Harrington et al. 2013) is a quantitative real-time sequencing method that is frequently used for the analysis of DNA methylation patterns and allows for the accurate measurement of methylation levels in a sequence of up to 100 bp (Tost and Gut 2007; Tost et al. 2003; Dupont et al. 2004). Pyrosequencing is based on the presence or absence of the incorporation of a nucleotide during primer extension (Ronaghi et al. 1998; Ronaghi 2001). In contrast to Sanger sequencing, which relies on the random incorporation of fluorescent ddNTPs during primer extension steps, only one specific nucleotide is present at any time in the pyrosequencing reaction. The pyrophosphate (PPi) released following nucleotide incorporation is used for the detection of incorporation. First, PPi functions as a substrate in combination with adenosine 5′ phosphosulfate (APS) for the ATP sulfurylase to produce ATP (Ahmadian et al. 2006). The latter is in turn used by luciferase to oxidize luciferin into oxyluciferin resulting in a light emission that is stoichiometrically proportional to the amount of incorporated nucleotide (Ahmadian et al. 2006). For DNA methylation analysis, a region of interest is amplified after bisulfite conversion with a standard PCR with one of the two primers being biotinylated. The biotinylated strand is captured on streptavidin-covered beads, the complementary strand is denatured and washed away, and a sequencing primer is annealed to the now single-stranded template before starting the pyrosequencing reaction. The methylation level is determined as the ratio of the signal corresponding to the incorporation of the nucleotides at a cytosine in a CpG dinucleotide corresponding to the methylated and unmethylated bases (i.e., C and T, or G and A when using a reverse primer). Thus, pyrosequencing signals report the average of all the molecules present in the reaction after amplification of the bisulfite-treated DNA. The limit of detection of pyrosequencing has been evaluated around 5% for the minor allele, which is far more sensitive than Sanger sequencing (Ogino et al. 2005). Additional advantages include the possibility to use cytosines outside CpG dinucleotides as an internal control for bisulfite conversion efficiency, short time to results and data analysis not requiring sophisticated bioinformatic expertise and resources. While the read-length of the pyrosequencing run is restricted to 100–120 base pairs, longer amplification products can be analyzed by serial pyrosequencing reactions, stripping of the DNA strand synthesized during the pyrosequencing reaction and annealing of a new sequencing primer (Tost et al. 2006). However, the low temperature (28 °C) at which pyrosequencing is performed due to the thermal instability of some enzymatic components, limits the length of the sequence that can be analyzed to ~300 base pairs, as longer amplification products with the low sequence complexity of bisulfite-treated DNA tend to form secondary structures

that increase background signals or impede binding of the sequencing primer. The minimally needed amount of input DNA required for pyrosequencing-based DNA methylation analysis allowing highly accurate quantification is ~10 ng (Dupont et al. 2004). Lower amounts might still yield useful information, but reproducibility should be verified using replicate measurements when possible. Furthermore, for absolute quantification the use of calibration curves mixing methylated and unmethylated DNA methylation standards is required. Pyrosequencing is also frequently used for the analysis of DNA from FFPE tissues (Newton et al. 2014); however, amplification products should be kept as short as possible in this application due to the extensive degradation of the DNA during fixation and bisulfite conversion. Pyrosequencing has shown high inter-laboratory reproducibility and correlation to amplicon sequencing making it well suited for DNA methylation-based biomarker development (BLUEPRINT Consortium 2016). Due to its quantitative accuracy, simplicity, and short time to results, pyrosequencing has become one of the most widely used technologies for locus-specific DNA methylation analysis. Following an oxidative bisulfite conversion, pyrosequencing can also be used to determine accurately the level of hydroxymethylation in specific regions of interest (Stewart et al. 2015).

16.6.3 MALDI Mass Spectrometry

Mass spectrometry (MS) provides an attractive solution for nucleic acid analysis in general and DNA methylation analysis in particular, as it enables direct, rapid, and quantitative detection of DNA products measuring the molecular weight, an intrinsic physical property of each molecule, rather than relying on an indirect readout, such as a fluorescent tag. Liquid Chromatography MS/MS is one of the most accurate methods to precisely quantify the global level of CpG methylation and its oxidative derivatives in samples of clinical interest (Berdasco et al. 2009; Godderis et al. 2015), while *M*atrix-*A*ssisted *L*aser *D*esorption/*I*onization time-of-flight *M*ass *S*pectrometry (MALDI-MS (Karas and Hillenkamp 1988)) has been among the most widely used instrumental platform for the analysis of DNA methylation patterns in specific regions of interest. The matrix usually is a low molecular weight organic acid with a strong absorption at the laser excitation wavelength. It contains the analyte molecules and is desorbed with a short laser pulse. The ionized nucleic acid molecules are extracted with an electric field and separated by their masses over charge ratio in the time of flight to a detector. The resolution of the current generation of MALDI mass spectrometers allows an easy distinction of nucleobase substitutions in the mass range of 1000–7000 Da, which corresponds to DNA sizes of 3–25 nucleobases. The methylation status is deduced from the proportional surface area of the peaks differing by 16 Da, corresponding to the difference between formerly methylated and unmethylated cytosine nucleotides after bisulfite treatment. The dynamic range of detection of MALDI-MS is between 2 and 3 orders of magnitude, yielding highly linear responses in titration experiments and low deviations (2–3%) from the expected values (Ross et al. 2000). Although amenable to very high throughput through a large degree of automation and highly parallel analyses, MALDI-MS does not permit genome-wide analyses as it requires the separate or multiplexed amplification of the different target loci. However, due to their multiplexing capabilities, the quantitative readout of the relative abundance of products, and a simple and reliable procedure the MALDI mass spectrometry-based assays are valuable tools for the identification and validation of methylation variable positions in a gene-targeted approach (Ragoussis et al. 2006). Therefore MALDI-MS-based methods with their single base resolution position themselves at the crucial follow-up stages for biomarker validation and large cohort analysis rather than biomarker discovery, as well as for large-scale investigations of candidate genes. Similar throughput for the analysis of specific CpG positions is not feasible with other available technologies for the analysis of single CpG positions.

The quantitative high-resolution scanning used in fragmentation-based approaches such as the EpiTYPER assay (Agena Biosciences, formerly Sequenom Biosciences) provides a quantitative readout for individual CpG sites in a target region of up to 600 base pairs with high accuracy (Ehrich et al. 2005). Due to the increased amplicon size compared to many other locus-specific analysis methods, this technology has been proven useful to define the boundaries in which differential DNA methylation patterns can be detected and to identify specific CpGs that have the greatest diagnostic potential. Starting from ~1 μg of bisulfite-treated DNA per sample, the region of interest is amplified using a reverse primer with an added T7-promotor sequence, which is subsequently used for in vitro transcription to generate single-stranded RNA, which is significantly more stable than DNA in MALDI analysis. The RNA transcript is then digested with an uracil-specific enzyme to create short DNA fragments of a few nucleobases to adapt the analyte size to the optimal detection window of the instrument which are then purified to remove counter-ions interfering with the MS analysis. The RNA fragments are loaded on a SpectroCHIP Array, a holder with hydrophilic anchors pre-loaded with a solution of the matrix, using a piezo-pipetting device and the matrix-embedded RNA fragments are subject to analysis on the MassARRAY Analyzer. Compared to other techniques that are able to achieve quantitative DNA methylation data on consecutive CpGs in a region of interest, its quantitative resolution of ~5% and a similar limit of detection for the minor methylation allele fraction is only rivaled by pyrosequencing (Tost and Gut 2007) and targeted bisulfite-sequencing approaches using NGS. The accuracy of the DNA methylation measurements notably exceeded the one of DNA methylation measurements with the Infinium array (BLUEPRINT Consortium 2016). In comparative studies, results obtained with the EpiTYPER do correlate well with those obtained by other quantitative methods such as pyrosequencing and bisulfite sequencing, but differ substantially from methylation-specific PCR (MSP)-based approaches (Alnaes et al. 2015; Claus et al. 2012). Systematic evaluation of each step in the workflow showed that most of the variability of the experiment was induced by either the bisulfite treatment or the subsequent PCR amplification, while the cleavage and the mass spectrometric analyses contributed much less to the variability in the observed quantitative measurements (Ehrich et al. 2007; Coolen et al. 2007). In more than 90% of the tested samples, a methylation difference of 10% was successfully detected with the mass spectrometric assay (Coolen et al. 2007). The procedure is amenable to the analysis of DNA extracted from fresh frozen, but also from FFPE tissues (Radpour et al. 2009). The EpiTYPER is one of the most widely used methods for the analysis of gene-specific DNA methylation patterns and it has been applied to the large-scale analysis of DNA methylation patterns in cancer (Radpour et al. 2009). Due to its high throughput capacities measuring 96–384 PCR products in parallel, it is also one of the most widely used methods for the validation of DNA methylation variations identified in Epigenome-Wide Association Studies (Zeilinger et al. 2013; Tobi et al. 2014; Zhang et al. 2014).

Once the methylation patterns have been characterized in detail, primer extension methods such as the commercial iPLEX assay (Ragoussis et al. 2006) can be employed to specifically target only those CpG sites with functional relevance or diagnostic potential in a specific biological context. The iPLEX assay offers routinely a degree of multiplexing in the low two digit range, but 27-plexes have been reported (Ragoussis et al. 2006). The key to this assay lies in the combination of the primer design for the upstream PCR and primer extension assay combined with the selection of terminating dideoxynucleotides. This epigenotyping method enables the multiplexed analysis of multiple CpG sites from different promoter regions making full use of the strength of the mass spectrometer in automation and throughput. Variation of primer extension assays using competitive primer extension with oligonucleotide standards has also been used for the absolute quantification of fetal DNA in maternal plasma with high analytical sensitivity and specificity amplifying specifically

hypermethylated fetal DNA resistant to a prior methylation-sensitive restriction digestion (Nygren et al. 2010).

16.6.4 Methylation-Specific PCR and Its Quantitative Variations

Methylation-specific PCR (MSP) and its quantitative real-time variations allow detecting methylated molecules in the presence of an excess of normal (and usually unmethylated) DNA (Herman et al. 1996). MSP allows the amplification of virtually any CpG sites after bisulfite treatment with three pairs of primers for amplification: complementary to the former methylated, the former unmethylated sequences, or to genomic, unconverted DNA, respectively (Herman et al. 1996). The latter could serve as control for complete bisulfite conversion, but is in practice rarely included in the experimental design. Primers need to hybridize to sequences with at least two methylation variable positions (CpGs) to obtain the necessary specificity for selective amplification. The presence or absence of an amplification product analyzed on a conventional agarose gel reveals the methylation status of the CpGs underlying the amplification primers (Fig. 16.2). MSP has been the most widely used technology for DNA methylation analysis, as it does not require any expensive instrumentation and a large number of samples can be rapidly assessed. The main advantage of MSP is the high sensitivity, which enables the detection of one allele in the presence of a 1000-fold excess of the other (Herman et al. 1996). However, MSP does not provide resolution at the individual nucleotide level and heterogeneous methylation patterns at the primer binding sites can induce failure of amplification (Alnaes et al. 2015; Vinarskaja et al. 2012), which contributes to an overall lower concordance compared to sequencing-based methods (BLUEPRINT Consortium 2016). Furthermore, the biased amplification leads to a more qualitative than quantitative result, making it difficult to distinguish different degrees of methylation at the target sites. It has been shown in comparative studies that MSP overestimates DNA methylation levels and due to a significant number of false-positive (and negative) results as well as the dichotomized result of MSP, associations between aberrant DNA methylation and clinical parameters are less easily identified in MSP data compared to quantitative DNA methylation technologies (Claus et al. 2012).

Quantification can be improved and false-positive results reduced by analyzing the MSP product after amplification in a real-time thermocycler by high-resolution melting analysis, an approach, which has been termed SMART-MSP (Kristensen et al. 2008, 2009) for *S*ensitive *M*elting *A*nalysis after *R*eal Time (Fig. 16.2). This approach makes use of a high-resolution melting analysis, which has been described in detail above and might be an alternative to the below described real-time approaches, if no probe can be designed or if the DNA methylation patterns of the amplification product are expected to be heterogeneous complicating the prediction of the annealing behavior of the probe. Furthermore, SMART-MSP might enable detection of DNA methylation patterns in samples difficult to amplify using standard HRM or with low levels of methylation.

Real-time PCR-based methods for DNA methylation analysis, such as MethyLight (Fig. 16.2), use the same principle as the TaqMan® assay (Holland et al. 1991). In addition to the two amplification primers, a probe which is dually labeled with a fluorescent reporter and a quencher dye hybridizes to a target sequence in the amplified region and improves sensitivity, as well as specificity compared to conventional MSP. The simple one-step procedure makes real-time methylation-specific PCRs rapid high-throughput assays for quantitative DNA methylation analysis, which are robust and quite resistant to carryover contamination. These approaches do not only provide information whether molecules with a certain methylation pattern are present in the sample—like conventional MSP—but also report on the fraction of them. Discrimination between methylated and unmethylated alleles can be achieved at different levels based on the primers and/or the hybridization probe (Eads et al. 2000). Although in principle primers and probes could be designed for different

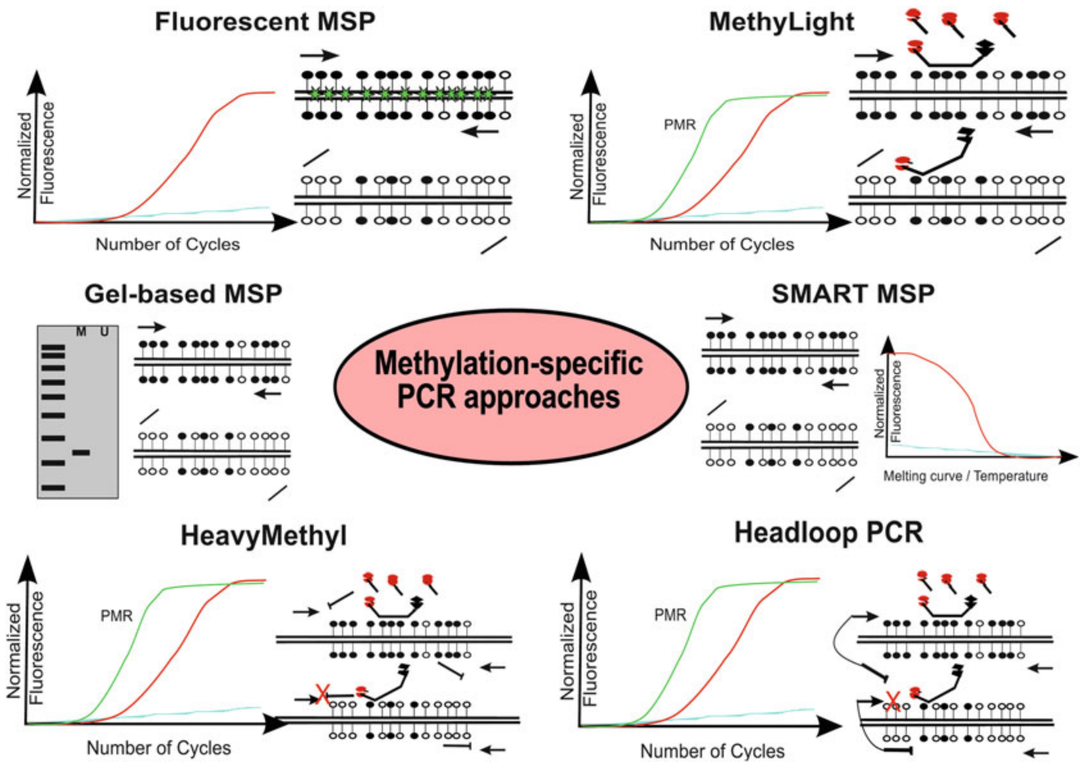


Fig. 16.2 Methylation-specific PCR methods for the sensitive detection of DNA methylation. For simplification, only amplification with a primer complementary to a completely methylated allele is shown. PMR: percentage of methylated allele, a completely methylated DNA standard that is used for the calculation of the percentage of methylation contained within a sample. All methods use sodium bisulfite treatment prior to PCR amplification. CpGs are depicted as lollipops, former unmethylated CpGs are shown as empty lollipops while filled ones correspond to former methylated CpGs. Although methylation is retained as a sequence difference after bisulfite treatment, lollipops are shown for easier differentiation of the alleles. Details of the different techniques are given in the text

combinations of methylated and unmethylated alleles, the most widely used approaches such as MethyLight use primers and probes that are specific for the same methylation patterns, mostly completely methylated molecules. Heterogeneous methylation patterns that display large variations between consecutive CpGs complementary to the primers or the probe will, however, also lead to a failure of the assay or biased quantitative results (Mikeska et al. 2010; Alnaes et al. 2015).

MethyLight can detect a single hypermethylated allele against a background of 10,000 unmethylated alleles (Eads et al. 2000; Lo et al. 1999). Absolute quantification of the number of molecules corresponding to the investigated pattern of methylation is achieved by measuring the ratio between the gene of interest and a reference gene, for example the *β-actin* (*ACTB*) gene. MethyLight yields highly precise and reproducible results with an average variation of ~0.8%, with slightly larger variations induced by different bisulfite treatments (Ogino et al. 2006). Nonetheless qMSP/MethyLight assays achieve an overall lower concordance compared to sequencing-based methods and there is a proportion of assays where these technologies detect DNA methylation changes, but with the opposite direction compared to assays providing absolute DNA methylation levels (BLUEPRINT

Consortium 2016). About twenty-fold-increased sensitivity (limit of detection and limit of quantification) can be obtained if MethyLight is not performed by conventional real-time, but by digital droplet PCR (ddPCR) (Wiencke et al. 2014; Yu et al. 2015). Furthermore, with ddPCR the quantitative accuracy is increased about sevenfold for some assays (Wiencke et al. 2014). The concept of using MethyLight on single molecules was devised some time ago using limiting dilution to perform the individual MethyLight reactions (Weisenberger et al. 2008). The use of commercially available ddPCR machines has significantly streamlined and simplified the assay procedure and the readout. This has also accelerated the development of clinically relevant DNA methylation assays based on ddPCR (Beinse et al. 2022; Vedeld et al. 2022; Sontag et al. 2022). Furthermore, absolute quantification of methylated alleles can be achieved by simply counting the positive droplets in ddPCR, and the use of Poisson statistics without the need for a standard curve. Nonetheless, internal reference standards can increase the accuracy of DNA methylation-based measurements (Pharo et al. 2018). Hence, amplification efficiency is less of a concern compared to the conventional MethyLight. This method might therefore be better suited if very few methylated alleles are expected to be present in a clinical sample. As ddPCR enters clinical laboratories for various applications including mutation detection and quantification, these instruments will probably be widely available. Additionally, its simplified technical use and the described advantages make the technology one of the promising approaches for DNA methylation analysis entering the field of personalized medicine. Nonetheless, these assays are significantly more expensive than normal MethyLight assays and the throughput with current ddPCR machines is much reduced.

The addition of a second probe marked with a different fluorescent dye to the qPCR-based MethyLight assay allows for the simultaneous detection of unconverted sequences that might co-amplify with the bisulfite-converted molecules avoiding potential false-positive results (ConLight (Rand et al. 2002)). Quantitative analysis of methylated alleles (QAMA) uses a TaqMan probe conjugated to a minor groove binder for discrimination at single base level by forming hyperstabilized duplexes with complementary DNAs (Zeschnigk et al. 2004). Methylated and unmethylated alleles are simultaneously quantified using two probes modified with two different fluorophores. Thus, the amplification of the bisulfite-treated DNA can be carried out with primers amplifying simultaneously the formerly methylated and unmethylated alleles and differentiation of the methylation status of alleles is achieved only at the probe level.

HeavyMethyl further increases the sensitivity and specificity of real-time PCR-based assays for the analysis of DNA methylation using methylation-dependent blocking oligonucleotides (Cottrell et al. 2004). In contrast to MethyLight, amplification primers are not specific for a certain methylation pattern, but positioned in sequence stretches containing no CpG positions (Fig. 16.2). Only the fluorescent probe is specific usually to a consistently hypermethylated sequence. The increased specificity and sensitivity are achieved through a second pair of non-extendable (3′ phosphorylated) oligonucleotides that hybridize specifically to a methylation pattern opposite to the investigated one, usually the unmethylated sequence. The annealing sites of these oligonucleotides overlap with the target sequences for PCR amplification and thereby efficiently block any amplification of the bisulfite-converted sequence corresponding to the undesired methylation pattern. HeavyMethyl was able to specifically detect 25 pg of in vitro methylated DNA in the background of 400 ng of unmethylated DNA (relative sensitivity up to 1: 8000). Similar to MethyLight, a reference standard is used to identify samples with negligible amounts of methylation. The use of four to five different oligonucleotides contributes significantly to the cost of the assay and the design might be more complex compared to the conventional MSP or MethyLight. However, HeavyMethyl shows the necessary sensitivity and specificity required for clinical applications and it is the underlying principle of the commercial DNA methylation-based diagnostic tests for

the (early) detection of colorectal cancer (Epi proColon, Epigenomics AG, (Church et al. 2014), targeting DNA methylation in the second intron of the *Septin9* gene) or lung cancer (Epi proLung, Epigenomics AG, (Ilse et al. 2014), targeting *SHOX2*).

An alternative approach uses amplification primers that carry a 5′-tail sequence complementary to a sequence that is present in the amplicon corresponding to a specific methylation pattern (Headloop PCR (Rand et al. 2005)). After incorporation of the primer in the synthesized PCR product, the tail folds back onto the template creating a secondary structure refractory to amplification (Fig. 16.2). For example, if the tail is complementary to an unmethylated sequence, only methylated molecules are amplified. The amplification is monitored in real time with *SYBR Green* or TaqMan probes. The sensitivity of the approach is similar to the others described above detecting a methylated allele in the presence of a 4000-fold excess of unmethylated ones.

Instead of using primers complementary to a specific methylation pattern, the *CO*amplification at *L*ower *D*enaturation temperature (COLD)-PCR-based approach makes use of the different GC content of the unmethylated and methylated molecules after bisulfite conversion. By lowering the denaturation temperature of the PCR, only the unmethylated molecules, which have a lower GC content, will be efficiently amplified, leading to an enrichment of the unmethylated sequences in a context of mainly methylated molecules (Castellanos-Rizaldos et al. 2014). Methylated molecules can be enriched using Enhanced (E)-*ice*-COLD-PCR (How Kit et al. 2013) and subsequently analyzed at single-nucleotide resolution using pyrosequencing (Mauger et al. 2018). In this technology, locked nucleic acid blocker probes are designed to prevent amplification of unmethylated bisulfite-converted DNA, thereby preferentially enriching the amplification of methylated DNA molecules during the E-*ice*-COLD-PCR reaction. The design of E-*ice*-COLD-PCR assays is fundamentally different from qMSP, MethyLight, or HeavyMethyl assays as it impedes the amplification of the normal, unmethylated state, but does not make any requirements on the degree or the patterns of DNA methylation in the amplified target region. All molecules different from the blocked pattern will be amplified and the subsequent sequencing-based analysis gives detailed information on the molecules that have been enriched. This difference suggests that E-*ice*-COLD-PCR could have an increased sensitivity, as it does not require any specific co-methylation patterns in the analyzed region, which would make it well suited for DNA methylation analysis in cell-free DNA.

As the amplicons of the different MSP variants are small (~100 bp), these methods usually work well with DNA of lower quality, such as DNA extracted from FFPE samples (Herman et al. 1996). No special equipment is required for the conventional methylation-specific PCR and real-time PCR machines are nowadays available at most research institutions for the quantitative analysis of gene expression. The similarity of the approaches described above to the real-time expression analysis also facilitates implementation of the technology, execution of the experiments, and interpretation of the results in laboratories not yet very familiar with DNA methylation analysis. The design of assays and optimization of amplification are probably the most important steps for MSP assays with notably the inclusion of CpGs in the 3′ end of the amplification primers to ensure specific primer annealing and efficient amplification of the desired locus (Li and Dahiya 2002; Brandes et al. 2007). Sensitivity and specificity vary largely between assays depending on the primers, conditions, and probes in case of techniques like MethyLight. Quantitative MSP assays allow high-throughput screening of a large number of clinical specimens in a single PCR step without complicated downstream analysis. Multiplexing with methylation independent controls normalizes for DNA input and parallel processing of calibration standards allows the assessment of run-to-run variability. However, these assays do not provide information on the DNA methylation level of individual CpGs and heterogeneous DNA methylation patterns can lead to a high rate of false-positive results (Mikeska et al. 2010; Alnaes et al. 2015; Claus et al. 2012).

16.7 DNA Methylation Analysis of Circulating Cell-Free DNA

DNA methylation of circulating cell-free (ccf) DNA has received a lot of attention in recent years, because of its potential as a stable and amplifiable biomarker for early diagnosis, prognosis, or response to treatment in various cancers and potentially other complex diseases (Beltran-Garcia et al. 2019; Taryma-Lesniak et al. 2020; How Kit et al. 2012). Biomarkers capable of distinguishing a disease state from healthy individuals must be specific, sensitive, and detectable in specimens obtained through minimally invasive procedures to be clinically applicable. Disease specific DNA molecules can be found in various body fluids, such as urine or sputum, or as ccf DNA molecules that can be isolated from the serum/plasma of cancer patients (Ignatiadis et al. 2021; Chakravarty and Solit 2021; Pantel and Alix-Panabieres 2019; Heitzer et al. 2019), individuals with autoimmune diseases (Chan et al. 2014), as well as individuals with many other complex diseases and physiological conditions under which cells undergo apoptosis and shed DNA molecules into the bloodstream (Lehmann-Werman et al. 2016; Zemmour et al. 2018; Moss et al. 2018). There are numerous applications for the analysis of DNA methylation patterns and changes in cell-free DNA, including the early detection and determination of the origin of a cancer, making use of the tissue-specificity of DNA methylation patterns, prognosis as well as the surveillance of cancer patients including the detection of minimal residual disease, relapse and response to treatment (Moss et al. 2018; Luo et al. 2021). Particularly, the widespread occurrence of DNA methylation changes during early phases of cancer is hoped to overcome sensitivity problems due to mutational heterogeneity in some cancer types (Luo et al. 2021). Furthermore, the analysis of DNA methylation changes should be less susceptible to confounding by polymerase-based errors, which might lead to false-positive point mutation calls.

A more specialized area for the analysis of DNA methylation patterns in cell-free circulating DNA is prenatal diagnosis. Analysis of differences in the DNA methylation patterns between the maternal and fetal circulating DNA molecules has been proposed as an alternative strategy to the analysis of DNA sequence-based variations (Lo et al. 2021; Sun et al. 2018; Papageorgiou et al. 2011). Circulating cell-free fetal DNA is shed through apoptosis from the placenta and can be isolated from maternal plasma or serum from 5 weeks gestation. A major analytical problem for the analysis of the circulating cell-free fetal DNA is the high risk of false negative results due to failure to extract or detect sufficient material and/or due to large individual variability in the total amount of cell-free DNA and the contribution of the fetal component to this total amount. Several studies have identified a number of genomic regions that are differentially methylated between the fetal DNA derived from the placental tissue and the maternal peripheral blood mononuclear cells (Chim et al. 2008; Papageorgiou et al. 2009). This epigenetic information can be used for diagnostic purposes by isolating the fetal DNA using methylation-sensitive restriction enzymes (Chim et al. 2005).

Initially MALDI mass spectrometry has been used to analyze placenta-specific DNA methylation profiles of genes located on different chromosomes to accurately determine the fetal portion of the circulating cell-free DNA isolated from maternal plasma and presented the first universal biomarker for fetal DNA quantification (Nygren et al. 2010). Target regions are co-amplified in the presence of known quantities of synthetic templates differing by a single nucleotide from the target regions enabling the accurate quantification of the total number of fetal copies, as well as the fraction of the fetal DNA in the maternal plasma with high specificity and sensitivity (100% and 99%, respectively) (Huang et al. 2006; Tsui et al. 2005).

Methods for the detection of DNA methylation in body fluids need to be highly sensitive as the target DNA molecules are present at only very low concentrations among an excess of DNA from healthy cells. Furthermore, only part of the molecules will carry the DNA methylation patterns of interest. Therefore, high analytical sensitivity of the method is crucial to detect the

low levels of tumor derived aberrantly methylated DNA molecules that are present in these specimens. The analytical specificity, i.e. the frequency of false positives obtained with the method, is also primordial for the use of a method in the clinics. It is therefore important to verify that the target regions are not methylated even at low levels in leukocyte DNA. Methylation-specific PCR (Hoon et al. 2004) and particularly methylation-specific real-time PCR-based methods such as MethyLight (Campan et al. 2011; Begum et al. 2011) and its ddPCR implementation (Yu et al. 2018a; Shemer et al. 2019), HeavyMethyl (Church et al. 2014), as well as Enhanced-*ice*-COLD-PCR (Mauger et al. 2018), methylation-sensitive high-resolution melting analysis (MS-HRM) (Yang et al. 2015a), and MRE-qPCR (Wielscher et al. 2015) have proven suitable for the detection of very low levels of aberrant methylation in circulating DNA (see also Table 16.2). Compared to the analysis of mutations in ccfDNA, the analysis of DNA methylation patterns has the advantage that DNA methylation changes are occurring at multiple CpGs in a target region, the analysis is therefore less likely to be influenced by polymerase errors during PCR (Goldstein et al. 2017).

The commercial Epi *pro*Colon test, which has been approved by the Chinese FDA in July 2015 and the US FDA in April 2016 uses the HeavyMethyl technology. It analyzes methylation in the *SEPT9* gene in cell-free circulating DNA enabling the population-wide screening for colorectal cancer (Church et al. 2014; Warren et al. 2011). Gene-specific assays analyzing DNA methylation changes in cell-free circulating DNA have recently been reviewed in detail (Warton and Samimi 2015).

MeDIP allows the analysis of DNA methylation in cell-free DNA avoiding further degradation through the bisulfite conversion, which makes it well suited for low-input analyses (1–10 ng) (Shen et al. 2018). Nonetheless the inherent problems of MeDIP remain such as the lack of single-nucleotide resolution and a semi-quantitative precision. Methylated DNA can also be enriched using a methyl-binding protein as in the MBD-seq approaches prior to NGS, but this requires relatively large volumes of plasma to obtain the required starting amount of 50 ng of cell-free DNA (Warton et al. 2014).

Next-generation sequencing approaches are becoming more and more used to identify and monitor the presence of mutations in cell-free DNA isolated from plasma (Newman et al. 2014; Crowley et al. 2013; Heitzer et al. 2015). Sequencing approaches of bisulfite-treated DNA isolated from plasma or serum are complicated by the fact that the bisulfite treatment will further degrade the DNA fragments reducing the amount of amplifiable DNA, but a number of successful examples have now been demonstrated (Chan et al. 2014; Lun et al. 2013). Most of the approaches for the whole-genome bisulfite sequencing described in this chapter are compatible with the DNA amount that can be obtained from a few mL of plasma, and the only adaption that is required is the omission of the fragmentation step as the cell-free DNA is already in a convenient size range. Bisulfite sequencing of the fetal methylome has been demonstrated using genetic differences to separate maternal and fetal derived molecules after sequencing. It allowed the analysis of ~100,000 loci and covered ~200,000 CpGs showing a strong resemblance of the fetal and placental methylomes allowing to reconstruct the fetal methylome from maternal plasma (Sun et al. 2018; Lun et al. 2013).

With the ample resources on DNA methylation patterns in a broad range of cancer types being available through the TCGA and ICGC portals, a number of studies and training algorithms have addressed the sensitive detection of different cancer types as well as to determine the tissue of origin of the detected methylated molecules (Kang et al. 2017; Li et al. 2018). While initial studies were limited both in scope, cancer types analyzed and number of patients included, they showed a promising diagnostic performance (Chan et al. 2013). Shallow WGBS sequencing (10M reads per sample) of ccfDNA outperformed the analysis of repetitive elements such as LINE1 retrotransposons in cell-free DNA (Tangkijvanich et al. 2007). Differences were assessed by binning of the methylation densities

(reads) in 1MB intervals and comparing methylation levels between patients and controls. The Circulation Cell-free Genome Atlas (CCGA) has performed WGBS and large-scale targeted DNA methylation Analysis in thousands of individuals showing improved of DNA methylation-based classifiers compared to genetic or copy-number variation-based classifiers with notably an improved limit of detection (Liu et al. 2020b). Based on a training and validation set of more than 2600 ccf samples (1493 cancer; 1135 non-cancer), a capture assay targeting more than 1,000,000 informative CpG positions for both the detection of cancer and the determination of the tissue of origin was devised and validated in more than 4000 additional ccf samples containing both cancer and non-cancer samples at equal proportions yielding a specificity of >98% and a stage-dependent sensitivity between 39 and 92%, with the tissue of origin correctly predicted in 93% of samples which scored positive for a cancer (Liu et al. 2020b).

16.8 Single-Cell DNA Methylation Analysis

Due to the large divergence of DNA methylomes between cell types and the cellular heterogeneity of tissues, the recent advances in single-cell omic technologies (Lynch and Ramalingam 2019; Mereu et al. 2020; Pierce et al. 2021; Chen et al. 2021; Ding et al. 2022; Melnekoff and Lagana 2022; Ren et al. 2018) and the insights gained from these studies have raised a lot of interest for single-cell DNA methylation analysis.

Working with single cells requires capturing individual cells. This is efficiently achieved with state-of-the-art cell isolation techniques. Sample preparation techniques for DNA methylation have been combined with multiplex assays using the SCRAM assay, which applies methylation-sensitive restriction enzymes after single-cell isolation and lysis, and locus-specific qPCR on a Fluidigm Biomark system, which allows the interrogation of 24 genomic positions in 48 single cells in one experiment (Lorthongpanich et al. 2013; Cheow et al. 2015). By using methylation-sensitive restriction enzymes, this approach avoids the degradation of the DNA by the bisulfite conversion, which can impede subsequent amplification if only little material is available, as in the case of single cells. A single-cell DNA methylation analysis method that combines bisulfite conversion with Sanger sequencing and Agena's Epityper also allows cost-effective analysis of a larger number of single cells (Gravina et al. 2015).

Techniques combining single-cell analysis with second generation sequencing have become a major field of development (Karemaker and Vermeulen 2018). Apart from the technical challenges, the main issue with single-cell analysis is the number of cells that need to be analyzed to obtain a representative picture of the biology and the amount of sequencing that can be afforded per cell. A balance needs to be found between these two parameters. Single-cell whole-genome bisulfite sequencing has been demonstrated using the PBAT approach and a whole-genome pre-amplification step, reducing the loss of DNA due to the fragmentation during bisulfite treatment (Smallwood et al. 2014). However, even with a large sequencing effort of 20M reads per cell, this approach has limited genomic coverage per cell, as it only recovers 8.5–48.4% of all CpG positions for each cell analyzed. This means that far more cells need to be included to capture sub-populations and large oversampling is necessary. On the other hand, it allows capturing non-CpG methylation. The pre-amplification step can be omitted, as shown in μWGBS approach, but this yields libraries with lower complexity and therefore lower coverage (Farlik et al. 2015). An alternative to preparing libraries for sequencing from individual cells, is to generate pools of a small, defined number of cells, sequence each pool and then use computational methods to deconvolute cell states due to the distortion in methylation detected between pools (Farlik et al. 2015). RRBS has also been applied to single cells and is a good match in terms of genomic coverage (~1.5 M CpGs), suitability with low input and reduction of the target size, allowing to assess the cell-to-cell heterogeneity of DNA methylation patterns, if no comprehensive

coverage of the genome is required (Guo et al. 2013, 2015). It also allows for the quantitative analysis of differential DNA methylation (Wang et al. 2015b), however a number of developmentally genomic regions such as imprinted genes are not sufficiently covered by this approach (Guo et al. 2013). Tailing and ligation-free single-cell DNA methylation analysis (TAILS) is based on bisulfite conversion, followed by a first round of random priming, dC tailing and a second round of target priming (Gu et al. 2019).

Multi-omic approaches such as COOL-seq, analyzing DNA methylation and chromatin accessibility in single cells, scTRIO-seq analzying mutations, the transcriptome and methylome of single cells (Bian et al. 2018), scM&T analzying the transcriptome and methylome of single cells (Angermueller et al. 2016), or scChaRM-seq analyzing the transcriptome, methylome, and chromatin accessibility (Yan et al. 2021) allow for the investigation of epigenetic and transcriptomic heterogeneity, as well as within cell correlation between multiple molecular properties. These approaches can be used to decipher the regulatory circuits of epigenetics in individual cells. With the increasing availability of scDNA methylation data, specific computational tools (Kapourani et al. 2021; Danese et al. 2021; Kapourani and Sanguinetti 2019) and data repositories, such as scMethbank (Zong et al. 2022), have been set-up providing the necessary infrastructure for the exploration of available single-cell methylation datasets. Nonetheless, compared to scRNA-seq or scATAC-seq, only a few computational tools have been developed and no standard analysis process has currently been established for genome-wide single-cell DNA methylation analysis. Furthermore, the available tools have not been thoroughly benchmarked against each other suggesting that there is need for further development and improvement.

Complementary to the above-described single-cell technologies, DNA methylation has also been detected in nanofluidic channels at the single-molecule level using fluorescent labeling of methylated DNA with MBD1 (Cipriany et al. 2010; Cipriany et al. 2012). This method combined with fluorescence-activated cell sorting, termed SCAN for *S*ingle *C*hromatin molecule *A*nalysis at the *N*anoscale, allows also selecting molecules with a distinct epigenetic pattern for further analysis using more resolutive technologies, such as qPCR and potentially NGS (Cipriany et al. 2012). It also offers the possibility of the simultaneous analysis of DNA methylation and chromatin modifications, as the labeling of the DNA and size of the nanochannels allow to isolate DNA together with its native chromatin, as described in more details in Sect. 16.11.

Although all these methods do provide a picture of the DNA methylation state at a given time point, they do not allow monitoring the dynamics of the DNA methylation at the single-cell level. This could be achieved with a GFP coupled-reporter construct for locus-specific DNA methylation changes with single-cell resolution that can be inserted at specific loci using the CRISPR/Cas9 technology (Stelzer et al. 2015). The methylation state of the surrounding sequence will influence the methylation state of the reporter construct, thereby allowing tracing the methylation level of the endogenous sequences and its dynamic changes during development and disease. Similarly, bimolecular anchor detector sensors, that were individually non-fluorescent constructs of proteins recognizing epigenetic modifications in living cells that yield fluorescence signals only when brought together complementing each other to form the fluorescent protein, can be used to determine the localization or co-localization of epigenetic marks at specific sequences and their dynamic behavior over time (Lungu et al. 2017).

16.9 Analysis of Cytosine Hydroxymethylation

Although cytosine 5-hydoxymethylation has been known since several decades, it has recently attracted much more attention, as it constitutes an intermediate in the active DNA demethylation process. Beyond this function, 5hmC is thought to play an active role in the regulation of gene

expression and altered patterns of hydroxymethylation have been found in different diseases, notably cancer and neurodegenerative diseases (Marshall et al. 2020; Lopez et al. 2017; Thomson and Meehan 2017; Xu et al. 2021b; Armstrong et al. 2019). In general, the total levels of 5-hydroxymethylcytosine observed across genomes are approximately 14-fold lower compared to 5-methylcytosine, although large variations between tissues exist (Globisch et al. 2010; Wagner et al. 2015; Ruzov et al. 2011). Global levels of 5hmC are routinely measured by either mass spectrometry, high performance liquid chromatography, dot blot or ELISA assays (Kriaucionis and Heintz 2009; Globisch et al. 2010; Wagner et al. 2015; Nestor et al. 2015; Chowdhury et al. 2014; Olova 2021). With the surge in interest to determine the exact location and relative abundance of hydroxymethylation, several technologies have been developed permitting its genome-wide or locus-specific analysis (Table 16.3, Fig. 16.3). However, although the potential confounding of DNA hydroxymethylation and methylation by bisulfite-based methods has been recognized early on (Nestor et al. 2010; Jin et al. 2010), most studies still tend to ignore this problem for both reasons of cost, but also because of the lower prevalence of 5hmC in the tissues.

A major difference between 5-methylcytosine and 5-hydroxymethylcytosine is the possibility of the latter to be modified by glycosylation, which is used as the underlying key principle of differentiation in several technologies (Fig. 16.3). The bulky adduct resulting from the glycosylation will protect modified 5hmC from deamination by the APOBEC cytidine deaminases as well as further oxidation by TET enzymes (Yu et al. 2012, 2018b; Vaisvila et al. 2021; Schutsky et al. 2018). Furthermore, 5hmC is a naturally occurring base in some bacteriophages and in these organisms it is often further modified by glycosylation by glycosyltransferases to protect against digestion by restriction endonucleases present in the host (Vrielink et al. 1994). However, a number of restriction enzymes have recently been identified which specifically recognize and cleave 5hmC-containing sequences after glycosylation, with PvuRts1I being the first enzyme identified (Fig. 16.3) (Borgaro and Zhu 2013; Wang et al. 2011). These enzymes cleave at a defined distance, normally 11–13 nucleotides 3′ from the modified cytosine. Aba-seq uses the enzymatic properties of AbaSI (AbaSDFI), a member of the PvuRts1I restriction enzyme family shown to exhibit high and improved specificity for 5hmC over 5mC and unmethylated cytosines compared to PvuRts1I (Wang et al. 2011). Starting with 2 μg of DNA, 5hmCs are glycosylated and subsequently cleaved, prior to the ligation of biotinylated adaptors, which allow the subsequent capture of sequences containing hydroxymethylated fragments. While initially only locus-specific analyses have been performed using this approach (Wang et al. 2011), its applicability to genome-wide analyses using NGS starting from as little as 100 ng of input DNA has also been demonstrated (Sun et al. 2013; Gross et al. 2015). This method allows to cover ~58% of all potentially hydroxymethylated cytosines in the genome (Wang et al. 2011). In contrast to the affinity-based enrichment methods, where the read numbers correlate with the density of hydroxymethylation, Aba-seq might be better suited for the identification of regions containing few hydroxymethylation marks, despite its limitation in coverage (Sun et al. 2013).

In addition, conventional restriction enzymes such as MspI can be used to differentiate between the two nucleosides after glycosylation of 5-hydroxymethylcytosines as the endonuclease activity is blocked by the glucosylation, but not methylation at the CpG cytosine (Kinney et al. 2011). *R*educed *R*epresentation 5-*H*ydroxymethylcytosine *P*rofiling (RRHP) makes use of the same principle (Petterson et al. 2014). Genomic DNA is digested twice with MspI; digestion of the DNA is separated by a 5hmC glucosylation step prior to the size selection and sequencing. RRHP exploits the β-glucosyltransferase to inhibit the enzymatic cleavage of the adapters ligated to a genomic library, allowing only fragments with glucosylated 5hmC residues at adapter junctions to be amplified and sequenced, thus providing a

Table 16.3 Technologies for the genome-wide analysis of 5-hydroxymethylcytosine

Technology	The technology behind	Single-nucleotide resolution (yes/no)	Absolute quantification (yes/no)	DNA input	Required number of reads per sample in M	Coverage of CpGs (theoretical)	Comments	References
Aba-seq	Glycosylation of 5hmC, followed by restriction digestion with AbaSI, ligation of biotinylated adaptors, size reduction of fragments and library preparation	Yes, for most fragments (82%)	No	100 ng–2 μg	200 M	58%	Simple assay procedure, identifies 5hmC in regions with little 5hmCs, replicate measurements required for good coverage	Sun et al. (2013), Gross et al. (2015)
RRHP	Genomic DNA is digested with the Msp1, 5hmC is glycosylated, and DNA is digested again with Msp1, prior to size selection and sequencing	Yes, but only at restriction sites	No	500 ng	20–30 M	15% (mainly in CpG-rich regions)	Reproducible data, but limited to Msp1 sites, limited coverage in gene regulatory regions with low CpG density	Petterson et al. (2014)
5hmC-MeDIP	Immunoprecipitation with an antibody specific for 5-hydroxymethylcytosine	No	No	1–4 μg	20–40 M	100%	Preferential enrichment of 5hmC in regions with high density of 5hmCs and certain sequence contexts, low (inter-laboratory) reproducibility	Williams et al. (2011), Wu et al. (2011), Thomson et al. (2013)
CMS-MeDIP	Immunoprecipitation with an antibody specific for 5-cytosinemethylenesulfonate, an intermediate product of the bisulfite conversion of 5hmC	No	No	5 μg	40 M	100%	Increased reproducibility compared to 5hmC-MeDIP, but still biased for hydroxymethylation density	Pastor et al. (2011)
GLIB	Glycosylation of 5hmC followed by oxidation and attachment of a biotin group used for subsequent enrichment and sequencing	Yes/No	No	1 μg	40 M	100%	No dependency on hydroxymethylation density, complex multi-step protocol, only relative enrichment can be measured	Pastor et al. (2011)
hMeSeal	Glycosylation of 5hmC with an azide containing glucose moiety followed by attachment of a biotin group used for subsequent enrichment and sequencing	Yes/No	No	1 μg	40–80 M	100%	Streamlined protocol, chemical reagents available as a kit, only relative enrichment can be measured	Song et al. (2011)

(continued)

Table 16.3 (continued)

Technology	The technology behind	Single-nucleotide resolution (yes/no)	Absolute quantification (yes/no)	DNA input	Required number of reads per sample in M	Coverage of CpGs (theoretical)	Comments	References
TAB-seq	Glycosylation of 5hmC, oxidation of all methylated cytosines to 5caC with a recombinant TET enzyme (5hmC are protected), bisulfite treatment followed by NGS sequencing	Yes	Yes	3 μg	600 M	100%	Quantitative high-resolution 5hmC profiles requires large amount of sequencing to get accurate estimation of 5hmC levels at all cytosines, direct readout of 5hmC, more complex protocol compared to OXBS-seq	Yu et al. (2012)
OXBS-seq	Part of the samples is analyzed by standard (whole-genome) bisulfite sequencing, while in the other part 5hmC is oxidized to formylcytosine prior to bisulfite conversion. 5hmC levels are obtained by subtracting methylation levels from both reactions	Yes	Yes	400 ng–1 μg	1200 M	100%	Quantitative high-resolution 5hmC profiles requires large amount of sequencing to get accurate estimation of 5hmC levels at all cytosines as WGBS is required in addition (doubled compared to TAB-seq), combination with Infinium BeadChips or RRBS are alternative strategies, 5hmC levels are inferred, not measured, potential oxidative damage to DNA	Booth et al. (2013), Field et al. (2015)

Abbreviations: *5hmC* 5-hydoxymethylcytosine, *Aba-seq* AbaSI-coupled sequencing, *RRHP* reduced representation hydroxymethylcytosine profiling, *HELP-GT* HpaII tiny fragment enrichment by ligation-mediated PCR-glycosyl transferase assay, *MeDIP* methylated DNA immunoprecipitation, *CMS* cytosine methylene sulfonate, *GLIB* glucosylation, perIodate oxidation, biotinylation, *hMeSEAL* 5hmC selective chemical labeling, *TAB-seq* TET-assisted bisulfite sequencing, *OxBs-seq, OxBS* oxidative bisulfite sequencing

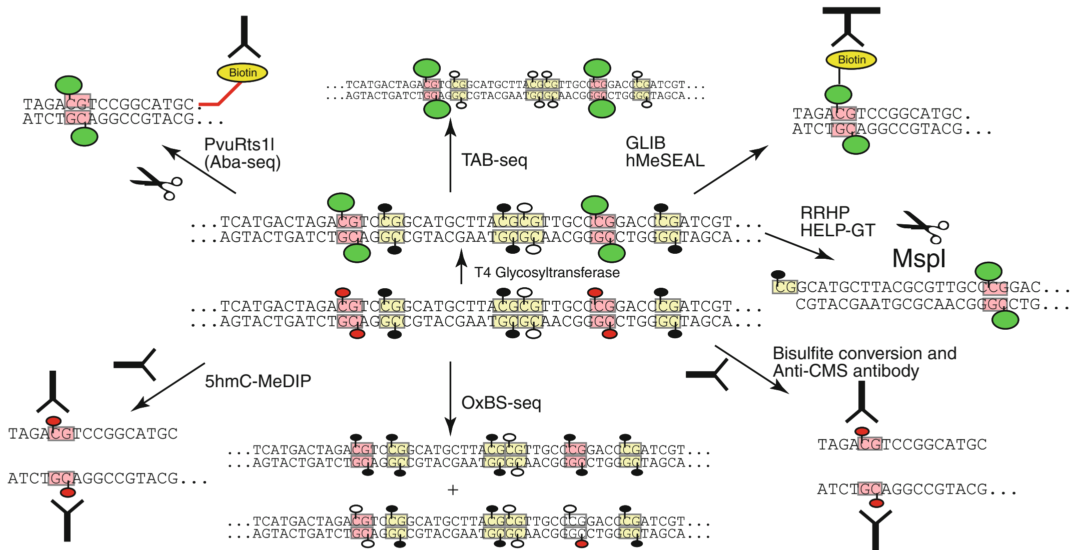


Fig. 16.3 Overview (simplified) of the different commonly used assays for the analysis of 5hmC. Details are given in the text and in Table 16.3. CpGs are depicted as lollipops, unmethylated CpGs are shown as empty lollipops while filled ones correspond to methylated CpGs, lollipops filled in red denote 5-hydroxymethylcytosines, while those in green depict hydroxymethylated and glycosylated cytosines

positive display of hydroxymethylation (Fig. 16.3). While this protocol is quite robust and yields high quality data with relatively few reads (20-30M), allowing a large number of samples to be simultaneously analyzed on a HiSeq/NovaSeq instrument, the dependence on MspI restriction sites allows to analyze only ~15% of CpG sites in the human genome. Using alternative restriction enzymes allows for the analysis of cytosines in other sequence contexts including non-CpG methylation (Sun et al. 2016).

In contrast to the bisulfite-based methods, affinity-based enrichments such as MeDIP, MIRA and several of the MBDs including MBD1, 2, and 4 and MECP2 do not recognize hydroxymethylated cytosines (Jin et al. 2010). A variety of new affinity-based methods have been devised to profile specifically 5-hydroxymethylation genome-wide using either antibodies against 5hmC or an intermediate product of its bisulfite conversion (5-methylenesulphonate) or glucosylation of 5hmC followed by biotinylation (Fig. 16.3).

Similar to the above-described MeDIP-seq protocols for the enrichment of methylated cytosines, antibodies raised against hydroxymethylated cytosine have been used to profile 5-hydroxymethylation genome-wide (Skvortsova et al. 2017; Williams et al. 2011; Wu et al. 2011; Ficz et al. 2011). However, despite the simplicity and cost-effectiveness of the antibody-based enrichment, this protocol was found to allow at best a semi-quantitative profiling of 5hmC and to enrich preferentially regions with a high density of hydroxymethylated cytosines and to display large inter-laboratory variations. In addition, there might be an enrichment of some sequence contexts such as tandem repeat sequences independent of the presence of hydroxymethylation (Pastor et al. 2011; Matarese et al. 2011; Thomson et al. 2013). The treatment of 5hmC with sodium bisulfite yields 5-cytosinemethylenesulfonate (CMS) as an intermediate product, which can be used to isolate sodium bisulfite-converted 5hmC using anti-5-methylenesulfonate antibodies (Pastor et al. 2011).

As an alternative strategy, the GLIB approach involves the glucosylation of 5-hydroxymethylcytosines, oxidation with

periodate, and biotinylation of 5mhC (Pastor et al. 2011). A glucose moiety is added to 5hmC by a glucosyltransferase, and the vicinal hydroxyl groups are subsequently oxidized to aldehydes by treatment with sodium periodate. Afterwards, biotin molecules are added to the newly formed aldehyde groups (Pastor et al. 2011). The GLIB approach allowed for the effective pull-down of >90% of fragments containing a single 5hmC, while the anti-CMS-based enrichment was more dependent on the density of 5hmC, but still achieved lower background levels compared to the conventional 5hmC-MeDIP (Pastor et al. 2011).

In a similar approach to GLIB, the glucose moiety containing an azide group can subsequently be used in a click-chemistry step to attach biotin molecules to 5hmC positions and enrich and sequence 5hmC-containing sequences (hMeSeal (Song et al. 2011)). This approach requires fewer steps and induces less DNA damage when compared to GLIB (Song et al. 2011). In direct comparisons, approaches based on chemical labeling proved to be more specific than antibody-based methods (Thomson et al. 2013) and have been successfully applied to minute amounts of biological material (1000 cells (Han et al. 2016)). For the quantification of the global level of 5hmC, a similar approach transferring a radioactively labeled glucose moiety can be used (Szwagierczak et al. 2010). The selective chemical labeling (SLC) exonuclease protocol combines the glycosylation and azide based biotinylation with an exonuclease digestion, where the exonuclease is blocked at the first modified 5hmC (Serandour et al. 2016). While this protocol allows the determination of hydroxymethylation in all sequence contexts at single CpG resolution, several replicates of this multi-step procedure are required to obtain a reliable and comprehensive coverage.

However, most of the above-described methods do not provide single-nucleotide resolution and due to the enrichment step do not allow for precise quantification of the hydroxymethylation level. The relative low abundance of 5hmC requires sensitive and preferably single-nucleotide resolution methods for its detection. Currently, two approaches are commonly used: TAB-seq and oxidative bisulfite sequencing. Both technologies can be performed at the genome-wide scale as well as for the targeted sequencing of regions of interest.

The TET-assisted bisulfite sequencing (TAB-seq) protocol uses also the differential potential of hydroxymethylated and methylated cytosines to glycosylation. Hydroxymethylated cytosines can be glycosylated and are thus protected from oxidation with recombinant TET-enzyme, which converts methylated cytosines to carboxymethyl-cytosines, which are subsequently deaminated to uracils using a conventional bisulfite treatment (Fig. 16.3) (Yu et al. 2012, 2018b). Of note, the glycosylation is not complete and its efficiency has been estimated at 75–92% (Yu et al. 2012), thus leading to an underestimation of the hydroxymethyl content of a sample, which can be mitigated through biological and technical replicates. TAB-seq allows a positive readout of hydroxymethylation at single-nucleotide resolution, as all remaining cytosines in the bisulfite-converted sequences should correspond to hydroxymethylated cytosines. Of note, TAB-seq has also been combined with Roche's SeqCap Epi CpGiant capture probes to sequence regions of interest to a much higher depth (Li et al. 2015a).

ACE-seq provides an alternative bisulfite-free method for the analysis of 5 hydroxymethylation based on the glycosylation of 5hmC followed by the enzymatic deamination of unmethylated and 5-methylated cytosines to uracil using the APOBEC3A cytosine deaminase (Schutsky et al. 2018). This method will only detect hydroxymethylated cytosine as methylated cytosines are deaminated with a similar efficiency as unmethylated ones and thus allows similarly to TAB-seq for a positive readout of DNA hydroxymethylation. Enzymatic Methyl (EM)-seq can be adapted to the analysis of hydroxymethylation using a similar principle by protecting hydroxymethylated cytosines through glycosylation followed by APOBEC mediated deamination (Vaisvila et al. 2021). Global

hydroxymethylation levels as assessed by EM-seq correlated with mass spectrometric measurements (Sun et al. 2021)

Oxidative bisulfite (OxBS-seq) sequencing is based on the selective and efficient oxidation of 5-hydroxymethylated cytosines to 5-formylcytosines using potassium perruthenate, while 5-methylcytosines are resistant to this oxidation reaction (Fig. 16.3) (Booth et al. 2012, 2013). In a subsequent bisulfite conversion, 5-formylcytosines are deaminated to uracils, like unmodified cytosines, and only methylated cytosines appear as cytosines in the sequence readout. Levels of hydroxymethylation can therefore be deduced by subtracting methylation levels of the OxBS-seq reaction from a standard bisulfite sequence which can be performed with the help of a number of specialized bioinformatic packages based on different statistical approaches (Kiihl 2021; Xu et al. 2016; Houseman et al. 2016). This approach has created considerable interest, as it allows adapting current workflows for whole-genome bisulfite sequencing for the detection of 5-hydroxymethylcytosine. However, it should be underlined that OxBS-Seq necessitates a standard bisulfite-treated reference sequence, which requires that two whole-genome bisulfite sequencing analyses have to be performed. Furthermore, due to the low abundance of 5-hydroxymethylation in most tissues an increased coverage might be required to reliably identify changes in hydroxymethylation for both TAB-seq and oxBS-seq, especially in regions with a low level of 5hmC, imposing significant cost challenges on the use of both protocols for the routine analysis of 5hmC for complete genome analyses. Approaches targeting the selected regions of the genome by either RRBS and/or amplicon sequencing have therefore been combined with the TAB and oxBS chemistry (Rydbirk et al. 2020; Skvortsova et al. 2017; Booth et al. 2012).

As described above in the paragraph on technologies for the genome-wide analysis of DNA methylation patterns, the Illumina Infinium BeadChips have become a very popular technology for the analysis of DNA methylation patterns in humans, especially for studies analyzing large cohorts. Two different approaches have been devised to use this platform also for the analysis of hydroxymethylation. The 450K/EPIC BeadChips also allow for a genome-wide analysis of hydroxymethylation at a fraction of the cost of the whole-genome sequencing and bioinformatic analysis is simplified. In the TAB-array approach, the TAB-seq principle (Yu et al. 2012) is transferred to the 450K BeadChip (Nazor et al. 2014; Chopra et al. 2014). Hydroxymethylation is analyzed on the methylation BeadChips using the standard experimental procedure following after glucosylation, oxidation, and bisulfite conversion. The TAB array yields a positive readout of hydroxymethylation, but with hydroxymethylation being a relatively rare DNA modification, the distribution of the observed beta-values is profoundly different from the one obtained by the conventional bisulfite-based analysis with a nearly complete unimodal distribution close to a beta value of zero. Therefore, separate normalizations for beta-values obtained by bisulfite and TAB array need to be performed. The 450K/EPIC BeadChips have also been combined with oxidative bisulfite sequencing using the bisulfite-converted and OxBS-converted DNA as input into the standard Illumina workflow (Skvortsova et al. 2017; Stewart et al. 2015; Field et al. 2015). Technical replicates show high reproducibility as for standard DNA methylation analysis (Spearman correlation >0.98, (Skvortsova et al. 2017)) About 40% of the probes on the 450K array and 48% on the EPIC array, with its increased content of intergenic and enhancer sequences, carry detectable 5hmC levels when brain samples are analyzed (Rydbirk et al. 2020; Skvortsova et al. 2017). The overall levels of hydroxymethylation correlated well with the hydroxymethylation levels measured by liquid chromatography coupled with mass spectrometry, and the results obtained by qPCR in conjunction with glycosylation and restriction enzyme digestion (Stewart et al. 2015; Field et al. 2015). Similarly, high correlation is obtained for array data when compared to sequencing technologies with single-nucleotide resolution such as oxBS-pyrosequencing, oxidative bisulfite or TAB amplicon sequencing

(Rydbirk et al. 2020; Skvortsova et al. 2017; Stewart et al. 2015). There are nonetheless some issues with the quantitative accuracy in regions which contain very high levels of methylated cytosines, which partly mask potential low hydroxymethylation levels. On the other hand, the sensitivity in regions with overall low DNA methylation is increased for the arrays compared to whole-genome sequencing approaches, as in most cases the required sequencing depth to detect the low levels of hydroxymethylation is not achieved (Skvortsova et al. 2017). Depending on the abundance of 5hmC in the analyzed tissue, the low level of hydroxymethylation and the inherent technical variability of the BeadChip measurement require the use of multiple biological and technical replicates to ensure the reliable detection of hydroxymethylation (Field et al. 2015). While too few samples have been analyzed on the 450K BeadChip in any of the so far published studies to draw definite conclusions, the degree of correlation with other methods seemed to be slightly higher for the studies using the oxBS approach compared to the TAB array. Due to the highly quantitative nature of its readout (Tost and Gut 2007), the pyrosequencing technology is also ideally suited for the analysis of gene-specific patterns of hydroxymethylation following oxidation and bisulfite treatment (Stewart et al. 2015; Qui et al. 2015).

Single-molecule real-time sequencing on the Pacific Biosciences sequencer shows slight variations in the kinetics of the incorporation of nucleotides depending on DNA base modifications and, consequently, the SMRT technology has been used to directly differentiate 5hmc from 5mC and unmodified cytosines (see also Sect. 16.10) (Flusberg et al. 2010). As genome-wide analyses of methylation and hydroxymethylation by SMRT sequencing are not yet feasible (the current output of a SMRT cell is ~500 MB), SMRT sequencing has been combined with the above-described glycosylation-mediated enrichment, to determine the localization of 5hmC in about 150 MB of sequence (Song et al. 2012) as well as with EM-seq(Sun et al. 2021).

Hydroxymethylation can be further oxidized to 5-formyl- and subsequently 5-carboxylcytosine by the TET enzymes. Reduced bisulfite sequencing (redBS-seq), enables the genome-wide identification of 5-formylcytosine at single base-resolution using the chemical reduction of 5-formylcytosine to 5-hydroxymethylcytosine (Booth et al. 2014). Similar to the above-described oxBS approach, the level of 5fC is obtained by subtraction of the methylation level at cytosines detected in standard BS-seq, which includes, of course, both 5mC and 5hmC marks from the methylation levels obtained by redBS-seq. Combination of OxBS-seq, standard WGBS, and redB-seqS allows thus the identification of 5mC, 5hmC, and 5fC in the same sample, but requires a significant amount of sequencing. The same chemistry has also been combined with an enzymatic approach using the above-described PvuRts1I restriction enzyme, thus avoiding bisulfite conversion (Sun et al. 2015), and selective chemical labeling whereby 5hmC is modified prior to reduction of 5fC. Then, the newly created 5hmC are glycosylated, biotinylated, and sequenced as described above (Song et al. 2013).

16.10 Direct Readout of DNA Methylation

The direct readout of CpG methylation has been demonstrated as a proof-of-principle for two amplification-free single-molecule sequencing technologies. These technologies bear the promise of sequencing longer DNA molecules at a single-molecule level, at lower cost and higher speed than existing methods. They can provide information on DNA methylation, hydroxymethylation, and other DNA modifications in the same experiment, at the same time abolishing some of the biases that are inherent to the second-generation sequencing approaches, such as the GC content bias.

The single-molecule real-time (SMRT) sequencer from Pacific Biosciences performs sequencing with an immobilized polymerase at

the bottom of zero-mode waveguide wells in zeptoliter volumes, monitoring the incorporation of phospholinked nucleotides through the detection of fluorescent pulses (Song et al. 2012; Flusberg et al. 2010; Ardui et al. 2018). SMRT sequencing on the Pacific Biosciences sequencer displays slight variations in the kinetics of the incorporation of nucleotides depending on the DNA base modifications and the SMRT technology has been used to directly differentiate 5hmC from 5mC and unmodified cytosines (Flusberg et al. 2010). As the kinetics, i.e. the interval between the end of a sequencing pulse and the beginning of the subsequent sequencing pulse, does also depend on the sequence context, an unmethylated template generated through whole-genome amplification is required to define a baseline. While this first report has raised a great interest, very few other reports of direct detection of methylation or hydroxymethylation in mammalian genomes have been published since. Due to the particularly low signal-to-noise ratio for 5-methylcytosine compared to other DNA modifications like 6-methyladenine, the deconvolution of the kinetic profile has proven to be much more challenging than initially anticipated, especially for regions with high CG content, which are of course of particular interest for DNA methylation analysis. As carboxymethyl cytosine shows an increased signal compared to 5mC, a protocol using enzymatic oxidation using TET1 has been devised to facilitate detection, but will rely strongly on the completeness of the conversion (Clark et al. 2013). Nonetheless, SMRT sequencing has been used for the detection of epigenetic modifications in bacteria and prokaryotes, where 5-methylcytosine occurs along with 6-methyladenosine and 4-methylcytosine (Blow et al. 2016; Beaulaurier et al. 2015). Due to the small genome size, exhaustive coverage of > 100 X increases the confidence in the kinetic data and allows accurate assessment of the methylation status. In a recent example, SMRT sequencing-based DNA methylation analysis was combined with restriction enzyme-based enrichment of a CGG-repeat region of the human *FMR1* gene causing fragile X syndrome (Pham et al. 2016). The accurate methylation analysis of these repeats, which are expanded in disease, was hitherto not possible due to lack of technologies providing sufficiently long reads. However, the protocol requires large amounts of starting material and employs a complicated multi-step enrichment procedure for the selection of locus of interest due to the necessity of avoiding amplification prior to sequencing. Furthermore, only relative and semi-quantitative DNA methylation levels could be obtained and only one of the sequenced strands gave interpretable results, demonstrating that the method has not yet reached maturity for the analysis of human and other mammalian genomes.

SMRT sequencing has been combined with amplicon bisulfite sequencing thus not making use of the potential capacities for direct readout of methylation marks, but allowing for sequencing of much larger amplicons (1.5 kb) instead of a maximum of 300–600 bp with standard Illumina sequencing (Yang et al. 2015b; Yang and Scott 2017). The analysis of CpGs overlapping between amplification products showed high reproducibility of the levels of DNA methylation and a high correlation to Sure Select Human Methyl-seq capture sequencing and Infinium arrays. However, the degradation induced by the bisulfite conversion substantially restricts the read length. As genome-wide analyses of methylation and hydroxymethylation by SMRT sequencing are not yet feasible (the current output of a SMRT cell is ~500 MB), SMRT sequencing has been combined with the above-described glycosylation-mediated enrichment, to determine the localization of 5hmC in about 150 MB of sequence (Song et al. 2012) as well as with EM-seq (Sun et al. 2021), allowing for the identification of 5mC and 5hmC in amplicons with a length up to 5 kb.

Nanopores are an alternative approach and make use of ionic current spectroscopy (Clarke et al. 2009; Jain et al. 2016). The current is very sensitive to subtle structural changes in the interrogated DNA—such as DNA methylation. Nanopore sequencing has the potential to change profoundly the way DNA methylation is analyzed, as besides the absence of potentially artifact prone steps, such as bisulfite conversion

and PCR amplification, it should allow for the direct detection of DNA methylation and its oxidative derivatives and provide longer sequencing reads at lower cost and higher speed. Both solid-state and mutated MspA nanopores have been used for the detection of methylcytosine and its differentiation from hydroxymethylcytosine (Simpson et al. 2017; Wanunu et al. 2011; Shim et al. 2013; Laszlo et al. 2013; Manrao et al. 2011; Wescoe et al. 2014). Due to their different polarity, DNA flexibility, and duplex stability, solid-state nanopores are capable of differentiating between cytosine, methylated cytosines as well as its oxidative derivatives in synthetic templates (Wanunu et al. 2011). However, depending on the nanopore employed, repeated reads might be required to accurately differentiate between 5mC and 5hmC (Laszlo et al. 2013). Depending on the sequence context surrounding the CpG dinucleotides of interest, error rates between 2 and 12% were observed, suggesting that accurate methylation estimates can be achieved with read numbers of less than 20 molecules of the same locus (Schreiber et al. 2013). In addition, exonuclease-assisted nanopore sequencing of single molecules, for which a processive exonuclease produces nucleoside monophosphates subsequently presented to an alpha-hemolysin protein nanopore with a cyclodextrin adapter, has been shown to detect cytosine modifications in synthetic templates (Clarke et al. 2009; Wallace et al. 2010).

An alternative strategy to increase the ionic blockage induced by the methylation group is the selective labeling of methylcytosine with MBD1 or Kaiso zinc finger proteins, which form a small non-covalent complex with methylated cytosines (Shim et al. 2013; Shim et al. 2015). The binding of these proteins leads to a threefold increase in the blockage current compared to unmethylated DNA, allowing thereby the simultaneous detection of methylated or unmethylated DNA as well as a rough quantification of the methylation degree of the analyzed DNA fragment by counting the molecules with bound proteins. Similarly, mercury ions can form a reversible bridge between two mismatched DNA bases containing a thymine-thymine or thymine-uracil mismatch between the analytical target and a synthetic probe (Kang et al. 2013). When passing through a solid-state (graphene) nanopore, changes in the ion current due to the passage of the mercury ion can be recorded. (Methyl)cytosines are however not able to form this reversible interstrand MercuLock, thereby allowing to distinguish methylated cytosines from cytosines after bisulfite conversion (Kang et al. 2013). As the method requires specific probes to be designed for each target, it is more suited for locus-specific analyses rather than potentially genome-wide analysis and the use of bisulfite conversion does prohibit the differentiation of methylcytosine and hydroxymethylcytosine.

There has been tremendous progress in recent years for Nanopore sequencing, especially in terms of throughput and associated costs which were prohibiting its use for DNA methylation analysis in complex genome, nonetheless accuracy needs still to be improved. In addition, significant progress has been made for the reliable detection of epigenetic modifications from Nanopore sequencing data using improved methods for signal detection and deconvolution (Simpson et al. 2017; Rand et al. 2017). In a recent benchmark study, DNA methylation calls correlated well with bisulfite and enzymatic conversion methylation data and a high precision for detecting differential DNA methylation differences (>20%) was confirmed, albeit at much lower sensitivity compared to the second-generation sequencing approaches (Foox et al. 2021). As the complexity of the human genome still remains a challenging task for genome-wide DNA methylation analysis, Nanopore sequencing for DNA methylation analysis is commonly focused on specific regions of interest, especially transposable elements and those associated with repeat expansion, or the mitochondria (Ewing et al. 2020; Luth et al. 2022; Giesselmann et al. 2019; Bicci et al. 2021). Enrichment strategies comprise Cas9 or Cas12a targeted sequencing, by which specific genomic regions are enriched by two orders of magnitude through targeting with guide RNAs and subsequent adaptor ligation to achieve an amplification-free workflow

preserving DNA methylation patterns (Giesselmann et al. 2019; Gilpatrick et al. 2020). An alternative strategy relies on amplification of large genomic regions (5kb) following, for example, EM-seq (Sun et al. 2021) for the analysis of 5mC and following glycosylation 5hmC. In direct comparison between SMRT sequencing and Nanopore sequencing, SMRT sequencing gave 5mC estimates close to those obtained by Illumina sequencing, while Nanopore more frequently called inaccurate DNA methylation in line with the known shortcoming of Nanopore sequencing (Sun et al. 2021).

16.11 Combined Analysis of DNA Methylation and Other Epigenetic Modifications

16.11.1 Histone Modifications

The different layers of epigenetic modifications, post-transcriptional histone modifications, histone variants, and DNA methylation, are closely intertwined and stabilize each other to ensure the faithful propagation of epigenetic states over time and especially through cell division. Studies analyzing several layers of epigenetic modifications are still scarce and due to the requirement of a large amount of biological material often restricted to cell lines and cellular models. Furthermore, the combination of the data is performed rather at the analysis level by overlaying the profiles obtained in distinct experiments and developing probabilistic models on the occurrence of the epigenetic marks, rather than by a direct molecular readout. While many technologies have been developed for the comprehensive analysis of a single type of epigenetic modification, few can address the co-occurrence and interaction of different modifications.

Nucleosomes consist of ~147-bp-long DNA stretches wrapped around an octamer of histone proteins, and are connected through "linker DNA." The N-terminal tails of the histone proteins are the targets of many posttranslational modifications. The combinations of the different modifications and their multivalency determine the regulatory landscape of a genomic region and its effect on gene expression, giving rise to the concept of the "histone code" (Ernst and Kellis 2012; Hoffman et al. 2013; Rothbart and Strahl 2014). Chromatin immunoprecipitation (ChIP (Gilmour and Lis 1984)) is a well-established method in cellular biology to study the specific interaction between a protein of interest and genomic DNA and has been extensively used to identify transcription factor and regulatory protein binding sites in the genome (Gerstein et al. 2012).

Chromatin, which might be crosslinked to DNA, is extracted and randomly fragmented by sonication into 200–600 base pair fragments. Then, DNA–protein complexes are immunoprecipitated using an antibody against a specific protein or histone modification and protein A/G agarose resin. Finally, covalent cross-links are reversed by heating and DNA is purified after RNase A and Proteinase K treatment. At this point, a small amount of purified DNA is available that can be subsequently analyzed by qPCR for the detection and quantification of the analyzed modification at a locus of interest using specific primers. Alternatively, DNA can be analyzed by microarrays (ChIP-on-chip) or NGS (ChIP-seq,) for a genome-wide picture of the DNA-protein binding events (i.e., identification of all binding sites of a transcription factor; mapping of an histone modification on the entire genome at very high resolution) (Gerstein et al. 2012; Barski et al. 2007; Mikkelsen et al. 2007). ChIP-seq has since been performed at production scale in the ENCODE, modENCODE, and Roadmap epigenomics mapping projects (Roadmap Epigenomics et al. 2015; Landt et al. 2012; Partridge et al. 2020) and has become the workhorse for the genome-wide mapping of the occupancy of DNA by transcription factors and posttranslationally modified histones.

The combination of ChIP with DNA methylation analysis enables determination if a protein is (on average) bound to methylated DNA or unmethylated DNA at a given genomic locus. This approach permits studying a putative methyl-binding protein or the association of a specific histone modification with methylated or

non-methylated DNA by analyzing the methylation levels of ChIPed DNA at single-nucleotide resolution using bisulfite conversion followed by NGS or pyrosequencing (Moison et al. 2013, 2014; Kagey et al. 2010; Thomson et al. 2010). For a more qualitative analysis, methylated DNA can also be specifically detected combining ChIP with methylation-specific PCR (Zinn et al. 2007). Other studies have combined ChIP with genome-wide sequencing methods to address the genome-wide relationship between DNA methylation and specific histone modifications (ChIP-BS-seq or BisChIP-seq; (Brinkman et al. 2012; Statham et al. 2012)). For this approach, a relatively large number of cells are used for immunoprecipitation to obtain sufficient material for the bisulfite conversion reaction, as well as the ChIP-seq reaction, or several ChIP reactions are pooled prior to library generation. After library preparation with methylated adaptors, part of the library is bisulfite converted, PCR amplified, and the DNA methylation patterns of the immunoprecipitated fragments are analyzed using second-generation sequencing. This analysis permits the simultaneous analysis of DNA methylation associated with histones marked with a set of specific posttranslational modifications on the same pools of cells, but not on the same nucleosome. One of the major drawbacks of ChIP is the requirement for relatively large amounts of input material as well as a relatively low signal-to-noise-ratio. Therefore, alternative approaches relying mainly on enzymatic methods have been developed, including DamID (DNA adenine methyltransferase identification), ChEC (chromatin endogenous cleavage), ChIC (chromatin immune cleavage), or CUT&RUN, all of which use various enzymes to specifically cleave the genome next to a binding site of a protein of interest (van Steensel and Henikoff 2000; Schmid et al. 2004; Skene et al. 2018; Meers et al. 2019). CUT&Tag is based on an antibody-directed tagmentation by a protein A-Tn5 transposase fusion (Kaya-Okur et al. 2019, 2020). Following the tagmentation, DNA fragments are released and purified, followed by PCR amplification to obtain libraries ready for sequencing, simplifying largely the protocol of ChIP-seq and making it to amenable to the analysis of very few cells and even single cells. In contrast to ChIP-seq, which is restricted to the use of a single antibody, this approach can also be used to profile several targets in parallel using antibody specific barcoded sequencing adaptors (Gopalan et al. 2021). In analogy to the ChIP-BS-seq approaches, CUT&Tag has been combined with bisulfite conversion of the transposed fragments, allowing the correlation of DNA methylation with ChIP modification profiles (Li et al. 2021). Similar protocols making use of the Tn5 transposase include ChIL-seq, ACT-seq, or CoBATCH (Harada et al. 2019; Carter et al. 2019; Wang et al. 2019), but have not yet been used for simultaneous DNA methylation analysis.

The above-described SCAN approach (Cipriany et al. 2010, 2012) allows isolating single methylated or unmethylated DNA molecules and enables characterization of epigenetic states analyzing the co-occurrence or absence of DNA methylation and histone modification using fluorescently labeled antibodies against the targeted histone modifications in combination with MBD1 for the detection of DNA methylation (Murphy et al. 2013). H3K9me3 was detected together with DNA methylation on the very same nucleosome, while DNA methylation and H3K27me3 were mutually exclusive under normal physiological conditions, but this epigenetic regulation became deregulated in cancer or upon cell transformation (Murphy et al. 2013). The use of quantum dots instead of fluorescent dye and a potential parallelization of the nanofluidic devices will allow a higher throughput of single-molecule analysis with increased multiplexing capabilities. Furthermore, due to the advances in low-input/single-cell sequencing described in Sect. 16.8, sequencing of the molecules with a given combinatorial pattern of epigenetic modifications for their identification has already become feasible as shown for the combinatorial decoding of histone modifications on single molecules using fluorescently labeled antibodies and total internal reflection microscopy (Shema et al. 2016).

16.11.2 Nucleosome Positioning

Positioning of nucleosomes and remodeling of chromatin play key roles in the coordination of the correct gene expression program. Positioning of nucleosomes depends on (among others) the underlying DNA sequence, ATP-dependent nucleosome remodelers, DNA-binding proteins, the RNA polymerase II transcription machinery, and their interactions. As a result, the core enhancer, promoter, and terminator regions of genes are typically depleted of nucleosomes, whereas most of the genomic DNA is occupied (Struhl and Segal 2013). Consequently, the analysis of chromatin accessibility and nucleosome positioning is essential for the understanding of transcriptional regulation and can be used for the analysis and identification of gene regulatory elements and their changes in disease. In most cases, information on nucleosome positioning is obtained by enzymatic digestion (MNAse-seq), chemical cleavage (CC-seq), or immunoprecipitation of chromatin followed by next-generation sequencing of the resulting DNA fragments or derived from chromatin accessibility profiles obtained with DNAseI-seq or ATAC-seq. While well suited for their purpose, these methods do not yield any information on DNA methylation patterns. The DNA *M*ethyltransferase *A*ccessibility *P*rotocol for *i*ndividual *t*emplates (MAPit) or *N*ucleosome *o*ccupancy and *ME*thylome sequencing (NoME-seq) makes use of DNA methyltransferase footprinting to determine the nucleosome positioning, while at the same time retaining the original DNA methylation patterns, thereby enabling the correlated analysis of these two epigenetic hallmarks and the corresponding chromatin configurations (Kelly et al. 2012; Pardo et al. 2011). These approaches use the GpC methyltransferase (M.CviPI (Xu et al. 1998)) in the presence of the universal methylation donor *S*-adenosyl-L-methionine that methylates accessible GC dinucleotides, i.e. sequences that are not protected by nucleosomes or other tight binding proteins. Methylation at CpG dinucleotides yields information about the DNA methylation patterns, while methylation at GpC dinucleotides informs on the chromatin accessibility at the same locus (Jessen et al. 2004). Locus-specific analyses can subsequently be performed by cloning and sequencing after GC methyltransferase treatment (Pardo et al. 2011; You et al. 2011; Taberlay et al. 2011). Genome-wide analyses can also be performed by ligating adaptors to the fragmented and methylated DNA followed by bisulfite conversion and sequencing (Kelly et al. 2012).

16.12 Conclusions

Sequencing-based approaches have revolutionized the analysis of the epigenome allowing the investigation of multiple gene regulatory levels including DNA methylation, coding and non-coding RNA expression and its nascent production, location of RNA polymerases, transcription factors and other DNA-binding proteins, histone modifications, chromatin accessibility, as well as the spatial organization of the genome. Many kits for library preparation for DNA methylation analysis have been commercialized, but benchmark studies have revealed inherent biases of many of the protocols. While still expensive, the deployment of new sequencing platforms with the possibility to perform whole-genome bisulfite sequencing will further decrease the cost and turn-around time. New protocols replacing bisulfite conversion with enzymatic methods might overcome some of the challenges associated with the strong DNA degradation and allow the analysis of DNA methylation from decreased amounts of input. In fact, the challenge of DNA methylation analysis has shifted from the data generation to the data analysis. Integration of molecular data with other publicly available genomic and epigenomic large-scale data sets to analyze the functional consequences of alterations at one molecular level, as well as appropriate statistical analysis to decipher higher-order regulatory circuits and the spatial

and temporal organization of the genome is currently the major challenge. In addition, expected continued improvement in sequencing technology and reduction in sequencing costs will enable the routine analysis of methylomes in combination with genomic data. Furthermore, the possibility to determine the methylome of cell-free circulating DNA might be a powerful tool for the early detection of cancers and determination of their origin, but might also be useful for the prediction of treatment response in many complex diseases of several organs without disease causing mutations, such as autoimmune and inflammatory diseases.

Direct readout technologies of epigenetic modifications have been devised and are actively investigated in many laboratories. Although the data analysis has been largely improved and solutions for targeted sequencing using third-generation sequencers have been developed, their time to full technical maturity is difficult to estimate. They promise further decrease in cost and differentiation of methylcytosine from its oxidative derivatives.

The implementation of the first DNA methylation-based biomarkers has been significantly slowed down due to the use of methods with varying sensitivity and readouts with arbitrary cut-offs leading to results that were difficult to integrate and combine between studies. However, the recent concentration on a few powerful methods with quantitative and often single-nucleotide resolution, will enable a much fast progress for the use of locus-specific DNA methylation technologies for clinical diagnosis and prognosis.

Acknowledgements The work in the laboratory of Jörg Tost is supported by the following grants unrelated to the presented work: The French National Research Agency ANR (ANR-13-EPIG-0003-05, ANR-13-CESA-0011-05, ANR-18-RAR3-0001, and RESCUER-19-SYME-0001-05), the European Union's Horizon 2020 Research and Innovation Programme under Grant agreement No. 847912, the French Medical Research Foundation FRM (ENV2020-04011870), a Sirius research award (UCB Pharma S.A.), a Passerelle research award (Pfizer), iCARE (MSD Avenir), funding from DBV Technologies and the institutional budget of the CNRGH.

References

Aberg KA, McClay JL, Nerella S, Xie LY, Clark SL, Hudson AD, Bukszar J, Adkins D, Consortium SS, Hultman CM, Sullivan PF, Magnusson PK, van den Oord EJ (2012) MBD-seq as a cost-effective approach for methylome-wide association studies: demonstration in 1500 case–control samples. Epigenomics 4: 605–621

Aberg KA, Xie L, Chan RF, Zhao M, Pandey AK, Kumar G, Clark SL, van den Oord EJ (2015) Evaluation of methyl-binding domain based enrichment approaches revisited. PLoS One 10:e0132205

Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, Bock C, Boehm B, Campo E, Caricasole A, Dahl F, Dermitzakis ET, Enver T, Esteller M, Estivill X, Ferguson-Smith A, Fitzgibbon J, Flicek P, Giehl C, Graf T, Grosveld F, Guigo R, Gut I, Helin K, Jarvius J, Kuppers R, Lehrach H, Lengauer T, Lernmark A, Leslie D, Loeffler M, Macintyre E, Mai A, Martens JH, Minucci S, Ouwehand WH, Pelicci PG, Pendeville H, Porse B, Rakyan V, Reik W, Schrappe M, Schubeler D, Seifert M, Siebert R, Simmons D, Soranzo N, Spicuglia S, Stratton M, Stunnenberg HG, Tanay A, Torrents D, Valencia A, Vellenga E, Vingron M, Walter J, Willcocks S (2012) BLUEPRINT to decode the epigenetic signature written in blood. Nat Biotechnol 30:224–226

Adey A, Shendure J (2012) Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. Genome Res 22:1139–1143

Adey A, Morrison HG, Asan XX, Kitzman JO, Turner EH, Stackhouse B, MacKenzie AP, Caruccio NC, Zhang X, Shendure J (2010) Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol 11:R119

Adusumalli S, Mohd Omar MF, Soong R, Benoukraf T (2015) Methodological aspects of whole-genome bisulfite sequencing analysis. Brief Bioinform 16: 369–379

Ahmadian A, Ehn M, Hober S (2006) Pyrosequencing: history, biochemistry and future. Clin Chim Acta 363: 83–94

Akalin A, Garrett-Bakelman FE, Kormaksson M, Busuttil J, Zhang L, Khrebtukova I, Milne TA, Huang Y, Biswas D, Hess JL, Allis CD, Roeder RG, Valk PJ, Lowenberg B, Delwel R, Fernandez HF, Paietta E, Tallman MS, Schroth GP, Mason CE, Melnick A, Figueroa ME (2012) Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. PLoS Genet 8:e1002781

Alders M, Bliek J, vd Lip K, vd Bogaard R, Mannens M (2009) Determination of KCNQ1OT1 and H19 methylation levels in BWS and SRS patients using methylation-sensitive high-resolution melting analysis. Eur J Hum Genet 17:467–473

Allum F, Shao X, Guenard F, Simon MM, Busche S, Caron M, Lambourne J, Lessard J, Tandre K, Hedman AK, Kwan T, Ge B, Ronnblom L, McCarthy MI, Deloukas P, Richmond T, Burgess D, Spector TD, Tchernof A, Marceau S, Lathrop M, Vohl MC, Pastinen T, Grundberg E, Multiple Tissue Human Expression Resource Consortium (2015) Characterization of functional methylomes by next-generation capture sequencing identifies novel disease-associated variants. Nat Commun 6:7211

Alnaes GI, Ronneberg JA, Kristensen VN, Tost J (2015) Heterogeneous DNA Methylation Patterns in the GSTP1 Promoter Lead to Discordant Results between Assay Technologies and Impede Its Implementation as Epigenetic Biomarkers in Breast Cancer. Genes (Basel) 6:878–900

Andersen GB, Hager H, Hansen LL, Tost J (2014) Improved reproducibility in genome-wide DNA methylation analysis for PAXgene-fixed samples compared with restored formalin fixation and paraffin-embedding DNA. Anal Biochem 468:50–58

Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, Hu TX, Krueger F, Smallwood S, Ponting CP, Voet T, Kelsey G, Stegle O, Reik W (2016) Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat Methods 13:229–232

Ardui S, Ameur A, Vermeesch JR, Hestand MS (2018) Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic Acids Res 46:2159–2168

Armstrong MJ, Jin Y, Allen EG, Jin P (2019) Diverse and dynamic DNA modifications in brain and diseases. Hum Mol Genet 28:R241–R253

Arneson A, Haghani A, Thompson MJ, Pellegrini M, Kwon SB, Vu H, Maciejewski E, Yao M, Li CZ, Lu AT, Morselli M, Rubbi L, Barnes B, Hansen KD, Zhou W, Breeze CE, Ernst J, Horvath S (2022) A mammalian methylation array for profiling methylation levels at conserved sequences. Nat Commun 13: 783

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30:1363–1369

Balch C, Yan P, Craft T, Young S, Skalnik DG, Huang TH, Nephew KP (2005) Antimitogenic and chemosensitizing effects of the methylation inhibitor zebularine in ovarian cancer. Mol Cancer Ther 4:1505–1514

Balic M, Pichler M, Strutz J, Heitzer E, Ausch C, Samonigg H, Cote RJ, Dandachi N (2009) High quality assessment of DNA methylation in archival tissues from colorectal cancer patients using quantitative high-resolution melting analysis. J Mol Diagn 11:102–108

Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH, Xie B, Daley GQ, Church GM (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol 27:361–368

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. Cell 129:823–837

Bastian PJ, Palapattu GS, Lin X, Yegnasubramanian S, Mangold LA, Trock B, Eisenberger MA, Partin AW, Nelson WG (2005) Preoperative serum DNA GSTP1 CpG island hypermethylation and the risk of early prostate-specific antigen recurrence following radical prostatectomy. Clin Cancer Res 11:4037–4043

Battram T, Yousefi P, Crawford G, Prince C, Sheikhali Babaei M, Sharp G, Hatcher C, Vega-Salas M, Khodabakhsh S, Whitehurst O, Langdon R, Mahoney L, Elliott H, Mancano G, Lee M, Watkins S, Lay A, Hemani G, Gaunt T, Relton C, Staley J, Suderman M (2022) The EWAS Catalog: a database of epigenome-wide association studies. Wellcome Open Res 7:41

Baylin SB, Jones PA (2011) A decade of exploring the cancer epigenome—biological and translational implications. Nat Rev Cancer 11:726–734

Beaulaurier J, Zhang XS, Zhu S, Sebra R, Rosenbluh C, Deikus G, Shen N, Munera D, Waldor MK, Chess A, Blaser MJ, Schadt EE, Fang G (2015) Single molecule-level detection and long read-based phasing of epigenetic variations in bacterial methylomes. Nat Commun 6:7438

Begum S, Brait M, Dasgupta S, Ostrow KL, Zahurak M, Carvalho AL, Califano JA, Goodman SN, Westra WH, Hoque MO, Sidransky D (2011) An epigenetic marker panel for detection of lung cancer using cell-free serum DNA. Clin Cancer Res 17:4494–4503

Beinse G, Borghese B, Metairie M, Just PA, Poulet G, Garinet S, Parfait B, Didelot A, Bourreau C, Agueeff N, Lavolle A, Terris B, Chapron C, Goldwasser F, Leroy K, Blons H, Laurent-Puig P, Taly V, Alexandre J (2022) Highly specific droplet-digital PCR detection of universally methylated circulating tumor DNA in endometrial carcinoma. Clin Chem 68(6):782–793

Beltran-Garcia J, Osca-Verdegal R, Mena-Molla S, Garcia-Gimenez JL (2019) Epigenetic IVD tests for personalized precision medicine in cancer. Front Genet 10:621

Berdasco M, Fraga MF, Esteller M (2009) Quantification of global DNA methylation by capillary electrophoresis and mass spectrometry. Methods Mol Biol 507:23–34

Bestor TH, Ingram VM (1983) Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proc Natl Acad Sci USA 80: 5559–5563

Bian S, Hou Y, Zhou X, Li X, Yong J, Wang Y, Wang W, Yan J, Hu B, Guo H, Wang J, Gao S, Mao Y, Dong J, Zhu P, Xiu D, Yan L, Wen L, Qiao J, Tang F, Fu W (2018) Single-cell multiomics sequencing and analyses of human colorectal cancer. Science 362:1060–1063

Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL (2009) Genome-wide DNA

methylation profiling using Infinium((R)) assay. Epigenomics 1:177–200
Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan JB, Shen R (2011) High density DNA methylation array with single CpG site resolution. Genomics 98: 288–295
Bicci I, Calabrese C, Golder ZJ, Gomez-Duran A, Chinnery PF (2021) Single-molecule mitochondrial DNA sequencing shows no evidence of CpG methylation in human cells and tissues. Nucleic Acids Res 49: 12757–12768
Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16:6–21
Bird AP, Southern EM (1978) Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis. J Mol Biol 118:27–47
Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, Froula J, Kang DD, Malmstrom RR, Morgan RD, Posfai J, Singh K, Visel A, Wetmore K, Zhao Z, Rubin EM, Korlach J, Pennacchio LA, Roberts RJ (2016) The epigenomic landscape of prokaryotes. PLoS Genet 12:e1005854
BLUEPRINT Consortium (2016) Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. Nat Biotechnol 34: 726–737
Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, Jager N, Gnirke A, Stunnenberg HG, Meissner A (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol 28:1106–1114
Bollati V, Baccarelli A, Hou L, Bonzini M, Fustinoni S, Cavallo D, Byun HM, Jiang J, Marinelli B, Pesatori AC, Bertazzi PA, Yang AS (2007) Changes in DNA methylation patterns in subjects exposed to low-dose benzene. Cancer Res 67:876–880
Bönsch D, Lenz B, Reulbach U, Kornhuber J, Bleich S (2004) Homocysteine associated genomic DNA hypermethylation in patients with chronic alcoholism. J Neural Transm 111:1611–1616
Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, Balasubramanian S (2012) Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science 336:934–937
Booth MJ, Ost TW, Beraldi D, Bell NM, Branco MR, Reik W, Balasubramanian S (2013) Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. Nat Protoc 8:1841–1851
Booth MJ, Marsico G, Bachman M, Beraldi D, Balasubramanian S (2014) Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. Nat Chem 6:435–440
Borgaro JG, Zhu Z (2013) Characterization of the 5-hydroxymethylcytosine-specific DNA restriction endonucleases. Nucleic Acids Res 41:4198–4206
Borno ST, Fischer A, Kerick M, Falth M, Laible M, Brase JC, Kuner R, Dahl A, Grimm C, Sayanjali B, Isau M, Rohr C, Wunderlich A, Timmermann B, Claus R, Plass C, Graefen M, Simon R, Demichelis F, Rubin MA, Sauter G, Schlomm T, Sultmann H, Lehrach H, Schweiger MR (2012) Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation. Cancer Discov 2: 1024–1035
Boyle P, Clement K, Gu H, Smith ZD, Ziller M, Fostel JL, Holmes L, Meldrim J, Kelley F, Gnirke A, Meissner A (2012) Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. Genome Biol 13:R92
Brandes JC, Carraway H, Herman JG (2007) Optimal primer design using the novel primer design program: MSPprimer provides accurate methylation analysis of the ATM promoter. Oncogene 26:6229–6237
Brena RM, Auer H, Kornacker K, Hackanson B, Raval A, Byrd JC, Plass C (2006) Accurate quantification of DNA methylation using combined bisulfite restriction analysis coupled with the Agilent 2100 Bioanalyzer platform. Nucleic Acids Res 34:e17
Brinkman AB, Simmer F, Ma K, Kaan A, Zhu J, Stunnenberg HG (2010) Whole-genome DNA methylation profiling using MethylCap-seq. Methods 52:232–236
Brinkman AB, Gu H, Bartels SJ, Zhang Y, Matarese F, Simmer F, Marks H, Bock C, Gnirke A, Meissner A, Stunnenberg HG (2012) Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res 22:1128–1138
Brouilette S, Kuersten S, Mein C, Bozek M, Terry A, Dias KR, Bhaw-Rosun L, Shintani Y, Coppen S, Ikebe C, Sawhney V, Campbell N, Kaneko M, Tano N, Ishida H, Suzuki K, Yashiro K (2012) A simple and novel method for RNA-seq library preparation of single cell cDNA analysis by hyperactive Tn5 transposase. Dev Dyn 241:1584–1590
Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10:1213–1218
Bujold D, Morais DAL, Gauthier C, Cote C, Caron M, Kwan T, Chen KC, Laperle J, Markovits AN, Pastinen T, Caron B, Veilleux A, Jacques PE, Bourque G (2016) The International Human Epigenome Consortium Data Portal. Cell Syst 3:496–499 e492
Bundo M, Sunaga F, Ueda J, Kasai K, Kato T, Iwamoto K (2012) A systematic evaluation of whole genome amplification of bisulfite-modified DNA. Clin Epigenetics 4:22
Campan M, Weisenberger DJ, Trinh B, Laird PW (2009) MethyLight. Methods Mol Biol 507:325–337
Campan M, Moffitt M, Houshdaran S, Shen H, Widschwendter M, Daxenbichler G, Long T,

Marth C, Laird-Offringa IA, Press MF, Dubeau L, Siegmund KD, Wu AH, Groshen S, Chandavarkar U, Roman LD, Berchuck A, Pearce CL, Laird PW (2011) Genome-scale screen for DNA methylation-based detection markers for ovarian cancer. PLoS ONE 6: e28141

Candiloro IL, Mikeska T, Dobrovic A (2011) Assessing combined methylation-sensitive high resolution melting and pyrosequencing for the analysis of heterogeneous DNA methylation. Epigenetics 6:500–507

Carter B, Ku WL, Kang JY, Hu G, Perrie J, Tang Q, Zhao K (2019) Mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACT-seq). Nat Commun 10:3747

Castellanos-Rizaldos E, Milbury CA, Karatza E, Chen CC, Makrigiorgos GM, Merewood A (2014) COLD-PCR amplification of bisulfite-converted DNA allows the enrichment and sequencing of rare un-methylated genomic regions. PLoS One 9:e94103

Chakravarty D, Solit DB (2021) Clinical cancer genomic profiling. Nat Rev Genet 22:483–501

Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Correa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier MC, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee TH, Thi VH, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CH, Wang X, Canaguier A, Chauveau A, Berard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury JM, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C, Wincker P (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science 345:950–953

Chan KC, Jiang P, Chan CW, Sun K, Wong J, Hui EP, Chan SL, Chan WC, Hui DS, Ng SS, Chan HL, Wong CS, Ma BB, Chan AT, Lai PB, Sun H, Chiu RW, Lo YM (2013) Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. Proc Natl Acad Sci USA 110:18761–18768

Chan RW, Jiang P, Peng X, Tam LS, Liao GJ, Li EK, Wong PC, Sun H, Chan KC, Chiu RW, Lo YM (2014) Plasma DNA aberrations in systemic lupus erythematosus revealed by genomic and methylomic sequencing. Proc Natl Acad Sci USA 111:E5302–E5311

Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8: 203–209

Chen Y, Song J, Ruan Q, Zeng X, Wu L, Cai L, Wang X, Yang C (2021) Single-cell sequencing methodologies: from transcriptome to multi-dimensional measurement. Small Methods 5:e2100111

Cheow LF, Quake SR, Burkholder WF, Messerschmidt DM (2015) Multiplexed locus-specific analysis of DNA methylation in single cells. Nat Protoc 10:619–631

Chhibber A, Schroeder BG (2008) Single-molecule polymerase chain reaction reduces bias: application to DNA methylation analysis by bisulfite sequencing. Anal Biochem 377:46–54

Chim SS, Tong YK, Chiu RW, Lau TK, Leung TN, Chan LY, Oudejans CB, Ding C, Lo YM (2005) Detection of the placental epigenetic signature of the maspin gene in maternal plasma. Proc Natl Acad Sci USA 102:14753–14758

Chim SS, Jin S, Lee TY, Lun FM, Lee WS, Chan LY, Jin Y, Yang N, Tong YK, Leung TY, Lau TK, Ding C, Chiu RW, Lo YM (2008) Systematic search for placental DNA-methylation markers on chromosome 21: toward a maternal plasma-based epigenetic test for fetal trisomy 21. Clin Chem 54:500–511

Choi SH, Worswick S, Byun HM, Shear T, Soussa JC, Wolff EM, Douer D, Garcia-Manero G, Liang G, Yang AS (2009) Changes in DNA methylation of tandem DNA repeats are different from interspersed repeats in cancer. Int J Cancer 125:723–729

Choi JH, Li Y, Guo J, Pei L, Rauch TA, Kramer RS, Macmil SL, Wiley GB, Bennett LB, Schnabel JL, Taylor KH, Kim S, Xu D, Sreekumar A, Pfeifer GP, Roe BA, Caldwell CW, Bhalla KN, Shi H (2010) Genome-wide DNA methylation maps in follicular lymphoma cells determined by methylation-enriched bisulfite sequencing. PLoS One 5:e13020

Chopra P, Papale LA, White AT, Hatch A, Brown RM, Garthwaite MA, Roseboom PH, Golos TG, Warren ST, Alisch RS (2014) Array-based assay detects genome-wide 5-mC and 5-hmC in the brains of humans, non-human primates, and mice. BMC Genomics 15:131

Chowdhury B, Cho IH, Hahn N, Irudayaraj J (2014) Quantification of 5-methylcytosine, 5-hydroxymethylcytosine and 5-carboxylcytosine from the blood of cancer patients by an enzyme-based immunoassay. Anal Chim Acta 852:212–217

Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR, Castanos-Velez E, Blumenstein BA, Rosch T, Osborn N, Snover D, Day RW, Ransohoff DF, PRESEPT Clinical Study Steering Committee, Investigators and Study Team (2014) Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut 63: 317–325

Cipriany BR, Zhao R, Murphy PJ, Levy SL, Tan CP, Craighead HG, Soloway PD (2010) Single molecule epigenetic analysis in a nanofluidic channel. Anal Chem 82:2480–2487

Cipriany BR, Murphy PJ, Hagarman JA, Cerf A, Latulippe D, Levy SL, Benitez JJ, Tan CP, Topolancik J, Soloway PD, Craighead HG (2012) Real-time analysis and selection of methylated DNA by fluorescence-activated single molecule sorting in a nanofluidic channel. Proc Natl Acad Sci USA 109: 8477–8482

Clark C, Palta P, Joyce CJ, Scott C, Grundberg E, Deloukas P, Palotie A, Coffey AJ (2012) A comparison of the whole genome approach of MeDIP-seq to the targeted approach of the Infinium HumanMethylation450 BeadChip((R)) for methylome profiling. PLoS One 7:e50233

Clark TA, Lu X, Luong K, Dai Q, Boitano M, Turner SW, He C, Korlach J (2013) Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. BMC Biol 11:4

Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H (2009) Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol 4:265–270

Claus R, Wilop S, Hielscher T, Sonnet M, Dahl E, Galm O, Jost E, Plass C (2012) A systematic comparison of quantitative high-resolution DNA methylation analysis and methylation-specific PCR. Epigenetics 7: 772–780

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452:215–219

Coolen MW, Statham AL, Gardiner-Garden M, Clark SJ (2007) Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. Nucleic Acids Res 35:e119

Cottrell SE, Distler J, Goodman NS, Mooney SH, Kluth A, Olek A, Schwope I, Tetzner R, Ziebarth H, Berlin K (2004) A real-time PCR assay for DNA-methylation using methylation-specific blockers. Nucleic Acids Res 32:e10

Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A (2013) Liquid biopsy: monitoring cancer-genetics in the blood. Nat Rev Clin Oncol 10:472–484

Danese A, Richter ML, Chaichoompu K, Fischer DS, Theis FJ, Colome-Tatche M (2021) EpiScanpy: integrated single-cell epigenomic analysis. Nat Commun 12:5228

Daugaard I, Kjeldsen TE, Hager H, Hansen LL, Wojdacz TK (2015) The influence of DNA degradation in formalin-fixed, paraffin-embedded (FFPE) tissue on locus-specific methylation assessment by MS-HRM. Exp Mol Pathol 99:632–640

Daviaud C, Renault V, Mauger F, Deleuze JF, Tost J (2018) Whole-genome bisulfite sequencing using the Ovation(R) ultralow methyl-seq protocol. Methods Mol Biol 1708:83–104

de la Calle-Fabregat C, Rodriguez-Ubreva J, Ciudad L, Ramirez J, Celis R, Azuaga AB, Cuervo A, Graell E, Perez-Garcia C, Diaz-Torne C, Salvador G, Gomez-Puerta JA, Haro I, Sanmarti R, Canete JD, Ballestar E (2022) The synovial and blood monocyte DNA methylomes mirror prognosis, evolution and treatment in early arthritis. JCI Insight 7:e158783

De Meyer T, Mampaey E, Vlemmix M, Denil S, Trooskens G, Renard JP, De Keulenaer S, Dehan P, Menschaert G, Van Criekinge W (2013) Quality evaluation of methyl binding domain based kits for enrichment DNA-methylation sequencing. PLoS One 8: e59068

Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, Egli D, Maherali N, Park IH, Yu J, Daley GQ, Eggan K, Hochedlinger K, Thomson J, Wang W, Gao Y, Zhang K (2009) Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat Biotechnol 27:353–360

Diep D, Plongthongkum N, Gore A, Fung HL, Shoemaker R, Zhang K (2012) Library-free methylation sequencing with bisulfite padlock probes. Nat Methods 9:270–272

Dikow N, Nygren AO, Schouten JP, Hartmann C, Kramer N, Janssen B, Zschocke J (2007) Quantification of the methylation status of the PWS/AS imprinted region: comparison of two approaches based on bisulfite sequencing and methylation-sensitive MLPA. Mol Cell Probes 21:208–215

Ding J, Sharon N, Bar-Joseph Z (2022) Temporal modelling using single-cell transcriptomics. Nat Rev Genet 23(6):355–368

Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, Graf S, Johnson N, Herrero J, Tomazou EM, Thorne NP, Backdahl L, Herberth M, Howe KL, Jackson DK, Miretti MM, Marioni JC, Birney E, Hubbard TJ, Durbin R, Tavare S, Beck S (2008) A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat Biotechnol 26:779–785

Dumenil TD, Wockner LF, Bettington M, McKeone DM, Klein K, Bowdler LM, Montgomery GW, Leggett BA, Whitehall VL (2014) Genome-wide DNA methylation analysis of formalin-fixed paraffin embedded colorectal cancer tissue. Genes Chromosomes Cancer 53:537–548

Dupont JM, Tost J, Jammes H, Gut IG (2004) De novo quantitative bisulfite sequencing using the pyrosequencing technology. Anal Biochem 333:119–127

Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, Shibata D, Danenberg PV, Laird PW (2000) MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res 28:e32

Edwards JR, O'Donnell AH, Rollins RA, Peckham HE, Lee C, Milekic MH, Chanrion B, Fu Y, Su T, Hibshoosh H, Gingrich JA, Haghighi F, Nutter R, Bestor TH (2010) Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. Genome Res 20:972–980

Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, Cantor CR, Field JK, van den Boom D (2005) Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci USA 102:15785–15790

Ehrich M, Zoll S, Sur S, van den Boom D (2007) A new method for accurate assessment of DNA quality after bisulfite treatment. Nucleic Acids Res 35:e29

Eick D, Fritz HJ, Doerfler W (1983) Quantitative determination of 5-methylcytosine in DNA by reverse-phase high-performance liquid chromatography. Anal Biochem 135:165–171

Ernst J, Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. Nat Methods 9:215–216

Ewing AD, Smits N, Sanchez-Luque FJ, Faivre J, Brennan PM, Richardson SR, Cheetham SW, Faulkner GJ (2020) Nanopore sequencing enables comprehensive transposable element epigenomic profiling. Mol Cell 80:915–928 e915

Farlik M, Sheffield NC, Nuzzo A, Datlinger P, Schonegger A, Klughammer J, Bock C (2015) Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. Cell Rep 10:1386–1397

Fazzari MJ, Greally JM (2004) Epigenomics: beyond CpG islands. Nat Rev Genet 5:446–455

Feber A, Wilson G, Zhang L, Presneau N, Idowu B, Down TA, Rakyan VK, Stupka E, Schiza V, Teschendorff A, Schroth GP, Flanagan A, Beck S (2011) Comparative methylome analysis of benign and malignant peripheral nerve sheath tumours. Genome Res 21:515–524

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature 473:398–402

Field SF, Beraldi D, Bachman M, Stewart SK, Beck S, Balasubramanian S (2015) Accurate measurement of 5-methylcytosine and 5-hydroxymethylcytosine in human cerebellum DNA by oxidative bisulfite on an array (OxBS-array). PLoS One 10:e0118202

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 7:461–465

Fogel O, Richard-Miceli C, Tost J (2017) Epigenetic changes in chronic inflammatory diseases. Adv Protein Chem Struct Biol 106:139–189

Foox J, Nordlund J, Lalancette C, Gong T, Lacey M, Lent S, Langhorst BW, Ponnaluri VKC, Williams L, Padmanabhan KR, Cavalcante R, Lundmark A, Butler D, Mozsary C, Gurvitch J, Greally JM, Suzuki M, Menor M, Nasu M, Alonso A, Sheridan C, Scherer A, Bruinsma S, Golda G, Muszynska A, Labaj PP, Campbell MA, Wos F, Raine A, Liljedahl U, Axelsson T, Wang C, Chen Z, Yang Z, Li J, Yang X, Wang H, Melnick A, Guo S, Blume A, Franke V, Ibanez de Caceres I, Rodriguez-Antolin C, Rosas R, Davis JW, Ishii J, Megherbi DB, Xiao W, Liao W, Xu J, Hong H, Ning B, Tong W, Akalin A, Wang Y, Deng Y, Mason CE (2021) The SEQC2 epigenomics quality control (EpiQC) study. Genome Biol 22:332

Fraga MF, Uriol E, Borja Diego L, Berdasco M, Esteller M, Canal MJ, Rodriguez R (2002) High-performance capillary electrophoretic method for the quantification of 5-methyl 2′-deoxycytidine in genomic DNA: application to plant, animal and human cancer tissues. Electrophoresis 23:1677–1681

Fraga MF, Ballestar E, Montoya G, Taysavang P, Wade PA, Esteller M (2003) The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. Nucleic Acids Res 31:1765–1774

Fraszczyk E, Spijkerman AMW, Zhang Y, Brandmaier S, Day FR, Zhou L, Wackers P, Dolle MET, Bloks VW, Gao X, Gieger C, Kooner J, Kriebel J, Picavet HSJ, Rathmann W, Schottker B, Loh M, Verschuren WMM, van Vliet-Ostaptchouk JV, Wareham NJ, Chambers JC, Ong KK, Grallert H, Brenner H, Luijten M, Snieder H (2022) Epigenome-wide association study of incident type 2 diabetes: a meta-analysis of five prospective European cohorts. Diabetologia 65:763–776

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA 89:1827–1831

Garcia-Prieto CA, Alvarez-Errico D, Musulen E, Bueno-Costa A, Vazquez BN, Vaquero A, Esteller M (2022) Validation of a DNA methylation microarray for 285,000 CpG sites in the mouse genome. Epigenetics in press

Garrett-Bakelman FE, Sheridan CK, Kacmarczyk TJ, Ishii J, Betel D, Alonso A, Mason CE, Figueroa ME, Melnick AM (2015) Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base pair resolution. J Vis Exp e52246

Gebhard C, Schwarzfischer L, Pham TH, Schilling E, Klug M, Andreesen R, Rehli M (2006) Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. Cancer Res 66:6118–6128

Genereux DP, Johnson WC, Burden AF, Stoger R, Laird CD (2008) Errors in the bisulfite conversion of DNA: modulating inappropriate- and failed-conversion frequencies. Nucleic Acids Res 36:e150

Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addleman N, Bhardwaj N, Boyle AP, Cayting P, Charos A, Chen DZ, Cheng Y, Clarke D, Eastman C, Euskirchen G, Frietze S, Fu Y, Gertz J, Grubert F, Harmanci A, Jain P, Kasowski M, Lacroute P, Leng J, Lian J, Monahan H, O'Geen H, Ouyang Z, Partridge EC, Patacsil D, Pauli F, Raha D, Ramirez L,

Reddy TE, Reed B, Shi M, Slifer T, Wang J, Wu L, Yang X, Yip KY, Zilberman-Schapira G, Batzoglou S, Sidow A, Farnham PJ, Myers RM, Weissman SM, Snyder M (2012) Architecture of the human regulatory network derived from ENCODE data. Nature 489:91–100

Ghantous A, Saffery R, Cros MP, Ponsonby AL, Hirschfeld S, Kasten C, Dwyer T, Herceg Z, Hernandez-Vargas H (2014) Optimized DNA extraction from neonatal dried blood spots: application in methylome profiling. BMC Biotechnol 14:60

Giesselmann P, Brandl B, Raimondeau E, Bowen R, Rohrandt C, Tandon R, Kretzmer H, Assum G, Galonska C, Siebert R, Ammerpohl O, Heron A, Schneider SA, Ladewig J, Koch P, Schuldt BM, Graham JE, Meissner A, Muller FJ (2019) Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing. Nat Biotechnol 37:1478–1481

Gilmour DS, Lis JT (1984) Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. Proc Natl Acad Sci USA 81:4275–4279

Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, Downs B, Sukumar S, Sedlazeck FJ, Timp W (2020) Targeted nanopore sequencing with Cas9-guided adapter ligation. Nat Biotechnol 38:433–438

Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, Koch S, Bruckl T, Biel M, Carell T (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One 5:e15367

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 27:182–189

Godderis L, Schouteden C, Tabish A, Poels K, Hoet P, Baccarelli AA, Van Landuyt K (2015) Global methylation and hydroxymethylation in DNA from blood and saliva in healthy volunteers. Biomed Res Int 2015: 845041

Goldstein A, Toro PV, Lee J, Silberstein JL, Nakazawa M, Waters I, Cravero K, Chu D, Cochran RL, Kim M, Shinn D, Torquato S, Hughes RM, Pallavajjala A, Carducci MA, Paller CJ, Denmeade SR, Kressel B, Trock BJ, Eisenberger MA, Antonarakis ES, Park BH, Hurley PJ (2017) Detection fidelity of AR mutations in plasma derived cell-free DNA. Oncotarget 8:15651–15662

Gonzalgo ML, Jones PA (1997) Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). Nucleic Acids Res 25:2529–2531

Gopalan S, Wang Y, Harper NW, Garber M, Fazzio TG (2021) Simultaneous profiling of multiple chromatin proteins in the same cells. Mol Cell 81:4736–4746 e4735

Gravina S, Ganapathi S, Vijg J (2015) Single-cell, locus-specific bisulfite sequencing (SLBS) for direct detection of epimutations in DNA methylation patterns. Nucleic Acids Res 43:e93

Gross JA, Pacis A, Chen GG, Barreiro LB, Ernst C, Turecki G (2015) Characterizing 5-hydroxymethylcytosine in human prefrontal cortex at single base resolution. BMC Genomics 16:672

Grunau C, Clark SJ, Rosenthal A (2001) Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res 29:e65

Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A (2011) Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat Protoc 6:468–481

Gu C, Liu S, Wu Q, Zhang L, Guo F (2019) Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes. Cell Res 29:110–123

Guo H, Zhu P, Wu X, Li X, Wen L, Tang F (2013) Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. Genome Res 23:2126–2135

Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, Zhong C, Hu S, Le T, Fan G, Zhu H, Chang Q, Gao Y, Ming GL, Song H (2014) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat Neurosci 17:215–222

Guo H, Zhu P, Guo F, Li X, Wu X, Fan X, Wen L, Tang F (2015) Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. Nat Protoc 10:645–659

Gupta S, Jaworska-Bieniek K, Narod SA, Lubinski J, Wojdacz TK, Jakubowska A (2014) Methylation of the BRCA1 promoter in peripheral blood DNA is associated with triple-negative and medullary breast cancer. Breast Cancer Res Treat 148:615–622

Han D, Lu X, Shih AH, Nie J, You Q, Xu MM, Melnick AM, Levine RL, He C (2016) A highly sensitive and robust method for genome-wide 5hmC profiling of rare cell populations. Mol Cell 63:711–719

Han Y, Zheleznyakova GY, Marincevic-Zuniga Y, Kakhki MP, Raine A, Needhamsen M, Jagodic M (2021) Comparison of EMseq and PBAT methylome library methods for low-input DNA. Epigenetics in press

Harada A, Maehara K, Handa T, Arimura Y, Nogami J, Hayashi-Takanaka Y, Shirahige K, Kurumizaka H, Kimura H, Ohkawa Y (2019) A chromatin integration labelling method enables epigenomic profiling with lower input. Nat Cell Biol 21:287–296

Harrington CT, Lin EI, Olson MT, Eshleman JR (2013) Fundamentals of pyrosequencing. Arch Pathol Lab Med 137:1296–1303

Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE, Fouse SD, Delaney A, Zhao Y, Olshen A, Ballinger T, Zhou X, Forsberg KJ, Gu J, Echipare L, O'Geen H, Lister R,

Pelizzola M, Xi Y, Epstein CB, Bernstein BE, Hawkins RD, Ren B, Chung WY, Gu H, Bock C, Gnirke A, Zhang MQ, Haussler D, Ecker JR, Li W, Farnham PJ, Waterland RA, Meissner A, Marra MA, Hirst M, Milosavljevic A, Costello JF (2010) Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol 28(10):1097–1105

Harrison A, Parle-McDermott A (2011) DNA methylation: a timeline of methods and applications. Front Genet 2:74

Heitzer E, Ulz P, Geigl JB (2015) Circulating tumor DNA as a liquid biopsy for cancer. Clin Chem 61:112–123

Heitzer E, Haque IS, Roberts CES, Speicher MR (2019) Current and future perspectives of liquid biopsies in genomics-driven oncology. Nat Rev Genet 20:71–88

Henkhaus RS, Kim SJ, Kimonis VE, Gold JA, Dykens EM, Driscoll DJ, Butler MG (2012) Methylation-specific multiplex ligation-dependent probe amplification and identification of deletion genetic subtypes in Prader-Willi syndrome. Genet Test Mol Biomarkers 16:178–186

Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci USA 93:9821–9826

Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M, Heyn H, Hvilsom C, Navarro A, Esteller M, Sharp AJ, Marques-Bonet T (2013) Dynamics of DNA methylation in recent human and great ape evolution. PLoS Genet 9:e1003763

Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M, Zhang MQ, Ye K, Bhattacharjee A, Brizuela L, McCombie WR, Wigler M, Hannon GJ, Hicks JB (2009) High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. Genome Res 19:1593–1605

Hoffman MM, Ernst J, Wilder SP, Kundaje A, Harris RS, Libbrecht M, Giardine B, Ellenbogen PM, Bilmes JA, Birney E, Hardison RC, Dunham I, Kellis M, Noble WS (2013) Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res 41: 827–841

Holland PM, Abramson RD, Watson R, Gelfand DH (1991) Detection of specific polymerase chain reaction product by utilizing the 5′-3′ exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci USA 88:7276–7280

Holmes EE, Jung M, Meller S, Leisse A, Sailer V, Zech J, Mengdehl M, Garbe LA, Uhl B, Kristiansen G, Dietrich D (2014) Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. PLoS One 9:e93933

Homig-Holzel C, Savola S (2012) Multiplex ligation-dependent probe amplification (MLPA) in tumor diagnostics and prognostics. Diagn Mol Pathol 21: 189–206

Hoon DS, Spugnardi M, Kuo C, Huang SK, Morton DL, Taback B (2004) Profiling epigenetic inactivation of tumor suppressor genes in tumors and plasma from cutaneous melanoma patients. Oncogene 23:4014–4022

Hop PJ, Zwamborn RAJ, Hannon E, Shireby GL, Nabais MF, Walker EM, van Rheenen W, van Vugt J, Dekker AM, Westeneng HJ, Tazelaar GHP, van Eijk KR, Moisse M, Baird D, Al Khleifat A, Iacoangeli A, Ticozzi N, Ratti A, Cooper-Knock J, Morrison KE, Shaw PJ, Basak AN, Chio A, Calvo A, Moglia C, Canosa A, Brunetti M, Grassano M, Gotkine M, Lerner Y, Zabari M, Vourc'h P, Corcia P, Couratier P, Mora Pardina JS, Salas T, Dion P, Ross JP, Henderson RD, Mathers S, McCombe PA, Needham M, Nicholson G, Rowe DB, Pamphlett R, Mather KA, Sachdev PS, Furlong S, Garton FC, Henders AK, Lin T, Ngo ST, Steyn FJ, Wallace L, Williams KL, Consortium B, Brain MC, Neto MM, Cauchi RJ, Blair IP, Kiernan MC, Drory V, Povedano M, de Carvalho M, Pinto S, Weber M, Rouleau GA, Silani V, Landers JE, Shaw CE, Andersen PM, McRae AF, van Es MA, Pasterkamp RJ, Wray NR, McLaughlin RL, Hardiman O, Kenna KP, Tsai E, Runz H, Al-Chalabi A, van den Berg LH, Van Damme P, Mill J, Veldink JH, Heijmans BT, PAC TH, van Meurs J, Jansen R, Franke L, Boomsma DI, Pool R, van Dongen J, Hottenga JJ, van Greevenbroek MMJ, Stehouwer CDA, van der Kallen CJH, Schalkwijk CG, Wijmenga C, Franke L, Zhernakova S, Tigchelaar EF, Slagboom PE, Beekman M, Deelen J, van Heemst D, Veldink JH, van den Berg LH, van Duijn CM, Hofman BA, Isaacs A, Uitterlinden AG, van Meurs J, Jhamai PM, Verbiest M, Suchiman HED, Verkerk M, van der Breggen R, van Rooij J, Lakenberg N, Mei H, van Iterson M, van Galen M, Bot J, Zhernakova DV, Jansen R, van't Hof P, Deelen P, Nooren I, PAC TH, Heijmans BT, Moed M, Franke L, Vermaat M, Zhernakova DV, Luijk R, Jan Bonder M, van Iterson M, Deelen P, van Dijk F, van Galen M, Arindrarto W, Kielbasa SM, Swertz MA, van Zwet EW, Jansen R, PAC TH, Heijmans BT, Al-Chalabi A, Wray NR, Bensimon G, Hardiman O, Chio A, Veldink JH, Smith GD, Mill J (2022) Genome-wide study of DNA methylation shows alterations in metabolic, inflammatory, and cholesterol pathways in ALS. Sci Transl Med 14:eabj0264

Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13:86

Houseman EA, Johnson KC, Christensen BC (2016) OxyBS: estimation of 5-methylcytosine and 5-hydroxymethylcytosine from tandem-treated oxidative bisulfite and bisulfite DNA. Bioinformatics 32: 2505–2507

How Kit A, Nielsen HM, Tost J (2012) DNA methylation based biomarkers: practical considerations and applications. Biochimie 94:2314–2337

How Kit A, Mazaleyrat N, Daunay A, Nielsen HM, Terris B, Tost J (2013) Sensitive detection of KRAS mutations using enhanced-ice-COLD-PCR mutation enrichment and direct sequence identification. Hum Mutat 34:1568–1580

Huang DJ, Nelson MR, Zimmermann B, Dudarewicz L, Wenzel F, Spiegel R, Nagy B, Holzgreve W, Hahn S (2006) Reliable detection of trisomy 21 using MALDI-TOF mass spectrometry. Genet Med 8:728–734

Huang J, Renault V, Sengenes J, Touleimat N, Michel S, Lathrop M, Tost J (2012) MeQA: a pipeline for MeDIP-seq data quality assessment and analysis. Bioinformatics 28:587–588

Huang HY, Li J, Tang Y, Huang YX, Chen YG, Xie YY, Zhou ZY, Chen XY, Ding SY, Luo MF, Jin CN, Zhao LS, Xu JT, Zhou Y, Lin YC, Hong HC, Zuo HL, Hu SY, Xu PY, Li X, Huang HD (2021) MethHC 2.0: information repository of DNA methylation and gene expression in human cancer. Nucleic Acids Res 49: D1268–D1275

Hur K, Cejas P, Feliu J, Moreno-Rubio J, Burgos E, Boland CR, Goel A (2014) Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. Gut 63:635–646

Ignatiadis M, Sledge GW, Jeffrey SS (2021) Liquid biopsy enters the clinic—implementation issues and future challenges. Nat Rev Clin Oncol 18:297–312

Ilse P, Biesterfeld S, Pomjanski N, Wrobel C, Schramm M (2014) Analysis of SHOX2 methylation as an aid to cytology in lung cancer diagnosis. Cancer Genomics Proteomics 11:251–258

Inamura K, Yamauchi M, Nishihara R, Lochhead P, Qian ZR, Kuchiba A, Kim SA, Mima K, Sukawa Y, Jung S, Zhang X, Wu K, Cho E, Chan AT, Meyerhardt JA, Harris CC, Fuchs CS, Ogino S (2014) Tumor LINE-1 methylation level and microsatellite instability in relation to colorectal cancer prognosis. J Natl Cancer Inst 106:dju195

Irizarry RA, Ladd-Acosta C, Carvalho B, Wu H, Brandenburg SA, Jeddeloh JA, Wen B, Feinberg AP (2008) Comprehensive high-throughput arrays for relative methylation (CHARM). Genome Res 18:780–790

Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash JB, Sabunciyan S, Feinberg AP (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet 41:178–186

Ivanov M, Kals M, Kacevska M, Metspalu A, Ingelman-Sundberg M, Milani L (2013) In-solution hybrid capture of bisulfite-converted DNA for targeted bisulfite sequencing of 174 ADME genes. Nucleic Acids Res 41:e72

Jain M, Olsen HE, Paten B, Akeson M (2016) The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17:239

Jasmine F, Rahaman R, Roy S, Raza M, Paul R, Rakibuz-Zaman M, Paul-Brutus R, Dodsworth C, Kamal M, Ahsan H, Kibriya MG (2012) Interpretation of genome-wide infinium methylation data from ligated DNA in formalin-fixed, paraffin-embedded paired tumor and normal tissue. BMC Res Notes 5:117

Jedynak P, Tost J, Calafat AM, Bourova-Flin E, Broseus L, Busato F, Forhan A, Heude B, Jakobi M, Schwartz J, Slama R, Vaiman D, Lepeule J, Philippat C (2022) Pregnancy exposure to phthalates and DNA methylation in male placenta—an epigenome-wide association study. Environ Int 160:107054

Jelinek J, Liang S, Lu Y, He R, Ramagli LS, Shpall EJ, Estecio MR, Issa JP (2012) Conserved DNA methylation patterns in healthy blood cells and extensive changes in leukemia measured by a new quantitative technique. Epigenetics 7:1368–1378

Jelinek J, Lee JT, Cesaroni M, Madzo J, Liang S, Lu Y, Issa JJ (2018) Digital restriction enzyme analysis of methylation (DREAM). Methods Mol Biol 1708:247–265

Jenkins TG, Aston KI, Pflueger C, Cairns BR, Carrell DT (2014) Age-associated sperm DNA methylation alterations: possible implications in offspring disease susceptibility. PLoS Genet 10:e1004458

Jessen WJ, Dhasarathy A, Hoose SA, Carvin CD, Risinger AL, Kladde MP (2004) Mapping chromatin structure in vivo using DNA methyltransferases. Methods 33: 68–80

Ji L, Sasaki T, Sun X, Ma P, Lewis ZA, Schmitz RJ (2014) Methylated DNA is over-represented in whole-genome bisulfite sequencing data. Front Genet 5:341

Jin SG, Kadam S, Pfeifer GP (2010) Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res 38:e125

Jung M, Kadam S, Xiong W, Rauch TA, Jin SG, Pfeifer GP (2015) MIRA-seq for DNA methylation analysis of CpG islands. Epigenomics 7:695–706

Kabesch M, Tost J (2020) Recent findings in the genetics and epigenetics of asthma and allergy. Semin Immunopathol 42:43–60

Kacmarczyk TJ, Fall MP, Zhang X, Xin Y, Li Y, Alonso A, Betel D (2018) "Same difference": comprehensive evaluation of four DNA methylation measurement platforms. Epigenetics Chromatin 11:21

Kagey JD, Kapoor-Vazirani P, McCabe MT, Powell DR, Vertino PM (2010) Long-term stability of demethylation after transient exposure to 5-aza-2′-deoxycytidine correlates with sustained RNA polymerase II occupancy. Mol Cancer Res 8:1048–1059

Kaminsky ZA, Assadzadeh A, Flanagan J, Petronis A (2005) Single nucleotide extension technology for quantitative site-specific evaluation of metC/C in GC-rich regions. Nucleic Acids Res 33:e95

Kang I, Wang Y, Reagan C, Fu Y, Wang MX, Gu LQ (2013) Designing DNA interstrand lock for locus-specific methylation detection in a nanopore. Sci Rep 3:2381

Kang S, Li Q, Chen Q, Zhou Y, Park S, Lee G, Grimes B, Krysan K, Yu M, Wang W, Alber F, Sun F, Dubinett SM, Li W, Zhou XJ (2017) CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. Genome Biol 18:53

Kantarjian H, Issa JP, Rosenfeld CS, Bennett JM, Albitar M, DiPersio J, Klimek V, Slack J, de Castro C, Ravandi F, Helmer R 3rd, Shen L, Nimer SD, Leavitt R, Raza A, Saba H (2006) Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study. Cancer 106:1794–1803

Kapourani CA, Sanguinetti G (2019) Melissa: Bayesian clustering and imputation of single-cell methylomes. Genome Biol 20:61

Kapourani CA, Argelaguet R, Sanguinetti G, Vallejos CA (2021) scMET: Bayesian modeling of DNA methylation heterogeneity at single-cell resolution. Genome Biol 22:114

Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10000 daltons. Anal Chem 60:2299–2301

Karemaker ID, Vermeulen M (2018) Single-cell DNA methylation profiling: technologies and biological applications. Trends Biotechnol 36:952–965

Karimi M, Johansson S, Stach D, Corcoran M, Grander D, Schalling M, Bakalkin G, Lyko F, Larsson C, Ekstrom TJ (2006) LUMA (LUminometric Methylation Assay)—a high throughput method to the analysis of genomic DNA methylation. Exp Cell Res 312:1989–1995

Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff S (2019) CUT & Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun 10:1930

Kaya-Okur HS, Janssens DH, Henikoff JG, Ahmad K, Henikoff S (2020) Efficient low-cost chromatin profiling with CUT & Tag. Nat Protoc 15:3264–3283

Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA (2012) Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res 22:2497–2506

Kiihl SF (2021) Bioinformatic estimation of DNA methylation and hydroxymethylation proportions. Methods Mol Biol 2272:141–162

Kinney SM, Chin HG, Vaisvila R, Bitinaite J, Zheng Y, Esteve PO, Feng S, Stroud H, Jacobsen SE, Pradhan S (2011) Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes. J Biol Chem 286:24685–24693

Kirschner SA, Hunewald O, Meriaux SB, Brunnhoefer R, Muller CP, Turner JD (2016) Focussing reduced representation CpG sequencing through judicious restriction enzyme choice. Genomics 107:109–119

Kitazawa H, Okuno Y, Muramatsu H, Aoki K, Murakami N, Wakamatsu M, Suzuki K, Narita K, Kataoka S, Ichikawa D, Hamada M, Taniguchi R, Kawashima N, Nishikawa E, Narita A, Nishio N, Hama A, Loh ML, Stieglitz E, Kojima S, Takahashi Y (2021) Simple and robust methylation test for risk stratification of patients with juvenile myelomonocytic leukemia. Blood Adv 5:5507–5518

Klughammer J, Kiesel B, Roetzer T, Fortelny N, Nemc A, Nenning KH, Furtner J, Sheffield NC, Datlinger P, Peter N, Nowosielski M, Augustin M, Mischkulnig M, Strobel T, Alpar D, Erguner B, Senekowitsch M, Moser P, Freyschlag CF, Kerschbaumer J, Thome C, Grams AE, Stockhammer G, Kitzwoegerer M, Oberndorfer S, Marhold F, Weis S, Trenkler J, Buchroithner J, Pichler J, Haybaeck J, Krassnig S, Mahdy Ali K, von Campe G, Payer F, Sherif C, Preiser J, Hauser T, Winkler PA, Kleindienst W, Wurtz F, Brandner-Kokalj T, Stultschnig M, Schweiger S, Dieckmann K, Preusser M, Langs G, Baumann B, Knosp E, Widhalm G, Marosi C, Hainfellner JA, Woehrer A, Bock C (2018) The DNA methylation landscape of glioblastoma disease progression shows extensive heterogeneity in time and space. Nat Med 24:1611–1624

Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y, Suzuki Y, Kono T (2012) Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet 8:e1002440

Kobayashi H, Sakurai T, Miura F, Imai M, Mochiduki K, Yanagisawa E, Sakashita A, Wakai T, Suzuki Y, Ito T, Matsui Y, Kono T (2013) High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. Genome Res 23:616–627

Koch A, Joosten SC, Feng Z, de Ruijter TC, Draht MX, Melotte V, Smits KM, Veeck J, Herman JG, Van Neste L, Van Criekinge W, De Meyer T, van Engeland M (2018) Analysis of DNA methylation in cancer: location revisited. Nat Rev Clin Oncol 15:459–466

Korbie D, Lin E, Wall D, Nair SS, Stirzaker C, Clark SJ, Trau M (2015) Multiplex bisulfite PCR resequencing of clinical FFPE DNA. Clin Epigenetics 7:28

Kremer D, Metzger S, Kolb-Bachofen V, Kremer D (2012) Quantitative measurement of genome-wide DNA methylation by a reliable and cost-efficient enzyme-linked immunosorbent assay technique. Anal Biochem 422:74–78

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

Kristensen LS, Mikeska T, Krypuy M, Dobrovic A (2008) Sensitive Melting Analysis after Real Time- Methylation Specific PCR (SMART-MSP): high-throughput and probe-free quantitative DNA methylation detection. Nucleic Acids Res 36:e42

Kristensen LS, Wojdacz TK, Thestrup BB, Wiuf C, Hager H, Hansen LL (2009) Quality assessment of DNA derived from up to 30 years old formalin fixed paraffin embedded (FFPE) tissue for PCR-based methylation analysis using SMART-MSP and MS-HRM. BMC Cancer 9:453

Kriukiene E, Labrie V, Khare T, Urbanaviciute G, Lapinaite A, Koncevicius K, Li D, Wang T, Pai S, Ptak C, Gordevicius J, Wang SC, Petronis A, Klimasauskas S (2013) DNA unmethylome profiling by covalent capture of CpG sites. Nat Commun 4:2190

Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27:1571–1572

Lan X, Adams C, Landers M, Dudas M, Krissinger D, Marnellos G, Bonneville R, Xu M, Wang J, Huang TH, Meredith G, Jin VX (2011) High resolution detection and analysis of CpG dinucleotides methylation using MBD-Seq technology. PLoS One 6:e22226

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L, Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shoresh N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ, Snyder M (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res 22:1813–1831

Laszlo AH, Derrington IM, Brinkerhoff H, Langford KW, Nova IC, Samson JM, Bartlett JJ, Pavlenok M, Gundlach JH (2013) Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. Proc Natl Acad Sci USA 110: 18904–18909

Le T, Kim KP, Fan G, Faull KF (2011) A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. Anal Biochem 412:203–209

Lee EJ, Pei L, Srivastava G, Joshi T, Kushwaha G, Choi JH, Robertson KD, Wang X, Colbourne JK, Zhang L, Schroth GP, Xu D, Zhang K, Shi H (2011) Targeted bisulfite sequencing by solution hybrid selection and massively parallel sequencing. Nucleic Acids Res 39: e127

Lee YK, Jin S, Duan S, Lim YC, Ng DP, Lin XM, Yeo GS, Ding C (2014) Improved reduced representation bisulfite sequencing for epigenomic profiling of clinical samples. Biol Proced Online 16:1

Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheim J, Vaknin-Dembinsky A, Rubertsson S, Nellgard B, Blennow K, Zetterberg H, Spalding K, Haller MJ, Wasserfall CH, Schatz DA, Greenbaum CJ, Dorrell C, Grompe M, Zick A, Hubert A, Maoz M, Fendrich V, Bartsch DK, Golan T, Ben Sasson SA, Zamir G, Razin A, Cedar H, Shapiro AM, Glaser B, Shemer R, Dor Y (2016) Identification of tissue-specific cell death using methylation patterns of circulating DNA. Proc Natl Acad Sci USA 113: E1826–E1834

Leong KJ, James J, Wen K, Taniere P, Morton DG, Bach SP, Matthews GM (2013) Impact of tissue processing, archiving and enrichment techniques on DNA methylation yield in rectal carcinoma. Exp Mol Pathol 95: 343–349

Li LC, Dahiya R (2002) MethPrimer: designing primers for methylation PCRs. Bioinformatics 18:1427–1431

Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, Zheng H, Yu J, Wu H, Sun J, Zhang H, Chen Q, Luo R, Chen M, He Y, Jin X, Zhang Q, Yu C, Zhou G, Huang Y, Cao H, Zhou X, Guo S, Hu X, Li X, Kristiansen K, Bolund L, Xu J, Wang W, Yang H, Wang J, Li R, Beck S, Zhang X (2010a) The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol 8:e1000533

Li N, Ye M, Li Y, Yan Z, Butcher LM, Sun J, Han X, Chen Q, Zhang X, Wang J (2010b) Whole genome DNA methylation analysis based on high throughput sequencing technology. Methods 52:203–212

Li Q, Suzuki M, Wendt J, Patterson N, Eichten SR, Hermanson PJ, Green D, Jeddeloh J, Richmond T, Rosenbaum H, Burgess D, Springer NM, Greally JM (2015a) Post-conversion targeted capture of modified cytosines in mammalian and plant genomes. Nucleic Acids Res 43:e81

Li D, Zhang B, Xing X, Wang T (2015b) Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. Methods 72:29–40

Li H, Hedmer M, Wojdacz T, Hossain MB, Lindh CH, Tinnerberg H, Albin M, Broberg K (2015c) Oxidative stress, telomere shortening, and DNA methylation in relation to low-to-moderate occupational exposure to welding fumes. Environ Mol Mutagen 56:684–693

Li H, Akerman G, Liden C, Alhamdow A, Wojdacz TK, Broberg K, Albin M (2016) Alterations of telomere length and DNA methylation in hairdressers: A cross-sectional study. Environ Mol Mutagen 57:159–167

Li W, Li Q, Kang S, Same M, Zhou Y, Sun C, Liu CC, Matsuoka L, Sher L, Wong WH, Alber F, Zhou XJ (2018) CancerDetector: ultrasensitive and non-invasive cancer detection at the resolution of individual reads using cell-free DNA methylation sequencing data. Nucleic Acids Res 46:e89

Li R, Grimm SA, Wade PA (2021) CUT & Tag-BS for simultaneous profiling of histone modification and DNA methylation with high efficiency and low cost. Cell Rep Methods 1:100118

Libertini E, Heath SC, Hamoudi RA, Gut M, Ziller MJ, Herrero J, Czyz A, Ruotti V, Stunnenberg HG, Frontini M, Ouwehand WH, Meissner A, Gut IG, Beck S (2016) Saturation analysis for whole-genome bisulfite sequencing data. Nat Biotechnol 34:691–693

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133:523–536

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:315–322

Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, Antosiewicz-Bourget J, O'Malley R, Castanon R, Klugman S, Downes M, Yu R, Stewart R, Ren B, Thomson JA, Evans RM, Ecker JR (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature 471:68–73

Liu Z, Liu S, Xie Z, Blum W, Perrotti D, Paschka P, Klisovic R, Byrd J, Chan KK, Marcucci G (2007) Characterization of in vitro and in vivo hypomethylating effects of decitabine in acute myeloid leukemia by a rapid, specific and sensitive LC-MS/MS method. Nucleic Acids Res 35:e31

Liu Y, Siejka-Zielinska P, Velikova G, Bi Y, Yuan F, Tomkova M, Bai C, Chen L, Schuster-Bockler B, Song CX (2019) Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. Nat Biotechnol 37:424–429

Liu Y, Cheng J, Siejka-Zielinska P, Weldon C, Roberts H, Lopopolo M, Magri A, D'Arienzo V, Harris JM, McKeating JA, Song CX (2020a) Accurate targeted long-read DNA methylation and hydroxymethylation sequencing with TAPS. Genome Biol 21:54

Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, CCGA Consortium (2020b) Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann Oncol 31:745–759

Lo YM, Wong IH, Zhang J, Tein MS, Ng MH, Hjelm NM (1999) Quantitative analysis of aberrant p16 methylation using real-time quantitative methylation-specific polymerase chain reaction. Cancer Res 59:3899–3903

Lo YMD, Han DSC, Jiang P, Chiu RWK (2021) Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. Science 372:eaaw3616

Lopez V, Fernandez AF, Fraga MF (2017) The role of 5-hydroxymethylcytosine in development, aging and age-related diseases. Ageing Res Rev 37:28–38

Lorthongpanich C, Cheow LF, Balu S, Quake SR, Knowles BB, Burkholder WF, Solter D, Messerschmidt DM (2013) Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. Science 341:1110–1112

Lu H, Yuan Z, Tan T, Wang J, Zhang J, Luo HJ, Xia Y, Ji W, Gao F (2015) Improved tagmentation-based whole-genome bisulfite sequencing for input DNA from less than 100 mammalian cells. Epigenomics 7:47–56

Lun FM, Chiu RW, Sun K, Leung TY, Jiang P, Chan KC, Sun H, Lo YM (2013) Noninvasive prenatal methylomic analysis by genomewide bisulfite sequencing of maternal plasma DNA. Clin Chem 59:1583–1594

Lungu C, Pinter S, Broche J, Rathert P, Jeltsch A (2017) Modular fluorescence complementation sensors for live cell detection of epigenetic signals at endogenous genomic sites. Nat Commun 8:649

Luo H, Wei W, Ye Z, Zheng J, Xu RH (2021) Liquid biopsy of methylation biomarkers in cell-free DNA. Trends Mol Med 27:482–500

Luth T, Labeta J, Schaake S, Wohlers I, Pozojevic J, Jamora RDG, Rosales RL, Bruggemann N, Saranza G, Diesta CCE, Schluter K, Tse R, Reyes CJ, Brand M, Busch H, Klein C, Westenberger A, Trinh J (2022) Elucidating hexanucleotide repeat number and methylation within the X-linked dystonia-Parkinsonism (XDP)-related SVA retrotransposon in TAF1 with nanopore sequencing. Genes (Basel) 13

Lutsik P, Feuerbach L, Arand J, Lengauer T, Walter J, Bock C (2011) BiQ Analyzer HT: locus-specific analysis of DNA methylation by high-throughput bisulfite sequencing. Nucleic Acids Res 39:W551–W556

Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R (2010) The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biol 8:e1000506

Lynch M, Ramalingam N (2019) Integrated fluidic circuits for single-cell omics and multi-omics applications. Adv Exp Med Biol 1129:19–26

Manrao EA, Derrington IM, Pavlenok M, Niederweis M, Gundlach JH (2011) Nucleotide discrimination with DNA immobilized in the MspA nanopore. PLoS One 6:e25723

Marques-Rocha JL, Milagro FI, Mansego ML, Mourao DM, Martinez JA, Bressan J (2016) LINE-1 methylation is positively associated with healthier lifestyle but inversely related to body fat mass in healthy young individuals. Epigenetics 11:49–60

Marshall LL, Killinger BA, Ensink E, Li P, Li KX, Cui W, Lubben N, Weiland M, Wang X, Gordevicius J, Coetzee GA, Ma J, Jovinge S, Labrie V (2020) Epigenomic analysis of Parkinson's disease neurons identifies Tet2 loss as neuroprotective. Nat Neurosci 23:1203–1214

Martin-Herranz DE, Ribeiro AJM, Krueger F, Thornton JM, Reik W, Stubbs TM (2017) cuRRBS: simple and robust evaluation of enzyme combinations for reduced representation approaches. Nucleic Acids Res 45:11559–11569

Masser DR, Berg AS, Freeman WM (2013) Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. Epigenetics Chromatin 6:33

Matarese F, Carrillo-de Santa Pau E, Stunnenberg HG (2011) 5-Hydroxymethylcytosine: a new kid on the epigenetic block? Mol Syst Biol 7:562

Mauger F, Kernaleguen M, Lallemand C, Kristensen VN, Deleuze JF, Tost J (2018) Enrichment of methylated molecules using enhanced-ice-co-amplification at lower denaturation temperature-PCR (E-ice-COLD-

PCR) for the sensitive detection of disease-related hypermethylation. Epigenomics 10:525–537

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466: 253–257

McCartney DL, Hillary RF, Conole ELS, Banos DT, Gadd DA, Walker RM, Nangle C, Flaig R, Campbell A, Murray AD, Maniega SM, Valdes-Hernandez MDC, Harris MA, Bastin ME, Wardlaw JM, Harris SE, Porteous DJ, Tucker-Drob EM, McIntosh AM, Evans KL, Deary IJ, Cox SR, Robinson MR, Marioni RE (2022) Blood-based epigenome-wide analyses of cognitive abilities. Genome Biol 23:26

McInroy GR, Beraldi D, Raiber EA, Modrzynska K, van Delft P, Billker O, Balasubramanian S (2016) Enhanced methylation analysis by recovery of unsequenceable fragments. PLoS One 11:e0152322

Meers MP, Bryson TD, Henikoff JG, Henikoff S (2019) Improved CUT & RUN chromatin profiling tools. elife 8:e46314

Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res 33:5868–5877

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454:766–770

Melnekoff DT, Lagana A (2022) Single-cell sequencing technologies in precision oncology. Adv Exp Med Biol 1361:269–282

Melnikov AA, Gartenhaus RB, Levenson AS, Motchoulskaia NA, Levenson Chernokhvostov VV (2005) MSRE-PCR for analysis of gene-specific DNA methylation. Nucleic Acids Res 33:e93

Mereu E, Lafzi A, Moutinho C, Ziegenhain C, McCarthy DJ, Alvarez-Varela A, Batlle E, Sagar GD, Lau JK, Boutet SC, Sanada C, Ooi A, Jones RC, Kaihara K, Brampton C, Talaga Y, Sasagawa Y, Tanaka K, Hayashi T, Braeuning C, Fischer C, Sauer S, Trefzer T, Conrad C, Adiconis X, Nguyen LT, Regev A, Levin JZ, Parekh S, Janjic A, Wange LE, Bagnoli JW, Enard W, Gut M, Sandberg R, Nikaido I, Gut I, Stegle O, Heyn H (2020) Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. Nat Biotechnol 38:747–755

Miceli-Richard C, Wang-Renault SF, Boudaoud S, Busato F, Lallemand C, Bethune K, Belkhir R, Nocturne G, Mariette X, Tost J (2016) Overlap between differentially methylated DNA regions in blood B lymphocytes and genetic at-risk loci in primary Sjogren's syndrome. Ann Rheum Dis 75: 933–940

Mikeska T, Candiloro IL, Dobrovic A (2010) The implications of heterogeneous DNA methylation for the accurate quantification of methylation. Epigenomics 2:561–573

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448:553–560

Miller OJ, Schnedl W, Allen J, Erlanger BF (1974) 5-Methylcytosine localised in mammalian constitutive heterochromatin. Nature 251:636–637

Miner BE, Stoger RJ, Burden AF, Laird CD, Hansen RS (2004) Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. Nucleic Acids Res 32:e135

Miura F, Ito T (2015) Highly sensitive targeted methylome sequencing by post-bisulfite adaptor tagging. DNA Res 22:13–18

Miura F, Enomoto Y, Dairiki R, Ito T (2012) Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. Nucleic Acids Res 40:e136

Moison C, Senamaud-Beaufort C, Fourriere L, Champion C, Ceccaldi A, Lacomme S, Daunay A, Tost J, Arimondo PB, Guieysse-Peugeot AL (2013) DNA methylation associated with polycomb repression in retinoic acid receptor beta silencing. FASEB J 27:1468–1478

Moison C, Assemat F, Daunay A, Tost J, Guieysse-Peugeot AL, Arimondo PB (2014) Synergistic chromatin repression of the tumor suppressor gene RARB in human prostate cancers. Epigenetics 9:477–482

Moran S, Vizoso M, Martinez-Cardus A, Gomez A, Matias-Guiu X, Chiavenna SM, Fernandez AG, Esteller M (2014) Validation of DNA methylation profiling in formalin-fixed paraffin-embedded samples using the Infinium HumanMethylation450 Microarray. Epigenetics 9:829–833

Moran S, Arribas C, Esteller M (2016) Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics 8:389–399

Morris TJ, Beck S (2015) Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. Methods 72:3–8

Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, Beck S (2014) ChAMP: 450k chip analysis methylation pipeline. Bioinformatics 30:428–430

Morrison J, Koeman JM, Johnson BK, Foy KK, Beddows I, Zhou W, Chesla DW, Rossell LL, Siegwald EJ, Adams M, Shen H (2021) Evaluation of

whole-genome DNA methylation sequencing library preparation protocols. Epigenetics Chromatin 14:28

Morselli M, Farrell C, Rubbi L, Fehling HL, Henkhaus R, Pellegrini M (2021) Targeted bisulfite sequencing for biomarker discovery. Methods 187:13–27

Moss J, Magenheim J, Neiman D, Zemmour H, Loyfer N, Korach A, Samet Y, Maoz M, Druid H, Arner P, Fu KY, Kiss E, Spalding KL, Landesberg G, Zick A, Grinshpun A, Shapiro AMJ, Grompe M, Wittenberg AD, Glaser B, Shemer R, Kaplan T, Dor Y (2018) Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun 9:5068

Muller F, Scherer M, Assenov Y, Lutsik P, Walter J, Lengauer T, Bock C (2019) RnBeads 2.0: comprehensive analysis of DNA methylation data. Genome Biol 20:55

Mund C, Hackanson B, Stresemann C, Lubbert M, Lyko F (2005) Characterization of DNA demethylation effects induced by 5-Aza-2′-deoxycytidine in patients with myelodysplastic syndrome. Cancer Res 65:7086–7090

Murphy PJ, Cipriany BR, Wallin CB, Ju CY, Szeto K, Hagarman JA, Benitez JJ, Craighead HG, Soloway PD (2013) Single-molecule analysis of combinatorial epigenomic states in normal and tumor cells. Proc Natl Acad Sci USA 110:7772–7777

Nair SS, Coolen MW, Stirzaker C, Song JZ, Statham AL, Strbenac D, Robinson MD, Clark SJ (2011) Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. Epigenetics 6: 34–44

Nair SS, Luu PL, Qu W, Maddugoda M, Huschtscha L, Reddel R, Chenevix-Trench G, Toso M, Kench JG, Horvath LG, Hayes VM, Stricker PD, Hughes TP, White DL, Rasko JEJ, Wong JJ, Clark SJ (2018) Guidelines for whole genome bisulphite sequencing of intact and FFPET DNA on the Illumina HiSeq X Ten. Epigenetics Chromatin 11:24

Nazor KL, Boland MJ, Bibikova M, Klotzle B, Yu M, Glenn-Pratola VL, Schell JP, Coleman RL, Cabral-da-Silva MC, Schmidt U, Peterson SE, He C, Loring JF, Fan JB (2014) Application of a low cost array-based technique —TAB-array—for quantifying and mapping both 5mC and 5hmC at single base resolution in human pluripotent stem cells. Genomics 104:358–367

Nestor C, Ruzov A, Meehan R, Dunican D (2010) Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. Biotechniques 48: 317–319

Nestor CE, Ottaviano R, Reinhardt D, Cruickshanks HA, Mjoseng HK, McPherson RC, Lentini A, Thomson JP, Dunican DS, Pennings S, Anderton SM, Benson M, Meehan RR (2015) Rapid reprogramming of epigenetic and transcriptional profiles in mammalian culture systems. Genome Biol 16:11

Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, Liu CL, Neal JW, Wakelee HA, Merritt RE, Shrager JB, Loo BW Jr, Alizadeh AA, Diehn M (2014) An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med 20:548–554

Newton K, Jorgensen NM, Wallace AJ, Buchanan DD, Lalloo F, McMahon RF, Hill J, Evans DG (2014) Tumour MLH1 promoter region methylation testing is an effective prescreen for Lynch Syndrome (HNPCC). J Med Genet 51:789–796

Nikolaidis G, Raji OY, Markopoulou S, Gosney JR, Bryan J, Warburton C, Walshaw M, Sheard J, Field JK, Liloglou T (2012) DNA methylation biomarkers offer improved diagnostic efficiency in lung cancer. Cancer Res 72:5692–5701

Nones K, Waddell N, Song S, Patch AM, Miller D, Johns A, Wu J, Kassahn KS, Wood D, Bailey P, Fink L, Manning S, Christ AN, Nourse C, Kazakoff S, Taylor D, Leonard C, Chang DK, Jones MD, Thomas M, Watson C, Pinese M, Cowley M, Rooman I, Pajic M, Apgi BG, Malpaga A, Corbo V, Crippa S, Falconi M, Zamboni G, Castelli P, Lawlor RT, Gill AJ, Scarpa A, Pearson JV, Biankin AV, Grimmond SM (2014) Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signaling. Int J Cancer 135:1110–1118

Nygren AO, Ameziane N, Duarte HM, Vijzelaar RN, Waisfisz Q, Hess CJ, Schouten JP, Errami A (2005) Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. Nucleic Acids Res 33: e128

Nygren AO, Dean J, Jensen TJ, Kruse S, Kwong W, van den Boom D, Ehrich M (2010) Quantification of fetal DNA by use of methylation-based DNA discrimination. Clin Chem 56:1627–1635

Oakes CC, La Salle S, Robaire B, Trasler JM (2006) Evaluation of a quantitative DNA methylation analysis technique using methylation-sensitive/dependent restriction enzymes and real-time PCR. Epigenetics 1: 146–152

Ogino S, Kawasaki T, Brahmandam M, Yan L, Cantor M, Namgyal C, Mino-Kenudson M, Lauwers GY, Loda M, Fuchs CS (2005) Sensitive sequencing method for KRAS mutation detection by Pyrosequencing. J Mol Diagn 7:413–421

Ogino S, Kawasaki T, Brahmandam M, Cantor M, Kirkner GJ, Spiegelman D, Makrigiorgos GM, Weisenberger DJ, Laird PW, Loda M, Fuchs CS (2006) Precision and performance characteristics of bisulfite conversion and real-time PCR (MethyLight) for quantitative DNA methylation analysis. J Mol Diagn 8:209–217

Olova NN (2021) ELISA-based quantitation of global 5hmC levels. Methods Mol Biol 2272:45–63

Olova N, Krueger F, Andrews S, Oxley D, Berrens RV, Branco MR, Reik W (2018) Comparison of whole-

genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. Genome Biol 19:33
Ong ML, Tan PY, MacIsaac JL, Mah SM, Buschdorf JP, Cheong CY, Stunkel W, Chan L, Gluckman PD, Chng K, Kobor MS, Meaney MJ, Holbrook JD (2014) Infinium monkeys: Infinium 450K array for the Cynomolgus macaque (Macaca fascicularis). G3 4:1227–1234
Paliwal A, Temkin AM, Kerkel K, Yale A, Yotova I, Drost N, Lax S, Nhan-Chang CL, Powell C, Borczuk A, Aviv A, Wapner R, Chen X, Nagy PL, Schork N, Do C, Torkamani A, Tycko B (2013) Comparative anatomy of chromosomal domains with imprinted and non-imprinted allele-specific DNA methylation. PLoS Genet 9:e1003622
Pantel K, Alix-Panabieres C (2019) Liquid biopsy and minimal residual disease—latest advances and implications for cure. Nat Rev Clin Oncol 16:409–424
Papageorgiou EA, Fiegler H, Rakyan V, Beck S, Hulten M, Lamnissou K, Carter NP, Patsalis PC (2009) Sites of differential DNA methylation between placenta and peripheral blood: molecular markers for noninvasive prenatal diagnosis of aneuploidies. Am J Pathol 174:1609–1618
Papageorgiou EA, Karagrigoriou A, Tsaliki E, Velissariou V, Carter NP, Patsalis PC (2011) Fetal-specific DNA methylation ratio permits noninvasive prenatal diagnosis of trisomy 21. Nat Med 17:510–513
Pardo CE, Darst RP, Nabilsi NH, Delmas AL, Kladde MP (2011) Simultaneous single-molecule mapping of protein-DNA interactions and DNA methylation by MAPit. Curr Protoc Mol Biol Chapter 21:Unit 21.22
Partridge EC, Chhetri SB, Prokop JW, Ramaker RC, Jansen CS, Goh ST, Mackiewicz M, Newberry KM, Brandsmeier LA, Meadows SK, Messer CL, Hardigan AA, Coppola CJ, Dean EC, Jiang S, Savic D, Mortazavi A, Wold BJ, Myers RM, Mendenhall EM (2020) Occupancy maps of 208 chromatin-associated proteins in one human cell type. Nature 583:720–728
Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, Tahiliani M, Daley GQ, Liu XS, Ecker JR, Milos PM, Agarwal S, Rao A (2011) Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature 473:394–397
Peat JR, Dean W, Clark SJ, Krueger F, Smallwood SA, Ficz G, Kim JK, Marioni JC, Hore TA, Reik W (2014) Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation. Cell Rep 9:1990–2000
Pelizzola M, Koga Y, Urban AE, Krauthammer M, Weissman S, Halaban R, Molinaro AM (2008) MEDME: an experimental and analytical methodology for the estimation of DNA methylation levels based on microarray derived MeDIP-enrichment. Genome Res 18:1652–1659
Petterson A, Chung TH, Tan D, Sun X, Jia XY (2014) RRHP: a tag-based approach for 5-hydroxymethylcytosine mapping at single-site resolution. Genome Biol 15:456
Pham TT, Yin J, Eid JS, Adams E, Lam R, Turner SW, Loomis EW, Wang JY, Hagerman PJ, Hanes JW (2016) Single-locus enrichment without amplification for sequencing and direct detection of epigenetic modifications. Mol Gen Genomics 291:1491–1504
Pharo HD, Andresen K, Berg KCG, Lothe RA, Jeanmougin M, Lind GE (2018) A robust internal control for high-precision DNA methylation analyses by droplet digital PCR. Clin Epigenetics 10:24
Pichon F, Shen Y, Busato F, Jochems SP, Jacquelin B, Grand RL, Deleuze JF, Muller-Trutwin M, Tost J (2021) Analysis and annotation of DNA methylation in two nonhuman primate species using the Infinium Human Methylation 450K and EPIC BeadChips. Epigenomics 13:169–186
Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Djik S, Muhlhausler B, Stirzaker C, Clark SJ (2016) Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol 17:208
Pierce SE, Kim SH, Greenleaf WJ (2021) Finding needles in a haystack: dissecting tumor heterogeneity with single-cell transcriptomic and chromatin accessibility profiling. Curr Opin Genet Dev 66:36–40
Pogribny I, Yi P, James SJ (1999) A sensitive new method for rapid detection of abnormal methylation patterns in global DNA and within CpG islands. Biochem Biophys Res Commun 262:624–628
Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 463: 1101–1105
Priolo M, Sparago A, Mammi C, Cerrato F, Lagana C, Riccio A (2008) MS-MLPA is a specific and sensitive technique for detecting all chromosome 11p15.5 imprinting defects of BWS and SRS in a single-tube experiment. Eur J Hum Genet 16:565–571
Qui Y, Yang Q, Sui F, Lu R, Dang S, Ji M, He N, Shi B, Hou P (2015) A strategy for accurate quantification of 5-methylcytosine and 5-hydroxymethylcytosine at CpG sites within gene promoter. J Biomed Nanotechnol 11:1016–1026
Radpour R, Kohler C, Haghighi MM, Fan AX, Holzgreve W, Zhong XY (2009) Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array. Oncogene 28: 2969–2978
Ragoussis J, Elvidge GP, Kaur K, Colella S (2006) Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry in genomics research. PLoS Genet 2: e100
Raine A, Manlig E, Wahlberg P, Syvanen AC, Nordlund J (2017) SPlinted Ligation Adapter Tagging (SPLAT), a novel library preparation method for whole genome bisulphite sequencing. Nucleic Acids Res 45:e36

Rand K, Qu W, Ho T, Clark SJ, Molloy P (2002) Conversion-specific detection of DNA methylation using real-time polymerase chain reaction (ConLight-MSP) to avoid false positives. Methods 27:114–120

Rand KN, Ho T, Qu W, Mitchell SM, White R, Clark SJ, Molloy PL (2005) Headloop suppression PCR and its application to selective amplification of methylated DNA sequences. Nucleic Acids Res 33:e127

Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akeson M, Paten B (2017) Mapping DNA methylation with high-throughput nanopore sequencing. Nat Methods 14:411–413

Rauch T, Pfeifer GP (2005) Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. Lab Invest 85: 1172–1180

Ren X, Kang B, Zhang Z (2018) Understanding tumor ecosystems by single-cell sequencing: promises and limitations. Genome Biol 19:211

Riebler A, Menigatti M, Song JZ, Statham AL, Stirzaker C, Mahmud N, Mein CA, Clark SJ, Robinson MD (2014) BayMeth: improved DNA methylation quantification for affinity capture sequencing data using a flexible Bayesian approach. Genome Biol 15: R35

Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M (2015) Integrative analysis of 111 reference human epigenomes. Nature 518:317–330

Robinson MD, Stirzaker C, Statham AL, Coolen MW, Song JZ, Nair SS, Strbenac D, Speed TP, Clark SJ (2010) Evaluation of affinity-based genome-wide DNA methylation data: effects of CpG density, amplification bias, and copy number variation. Genome Res 20:1719–1729

Roessler J, Ammerpohl O, Gutwein J, Hasemeier B, Anwar SL, Kreipe H, Lehmann U (2012) Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc. BMC Res Notes 5:210

Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. Genome Res 11:3–11

Ronaghi M, Uhlen M, Nyren P (1998) A sequencing method based on real-time pyrophosphate. Science 281:363–365

Ross P, Hall L, Haff LA (2000) Quantitative approach to single-nucleotide polymorphism analysis using MALDI-TOF mass spectrometry. Biotechniques 29: 620–629

Rothbart SB, Strahl BD (2014) Interpreting the language of histone and DNA modifications. Biochim Biophys Acta 1839:627–643

Rougier N, Bourc'his D, Gomes DM, Niveleau A, Plachot M, Paldi A, Viegas-Pequignot E (1998) Chromosome methylation patterns during mammalian preimplantation development. Genes Dev 12:2108–2113

Rousseaux S, Seyve E, Chuffart F, Bourova-Flin E, Benmerad M, Charles MA, Forhan A, Heude B, Siroux V, Slama R, Tost J, Vaiman D, Khochbin S, Lepeule J, Group EM-CCS (2020) Immediate and durable effects of maternal tobacco consumption alter placental DNA methylation in enhancer and imprinted gene-containing regions. BMC Med 18:306

Ruzov A, Tsenkina Y, Serio A, Dudnakova T, Fletcher J, Bai Y, Chebotareva T, Pells S, Hannoun Z, Sullivan G, Chandran S, Hay DC, Bradley M, Wilmut I, De Sousa P (2011) Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. Cell Res 21:1332–1342

Rydbirk R, Folke J, Busato F, Roche E, Chauhan AS, Lokkegaard A, Hejl AM, Bode M, Blaabjerg M, Moller M, Danielsen EH, Brudek T, Pakkenberg B, Tost J, Aznar S (2020) Epigenetic modulation of AREL1 and increased HLA expression in brains of multiple system atrophy patients. Acta Neuropathol Commun 8:29

Salvaing J, Li Y, Beaujean N, O'Neill C (2015) Determinants of valid measurements of global changes in 5′-methylcytosine and 5′-hydroxymethylcytosine by immunolocalisation in the early embryo. Reprod Fertil Dev 27:755–764

Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M (2011) Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics 6:692–702

Schmid M, Durussel T, Laemmli UK (2004) ChIC and ChEC; genomic mapping of chromatin proteins. Mol Cell 16:147–157

Schmidl C, Rendeiro AF, Sheffield NC, Bock C (2015) ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat Methods 12:963–965

Schreiber J, Wescoe ZL, Abu-Shumays R, Vivian JT, Baatar B, Karplus K, Akeson M (2013) Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along

individual DNA strands. Proc Natl Acad Sci USA 110: 18910–18915
Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, Rajagopal N, Nery JR, Urich MA, Chen H, Lin S, Lin Y, Jung I, Schmitt AD, Selvaraj S, Ren B, Sejnowski TJ, Wang W, Ecker JR (2015) Human body epigenome maps reveal noncanonical DNA methylation variation. Nature 523:212–216
Schutsky EK, DeNizio JE, Hu P, Liu MY, Nabel CS, Fabyanic EB, Hwang Y, Bushman FD, Wu H, Kohli RM (2018) Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. Nat Biotechnol 36:1083–1090
Scott RH, Douglas J, Baskcomb L, Nygren AO, Birch JM, Cole TR, Cormier-Daire V, Eastwood DM, Garcia-Minaur S, Lupunzina P, Tatton-Brown K, Bliek J, Maher ER, Rahman N (2008) Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) robustly detects and distinguishes 11p15 abnormalities associated with overgrowth and growth retardation. J Med Genet 45:106–113
Selker EU, Tountas NA, Cross SH, Margolin BS, Murphy JG, Bird AP, Freitag M (2003) The methylated component of the Neurospora crassa genome. Nature 422: 893–897
Sengenes J, Daunay A, Charles MA, Tost J (2010) Quality control and single nucleotide resolution analysis of methylated DNA immunoprecipitation products. Anal Biochem 407:141–143
Serandour AA, Avner S, Mahe EA, Madigou T, Guibert S, Weber M, Salbert G (2016) Single-CpG resolution mapping of 5-hydroxymethylcytosine by chemical labeling and exonuclease digestion identifies evolutionarily unconserved CpGs as TET targets. Genome Biol 17:56
Serizawa RR, Ralfkiaer U, Dahl C, Lam GW, Hansen AB, Steven K, Horn T, Guldberg P (2010) Custom-designed MLPA using multiple short synthetic probes: application to methylation analysis of five promoter CpG islands in tumor and urine specimens from patients with bladder cancer. J Mol Diagn 12:402–408
Serre D, Lee BH, Ting AH (2010) MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res 38:391–399
Shapiro R, DiFate V, Welcher M (1974) Deamination of cytosine derivatives by bisulfite. Mechanism of the reaction. J Am Chem Soc 96:906–912
Shema E, Jones D, Shoresh N, Donohue L, Ram O, Bernstein BE (2016) Single-molecule decoding of combinatorially modified nucleosomes. Science 352: 717–721
Shemer R, Magenheim J, Dor Y (2019) Digital droplet PCR for monitoring tissue-specific cell death using DNA methylation patterns of circulating cell-free DNA. Curr Protoc Mol Biol 127:e90
Shen L, Guo Y, Chen X, Ahmed S, Issa JP (2007) Optimizing annealing temperature overcomes bias in bisulfite PCR methylation analysis. Biotechniques 42:48–52
Shen SY, Singhania R, Fehringer G, Chakravarthy A, Roehrl MHA, Chadwick D, Zuzarte PC, Borgida A, Wang TT, Li T, Kis O, Zhao Z, Spreafico A, Medina TDS, Wang Y, Roulois D, Ettayebi I, Chen Z, Chow S, Murphy T, Arruda A, O'Kane GM, Liu J, Mansour M, McPherson JD, O'Brien C, Leighl N, Bedard PL, Fleshner N, Liu G, Minden MD, Gallinger S, Goldenberg A, Pugh TJ, Hoffman MM, Bratman SV, Hung RJ, De Carvalho DD (2018) Sensitive tumour detection and classification using plasma cell-free DNA methylomes. Nature 563:579–583
Shim J, Humphreys GI, Venkatesan BM, Munz JM, Zou X, Sathe C, Schulten K, Kosari F, Nardulli AM, Vasmatzis G, Bashir R (2013) Detection and quantification of methylation in DNA using solid-state nanopores. Sci Rep 3:1389
Shim J, Kim Y, Humphreys GI, Nardulli AM, Kosari F, Vasmatzis G, Taylor WR, Ahlquist DA, Myong S, Bashir R (2015) Nanopore-based assay for detection of methylation in double-stranded DNA fragments. ACS Nano 9:290–300
Shirane K, Toh H, Kobayashi H, Miura F, Chiba H, Ito T, Kono T, Sasaki H (2013) Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. PLoS Genet 9:e1003439
Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W (2017) Detecting DNA cytosine methylation using nanopore sequencing. Nat Methods 14:407–410
Singer-Sam J, LeBon JM, Tanguay RL, Riggs AD (1990) A quantitative HpaII-PCR assay to measure methylation of DNA from a small number of cells. Nucleic Acids Res 18:687
Skene PJ, Henikoff JG, Henikoff S (2018) Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nat Protoc 13:1006–1019
Skvortsova K, Zotenko E, Luu PL, Gould CM, Nair SS, Clark SJ, Stirzaker C (2017) Comprehensive evaluation of genome-wide 5-hydroxymethylcytosine profiling approaches in human DNA. Epigenetics Chromatin 10:16
Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, Andrews SR, Stegle O, Reik W, Kelsey G (2014) Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods 11:817–820
Smith ZD, Gu H, Bock C, Gnirke A, Meissner A (2009) High-throughput bisulfite sequencing in mammalian genomes. Methods 48:226–232
Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang B, Godley LA, Hicks LM, Lahn BT, Jin P, He C (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol 29:68–72

Song CX, Clark TA, Lu XY, Kislyuk A, Dai Q, Turner SW, He C, Korlach J (2012) Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. Nat Methods 9:75–77

Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, Street C, Li Y, Poidevin M, Wu H, Gao J, Liu P, Li L, Xu GL, Jin P, He C (2013) Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell 153:678–691

Sontag S, Bocova L, Hubens WHG, Nuchtern S, Schnitker M, Look T, Schroder KM, Plumakers B, Tharmapalan V, Wessiepe M, Kraus T, Kramer J, Rink L, Koschmieder S, Wagner W (2022) Toward clinical application of leukocyte counts based on targeted DNA methylation analysis. Clin Chem 68:646–656

Srinivasan M, Sedmak D, Jewell S (2002) Effect of fixatives and tissue processing on the content and integrity of nucleic acids. Am J Pathol 161:1961–1971

Statham AL, Robinson MD, Song JZ, Coolen MW, Stirzaker C, Clark SJ (2012) Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. Genome Res 22:1120–1127

Stelzer Y, Shivalila CS, Soldner F, Markoulaki S, Jaenisch R (2015) Tracing dynamic changes of DNA methylation at single-cell resolution. Cell 163:218–229

Stewart FJ, Panne D, Bickle TA, Raleigh EA (2000) Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. J Mol Biol 298:611–622

Stewart SK, Morris TJ, Guilhamon P, Bulstrode H, Bachman M, Balasubramanian S, Beck S (2015) oxBS-450K: a method for analysing hydroxymethylation using 450K BeadChips. Methods 72:9–15

Struhl K, Segal E (2013) Determinants of nucleosome positioning. Nat Struct Mol Biol 20:267–273

Sun Z, Terragni J, Borgaro JG, Liu Y, Yu L, Guan S, Wang H, Sun D, Cheng X, Zhu Z, Pradhan S, Zheng Y (2013) High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. Cell Rep 3:567–576

Sun Z, Dai N, Borgaro JG, Quimby A, Sun D, Correa IR Jr, Zheng Y, Zhu Z, Guan S (2015) A sensitive approach to map genome-wide 5-hydroxymethylcytosine and 5-formylcytosine at single-base resolution. Mol Cell 57:750–761

Sun X, Chung TH, Tan D, Kim A (2016) Practical guidelines and consideration of using RRHP for 5hmC detection. Epigenomics 8:225–235

Sun K, Lun FMF, Leung TY, Chiu RWK, Lo YMD, Sun H (2018) Noninvasive reconstruction of placental methylome from maternal plasma DNA: Potential for prenatal testing and monitoring. Prenat Diagn 38:196–203

Sun Z, Vaisvila R, Hussong LM, Yan B, Baum C, Saleh L, Samaranayake M, Guan S, Dai N, Correa IR Jr, Pradhan S, Davis TB, Evans TC Jr, Ettwiller LM (2021) Nondestructive enzymatic deamination enables single-molecule long-read amplicon sequencing for the determination of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Genome Res 31(2):291–300

Suzuki M, Liao W, Wos F, Johnston AD, DeGrazia J, Ishii J, Bloom T, Zody MC, Germer S, Greally JM (2018) Whole-genome bisulfite sequencing with improved accuracy and cost. Genome Res 28:1364–1371

Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res 38:e181

Taberlay PC, Kelly TK, Liu CC, You JS, De Carvalho DD, Miranda TB, Zhou XJ, Liang G, Jones PA (2011) Polycomb-repressed genes have permissive enhancers that initiate reprogramming. Cell 147:1283–1294

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324:930–935

Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D, Beck S, Butcher LM (2012) Methylome analysis using MeDIP-seq with low DNA concentrations. Nat Protoc 7:617–636

Tanas AS, Borisova ME, Kuznetsova EB, Rudenko VV, Karandasheva KO, Nemtsova MV, Izhevskaya VL, Simonova OA, Larin SS, Zaletaev DV, Strelnikov VV (2017) Rapid and affordable genome-wide bisulfite DNA sequencing by XmaI-reduced representation bisulfite sequencing. Epigenomics 9:833–847

Tangkijvanich P, Hourpai N, Rattanatanyong P, Wisedopas N, Mahachai V, Mutirangura A (2007) Serum LINE-1 hypomethylation as a potential prognostic marker for hepatocellular carcinoma. Clin Chim Acta 379:127–133

Taryma-Lesniak O, Sokolowska KE, Wojdacz TK (2020) Current status of development of methylation biomarkers for in vitro diagnostic IVD applications. Clin Epigenetics 12:100

Thirlwell C, Eymard M, Feber A, Teschendorff A, Pearce K, Lechner M, Widschwendter M, Beck S (2010) Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip. Methods 52:248–254

Thomson JP, Meehan RR (2017) The application of genome-wide 5-hydroxymethylcytosine studies in cancer research. Epigenomics 9:77–91

Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr AR, Deaton A, Andrews R, James KD, Turner DJ, Illingworth R, Bird A (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 464:1082–1086

Thomson JP, Hunter JM, Nestor CE, Dunican DS, Terranova R, Moggs JG, Meehan RR (2013) Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques. Nucleic Acids Res 41:e206

Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, Teschendorff AE (2017) ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics 33:3982–3984

Tobi EW, Goeman JJ, Monajemi R, Gu H, Putter H, Zhang Y, Slieker RC, Stok AP, Thijssen PE, Muller F, van Zwet EW, Bock C, Meissner A, Lumey LH, Eline Slagboom P, Heijmans BT (2014) DNA methylation signatures link prenatal famine exposure to growth and metabolism. Nat Commun 5: 5592

Tost J (2008a) Epigenetics. Horizon Scientific Press, Norwich, UK

Tost J (2008b) Methods for the genome-wide and gene-specific analysis of DNA methylation levels and patterns. In: Tost J (ed) Epigenetics. Horizon Scientific Press, Norwich, pp 63–104

Tost J, Gut IG (2007) DNA methylation analysis by pyrosequencing. Nat Protoc 2:2265–2275

Tost J, Dunker J, Gut IG (2003) Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. Biotechniques 35:152–156

Tost J, El Abdalaoui H, Gut IG (2006) Serial pyrosequencing for quantitative DNA methylation analysis. Biotechniques 40:721–726

Touleimat N, Tost J (2012) Complete pipeline for Infinium ((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics 4:325–341

Tsui NB, Chiu RW, Ding C, El-Sheikhah A, Leung TN, Lau TK, Nicolaides KH, Lo YM (2005) Detection of trisomy 21 by quantitative mass spectrometric analysis of single-nucleotide polymorphisms. Clin Chem 51: 2358–2362

Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR (2015) MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. Nat Protoc 10:475–483

Vaca-Paniagua F, Oliver J, Nogueira da Costa A, Merle P, McKay J, Herceg Z, Holmila R (2015) Targeted deep DNA methylation analysis of circulating cell-free DNA in plasma using massively parallel semiconductor sequencing. Epigenomics 7:353–362

Vaisvila R, Ponnaluri VKC, Sun Z, Langhorst BW, Saleh L, Guan S, Dai N, Campbell MA, Sexton BS, Marks K, Samaranayake M, Samuelson JC, Church HE, Tamanaha E, Correa IR Jr, Pradhan S, Dimalanta ET, Evans TC Jr, Williams L, Davis TB (2021) Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. Genome Res 31:1280–1289

Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, Weigman VJ, Marynen P, Zetterberg A, Naume B, Perou CM, Borresen-Dale AL, Kristensen VN (2010) Allele-specific copy number analysis of tumors. Proc Natl Acad Sci USA 107: 16910–16915

van Steensel B, Henikoff S (2000) Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nat Biotechnol 18:424–428

Varinli H, Statham AL, Clark SJ, Molloy PL, Ross JP (2015) COBRA-Seq: sensitive and quantitative methylome profiling. Genes (Basel) 6:1140–1163

Varley KE, Mitra RD (2010) Bisulfite Patch PCR enables multiplexed sequencing of promoter methylation across cancer samples. Genome Res 20:1279–1287

Vedeld HM, Grimsrud MM, Andresen K, Pharo HD, von Seth E, Karlsen TH, Honne H, Paulsen V, Farkkila MA, Bergquist A, Jeanmougin M, Aabakken L, Boberg KM, Folseraas T, Lind GE (2022) Early and accurate detection of cholangiocarcinoma in patients with primary sclerosing cholangitis by methylation markers in bile. Hepatology 75:59–73

Verhoef L, Floore AN, Doorn S, Cuschieri K, Bhatia R, Hesselink AT, Meijer C, Steenbergen RDM, Heideman DAM (2021) Direct bisulphite conversion of cervical samples for DNA methylation analysis. Epigenetics in press

Vinarskaja A, Goering W, Ingenwerth M, Schulz WA (2012) ID4 is frequently downregulated and partially hypermethylated in prostate cancer. World J Urol 30: 319–325

Voss KH, Ross KP, Nonay NJ, Dovichi NJ (1998) Combating PCR bias in bisulfite-based cytosine methylation analysis. Betaine-modified cytosine deamination PCt. Anal Chem 70:3818–3823

Vrielink A, Ruger W, Driessen HP, Freemont PS (1994) Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. EMBO J 13: 3413–3422

Wagner M, Steinbacher J, Kraus TF, Michalakis S, Hackner B, Pfaffeneder T, Perera A, Muller M, Giese A, Kretzschmar HA, Carell T (2015) Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. Angew Chem Int Ed Engl 54:12511–12514

Walker DL, Bhagwate AV, Baheti S, Smalley RL, Hilker CA, Sun Z, Cunningham JM (2015) DNA methylation profiling: comparison of genome-wide sequencing methods and the Infinium Human Methylation 450 Bead Chip. Epigenomics 7:1287–1302

Wallace EV, Stoddart D, Heron AJ, Mikhailova E, Maglia G, Donohoe TJ, Bayley H (2010) Identification of epigenetic DNA modifications with a protein nanopore. Chem Commun (Camb) 46:8195–8197

Wang H, Guan S, Quimby A, Cohen-Karni D, Pradhan S, Wilson G, Roberts RJ, Zhu Z, Zheng Y (2011) Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. Nucleic Acids Res 39:9294–9305

Wang Q, Gu L, Adey A, Radlwimmer B, Wang W, Hovestadt V, Bahr M, Wolf S, Shendure J, Eils R, Plass C, Weichenhan D (2013) Tagmentation-based

whole-genome bisulfite sequencing. Nat Protoc 8: 2022–2032

Wang S, Lv J, Zhang L, Dou J, Sun Y, Li X, Fu X, Dou H, Mao J, Hu X, Bao Z (2015a) MethylRAD: a simple and scalable method for genome-wide DNA methylation profiling using methylation-dependent restriction enzymes. Open Biol 5:150130

Wang K, Li X, Dong S, Liang J, Mao F, Zeng C, Wu H, Wu J, Cai W, Sun ZS (2015b) Q-RRBS: a quantitative reduced representation bisulfite sequencing method for single-cell methylome analyses. Epigenetics 10:775–783

Wang Q, Xiong H, Ai S, Yu X, Liu Y, Zhang J, He A (2019) CoBATCH for high-throughput single-cell epigenomic profiling. Mol Cell 76:206–216 e207

Wanunu M, Cohen-Karni D, Johnson RR, Fields L, Benner J, Peterman N, Zheng Y, Klein ML, Drndic M (2011) Discrimination of methylcytosine from hydroxymethylcytosine in DNA molecules. J Am Chem Soc 133:486–492

Warnecke PM, Stirzaker C, Melki JR, Millar DS, Paul CL, Clark SJ (1997) Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. Nucleic Acids Res 25:4422–4426

Warnecke PM, Stirzaker C, Song J, Grunau C, Melki JR, Clark SJ (2002) Identification and resolution of artifacts in bisulfite sequencing. Methods 27:101–107

Warren JD, Xiong W, Bunker AM, Vaughn CP, Furtado LV, Roberts WL, Fang JC, Samowitz WS, Heichman KA (2011) Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. BMC Med 9: 133

Warton K, Samimi G (2015) Methylation of cell-free circulating DNA in the diagnosis of cancer. Front Mol Biosci 2:13

Warton K, Lin V, Navin T, Armstrong NJ, Kaplan W, Ying K, Gloss B, Mangs H, Nair SS, Hacker NF, Sutherland RL, Clark SJ, Samimi G (2014) Methylation-capture and Next-Generation Sequencing of free circulating DNA from human plasma. BMC Genomics 15:476

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet 37:853–862

Weichenhan D, Wang Q, Adey A, Wolf S, Shendure J, Eils R, Plass C (2018) Tagmentation-based library preparation for low DNA input whole genome bisulfite sequencing. Methods Mol Biol 1708:105–122

Weisenberger DJ, Trinh BN, Campan M, Sharma S, Long TI, Ananthnarayan S, Liang G, Esteva FJ, Hortobagyi GN, McCormick F, Jones PA, Laird PW (2008) DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. Nucleic Acids Res 36:4689–4698

Wescoe ZL, Schreiber J, Akeson M (2014) Nanopores discriminate among five C5-cytosine variants in DNA. J Am Chem Soc 136:16582–16587

White HE, Hall VJ, Cross NC (2007) Methylation-sensitive high-resolution melting-curve analysis of the SNRPN gene as a diagnostic screen for Prader-Willi and Angelman syndromes. Clin Chem 53:1960–1962

Wiebers JL (1976) Detection and identification of minor nucleotides in intact deoxyribonucleic acids by mass spectrometry. Nucleic Acids Res 3:2959–2970

Wielscher M, Vierlinger K, Kegler U, Ziesche R, Gsur A, Weinhausel A (2015) Diagnostic performance of plasma DNA methylation profiles in lung cancer, pulmonary fibrosis and COPD. EBioMedicine 2:927–934

Wiencke JK, Bracci PM, Hsuang G, Zheng S, Hansen H, Wrensch MR, Rice T, Eliot M, Kelsey KT (2014) A comparison of DNA methylation specific droplet digital PCR (ddPCR) and real time qPCR with flow cytometry in characterizing human T cells in peripheral blood. Epigenetics 9:1360–1365

Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, Helin K (2011) TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature 473:343–348

Wojdacz TK, Dobrovic A (2007) Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. Nucleic Acids Res 35:e41

Wojdacz TK, Hansen LL (2006) Reversal of PCR bias for improved sensitivity of the DNA methylation melting curve assay. Biotechniques 41:274, 276, 278

Wojdacz TK, Dobrovic A, Hansen LL (2008a) Methylation-sensitive high-resolution melting. Nat Protoc 3:1903–1908

Wojdacz TK, Dobrovic A, Algar EM (2008b) Rapid detection of methylation change at H19 in human imprinting disorders using methylation-sensitive high-resolution melting. Hum Mutat 29:1255–1260

Wojdacz TK, Borgbo T, Hansen LL (2009) Primer design versus PCR bias in methylation independent PCR amplifications. Epigenetics 4:231–234

Worm J, Aggerholm A, Guldberg P (2001) In-tube DNA methylation profiling by fluorescence melting curve analysis. Clin Chem 47:1183–1189

Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y (2011) Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev 25:679–684

Wu X, Li G, Xie R (2018) Decoding the role of TET family dioxygenases in lineage specification. Epigenetics Chromatin 11:58

Xiao Y, Yu F, Pang L, Zhao H, Liu L, Zhang G, Liu T, Zhang H, Fan H, Zhang Y, Pang B, Li X (2015) MeSiC: a model-based method for estimating 5 mC levels at single-CpG RESOLUTION from MeDIP-seq. Sci Rep 5:14699

Xiong Z, Laird PW (1997) COBRA: a sensitive and quantitative DNA methylation assay. Nucleic Acids Res 25: 2532–2534

Xiong Z, Yang F, Li M, Ma Y, Zhao W, Wang G, Li Z, Zheng X, Zou D, Zong W, Kang H, Jia Y, Li R,

Zhang Z, Bao Y (2022) EWAS Open Platform: integrated data, knowledge and toolkit for epigenome-wide association study. Nucleic Acids Res 50:D1004–D1009

Xu M, Kladde MP, Van Etten JL, Simpson RT (1998) Cloning, characterization and expression of the gene coding for a cytosine-5-DNA methyltransferase recognizing GpC. Nucleic Acids Res 26:3961–3966

Xu Z, Taylor JA, Leung YK, Ho SM, Niu L (2016) oxBS-MLE: an efficient method to estimate 5-methylcytosine and 5-hydroxymethylcytosine in paired bisulfite and oxidative bisulfite treated DNA. Bioinformatics 32: 3667–3669

Xu Z, Niu L, Taylor JA (2021a) The ENmix DNA methylation analysis pipeline for Illumina BeadChip and comparisons with seven other preprocessing pipelines. Clin Epigenetics 13:216

Xu L, Zhou Y, Chen L, Bissessur AS, Chen J, Mao M, Ju S, Chen L, Chen C, Li Z, Zhang X, Chen F, Cao F, Wang L, Wang Q (2021b) Deoxyribonucleic acid 5-hydroxymethylation in cell-free deoxyribonucleic acid, a novel cancer biomarker in the era of precision medicine. Front Cell Dev Biol 9:744990

Yamada Y, Watanabe H, Miura F, Soejima H, Uchiyama M, Iwasaka T, Mukai T, Sakaki Y, Ito T (2004) A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. Genome Res 14:247–266

Yamazaki J, Toyomaki H, Nakayama SMM, Yabe J, Muzandu K, Jelinek J, Yokoyama S, Ikenaka Y, Takiguchi M, Ishizuka M (2021) Genome-wide DNA methylation analysis of dogs with high lead exposure living near a lead mining area in Kabwe, Zambia. Environ Pollut 286:117229

Yan R, Gu C, You D, Huang Z, Qian J, Yang Q, Cheng X, Zhang L, Wang H, Wang P, Guo F (2021) Decoding dynamic epigenetic landscapes in human oocytes using single-cell multi-omics sequencing. Cell Stem Cell 28(1641-1656):e1647

Yang Y, Scott SA (2017) DNA methylation profiling using long-read single molecule real-time bisulfite sequencing (SMRT-BS). Methods Mol Biol 1654: 125–134

Yang AS, Estecio MR, Doshi K, Kondo Y, Tajara EH, Issa JP (2004) A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. Nucleic Acids Res 32:e38

Yang X, Dai W, Kwong DL, Szeto CY, Wong EH, Ng WT, Lee AW, Ngan RK, Yau CC, Tung SY, Lung ML (2015a) Epigenetic markers for noninvasive early detection of nasopharyngeal carcinoma by methylation-sensitive high resolution melting. Int J Cancer 136:E127–E135

Yang Y, Sebra R, Pullman BS, Qiao W, Peter I, Desnick RJ, Geyer CR, DeCoteau JF, Scott SA (2015b) Quantitative and multiplexed DNA methylation analysis using long-read single-molecule real-time bisulfite sequencing (SMRT-BS). BMC Genomics 16:350

You JS, Kelly TK, De Carvalho DD, Taberlay PC, Liang G, Jones PA (2011) OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. Proc Natl Acad Sci USA 108:14497–14502

Yu M, Hon GC, Szulwach KE, Song CX, Jin P, Ren B, He C (2012) Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. Nat Protoc 7:2159–2170

Yu M, Carter KT, Makar KW, Vickers K, Ulrich CM, Schoen RE, Brenner D, Markowitz SD, Grady WM (2015) MethyLight droplet digital PCR for detection and absolute quantification of infrequently methylated alleles. Epigenetics 10:803–809

Yu M, Heinzerling TJ, Grady WM (2018a) DNA methylation analysis using droplet digital PCR. Methods Mol Biol 1768:363–383

Yu M, Han D, Hon GC, He C (2018b) Tet-assisted bisulfite sequencing (TAB-seq). Methods Mol Biol 1708: 645–663

Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, Strauch K, Waldenberger M, Illig T (2013) Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One 8:e63812

Zemmour H, Planer D, Magenheim J, Moss J, Neiman D, Gilon D, Korach A, Glaser B, Shemer R, Landesberg G, Dor Y (2018) Non-invasive detection of human cardiomyocyte death using methylation patterns of circulating DNA. Nat Commun 9:1443

Zeschnigk M, Bohringer S, Price EA, Onadim Z, Masshofer L, Lohmann DR (2004) A novel real-time PCR assay for quantitative analysis of methylated alleles (QAMA): analysis of the retinoblastoma locus. Nucleic Acids Res 32:e125

Zhang Y, Jeltsch A (2010) The application of next generation sequencing in DNA methylation analysis. Genes (Basel) 1:85–101

Zhang Y, Rohde C, Tierling S, Jurkowski TP, Bock C, Santacruz D, Ragozin S, Reinhardt R, Groth M, Walter J, Jeltsch A (2009) DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. PLoS Genet 5:e1000438

Zhang L, Szulwach KE, Hon GC, Song CX, Park B, Yu M, Lu X, Dai Q, Wang X, Street CR, Tan H, Min JH, Ren B, Jin P, He C (2013) Tet-mediated covalent labelling of 5-methylcytosine for its genome-wide detection and sequencing. Nat Commun 4:1517

Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H (2014) F2RL3 methylation as a biomarker of current and lifetime smoking exposures. Environ Health Perspect 122:131–137

Zhou W, Triche TJ Jr, Laird PW, Shen H (2018) SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. Nucleic Acids Res 46:e123

Zhou L, Ng HK, Drautz-Moses DI, Schuster SC, Beck S, Kim C, Chambers JC, Loh M (2019) Systematic evaluation of library preparation methods and sequencing

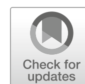
platforms for high-throughput whole genome bisulfite sequencing. Sci Rep 9:10383

Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, Gnirke A, Meissner A (2013) Charting a dynamic DNA methylation landscape of the human genome. Nature 500:477–481

Ziller MJ, Hansen KD, Meissner A, Aryee MJ (2015) Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. Nat Methods 12:230–232

Zinn RL, Pruitt K, Eguchi S, Baylin SB, Herman JG (2007) hTERT is expressed in cancer cell lines despite promoter DNA methylation by preservation of unmethylated DNA and active chromatin around the transcription start site. Cancer Res 67:194–201

Zong W, Kang H, Xiong Z, Ma Y, Jin T, Gong Z, Yi L, Zhang M, Wu S, Wang G, Bao Y, Li R (2022) scMethBank: a database for single-cell whole genome DNA methylation maps. Nucleic Acids Res 50:D380–D386



17 Inhibitors of DNA Methylation

Marie Lopez, Julie Gilbert, Jean Contreras, Ludovic Halby, and Paola B. Arimondo

Abstract

DNA methylation is involved in numerous biological processes and is deregulated in human diseases. The modulation of the activity of the enzymes and proteins in charge of DNA methylation, for example, DNA methyltransferases (DNMTs), can represent a powerful strategy to alter DNA methylation patterns and restore biological processes that are aberrant in diseases. In this chapter, we present examples of inhibitors of DNMTs (DNMTi). We review their fields of application either as therapeutic molecules, for example, in cancers, cardiovascular, neurological, and infectious diseases or as bioengineering tools. Finally, novel strategies to target DNA methylation and overcome the limits of single DNMT inhibitors will be described. These strategies consist in either targeting the methyl group reader proteins rather than targeting directly DNMTs or to combine within the same molecule a DNMT inhibitor with an additional active moiety, e.g., HDAC inhibitor, to improve efficacy and lower secondary effect of such drug.

Keywords

DNA methyltransferases · A · DNMT inhibitors · Methylbinding proteins · MBD · therapeutic strategies · chemical biology

M. Lopez · J. Gilbert
Institut des Biomolécules Max Mousseron (IBMM), CNRS, Université de Montpellier, ENSCM UMR 5247, Montpellier, France
e-mail: marie.lopez@cnrs.fr; julie.gilbert@enscm.fr

J. Contreras
Epigenetic Chemical Biology, Department of Structural Biology and Chemistry, Institut Pasteur, Université de Paris, CNRS UMR3523 Chem4Life, Paris, France

Ecole Doctorale FIRE ED 474, Université de Paris, PSL, Paris, France
e-mail: jean.contreras@pasteur.fr

L. Halby · P. B. Arimondo (✉)
Epigenetic Chemical Biology, Department of Structural Biology and Chemistry, Institut Pasteur, Université de Paris, CNRS UMR3523 Chem4Life, Paris, France
e-mail: ludovic.halby@pasteur.fr; paola.arimondo@cnrs.fr

Abbreviations

5aza	5-azacytidine
5azadC	5-aza-2′-deoxycytidine
AD	Alzheimer's disease
ALI	Acute lung injury
ALL	Acute lymphoblastic leukemia
ALS	Amyotrophic lateral sclerosis
AML	Acute myeloid leukemia
Ara-C	Cytarabine
ASMA	Alpha-smooth muscle actin
ATRA	All-*trans* retinoic acid
BBB	Blood-brain barrier
BDNF	Brain-derived neurotrophic factor

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_17

BM	Bone marrow
CMML	Chronic myelomonocytic leukemia
CNS	Central nervous system
COMT	Catechol-*O*-methyltransferase
CpG	Cytosine-guanine dinucleotide
CVD	Cardiovascular disease
DNMT	DNA methyltransferase
DNMTi	DNMT inhibitor
EC	Endothelial cells
EGCG	(–)-epigallocatechin-3-gallate
EMA	European Medicines Agency
ER	Estrogen receptor
ESC	Embryonic stem cells
FDA	U. S. Food and Drug Administration
GABA	γ-aminobutyric acid
GAD	Glutamic acid decarboxylase
HCV	Hepatitis C virus
HDAC	Histone deacetylase
HDACi	HDAC inhibitor
HIV	Human immunodeficiency virus
HPV	Human papilloma virus
HTS	High-throughput screening
INFγ	Interferon-γ
IVF	In vitro fertilization
LPS	Lipopolysaccharide
LTP	Long-term potentiation
MDS	Myelodysplasic syndrome
MeCP2	Methyl-CpG-binding protein 2
mPFC	Medial prefrontal cortex
MSC	Mesenchymal stromal cells
NHL	Non-Hodgkin's lymphomas
NMS	Neonatal maternal separation
NSC	Neuronal stem cell
PARP	Poly-(ADP-ribose) polymerase
PARPi	PARP inhibitor
PD	Parkinson's disease
PTSD	Post-traumatic stress disorder
SAH	*S*-adenosyl-L-homocysteine
SAHA	Suberoylanilide hydroxamic acid, vorinostat (Zolinza®)
SAM	*S*-adenosyl-L-methionine
SAR	Seasonal allergic rhinitis
SFRP4	Secreted frizzled-related protein
SHS	Second-hand smoke
Th	T-helper
TSA	Trichostatin A
TSG	Tumor suppressor gene
VPA	Valproic acid
XRCC1	X-ray repair cross-complementing protein

17.1 How to Inhibit DNA Methyltransferases

The DNA methylation mechanism involves three main actors: DNA, DNA methyltransferase (DNMT) and the co-factor *S*-adenosyl-L-methionine (SAM). DNMTs scan the DNA, recognize CpG sites and flip out 2′-deoxycytidine into the catalytic pocket of the enzyme. A cysteine residue (C1226 in human DNMT1, C711 in human DNMT3A and C652 in human DNMT3B) binds to position 6 of the cytosine (Fig. 17.1 top), then transfer of the methyl group from the SAM occurs on the C5 position of the cytosine (Fig. 17.1, gray arrow), the cysteine is released by β-elimination and the enzyme is ready to start a new catalytic cycle (Fig. 17.1 top).

DNMTs can be inhibited by targeting different parts of the catalytic pocket (the deoxycytidine binding site, the SAM binding site or both), by targeting allosteric sites or by interfering with their binding to DNA. When the DNA is the target, for example, by DNA binders, it is important to design specific compounds that do not inhibit other enzymes acting on DNA. A possibility is to use DNA ligands specific for CpGs sites. An interesting alternative are compounds that recognize the cytidine binding site in the DNMT catalytic pocket and, in addition, bear a chemical moiety able to react with the catalytic cysteine. Cytidine analogs, leading to a suicidal covalent complex targeting, are the best example and are described in Sect. 17.2. Another strategy is to target the co-factor binding pocket by SAM analogs, for example. Since SAM is the most widely used enzyme co-factor after ATP (Struck et al. 2012), a challenge lies in the design of compounds specific for the DNMT SAM-binding site and not binding to other methyltransferases. The design of SAM analogs that also bind to the 2′-deoxycytidine binding site, such as bisubstrate inhibitors (Bon et al. 2020),

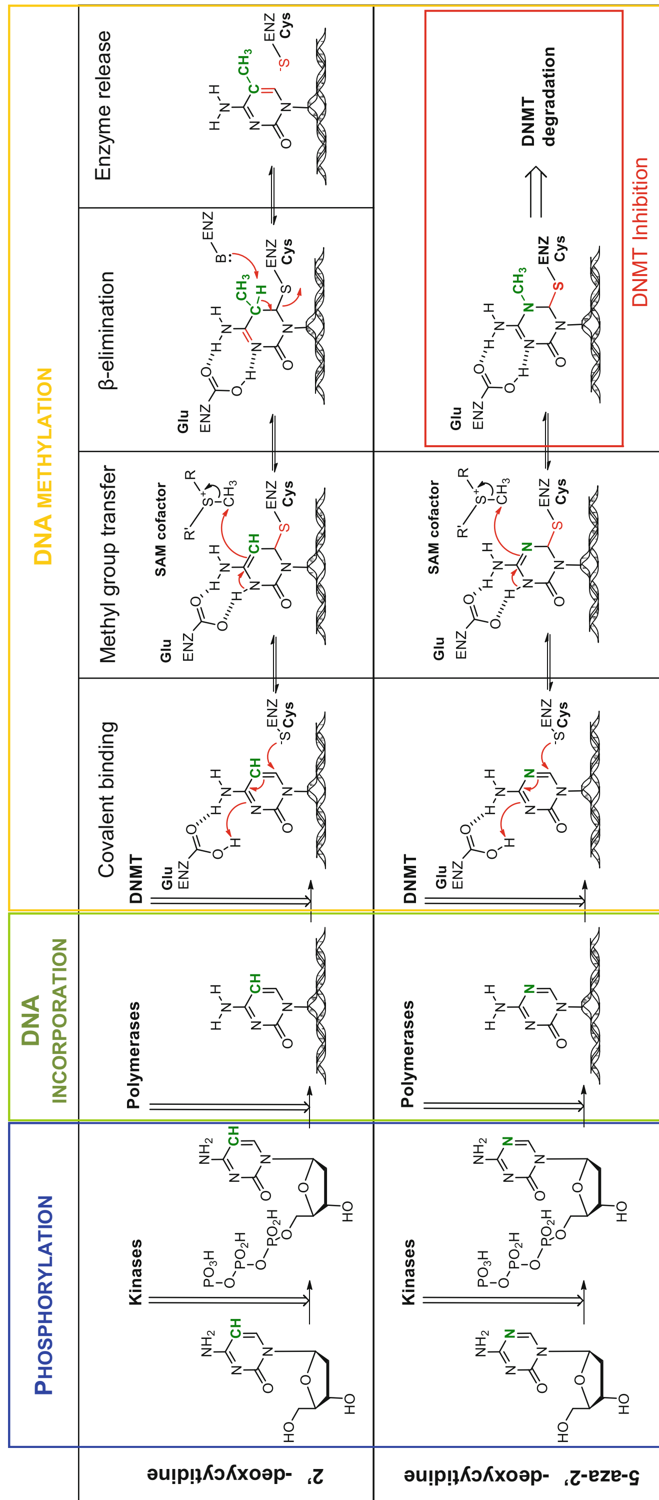


Fig. 17.1 Schematic representation of DNMT catalytic cycle mechanism and its inhibition by 5azadC

could confer this specificity and in addition it could potentially give higher affinity compounds and more potent inhibitors. Below, we list the principal inhibitors representing the different possible mechanisms of inhibition of DNMTs.

17.2 Chemistry and Structure of DNMT Inhibitors

DNMT inhibitors (DNMTi) are grouped in two families: the nucleoside inhibitors and the non-nucleoside ones (Fig. 17.2). To date, two compounds, **5-azacytidine** (**5aza**, azacytidine, Vidaza®) and **5-aza-2′-deoxycytidine** (**5azadC**, decitabine, Dacogen™ and in combination Inqovi®) (Fig. 17.2), are approved by the FDA for the treatment of hematological cancers. 5Aza and 5azadC were the first DNMTi and their story began with their synthesis in 1964 (Sorm et al. 1964). They were initially tested as anti-metabolite agents against acute myelogenous leukemia (AML) (Cihak 1974; Sorm et al. 1964), showing anticancer properties (Evans and Hanka 1968). However, it was in the late 1970s–early 1980s that a big turn was taken in their use when Jones and Taylor showed that non-toxic doses of 5aza and 5azadC induced cell differentiation by DNMT inhibition (Jones and Taylor 1980; Taylor and Jones 1979).

A new paradigm was found, as these compounds were shown to be able to reprogram cells, i.e., induce silenced genes expression and thus restore normal cell functions. The *aza*-nucleosides, once phosphorylated in cells and incorporated into DNA, replace the cytosine in the catalytic pocket of the enzyme and act as suicide substrates of DNMTs (cf. Fig. 17.1). In DNA, at the CpG sites, 5azadC is flipped out from the double-helix into the catalytic pocket, the catalytic cysteine binds covalently to the C6 position, but the β-elimination, essential to release the enzyme, cannot occur after the methyl group transfer because of the presence of a nitrogen atom at the C5 position. The enzyme is thus irreversibly trapped on the DNA and further degraded by the proteasome (Santi et al. 1984), inducing DNA demethylation.

Based on these findings, 5aza (Vidaza®) and 5azadC (Dacogen™) were approved in 2004 and 2006, respectively, by the FDA for the treatment of acute myelodysplastic syndrome (MDS), and chronic myelomonocytic leukemia (CMML). However, these drugs are chemically and metabolically unstable, present a low bioavailability (Chan et al. 1979; Notari and Deyoung 1975), and, as they are incorporated at every deoxycytidine, are not selective (Karahoca and Momparler 2013). Nevertheless, the proof of concept for DNMTi as a therapeutic target was established, and since then many efforts have been dedicated to identify novel DNMT inhibitors.

First, the chemical instability of 5aza and 5azadC was addressed. The 5-azacytosine moiety was shown to be sensitive to rapid hydrolysis under mild conditions over two steps; a reversible ring opening to the N-formylguanylribosylurea, followed by an irreversible formation of guanylribosylurea specie (Chan et al. 1979). More stable analogs were designed, like **5,6-dihydro-5-azacytidine**, **2′-deoxy-5,6-dihydro-5-azacytidine**, **2′-deoxy-5-fluorocytidine** and **zebularine** (cf. Fig. 17.2). **5,6-Dihydro-5-azacytidine** showed weak inhibition of DNA methylation and was withdrawn from clinical trials (Yogelzang et al. 1997). Its analog **2′-deoxy-5,6-dihydro-5-azacytidine** was shown not to bind covalently to DNMTs, but simply to occupy the cytidine pocket (Sheikhnejad et al. 1999), resulting in methylation inhibition in several cell lines (CRF-CEM and HL60) with very low cell toxicity (Matoušová et al. 2011). **2′-Deoxy-5-fluorocytidine** was described as more stable in aqueous media and as potent in vitro, but it is currently rather recognized as a pro-drug of a thymidylate synthase inhibitor, because it is metabolized into 5-fluorouridine (Boothman et al. 1989). **Zebularine** is also a stable compound with a weaker inhibition activity and cytotoxicity than azacytosine analogs (Flotho et al. 2009). Its mode of action is different from 5aza and 5azadC, since it forms a very stable, but reversible complex with DNMTs that showed a slow dissociation kinetic (Champion et al. 2010; van Bemmel et al. 2009).

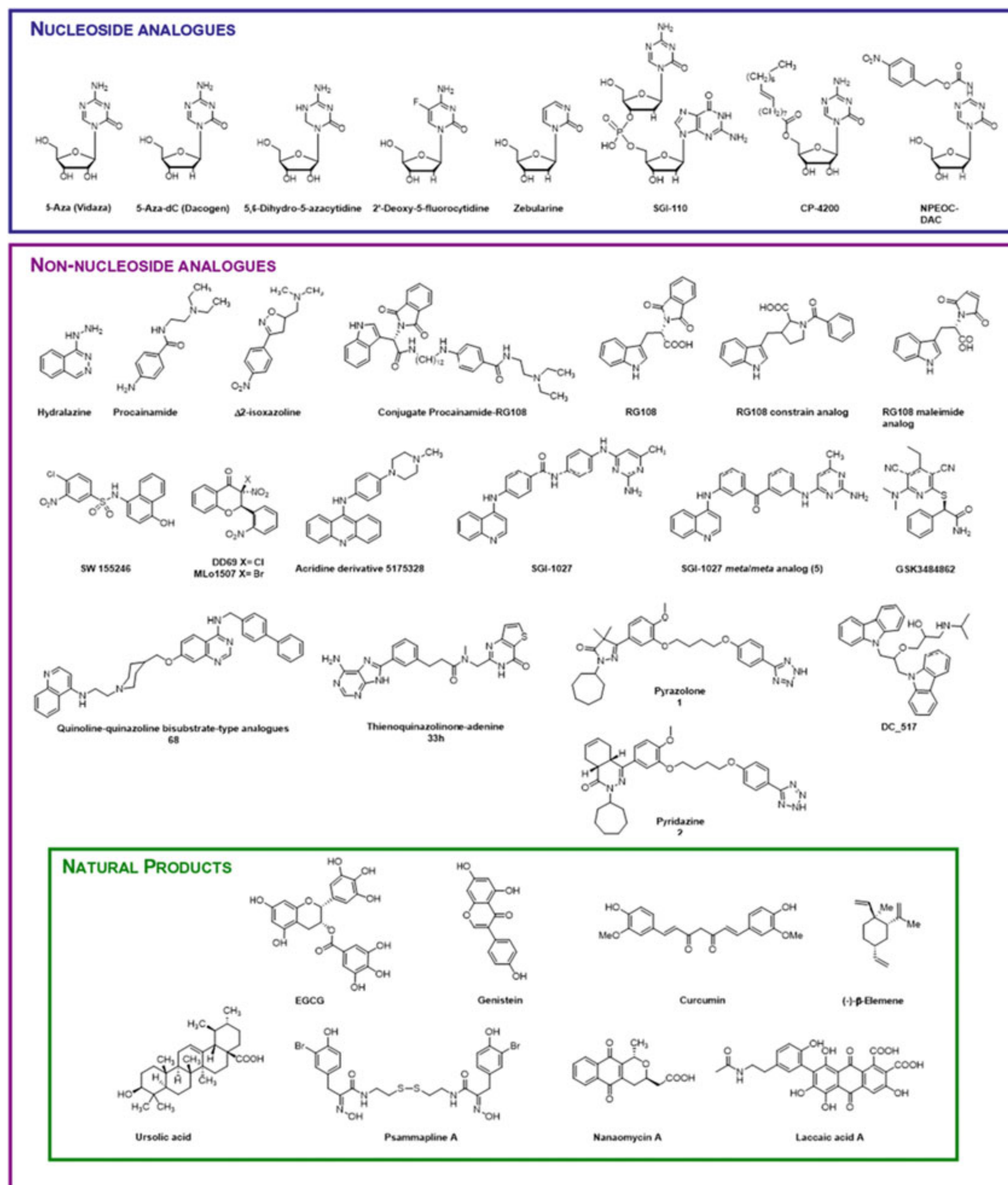


Fig. 17.2 Selection of compounds described as DNMT inhibitors

Second, a pro-drug approach was chosen aiming at releasing the active molecule in the organism. This approach resulted in the generation of **NPEOC-DAC** (Byun et al. 2008), **CP-4200** (Brueckner et al. 2010) and **SGI-110** (Chuang et al. 2010; Yoo et al. 2007). **SGI-110** or **guadecitabine**, composed of a 5azad-CpG dinucleotide, is a pro-drug of 5azadC, it entered clinical trials and is the most promising pro-drug with a comparable in vitro and in vivo DNMT inhibition activity. The advantage of SGI-110 is that it is less sensitive to cytidine deaminase than

5azadC, which improves its metabolic stability.

Third, non-nucleoside compounds were investigated. However, most of them lack specificity, show weak activity against the enzyme or do not induce a strong DNA methylation inhibition in cellular models. Several natural products were reported as DNMTi (Fig. 17.2), for example, **(–)-epigallocatechin-3-gallate (EGCG)** (Fang et al. 2003) and **laccaic acid** (Fagan et al. 2013). These natural products lack specificity and their mode of action as direct DNMT inhibitor was afterwards questioned, as the one of **genistein** (Fang et al. 2005), **psammaplin A** (Baud et al. 2012; Pereira et al. 2012) and **curcumin** (Liu et al. 2009). Indeed, indirect or off-target effects seem to be responsible for the observed inhibition of DNA methylation. All these compounds are known as multitarget compounds (Lopez et al. 2015; Pechalrieu et al. 2017). Similarly, Hann's team highlighted an indirect decrease in DNMT1 expression in cancer cells (lung cancer and hepatocarcinoma) via ERK1/2- and AMPKa-mediated inhibition of the transcription factor Sp1 after treatment with two natural products, **β-elemene** (Zhao et al. 2015) and **ursolic acid** (Yie et al. 2015). **Nanaomycin A** is a more specific compound, because it was found to target specifically DNMT3B, and shown to induce genomic DNA methylation inhibition in colon cancer cell line (Kuck et al. 2010a). Nevertheless, in a recent study, Penter et al. did not observe any DNA methylation inhibition by nanaomycin A in four neuroblastoma cell lines, although they observed an interesting synergistic activity when used in combination with doxorubicin (Penter et al. 2015). To conclude, none of the natural products described to date clearly targets the DNMTs directly in cells and their mechanism of action is controversial (Li and Tollefsbol 2010; Medina-Franco et al. 2011; Suh and Pezzuto 2012).

The same difficulty to identify potent and specific non-nucleoside DNMTi is found among the synthetic compounds. For example, drug repurposing showed that **hydralazine** and **procainamide** are week DNMTi (Candelaria et al. 2012; Chuang et al. 2005). These two FDA-approved drugs are vasodilator and anti-arrhythmic compounds, respectively, that were described as DNMTi because patients developed autoimmune disease (*lupus erythematosus*) correlated to an abnormal DNA methylation pattern of T-cells (Cornacchia et al. 1988). Despite the weak DNMT inhibition activity of hydralazine, a molecular modeling study (Singh et al. 2009a) showed that hydralazine could be involved in a complex network of hydrogen bonds into the cytidine pocket. Today, hydralazine is still extensively studied, especially in combination with valproic acid (VPA), an HDAC inhibitor (Dueñas-Gonzalez et al. 2014). Procainamide was also described to potentially interact within the catalytic side by molecular modeling (Singh et al. 2009a), while biophysical studies previously demonstrated that it acted as DNA ligand and decreased the processivity of DNMTs (Lee et al. 2005). Procainamide and procaine were indeed described as DNA ligands with a certain specificity for CG-rich regions, potentially targeting the DNA binding of the DNMT as described above in Sect. 17.1 (Villar-Garea et al. 2003). Procainamide was an interesting starting point for drug design and several constrained derivatives were synthesized and tested (Castellano et al. 2008, 2011). In parallel, procainamide conjugated to RG108, an in silico identified pharmacophore (see paragraph below), induced a synergy in the inhibition of the enzymatic activity of DNMT3A and DNMT1 (Halby et al. 2012). Initially, it was hypothesized that the procainamide moiety of these compounds would bind to DNA and RG108 moiety would bind to DNMT catalytic pocket. Later, a molecular modeling study suggested that the conjugates could occupy both SAM and cytidine pockets (Yoo et al. 2013), explaining the increased activity (cf. Sect. 17.1).

In parallel, virtual screening-based studies led to the identification of **RG108** (Brueckner et al. 2005; Siedlecki et al. 2006; Stresemann et al. 2006), **DC_517** (Chen et al. 2014), **thienoquinazolinone-adenine 33h** (Newton et al. 2020) and **pyrazolone/pyridazine** derivatives (Huang et al. 2021). Recent experimental studies revealed that RG108 has only a weak activity against human DNMTs (Halby et al. 2012; Suzuki et al. 2010)

and more potent derivatives were since synthesized, such as **maleimide** (Suzuki et al. 2010), **constraint analogs** (Asgatay et al. 2014), and other derivatives (Rotili et al. 2014; Rondelet et al. 2017). Nevertheless, their ability to inhibit genomic DNA methylation has still to be established for all these inhibitors. New generation compounds seem to be more promising, as they showed DNA methylation inhibition and/or gene re-expression in cells. **DC_517** and **Cpd 33h** showed micromolar inhibition activity against DNMT1 (IC_{50} = 1.7 μM) (Chen et al. 2014) and DNMT3B (IC_{50} = 8 μM) (Newton et al. 2020), respectively. DC_517 reduced proliferation and induced apoptosis in HCT116 colorectal cancer cell line (Chen et al. 2014). Regarding pyrazolone/pyridazine derivatives, enzymology studies interestingly suggested that they act as allosteric inhibitors with high-micromolar range activity on purified DNMTs (Huang et al. 2021) but their efficacy in cells needs to be confirmed. Virtual screening followed by structure optimization also led to the discovery of **GSK3484862** (Pappalardi et al. 2021), a highly potent DNMT1-selective inhibitor with IC_{50} = 0.23 μM vs. IC_{50} > 50 μM for DNMT3A/B. In cells, GSK3484862 decreased promoter methylation and reinduced expression of VIM gene, and it also induced global decrease of DNA methylation levels in murine embryonic stem cells with limited toxicity (Azevedo Portilho et al. 2021).

Several High-Throughput Screening (HTS) campaigns were performed and allowed to identify new families of DNMTi. **3-Chloro-3-nitroflavanones** were identified (Ceccaldi et al. 2011) with sub-micromolar activity against Dnmt3a/3L complex and exhibited a phenotype in zebrafish embryos similar to the one observed with the 5aza. This scaffold was further optimized and **3-bromo-3-nitroflavanones** showed improved stability and cellular activity in several cancer models in vitro and ex vivo (Lobo et al. 2021; Marques-Magalhães et al. 2021; Pechalrieu et al. 2020). A naphtoquinone, diclone, and other flavonoids were identified in another screening campaigns, opening the path to study the impact on the epigenome of plants, animals and humans when addressing the toxicology of pesticides (Ceccaldi et al. 2013). Molecular modeling studies suggested that the most potent compounds occupy both the cytosine and SAM pockets (Ceccaldi et al. 2011). **SW155246**, an aromatic sulfonamide, was also identified by HTS (Kilgore et al. 2013). Interestingly, this compound was able to induce a weak methylation inhibition and reactivation of tumor suppressor genes (TSGs) in human lung carcinoma. Acridine derivatives, known to intercalate into DNA, were described to modulate DNA methylation (Hossain et al. 2013). Among them, derivative **517328** was the most potent analog inhibiting DNMT1 in vitro (~30% inhibition at 10 μM) and DNA methylation of TSGs in colon and pancreatic cancer cell lines. This compound demonstrated that DNA binders inhibit DNMTs, however, the specificity of such inhibitors remains an important limitation to their use. **SGI-1027** is another weak DNA ligand that was identified to be a DNMT1 inhibitor (Datta et al. 2009). Molecular modeling studies suggested that SGI-1027 could inhibit DNMTs by occupying both cytidine and SAM pockets (Yoo et al. 2013) and Gros et al. (2015) confirmed by biophysical studies that it interacts with DNA and functions as a SAM non-competitive but DNA-competitive inhibitor. SGI-1027 was described to induce DNMT1 degradation and TSG re-expression in colon cancer cell lines and was quickly considered as the most promising starting point for DNMTi design. Several derivatives (Gamage et al. 2013; Rilova et al. 2014; Valente et al. 2014) succeeded in increasing significantly DNMT inhibition potency with the *meta*/*meta* analog, but no potent inhibition of genomic DNA methylation or TSG re-expression has so far been shown. Further investigations to characterize the mode of action of this compound family indicated that the *meta*/*meta* analog strongly interacted with DNA and inhibited DNMT by DNA interaction and destabilization of the DNMT/DNA/SAM complex (Gros et al. 2015).

Finally, Halby et al. carried out a rational design based on bisubstrate analogs and identified **quinoline-quinazoline** DNMTi with compound **68** as their lead compound. This DNMTi strongly inhibits DNMT3A ($EC_{50} = 1.1$ μM) with a good selectivity over DNMT1 ($EC_{50} = 100$ μM). Besides it potently reduced promoter methylation of *CDKN2A* together with an increase of its expression level (Halby et al. 2017).

To conclude, to date only nucleoside DNMTi were FDA-approved and SGI-110 is a promising pro-drug of 5azadC, evaluated in clinical trials both in hematological and solid cancers. An increasing number of new non-nucleoside DNMTi was published in the last few years, but most lack selectivity or demonstrated impact on genomic DNA methylation and TSG re-expression, except for GSK3484862, an allosteric inhibitor, that is to date the most promising chemical probe for DNMT. Several strategies are therefore implemented to discover more potent DNMTi.

17.3 Potential Applications of DNMT Inhibitors

DNA methylation is crucial for the control of gene expression and cell proliferation. Indeed, DNA methylation of the CpG islands of gene promoters leads to the silencing of the corresponding gene. This regulation participates in the dynamics of the gene expression regulation in cells and is crucial for normal cell functions. However, many factors, including diet, stress, environmental conditions, etc., can lead to abnormal DNA methylation patterns. These deregulations are often responsible for dysfunctions and development of diseases such as cancers, neurological and cardiovascular diseases, and abnormal plant growth (Fig. 17.3). As all epigenetic modifications, DNA methylation is reversible and therefore can be chemically reversed. The main application of DNMTi is by far their use as anticancer agents. However, the importance of DNA methylation in various biological contexts has led to increasing research in the context of other pathologies, such as neurodegenerative and cardiovascular diseases and in other fields, such as plant growth optimization.

17.3.1 DNMTi Application in Cancers

In cancer, hypermethylation of TSG promoter regions is observed together with a global hypomethylation (Esteller 2008). The hypermethylation of promoters of genes, such as *P53*, *P16*, *P15*, *RARβ2*, *HIC1* and *RASSF1A*, results in their silencing and participates in tumor formation, maintenance, and proliferation.

17.3.1.1 Nucleoside Analogs

As Single Agent

As described above, 5aza and 5azadC are the most extensively used DNMTi. These nucleoside analogs are able to decrease TSG-promoter methylation levels in cancer cells and induce TSG re-expression, resulting in cell reprogramming and, eventually, cell cycle arrest and apoptosis (Fahy et al. 2012). A historical overview of their development and application is described in Issa and Kantarjian (2009). As single agents, 5aza or 5azadC are FDA- and EMA-approved to treat certain forms of leukemia. To date there is no approval of 5aza or 5azadC monotherapies for solid tumors and only phase I and II clinical trials are ongoing, for example, for liver, colorectal, breast prostate, pancreatic and non-small cell lung cancers and head neck carcinoma, (https://clinicaltrials.gov/ct2/home). Table 17.1 reports some examples of current clinical trials.

As described above, to overcome 5aza and 5azadC drawbacks (i.e., chemical instability, poor delivery and side effects), other nucleoside analogs were developed (Fig. 17.2), but they need to be used at higher doses due to their lower efficacy and up to date no clinical trials have been undertaken with these compounds.

Despite their promise for non-solid tumor treatment, the direct use of nucleoside analogs as a single therapeutic agent is quite limited considering their instability, their side effects and their lower therapeutic index in solid tumors.

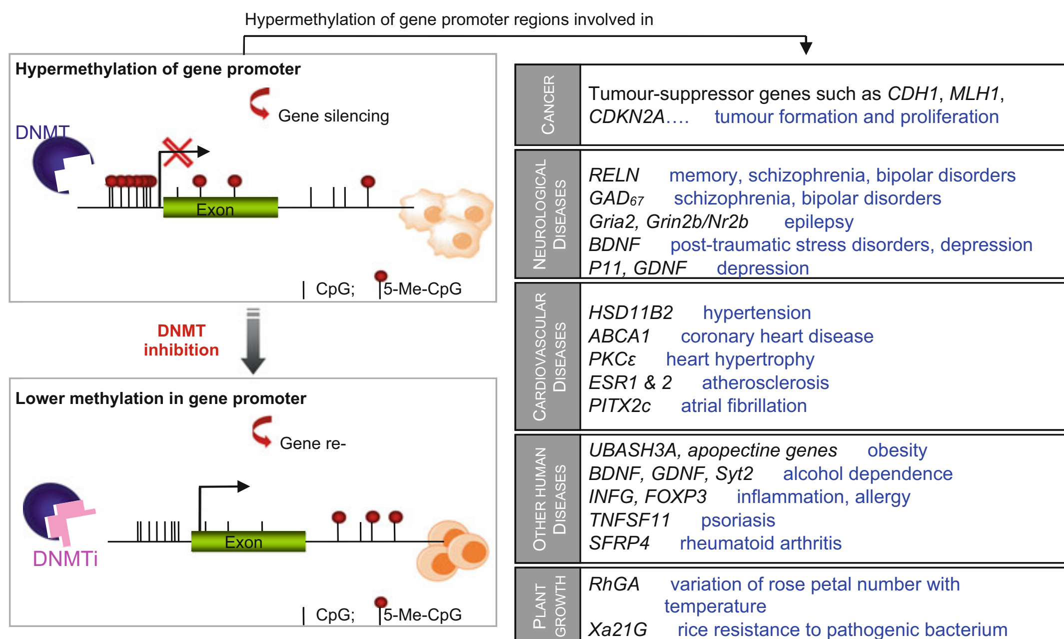


Fig. 17.3 Schematic representation of the mode of action of DNMTi as gene re-expressing agents and examples of DNA methylation-induced gene silencing in diverse pathologies

Pro-drugs of Nucleoside Analogs

Pro-drugs of 5aza and 5azadC, which delay the release of the active molecule in the organism and lower its degradation, showed a real improvement of the drugs. CP-4200 is a lipophilic ester of 5aza patented for its better cellular uptake and activity in vivo in an orthotropic acute lymphoblastic leukemia (ALL) mouse model (Silverman et al. 2009), but no clinical trial was undertaken up to date. SGI-110 (guadecitabine) is a 5azadC pro-drug, which avoids its decomposition by deaminase (Chuang et al. 2010). In the last few years, many clinical trials were started with SGI-110 alone (for example, in phase III clinical trials for AML (Acute Myeloid Leukemia), MDS (Myelodysplastic Syndrome) and CMML (Chronic Myelomonocytic Leukemia), in phase II for gastrointestinal cancer and in phase I for testicular cancer) and in combination with immunotherapies (for example, in phase II for kidney, ovarian, melanoma and (non-) small cell lung cancers).

Nucleoside DNMTi in Combination with Other Drugs

With Other Epidrugs

Nucleoside DNMTi were studied in combinations with other epidrugs, for example, 5azadC was combined with trichostatin A (TSA), an HDACi, in colorectal carcinoma cell line. Whereas TSA alone was not able to cause re-expression of *MLH1*, *TIMP3*, *P15* and *P16* genes, a pre-treatment with a low dose of 5azadC led to their re-expression (Cameron et al. 1999). The combination of 5azadC and valproic acid (VPA), another HDACi, was evaluated in AML and MDS (Yang et al. 2005) and tested in phase II clinical trials (NCT00414310). However, no improvement in outcome was reported. The association of 5aza or 5azadC with the FDA-approved HDACi vorinostat (SAHA) was proven to be of interest in hematological cancers (Silverman et al. 2008), in colorectal cancer cells (Najem et al. 2019) and in vivo in a colon carcinoma CT26

Table 17.1 Examples of DNMTi FDA-approved and in phase II and III clinical trials as a single drug or in combination

	Status	Drugs	Type of cancer targeted (phase of study)
as a single agent	FDA-approved	**5aza (Vidaza®)**	**AML, MDS**
		5azadC (Dacogen™)	**AML, MDS, CMML**
	in clinical trials	Guadecitabine	AML (III), MDS (III), CMML (III), Gastrointestinal Stromal Tumors (II)
		5Aza	T-cell lymphoma (III), DLBCL (III)
		Disulfiram (DSF) + Copper	Metastatic breast (II) and pancreatic (II) cancers, Glioblastoma (III), Non-small Cell Lung Cancer (III)
in combination	with other epi-drug	5AzadC + SAHA	haematological cancers (II)
		5AzadC + VPA	AML (II), MDS (II)
		5Aza + entinostat	Colon and rectal cancer (II), non-small lung cancer (II)
	with chemotherapies	5AzadC + cedazuridine (Inqovi®)	**MDS (FDA-approved)**
		5Aza or 5AzadC + ATRA	AML (II), MDS (II), prostate cancer (II)
		5Aza or 5AzadC + Venetoclax	AML (III), MDS (III), NHL (II)
		5Aza + Glasdegib	AML (III), MDS (III), CMML (III)
		5AzadC + Ivosidenib	AML (III)
	with immunotherapies	5Aza + Magrolimab (anti CD-47)	AML (III), MDS (III)
		5AzadC + Talacotuzumab (anti CD-123)	AML (III), MSD (II)
		5AzadC + Camrelizumab (anti PD-1)	Hodgkin Lymphoma (III)

Phase of clinical trials are mentioned in brackets

experimental lung metastasis mouse model (Yang et al. 2012). In addition, 5aza was reported to be active in combination with entinostat and mocetinostat, benzamide inhibitors of HDACi (Fandy et al. 2009) and these combinations are tested in phase I and II clinical trials in hematological cancers. 5aza/entinostat combination is also studied in solid cancers, such as advanced non-small cell lung cancer (NCT01886573) (Juergens et al. 2011), advanced breast cancer (NCT01349959) and metastatic colorectal cancer (NCT01105377) (see Table 17.1 for some examples). Combination of 5aza/pracinostat, an hydroxamic acid-based HDACi, is in phase III for the treatment of AML (NCT03151408).

With "Classical" Chemo- and Immunotherapies

Great promises arise from the use of nucleoside DNMTi in combination with chemotherapies or immunotherapies in tumors (Table 17.1). The rational of these combination approaches relies on the fact that, due to their capacity to globally affect cells and restore cell functions, DNMTi can increase the sensitivity to other anticancer agents, resulting in higher efficacy (Ahuja et al. 2014, 2016; Azad et al. 2013; Cameron et al. 1999). A plethora of clinical trials, involving multi-anticancer agents in combination with both nucleoside DNMTi (5aza and 5azadC), are currently ongoing, mainly in hematologic cancers with 39 phase III/IV active clinical trials. Many phase III/IV trials have been undertaken in the last few years. 5aza and 5azadC are in phase III in combination with "classical" chemotherapies such as doxorubicin, all-*trans* retinoic acid (ATRA) or cytarabine. Both nucleosides are also in phase III and IV in combination with Venetoclax, a Bcl2 inhibitor, Glasdegib, a Hedgehog pathway inhibitor, Ivosidenib, an IDH1 inhibitor, Eprenetapopt, a *P53* reactivator, or the kinase

inhibitors Midostaurin, Gilteritinib, Clofarabine, exclusively against hematological cancers, i.e., AML, MDS or CMML. Besides this, several combinations of nucleoside DNMTi with monoclonal antibodies, such as anti-CD47 (Magrolimab), anti-CD3 (Visilizumab), anti-CD123 (Talacotuzumab), anti-TIM-3 (Sabatolimab) or anti-PD-1 (Camrelizumab), have reached phase III and IV. Finally, a combination of 5azadC and Cedazuridine, a cytidine deaminase inhibitor, was approved in 2020 for the treatment of MDS (Inqovi®).

In parallel, several studies described an enhancement of radiosensitivity by 5azadC treatment in lung cancer A549 and glioblastoma U373MG cells (Kim et al. 2012), SaOS2, HOS, and U2OS osteosarcoma cells (Li et al. 2014) and MDA-MB-231 and MDA-MB-435 breast cancer cell lines (Wang et al. 2013a). In all cases, the increase in radiosensitivity was correlated with cell cycle-, apoptosis- and radiosensitivity-related gene up-regulation and G2/M cell arrest.

In summary, nucleoside DNMTi are frequently used for hematological cancers. New clinical trials are ongoing to optimize the treatment schedule using a low-dose strategy, and evaluate DNMTi efficacy in solid tumors. SGI-110 could eventually overcome the compound stability limitation. In parallel, their use in combination with other anticancer agents is giving promising results increasing the efficacy of the treatments and decreasing the side effects, since lower doses of each drug are used.

17.3.1.2 Non-nucleoside DNMTi

As described above, several non-nucleoside DNMTi have been reported, but none obtained the same potency and validation as 5aza and 5azadC. Further studies and clinical trials are required to better understand the therapeutic potential of DNA methylation (Erdmann et al. 2015). Hydralazine (de la Cruz-Hernandez et al. 2011; Graça et al. 2014b) failed as single therapeutic agent in phase II for breast and rectal cancer (Wang et al. 2009). Its combination with VPA (Duenas-Gonzalez et al. 2008) was completed in phase II against cervical cancer (NCT00404326) and refractory solid tumors (NCT00404508) more than 10 years ago and no further studies had been started since.

Indirect effects on DNA methylation are also of therapeutic interest. In the presence of disulfiram, in vitro DNMT1 activity was shown to be reduced, and global hypomethylation was observed in prostate cancer cell lines with decrease in *APC* and *RAR-β* promoter methylation accompanied by their re-expression. Xenograft tumor volumes were shown to be reduced under disulfiram treatment independently of the dose (from 10 to 40 mg kg^{-1}) (Lin et al. 2011). However, a completed Disulfiram phase II study prostate cancer showed a limited impact (NCT01118741). Combination of disulfiram with copper improved its potency in refractory cancers (Li et al. 2020a). This combination is currently in phase II against metastatic breast (NCT03323346) and pancreatic (NCT03714555) cancers and in phase III against glioblastoma (NCT02678975).

RG108, found by virtual screening, was shown to inhibit methylation in NALM6 (leukemia cell line) and in HCT116 (colorectal cancer cell line) inducing re-expression of *P16* and *TIMP3* (Stresemann et al. 2006). In prostate cancer cell lines (LNCaP and 22Rv1), it decreased DNMT activity together with cell growth inhibition and apoptosis (Graça et al. 2014a). However, the mechanism by which RG108 treatment resulted in a decrease of the methylation level in cells remains to be understood. Nevertheless, since it is commercially available, it was used as a tool to inhibit DNA methylation in several cellular models (see below).

17.3.2 DNMTi Application in Neurological and Psychiatric Disorders

DNA methylation, together with other epigenetic modifications, is also deregulated in neurological diseases and psychiatric disorders, such as schizophrenia, epilepsy or bipolar disorder (for review, Mohd Murshid et al. 2022). However, to date, no DNMTi has been approved for these diseases, the main hurdle being the poor capability of the

existing DNMTi to cross the blood-brain barrier (BBB). Nevertheless, interesting results listed here below highlight the great interest of targeting DNA methylation in neurological and psychiatric disorders.

17.3.2.1 Memory Formation

Memory loss is involved in many neurodegenerative diseases, such as Alzheimer's and Parkinson's diseases, schizophrenia or post-traumatic stress disorders. DNA methylation is implicated in memory formation, as it regulates gene transcription in central nervous system (CNS) and is required in long-term memory formation (Day and Sweatt 2010). Additionally, DNA methylation was demonstrated to be involved in synaptic plasticity as *Dnmt1/Dnmt3a* double knockout in mice exhibit long-term potentiation (LTP) deficiencies (Feng et al. 2010). To better understand the implication of DNA methylation in various brain regions, rats were treated with RG108 by using the object-in-place paradigm, a test based on the memory of the positioning of objects in an enclosed space, which requires both hippocampus and perirhinal cortex brain regions. Using this model, Michnick et al. demonstrated that DNA methylation was required for long-term but not short-term memory (Mitchnick et al. 2015). In parallel, expression of *Pp1*, a memory suppressor gene, and *Reln*, coding for reelin, a positive memory regulator, was shown to be controlled by promoter methylation. Following fear conditioning, an increase in *Pp1* promoter methylation and a decrease in *Reln* promoter methylation levels were observed in adult rat hippocampus. Infusion of 5aza in adult mouse brain inverted these effects, decreased *Pp1* methylation and enhanced the low methylation level of *Reln*, which inversely correlated with the expression of both genes (Miller and Sweatt 2007). In rats trained for contextual fear memory, the authors observed that intra-anterior cingulate cortex infusions of 5aza or zebularine, 30 days post-training, disrupted remote memory (Miller et al. 2010). Reelin is also implicated in synaptic plasticity that is involved in storage memory. Levenson et al. demonstrated that zebularine induced a significant methylation decrease in one of the two CpG islands of the *Reln* promoter, whereas the other CpG methylation level remained steady (Levenson et al. 2006). The impact of DNMTi in synaptic plasticity was studied by Nelson et al. in hippocampus slices, where 5azadC was reported to decrease the genomic DNA methylation level, concomitantly with a reduction of miniature excitatory post-synaptic current frequencies in neurons, impacting neuronal activity (Nelson et al. 2008). Tetrodotoxin, known to decrease neuronal activity, also induced a loss of methylation and the same effects were observed both by using RG108 or by knocking down *Dnmt1* and *Dnmt3a* (Meadows et al. 2015). Treatment of honeybees with RG108 also induced relearning and bees were able to discriminate between a known and a new odor (Biergans et al. 2016).

Altogether, these studies show that DNA methylation is implicated in memory regulation, long-term memory and synaptic plasticity. DNMTi were shown to interfere with the long-term memory and to enhance synapse receptiveness in neurons (synaptic upscaling). Consequently, DNMTi were shown to impair memory, suggesting a positive effect on post-traumatic stress disorders (PTSD) caused by long-lasting traumatic memory (see below). Animal studies using 5azadC were in most cases carried out by direct injection in brain regions or on slice cultures. However, 5azadC injected intraperitoneally in mice led also to an increase of *Bdnf* expression and a lower global DNA methylation in the hippocampus (Sales et al. 2011). These effects on memory functions were hypothesized as limiting factor for the use of decitabine in MDS patients (Aydin et al. 2012).

17.3.2.2 Schizophrenia

Schizophrenia is a cognition disorder often characterized by hallucination, paranoia and failure to adopt a "normal" behavior in social situations. It is now quite clearly established that a deregulation of the GABA ((γ-aminobutyric acid)ergenic/glutamatergic) network in the hippocampus and cortex is characteristic of psychotic disorders, including schizophrenia (Lewis et al. 2005). More specifically, a down-regulation of

GABAergic genes, such as *glutamic acid decarboxylase$_{67}$* (*GAD$_{67}$*) and *RELN*, was measured in post-mortem samples of schizophrenia patients (Guidotti et al. 2000). This down-regulation was correlated with an hypermethylation of their CpG island promoter regions (Chen et al. 2002; Grayson et al. 2005) and with an increase in DNMT1 (Veldic et al. 2004) and DNMT3A (Zhubi et al. 2009) mRNA expression in schizophrenia post-mortem cortical GABAergic neurons. In a model of prenatal-stress mice exhibiting a schizophrenia-like behavior, high levels of Dnmt1 and Dnmt3a expression occurred in GABAergic neurons (Matrisciano et al. 2013). Additionally, a knock-down of *Dnmt1* in mouse primary cortical cultures showed that *Reln* expression is controlled by Dnmt1 (Noh et al. 2005).

In this context, DNMTi were tested to decrease the methylation level and restore a normal expression of *GAD$_{67}$* and *RELN*. 5aza, zebularine, and procainamide were tested in cell cultures and an increase in the expression of *GAD$_{67}$* and *RELN* and a decrease in *DNMT1* expression were observed (Kundakovic et al. 2007). However, 5aza and zebularine, the most active molecules in the cellular context, present limiting potential to cross the blood-brain barrier (BBB).

17.3.2.3 Bipolar Disorders

Bipolar disorders are maniac-depressive disorders characterized by a succession of elevated mood and depression periods. Aberrant DNA methylation patterns are implicated in the etiology of bipolar disorders and, very similarly to schizophrenia, patients with bipolar disorders show a down-regulation of *GAD$_{67}$* and *RELN* (Guidotti et al. 2000). Because of the lack of relevant animal models to study bipolar disorders, very few studies involving DNMTi have been carried out. Mainly, VPA, an HDACi, was showed to decrease DNA methylation without decrease in DNMT expression (Aizawa and Yutaka 2015). VPA is FDA approved for bipolar disorders and in clinical studies in combination with antipsychotics.

17.3.2.4 Epilepsy

Epilepsy is a group of neurological diseases for which causes are unknown in most cases. Epileptic patients are subject to recurrent unprovoked seizures, which are brief to quite long periods of intense shaking.

In temporal lobe epilepsy, the most common epilepsy syndrome in human, low levels of reelin were observed in brain specimens (Heinrich et al. 2006) in correlation with a hypermethylation of its promoter region (Kobow et al. 2009). A genome-wide study revealed a global DNA hypermethylation in chronic epileptic rat correlated to a decrease in reelin gene expression at the mRNA level (Kobow et al. 2013). Moreover, *Gria2*, a Glu2A subunit of ionotropic glutamate receptor, has been identified to play a role in epilepsy, and its hypermethylation and corresponding decrease in mRNA expression was correlated to the intensity and frequency of seizure in rats (Machnes et al. 2013). In an equivalent rat model of kainic acid-induced epilepsy, bisulfite sequencing showed a hypermethylation of *Grin2b/Nr2b*, resulting in a lower level of the GRIN2B protein, another glutamate receptor subunit, together with a hypomethylation of the *Bdnf* gene in epileptic hippocampus (Parrish et al. 2013). RG108 treatment on kainite-treated hippocampal slice cultures showed an inhibition of the *Gria2* hypermethylation and a beneficial increase in Gria2 activity (Machnes et al. 2013). Zebularine also decreased the methylation level of *Grin2b/Nr2b,* concomitantly with an increase in GRIN2B protein level (Parrish et al. 2013). All these findings are in favor of applying DNMTi to epilepsy treatment. It was also demonstrated that treatment with RG108 led to a lower methylation level of *RASgrf1*gene and increase its expression, which was associated with the suppression of epileptic seizure in epileptic mice model (Chen et al. 2017).

17.3.2.5 Post-traumatic Stress Disorder

Post-traumatic stress disorders (PTSD) appear in patients who experienced a psychologically traumatic event, such as violent death witnessing, domestic violence, child abuse, etc. and result in

severe anxiety associated with memory avoidance, flashbacks, nightmares and emotional arousal. For example, war veterans commonly suffer from PTSD.

In PTSD patients, altered DNA methylation was observed in the peripheral blood immune cells. Together with other genes (Uddin et al. 2010), *BDNF* was reported to be hypermethylated (Smith et al. 2011). In rat models, epigenetic modifications were identified to play a role in fear memory and Miller et al. demonstrated that *Dnmt3a* and *3b* are up-regulated, whereas level of Dnmt1 remains steady in rat hippocampus in response to contextual fear conditioning (Miller and Sweatt 2007). Models of maltreated rats compared to normal animals also showed an increase in DNA methylation levels of *Bdnf*, associated with a decrease of *Bdnf* expression. This modification is a lasting effect, which was observed in adulthood and passed to the next generation (Roth et al. 2009). DNA methylation of *BDNF* contributes to learned fear and BDNF is important for the persistence of the pathological fear. Therefore, DNMTi could be envisaged to erase the memory of the trauma in PTSD patients (Zovkic and Sweatt 2013). Administration of DNMTi (5aza or zebularine) induces *Bdnf* expression even in adult animals. In PTSD, an overconsolidation of the fear memory is observed and nucleoside DNMTi and RG108 were able to block the contextual fear memory formation. RG108 was also tested directly in rat brains and shown to abolish long-term fear memory (Miller et al. 2010).

In rodent, neonatal maternal separation (NMS) attenuates neuron differentiation and can lead to stress-induced behavioral disorders. It results in a decrease in *RARα* expression concomitant with an increase in the *RARα* promoter methylation. *Dnmt1* was shown to be up-regulated, whereas no change was measured in *Dnmt3a* and *Dnmt3b* expression in adult dentate gyrus-derived neural precursor cells (ADP). 5azadC was capable of increasing the neuronal differentiation and decreasing methylation of *RARα* (Boku et al. 2015).

17.3.2.6 Depression

Sadness, low mood, loss of motivation, and low self-esteem characterize pathological depression, which in the most severe cases can lead to suicide. For example, 38,000 cases of suicide were reported in 2010 in the USA, which represents more than the number of human beings killed by car accident (34,000) or twice more than victims of homicide (16,000) (Nature Editorials 2014). About 60% of the subjects who committed suicide had depression or related mood disorders.

Until recently, the only epidrugs used to treat depression were HDACi (Sun et al. 2013). However, a couple of years ago, DNA methylation was proven to be involved in this disorder, and the environment was shown to play a major role, particularly early-life environmental stress (Booij et al. 2013). Dnmt3a was identified in a mouse model as regulating emotional behavior (LaPlant et al. 2010). In patients with severe depression and previous suicidal attempts, *BDNF* hypermethylation was proposed as a biological marker of suicidal behavior (Kang et al. 2013). Additionally, the gene encoding for P11 protein, a modulator of neuronal function involved in depression, was demonstrated to be hypermethylated in rodent and human depression. Interestingly, after treatment with anti-depressant, such as escitalopram (a serotonin reuptake inhibitor), the methylation of *P11* was shown to return to normal (Melas et al. 2012). Hypermethylation of the promoter region of *TRKB*, encoding for tropomyosin receptor kinase B (TrkB), a BDNF receptor, was also reported in suicide completers accompanied by a decrease of mRNA expression (Ernst et al. 2009).

Sales et al. observed that 5aza, 5azadC and RG108 possessed anti-depressant-like effects in rats increasing the mobility in the tail suspension test and decreasing the immobility time in the forced swimming test (Sales et al. 2011). This was correlated to an increase of *BDNF* expression level. In a model of chronic ultra-middle stress exposure in mice, representative of a depression-like behavior, zebularine and RG108 reverted the depression-like behavior with an increase in *Gdnf* expression (Uchida et al. 2011).

17.3.2.7 X-Chromosome-Related Diseases and Autism Disorders

Patients with Rett syndrome, predominantly occurring in women, show a normal development for the first 6–18 months, then progressively loose speech and hand use, associated with seizure and autism. Rett syndrome is an X-chromosome-related neurological disease caused by a mutation in the MeCP2 (methyl-CpG-binding protein 2) protein (Amir et al. 1999; Guy et al. 2007). This protein is known to bind methylated DNA, repress transcription, and to be involved in gene regulation *via* the recruitment of the co-activator CREB1 (Lyst and Bird 2015). Although DNA methylation pattern is described as crucial for MeCP2 recognition, no DNMTi was used to treat Rett syndrome. However, treatment of hippocampal cells of MeCP2 knockout mice with methyl donor SAM was able to rescue neurotransmission event frequencies and thus to partially compensate MeCP2 loss of function (Nelson et al. 2008).

Additionally, in X-fragile syndrome, another related autism disorder, hypermethylation in the promoter region of the *FMR1* gene is observed (Sutcliffe et al. 1992). The treatment of X-fragile cell lines with 5azadC was reported to slightly re-induce *FRM1* expression (Tabolacci et al. 2005).

17.3.2.8 Parkinson's and Alzheimer's Diseases

Parkinson's disease (PD) is a CNS disease due to the degeneration of dopamine-producing neurons, and it affects more than 4 million people worldwide. The movement-related symptoms include shaking and dementia and can occur in a later stage of the disease. Hypomethylation is observed in PD patients in the CpG-rich island of *SNCA*, gene coding for α-synuclein (Jowaed et al. 2010). α-Synuclein is a protein involved in aberrant soluble oligomers that lead to neurons death. The reduced levels of *SNCA* methylation were shown to result from a "sequestration" of DNMT1 in the cytosol in an α-synuclein-transgenic mice and post-mortem PD brains (Desplats et al. 2011). 5azadC was tested on dopaminergic neurons and resulted in viability decrease and increase of apoptosis associated with an up-regulation of α-synuclein (Wang et al. 2013b). DNMTi can thus have deleterious effects for PD patients.

Alzheimer's disease (AD) is a neurodegenerative disorder resulting in severe dementia. The methylation profile of AD disease patients is controversial as both hypo- and hypermethylation were reported (Coppieters and Dragunow 2011). Nuclear immunostaining of DNMT1 and other component of the methylation complex such as MBD2/3 was significantly diminished in neurons (Mastroeni et al. 2010), where a loss of methylation of the amyloïd precursor proteins was observed. A decrease in *BACE* and *PSEN1* (encoding for presenilin 1) methylation was also shown in AD patients and was associated with a potential over-expression of amyloid β-peptides (Mastroeni et al. 2010; Scarpa et al. 2003). A cocktail of DNA methylation enhancers, such as folate, is currently in phase III clinical trial (NCT00056225). Moreover, a dramatic global hypermethylation in the gray matter of post-mortem human brain tissues was also observed in AD patients (Coppieters et al. 2014) together with the increase of methylation in certain regions such as *MCF2L* and *ANK1* genes (De Jager et al. 2014). Additional data are crucial to confirm the hypermethylation profile of AD patient and to potentially envisage the use of DNMTi for AD treatment.

17.3.2.9 Aging-Related Senescence and Amyotrophic Lateral Sclerosis

In amyotrophic lateral sclerosis (ALS), motor neurons death leads to muscle weakness and evolves to breathing difficulties. This neurodegenerative disease, often associated with aging, has a 3–5-year survival prognostic (Cleveland and Rothstein 2001). In motor cortex tissues and spinal cord motor neurons of ALS patients, levels of DNMT1 and DNMT3A were shown to be increased (Chestnut et al. 2011; Cho et al. 2010). Oh et al. reported an increase in DNMT1 and DNMT3A expression in ALS mesenchymal

stromal cells (ALS-MSCs) compared to normal MSCs. RG108 treatment of ALS-MSCs isolated form ALS patients' bone marrow induced anti-senescence factors (*TERT*, *VEGF*, *ANG*) together with the down-regulation of senescence factors (*ATM*, *P21*). Additionally, a significant improvement in ALS-MSC migration and their differentiation into neurons was observed (Oh et al. 2015b). Thus, RG108 treatment is promising for a more efficient ALS treatment by autologous cells therapies.

In addition, Oh et al. evaluated the effect of RG108 on human bone marrow mesenchymal stromal cells (hBM-MSCs), which are used in cell therapies. They confirmed the anti-senescence effect of RG108 in in vitro cultures with an increase in the expression of anti-senescence factors (*TERT*, *VEGF*, *bFGF* and *ANG*) and a decrease of senescence-related factors (*ATM*, *P21* and *P53*) (Oh et al. 2015a). Thus, an optimized dose of RG108 (5 μM), for which cell viability is maintained, could greatly improve hBM-MSC potency, which could constitute a real progress in the improvement of the stem cell therapies.

17.3.2.10 Neuronal Stem Cell

Neuronal stem cell (NSC) cultures are potential sources of transplantable cells to treat neurodegenerative diseases. DNA methylation was proven to be essential for NSC differentiation and proliferation. Folic acid was shown to stimulate neonatal rat NSC proliferation in vitro (Li et al. 2013), whereas zebularine treatment resulted in the attenuation of their proliferation (Lin et al. 2014; Luo et al. 2013) and a reduction of their migration (Singh et al. 2009b). Hence DNMTi are deleterious for NSC growth, but the use of DNMT activators can provide an interesting way to optimize NSC cultures.

DNA methylation was shown to be implicated in many neurological diseases and psychiatric disorders. The consequences of DNA methylation depend on the disease and the use of DNMTi can be, in a few cases, unfavorable, like in Alzheimer's disease, but it can also have a positive outcome in pathologies like schizophrenia, depression and post-traumatic stress disorders. Despite the high interest in controlling DNMT activity, the path to use DNMTi as drugs to treat neurological diseases and psychiatric disorders is still long. First, it is made difficult by the lack of animal models to study these pathologies. Second, active molecules must cross the blood-brain barrier, and, for example, 5aza or zebularine are not able to do so. Additionally, toxicity is not acceptable for psychiatric diseases, which often require chronic and life-long treatment. Finally, the specificity of action is also an issue as DNA methylation is involved in biological processes in all tissues and organs. Despite these limitations, the better comprehension of DNA methylation in neurological diseases will led to new therapeutic strategies and new small molecule DNMTi are urgently needed for neurodegenerative diseases and psychiatric disorders (Szyf 2015).

17.3.3 DNMTi Application in Cardiovascular Diseases

Cardiovascular diseases (CVDs) are the leading cause of death in most of the developed countries and environmental factors such as diet, smoking habits, or impaired metabolism are critical risk factors. Therefore, it is not surprising that epigenetic modifications, including DNA methylation, are involved in these pathologies.

DNA methylation patterns were reported to be altered in several CVDs. In hypertension, the hypermethylation of the *HSD11B2* promoter, a gene affecting blood pressure, was observed (Friso et al. 2008). Hypermethylation was also reported for the ATP-binding cassette transporter A1 gene (*ABCA1*) in coronary heart disease (Guay et al. 2014) and, in an atherosclerosis Apoe$^{-/-}$ mouse model, the hypermethylation of specific vascular homeostasis genes was described (Zaina et al. 2014). A study on about 300 Singapore Chinese subjects also proposed DNA methylation as CVD risk biomarker, specifically the *Alu*/*STAT2* methylation level (Kim et al. 2010). In parallel, in the early-stage atherosclerosis Apoe$^{-/-}$ mouse model a global hypomethylation was reported in peripheral blood mononuclear cells (Lund et al. 2004).

Concerning the use of DNMTi in CVDs, in the context of norepinephrine-induced heart hypertrophy, a 6-days treatment with 5azadC restored a normal protein expression profile of the whole cardiac proteome and rescued the phenotype of norepinephrine-treated rats (Xiao et al. 2014). 5azadC was also shown to facilitate inorganic phosphorus mineralization of human aortic smooth muscle cells (HASMCs) via up-regulation of the alkaline phosphatase (ALP), a risk factor for cardiovascular disease in patients with chronic kidney disease (CKD) (Azechi et al. 2014).

Atherosclerosis is characterized by an inflammation of arterial wall and it is the major cause of stroke and heart attack. Gene expression in endothelial cells changes dramatically when submitted to blood flow variations and disturbed flow is pro-atherogenic. In mice, disturbed blood flow induces *Dnmt* expression, moreover several down-regulated mechano-sensitive genes were identified and their expression was shown to be under their promoter methylation control and reversed by 5azadC (Dunn et al. 2014). Besides, estrogen receptors (ERα, ERβ) are artheroprotective and hypermethylation of *ESR1* (ERα) and *ESR2* (ERβ) promoters, associated with their silencing, were shown in atherosclerotic tissues and in senescing cells. Treatment with 5azadC was carried out on smooth muscle cells (SMCs) and endothelial cells (ECs), resulting in the decrease of estrogen receptor gene promoter methylation accompanied by an increase of ER levels in both vascular cell lines. Additionally, the combination 5azadC/TSA showed some synergetic effect, while TSA alone had no effect (Kim et al. 2007).

Ischemia is a decrease in blood supply to tissue, which results in hypoxia and leads to cardiac fibrosis. In hypoxia-induced pro-fibrotic state of human cardiac fibroblast, an increase in *DNMT1* and *DNMT3B* expression and a global hypermethylation were reported. The expression of pro-fibrotic genes, such as *alpha-smooth muscle actin* (*ASMA*) or *collagen 1*, increased and this was enhanced by treatment with TGFβ, a pro-fibrotic cytokine. Interestingly, treatment with 5azadC significantly reduced TGFβ effects and levels of ASMA and collagen 1, decreasing pro-fibrotic effects and positioning DNMT as potentially valuable therapeutic target in ischemic heart disease (Watson et al. 2014). Finally, exploring the influence of DNMTi on mouse embryonic stem cells (mESCs), zebularine was shown to re-express cardiac-specific genes, such as *Nkx2.5* and *Gata4* (Horrillo et al. 2013).

DNA methylation is also implicated in heart failure. Indeed, a higher level of homeobox gene *PITX2c* promoter methylation accompanied with increased DNMT1 and decreased PITX2 protein levels was identified in heart disease atrium. In the HL-1 cell line, 5azadC reversed *PITX2c* promoter methylation and increased the PITX2 and KIR2.1 protein level (Kao et al. 2013), which could have a positive outcome in case of heart failure.

Recently, DNA methylation and expression levels of COX2 in human heart mesenchymal stem cells (HMSCs) were shown to be involved in the senescence and aging of HMSCs. Treatment with 5azadC lowered COX2 methylation level, restored its expression level, promoting HMSCs proliferation and delating aging (Sun et al. 2021b). This highlighted the potential role of COX2 as a new therapeutic target against cardiovascular diseases. Despite the evidence of the role of DNA methylation in CVDs (Chaturvedi and Tyagi 2014), no epidrug is currently in clinical trial for CVDs.

17.3.4 DNMTi Application in Other Human Pathologies

17.3.4.1 Obesity

Obesity has a high prevalence in industrialized countries where high caloric diet is common. This major public health problem can lead to diabetes or cardiovascular diseases. Following a 5 days high fat diet, DNA methylation changes of the transcriptional co-activator PGC-1α involved in oxidative energy metabolism were observed (Brøns et al. 2010). Variations of DNA methylation were also reported by Wang et al. who analyzed DNA methylation in peripheral blood leukocytes and determined that *UBASH3A*

was hypermethylated, whereas *TRIM3* was hypomethylated in obesity cases (Wang et al. 2010b). Variability of DNA methylation in individual adipose tissues was shown to influence their response to caloric restriction in terms of weight loss (Bouchard et al. 2010). A comparison of methylation levels in abdominal adipose tissues before and after gastric bypass and weight loss showed a decrease of DNA methylation in both tissues after gastric bypass (Benton et al. 2015). Expression of adiponectin, a protein regulating glucose and lipid metabolism (Yamauchi et al. 2002), was inversely correlated with insulin resistance, type 2 diabetes and cardiovascular diseases (Kadowaki et al. 2006). Kim et al. established that adiponectin expression was under epigenetic control by hypermethylation of the R2 promoter region observed in high-fat diet obese mice compared to normal-diet lean mice. RG108 was tested in an obese mice model (*db/db* mice) and increased adiponectin expression levels were shown to lead to an improvement of glucose intolerance and insulin resistance (Kim et al. 2015). These recent results are promising for obesity-related disease therapeutics.

17.3.4.2 Alcohol Addiction

Alcohol abuse with its associated diseases and behaviors is responsible for about 6% of the death worldwide (WHO 2014 Global status report on alcohol and health). In the UK, alcohol was classified as the most harmful drug for oneself and others (Nutt et al. 2010). The influence of alcohol consumption on DNA methylation is controversial and seems to be highly dependent on the studied tissues. For example, DNA in post-mortem human brains was reported to be hypomethylated in alcoholics (Ponomarev et al. 2012). This hypomethylation was shown to result in the increased expression of long-term repeat-containing human endogenous retrovirus. A decrease in DNMT3A and DNMT3B levels was also shown in post-mortem alcoholic patient brains compared to healthy individuals. DNMT3B expression was reported to be inversely proportional to blood alcohol concentration, whereas no variation was reported for DNMT1 expression level (Bönsch et al. 2006). Genome-wide DNA hypermethylation in lymphoblasts of 165 female subjects showed a correlation of the increase in DNA methylation with the drinking frequency (Philibert et al. 2012).

DNMTi were studied for the treatment of alcohol addiction and encouraging outcomes were described. In an alcohol-exposed rat model submitted to the two-bottle choice test and treated or not with 5aza, Warnault et al. found that for 5aza-treated animals the rate of alcohol intake was significantly lower compared to non-treated alcohol-dependent rats (Warnault et al. 2013). These results also showed that inhibition of DNA methylation enabled to specifically reduce alcohol consumption and preference but did not influence other rewarding substance intake. DNA methylation was suggested to increase the expression of endogenous factors interfering with alcohol drinking behavior, such as BDNF (Logrip et al. 2009) and GDNF (Carnicella et al. 2008) that prevent the escalation from moderate to excessive consumption. Barbier et al. studied long-term behavior in alcohol-conditioned rats after 3 weeks of abstinence and established a causal relationship between DNA methylation, alcohol intake and seeking behavior. They measured an increase in Dnmt1 and no change in Dnmt3A and Dnmt3B levels. Using intracerebroventricular infusion of RG108, they lowered the DNA methylation level in mPFC (medial prefrontal cortex) of alcohol-conditioned rats, while no change was observed in controls. DNMTi also prevented escalation of alcohol consumption. Seven genes involved in neurotransmission and coding for synaptic proteins were down-regulated in alcohol-dependent rats. Among them, synaptotagmin 2 (*Syt2*) presented a higher methylation level in alcohol post-dependent rats and a synaptic transmission deregulation was reported in post-dependent mPFC neurons. A *Syt2* knock-down resulted in compulsive-like drinking behavior. Interestingly, *Syt2* normal methylation level and expression were restored by RG108 treatment (Barbier et al. 2015).

Although variations of DNA methylation levels in the brain of alcoholic patients seem to be heterogeneous, the results of DNMTi on

preclinical animal models are an interesting starting point for the use of DNMTi to limit alcohol dependence.

17.3.4.3 Inflammation and Allergy

Inflammation is a reaction of the immune system to harmful stimuli. In allergy, also known as hypersensitivity (e.g., hay fever, allergic asthma, etc.), inflammation is triggered by an unsuitable immune response to harmless environmental cues. Similarly, in auto-immune diseases, an abnormal immune response is directed towards tissues normally present in the organism. Among the $CD4^{+}$ T-cells (T-helper (Th) cells), T-helper 1 (Th1) express specifically interferon-γ (INFγ) (Brand et al. 2012), whereas T-helper 2 (Th2) express interleukin 4 (IL4), IL5 and IL13. The INFγ (Winders et al. 2004) and IL4 (Kwon et al. 2008) cytokine expression is under DNA methylation control and differentiation of naïve T-cell into Th1 or Th2 cells results in a modification of the methylation levels of CpG sites in *INFG* coding for INFγ (White et al. 2006). The level of *INFG* promoter methylation in CD4+ lymphocytes of adults and children with asthma was shown to depend on age, sex and tissues (Lovinsky-Desir et al. 2014). However, in this study, the authors hypothesized that the increase of *INFG* promoter methylation might not be the only reason of its lower expression and further investigation is required. In mouse models, *Dnmt3a* and *Dnmt3b* were more expressed in Th2 cells compared to Th1 cells, whereas no difference in *Dnmt1* expression was observed (Yu et al. 2012). Moreover, in the same study, *INFG* and *IL4* expression was lower in mutant mice lacking Dnmt3a or Dnmt3a/3b but not in mutants lacking only Dnmt3b. Altogether these data highlight the importance of DNA methylation in the regulation of cytokine expression in $CD4^{+}$ T-cells and in inflammation reactions.

In inflammation, the balance of T-cell sub-type is disturbed with a shift from Th1 towards Th2 response. As a result, the ratio INFγ vs. IL4 production is modified, which can be considered as an indicator of inflammation risk (Shahid et al. 2002). In patients with bronchial asthma, T-cell differentiation into Th2 sub-type and IL4-associated cytokine are much higher than in controls (Kwon et al. 2008). Although genetics predispositions are involved in asthma (Lloyd and Hawrylowicz 2009), environment such as pollution, cigarette smoke, etc., is also well-known to have a crucial influence, which rises the importance of epimutations in asthmatic patients. A comparison of 21 monozygotic adult twins discordant for asthma showed a decrease in *INFG* and *FOXP3* expression in the case of asthma with a greater decrease for cases with second-hand smoke (SHS) during childhood, which was shown to significantly affect *INFG* and *FOXP3* methylation levels (Kohli et al. 2012; Runyon et al. 2012). A similar study with monozygotic twins concluded the importance of DNA methylation of genes involved in immune response associated with psoriasis, like *TNFSF11*, in $CD4^{+}$ T-cells (Gervin et al. 2012). It is important to underline the significance and increased interest in epidemiology studies on the epigenome of monozygotic twins, starting from the seminal paper of Fraga et al. in 2005 (Fraga et al. 2005).

Concerning seasonal allergic rhinitis (SAR), peripheral blood mononuclear cells from SAR patients and healthy controls were challenged with allergen. Interestingly, the DNA methylation profile of SAR patients varied during and outside pollen season (Nestor et al. 2014). DNMTi were tested in vitro and in vivo for their ability to control inflammation. In asthma, upon allergen sensitization/challenge by ovalbumin treatment a comparison of sensitized and non-sensitized mice showed an increase in *INFG* promoter methylation correlating with a decrease in INFγ level. This increase was reversed by 5azadC treatment (Brand et al. 2012).

In acute lung injury (ALI), an inflammatory lung disease mainly caused by sepsis, the combination of 5azadC and TSA reduced mortality level in LPS (lipopolysaccharide)-induced ALI mouse mode with a 80% survival rate (Thangavel et al. 2014). Primary bone marrow-derived macrophages of LPS-treated mice expressed higher levels of pro-inflammatory chemokines

and cytokines, which were decreased upon 5azadC/TSA treatment, resulting in cell survival increase (Thangavel et al. 2015).

Rheumatoid arthritis is a chronic inflammatory auto-immune disease involving synovial inflammation and causing joint pains. In fibroblast-like synoviocytes (FLS) of arthritic rats, the methyl-CpG-binding protein MeCP2 was selectively over-expressed, whereas secreted frizzled-related protein (SFRP4) was down-regulated. A knock-down of MeCP2 resulted in an enhancement of the SFRP4 level. SFRP4 is known to activate the Wnt pathway, which is involved in FSL abnormal proliferation. Treatment of FLS with 5azadC resulted in an increased *SFRP4* expression and a decreased cell proliferation (Miao et al. 2013). Finally, a strong increase in promoter methylation of Th2 genes (*IL4* and *IL5*) with an opposite effect for Th1 genes (*INFG* and *IL10*) was observed in PMBC of children allergic to cow's milk compared to healthy ones (Berni Canani et al. 2015).

In combination, these studies described the potential of DNMTi in the development of new anti-inflammatory treatment in asthma, acute lung injury, and rheumatoid arthritis.

17.3.4.4 Infectious Diseases

Viral Infections

In 2014, 35 million people were affected by HIV (human immunodeficiency virus)-1 worldwide. Most infected people have a normal quality of life when treated with anti-retrovirus therapy, but an interruption in the treatment can reactivate latent viruses (Van Lint et al. 2013). Efforts have been made to improve the anti-HIV therapies. It was shown that methylation of proviral DNA of HIV (Kumar et al. 2015) or human T-lymphotropic virus (HTLV) (Saggioro et al. 1991) is involved in latency establishment and in their escape from the immune system. In HIV-infected cells, an increase of DNMT1 (Mikovits et al. 1998) and DNMT3A/3B levels (Chandel et al. 2015) has been observed. However, the influence of DNA methylation is quite controversial and a hypermethylation of HIV-1 5′ long terminal regions (LTR) was reported in latent HIV reservoirs of patients who did not present viremia compared to viremic patients (Blazkova et al. 2009). A strategic option could be to activate HIV in their reservoir to kill infected cells, while anti-retrovirals are used to block new infections. In this context, Fernandez et al. demonstrated that, in association with TNFα known to activate HIV replication, the use of 5azadC activated HIV twice more than the single use of TNFα in J-Lat cells. However, this seems to be cell line specific, as in other cell lines, such as J1.1 and U1, 5azadC inhibited HIV activation (Fernandez and Zeichner 2010). This variability highlights the complexity in the dynamic of DNA methylation and the need of novel chemical inhibitors to probe the role of DNA methylation in different biological contexts.

HIV is also known to be associated with certain cancers such as non-Hodgkin's lymphomas (NHL). In HIV-positive NHL patients, a clear up-regulation of DNMTs was observed. In HIV-positive aggressive B-cell lymphomas, the HIV-TAT protein was shown to be secreted by infected cells and to result in cell proliferation. In this context, an up-regulation of DNMTs was observed and 5azadC treatment was able to re-express *P16* (Luzzi et al. 2014).

Similarly, human papillomavirus (HPV)-positive lung cancers exhibited high levels of DNMT3B (Lin et al. 2005) and treatment with 5azadC was able to restore normal levels of *E-cadherin* expression in the presence of HPV (D'Costa et al. 2012). These findings could be exploited to improve the treatment of virus-induced cancers like HIV-positive NHL, papilloma virus-positive cancers or Epstein-Barr virus-positive Burkitt's lymphoma (Paschos et al. 2009).

Hepatitis C virus (HCV) was also shown to lower E-cadherin levels and induce an hypermethylation of the *E-cadherin* promoter, which is known to induce morphological changes, alter cell-cell adhesion and induce epithelial-mesenchymal transition, critical in tumorigenesis (Park and Jang 2014). DNA methylation was reported to be deregulated in HCV. DNMT activity was shown to be essential for HCV cell infection and the use of 5aza or

5azadC induced a significant decrease in HCV infection (Chen et al. 2013). Therefore, DNMTs can represent potential targets for HCV treatment.

Bacterial Infections

Epigenetic changes regulate the bacteria-specific innate immune response, as described for inflammation (cf. Sect. 17.3.4.3). Alterations of DNMT expression levels were measured in mice provided with *E. coli* or *E. coli*-LPS contaminated water compared to control mice. This highlighted the role of bacterial LPS in the alteration of the epigenetic response (Kovalchuk et al. 2013). In *E. coli*-infected human uroepithelial cells, DNMT activity was demonstrated to be more than 10-fold enhanced compared to non-infected cells (Tolg et al. 2011). DNMT1 expression was also altered following *Porphyromonas gingivalis* or *Fusobacterium nucleatum* infection (Yin and Chung 2011). In parallel, pre-treatment of gingival epithelial cells with 5aza significantly reduced the up-regulation of cytokine genes *IL6* and *CXCL1*, observed upon *P. gingivalis* or *F. nucleatum* exposure. However, no reduction in bacterial invasion of gingival epithelial cells was observed (Drury and Chung 2015). Hence, although DNA methylation is involved in bacterial infection, the use of DNMTi is still to be studied.

Parasite Infections

DNA methylation was shown to be involved in several parasite infections. Quinazoline-based DNMTi, show oral efficacy in *P. berghei* infected mice (Bouchut et al. 2019) and quinoline–quinazoline DNMTi induced parasite death including artemisinin-resistant pathogens together with a decrease in methylation level (Nardella et al. 2020). DNA methylation inhibition was also studied in *Biomphalaria glabrata*, a mollusk, host of the pathogen *Schistosoma mansoni*. In this invertebrate model, 3-bromo-3-nitroflavanones were shown to reduce DNA methylation in *B. glabrata* and its offspring together with phenotype modification but with no noticeable toxicity. Interestingly, identified hypomethylated loci were associated with reduced gene expression (Luviano et al. 2021).

The importance of the epigenetic modulation of the host in case of infective disease starts to be established and this could open the door to the use of epidrugs as anti-infectious agents. This would certainly be facilitated by the discovery of more potent and selective small molecule DNMTi.

17.3.4.5 Embryo Growth

The establishment and maintenance of DNA methylation patterns are crucial in the early development of the human embryo (Guo et al. 2014). Altered DNMT activity in the embryo can lead to early pregnancy loss (Yin et al. 2012). Early DNA methylation pattern establishment is also important in the in vitro fertilization (IVF) process and development of pre-implanted embryos was shown to be dependent on correct methylation (Dobbs et al. 2013). To better understand and improve the success rate in IVF process, studies were carried out on DNA methylation pattern in animal model embryos. *S*-Adenosyl-L-homocysteine (SAH), a common inhibitor of methyltransferase, was used to study in vitro mouse embryo development. Administered at a period of time preceding de novo methylation, SAH results in a decrease of DNA methylation and improvement of development competency of in vitro cloned embryos compared to non-treated ones (Jafari et al. 2011). Treatment of buffalo skin fibroblast donor cells with the combination 5azadC/TSA decreased global DNA methylation, apoptosis and improved the development of cloned embryos (Saini et al. 2014). RG108 in combination with an HDAC inhibitor showed synergetic effects on somatic cell nuclear transfer, a technique that is known to have limited efficiency in mammalian cloning and in which the DNA methylation profile was shown to be incompletely reprogrammed (Peat and Reik 2012). RG108 at high doses (500 μM) increased embryoblasts compared to non-treated controls, which could be promising for the improvement in cloning efficiency (Li et al. 2011; Watanabe et al. 2013).

17.3.5 DNMTi Application in Metabolite Production

To date, no direct therapeutic studies targeting DNA methylation in fungal infection were carried out. However, DNA methylation was found to be involved in fungal biosynthetic genes repression. When cultured in the presence of 50 μM of 5aza, atlantic-forest-soil-derived *Penicillium citreonigrum* produced fungal exudates, formally known as guttates, secondary metabolites strongly enriched in several specific components, as the azaphilone family, and exhibited new secondary metabolites from the atlantinone family (Wang et al. 2010a). Likewise, supplementation of the *Alternaria sp.* fungus culture medium by 5aza induced the production of toxic metabolites, such as alternariol, altenusin, alternariol-5-*O*-methyl ether, 3′-hydroxyalternariol-5-*O*-methyl ether, which are known as plant-disease inducers and that are dormant in normal culture conditions (Sun et al. 2012).

Additionally, DNMTi and HDACi can be added to bacterial cultures to diversify the metabolite production. For example, 5aza was added to marine fungus *Leucostoma perisoonii* culture medium to optimize the production of cytosporones; two unknown cytosporones were obtained, one of which exhibiting some activity against *Plasmodium falciparum* (Beau et al. 2012). This modulation of the biosynthesis pathways through DNA methylation tailoring could lead to the production and identification of novel molecules that can be therapeutically active compounds

17.3.6 DNMTi in Plants

Variations in plant growth conditions, such as temperature, induce epimutations, which can lead to significant modifications in the phenotype, as, for example, the variation of petal number of the *Rosa hybrid*. This modification was related to DNA hypermethylation of the *RhAG* promoter at low temperature (Cortijo et al. 2014; Ma et al. 2015). In contrast to humans, cytosine methylation in plants does not occur only in a CpG context and 30% of the cytosines are methylated compared to 5–8% in humans (Finnegan et al. 1998). In *Japonica* rice (*Oryza sativa L.*), 5aza or 5azadC (0.3 mM for 16 h to 3 days) were shown to decrease the DNA methylation level and to result in dwarfism. The effect of 5azadC (70% dwarfism effect) was more pronounced than with 5aza (30% dwarfism effect). Unlike in mammals, the DNA methylation profile is not erased during gametogenesis (Kinoshita et al. 2004, 2007), which results in heritable DNA methylation changes in progenies, thus 5aza induced dwarfism in rice was shown to be maintained up to the third generation (Sano et al. 1990). Upon treatment with 5azadC, a lower methylation at the *Xa21G* promoter was measured. *Xa21G* is a resistance gene to the pathogenic bacterium, *X. oryzae pv. Oryzae* (Ronald 1997), and, indeed, the new rice line obtained upon 5aza treatment was proven to possess a resistance trait towards bacterial infection (Akimoto et al. 2007).

Thus, DNMTi can program new features in plants and cultures, like the disease-resistance trait described above or flowering (Kondo et al. 2007). However, as long as no plant-specific DNMTi is identified, attention must be taken that these treatments do not affect the methylation profile of animals and human beings.

17.4 Innovative Indirect and Combined Approaches to Better Target DNA Methylation

Since direct targeting of DNMTs is highly challenging, other approaches can be envisaged. Rather than targeting DNA methylation writers, DNA methylation readers responsible for sensing the presence of methylated DNA and triggering signaling cascades can also be considered. Besides, DNMTi can also be coupled with other active moiety to lead to bifunctional inhibitors (Fig. 17.4).

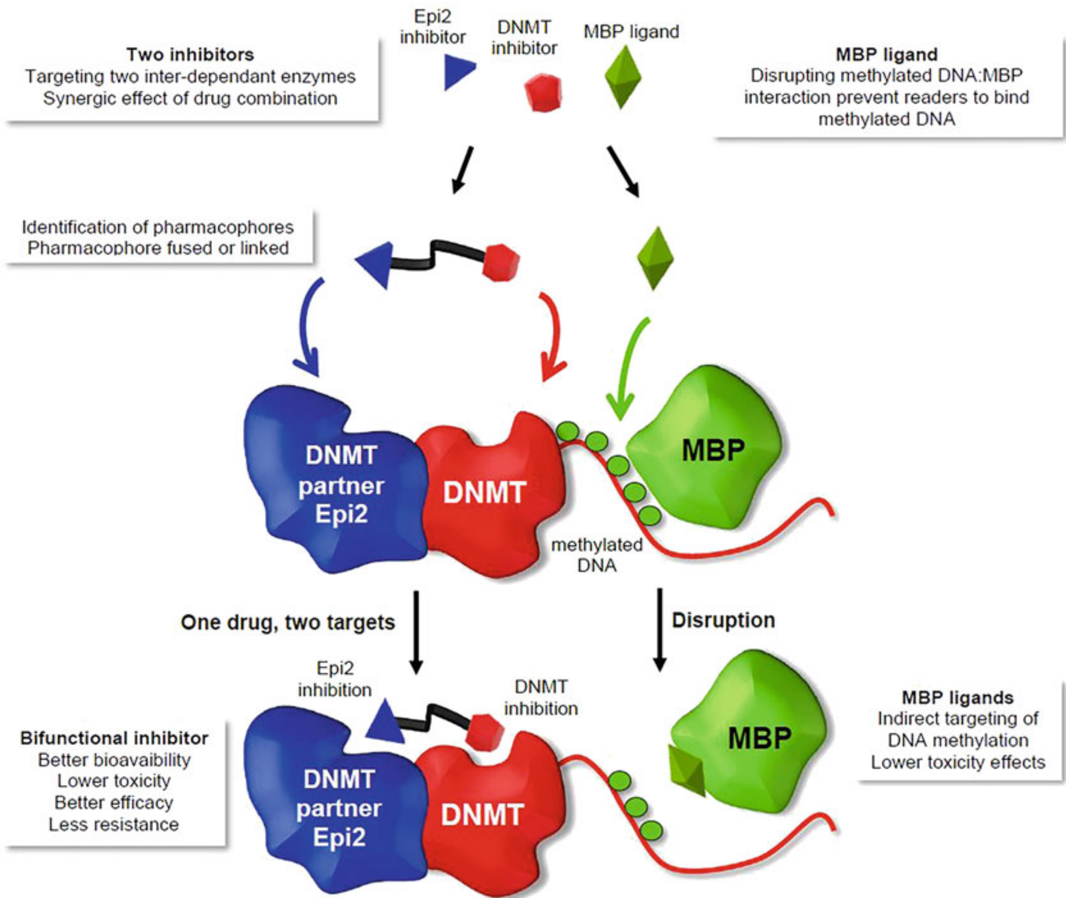


Fig. 17.4 Alternative strategies to target DNA methylation

17.4.1 Methyl-CpG Binding Proteins: Nature and Probes

In addition to the writers of DNA methylation, the DNMT enzymes, and the erasers, the TET enzymes, the readers proteins, known as Methyl-CpG Binding Proteins (MBPs), recognize and bind to methylated cytosine and transform this modification into a biological signal (Ludwig et al. 2016). As DNA methylation is involved in numerous biological processes and in diseases, the MBPs play a global role as well (Sidhu and Capalash 2017; Stirzaker et al. 2017; van Roy and McCrea 2005). Indeed, abnormal MBPs expressions and structures have been reported in psychiatric disorders (Gigek et al. 2016), in different cancer types (Gigek et al. 2016; Li et al. 2020b; Prokhortchouk et al. 2006) and in inflammatory diseases (Robinson et al. 2019), for example. Three different subfamilies of MBPs are known at the date: Methyl-CpG Binding Domain Proteins (MBDs), the SET and RING Associated (SRA) proteins and the Kaiso proteins (Sidhu and Capalash 2017).

17.4.1.1 Methyl-CpG Binding Domain Proteins (MBD)

The Methyl-CpG Binding Domain Proteins (MBD) subfamily is the largest group of MBPs with seven members in total: MeCP2 (Lewis et al. 1992), MBD1 (Fujita et al. 1999), MBD2 (Hendrich and Bird 1998), MBD3 (Hendrich and Bird 1998), MBD4 (Hendrich and Bird 1998), MBD5 (Baymaz et al. 2014), MBD6

(Baymaz et al. 2014). All of them are characterized by an MBD domain comprising approximately 70 amino acids, which mediates their binding to methylated DNA (except for MBD3). MBD4 has in addition an enzymatic activity through its glycosylase domain (Hendrich et al. 1999). Furthermore, four other proteins possess an MBD binding domain in their structure (Du et al. 2015): SETDB1 and SETDB2, which also have a SET domain bearing an histone methyltransferase activity (Falandry et al. 2010; Schultz et al. 2002), BAZ2A (also known as TIPS5) and BAZ2B, which have a bromodomain explaining their affinity for acetylated histones (Jones et al. 2000).

The MBD proteins are involved in several protein regulatory complexes, such as the NuRD complex, which mainly suppresses genes transcription (Xue et al. 1998). Because of their role in mediating the DNA methylation signal, they constitute an interesting target for chemical modulation, as, for example, the disruption of the binding of the MBD domain to methylated DNA could potentially reactivate silenced genes.

Currently, only three binders of the MBD domain have been identified in the literature, all are targeting MBD2: NF449, aurintricarboxylic acid (Wyhs et al. 2014) and KCC-07 (Zhu et al. 2018) (cf. Fig. 17.5). The first two first compounds were identified in an time-resolved (TR)-FRET screening assay for compounds that disrupt the interaction between the MBD domain of hMBD2 and methylated DNA (IC_{50} of 290 nM for NF449 and 2.7 nM for aurintricarboxilic acid) (Tang et al. 2021). Previously, these compounds were known as a P2X1 receptor antagonist and a well-known inhibitor of multiple DNA-protein interactions, respectively. However, both compounds present features that are not favorable for chemical probes or drugs as they interfere with multiple process, lack complete selectivity and bear reactive moieties (El-Ajouz et al. 2012; Roos et al. 2017; Sun et al. 2014). Compound KCC-07 was previously described to interfere with the binding of MBD2-MBD domain to a methylated 45mer DNA with an IC_{50} of 1.55 μM (determined by ELISA), while inhibition of MeCP2-DNA binding was undetectable (Reichert 2010). Interestingly, the compound showed in vivo activity extending the lifespan of medulloblastoma xenografted mice by 6 days and restoring Brain-specific angiogenesis inhibitor (BAI1) expression (Zhu et al. 2018). These results are promising for the use of chemical compounds that target MBDs.

In addition, two compounds have been identified to bind an MBD2 Intrinsic Disordered Domain (IDD) outside the MBD domain: ABA and APC (Kim et al. 2019) (Fig. 17.5). These compounds disrupt the recruitment of P66α by MBD2, which is a key factor of the NuRD complex formation (Gnanapragasam Merlin et al. 2011) with an IC_{50} of 1.93 μM for ABA and 1.75 μM for APC (determined by FRET) (Schultz et al. 2002).

17.4.1.2 SET and RING Associated (SRA) Domain Proteins

The SRA proteins are also able to bind methylated and hydroxymethylated DNA, in particular hemimethylated DNA, through the SRA domain. UHRF1 (Fujimori et al. 1998) and UHRF2 (Fujimori et al. 1998) are the two human members of this family. These proteins can recruit epigenetic silencing factors such as DNMT1 and HDAC1 and promote gene silencing (Niinuma et al. 2019). It has been shown that they play a key role in the recruitment of DNMT1 to hemimethylated DNA (Felle et al. 2011). These proteins act as "pivots" in the gene silencing machinery since they are able to recruit different types of proteins and enzymes to ensure epigenetic DNA methylation, histone methylation and acetylation and even ubiquitin-related actors (Choudhry et al. 2018).

So far, no binder of UHRF1 and UHRF2 has been identified that disrupts their interaction with methylated DNA. The described inhibitors for UHRF1 target mainly its Tandem Tudor Domain and Plant Homeodomain (TTD and PHD, respectively), which mediate the binding to silencing-related proteins such as DNMT1 (Achour et al. 2008) and HDAC1 (Unoki et al. 2004). The most relevant molecules are NV03 (Senisterra et al. 2018), NSC232003 (Myrianthopoulos et al.

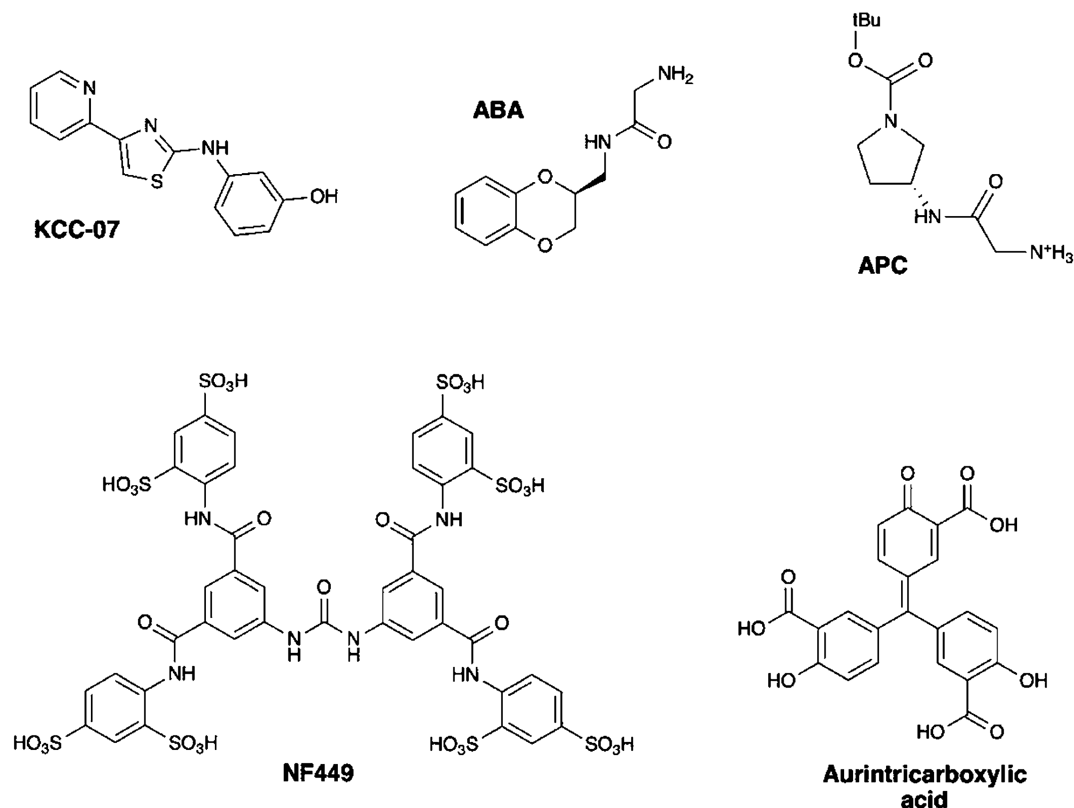


Fig. 17.5 MBD2 targeting compounds

2016), BPC (Houliston et al. 2017) and 2,4-lutidine (Chang et al. 2021) (Fig. 17.6) with an affinity to the TTD domain of $K_d = 2$ μM for NV03 (determined by ITC and Fluorescence polarization) (Senisterra et al. 2018); 50 μM for BPC (determined by ITC and DSF) (Houliston et al. 2017) and 29 μM for 1,4-lutidine (measured by TR-FRET) (Chang et al. 2021). NSC232003, which interacts with UHRF1, the global number of methylated CpGs of a population of U251 glioblastoma cells by 50% (Myrianthopoulos et al. 2016). However, the specificity of these molecules towards UHRF1 over UHRF2 has not been established.

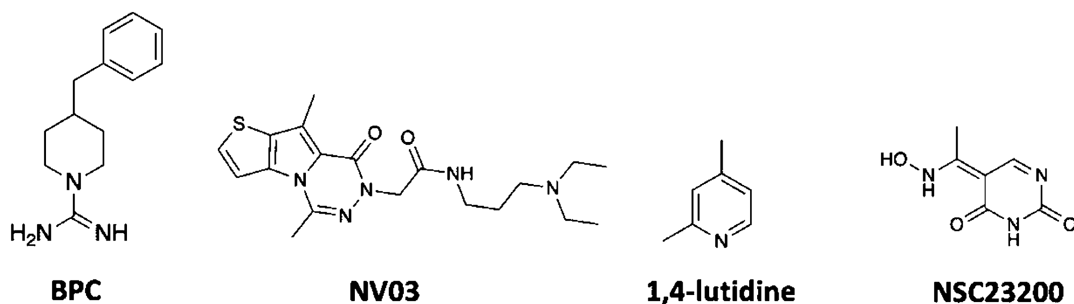


Fig. 17.6 UHRF1 binders

17.4.1.3 Kaiso Proteins

The Kaiso protein subfamily is composed of three different members: Kaiso (Daniel Juliet and Reynolds 1999), ZBTB4 (Filion et al. 2006) and ZTB38 (Sasai et al. 2005). These are Zinc Finger Domains (ZF) proteins that bind methylated and unmethylated DNA in a sequence-specific manner (Donaldson et al. 2012). They are known to recruit HDAC-dependent transcriptional repressors (for ZBTB4 and ZBTB38) (Sasai et al. 2005) or the N-CoR repressive complex (for Kaiso) (Yoon et al. 2003). The implication of these proteins in diseases remarkably differs from one another, going from oncogenesis (Buck-Koehntop Bethany et al. 2012), to abnormal inflammatory mechanisms (Ocskó et al. 2018) and even cell cycle dysregulations (Blue et al. 2018). To this day, no inhibitor or probe has been identified, even if in silico screening assays and computational docking simulations have been realized (Chikan and Vipperla 2015). This subfamily is interesting since its members do not share similar biological outcomes (Filion et al. 2006) and their depletion in mice has dramatically opposite consequences between Kaiso (viable and fertile) (Prokhortchouk et al. 2006) and the two other members (cell damage and genomic instability) (Marchal et al. 2018; Roussel-Gervais et al. 2017).

As illustrated, chemical probes for MBPs are still at an early stage with few active molecules identified to date. Mostly they lack specificity and are not potent enough for cellular or in vivo assessments. Nevertheless, these proteins remain an interesting chemical target since they could allow studying complex epigenetic mechanisms and widen the spectra of therapeutic approaches for human pathologies.

17.4.2 Bifunctional Inhibitors Involving DNMTi

Another strategy to address the lack of efficacy and selectivity of DNMTi is the use of bifunctional compounds. This promising strategy consists in designing bifunctional inhibitors formed by the connection of at least two different pharmacophores in a single drug (Tomaselli et al. 2020). Such compounds have many advantages compared to the use of drug combination. First, selectivity might be improved, pharmacological properties can be modulated and toxicity can be lowered. Besides, both inhibitors are simultaneously present in the same tissue that can help to reach greater efficacy against advanced-stage disease. Bifunctional compounds also limit resistance mechanisms (de Lera and Ganesan 2016).

Due to the lack of reported potent DNMTi, it is only recently that a few bifunctional inhibitors, including a DNMTi moiety, have been reported. Synthesis and biological evaluation of four different DNMT/HDAC and DNMT/G9a bifunctional inhibitors have recently been described (Fig. 17.7). Because of the co-location of these two targets in the same protein complex (Hervouet et al. 2018), their involvement in the same diseases (Esteller 2008), and the synergic effect of drug combinations, targeting these two enzymes exhibit good potency. The biggest challenge in designing hybrid inhibitors is to maintain or improve activity against both enzymes.

Histone deacetylases (HDAC) are a family of 18 enzymes, divided in 4 classes, which catalyze deacetylation of histone-tail *N*-terminal lysines leading to chromatin compaction (Ruijter et al. 2003). In some diseases, particularly in cancer, histone hypoacetylation combined with DNA hypermethylation, cause transcriptional repression (Kondo 2009). HDAC inhibitors have all the same general structure characterized by a Zinc Binding Group (ZBG) that chelate the zincite ion in the catalytic pocket of the enzyme, a hydrophobic linker and a cap group, which interacts at the entrance of the catalytic pocket. To date, four HDACi were approved: Vorinostat, Belinostat, Panobinostat, and Chidamide, where ZBG is a hydroxamic acid moiety (Ho et al. 2020).

In 2017, the first-in-class bifunctional DNMT/HDAC inhibitor was described (Yuan et al. 2017). Its design was based on NSC-319745 DNMTi pharmacophore functionalized with a hydroxamic acid moiety. NSC-319745 had been previously identified by virtual screening as a selective DNMT1 inhibitor with low activity

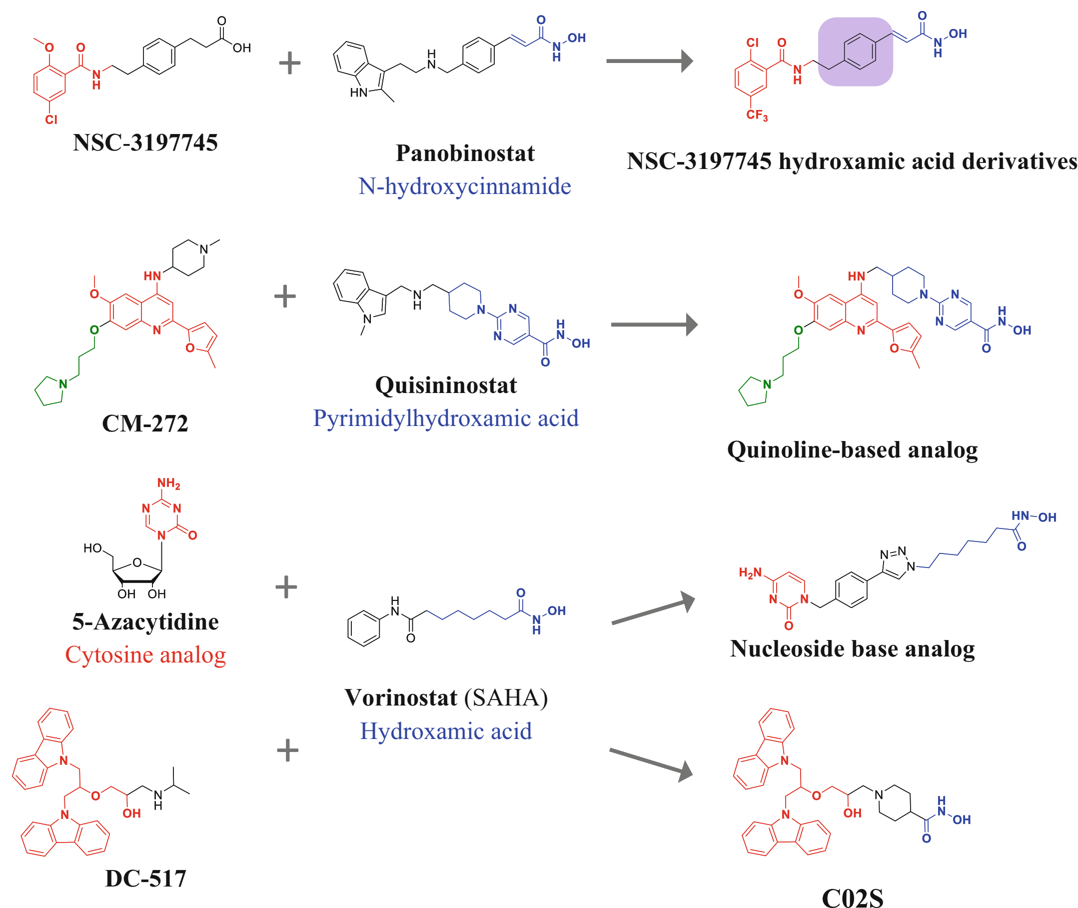


Fig. 17.7 Bifunctional epigenetic inhibitors containing DNMTi moieties

(34% inhibition at 100 μM on enzymatic assay) (Kuck et al. 2010b). Interestingly, despite its low efficacy, this DNMTi shares similarities with hydroxamate HDAC inhibitors. Changing carboxylic acid in hydroxamic acid, more precisely with an *N*-hydroxycinnamide moiety, and optimizing its structure allowed to improve the activity against target enzymes. The lead compound finally showed 70%, 23%, and 49% inhibition at 100 μM against DNMT1, 3A et 3B, respectively and IC_{50} at nanomolar range against two classes of HDAC in enzyme assays. Potent antiproliferative activity was measured in K562 and U937 human cancer cell lines ($IC_{50} = 2.9$ μM and 1.1 μM, respectively) as well as an increased H3K9 and H4K8 acetylation level. This bifunctional inhibitor was shown to initiate *P16* methylation inhibition and its re-expression and to induce apoptosis in leukemia model cancer cells U937.

Another HDAC/DNMT bifunctional inhibitor was described by Rabal's team (Rabal et al. 2021). They first worked on a series of quinoline-based analog as bifunctional DNMT/G9a inhibitor, where G9a is a lysine methyltransferase. Their hit compound exhibited IC_{50} of 8 nM against G9a and 382 nM against DNMT1 on purified enzymes (San José-Enériz et al. 2017) and it was shown to act as substrate competitive inhibitor of G9a and DNMTs (Rabal et al. 2018). They then designed a multitarget epigenetic inhibitor targeting HDAC, DNMT1

and optionally G9a. Based on structure design, docking and structure activity relationship SAR) analysis, they inserted a ZBG at the solvent exposed area of their inhibitor. They selected the pyrimidylhydroxamic acid moiety of quisininostat since its cap group mimics a quinoline base. Substitution of the quinoline moiety increased activity against DNMT and G9a and finally led to two potent compounds, one acting as a trifunctional G9a/HDAC/DNMT inhibitor and the second one as a bifunctional DNMT/HDAC inhibitor obtained by removing the lysine mimic moiety. Although they lost efficacy against DNMT1, compared with their previous hit compounds, these multifunctional compounds exhibited IC_{50} in nanomolar range against HDAC1, 2 and 3 and a potent antiproliferative activity on MM.1S multiple myeloma cell line. Due to their poor pharmacological property profiles, these compounds should be optimized to enable in vivo experiments.

By fragment-based rational drug design, Jiang's group worked on a series of hydroxamic acid derivatives of nucleoside bases. This work was inspired by cytosine and hydroxamic acid derivatives, FDA-approved DNMTi and HDACi, respectively. In both studies, both pharmacophores were linked through a triazole with different linker lengths. The first bifunctional inhibitor reported exhibited adenine as DNMT pharmacophore (Ren et al. 2019), then they completed the series varying the nucleobase (Sun et al. 2021a). The most potent compound, carrying a cytosine linked to the hydroxamic acid moiety with a linear six-carbon alkyl chain showed potent inhibitory activity against DNMT1 with IC_{50}= 6.39 μM and nanomolar activities against HDAC1 and HDAC6, comparable to Vorinostat. Western Blot in U937 lymphoma cell lines treated with this DNMT-HDAC inhibitor, showed an increase of H3K9 and H4K8 acetylation levels and a weak decrease of methylation levels in *P16* promoter region that led to an increase of *P16* expression and antiproliferative activity in a dose-dependent manner. Even though these bifunctional inhibitors displayed a good potency, they need further investigations for cancer therapy, especially regarding quantification of DNA methylation inhibition in cells, pharmacological properties as well as comparison with DNMTi and HDACi combinations.

The most potent bifunctional inhibitor known to date was described by Yuan et al. (2019). In this study, DC-517, a selective DNMT1 inhibitor described by Luo's group composed of two carbazole groups and a side chain (Chen et al. 2014), was selected as DNMTi. Previous docking experiments on this molecule suggested the addition of the hydroxamic acid moiety at the end of the side chain without removing the hydroxyl group, which interacts by hydrogen bonding in the DNMT1 catalytic pocket. Cell viability assays in breast cancer and epithelial cell lines MCF-7, A549, and MDA-MB-231, showed that the (*S*)-enantiomer C02S has the strongest antiproliferative effect, which was higher than compounds SAHA and SGI-1027. C02S achieved DNMT inhibition in the micromolar range against DNMT1. Its activities against DNMT3A and 3B were improved compared to DC-517. However, it induced significant DNMT1 degradation only after 24 h treatment in MCF-7 tumor cells. Moderate HDAC inhibition activity was also reported by hyperacetylation of H3K9 and H4K8 in MCF-7 cell line. C02S also inhibited *P16* and *P21* promoter methylation together with the re-expression of these TSGs. By flow cytometry assay and in PI-FITC/annexin assay, it was shown that C02S induced G0/G1 cell cycle arrest that led to apoptosis. It also possessed antiangiogenetic activity, confirmed by the inhibition of HUVEC tubule formation in a dose-dependent-manner, and inhibited tumor cells migration and invasion of MDA-MB-231 cells. In in vivo model, using mice xenograft 4T1 breast tumor models, C02S significantly reduced tumor volume and mass at 5 and 15 mg/kg/day treatment, that was comparable with the positive control group treated with a combination of SAHA and decitabine. Despite low HDAC inhibition activity (Yuan et al. 2019), C02S is the only

DNMTi-containing bifunctional inhibitor evaluated in vivo, which exhibited a significant effect on solid tumor.

To conclude, development of hybrid compounds that can inhibit simultaneously DNMT and HDAC is a very recent strategy which has a real potential in anti-tumor activity.

17.5 Limits and Hopes of DNMTi Applications and New Perspectives

The above examples illustrate the major role of DNMTs in the normal and abnormal functioning of cells and how their inhibition by DNMTi can change phenotypes, revealing the high potential of DNMTi both for therapeutic strategies and for the bioengineering of organisms. It also highlights how, depending on the context, the effects can be inversed. This is related to the dynamics of DNA methylation and its role in controlling gene expression. As illustrated by the above examples, DNMTi are just starting to be applied to diseases other than cancer and these strategies seem very promising. An increasing number of clinical trials are starting to explore epigenetic reprogramming by DNMTi in combination with other treatments, in particular, in solid tumors. Time will show if these studies can validate the interest of epidrugs beyond hematological cancers, the only diseases they are approved for.

Noteworthy, the only FDA-approved DNMTi are 5aza and 5azadC, suicide substrates that are incorporated into DNA and form an irreversible covalent complex with the enzyme. This triggers a subsequent DNMT degradation by the proteasome, inducing a potent inhibition of DNA methylation. It is not clear whether a non-covalent inhibitor can induce this level of demethylation (Erdmann et al. 2015). To obtain the same potency as covalent nucleoside inhibitors, non-nucleoside inhibitors could be an alternative, but they would need to be very specific for the DNMTs and today this specificity is lacking. Among non-nucleoside DNMTi, GSK3484862 showed the most promising results with high in vitro and in cell potency. Nevertheless, it is urgent to find novel compounds that are very potent inhibitors of DNMTs and bind strongly to the enzyme, in order to induce a strong inhibition in cells and eventually a down-regulation of the DNMTs.

Another feature of 5aza, 5azadC and their analogs is the fact that they are globally incorporated into the DNA instead of deoxycytosine and thus they are not specific for CpGs or certain genomic regions. This can hinder their use by inducing non-desired secondary effects. The use of low doses of the drugs has diminished these effects, however, repeated cycles are necessary for the epigenetic reprogramming, increasing the probability of their appearance. Non-nucleoside inhibitors have the advantage of not needing incorporation into DNA and thus potentially diminishing the side effects. In conclusions, it is important to pursue the search for new non-nucleoside inhibitors of DNA methylation that are potent, specific for the CpGs and of promoters silencing key genes in pathologies. Several strategies are being explored for the next generation of inhibitors, as allosteric inhibitors, protein-protein ligands and dual inhibitors (Erdmann et al. 2016).

An interesting alternative is to target the proteins that recognize methylated DNA and trigger the signaling pathways that silence DNA (DNA methylation reader proteins). While histone reader proteins are a well-explored target, the chemical targeting of the MBP is still in its infancy.

Finally, hybrid compounds bearing both a DNMTi moiety and another inhibitor moiety is also promising. Interestingly, targeting not a single enzyme but cancer-specific protein complexes can enable cancer cell selectivity. In addition, the merging of two drugs in the same molecules can facilitate the pharmacology compared to drug combinations. Challenges remain in the design of such compounds to maintain both scaffold activity and efficacy.

References

Achour M, Jacq X, Rondé P, Alhosin M, Charlot C, Chataigneau T et al (2008) The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. Oncogene 27(15):2187–2197. https://doi.org/10.1038/sj.onc.1210855

Ahuja N, Easwaran H, Baylin SB (2014) Harnessing the potential of epigenetic therapy to target solid tumors. J Clin Invest 124(1):56–63. https://doi.org/10.1172/jci69736

Ahuja N, Sharma AR, Baylin SB (2016) Epigenetic therapeutics: a new weapon in the war against cancer. Annu Rev Med 67(1):73–89. https://doi.org/10.1146/annurev-med-111314-035900

Aizawa SY, Yutaka (2015) Valproate administration to mice increases hippocampal p21 expression by altering genomic DNA methylation. Neuroreport 26(15): 915–920. https://doi.org/10.1097/WNR.0000000000000448

Akimoto K, Katakami H, Kim H-J, Ogawa E, Sano CM, Wada Y et al (2007) Epigenetic inheritance in rice plants. Ann Bot 100(2):205–217. https://doi.org/10.1093/aob/mcm110

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet 23(2):185–188

Asgatay S, Champion C, Marloie G, Drujon T, Senamaud-Beaufort C, Ceccaldi A et al (2014) Synthesis and evaluation of analogues of *N*-phthaloyl-L-tryptophan (RG108) as inhibitors of DNA methyltransferase 1. J Med Chem 57(2):421–434. https://doi.org/10.1021/jm401419p

Aydin B, Hocaoglu N, Gidener S (2012) Could decitabine treatment impair memory functions in humans? Med Hypotheses 79(5):639–641. https://doi.org/10.1016/j.mehy.2012.07.042

Azad N, Zahnow CA, Rudin CM, Baylin SB (2013) The future of epigenetic therapy in solid tumours—lessons from the past. Nat Rev Clin Oncol 10(5):256–266. https://doi.org/10.1038/nrclinonc.2013.42

Azechi T, Sato F, Sudo R, Wachi H (2014) 5-aza-2-′-Deoxycytidine, a DNA methyltransferase inhibitor, facilitates the inorganic phosphorus-induced mineralization of vascular smooth muscle cells. J Atheroscler Thromb 21(5):463–476. https://doi.org/10.5551/jat.20818

Azevedo Portilho N, Saini D, Hossain I, Sirois J, Moraes C, Pastor WA (2021) The DNMT1 inhibitor GSK-3484862 mediates global demethylation in murine embryonic stem cells. Epigenetics Chromatin 14(1):56. https://doi.org/10.1186/s13072-021-00429-0

Barbier E, Tapocik JD, Juergens N, Pitcairn C, Borich A, Schank JR et al (2015) DNA methylation in the medial prefrontal cortex regulates alcohol-induced behavior and plasticity. J Neurosci 35(15):6153–6164. https://doi.org/10.1523/jneurosci.4571-14.2015

Baud MG, Leiser T, Haus P, Samlal S, Wong AC, Wood RJ et al (2012) Defining the mechanism of action and enzymatic selectivity of psammaplin A against its epigenetic targets. J Med Chem 55(4):1731–1750. https://doi.org/10.1021/jm2016182

Baymaz HI, Fournier A, Laget S, Ji Z, Jansen PWTC, Smits AH et al (2014) MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. Proteomics 14(19): 2179–2189. https://doi.org/10.1002/pmic.201400013

Beau J, Mahid N, Burda WN, Harrington L, Shaw LN, Mutka T et al (2012) Epigenetic tailoring for the production of anti-infective cytosporones from the *marine fungus Leucostoma persoonii*. Mar Drugs 10(4): 762–774. https://doi.org/10.3390/md10040762

Benton M, Johnstone A, Eccles D, Harmon B, Hayes M, Lea R et al (2015) An analysis of DNA methylation in human adipose tissue reveals differential modification of obesity genes before and after gastric bypass and weight loss. Genome Biol 16(1):8

Berni Canani R, Paparo L, Nocerino R, Cosenza L, Pezzella V, Di Costanzo M et al (2015) Differences in DNA methylation profile of Th1 and Th2 cytokine genes are associated with tolerance acquisition in children with IgE-mediated cow's milk allergy. Clin Epigenetics 7(1):38. https://doi.org/10.1186/s13148-015-0070-8

Biergans SD, Claudianos C, Reinhard J, Galizia CG (2016) DNA methylation adjusts the specificity of memories depending on the learning context and promotes relearning in honeybees. Front Mol Neurosci 9:82. https://doi.org/10.3389/fnmol.2016.00082

Blazkova J, Trejbalova K, Gondois-Rey F, Halfon P, Philibert P, Guiguen A et al (2009) CpG methylation controls reactivation of HIV from latency. PLoS Pathog 5(8):e1000554. https://doi.org/10.1371/journal.ppat.1000554

Blue EE, Yu CE, Thornton TA, Chapman NH, Kernfeld E, Jiang N et al (2018) Variants regulating ZBTB4 are associated with age-at-onset of Alzheimer's disease. Genes Brain Behav 17(6):e12429. https://doi.org/10.1111/gbb.12429

Boku S, Toda H, Nakagawa S, Kato A, Inoue T, Koyama T et al (2015) Neonatal maternal separation alters the capacity of adult neural precursor cells to differentiate into neurons via methylation of retinoic acid receptor gene promoter. Biol Psychiatry 77(4):335–344. https://doi.org/10.1016/j.biopsych.2014.07.008

Bon C, Halby L, Arimondo PB (2020) Bisubstrate inhibitors: the promise of a selective and potent chemical inhibition of epigenetic "writers". Epigenomics 12 (17):1479–1482. https://doi.org/10.2217/epi-2020-0203

Bönsch D, Lenz B, Fiszer R, Frieling H, Kornhuber J, Bleich S (2006) Lowered DNA methyltransferase (DNMT-3b) mRNA expression is associated with genomic DNA hypermethylation in patients with chronic alcoholism. J Neural Transm 113(9): 1299–1304. https://doi.org/10.1007/s00702-005-0413-2

Booij L, Wang D, Lévesque ML, Tremblay RE, Szyf M (2013) Looking beyond the DNA sequence: the relevance of DNA methylation processes for the stress–diathesis model of depression. Philos Trans R Soc Lond B Biol Sci 368(1615):20120251. https://doi.org/10.1098/rstb.2012.0251

Boothman DA, Briggle TV, Greer S (1989) Exploitation of elevated pyrimidine deaminating enzymes for selective chemotherapy. Pharmacol Ther 42(1):65–88

Bouchard L, Rabasa-Lhoret R, Faraj M, Lavoie M-È, Mill J, Pérusse L et al (2010) Differential epigenomic and transcriptomic responses in subcutaneous adipose tissue between low and high responders to caloric restriction. Am J Clin Nutr 91(2):309–320. https://doi.org/10.3945/ajcn.2009.28085

Bouchut A, Rotili D, Pierrot C, Valente S, Lafitte S, Schultz J et al (2019) Identification of novel quinazoline derivatives as potent antiplasmodial agents. Eur J Med Chem 161:277–291. https://doi.org/10.1016/j.ejmech.2018.10.041

Brand S, Kesper DA, Teich R, Kilic-Niebergall E, Pinkenburg O, Bothur E et al (2012) DNA methylation of TH1/TH2 cytokine genes affects sensitization and progress of experimental asthma. J Allergy Clin Immunol 129(6):1602–1610. https://doi.org/10.1016/j.jaci.2011.12.963

Brøns C, Jacobsen S, Nilsson E, Rönn T, Jensen CB, Storgaard H et al (2010) Deoxyribonucleic acid methylation and gene expression of PPARGC1A in human muscle is influenced by high-fat overfeeding in a Birth-weight-dependent manner. J Clin Endocrinol Metab 95(6):3048–3056. https://doi.org/10.1210/jc.2009-2413

Brueckner B, Garcia Boy R, Siedlecki P, Musch T, Kliem HC, Zielenkiewicz P et al (2005) Epigenetic reactivation of tumor suppressor genes by a novel small-molecule inhibitor of human DNA methyltransferases. Cancer Res 65(14):6305–6311. https://doi.org/10.1158/0008-5472.can-04-2957

Brueckner B, Rius M, Markelova MR, Fichtner I, Hals PA, Sandvold ML et al (2010) Delivery of 5-azacytidine to human cancer cells by elaidic acid esterification increases therapeutic drug efficacy. Mol Cancer Ther 9(5):1256–1264. https://doi.org/10.1158/1535-7163.MCT-09-1202

Buck-Koehntop Bethany A, Stanfield Robyn L, Ekiert Damian C, Martinez-Yamout Maria A, Dyson HJ, Wilson Ian A et al (2012) Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. Proc Natl Acad Sci 109(38):15229–15234. https://doi.org/10.1073/pnas.1213726109

Byun HM, Choi SH, Laird PW, Trinh B, Siddiqui MA, Marquez VE et al (2008) 2′-Deoxy-N4-[2-(4-nitrophenyl)ethoxycarbonyl]-5-azacytidine: a novel inhibitor of DNA methyltransferase that requires activation by human carboxylesterase 1. Cancer Lett 266(2):238–248. https://doi.org/10.1016/j.canlet.2008.02.069

Cameron EE, Bachman KE, Myöhänen S, Herman JG, Baylin SB (1999) Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. Nat Genet 21(1):103–107. https://doi.org/10.1038/5047

Candelaria M, de la Cruz-Hernandez E, Taja-Chayeb L, Perez-Cardenas E, Trejo-Becerril C, Gonzalez-Fierro A et al (2012) DNA methylation-independent reversion of gemcitabine resistance by hydralazine in cervical cancer cells. PLoS One 7(3):e29181. https://doi.org/10.1371/journal.pone.0029181

Carnicella S, Kharazia V, Jeanblanc J, Janak PH, Ron D (2008) GDNF is a fast-acting potent inhibitor of alcohol consumption and relapse. Proc Natl Acad Sci U S A 105(23):8114–8119. https://doi.org/10.1073/pnas.0711755105

Castellano S, Kuck D, Sala M, Novellino E, Lyko F, Sbardella G (2008) Constrained analogues of procaine as novel small molecule inhibitors of DNA methyltransferase-1. J Med Chem 51(7):2321–2325. https://doi.org/10.1021/jm7015705

Castellano S, Kuck D, Viviano M, Yoo J, Lopez-Vallejo F, Conti P et al (2011) Synthesis and biochemical evaluation of Δ2-isoxazoline derivatives as DNA methyltransferase 1 inhibitors. J Med Chem 54(21):7663–7677

Ceccaldi A, Rajavelu A, Champion C, Rampon C, Jurkowska R, Jankevicius G et al (2011) C5-DNA methyltransferase inhibitors: from screening to effects on zebrafish embryo development. ChemBioChem 12(9):1337–1345

Ceccaldi A, Rajavelu A, Ragozin S, Sénamaud-Beaufort C, Bashtrykov P, Testa N et al (2013) Identification of novel inhibitors of DNA methylation by screening of a chemical library. ACS Chem Biol 8(3):543–548

Champion C, Guianvarc'h D, Sénamaud-Beaufort C, Jurkowska RZ, Jeltsch A, Ponger L et al (2010) Mechanistic insights on the inhibition of C5 DNA methyltransferases by zebularine. PLoS One 5(8):e12388. https://doi.org/10.1371/journal.pone.0012388

Chan KK, Giannini DD, Staroscik JA, Sadee W (1979) 5-Azacytidine hydrolysis kinetics measured by high-pressure liquid chromatography and 13C-NMR spectroscopy. J Pharm Sci 68(7):807–812. https://doi.org/10.1002/jps.2600680705

Chandel N, Ayasolla KS, Lan X, Sultana-Syed M, Chawla A, Lederman R et al (2015) Epigenetic modulation of human podocyte vitamin D receptor in HIV milieu. J Mol Biol 427(20):3201–3215. https://doi.org/10.1016/j.jmb.2015.07.011

Chang L, Campbell J, Raji IO, Guduru SKR, Kandel P, Nguyen M et al (2021) Discovery of small molecules targeting the tandem tudor domain of the epigenetic factor UHRF1 using fragment-based ligand discovery. Sci Rep 11(1):1121. https://doi.org/10.1038/s41598-020-80588-4

Chaturvedi P, Tyagi SC (2014) Epigenetic mechanisms underlying cardiac degeneration and regeneration. Int

J Cardiol 173(1):1–11. https://doi.org/10.1016/j.ijcard.2014.02.008
Chen Y, Sharma RP, Costa RH, Costa E, Grayson DR (2002) On the epigenetic regulation of the human reelin promoter. Nucleic Acids Res 30(13): 2930–2939. https://doi.org/10.1093/nar/gkf401
Chen C, Pan D, Deng AM, Huang F, Sun B-L, Yang R-G (2013) DNA methyltransferases 1 and 3B are required for hepatitis C virus infection in cell culture. Virology 441(1):57–65. https://doi.org/10.1016/j.virol.2013.03.005
Chen S, Wang Y, Zhou W, Li S, Peng J, Shi Z et al (2014) Identifying novel selective non-nucleoside DNA methyltransferase 1 inhibitors through docking-based virtual screening. J Med Chem 57(21):9028–9041. https://doi.org/10.1021/jm501134e
Chen X, Peng X, Wang L, Fu X, Zhou JX, Zhu B et al (2017) Association of RASgrf1 methylation with epileptic seizures. Oncotarget 8(28):46286–46297. https://doi.org/10.18632/oncotarget.18000
Chestnut BA, Chang Q, Price A, Lesuisse C, Wong M, Martin LJ (2011) Epigenetic regulation of motor neuron cell death through DNA methylation. J Neurosci 31(46):16619–16636. https://doi.org/10.1523/jneurosci.1639-11.2011
Chikan NA, Vipperla B (2015) KAISO inhibition: an atomic insight. J Biomol Struct Dyn 33(8): 1794–1804. https://doi.org/10.1080/07391102.2014.974072
Cho G-W, Noh M-Y, Kim HY, Koh S-H, Kim K-S, Kim SH (2010) Bone marrow-derived stromal cells from amyotrophic lateral sclerosis patients have diminished stem cell capacity. Stem Cells Dev 19(7):1035–1042. https://doi.org/10.1089/scd.2009.0453
Choudhry H, Zamzami MA, Omran Z, Wu W, Mousli M, Bronner C et al (2018) Targeting microRNA/UHRF1 pathways as a novel strategy for cancer therapy. Oncol Lett 15(1):3–10. https://doi.org/10.3892/ol.2017.7290
Chuang JC, Yoo CB, Kwan JM, Li TWH, Liang G, Yang AS et al (2005) Comparison of biological effects of non-nucleoside DNA methylation inhibitors versus 5-aza-2′-deoxycytidine. Mol Cancer Ther 4(10): 1515–1520. https://doi.org/10.1158/1535-7163.mct-05-0172
Chuang JC, Warner SL, Vollmer D, Vankayalapati H, Redkar S, Bearss DJ et al (2010) S110, a 5-Aza-2-′-deoxycytidine-containing dinucleotide, is an effective DNA methylation inhibitor *in vivo* and can reduce tumor growth. Mol Cancer Ther 9(5): 1443–1450. https://doi.org/10.1158/1535-7163.mct-09-1048
Cihak A (1974) Biological effects of 5-azacytidine in eukaryotes. Oncology 30(5):405–422
Cleveland DW, Rothstein JD (2001) From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS. Nat Rev Neurosci 2(11):806–819
Coppieters N, Dragunow M (2011) Epigenetics in Alzheimer's disease: a focus on DNA modifications. Curr Pharm Des 17(31):3398–3412. https://doi.org/10.2174/138161211798072544#sthash.ZNP9Kotu.dpuf
Coppieters N, Dieriks BV, Lill C, Faull RLM, Curtis MA, Dragunow M (2014) Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. Neurobiol Aging 35(6):1334–1344. https://doi.org/10.1016/j.neurobiolaging.2013.11.031
Cornacchia E, Golbus J, Maybaum J, Strahler J, Hanash S, Richardson B (1988) Hydralazine and procainamide inhibit T cell DNA methylation and induce autoreactivity. J Immunol 140(7):2197–2200
Cortijo S, Wardenaar R, Colomé-Tatché M, Gilly A, Etcheverry M, Labadie K et al (2014) Mapping the epigenetic basis of complex traits. Science 343(6175): 1145–1148. https://doi.org/10.1126/science.1248127
Daniel Juliet M, Reynolds AB (1999) The catenin p120ctn interacts with kaiso, a novel BTB/POZ domain zinc finger transcription factor. Mol Cell Biol 19(5): 3614–3623. https://doi.org/10.1128/MCB.19.5.3614
Datta J, Ghoshal K, Denny WA, Gamage SA, Brooke DG, Phiasivongsa P et al (2009) A new class of quinoline-based DNA hypomethylating agents reactivates tumor suppressor genes by blocking DNA methyltransferase 1 activity and inducing its degradation. Cancer Res 69(10):4277–4285. https://doi.org/10.1158/0008-5472.can-08-3669
Day JJ, Sweatt JD (2010) DNA methylation and memory formation. Nat Neurosci 13(11):1319–1323
D'Costa ZJ, Jolly C, Androphy EJ, Mercer A, Matthews CM, Hibma MH (2012) Transcriptional repression of E-cadherin by human papillomavirus type 16 E6. PLoS One 7(11):e48954. https://doi.org/10.1371/journal.pone.0048954
De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L et al (2014) Alzheimer's disease: early alterations in brain DNA methylation at *ANK1*, *BIN1*, *RHBDF2* and other loci. Nat Neurosci 17(9): 1156–1163. https://doi.org/10.1038/nn.3786. http://www.nature.com/neuro/journal/v17/n9/abs/nn.3786.html#supplementary-information
de la Cruz-Hernandez E, Perez-Cardenas E, Taja-Chayeb-L, Chavez-Blanco A, Trejo-Becerril C, Trujillo J et al (2011) DNA demethylating activity of hydralazine in cancer cell lines. Life Sci Med Res 2011:LSMR-30
de Lera AR, Ganesan A (2016) Epigenetic polypharmacology: from combination therapy to multitargeted drugs. Clin Epigenetics 8:105. https://doi.org/10.1186/s13148-016-0271-9
Desplats P, Spencer B, Coffee E, Patel P, Michael S, Patrick C et al (2011) α-synuclein sequesters dnmt1 from the nucleus: a novel mechanism for epigenetic alterations in Lewy body diseases. J Biol Chem 286(11):9031–9037. https://doi.org/10.1074/jbc.C110.212589
Dobbs KB, Rodriguez M, Sudano MJ, Ortega MS, Hansen PJ (2013) Dynamics of DNA methylation during early development of the preimplantation bovine embryo. PLoS One 8(6):e66230. https://doi.org/10.1371/journal.pone.0066230

Donaldson NS, Pierre CC, Anstey MI, Robinson SC, Weerawardane SM, Daniel JM (2012) Kaiso represses the cell cycle gene cyclin D1 via sequence-specific and methyl-CpG-dependent mechanisms. PLoS One 7(11): e50398. https://doi.org/10.1371/journal.pone.0050398

Drury JL, Chung WO (2015) DNA methylation differentially regulates cytokine secretion in gingival epithelia in response to bacterial challenges. Pathog Dis 73(2): 1–6. https://doi.org/10.1093/femspd/ftu005

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

Duenas-Gonzalez A, Candelaria M, Perez-Plascencia C, Perez-Cardenas E, de la Cruz-Hernandez E, Herrera LA (2008) Valproic acid as epigenetic cancer drug: Preclinical, clinical and transcriptional effects on solid tumors. Cancer Treat Rev 34(3):206–222. https://doi.org/10.1016/j.ctrv.2007.11.003

Dueñas-Gonzalez A, Coronel J, Cetina L, González-Fierro A, Chavez-Blanco A, Taja-Chayeb L (2014) Hydralazine–valproate: a repositioned drug combination for the epigenetic therapy of cancer. Expert Opin Drug Metab Toxicol 10(10):1433–1444. https://doi.org/10.1517/17425255.2014.947263

Dunn J, Qiu H, Kim S, Jjingo D, Hoffman R, Kim CW et al (2014) Flow-dependent epigenetic DNA methylation regulates endothelial gene expression and atherosclerosis. J Clin Invest 124(7):3187–3199. https://doi.org/10.1172/jci74792

El-Ajouz S, Ray D, Allsopp RC, Evans RJ (2012) Molecular basis of selective antagonism of the P2X1 receptor for ATP by NF449 and suramin: contribution of basic amino acids in the cysteine-rich loop. Br J Pharmacol 165(2):390–400. https://doi.org/10.1111/j.1476-5381.2011.01534.x

Erdmann A, Halby L, Fahy J, Arimondo PB (2015) Targeting DNA methylation with small molecules: what's next? J Med Chem 58(6):2569–2583. https://doi.org/10.1021/jm500843d

Erdmann A, Guianvarc'h D, Arimondo PB (2016) Structure-guided optimization of DNA methyltransferase inhibitors. In: Medina-Franco J (ed) Epi-informatics, discovery and development of small molecule epigenetic drugs and probes, 1st ed. Elsevier

Ernst C, Deleva V, Deng X, Sequeira A, Pomarenski A, Klempan T et al (2009) Alternative splicing, methylation state, and expression profile of tropomyosin-related kinase B in the frontal cortex of suicide completers. Arch Gen Psychiatry 66(1):22–32. https://doi.org/10.1001/archpsyc.66.1.22

Esteller M (2008) Epigenetics in cancer. N Engl J Med 358(11):1148–1159. https://doi.org/10.1056/NEJMra072067

Evans JS, Hanka LJ (1968) The *in vivo* activity of combinations of 5-azacytidine and cytidine on leukemia L-1210. Experientia 24(9):922–923

Fagan RL, Cryderman DE, Kopelovich L, Wallrath LL, Brenner C (2013) Laccaic acid A is a direct, DNA-competitive inhibitor of DNA methyltransferase 1. J Biol Chem 288(33):23858–23867

Fahy J, Jeltsch A, Arimondo PB (2012) DNA methyltransferase inhibitors in cancer: a chemical and therapeutic patent overview and selected clinical studies. Expert Opin Ther Pat 22(12):1427–1442. https://doi.org/10.1517/13543776.2012.729579

Falandry C, Fourel G, Galy V, Ristriani T, Horard B, Bensimon E et al (2010) CLLD8/KMT1F is a lysine methyltransferase that is important for chromosome segregation. J Biol Chem 285(26):20234–20241. https://doi.org/10.1074/jbc.M109.052399

Fandy TE, Herman JG, Kerns P, Jiemjit A, Sugar EA, Choi S-H et al (2009) Early epigenetic changes and DNA damage do not predict clinical response in an overlapping schedule of 5-azacytidine and entinostat in patients with myeloid malignancies. Blood 114(13): 2764–2773. https://doi.org/10.1182/blood-2009-02-203547

Fang MZ, Wang Y, Ai N, Hou Z, Sun Y, Lu H et al (2003) Tea polyphenol (-)-epigallocatechin-3-gallate inhibits DNA methyltransferase and reactivates methylation-silenced genes in cancer cell lines. Cancer Res 63(22):7563–7570

Fang MZ, Chen D, Sun Y, Jin Z, Christman JK, Yang CS (2005) Reversal of hypermethylation and reactivation of p16INK4a, RARβ, and MGMT genes by genistein and other isoflavones from soy. Clin Cancer Res 11(19):7033–7041. https://doi.org/10.1158/1078-0432.ccr-05-0406

Felle M, Joppien S, Németh A, Diermeier S, Thalhammer V, Dobner T et al (2011) The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1. Nucleic Acids Res 39(19):8355–8365. https://doi.org/10.1093/nar/gkr528

Feng J, Zhou Y, Campbell SL, Le T, Li E, Sweatt JD et al (2010) Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. Nat Neurosci 13(4):423–430. http://www.nature.com/neuro/journal/v13/n4/suppinfo/nn.2514_S1.html

Fernandez G, Zeichner S (2010) Cell line-dependent variability in HIV activation employing DNMT inhibitors. Virol J 7(1):266. https://doi.org/10.1186/1743-422X-7-266

Filion GJP, 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. Mol Cell Biol 26(1): 169–181. https://doi.org/10.1128/MCB.26.1.169-181.2006

Finnegan EJ, Genger RK, Peacock WJ, Dennis ES (1998) DNA methylation in plants. Annu Rev Plant Physiol Plant Mol Biol 49(1):223–247. https://doi.org/10.1146/annurev.arplant.49.1.223

Flotho C, Claus R, Batz C, Schneider M, Sandrock I, Ihde S et al (2009) The DNA methyltransferase inhibitors azacitidine, decitabine and zebularine exert differential effects on cancer gene expression in acute myeloid leukemia cells. Leukemia 23(6):1019–1028. http://www.nature.com/leu/journal/v23/n6/suppinfo/leu2008397s1.html

Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML et al (2005) Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A 102(30):10604–10609. https://doi.org/10.1073/pnas.0500398102

Friso S, Pizzolo F, Choi S-W, Guarini P, Castagna A, Ravagnani V et al (2008) Epigenetic control of 11 beta-hydroxysteroid dehydrogenase 2 gene promoter is related to human hypertension. Atherosclerosis 199(2):323–327. https://doi.org/10.1016/j.atherosclerosis.2007.11.029

Fujimori A, Matsuda Y, Takemoto Y, Hashimoto Y, Kubo E, Araki R et al (1998) Cloning and mapping of Np95 gene which encodes a novel nuclear protein associated with cell proliferation. Mamm Genome 9(12):1032–1035. https://doi.org/10.1007/s003359900920

Fujita N, Takebayashi S, Okumura K, Kudo S, Chiba T, Saya H et al (1999) Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms. Mol Cell Biol 19(9):6415–6426. https://doi.org/10.1128/MCB.19.9.6415

Gamage SA, Brooke DG, Redkar S, Datta J, Jacob ST, Denny WA (2013) Structure-activity relationships for 4-anilinoquinoline derivatives as inhibitors of the DNA methyltransferase enzyme DNMT1. Bioorg Med Chem 21(11):3147–3153. https://doi.org/10.1016/j.bmc.2013.03.033

Gervin K, Vigeland MD, Mattingsdal M, Hammerø M, Nygård H, Olsen AO et al (2012) DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes. PLoS Genet 8(1):e1002454. https://doi.org/10.1371/journal.pgen.1002454

Gigek CO, Chen ES, Smith MAC (2016) Methyl-CpG-binding protein (MBD) family: epigenomic read-outs functions and roles in tumorigenesis and psychiatric diseases. J Cell Biochem 117(1):29–38. https://doi.org/10.1002/jcb.25281

Gnanapragasam Merlin N, Scarsdale JN, Amaya Maria L, Webb Heather D, Desai Megha A, Walavalkar Ninad M et al (2011) p66α–MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2–NuRD complex. Proc Natl Acad Sci U S A 108(18):7487–7492. https://doi.org/10.1073/pnas.1015341108

Graça I, Sousa EJ, Baptista T, Almeida M, Ramalho-Carvalho J, Palmeira C et al (2014a) Anti-tumoral effect of the non-nucleoside DNMT inhibitor RG108 in human prostate cancer cells. Curr Pharm Des 20(11):1803–1811

Graça I, Sousa EJ, Costa-Pinheiro P, Vieira FQ, Torres-Ferreira J, Martins MG et al (2014b) Anti-neoplastic properties of hydralazine in prostate cancer. Oncotarget 5(15):5950–5964

Grayson DR, Jia X, Chen Y, Sharma RP, Mitchell CP, Guidotti A et al (2005) Reelin promoter hypermethylation in schizophrenia. Proc Natl Acad Sci U S A 102(26):9341–9346. https://doi.org/10.1073/pnas.0503736102

Gros C, Fleury L, Nahoum V, Faux C, Valente S, Labella D et al (2015) New insights on the mechanism of quinoline-based DNA methyltransferase inhibitors. J Biol Chem 290(10):6293–6302. https://doi.org/10.1074/jbc.M114.594671

Guay S-P, Legare C, Houde A-A, Mathieu P, Bosse Y, Bouchard L (2014) Acetylsalicylic acid, aging and coronary artery disease are associated with ABCA1 DNA methylation in men. Clin Epigenetics 6(1):14

Guidotti A, Auta J, Davis JM, DiGiorgi GV, Dwivedi Y, Grayson DR et al (2000) Decrease in reelin and glutamic acid decarboxylase67 (GAD67) expression in schizophrenia and bipolar disorder: a postmortem brain study. Arch Gen Psychiatry 57(11):1061–1069. https://doi.org/10.1001/archpsyc.57.11.1061

Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y et al (2014) The DNA methylation landscape of human early embryos. Nature 511(7511):606–610. https://doi.org/10.1038/nature13544. http://www.nature.com/nature/journal/v511/n7511/abs/nature13544.html#supplementary-information

Guy J, Gan J, Selfridge J, Cobb S, Bird A (2007) Reversal of neurological defects in a mouse model of Rett syndrome. Science 315(5815):1143–1147. https://doi.org/10.1126/science.1138389

Halby L, Champion C, Sénamaud-Beaufort C, Ajjan S, Drujon T, Rajavelu A et al (2012) Rapid synthesis of new DNMT inhibitors derivatives of procainamide. ChemBioChem 13(1):157–165

Halby L, Menon Y, Rilova E, Pechalrieu D, Masson V, Faux C et al (2017) Rational design of bisubstrate-type analogs as inhibitors of DNA methyltransferases in cancer cells. J Med Chem 60(11):4665–4679. https://doi.org/10.1021/acs.jmedchem.7b00176

Heinrich C, Nitta N, Flubacher A, Müller M, Fahrner A, Kirsch M et al (2006) Reelin deficiency and displacement of mature neurons, but not neurogenesis, underlie the formation of granule cell dispersion in the epileptic hippocampus. J Neurosci 26(17):4701–4713. https://doi.org/10.1523/jneurosci.5516-05.2006

Hendrich B, Bird A (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol 18(11):6538–6547. https://doi.org/10.1128/mcb.18.11.6538

Hendrich B, Hardeland U, Ng H-H, Jiricny J, Bird A (1999) The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature 401(6750):301–304. https://doi.org/10.1038/45843

Hervouet E, Peixoto P, Delage-Mourroux R, Boyer-Guittaut M, Cartron P-F (2018) Specific or not specific recruitment of DNMTs for DNA methylation, an epigenetic dilemma. Clin Epigenetics 10:17. https://doi.org/10.1186/s13148-018-0450-y

Ho TCS, Chan AHY, Ganesan A (2020) Thirty years of HDAC inhibitors: 2020 insight and hindsight. J Med Chem 63(21):12460–12484. https://doi.org/10.1021/acs.jmedchem.0c00830

Horrillo A, Pezzolla D, Fraga MF, Aguilera Y, Salguero-Aranda C, Tejedo JR et al (2013) Zebularine regulates early stages of mESC differentiation: effect on cardiac commitment. Cell Death Dis 4:e570. https://doi.org/10.1038/cddis.2013.88

Hossain M, Healey M, Lee C, Poh W, Yerram S, Patel K et al (2013) DNA-intercalators causing rapid re-expression of methylated and silenced genes in cancer cells. Oncotarget 4(2):298–309

Houliston RS, Lemak A, Iqbal A, Ivanochko D, Duan S, Kaustov L et al (2017) Conformational dynamics of the TTD-PHD histone reader module of the UHRF1 epigenetic regulator reveals multiple histone-binding states, allosteric regulation, and druggability. J Biol Chem 292(51):20947–20959. https://doi.org/10.1074/jbc.M117.799700

Huang S, Stillson NJ, Sandoval JE, Yung C, Reich NO (2021) A novel class of selective non-nucleoside inhibitors of human DNA methyltransferase 3A. Bioorg Med Chem Lett 40:127908. https://doi.org/10.1016/j.bmcl.2021.127908

Issa J-PJ, Kantarjian HM (2009) Targeting DNA methylation. Clin Cancer Res 15(12):3938–3946. https://doi.org/10.1158/1078-0432.ccr-08-2783

Jafari S, Hosseini MS, Hajian M, Forouzanfar M, Jafarpour F, Abedi P et al (2011) Improved *in vitro* development of cloned bovine embryos using *S*-adenosylhomocysteine, a non-toxic epigenetic modifying reagent. Mol Reprod Dev 78(8):576–584. https://doi.org/10.1002/mrd.21344

Jones PA, Taylor SM (1980) Cellular differentiation, cytidine analogs and DNA methylation. Cell 20(1):85–93. https://doi.org/10.1016/0092-8674(80)90237-8

Jones MH, Hamana N, Nezu J-i, Shimane M (2000) A novel family of bromodomain genes. Genomics 63(1):40–45. https://doi.org/10.1006/geno.1999.6071

Jowaed A, Schmitt I, Kaut O, Wüllner U (2010) Methylation regulates α-synuclein expression and is decreased in Parkinson's disease patients' brains. J Neurosci 30(18):6355–6359. https://doi.org/10.1523/jneurosci.6119-09.2010

Juergens RA, Wrangle J, Vendetti FP, Murphy SC, Zhao M, Coleman B et al (2011) Combination epigenetic therapy has efficacy in patients with refractory advanced non-small cell lung cancer. Cancer Discov 1(7):OF1–OF10. https://doi.org/10.1158/2159-8290.cd-11-0214

Kadowaki T, Yamauchi T, Kubota N, Hara K, Ueki K, Tobe K (2006) Adiponectin and adiponectin receptors in insulin resistance, diabetes, and the metabolic syndrome. J Clin Invest 116(7):1784–1792. https://doi.org/10.1172/jci29126

Kang H-J, Kim J-M, Lee J-Y, Kim S-Y, Bae K-Y, Kim S-W et al (2013) *BDNF* promoter methylation and suicidal behavior in depressive patients. J Affect Disord 151(2):679–685. https://doi.org/10.1016/j.jad.2013.08.001

Kao Y-H, Chen Y-C, Chung C-C, Lien G-S, Chen S-A, Kuo C-C et al (2013) Heart failure and angiotensin II modulate atrial *Pitx2c* promotor methylation. Clin Exp Pharmacol Physiol 40(6):379–384. https://doi.org/10.1111/1440-1681.12089

Karahoca M, Momparler RL (2013) Pharmacokinetic and pharmacodynamic analysis of 5-aza-2′-deoxycytidine (decitabine) in the design of its dose-schedule for cancer therapy. Clin Epigenetics 5(1):3. https://doi.org/10.1186/1868-7083-5-3

Kilgore JA, Du X, Melito L, Wei S, Wang C, Chin HG et al (2013) Identification of DNMT1 selective antagonists using a novel scintillation proximity assay. J Biol Chem 288(27):19673–19684. https://doi.org/10.1074/jbc.M112.443895

Kim J, Kim JY, Song KS, Lee YH, Seo JS, Jelinek J et al (2007) Epigenetic changes in estrogen receptor β gene in atherosclerotic cardiovascular tissues and in-vitro vascular senescence. Biochim Biophys Acta Mol Cell Res 1772(1):72–80. https://doi.org/10.1016/j.bbadis.2006.10.004

Kim M, Long TI, Arakawa K, Wang R, Yu MC, Laird PW (2010) DNA methylation as a biomarker for cardiovascular disease risk. PLoS One 5(3):e9692. https://doi.org/10.1371/journal.pone.0009692

Kim HJ, Kim JH, Chie EK, Da Young P, Kim IA, Kim IH (2012) DNMT (DNA methyltransferase) inhibitors radiosensitize human cancer cells by suppressing DNA repair activity. Radiat Oncol 7(1):39

Kim AY, Park YJ, Pan X, Shin KC, Kwak S-H, Bassas AF et al (2015) Obesity-induced DNA hypermethylation of the adiponectin gene mediates insulin resistance. Nat Commun 6:7585. https://doi.org/10.1038/ncomms8585

Kim MY, Na I, Kim JS, Son SH, Choi S, Lee SE et al (2019) Rational discovery of antimetastatic agents targeting the intrinsically disordered region of MBD2. Sci Adv 5(11):eaav9810. https://doi.org/10.1126/sciadv.aav9810

Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE et al (2004) One-way control of FWA imprinting in arabidopsis endosperm by DNA methylation. Science 303(5657):521–523. https://doi.org/10.1126/science.1089835

Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJJ, Koornneef M et al (2007) Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. Plant J 49(1):38–45. https://doi.org/10.1111/j.1365-313X.2006.02936.x

Kobow K, Jeske I, Hildebrandt M, Hauke J, Hahnen E, Buslei R et al (2009) Increased reelin promoter methylation is associated with granule cell dispersion in

human temporal lobe epilepsy. J Neuropathol Exp Neurol 68(4):356–364. https://doi.org/10.1097/NEN.0b013e31819ba737

Kobow K, Kaspi A, Harikrishnan KN, Kiese K, Ziemann M, Khurana I et al (2013) Deep sequencing reveals increased DNA methylation in chronic rat epilepsy. Acta Neuropathol 126(5):741–756. https://doi.org/10.1007/s00401-013-1168-8

Kohli A, Garcia M, Miller R, Maher C, Humblet O, Hammond S et al (2012) Secondhand smoke in combination with ambient air pollution exposure is associated with increasedx CpG methylation and decreased expression of IFN-γ in T effector cells and Foxp3 in T regulatory cells in children. Clin Epigenetics 4(1):17

Kondo Y (2009) Epigenetic cross-talk between DNA methylation and histone modifications in human cancers. Yonsei Med J 50(4):455–463

Kondo H, Miura T, Wada KC, Takeno K (2007) Induction of flowering by 5-azacytidine in some plant species: relationship between the stability of photoperiodically induced flowering and flower-inducing effect of DNA demethylation. Physiol Plant 131(3):462–469. https://doi.org/10.1111/j.1399-3054.2007.00965.x

Kovalchuk I, Walz P, Thomas J, Kovalchuk O (2013) The increased expression of proteins involved in proliferation, DNA repair and DNA methylation in spleen of mice exposed to E. coli O157:H7 lipopolysaccharide. Environ Mol Mutagen 54(6):421–428. https://doi.org/10.1002/em.21787

Kuck D, Caulfield T, Lyko F, Medina-Franco JL (2010a) Nanaomycin A selectively inhibits DNMT3B and reactivates silenced tumor suppressor genes in human cancer cells. Mol Cancer Ther 9(11):3015–3023. https://doi.org/10.1158/1535-7163.mct-10-0609

Kuck D, Singh N, Lyko F, Medina-Franco JL (2010b) Novel and selective DNA methyltransferase inhibitors: Docking-based virtual screening and experimental evaluation. Bioorg Med Chem 18(2):822–829. https://doi.org/10.1016/j.bmc.2009.11.050

Kumar A, Darcis G, Van Lint C, Herbein G (2015) Epigenetic control of HIV-1 post integration latency: implications for therapy. Clin Epigenetics 7(1):103. https://doi.org/10.1186/s13148-015-0137-6

Kundakovic M, Chen Y, Costa E, Grayson DR (2007) DNA methyltransferase inhibitors coordinately induce expression of the human reelin and glutamic acid decarboxylase 67 genes. Mol Pharmacol 71(3):644–653. https://doi.org/10.1124/mol.106.030635

Kwon N-H, Kim J-S, Lee J-Y, Oh M-J, Choi D-C (2008) DNA methylation and the expression of IL-4 and IFN-γ promoter genes in patients with bronchial asthma. J Clin Immunol 28(2):139–146. https://doi.org/10.1007/s10875-007-9148-1

LaPlant Q, Vialou V, Covington HE, Dumitriu D, Feng J, Warren BL et al (2010) Dnmt3a regulates emotional behavior and spine plasticity in the nucleus accumbens. Nat Neurosci 13(9):1137–1143. http://www.nature.com/neuro/journal/v13/n9/abs/nn.2619.html#supplementary-information

Lee BH, Yegnasubramanian S, Lin X, Nelson WG (2005) Procainamide is a specific inhibitor of DNA methyltransferase 1. J Biol Chem 280(49):40749–40756. https://doi.org/10.1074/jbc.M505593200

Levenson JM, Roth TL, Lubin FD, Miller CA, Huang I-C, Desai P et al (2006) Evidence that DNA (cytosine-5) methyltransferase regulates synaptic plasticity in the hippocampus. J Biol Chem 281(23):15763–15773. https://doi.org/10.1074/jbc.M511767200

Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F et al (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell 69(6):905–914. https://doi.org/10.1016/0092-8674(92)90610-o

Lewis DA, Hashimoto T, Volk DW (2005) Cortical inhibitory neurons and schizophrenia. Nat Rev Neurosci 6(4):312–324

Li Y, Tollefsbol TO (2010) Impact on DNA methylation in cancer prevention and therapy by bioactive dietary components. Curr Med Chem 17(20):2141–2151. https://doi.org/10.2174/092986710791299966

Li C, Terashita Y, Tokoro M, Wakayama S, Wakayama T (2011) Effect of DNA methyltransferase inhibitor, RG108, on in vitro development and ntES establishment rate in cloned mouse embryos. Reprod Fertil Dev 24:130. https://doi.org/10.1071/RDv24n1Ab36

Li W, Yu M, Luo S, Liu H, Gao Y, Wilson JX et al (2013) DNA methyltransferase mediates dose-dependent stimulation of neural stem cell proliferation by folate. J Nutr Biochem 24(7):1295–1301. https://doi.org/10.1016/j.jnutbio.2012.11.001

Li Y, Geng P, Jiang W, Wang Y, Yao J, Lin X et al (2014) Enhancement of radiosensitivity by 5-Aza-CdR through activation of G2/M checkpoint response and apoptosis in osteosarcoma cells. Tumor Biol 35(5):4831–4839. https://doi.org/10.1007/s13277-014-1634-5

Li H, Wang J, Wu C, Wang L, Chen Z-S, Cui W (2020a) The combination of disulfiram and copper for cancer treatment. Drug Discov Today 25(6):1099–1108. https://doi.org/10.1016/j.drudis.2020.04.003

Li L, Duan Q, Zeng Z, Zhao J, Lu J, Sun J et al (2020b) UHRF2 promotes intestinal tumorigenesis through stabilization of TCF4 mediated Wnt/β-catenin signaling. Int J Cancer 147(8):2239–2252. https://doi.org/10.1002/ijc.33036

Lin T-S, Lee H, Chen R-A, Ho M-L, Lin C-Y, Chen Y-H et al (2005) An association of DNMT3b protein expression with *P16INK4a* promoter hypermethylation in non-smoking female lung cancer with human papillomavirus infection. Cancer Lett 226(1):77–84. https://doi.org/10.1016/j.canlet.2004.12.031

Lin J, Haffner MC, Zhang Y, Lee BH, Brennen WN, Britton J et al (2011) Disulfiram is a DNA demethylating agent and inhibits prostate cancer cell

growth. Prostate 71(4):333–343. https://doi.org/10.1002/pros.21247

Lin N, Qin S, Luo S, Cui S, Huang G, Zhang X (2014) Homocysteine induces cytotoxicity and proliferation inhibition in neural stem cells via DNA methylation *in vitro*. FEBS J 281(8):2088–2096. https://doi.org/10.1111/febs.12764

Liu Z, Xie Z, Jones W, Pavlovicz RE, Liu S, Yu J et al (2009) Curcumin is a potent DNA hypomethylation agent. Bioorg Med Chem Lett 19(3):706–709. https://doi.org/10.1016/j.bmcl.2008.12.041

Lloyd CM, Hawrylowicz CM (2009) Regulatory T cells in asthma. Immunity 31(3):438–449. https://doi.org/10.1016/j.immuni.2009.08.007

Lobo J, Cardoso AR, Miranda-Goncalves V, Looijenga LHJ, Lopez M, Arimondo PB et al (2021) Targeting germ cell tumors with the newly synthesized flavanone-derived compound MLo1302 efficiently reduces tumor cell viability and induces apoptosis and cell cycle arrest. Pharmaceutics 13(1):73. https://doi.org/10.3390/pharmaceutics13010073

Logrip ML, Janak PH, Ron D (2009) Escalating ethanol intake is associated with altered corticostriatal BDNF expression. J Neurochem 109(5):1459–1468. https://doi.org/10.1111/j.1471-4159.2009.06073.x

Lopez M, Leroy M, Etievant C, Ausseil F, Gagnon A, Arimondo PB (2015) Drug discovery methods. In: Egger G, Arimondo PB (eds) Drug discovery in cancer epigenetics. Elsevier-Academic, Amsterdam, pp 63–95

Lovinsky-Desir S, Ridder R, Torrone D, Maher C, Narula S, Scheuerman M et al (2014) DNA methylation of the allergy regulatory gene interferon gamma varies by age, sex, and tissue type in asthmatics. Clin Epigenetics 6(1):9

Ludwig AK, Zhang P, Cardoso MC (2016) Modifiers and readers of DNA modifications and their impact on genome structure, expression, and stability in disease. Front Genet 7:115. https://doi.org/10.3389/fgene.2016.00115

Lund G, Andersson L, Lauria M, Lindholm M, Fraga MF, Villar-Garea A et al (2004) DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein E. J Biol Chem 279(28):29147–29154. https://doi.org/10.1074/jbc.M403618200

Luo S, Zhang X, Yu M, Yan H, Liu H, Wilson J et al (2013) Folic acid acts through DNA methyltransferases to induce the differentiation of neural stem cells into neurons. Cell Biochem Biophys 66(3):559–566. https://doi.org/10.1007/s12013-012-9503-6

Luviano N, Lopez M, Gawehns F, Chaparro C, Arimondo PB, Ivanovic S et al (2021) The methylome of Biomphalaria glabrata and other mollusks: enduring modification of epigenetic landscape and phenotypic traits by a new DNA methylation inhibitor. Epigenetics Chromatin 14(1):48. https://doi.org/10.1186/s13072-021-00422-7

Luzzi A, Morettini F, Gazaneo S, Mundo L, Onnis A, Mannucci S et al (2014) HIV-1 Tat induces DNMT over-expression through microRNA dysregulation in HIV-related non Hodgkin lymphomas. Infect Agent Cancer 9(1):41. https://doi.org/10.1186/1750-9378-9-41

Lyst MJ, Bird A (2015) Rett syndrome: a complex disorder with simple roots. Nat Rev Genet 16(5):261–275. https://doi.org/10.1038/nrg3897

Ma N, Chen W, Fan T, Tian Y, Zhang S, Zeng D et al (2015) Low temperature-induced DNA hypermethylation attenuates expression of *RhAG*, an *AGAMOUS* homolog, and increases petal number in rose (*Rosa hybrida*). BMC Plant Biol 15(1):237

Machnes ZM, Huang TCT, Chang PKY, Gill R, Reist N, Dezsi G et al (2013) DNA methylation mediates persistent epileptiform activity *in vitro* and *in vivo*. PLoS One 8(10):e76299. https://doi.org/10.1371/journal.pone.0076299

Marchal C, de Dieuleveult M, Saint-Ruf C, Guinot N, Ferry L, Olalla Saad ST et al (2018) Depletion of ZBTB38 potentiates the effects of DNA demethylating agents in cancer cells via CDKN1C mRNA up-regulation. Oncogenesis 7(10):82. https://doi.org/10.1038/s41389-018-0092-0

Marques-Magalhães Â, Graça I, Miranda-Gonçalves V, Henrique R, Lopez M, Arimondo PB et al (2021) Anti-neoplastic and demethylating activity of a newly synthetized flavanone-derived compound in Renal Cell Carcinoma cell lines. Biomed Pharmacother 141:111681. https://doi.org/10.1016/j.biopha.2021.111681

Mastroeni D, Grover A, Delvaux E, Whiteside C, Coleman PD, Rogers J (2010) Epigenetic changes in Alzheimer's disease: Decrements in DNA methylation. Neurobiol Aging 31(12):2025–2037. https://doi.org/10.1016/j.neurobiolaging.2008.12.005

Matoušová M, Votruba I, Otmar M, Tloušťová E, Günterová J, Mertlíková-Kaiserová H (2011) 2′-deoxy-5,6-dihydro-5-azacytidine—a less toxic alternative of 2′-deoxy-5-azacytidine: A comparative study of hypomethylating potential. Epigenetics 6(6):769–776. https://doi.org/10.4161/epi.6.6.16215

Matrisciano F, Tueting P, Dalal I, Kadriu B, Grayson DR, Davis JM et al (2013) Epigenetic modifications of GABAergic interneurons are associated with the schizophrenia-like phenotype induced by prenatal stress in mice. Neuropharmacology 68:184–194. https://doi.org/10.1016/j.neuropharm.2012.04.013

Meadows JP, Guzman-Karlsson MC, Phillips S, Holleman C, Posey JL, Day JJ et al (2015) DNA methylation regulates neuronal glutamatergic synaptic scaling. Sci Signal 8(382):ra61. https://doi.org/10.1126/scisignal.aab0715

Medina-Franco JL, López-Vallejo F, Kuck D, Lyko F (2011) Natural products as DNA methyltransferase inhibitors: a computer-aided discovery approach. Mol Divers 15(2):293–304. https://doi.org/10.1007/s11030-010-9262-5

Melas PA, Rogdaki M, Lennartsson A, Björk K, Qi H, Witasp A et al (2012) Antidepressant treatment is associated with epigenetic alterations in the promoter of P11 in a genetic model of depression. Int J Neuropsychopharmacol 15(5):669–679. https://doi.org/10.1017/s1461145711000940

Miao C-g, Huang C, Huang Y, Yang Y-y, He X, Zhang L et al (2013) MeCP2 modulates the canonical Wnt pathway activation by targeting SFRP4 in rheumatoid arthritis fibroblast-like synoviocytes in rats. Cell Signal 25(3):598–608. https://doi.org/10.1016/j.cellsig.2012.11.023

Mikovits JA, Young HA, Vertino P, Issa J-PJ, Pitha PM, Turcoski-Corrales S et al (1998) Infection with human immunodeficiency virus type 1 upregulates DNA methyltransferase, resulting in de novo methylation of the gamma interferon (IFN-γ) promoter and subsequent downregulation of IFN-γ production. Mol Cell Biol 18(9):5166–5177. https://doi.org/10.1128/mcb.18.9.5166

Miller CA, Sweatt JD (2007) Covalent modification of DNA regulates memory formation. Neuron 53(6): 857–869. https://doi.org/10.1016/j.neuron.2007.02.022

Miller CA, Gavin CF, White JA, Parrish RR, Honasoge A, Yancey CR et al (2010) Cortical DNA methylation maintains remote memory. Nat Neurosci 13(6): 664–666. http://www.nature.com/neuro/journal/v13/n6/abs/nn.2560.html#supplementary-information

Mitchnick KA, Creighton S, O'Hara M, Kalisch BE, Winters BD (2015) Differential contributions of de novo and maintenance DNA methyltransferases to object memory processing in the rat hippocampus and perirhinal cortex—a double dissociation. Eur J Neurosci 41(6):773–786. https://doi.org/10.1111/ejn.12819

Mohd Murshid N, Aminullah Lubis F, Makpol S (2022) Epigenetic changes and its intervention in age-related neurodegenerative diseases. Cell Mol Neurobiol 42(3): 577–595. https://doi.org/10.1007/s10571-020-00979-z

Myrianthopoulos V, Cartron PF, Liutkevičiūtė Z, Klimašauskas S, Matulis D, Bronner C et al (2016) Tandem virtual screening targeting the SRA domain of UHRF1 identifies a novel chemical tool modulating DNA methylation. Eur J Med Chem 114:390–396. https://doi.org/10.1016/j.ejmech.2016.02.043

Najem AS, Khawaja G, Hodroj HM, Rizk S (2019) Synergistic effect of epigenetic inhibitors decitabine and suberoylanilide hydroxamic acid on colorectal cancer in vitro. Curr Mol Pharmacol 12(4):281–300. https://doi.org/10.2174/1874467212666190313154531

Nardella F, Halby L, Hammam E, Erdmann D, Cadet-Daniel V, Peronet R et al (2020) DNA methylation bisubstrate inhibitors are fast-acting drugs active against artemisinin-resistant Plasmodium falciparum parasites. ACS Cent Sci 6(1):16–21. https://doi.org/10.1021/acscentsci.9b00874

Nature Editorials (2014) Suicide watch. Nature 506:131

Nelson ED, Kavalali ET, Monteggia LM (2008) Activity-dependent suppression of miniature neurotransmission through the regulation of DNA methylation. J Neurosci 28(2):395–406. https://doi.org/10.1523/jneurosci.3796-07.2008

Nestor CE, Barrenäs F, Wang H, Lentini A, Zhang H, Bruhn S et al (2014) DNA methylation changes separate allergic patients from healthy controls and may reflect altered CD4+ T-cell population structure. PLoS Genet 10(1):e1004059. https://doi.org/10.1371/journal.pgen.1004059

Newton AS, Faver JC, Micevic G, Muthusamy V, Kudalkar SN, Bertoletti N et al (2020) Structure-guided identification of DNMT3B inhibitors. ACS Med Chem Lett 11(5):971–976. https://doi.org/10.1021/acsmedchemlett.0c00011

Niinuma T, Kitajima H, Kai M, Yamamoto E, Yorozu A, Ishiguro K et al (2019) UHRF1 depletion and HDAC inhibition reactivate epigenetically silenced genes in colorectal cancer cells. Clin Epigenetics 11(1):70. https://doi.org/10.1186/s13148-019-0668-3

Noh JS, Sharma RP, Veldic M, Salvacion AA, Jia X, Chen Y et al (2005) DNA methyltransferase 1 regulates reelin mRNA expression in mouse primary cortical cultures. Proc Natl Acad Sci U S A 102(5): 1749–1754. https://doi.org/10.1073/pnas.0409648102

Notari RE, Deyoung JL (1975) Kinetics and mechanisms of degradation of the antileukemic agent 5-azacytidine in aqueous solutions. J Pharm Sci 64(7):1148–1157. https://doi.org/10.1002/jps.2600640704

Nutt DJ, King LA, Phillips LD (2010) Drug harms in the UK: a multicriteria decision analysis. Lancet 376(9752):1558–1565. https://doi.org/10.1016/s0140-6736(10)61462-6

Ocskó T, Tóth DM, Hoffmann G, Tubak V, Glant TT, Rauch TA (2018) Transcription factor Zbtb38 downregulates the expression of anti-inflammatory IL1r2 in mouse model of rheumatoid arthritis. Biochim Biophys Acta Gene Regul Mech 1861(11):1040–1047. https://doi.org/10.1016/j.bbagrm.2018.09.007

Oh YS, Jeong S-G, Cho G-W (2015a) Anti-senescence effects of DNA methyltransferase inhibitor RG108 in human bone marrow mesenchymal stromal cells. Biotechnol Appl Biochem 583–590. https://doi.org/10.1002/bab.1393

Oh YS, Kim SH, Cho G-W (2015b) Functional restoration of amyotrophic lateral sclerosis patient-derived mesenchymal stromal cells through inhibition of DNA methyltransferase. Cell Mol Neurobiol 613–620. https://doi.org/10.1007/s10571-015-0242-2

Pappalardi MB, Keenan K, Cockerill M, Kellner WA, Stowell A, Sherk C et al (2021) Discovery of a first-in-class reversible DNMT1-selective inhibitor with improved tolerability and efficacy in acute myeloid leukemia. Nat Cancer 2(10):1002–1017. https://doi.org/10.1038/s43018-021-00249-x

Park J, Jang KL (2014) Hepatitis C virus represses E-cadherin expression via DNA methylation to induce

epithelial to mesenchymal transition in human hepatocytes. Biochem Biophys Res Commun 446(2): 561–567. https://doi.org/10.1016/j.bbrc.2014.03.009

Parrish RR, Albertson AJ, Buckingham SC, Hablitz JJ, Mascia KL, Haselden WD et al (2013) Status epilepticus triggers early and late alterations in brain-derived neurotrophic factor and NMDA glutamate receptor GRIN2B DNA methylation levels in the hippocampus. Neuroscience 0:602–619. https://doi.org/10.1016/j.neuroscience.2013.06.029

Paschos K, Smith P, Anderton E, Middeldorp JM, White RE, Allday MJ (2009) Epstein-Barr virus latency in B cells leads to epigenetic repression and CpG methylation of the tumour suppressor gene bim. PLoS Pathog 5(6):e1000492. https://doi.org/10.1371/journal.ppat.1000492

Peat JR, Reik W (2012) Incomplete methylation reprogramming in SCNT embryos. Nat Genet 44(9): 965–966

Pechalrieu D, Etievant C, Arimondo PB (2017) DNA methyltransferase inhibitors in cancer: From pharmacology to translational studies. Biochem Pharmacol 129:1–13. https://doi.org/10.1016/j.bcp.2016.12.004

Pechalrieu D, Dauzonne D, Arimondo PB, Lopez M (2020) Synthesis of novel 3-halo-3-nitroflavanones and their activities as DNA methyltransferase inhibitors in cancer cells. Eur J Med Chem 186: 111829. https://doi.org/10.1016/j.ejmech.2019.111829

Penter L, Maier B, Frede U, Hackner B, Carell T, Hagemeier C et al (2015) A rapid screening system evaluates novel inhibitors of DNA methylation and suggests F-box proteins as potential therapeutic targets for high-risk neuroblastoma. Target Oncol 10(4): 523–533. https://doi.org/10.1007/s11523-014-0354-5

Pereira R, Benedetti R, Pérez-Rodriguez S, Nebbioso A, GarcÃia-Rodriguez J, Carafa V et al (2012) Indole-derived psammaplin A analogues as epigenetic modulators with multiple inhibitory activities. J Med Chem 55(22):9467–9491

Philibert RA, Plume JM, Gibbons FX, Brody GH, Beach SRH (2012) The impact of recent alcohol use on genome wide DNA methylation signatures. Front Genet 3:54. https://doi.org/10.3389/fgene.2012.00054

Ponomarev I, Wang S, Zhang L, Harris RA, Mayfield RD (2012) Gene coexpression networks in human brain identify epigenetic modifications in alcohol dependence. J Neurosci 32(5):1884–1897. https://doi.org/10.1523/jneurosci.3136-11.2012

Prokhortchouk A, Sansom O, Selfridge J, Caballero Isabel M, Salozhin S, Aithozhina D et al (2006) Kaiso-deficient mice show resistance to intestinal cancer. Mol Cell Biol 26(1):199–208. https://doi.org/10.1128/MCB.26.1.199-208.2006

Rabal O, San José-Enériz E, Agirre X, Sánchez-Arias JA, Vilas-Zornoza A, Ugarte A et al (2018) Discovery of reversible DNA methyltransferase and lysine methyltransferase G9a inhibitors with antitumoral in vivo efficacy. J Med Chem 61(15):6518–6545. https://doi.org/10.1021/acs.jmedchem.7b01926

Rabal O, San José-Enériz E, Agirre X, Sánchez-Arias JA, de Miguel I, Ordoñez R et al (2021) Design and synthesis of novel epigenetic inhibitors targeting histone deacetylases, DNA methyltransferase 1, and lysine methyltransferase G9a with in vivo efficacy in multiple myeloma. J Med Chem 64(6):3392–3426. https://doi.org/10.1021/acs.jmedchem.0c02255

Reichert ZR (2010) Targeting epigenetic repression by interfering with methyl-binding domain protein function. John Hopkins University, Baltimore, MA

Ren Y, Sun Q, Yuan Z, Jiang Y (2019) Combined inhibition of HDAC and DNMT1 induces p85α/MEK-mediated cell cycle arrest by dual target inhibitor 208 in U937 cells. Chin Chem Lett 30(6):1233–1236. https://doi.org/10.1016/j.cclet.2019.03.029

Rilova E, Erdmann A, Gros C, Masson V, Aussagues Y, Poughon-Cassabois V et al (2014) Design, synthesis and biological evaluation of 4-amino-N-(4-aminophenyl)benzamide analogues of quinoline-based SGI-1027 as inhibitors of DNA methylation. ChemMedChem 9(3):590–601. https://doi.org/10.1002/cmdc.201300420

Robinson SC, Chaudhary R, Jiménez-Saiz R, Rayner LGA, Bayer L, Jordana M et al (2019) Kaiso-induced intestinal inflammation is preceded by diminished E-cadherin expression and intestinal integrity. PLoS One 14(6):e0217220. https://doi.org/10.1371/journal.pone.0217220

Ronald PC (1997) The molecular basis of disease resistance in rice. Plant Mol Biol 35(1-2):179–186. https://doi.org/10.1023/a:1005750811983

Rondelet G, Fleury L, Faux C, Masson V, Dubois J, Arimondo PB et al (2017) Inhibition studies of DNA methyltransferases by maleimide derivatives of RG108 as non-nucleoside inhibitors. Future Med Chem 9 (13):1465–1481. https://doi.org/10.4155/fmc-2017-0074

Roos A, Dhruv HD, Mathews IT, Inge LJ, Tuncali S, Hartman LK et al (2017) Identification of aurintricarboxylic acid as a selective inhibitor of the TWEAK-Fn14 signaling pathway in glioblastoma cells. Oncotarget 8(7):12234–12246. https://doi.org/10.18632/oncotarget.14685

Roth TL, Lubin FD, Funk AJ, Sweatt JD (2009) Lasting epigenetic influence of early-life adversity on the BDNF gene. Biol Psychiatry 65(9):760–769. https://doi.org/10.1016/j.biopsych.2008.11.028

Rotili D, Tarantino D, Marrocco B, Gros C, Masson V, Poughon V et al (2014) Properly substituted analogues of BIX-01294 lose inhibition of G9a histone methyltransferase and gain selective anti-DNA methyltransferase 3A activity. PLoS One 9(5): e96941. https://doi.org/10.1371/journal.pone.0096941

Roussel-Gervais A, Naciri I, Kirsh O, Kasprzyk L, Velasco G, Grillo G et al (2017) Loss of the methyl-CpG-binding protein ZBTB4 alters mitotic checkpoint, increases aneuploidy, and promotes tumorigenesis.

Cancer Res 77(1):62–73. https://doi.org/10.1158/0008-5472.CAN-16-1181
Ruijter AJMD, Gennip AHV, Caron HN, Kemp S, Kuilenburg ABPV (2003) Histone deacetylases (HDACs): characterization of the classical HDAC family. Biochem J 370(3):737–749. https://doi.org/10.1042/bj20021321
Runyon RS, Cachola LM, Rajeshuni N, Hunter T, Garcia M, Ahn R et al (2012) Asthma discordance in twins is linked to epigenetic modifications of T cells. PLoS One 7(11):e48796. https://doi.org/10.1371/journal.pone.0048796
Saggioro D, Forino M, Chieco-Bianchi L (1991) Transcriptional block of HTLV-I LTR by sequence-specific methylation. Virology 182(1):68–75. https://doi.org/10.1016/0042-6822(91)90649-V
Saini M, Selokar NL, Agrawal H, Singla SK, Chauhan MS, Manik RS et al (2014) Treatment of buffalo (*Bubalus bubalis*) donor cells with trichostatin A and 5-aza-2′-deoxycytidine alters their growth characteristics, gene expression and epigenetic status and improves the in vitro developmental competence, quality and epigenetic status of cloned embryos. Reprod Fertil Dev. https://doi.org/10.1071/RD14176 [Epub ahead of print]
Sales AJ, Biojone C, Terceti MS, Guimarães FS, Gomes MVM, Joca SRL (2011) Antidepressant-like effect induced by systemic and intra-hippocampal administration of DNA methylation inhibitors. Br J Pharmacol 164(6):1711–1721. https://doi.org/10.1111/j.1476-5381.2011.01489.x
San José-Enériz E, Agirre X, Rabal O, Vilas-Zornoza A, Sanchez-Arias JA, Miranda E et al (2017) Discovery of first-in-class reversible dual small molecule inhibitors against G9a and DNMTs in hematological malignancies. Nat Commun 8(1):15424. https://doi.org/10.1038/ncomms15424
Sano H, Kamada I, Youssefian S, Katsumi M, Wabiko H (1990) A single treatment of rice seedlings with 5-azacytidine induces heritable dwarfism and undermethylation of genomic DNA. Mol Gen Genet 220(3):441–447. https://doi.org/10.1007/bf00391751
Santi DV, Norment A, Garrett CE (1984) Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. Proc Natl Acad Sci U S A 81(22):6993–6997
Sasai N, Matsuda E, Sarashina E, Ishida Y, Kawaichi M (2005) Identification of a novel BTB-zinc finger transcriptional repressor, CIBZ, that interacts with CtBP corepressor. Genes Cells 10(9):871–885. https://doi.org/10.1111/j.1365-2443.2005.00885.x
Scarpa S, Fuso A, D'Anselmi F, Cavallaro RA (2003) Presenilin 1 gene silencing by *S*-adenosylmethionine: a treatment for Alzheimer disease? FEBS Lett 541(1–3):145–148. https://doi.org/10.1016/S0014-5793(03)00277-1
Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ 3rd. (2002) SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev 16(8):919–932. https://doi.org/10.1101/gad.973302
Senisterra G, Zhu HY, Luo X, Zhang H, Xun G, Lu C et al (2018) Discovery of small-molecule antagonists of the H3K9me3 binding to UHRF1 tandem tudor domain. SLAS Discov Adv Sci Drug Discov 23(9):930–940. https://doi.org/10.1177/2472555218766278
Shahid SK, Kharitonov SA, Wilson NM, Bush A, Barnes PJ (2002) Increased interleukin-4 and decreased interferon-γ in exhaled breath condensate of children with asthma. Am J Respir Crit Care Med 165(9):1290–1293. https://doi.org/10.1164/rccm.2108082
Sheikhnejad G, Brank A, Christman JK, Goddard A, Alvarez E, Ford H Jr et al (1999) Mechanism of inhibition of DNA (cytosine C5)-methyltransferases by oligodeoxyribonucleotides containing 5,6-dihydro-5-azacytosine. J Mol Biol 285(5):2021–2034. https://doi.org/10.1006/jmbi.1998.2426
Sidhu H, Capalash N (2017) UHRF1: The key regulator of epigenetics and molecular target for cancer therapeutics. Tumor Biol 39(2):1010428317692205. https://doi.org/10.1177/1010428317692205
Siedlecki P, Boy RG, Musch T, Brueckner B, Suhai S, Lyko F et al (2006) Discovery of two novel, small-molecule inhibitors of DNA methylation. J Med Chem 49(2):678–683. https://doi.org/10.1021/jm050844z
Silverman LR, Verma A, Odchimar-Reissig R, Leblanc A, Najfeld V, Gabrilove J et al (2008) A phase I trial of epigenetic modulators vorinostat, in combination with azacitidine (azaC) in patients with the myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML): a study of the New York Cancer Consortium [Abstract]. Blood (ASH Annu Meet Abstr) 112:3656
Silverman LR, Holland J, Sandvold ML, Myhren F, Eriksen OH (2009) Analogues d'azacytidine et leurs utilisations. WO2009042767 A1, PCT/US2008/077673
Singh N, Dueñas-González A, Lyko F, Medina-Franco JL (2009a) Molecular modeling and molecular dynamics studies of hydralazine with human DNA methyltransferase 1. ChemMedChem 4(5):792–799. https://doi.org/10.1002/cmdc.200900017
Singh RP, Shiue K, Schomberg D, Zhou FC (2009b) Cellular epigenetic modifications of neural stem cell differentiation. Cell Transplant 18(10):1197–1211. https://doi.org/10.3727/096368909x12483162197204
Smith AK, Conneely KN, Kilaru V, Mercer KB, Weiss TE, Bradley B et al (2011) Differential immune system DNA methylation and cytokine regulation in post-traumatic stress disorder. Am J Med Genet B Neuropsychiatr Genet 156(6):700–708. https://doi.org/10.1002/ajmg.b.31212
Sorm F, Piskala A, Cihak A, Vesely J (1964) 5-Azacytidine, a new, highly effective cancerostatic. Experientia 20(4):202–203
Stirzaker C, Song JZ, Ng W, Du Q, Armstrong NJ, Locke WJ et al (2017) Methyl-CpG-binding protein MBD2

plays a key role in maintenance and spread of DNA methylation at CpG islands and shores in cancer. Oncogene 36(10):1328–1338. https://doi.org/10.1038/onc.2016.297

Stresemann C, Brueckner B, Musch T, Stopper H, Lyko F (2006) Functional diversity of DNA methyltransferase inhibitors in human cancer cell lines. Cancer Res 66(5):2794–2800. https://doi.org/10.1158/0008-5472.can-05-2821

Struck A-W, Thompson ML, Wong LS, Micklefield J (2012) *S*-adenosyl-methionine-dependent methyltransferases: highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. ChemBioChem 13(18):2642–2655

Suh N, Pezzuto JM (2012) Strawberry fields forever? Cancer Prev Res 5(1):30–33. https://doi.org/10.1158/1940-6207.capr-11-0534

Sun J, Awakawa T, Noguchi H, Abe I (2012) Induced production of mycotoxins in an endophytic fungus from the medicinal plant *Datura stramonium L.* Bioorg Med Chem Lett 22(20):6397–6400. https://doi.org/10.1016/j.bmcl.2012.08.063

Sun H, Kennedy PJ, Nestler EJ (2013) Epigenetics of the depressed brain: role of histone acetylation and methylation. Neuropsychopharmacology 38(1):124–137

Sun Y, Wang Z, Tao J, Wang Y, Wu A, Yang Z et al (2014) Yeast-based assays for the high-throughput screening of inhibitors of coronavirus RNA cap guanine-N7-methyltransferase. Antiviral Res 104:156–164. https://doi.org/10.1016/j.antiviral.2014.02.002

Sun Q, Dai Q, Zhang C, Chen Y, Zhao L, Yuan Z et al (2021a) Design, synthesis and antitumor evaluations of nucleoside base hydroxamic acid derivatives as DNMT and HDAC dual inhibitors. Chin Chem Lett 32(8):2479–2483. https://doi.org/10.1016/j.cclet.2021.02.004

Sun X, Wang Z, Cong X, Lv Y, Li Z, Rong L et al (2021b) Mitochondrial gene COX2 methylation and downregulation is a biomarker of aging in heart mesenchymal stem cells. Int J Mol Med 47(1):161–170. https://doi.org/10.3892/ijmm.2020.4799

Sutcliffe JS, Nelson DL, Zhang F, Pieretti M, Caskey CT, Saxe D et al (1992) DNA methylation represses FMR-1 transcription in fragile X syndrome. Hum Mol Genet 1(6):397–400. https://doi.org/10.1093/hmg/1.6.397

Suzuki T, Tanaka R, Hamada S, Nakagawa H, Miyata N (2010) Design, synthesis, inhibitory activity, and binding mode study of novel DNA methyltransferase 1 inhibitors. Bioorg Med Chem Lett 20(3):1124–1127

Szyf M (2015) Prospects for the development of epigenetic drugs for CNS conditions. Nat Rev Drug Discov 14(7):461–474. https://doi.org/10.1038/nrd4580

Tabolacci E, Pietrobono R, Moscato U, Oostra BA, Chiurazzi P, Neri G (2005) Differential epigenetic modifications in the FMR1 gene of the fragile X syndrome after reactivating pharmacological treatments. Eur J Hum Genet 13(5):641–648. https://doi.org/10.1038/sj.ejhg.5201393

Tang C, Niu Q, Cicka D, Du Y, Mo X, Fu H (2021) A time-resolved fluorescence resonance energy transfer screening assay for discovery of protein-protein interaction modulators. STAR Protoc 2(3):100804. https://doi.org/10.1016/j.xpro.2021.100804

Taylor SM, Jones PA (1979) Multiple new phenotypes induced in 10T12 and 3T3 cells treated with 5-azacytidine. Cell 17(4):771–779. https://doi.org/10.1016/0092-8674(79)90317-9

Thangavel J, Malik AB, Elias HK, Rajasingh S, Simpson AD, Sundivakkam PK et al (2014) Combinatorial therapy with acetylation and methylation modifiers attenuates lung vascular hyperpermeability in endotoxemia-induced mouse inflammatory lung injury. Am J Pathol 184(8):2237–2249. https://doi.org/10.1016/j.ajpath.2014.05.008

Thangavel J, Samanta S, Rajasingh S, Barani B, Xuan Y-T, Dawn B et al (2015) Epigenetic modifiers reduce inflammation and modulate macrophage phenotype during endotoxemia-induced acute lung injury. J Cell Sci 128(16):3094–3105. https://doi.org/10.1242/jcs.170258

Tolg C, Sabha N, Cortese R, Panchal T, Ahsan A, Soliman A et al (2011) Uropathogenic E. coli infection provokes epigenetic downregulation of CDKN2A (p16INK4A) in uroepithelial cells. Lab Invest 91(6):825–836. https://doi.org/10.1038/labinvest.2010.197

Tomaselli D, Lucidi A, Rotili D, Mai A (2020) Epigenetic polypharmacology: A new frontier for epi-drug discovery. Med Res Rev 40(1):190–244. https://doi.org/10.1002/med.21600

Uchida S, Hara K, Kobayashi A, Otsuki K, Yamagata H, Hobara T et al (2011) Epigenetic status of *Gdnf* in the ventral striatum determines susceptibility and adaptation to daily stressful events. Neuron 69(2):359–372. https://doi.org/10.1016/j.neuron.2010.12.023

Uddin M, Aiello AE, Wildman DE, Koenen KC, Pawelec G, de los Santos R et al (2010) Epigenetic and immune function profiles associated with posttraumatic stress disorder. Proc Natl Acad Sci U S A 107(20):9470–9475. https://doi.org/10.1073/pnas.0910794107

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

Valente S, Liu Y, Schnekenburger M, Zwergel C, Cosconati S, Gros C et al (2014) Selective non-nucleoside inhibitors of human DNA methyltransferases active in cancer including in cancer stem cells. J Med Chem 57(3):701–713. https://doi.org/10.1021/jm4012627

van Bemmel DM, Brank AS, Eritja R, Marquez VE, Christman JK (2009) DNA (Cytosine-C5) methyltransferase inhibition by oligodeoxyribonucleotides containing 2-(1H)-pyrimidinone (zebularine aglycon) at the enzymatic target site. Biochem Pharmacol 78(6):633–641. https://doi.org/10.1016/j.bcp.2009.05.017

Van Lint C, Bouchat S, Marcello A (2013) HIV-1 transcription and latency: an update. Retrovirology 10(1): 67. https://doi.org/10.1186/1742-4690-10-67

van Roy FM, McCrea PD (2005) A role for Kaiso-p120ctn complexes in cancer? Nat Rev Cancer 5(12):956–964. https://doi.org/10.1038/nrc1752

Veldic M, Caruncho HJ, Liu WS, Davis J, Satta R, Grayson DR et al (2004) DNA-methyltransferase 1 mRNA is selectively overexpressed in telencephalic GABAergic interneurons of schizophrenia brains. Proc Natl Acad Sci U S A 101(1):348–353. https://doi.org/10.1073/pnas.2637013100

Villar-Garea A, Fraga MF, Espada J, Esteller M (2003) Procaine is a DNA-demethylating agent with growth-inhibitory effects in human cancer cells. Cancer Res 63(16):4984–4989

Wang X, Chao L, Jin G, Ma G, Zang Y, Sun J (2009) Association between CpG island methylation of the WWOX gene and its expression in breast cancers. Tumor Biol 30(1):8–14

Wang X, Sena Filho JG, Hoover AR, King JB, Ellis TK, Powell DR et al (2010a) Chemical epigenetics alters the secondary metabolite composition of guttate excreted by an atlantic-forest-soil-derived Penicillium citreonigrum. J Nat Prod 73(5):942–948. https://doi.org/10.1021/np100142h

Wang X, Zhu H, Snieder H, Su S, Munn D, Harshfield G et al (2010b) Obesity related methylation changes in DNA of peripheral blood leukocytes. BMC Med 8:87. https://doi.org/10.1186/1741-7015-8-87

Wang L, Zhang Y, Li R, Chen Y, Pan X, Li G et al (2013a) 5-aza-2′-Deoxycytidine enhances the radiosensitivity of breast cancer cells. Cancer Biother Radiopharm 28(1):34–44. https://doi.org/10.1089/cbr.2012.1170.

Wang Y, Wang X, Li R, Yang Z-F, Wang Y-Z, Gong X-L et al (2013b) A DNA methyltransferase inhibitor, 5-Aza-2′-deoxycytidine, exacerbates neurotoxicity and upregulates Parkinson's disease-related genes in dopaminergic neurons. CNS Neurosci Ther 19(3): 183–190. https://doi.org/10.1111/cns.12059.

Warnault V, Darcq E, Levine A, Barak S, Ron D (2013) Chromatin remodeling—a novel strategy to control excessive alcohol drinking. Transl Psychiatry 3:e231. https://doi.org/10.1038/tp.2013.4

Watanabe A, Yamada Y, Yamanaka S (2013) Epigenetic regulation in pluripotent stem cells: a key to breaking the epigenetic barrier. Philos Trans R Soc B Biol Sci 368(1609):20120292. https://doi.org/10.1098/rstb.2012.0292

Watson CJ, Collier P, Tea I, Neary R, Watson JA, Robinson C et al (2014) Hypoxia-induced epigenetic modifications are associated with cardiac tissue fibrosis and the development of a myofibroblast-like phenotype. Hum Mol Genet 23(8):2176–2188. https://doi.org/10.1093/hmg/ddt614

White GP, Hollams EM, Yerkovich ST, Bosco A, Holt BJ, Bassami MR et al (2006) CpG methylation patterns in the IFNγ promoter in naive T cells: Variations during Th1 and Th2 differentiation and between atopics and non-atopics. Pediatr Allergy Immunol 17(8):557–564. https://doi.org/10.1111/j.1399-3038.2006.00465.x

Winders BR, Schwartz RH, Bruniquel D (2004) A distinct region of the murine IFN-γ promoter is hypomethylated from early T cell development through mature naive and Th1 cell differentiation, but is hypermethylated in Th2 cells. J Immunol 173(12): 7377–7384. https://doi.org/10.4049/jimmunol.173.12.7377

Wyhs N, Walker D, Giovinazzo H, Yegnasubramanian S, Nelson WG (2014) Time-resolved fluorescence resonance energy transfer assay for discovery of small-molecule inhibitors of methyl-cpg binding domain protein 2. J Biomol Screen 19(7):1060–1069. https://doi.org/10.1177/1087057114526433

Xiao D, Dasgupta C, Chen M, Zhang K, Buchholz J, Xu Z et al (2014) Inhibition of DNA methylation reverses norepinephrine-induced cardiac hypertrophy in rats. Cardiovasc Res 101(3):373–382. https://doi.org/10.1093/cvr/cvt264

Xue Y, Wong J, Moreno GT, Young MK, Côté J, Wang W (1998) NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol Cell 2(6):851–861. https://doi.org/10.1016/s1097-2765(00)80299-3

Yamauchi T, Kamon J, Minokoshi Y, Ito Y, Waki H, Uchida S et al (2002) Adiponectin stimulates glucose utilization and fatty-acid oxidation by activating AMP-activated protein kinase. Nat Med 8(11): 1288–1295. http://www.nature.com/nm/journal/v8/n11/suppinfo/nm788_S1.html

Yang H, Hoshino K, Sanchez-Gonzalez B, Kantarjian H, Garcia-Manero G (2005) Antileukemia activity of the combination of 5-aza-2′-deoxycytidine with valproic acid. Leuk Res 29(7):739–748 https://doi.org/10.1016/j.leukres.2004.11.022

Yang D, Torres CM, Bardhan K, Zimmerman M, McGaha TL, Liu K (2012) Decitabine and vorinostat cooperate to sensitize colon carcinoma cells to Fas ligand-induced apoptosis *in vitro* and tumor suppression *in vivo*. J Immunol 188(9):4441–4449. https://doi.org/10.4049/jimmunol.1103035

Yie Y, Zhao S, Tang Q, Zheng F, Wu J, Yang L et al (2015) Ursolic acid inhibited growth of hepatocellular carcinoma HepG2 cells through AMPKalpha-mediated reduction of DNA methyltransferase 1. Mol Cell Biochem 402(1-2):63–74. https://doi.org/10.1007/s11010-014-2314-x

Yin L, Chung WO (2011) Epigenetic regulation of human β-defensin 2 and CC chemokine ligand 20 expression in gingival epithelial cells in response to oral bacteria. Mucosal Immunol 4(4):409–419. https://doi.org/10.1038/mi.2010.83

Yin L-J, Zhang Y, Lv P-P, He W-H, Wu Y-T, Liu A-X et al (2012) Insufficient maintenance DNA methylation is associated with abnormal embryonic development. BMC Med 10(1):26. https://doi.org/10.1186/1741-7015-10-26

Yogelzang NJ, Herndon JE 2nd, Cirrincione C, Harmon DC, Antman KH, Corson JM et al (1997) Dihydro-5-azacytidine in malignant mesothelioma. A phase II trial demonstrating activity accompanied by cardiac toxicity. Cancer and Leukemia Group B. Cancer 79(11): 2237–2242

Yoo CB, Jeong S, Egger G, Liang G, Phiasivongsa P, Tang C et al (2007) Delivery of 5-aza-2′-deoxycytidine to cells using oligodeoxynucleotides. Cancer Res 67(13):6400–6408. https://doi.org/10.1158/0008-5472.CAN-07-0251

Yoo J, Choi S, Medina-Franco JL (2013) Molecular modeling studies of the novel inhibitors of DNA methyltransferases SGI-1027 and CBC12: implications for the mechanism of inhibition of DNMTs. PLoS One 8(4):e62152. https://doi.org/10.1371/journal.pone.0062152

Yoon H-G, Chan DW, Reynolds AB, Qin J, Wong J (2003) N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. Mol Cell 12(3):723–734. https://doi.org/10.1016/j.molcel.2003.08.008

Yu Q, Zhou B, Zhang Y, Nguyen ET, Du J, Glosson NL et al (2012) DNA methyltransferase 3a limits the expression of interleukin-13 in T helper 2 cells and allergic airway inflammation. Proc Natl Acad Sci U S A 109(2):541–546. https://doi.org/10.1073/pnas.1103803109

Yuan Z, Sun Q, Li D, Miao S, Chen S, Song L et al (2017) Design, synthesis and anticancer potential of NSC-319745 hydroxamic acid derivatives as DNMT and HDAC inhibitors. Eur J Med Chem 134:281–292. https://doi.org/10.1016/j.ejmech.2017.04.017

Yuan Z, Chen S, Gao C, Dai Q, Zhang C, Sun Q et al (2019) Development of a versatile DNMT and HDAC inhibitor C02S modulating multiple cancer hallmarks for breast cancer therapy. Bioorg Chem 87:200–208. https://doi.org/10.1016/j.bioorg.2019.03.027

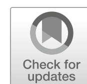
Zaina S, Heyn H, Carmona FJ, Varol N, Sayols S, Condom E et al (2014) DNA methylation map of human atherosclerosis. Circ Cardiovasc Genet 7(5):692–700. https://doi.org/10.1161/circgenetics.113.000441

Zhao S, Wu J, Zheng F, Tang Q, Yang L, Li L et al (2015) β-elemene inhibited expression of DNA methyltransferase 1 through activation of ERK1/2 and AMPKalpha signalling pathways in human lung cancer cells: the role of Sp1. J Cell Mol Med 19(3):630–641. https://doi.org/10.1111/jcmm.12476

Zhu D, Osuka S, Zhang Z, Reichert ZR, Yang L, Kanemura Y et al (2018) BAI1 suppresses medulloblastoma formation by protecting p53 from Mdm2-mediated degradation. Cancer Cell 33(6):1004–16.e5. https://doi.org/10.1016/j.ccell.2018.05.006

Zhubi A, Veldic M, Puri NV, Kadriu B, Caruncho H, Loza I et al (2009) An upregulation of DNA-methyltransferase 1 and 3a expressed in telencephalic GABAergic neurons of schizophrenia patients is also detected in peripheral blood lymphocytes. Schizophr Res 111(1–3):115–122. https://doi.org/10.1016/j.schres.2009.03.020

Zovkic IB, Sweatt JD (2013) Epigenetic mechanisms in learned fear: implications for PTSD. Neuropsychopharmacol Rev 38(1):77–93



Gene-Targeted DNA Methylation: Towards Long-Lasting Reprogramming of Gene Expression?

18

Fabian M. Cortés-Mancera, Federica Sarno, Désirée Goubert, and Marianne G. Rots

Abstract

DNA methylation is an essential epigenetic mark, strongly associated with gene expression regulation. Aberrant DNA methylation patterns underlie various diseases and efforts to intervene with DNA methylation signatures are of great clinical interest. Technological developments to target writers or erasers of DNA methylation to specific genomic loci by epigenetic editing resulted in successful gene expression modulation, also in in vivo models. Application of epigenetic editing in human health could have a huge impact, but clinical translation is still challenging. Despite successes for a wide variety of genes, not all genes mitotically maintain their (de)-methylation signatures after editing, and reprogramming requires further understanding of chromatin context-dependency. In addition, difficulties of current delivery systems and off-target effects are hurdles to be tackled. The present review describes findings towards effective and sustained DNA (de)methylation by epigenetic editing and discusses the need for multi-effector approaches to achieve highly efficient long-lasting reprogramming.

Keywords

Zinc finger · TALE · CRISPR-dCas9 · Epigenetic editing · DNMT · CpG methylation

18.1 Introduction

The epigenetic concept was first described by Conrad Waddington early in 1942, when he conducted experiments to understand phenotypic plasticity during embryonic development (Felsenfeld 2014). The definition has evolved over time to one of the current understandings of epigenetics as "the study of heritable changes in gene function that occur independent of changes in the primary DNA sequence"

Fabian M. Cortés-Mancera and Federica Sarno contributed equally.

F. M. Cortés-Mancera
Epigenetic Editing, Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Grupo de Investigación e Innovación Biomédica, Departamento de Ciencias Aplicadas, Instituto Tecnológico Metropolitano, Medellín, Colombia
e-mail: fabiancortes@itm.edu.co

F. Sarno
Epigenetic Editing, Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Dipartimento di Medicina di Precisione, Universita degli Studi della Campania "Luigi Vanvitelli" Napoli, Napoli, Italy
e-mail: f.sarno@umcg.nl

D. Goubert · M. G. Rots (✉)
Epigenetic Editing, Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
e-mail: m.g.rots@umcg.nl

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_18

(Nicoglou and Merlin 2017). The heritable modifications that epigenetics refer to correspond to biochemical changes on DNA and histone proteins. These changes influence the chromatin structure and thereby the expression of genes, even when the initial trigger has gone, and without underlying DNA sequence alterations. The main covalent chemical modification on the DNA molecule itself is methylation of cytosines, mostly in the context of CpGs dinucleotides (Petryk et al. 2021). Posttranslational modifications (e.g., methylation, acetylation), mainly on histone tails, provide another class of epigenetic signatures (Huo et al. 2021).

Strong observational evidence has been obtained on how epigenetic modifications associate with gene expression. To pinpoint an actual causative role of a particular epigenetic modification at a given genomic site, epigenetic editing tools have been exponentially exploited (de Groote et al. 2012; Jurkowski et al. 2015; Nakamura et al. 2021b). Epigenetic editing refers to the technology of actively rewriting epigenetic signatures at a genomic locus of interest. Towards this end, molecular tools have been generated (Jurkowski et al. 2015) consisting of a DNA-binding platform, which can be engineered to achieve locus-specific targeting, fused to an epigenetic effector domain (see Fig. 18.1). The first programmable protein-based DNA-binding platform used for endogenous gene targeting exploited the modular zinc finger (ZF) protein transcription factors, followed by transcription activator-like effectors (TALEs), and more recently the RNA-directed clustered regulatory interspaced palindromic repeats (CRISPRs) system (Stolzenburg et al. 2016).

ZF proteins, the largest group of naturally occurring transcription factors in the human genome, consist of approximately 30 amino acid-sized modules, each recognizing 3–4 bps in the major groove of double-stranded DNA (Sgro and Blancafort 2020). Mechanistically, the alpha-helix amino acids at positions -1, 3, and 6 can be engineered to recognize the third, second, and first base pair of a $5'-3'$ target sequence. Fusing together various of these modules resulted in effective tools targeting numerous genes in preclinical research and several ZF fusions have been clinically tested for ex vivo (and were the first tested in vivo (Ledhord 2018)) gene editing purposes. Next to their use as "*molecular scissors*" (when fused to nucleases), ZFs were used in pioneering studies of gene expression modulation by fusing transcriptional activators/repressors (Artificial Transcription Factors) to target a wide variety of endogenous genes (de Groote et al. 2012). The relatively compact size and scarce immunogenicity of ZFs are a major advantage compared to other DNA-targeting proteins.

TALEs provide another class of programmable DNA-binding tools and are derived from pathogenic bacteria that naturally modulate plant gene expression (Becker and Boch 2021). TALEs consist of individual protein modules that mediate binding to the target DNA site. Subsequently, transcriptional activators/repressors, or nucleases can be fused to the TALE DNA-binding domain for targeted gene expression modulation (Jain et al. 2021).

The more recent introduction of the versatile CRISPR-Cas9 system made gene targeting readily available for any laboratory with cloning facilities. CRISPR-Cas9 is derived from the bacterial defense system that recognizes foreign DNA. The nuclease activity of Cas9 is guided to a particular target sequence in the host genome via single-guide RNA (sgRNA)-DNA base pairing (see Fig. 18.1). As the DNA-binding specificity of earlier platforms (e.g., ZFs or TALEs) is provided by the engineered DNA-binding part within the fusions, for every new target sequence a new fusion protein needs to be designed. Target specificity of CRISPR-Cas9 is provided by separate sgRNAs, which are also simpler and less expensive to design, making this system much more flexible.

All three systems have been successfully exploited for epigenetic editing through the engineering of fusion proteins with epigenetic effector domains (Epi-editors) (Sgro and Blancafort 2020). In the case of CRISPR-Cas9, the epi-editor is cloned as a fusion to Cas9 proteins lacking the endonuclease activity (deactivated Cas9, dCas9). Upon delivery into target cells, the DNA-binding platform-fusion will bind to

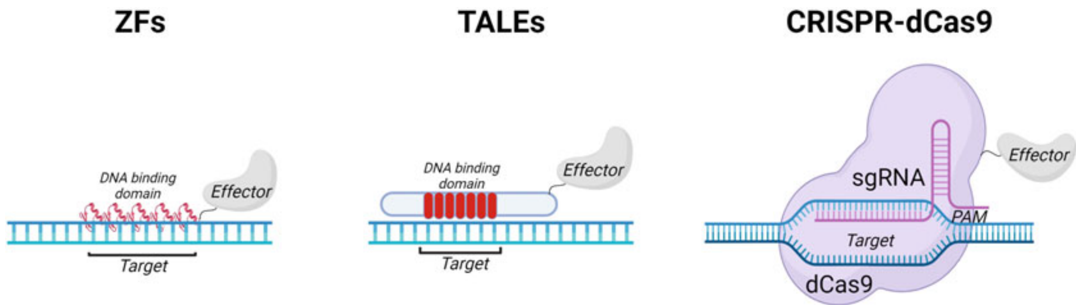


Fig. 18.1 Schematic representation of modular systems used in epigenetic editing. Epigenetic effector domains are recruited to the target DNA sequence by a DNA-binding platform: *ZFs* zinc finger proteins, *TALEs* transcription activator-like effectors or *CRISPR-dCas* the Clustered Regulatory InterSpaced Palindromic Repeat platform with *dCas9* deactivated Cas9 protein, *sgRNA* single-guide RNA, *PAM* proto-spacer adjacent motif. Figure made in https://biorender.com

the target sequence and exert its (enzymatic) activity. Initially, the assumed inaccessibility of heterochromatic genes, the unclear causative role of epigenetic marks on gene expression, as well as the unknown stability of edited marks was thought to hamper successful expression modulation of (silenced) genes. Pioneering studies and the general acceptance of CRISPR as a straightforward DNA-targeting approach, set the stage for the broad application of epigenetic editing as a research tool, e.g., to assess causative roles of epigenetic marks (Wang et al. 2021; Policarpi et al. 2021) and as potential therapeutic approach (Sgro and Blancafort 2020; Nakamura et al. 2021b).

The first well studied epigenetic mechanism is DNA methylation, predominately occurring on cytosine in the context of CpG (5mC), although methylation in non-CpG context has also been described (Ehrlich 2019). This epigenetic modification is important in stable (re)programming of expression patterns during development and cell differentiation, genome integrity and X chromosome inactivation, in health and disease (Ehrlich 2019; Petryk et al. 2021). In promoter regions, CpG dinucleotides often cluster in so-called CpG islands (CGIs), and more than half of the human gene promoters contain a CGI. These CpG-rich promoters are usually unmethylated, with a few exceptions, including tissue-specific methylation during development (Greenberg and Bourc'his 2019). Gene promoters with high levels of DNA methylation are generally transcriptionally inactive, while hypermethylated gene bodies generally associate with actively transcribed genes (Jeziorska et al. 2017).

DNA methyltransferase enzymes (DNMTs) generate this epigenetic mark. Specifically, DNMT1 is responsible for the methylation maintenance process coupled to DNA replication targeting hemimethylated strands (Petryk et al. 2021). DNMT3A and DNMT3B are capable of establishing new methylation patterns on previously unmodified cytosines, mainly in the CpG context. DNMT3L does not possess enzymatic activity but works as a coactivator of DNMT3A or 3B (Petryk et al. 2021). On the other hand, a family of enzymes called ten-eleven translocation proteins (TET1, TET2, and TET3) (Wu et al. 2018) possess dioxygenases activity that can convert methylated cytosine to 5-hydroxymethylcytosine (5hmC), followed by 5-formylcytosine (5fC) formation, and then 5-carboxylcytosine (5caC). Finally, 5fC and 5caC are removed by thymine DNA glycosylase (TDG), and cytosine is reestablished by base excision repair (BER) mechanism (Onodera et al. 2021).

Thanks to the programmable protein-based DNA-binding platforms, targeting (de)-methylation at specific loci is achievable and can be applied in a huge variety of physiological and

pathological contexts. A better understanding of factors that promote on-target epigenetic effects, and induce the desired long-lasting transcriptional states will facilitate further breakthroughs and the clinical application of epigenetic editing. Here, we will discuss findings on the use of epigenetic editing in exploring causative roles of DNA methylation and gene expression, with a specific focus on in vivo models and on the understanding of achieving long-lasting effects on gene expression levels.

18.2 Locus-Specific DNA Methylation Editing

18.2.1 Targeted DNA Methylation

Targeting DNA methyltransferases (MTase) to given genomic loci by epigenetic editing provides unique tools to investigate the causal role of DNA methylation in the modulation of gene expression (see Fig. 18.2), and to exploit this mechanism to combat diseases (Sgro and Blancafort 2020). The first proof of concept of targeted DNA methylation inhibiting gene expression was reported by Xu and Bestor in 1997, who constructed a fusion protein consisting of an engineered ZF and the prokaryotic DNA MTase *M.SssI* to induce DNA methylation on a p21 synthetic oligonucleotide promoter target (Xu and Bestor 1997). Several subsequent studies of targeted DNA methylation using human or bacterial DNA methyltransferases confirmed that induction of DNA methylation results in transcriptional repression in an exogenous system or non-mammalian genomes reviewed by us earlier (Stolzenburg et al. 2016). Genome-wide studies, however, pointed out that not all genes are equally permissive to methylation-induced gene silencing (Galonska et al. 2018; Broche et al. 2021). Moreover, cell heterogeneity, with even unexpected gene expression upregulation in response to DNA methylation editing, is incompletely understood (Vizoso and Van Rheenen 2021).

In 2012 and 2013, the endogenous repression of human genes by targeted DNA methylation was reported for the first time in two independent publications, targeting the vascular endothelial cell growth factor A (*VEGF-A*) promoter (Siddique et al. 2013), and *SOX2* and *MASPIN* oncogenes (Rivenbark et al. 2012). These studies used designed ZF proteins fused to the catalytic domain of the murine or human DNA methyltransferase 3A, respectively. The former report also demonstrated a twofold enhanced methylation activity by the fusion of DNMT3A and DNMT3L single chain dual effector (ZN-DNMT3A-3L) compared to ZN-DNMT3A alone (28 versus 14%, respectively). The increase is explained by the ability of the non-enzymatic DNMT3L to not only enhance the activity of other DNMTs, but also to recruit endogenous DNMTs (O'Geen et al. 2019). This synergy between DNMT3A and 3L was confirmed by various subsequent studies (Stepper et al. 2017, O'Geen et al. 2019, Tarjan et al. 2019; Nakamura et al. 2021a). Although the DNTM3A/3L fusion was frequently used in editing studies (Saunderson et al. 2017; Shayevitch et al. 2018; Hofacker et al. 2020), effective gene repression was also obtained by targeting DNMT3A catalytic domain (DNMT3A-CD) only (Bernstein et al. 2015; Vojta et al. 2016; McDonald et al. 2016; Qu et al. 2018; Josipovic et al. 2019; Tiane et al. 2021), or DNMT3A full length (Liu et al. 2016). Even targeting DNMT3L alone was sufficient to induce gene repression (O'Geen et al. 2019; Nakamura et al. 2021b), although not in all contexts (Amabile et al. 2016). Compared to the effective targeting of the long isoform (DNMT3A1) or the short isoform (DNMT3A2) using the dCas9-SunTag system (see Fig. 18.3), transient targeting of multiple copies of the catalytic domain (DNMT3A-CD) alone, resulted in no significant methylation or gene expression changes on *HOXA5* (Huang et al. 2017), indicating context-dependent effects. Compared to dCas9-DNMT3A, DNMT3B exhibited a lower methylation activity when targeted to the endogenous urokinase (*uPA*) promoter in HEK293T cells. Also for DNMT1, although DNMT1 recruitment had been shown to induce DNA methylation (Van et al. 2021), Lin and coworkers could not demonstrate changes in

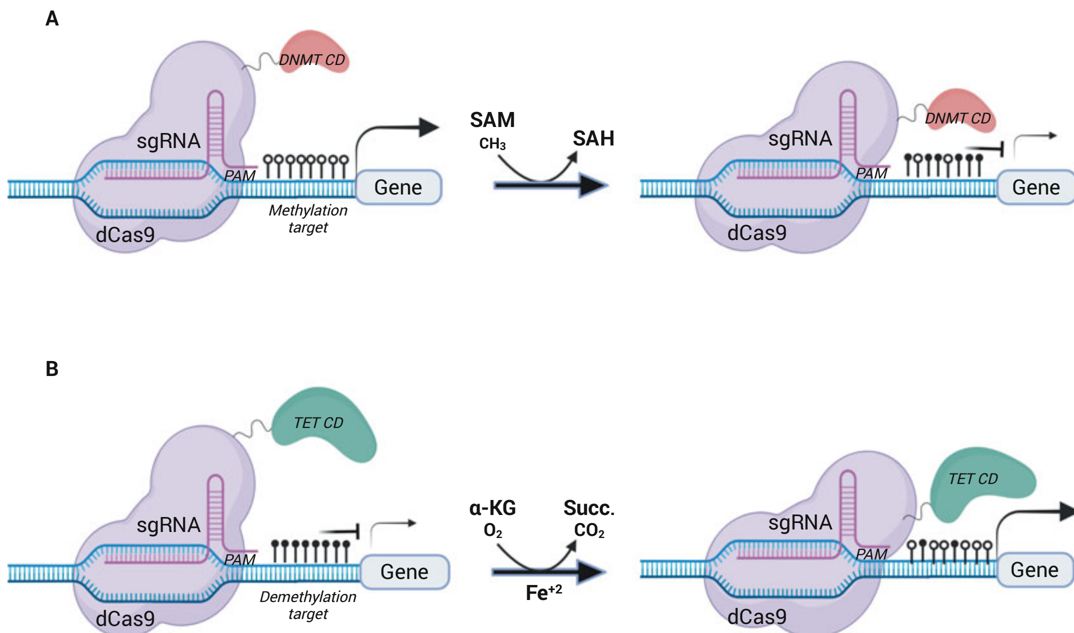


Fig. 18.2 Gene expression regulation via CRISPR-dCas9 targeting (de)methylation. (**a**) Representation of dCas9-DNMT (DNA methyltransferase) writing methylation at the target promoter region to induce gene expression downregulation. (**b**) Representation of dCas9-TET (Ten-eleven translocation methylcytosine dioxygenase) oxidizing (erasing) the methyl group at 5mC to induce gene re-expression. SAM *S*-adenosylmethionine, SAH *S*-adenosylhomocysteine, α-KG alpha-ketoglutarate, Succ succinate. Figure made in https://biorender.com

methylation levels in cells transfected to express dCas9-DNMT1 (Lin et al. 2018), suggesting that DNMT1 is less suitable for methylation editing.

The higher activity of DNMT3A was, however, also associated with off-target methylation. Although off-target effects can be sgRNA-driven (Zhang et al. 2015; McDonald et al. 2016), some studies indicate sgRNA-independent off-targeting (Lin et al. 2018; Galonska et al. 2018; Hofacker et al. 2020) via effector overexpression and/or interactions with endogenous de novo methylation enzymes. In this respect, Galonska and coworkers confirmed that increasing the pool of transduced sgRNAs spanning multi-loci regions to achieve simultaneous dCas9 recruitment did not reduce off-target effects (Galonska et al. 2018). Some reports described that increasing the efficiency of inducing local methylation (e.g., by dCas9-SunTag) improved the specificity (Huang et al. 2017; Pflueger et al. 2018). However, Hofacker and coworkers did not confirm improved specificity for the SunTag system when targeting *ISG15*, using the endogenous *VEGFA* promoter as an off-target reporter. Transfection of dCas9-DNMT3A-DNMT3L (dC) or dCas9-SunTag-DNMT3A/DNMT3L resulted in similar *ISG15* methylation levels (around 80%), while off-target *VEGF-A* methylation was higher for dCas9-SunTag (53%) versus dC (36%) (Hofacker et al. 2020). Therefore, constructs carrying different single mutations affecting DNA binding (K766E, K844E, R887E and R831E variants) were evaluated to improve methylation targeting specificity. Compared to wild-type dCas9-SunTag-DNMT3A/DNMT3L, residual on-target methylation activity of mutated effectors remained high (56 to 77% on *ISG15*), while methylation on *VEGFA* dramatically decreased. The R831E mutant provided the highest specificity with approximately 5% off-target methylation at the *VEGFA* promotor versus around 50% for the

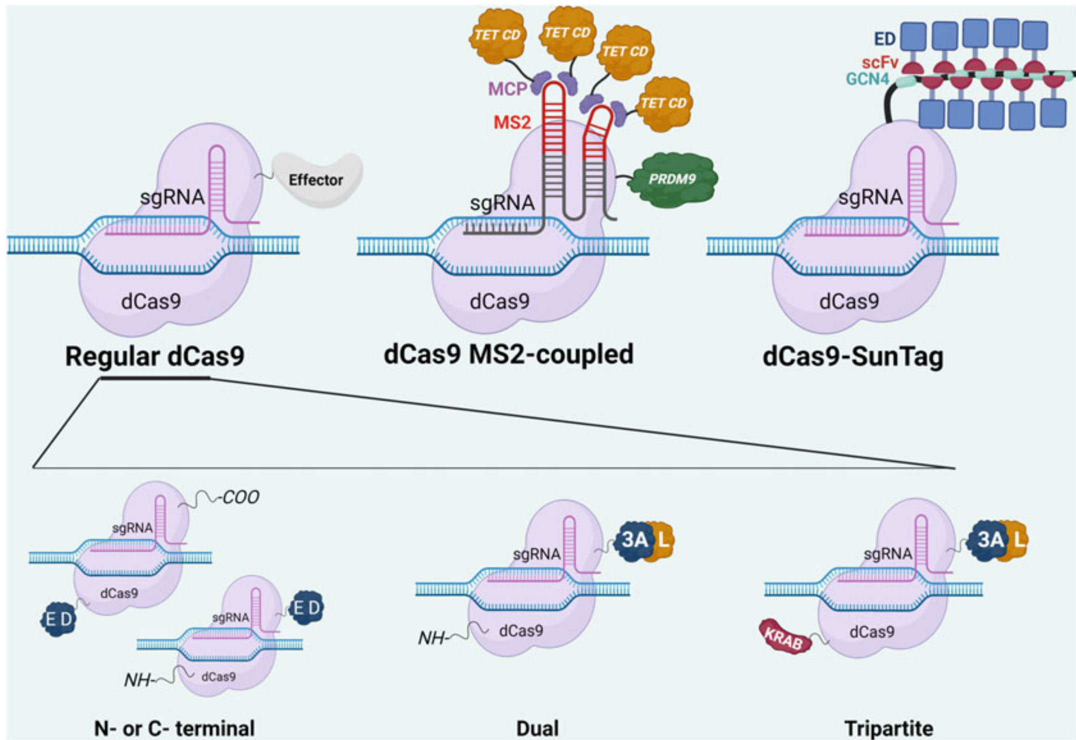


Fig. 18.3 Representation of enhanced CRISPR-dCas9 tools. At the top, three commonly used dCas9 tools. ***dCas9 MS2-coupled***: sgRNA is engineered to harbor RNA motifs (**MS2**) that can be recognized by RNA-binding proteins (**MCP**) fused to epigenetic effector domains such as TETCD to synergize with, for example, dCas9-PRDM9 writing H3K4me3 (Cano-Rodriguez et al. 2016). ***dCas9-SunTag***: dCas9 is fused to **GCN4** repeats that can recruit effector domains (**ED**) fused to a GCN4 recognizing single chain antibody (**scFv**) (Pflueger et al. 2018). At the bottom, options in effector configuration diversity (N/C-terminal orientation, different numbers/ combinations) are shown. Figure made in https://biorender.com

wild-type enzyme, as confirmed using a genome-wide approach.

Other strategies to reduce off-target methylation include the usage of the prokaryotic MTase *M.SssI* variant MQ1^{Q147L} that does not recruit endogenous mammalian DNA methyltransferases, and demonstrated less off-target effects compared to wild type at endogenous loci (Lei et al. 2017). Alternatively, a split version of the *M.SssI* MTase was shown to generate efficient targeted DNA methylation, with less off-target effects when compared to dCas9 fused to full-length *M.SssI* (Xiong et al. 2017). Recently, Ślaska-Kiss and colleagues studied *M.SssI* variants fused to zinc fingers or dCas9, and demonstrated in *E. coli* cells that methylation specificity on plasmids was predominantly influenced by mutations affecting catalytic activity rather than DNA-binding affinity of the MTase domain (Slaska-Kiss et al. 2021).

To further improve the toolbox of targeted methylation, spatiotemporal control has been exploited to enhance site specificity by cloning light-inducible protein pairs to DNA-binding modules and to a DNA methyltransferase. Indeed, Lo and coworkers engineered DNMT3A-CRY2-EGFP and TALE-CINB1-mCherrry constructs to control *Ascl1* promoter methylation changes by exploiting the optogenetic blue light inducible dimerizing of cryptochrome-2 (CRY2) and its interacting protein (CIB1). Upon blue light exposure, DNMT3A-CRY2 paired to TALE-CINB1

and effectively induced highly specific DNA methylation and subsequent decrease in gene expression (Lo et al. 2017).

18.2.2 Targeted DNA Demethylation

To exploit the reversibility of DNA methylation in a gene-targeted manner, Ten–eleven translocation (TET) dioxygenase enzymes offer unique tools for DNA demethylation (see Fig. 18.2). Using the ZF or TALE platforms, the first TET-editing reports compared the potency of the three different TET domains (Chen et al. 2014), and demonstrated the improved efficacy of the catalytic domain (CD) over full length (Maeder et al. 2013), in inducing active DNA demethylation and subsequent transcriptional upregulation. Using CRISPR-dCas9, effective demethylation was further demonstrated for various genes (Choudhury et al. 2016; Amabile et al. 2016; Xu et al. 2016; Okada et al. 2017), and the approach was rapidly translated to in vivo models as described hereafter (Liu et al. 2016; Morita et al. 2016; Xu et al. 2018; Ou et al. 2019; Wang et al. 2019; Horii et al. 2020; Hanzawa et al. 2020).

The SunTag system (Morita et al. 2016) and MS2 elements inserted into sgRNAs (Xu et al. 2016) were shown to enhance the effect of targeted demethylation via dCas9-TET (see Fig. 18.3). Also combining TET demethylation activity with VP64 activation showed promise, as demonstrated for *CDKL5*, a gene causative for an infantile epilepsy in human neuronal-like cells (Halmai et al. 2020). As known from literature, a significant number of X-linked genes escape from X chromosome inactivation and are associated with a distinct epigenetic signature like reduced DNA promoter methylation. Halmai and coworkers created such escape by removing DNA methylation on the promoter of the *CDKL5* promoter. The dCas-TET1 targeting caused a significant reactivation of the inactive allele (Halmai et al. 2020), which was further improved by dCas9-TET1 and dCas9-VP64 co-treatment resulting in reactivation of the inactive allele to levels of >60% of the active allele. This artificial escape study confirmed earlier observations of synergism between TET and transcriptional activation domains, such as VPR (VP64-p65-Rta). Interestingly, despite a more effective demethylation by TET alone compared to the combination, a synergism with respect to increased re-expression of *Hnf1a* was observed (Josipovic et al. 2019).

Targeted demethylation of DNA can also be induced using the plant-derived ROS effector (Devesa-Guerra et al. 2020) or Thymidine DNA Glycosylase (Gregory et al. 2013) or even by dCas9 alone or with an inactive enzyme as recently demonstrated (Sapozhnikov and Szyf 2021). The latter authors studied several proximal promoters, including the hypermethylated *IL33* gene. After transient transfection experiments, dCas9-TET or a catalytically inactive mutated version (dCas9-dead-TET) caused hypomethylation and induction of *IL33* gene expression, suggesting a mechanism independent of TET oxygenase activity. The authors demonstrated that hypomethylation was related to DNMTs blockage, which is consistent with previous reports that showed mild hypomethylation induced by binding of dCas9-TET catalytically inactive mutants (Maeder et al. 2013; Xu et al. 2016; Morita et al. 2016), as is also known to occur upon binding of some transcription factors (Suzuki et al. 2017). Similarly, for engineered ZFs, hypomethylation was observed for targeted CpGs (Chen et al. 2014; Huisman et al. 2016). Sapozhnikov and Szyf also highlighted some important aspects with respect to promoter methylation and gene activation: demethylation of CGG repeats in the *IL33* promoter region resulted in gene re-expression, while demethylation in the proximal promoter region of other genes was not enough to induce their expression (e.g., *SERPINB5*, *TNF*). These genes required demethylation also of other regions (*cis* or *trans*) to induce gene expression. Such data illustrate the importance of studying demethylation of specific sites to better understand their relative contribution to gene expression and cause-effect dynamics. Moreover, despite effective demethylation and re-expression, the cellular functional effects might not be as expected, as was the case for

dCas-TET1 induced re-expression of *FoxP3*: despite an effective increment in *FoxP3* gene expression, no increase in the functional regulatory T cell population was observed (Kressler et al. 2020).

18.3 Sustained Transcriptional States upon DNA Methylation Editing

18.3.1 Long-Lasting Transcriptional Repression

Given the maintenance of DNA methylation during cell division (and the for a long time presumed absence of active DNA demethylases), CpG methylation was initially considered a stable epigenetic mark associated with persistent silencing (Petryk et al. 2021). Currently, it is generally accepted that also this epigenetic signal is highly dynamic.

To evaluate the long-term effect of dCas9-DNMT3A without interference from the endogenous DNMT enzymes, Galonska and coworkers made use of DNMT3A/B double knockout (DKO) embryonic stem (ES) cells and DNMT1 transient repression (Galonska et al. 2018). Transient induction of dCas9-DNMT3A increased global methylation in DKO cells with a preference for hypermethylated elements or H3K27ac-enriched regions in wild-type ES cells, such as exons and repetitive elements. In contrast, unmethylated sites, such as CpG islands associated with transcription start sites, remained generally hypomethylated (Galonska et al. 2018). In these maintenance competent cells, methylation was only retained at a subset of lowly transcribed genes after 7 days post-transfection at regions devoid of histone 3 K4me3 (Galonska et al. 2018). Also in wild-type HEK293 cells, where DNA methylation was written at thousands of CGIs upon 3 days of doxycycline-induced ZN-DNMT3A expression, the introduced methylation was rapidly lost at most of them (90%) (Broche et al. 2021). The partially stable methylated CGIs (∼1000) were enriched in H3K27me3, reduced in H3K4me3 and H3K27ac, and without differences in K9me3, confirming a role for the native chromatin contexts determining permissiveness for stable editing (see Fig. 18.4).

The first pioneering studies already indicated the context-dependency of maintenance of DNA methylation (Stolzenburg et al. 2015; Kungulovski et al. 2015; Vojta et al. 2016). Stolzenburg and coworkers reported a persistent tumor repression linked to sustained DNA methylation on the *SOX2* oncogene promoter using ZF-DNMT3A effector in breast tumor cells, which was not observed for the ZF-KRAB fusion. Comparing different epigenetic effector domains (EED, DNMT3B, HDAC4) with the transcriptional repressor KRAB, also Bintu and coworkers demonstrated differential dynamics of repression, with epigenetic modulators being relatively ineffective also long-term, except for DNMT3B that induced sustained silencing up to 30 days (Bintu et al. 2016). Vizoso and van Rheenen provided evidence that targeted methylation of DNA, introduced by CRISPR-dCas9-DNMT3ACD, can be inherited by daughter cells for over 48 cell divisions. The authors used methyl-specific PCR (MS-PCR) to follow up sorted single clones, and bisulfite sequencing to confirm, and indicated long-term DNA methylation for 14 out of 18 clones at day 22. Two of these HEK293 clones, randomly selected, were again clonally expanded and the 24 subclones mostly maintained methylation values after an additional 22 days of culture. Taking advantage of dCas9 system coupled to DNMT3ACD plus C-terminal DNMT3L effector (dCas9-3ACD-^{C}3L) Saunderson and coworkers targeted the p16 promoter in primary breast cells. Also here, up to 35 days post-transient transfection, maintenance of p16 CpG hypermethylation and transcript downregulation was demonstrated when compared to dCas9-3ACD-^{C}3LΔ mutant, with sustained effects on cell proliferation and senescence processes (Saunderson et al. 2017).

Yet, writing DNA methylation does not necessarily result in long-term effects (Kungulovski et al. 2015; McDonald et al. 2016; Broche et al. 2021). Rewriting a combination of classes of epigenetic marks might provide a synergistic

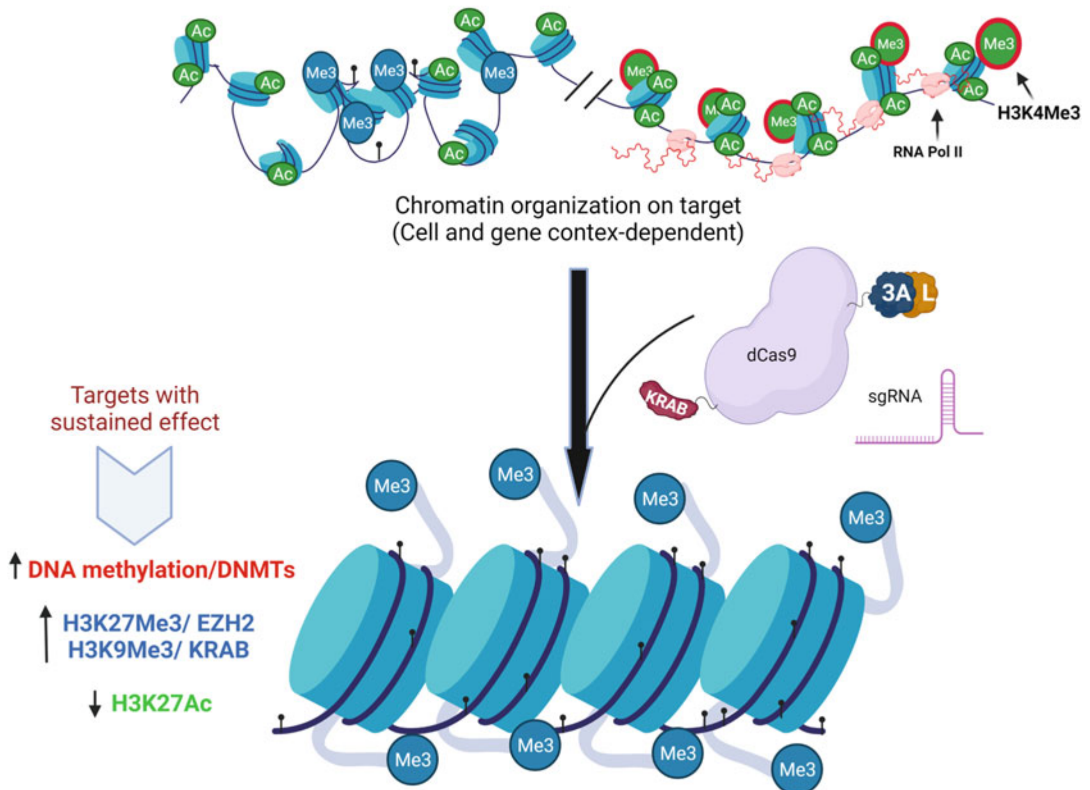


Fig. 18.4 Writing epigenetic marks to induce sustained transcriptional effects. A tripartite CRISPR-dCas9 configuration used for epigenetic long-lasting effects (KRAB-dCas9-DNMT3ACD/3L) is shown. Target genes showing mitotically sustained transcriptional reprogramming are commonly correlated with increase in DNA methylation and repressive marks on histone 3: H3K27Me3 or H3K9Me3, for EZH2 or KRAB effectors, respectively. Black hairpin decorations represent DNA methylation. Figure made in https://biorender.com

and more predictable approach towards inducing sustained silencing for subsets of genes. In this respect, an elegant system based on endogenous recruitment of epigenetic players at specific loci by nanobodies (single-domain antibodies), demonstrated that co-recruitment of DNMT1 synergistically improved the sustained downregulation of a reporter gene, induced by KRAB, DNMT3A, HP1 or HDAC4 (Van et al. 2021). The first proof of the combinational enhancement via targeting KRAB and de novo DNMT3A and DNMT3L effectors was described by Amabile and coworkers who demonstrated sustained silencing of three somatic genes (Amabile et al. 2016). Tarjan and coworkers demonstrated that dCas9-KRAB, dCas9-DNMT3A or dCas9-DNMT3A3L can selectively displace the protein insulator CTCF, with dCas9-KRAB achieving 83% of CTCF binding reduction, but the effect was not sustained. When dCas9-DNMT3A or dCas9-DNMT3A3L were transiently transfected, 20–40% of DNA methylation was detected over the targeted CTCF motif, with DNMT3A3L being more effective than DNMT3A (Tarjan et al. 2019). Here, the DNA methylation on the CTCF motif persisted (~20%) upon serial passage (12 days), when the dCas9 fusions were no longer detected, congruent with ~20% reduction in CTCF binding. Again, combined treatment with single chain double effector dCas9-DNMT3A3L plus dCas9-KRAB resulted in an enhancement of CTCF displacement and in a longer sustained response (up to 27 days) (Tarjan et al. 2019).

Similarly, other reports confirmed the effectiveness of co-targeting KRAB, DNMT3A and 3L effectors to achieve sustained epigenetics changes (Mlambo et al. 2018; Nakamura et al. 2021b; Nuñez et al. 2021), however again not all genes were responsive (Mlambo et al. 2018; Nuñez et al. 2021). Using a genome-wide screen and growth as read-out, Nuñez and coworkers indicated the general applicability of transient CRISPR-Off treatment (DNMT3A, DNMT3L and KRAB fused to one dCas9 protein) to induce effective and persistent gene silencing. Interestingly, although the long-lasting silencing was not obtained for all genes, CRISPR-Off was even effective for genes lacking canonical CpG islands or with a low CpG density (Nuñez et al. 2021).

To investigate the mechanisms of maintenance in more detail, Nakamura and coworkers generated a stable cell line (HEK293T) with GFP expression under the SV40 promoter regulation, and SV40-targeting guide RNAs. This reporter allowed to evaluate gene expression effects without the context-dependent restrictions of endogenous targets, which affect accessibility and activity of CRISPR-dCas9 (Nakamura et al. 2021b). Plasmids were transiently transfected, individually or combined (dCas9-KRAB, DNMT3A, and DNMT3L) to determine the best combination and the optimal positional configuration. To evaluate the long-lasting reprogramming, cells were cultured and periodically harvested up to 30 days post-transfection with Zeocin treatments for effector enrichment during these experiments. dCas9-KRAB significantly repressed GFP expression shortly after transient transfection, with subsequent recovery of expression at longer time scales. DNA methyltransferase domains individually exhibited minor ability to generate stable silencing. When cells were cotransfected using all three dCas9 effectors, a strong reduction in GFP expression was observed for weeks post-transfection. DNA methylation analysis showed a localized hypermethylation around the TSS and more extended repressive histone marks (H3K9me3) +/– 500 bp. After experimental pairwise domain analysis and testing modular swapping combinations, Nakamura and coworkers demonstrated that C-termini configuration for DNMTs, with first DNMT3L followed by 3A, was more effective for silencing. The addition of KRAB at the N-terminus showed the highest levels of stable gene repression, and KRAB swapping by SID effector (small temporary repression), or ZIM3/KRAB effector (twofold greater maximal repression) did not further improve sustained gene repression.

Exchanging KRAB for Ezh2 (Enhancer Zeste Homolog 2) did proof effective for a gene unresponsive to KRAB/3A/3L combinations: O'Geen and coworkers confirmed that combinatorial treatment with KRAB amino-terminal fused (KRAB-dCas9) and DNMT3A-dCas9 combined with ectopically overexpressed DNMT3L was able to initiate long-term repression for six out of seven targeted genes (O'Geen et al. 2019), but the combination failed to maintain persistence at *HER2* in HCT116 cells. The dCas9 treatment combinations (KRAB + DNMT3A + DNMT3L), triggered a strong burst in H3K9me3 at the target locus, but the repressive H3K9me3 mark was completely lost after 24 days. On the other hand, histone methyltransferase Ezh2 co-treatments (Ezh2-dCas9 + DNMT3A + DNMT3L) led to a long-term *HER2* repression (O'Geen et al. 2019), with both DNA and histone methylation (H3K27me3) marks maintained through approximately 57 cell divisions. Interestingly, full-length DNMT3L was essential for Ezh2-dCas9 mediated long-term repression, and the Carboxy-terminal hybrid dCas9-DNMT3L lacking the ADD domain fused to the DNMT3A catalytic domain (dCas9-DNMT3A/L) was unable to establish long-term epigenetic memory. This report again indicated that DNA or histone methylation alone are not always sufficient for long-term repression, but that the combination of epigenetic marks is important for predictable establishment and maintenance of epigenetic memory.

18.3.2 Sustained Gene Re-expression

Long-lasting effects on gene modulation via actively inducing locus-targeted DNA

demethylation have also been reported. For example, Nakamura and coworkers assessed the possibility of dCas9-reprogrammed genes to be reactivated by transient expression of various dCas9-fusions, including dCas9-TET1 and TET2. Five days post-transfection, dCas-VP64, -VPR and -p300 demonstrated the strongest gene reactivation, with negligible effect for most of the tested epigenetic effectors (full length or catalytic domains; cloned at dCas9 N- or C-termini) (KDM3A, KDM4D, KDM7B, TDG). Targeting dCas9-TET1 and -TET2 did induce GFP re-expression, and more importantly, this re-expression was stably maintained for up to 60 days, while the dCasVP64, -VPR and -p300 reactivation was transient (Nakamura et al. 2021b). Also in CHO cells, Marx and coworkers demonstrated that by using dCas9-SunTag-TET1CD targeting a constitutively silenced gene (Beta-galactoside alpha-2,6-sialyltransferase 1 -*ST6GAL1*), a stable re-expression for more than 80 days was achieved (Marx et al. 2018). A stable reactivation induced by transient dCas9-TET1-CD expression was also confirmed for an enhancer involved in *FOXP3* expression in human T-cells, although this persistent demethylation status was not sufficient to induce a stable CD4+ regulatory T-cells (Tregs) phenotype (Kressler et al. 2020). In this respect, Okada and coworkers demonstrated that despite a partial lentiviral TET1-induced demethylation of this enhancer region of *Foxp3*, no stable gene expression was induced in mouse primary T-cells, while promoter-targeted dCas9-p300 did result in partially maintained *Foxp3* expression and functionality (Okada et al. 2017). Also for *Fgf21*, DNA promoter re-methylation occurred as measured 14 days after scFv-TET1CD transient transfection (Hanzawa et al. 2020). So, not all genes were equally permissive to sustained re-expression by targeting TET alone.

In fact, sustained re-expression was obtained only after simultaneous targeting of TET1-dCas9 and VPR-dCas9, inducing a persistent upregulation up to 30 days which was not achieved for either dCas9-fusion construct alone (Josipovic et al. 2019). Also Nuñez et al. demonstrated that combinations of TET1-dCas9 recruiting p65-AD (activation domain of NFkB subunit) and/or Rta (transcriptional activation domain of Epstein-Barr virus) via the MS2 system increased effectivity of targeting TET1 in re-expressing genes earlier silenced by KRAB-3A3L CRISPR-off (Nuñez et al. 2021). This study again elegantly showed that repressive epigenetic states can readily be reverted using epigenetic editing in a sustained manner.

18.4 In Vivo Transcriptional Modulation via DNA Methylation Epigenetic Editing

The technology of genome editing is rapidly advancing into the clinic with over 40 ZNF, TALEN and CRISPR-Cas9 studies ongoing (https://clinicaltrial.gov). Although mainly ex vivo, the first in vivo studies have been initiated making use of lentiviral vectors or AAVs (Adenoviral Associated Vectors). Since inducing mutations in the human genome, however, is subject of societal debate, epigenetic editing, which maintains integrity on the genome sequence without introducing mutations, is explored as a more versatile and less invasive approach, with potentially equal efficiency. Despite similar limitations, including off-targets and delivery effectivity, in vivo preclinical transcriptional modulation studies have shown therapeutic effectiveness. Indeed, artificial transcriptional factors (targeting KRAB, VP64, e.g., in CRISPRi/a (Geel et al. 2018; Nakamura et al. 2021a) have induced gene expression modulation in vivo, such as gene silencing in mouse brains (Zheng et al. 2018), or activation (Bustos et al. 2017), also in mouse models of muscular dystrophy/diabetes (Liao et al. 2017), cancer (Kretzmann et al. 2019) or obesity (Matharu et al. 2019). Unless stably expressed, such agents are thought to act transiently. Using gene targeting platforms to induce epigenetic modifications of DNA and histones bears the promise for gene expression modulation to be maintained for a long time. However, only few

studies actually examined the in vivo effects of epigenetic writer or eraser effector domains (Gomez et al. 2019).

As discussed already in this review, aberrant DNA methylation is associated with disease development. Despite large and ongoing efforts of the scientific world to demonstrate that modulating DNA methylation interferes with dysregulated gene expression profiles, clinical applications of interfering with DNA methylation are limited to two inhibitors of DNMTs (azacitidine (Vidaza) and decitabine (Dacogen)), which are FDA approved to treat hematological malignancies. However, these hypomethylating agents have some limitations, including a low response rate, short duration of action, and lack of specificity (Berdasco and Esteller 2019). Gene-specific DNA (de)methylation tools are thus important in assessing the causal correlation between DNA methylation status, biological function and disease development. Additionally, DNA methylation editing tools open interesting avenues to, e.g., compensate for genetic mutations, prevent therapy resistance or otherwise interfere with pathophysiology. Eventually, investing in effective DNA methylation editing techniques gives therapeutic possibilities for the numerous diseases related with aberrant up- and downregulated gene expression levels.

The few in vivo DNA methylation epigenetic editing studies available to date, described below, show promising effects, demonstrating its exciting application to create innovative disease models as well as its potential therapeutic role in the clinic. The first published mouse studies made use of injecting stable, ex vivo transduced, inducible ZF-DNMT3a expressing tumor cells. These xenograft models clearly demonstrated the correlation between tumor growth and methylation state of either the *p16* (Cui et al. 2015) or the *SOX2* (Stolzenburg et al. 2015) promoters. Similarly, the role of *Crmp4* in inducing metastases was demonstrated in prostate cancer with all control mice developing metastases, whereas 8 out of 9 animals injected with prostate cancer cells expressing a TALE-TET1 fusion designed to target the gene did not (Li et al. 2015). Using the CRISPR system, a putative tumor suppressor gene was functionally validated in a colon cancer mouse model. Targeting TET1CD to the *SARI* promoter resulted in specific demethylation and substantial gene activation of *SARI*, which is frequently downregulated in several cancers (Wang et al. 2019). Injection of transfected cancer cells into the flank of nude mice resulted in smaller tumors compared to the controls, and less angiogenesis was observed as well. Although delivery issues hamper clinical translation of such methylation editing approaches in oncology, these tools offer unique opportunities to create disease models to better understand cancer biology (Weichenhan et al. 2020).

Before the adoption of epigenetic editing, no tools were available to directly demonstrate the correlation between epigenetic changes and disease. In recent years, the DNA methylation editing approach has gained attention to create epigenome-modified animals to explore epimutations in (epigenetic) diseases. For example, to understand the role of aberrant expression of the *H19-Igf2* genes, regulated by allele-specific DNA methylation in Silver-Russell syndrome (SRS), an imprinting mouse model was created by demethylating the paternally imprinted allele (Horii et al. 2020). In this study, three different methods were compared for efficiency: reprogramming ESCs, transient transfection or stable integration of the editor-expression cassette in fertilized ovules.

The first method involved transient transfection of ESCs with dCas9-SunTag/scFv-GFP-TET1CD implanted in the uterus after 4 weeks. Even though the extent of demethylation in almost all the animals obtained was higher compared to the other two methods (75% of target sequences were demethylated), the epigenetic changes of the genomic imprinting induced by the editing were not stably inherited. The second method generated animals by transient transfection of epigenetic editor mRNA into fertilized eggs. Compared to the previous one, this approach is applicable to most animal species. However, the modification observed at the blastocyst stage was low in frequency as well as in degree of demethylation, reflecting the instability of the reprogrammed epigenetic signature in vivo.

The third approach was based on continuous modification of the epigenome of animals by stable expression of epigenetic editors by transgenes introduced at the Rosa26 locus in fertilized ovules. Although a lower percentage (50–67%) of newborn mice as compared to the first method showed transgene integration, the integration was associated with significant demethylation at seven CpG sites in the H19-DMR promoter region. Importantly, these epigenetic changes were inherited by the next generation, creating an SRS mouse model. Comparison of the three mouse models generated demonstrated that stable integration upon dCas9-ED-sgRNA delivery is a realistic approach with a high percentage of vector-integrated animals, which showed a constant expression of the epi-editor over time. However, off-target effects are a serious problem. In fact, the stable expression of epigenome-modifying factors induced DNA demethylation in two predicted off-target regions for gRNA of H19DMR_10 (2 mismatches) and H19DMR_11 (2 mismatches). This indicates that this approach could increase the risk of off-target epigenome modification.

Alternatively, zygote microinjection of CRISPR-dCas9 tools has been used to create animal models of imprinting (Lei et al. 2017) and neurological (Lu et al. 2020) disorders. In the first, in vivo locus-specific DNA methylation was inherited for up to 3 weeks from mouse birth. Targeting CpGs of the imprinted locus of *Igf2/H19* in mice, dCas-MQ1^{Q147L} stably increased DNA methylation demonstrating the possibility to modify the methylation status of a specific gene in the early stage of embryonic development, which was maintained during cellular differentiation processes (Lei et al. 2017). This is a clear demonstration of the potency to use dCas9-MQ1^{Q147L} to introduce site-specific DNA methylation with high activity and specificity. It suggests its broad applications for the study of gene dysregulation in various disease contexts.

Zygote microinjection was also used to create a disease model for autism spectrum disorders (ASD): targeting *Mecp2* by microinjecting dCas9-DNMT3A/3L decreased the expression of *Mecp2* resulting in ASD behavior as measured up to 8 weeks after birth. These data demonstrated that DNA methylation at the *Mecp2* promoter contributes to ASD pathology and suggest that changing *Mecp2* gene expression improves treatment outcomes in individuals with ASD. The authors also applied AAV infection to express dCas9-DNMT3A/3L in the hippocampus, thereby highlighting epigenetic editing opportunities for therapeutic intervention (Lu et al. 2020).

Effective interference using epigenetic editing was also demonstrated at a later developmental stage (in utero). dCas9-SunTag-TET1CD was successfully introduced in isolated neural precursor cells (NPCs) from mouse embryos by electroporation to reactivate the expression of *Gfap* in order to induce the differentiation of NPCs into astrocytes (Morita et al. 2016). As one cytosine in the *Gfap* gene promoter is methylated in most cell types, except for astrocytes, targeted demethylation of this site was hypothesized to play a critical role in the differentiation of NPCs into astrocytes. Implantation of transfected NPCs into the ventricular zone of mouse fetal brain in utero resulted in increased expression of *Gfap*. With this article, the authors demonstrated the feasibility of implanting functionally reprogrammed cells in vivo early in development.

Using a lentiviral delivery approach, Liu and coworkers confirmed the possibility to effectively alter the methylation status and regulate the expression of a neurological gene in adult mice. Microinjection of dCas9-TET1 in the brains of GFP-transgenic mice to induce demethylation of the *Snrpn* promoter driving GFP resulted in 70% activation of GFP (Liu et al. 2016). This study set the stage to address Fragile X syndrome (FXS), the most common form of mental disability, associated with methylation-induced silencing of the *Fmr1* gene. To date, there is no effective cure for this disease. FX52 neuronal precursor cells (NPCs) were infected to express dCas9-TET1 targeting *Fmr1*, and then implanted in newborn mice brains, to study the effect of DNA methylation on *Fmr1* gene expression in vivo (Liu et al. 2018). In mice lacking *Fmr1* expression, dCas9-TET1 opened the heterochromatin state of the

Fmr1 promoter region, inducing its expression up to 1–3 months after NPCs transplantation. The increase in gene expression restored the normal condition of FXS neurons, reversing the abnormal electrophysiological phenotype, which is close to a possible therapeutic application (Liu et al. 2018). These results, retained in adult mice upon implantation in newborns, open new possibilities in this field, not only to better understand the physiology of the disease, but also to investigate its use as a potential therapeutic approach.

DNA methylation editing findings further demonstrate that epigenetic mechanisms drive pathology in neurodevelopmental disorders and confirm various neuroepigenetic editing studies using other epigenetic effector domains (Xu and Heller 2019), even in inducing differential splicing (Xu et al. 2021), which point out the use of epigenetic editing as a promising therapeutic approach for neurodevelopmental disorders. Other pathophysiologies addressed in in vivo DNA methylation editing studies concern metabolic disorders (Ou et al. 2019; Hanzawa et al. 2020) and fibrosis (Xu et al. 2018). To further understand the role of DNA demethylation on the obesity-related fibroblast growth factor 21 (*Fgf21*) gene expression in the liver, dCas9-SunTag and scFv-TET1CD were introduced into the liver of PPARα-KO mice by hydrodynamic injection into the tail vein (HTVi) (Hanzawa et al. 2020). PPARα, a nuclear receptor regulating the transcription of major genes related to hepatic metabolism, is thought to induce *Fgf21* expression via DNA demethylation, but the exact mechanism is unclear. The use of non-specific DNA methyltransferase inhibitors that demethylate the genome globally only indirectly helps to understand such specific gene regulation. Epigenetic editing, uniquely suited to address a single gene, allowed to unravel the role of epigenetic regulation mechanisms. The *Fgf21* PPARα-KO model validated that altered DNA methylation of *Fgf21* is indeed causally related to the biological activation.

Another in vivo study addressing metabolic diseases exploited the TALE platform to target TET1 to the methylated promoter of *ICR2* gene, which upon re-expression repressed p57, inducing growth of β cells, which are dysfunctional in diabetes (Ou et al. 2019). Transplantation of the TALE-TET1 expressing β cells was shown to increase proliferation, and this ex vivo approach comes very close to a possible therapeutic application for diabetic patients.

Although the above DNA demethylation in vivo studies exploited TET1 as effector domain, TET3CD was also successfully used to induce the reactivation of *Rasal1* and *Klotho* in interstitial fibroblasts and in renal tubular epithelial cells, respectively, in the unilateral ureter obstruction mouse model of nephropathy. Both genes are highly hypermethylated in these cells and their downregulation is associated with fibrosis. Using lentiviral delivery (intraparenchymal for *Rasal1*, ureter retrograde for *Klotho*), a high-fidelity dCas9 fusion (dHFCas9-TET3CD) decreased off-targets by 85% compared to conventional dCas9. Targeting the two fibrotic genes led to a reduction of 50% and 25% in the production of fibroblasts, respectively and subsequently reduced renal fibrosis (Xu et al. 2018). Combined with the ongoing efforts to improve maintenance, specificity and delivery, more in vivo preclinical studies are expected to further spark the interest for epigenetic editing, not only in providing potent disease models, but to be considered as a versatile therapeutic approach in the fight against currently uncurable diseases.

18.5 Further Considerations

Application of epigenetic editing technology in human health is desirable, as it opens novel avenues for diseases where currently no treatment or cure options are within sight. Clinical translation, however, is still challenging, although ongoing developments in applying CRISPR-Cas gene editing will certainly pave the way in overcoming delivery and off-target issues. Viruses are frequently used for efficient delivery. To circumvent the potentially harmful host genome integrations by lentiviruses, AAVs have been shown to effectively deliver dCas constructs (Thakore et al. 2018; Kemaladewi et al. 2019; Lu et al. 2020;

Matharu et al. 2019) and to exhibit low immunogenicity (Levy et al. 2020; Wu et al. 2021). Despite this, AAVs come with some limitations specific to epigenetic editing. The size of AAV restricts its application for in vivo epigenetic editing due to the inability to carry large transgenes needed to encode the fusions of the epigenetic effector domains (Colella et al. 2018). Based on the hit-and-run promise of epigenetic editing (Amabile et al. 2016; Saunderson et al. 2017), episomally maintained AAVs might not be needed for effective therapeutic effects and transient administration of proteins directly (Bailus et al. 2016) or by, e.g., lipid nanoparticles containing protein/RNA/DNA could thus be useful for future applications with effectivity shown in the first in vivo CRISPR-Cas9 trial (Gillmore et al. 2021). Indeed, advances were obtained in delivery technologies, with physical (electroporation, microinjection), chemical (lipids, polymers, nanomaterials) and biological alternatives, besides (viral) vectors. As alternative to using DNA as cargo, direct delivery of the sgRNA and dCas fusion mRNA (or protein as ribonucleoprotein (RNP)) is a very interesting and promising approach for in vivo application delivery (Wei et al. 2020; Qiu et al. 2021) as lower controllable cellular levels might reduce the off-target effects. Delivery systems based on extracellular vesicles (EVs) have shown to be an interesting approach for therapeutic genome editing (Chen et al. 2021). Also for CRISPRa delivery, applicability of EVs as vehicles has been demonstrated in mice by incorporating sgRNA and dCas9 proteins (Lainscek et al. 2018). More recently, further preclinical proof of EV-mediated delivery of CRISPR-dCas9-VP64 was reported for liver fibrosis treatment (Luo et al. 2021).

To further improve selectivity, light-inducible approaches seem versatile, responsive, precise and reversible (Wu et al. 2021); however, short wave excitation limits its application at in vivo level. To get over this hurdle, near-infrared optical control has been proposed (Chen et al. 2020). On the other hand, concerns regarding off-target effects might turn out to be less significant for epigenetic editing versus genetic engineering: Cas9-mediated double-strand breaks can be induced by the (unspecific) binding of one Cas9 molecule, while various events are thought to be required in epigenetic editing to achieve gene expression modulation. Indeed, various combinations of effector domains are required for sustained expression modulation (Amabile et al. 2016; Josipovic et al. 2019; O'Geen et al. 2019; Halmai et al. 2020; Nuñez et al. 2021; Nakamura et al. 2021b), offering options to further reduce the off-target toxic effects. Importantly, the (off-target) stable reprogramming can be reversed by targeting counteracting enzymes (Amabile et al. 2016; Nuñez et al. 2021), allowing possibilities to reset the intervention. So, although the goal to reach to a system that allows straightforward and very efficient sustained gene expression modulation, with low off-target and immunological effects, seems far, companies are founded and developments are promising with exciting results obtained.

18.6 Conclusions

The study of DNA methylation in vivo is rapidly developing, and helps to understand epigenetic dysregulations at the single gene level and its association with disease. By direct interference at the level of DNA methylation, restoration of cellular function can be induced. As discussed in this review, some stumbling blocks have slowed the development of epigenetic editing, but ongoing technological improvements (especially sustained reprogramming) and the increasing list of preclinical therapeutic successes spark a wide interest to develop methylation-based epigenetic editing strategies for a wide variety of diseases. As any genomic locus can be targeted, epigenetic editing might open avenues for diseases without any current treatment options.

Acknowledgements The authors thank Sabine Stolzenburg for her assistance in writing the chapter on targeted methylation in the earlier book version (Stolzenburg et al. 2016). FCM acknowledges his Scholarship funding from the Colombian Ministry of Science, Technology and Innovation (Minciencias -COLCIENCIAS/COLFUTURO-Doctorados en el exterior 2017 N°783) and Instituto Tecnológico Metropolitano

(ITM). FS is supported by VALERE program, Vanvitelli per la Ricerca. H2020 European Cooperation in Science and Technology (COST) Training Actions (www.INC-COST.eu and www.EpiChemBio.eu) are acknowledged for facilitating network activities.

References

Amabile A, Migliara A, Capasso P, Biffi M, Cittaro D, Naldini L, Lombardo A (2016) Inheritable silencing of endogenous genes by hit-and-run targeted epigenetic editing. Cell 167:219–232 e14

Bailus BJ, Pyles B, Mcalister MM, O'Geen H, Lockwood SH, Adams AN, Nguyen JT, Yu A, Berman RF, Segal DJ (2016) Protein delivery of an artificial transcription factor restores widespread Ube3a expression in an Angelman syndrome mouse brain. Mol Ther 24:548–555

Becker S, Boch J (2021) TALE and TALEN genome editing technologies. Gene Genome Editing 2:100007

Berdasco M, Esteller M (2019) Clinical epigenetics: seizing opportunities for translation. Nat Rev Genet 20:109–127

Bernstein DL, Le Lay JE, Ruano EG, Kaestner KH (2015) TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. J Clin Invest 125:1998–2006

Bintu L, Yong J, Antebi YE, Mccue K, Kazuki Y, Uno N, Oshimura M, Elowitz MB (2016) Dynamics of epigenetic regulation at the single-cell level. Science 351: 720–724

Broche J, Kungulovski G, Bashtrykov P, Rathert P, Jeltsch A (2021) Genome-wide investigation of the dynamic changes of epigenome modifications after global DNA methylation editing. Nucleic Acids Res 49:158–176

Bustos FJ, Ampuero E, Jury N, Aguilar R, Falahi F, Toledo J, Ahumada J, Lata J, Cubillos P, Henríquez B, Guerra MV, Stehberg J, Neve RL, Inestrosa NC, Wyneken U, Fuenzalida M, Härtel S, Sena-Esteves M, Varela-Nallar L, Rots MG, Montecino M, van Zundert B (2017) Epigenetic editing of the Dlg4/PSD95 gene improves cognition in aged and Alzheimer's disease mice. Brain 140:3252–3268

Cano-Rodriguez D, Gjaltema R, Jilderda R, Jellema P, Dokter-Fokkens J, Ruiters M, Rots MG (2016) Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. Nat Commun 7:1–11

Chen H, Kazemier HG, De Groote ML, Ruiters MH, Xu GL, Rots MG (2014) Induced DNA demethylation by targeting ten-eleven translocation 2 to the human ICAM-1 promoter. Nucleic Acids Res 42:1563–1574

Chen Y, Yan X, Ping Y (2020) Optical manipulation of CRISPR/Cas9 functions: from ultraviolet to near-infrared light. ACS Mater Lett 2(6):644–653

Chen H, Wang L, Zeng X, Schwarz H, Nanda HS, Peng X, Zhou Y (2021) Exosomes, a new star for targeted delivery. Front Cell Dev Biol 9:751079

Choudhury SR, Cui Y, Lubecka K, Stefanska B, Irudayaraj J (2016) CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. Oncotarget 7:46545–46556

Colella P, Ronzitti G, Mingozzi F (2018) Emerging issues in AAV-mediated in vivo gene therapy. Mol Ther Methods Clin Dev 8:87–104

Cui C, Gan Y, Gu L, Wilson J, Liu Z, Zhang B, Deng D (2015) P16-specific DNA methylation by engineered zinc finger methyltransferase inactivates gene transcription and promotes cancer metastasis. Genome Biol 16:252

de Groote ML, Verschure PJ, Rots MG (2012) Epigenetic Editing: targeted rewriting of epigenetic marks to modulate expression of selected target genes. Nucleic Acids Res 40:10596–10613

Devesa-Guerra I, Morales-Ruiz T, Perez-Roldan J, Parrilla-Doblas JT, Dorado-Leon M, Garcia-Ortiz MV, Ariza RR, Roldan-Arjona T (2020) DNA methylation editing by CRISPR-guided excision of 5-methylcytosine. J Mol Biol 432:2204–2216

Ehrlich M (2019) DNA hypermethylation in disease: mechanisms and clinical relevance. Epigenetics 14: 1141–1163

Felsenfeld G (2014) A brief history of epigenetics. Cold Spring Harb Perspect Biol 6

Galonska C, Charlton J, Mattei AL, Donaghey J, Clement K, Gu H, Mohammad AW, Stamenova EK, Cacchiarelli D, Klages S, Timmermann B, Cantz T, Scholer HR, Gnirke A, Ziller MJ, Meissner A (2018) Genome-wide tracking of dCas9-methyltransferase footprints. Nat Commun 9:597

Geel TM, Ruiters MHJ, Cool RH, Halby L, Voshart DC, Andrade Ruiz L, Niezen-Koning KE, Arimondo PB, Rots MG (2018) The past and presence of gene targeting: from chemicals and DNA via proteins to RNA. Philos Trans R Soc Lond B Biol Sci 373

Gillmore JD, Gane E, Taubel J, Kao J, Fontana M, Maitland ML, Seitzer J, O'Connell D, Walsh KR, Wood K, Phillips J, Xu Y, Amaral A, Boyd AP, Cehelsky JE, Mckee MD, Schiermeier A, Harari O, Murphy A, Kyratsous CA, Zambrowicz B, Soltys R, Gutstein DE, Leonard J, Sepp-Lorenzino L, Lebwohl D (2021) CRISPR-Cas9 in vivo gene editing for transthyretin amyloidosis. N Engl J Med 385:493–502

Gomez JA, Beitnere U, Segal DJ (2019) Live-animal epigenome editing: convergence of novel techniques. Trends Genet 35:527–541

Greenberg MVC, Bourc'his D (2019) The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol 20:590–607

Gregory DJ, Zhang Y, Kobzik L, Fedulov AV (2013) Specific transcriptional enhancement of inducible nitric oxide synthase by targeted promoter demethylation. Epigenetics 8:1205–1212

Halmai J, Deng P, Gonzalez CE, Coggins NB, Cameron D, Carter JL, Buchanan FKB, Waldo JJ, Lock SR, Anderson JD, O'Geen H, Segal DJ, Nolta J, Fink KD (2020) Artificial escape from XCI by DNA methylation editing of the CDKL5 gene. Nucleic Acids Res 48:2372–2387

Hanzawa N, Hashimoto K, Yuan X, Kawahori K, Tsujimoto K, Hamaguchi M, Tanaka T, Nagaoka Y, Nishina H, Morita S, Hatada I, Yamada T, Ogawa Y (2020) Targeted DNA demethylation of the Fgf21 promoter by CRISPR/dCas9-mediated epigenome editing. Sci Rep 10:5181

Hofacker D, Broche J, Laistner L, Adam S, Bashtrykov P, Jeltsch A (2020) Engineering of effector domains for targeted DNA methylation with reduced off-target effects. Int J Mol Sci 21

Horii T, Morita S, Hino S, Kimura M, Hino Y, Kogo H, Nakao M, Hatada I (2020) Successful generation of epigenetic disease model mice by targeted demethylation of the epigenome. Genome Biol 21:77

Huang YH, Su J, Lei Y, Brunetti L, Gundry MC, Zhang X, Jeong M, Li W, Goodell MA (2017) DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A. Genome Biol 18:176

Huisman C, Van Der Wijst MG, Schokker M, Blancafort P, Terpstra MM, Kok K, Van Der Zee AG, Schuuring E, Wisman GB, Rots MG (2016) Re-expression of selected epigenetically silenced candidate tumor suppressor genes in cervical cancer by TET2-directed demethylation. Mol Ther 24:536–547

Huo M, Zhang J, Huang W, Wang Y (2021) Interplay among metabolism, epigenetic modifications, and gene expression in cancer. Front Cell Dev Biol 9: 793428

Jain S, Shukla S, Yang C, Zhang M, Fatma Z, Lingamaneni M, Abesteh S, Lane ST, Xiong X, Wang Y, Schroeder CM, Selvin PR, Zhao H (2021) TALEN outperforms Cas9 in editing heterochromatin target sites. Nat Commun 12:606

Jeziorska DM, Murray RJS, De Gobbi M, Gaentzsch R, Garrick D, Ayyub H, Chen T, Li E, Telenius J, Lynch M, Graham B, Smith AJH, Lund JN, Hughes JR, Higgs DR, Tufarelli C (2017) DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease. Proc Natl Acad Sci U S A 114:E7526–E7535

Josipovic G, Tadic V, Klasic M, Zanki V, Beceheli I, Chung F, Ghantous A, Keser T, Madunic J, Boskovic M, Lauc G, Herceg Z, Vojta A, Zoldos V (2019) Antagonistic and synergistic epigenetic modulation using orthologous CRISPR/dCas9-based modular system. Nucleic Acids Res 47:9637–9657

Jurkowski TP, Ravichandran M, Stepper P (2015) Synthetic epigenetics-towards intelligent control of epigenetic states and cell identity. Clin Epigenetics 7:18

Kemaladewi DU, Bassi PS, Erwood S, Al-Basha D, Gawlik KI, Lindsay K, Hyatt E, Kember R, Place KM, Marks RM, Durbeej M, Prescott SA, Ivakine EA, Cohn RD (2019) A mutation-independent approach for muscular dystrophy via upregulation of a modifier gene. Nature 572:125–130

Kressler C, Gasparoni G, Nordstrom K, Hamo D, Salhab A, Dimitropoulos C, Tierling S, Reinke P, Volk HD, Walter J, Hamann A, Polansky JK (2020) Targeted de-methylation of the FOXP3-TSDR is sufficient to induce physiological FOXP3 expression but not a functional treg phenotype. Front Immunol 11: 609891

Kretzmann JA, Evans CW, Moses C, Sorolla A, Kretzmann AL, Wang E, Ho D, Hackett MJ, Dessauvagie BF, Smith NM, Redfern AD, Waryah C, Norret M, Iyer KS, Blancafort P (2019) Tumour suppression by targeted intravenous non-viral CRISPRa using dendritic polymers. Chem Sci 10:7718–7727

Kungulovski G, Nunna S, Thomas M, Zanger UM, Reinhardt R, Jeltsch A (2015) Targeted epigenome editing of an endogenous locus with chromatin modifiers is not stably maintained. Epigenetics Chromatin 8:12

Lainscek D, Kadunc L, Keber MM, Bratkovic IH, Romih R, Jerala R (2018) Delivery of an artificial transcription regulator dCas9-VPR by extracellular vesicles for therapeutic gene activation. ACS Synth Biol 7:2715–2725

Ledhord H (2018) First test of in-body gene editing shows promise. Nature. https://doi.org/10.1038/d41586-018-06195-6

Lei Y, Zhang X, Su J, Jeong M, Gundry MC, Huang YH, Zhou Y, Li W, Goodell MA (2017) Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. Nat Commun 8:16026

Levy JM, Yeh WH, Pendse N, Davis JR, Hennessey E, Butcher R, Koblan LW, Comander J, Liu Q, Liu DR (2020) Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. Nat Biomed Eng 4:97–110

Li K, Pang J, Cheng H, Liu WP, Di JM, Xiao HJ, Luo Y, Zhang H, Huang WT, Chen MK, Li LY, Shao CK, Feng YH, Gao X (2015) Manipulation of prostate cancer metastasis by locus-specific modification of the CRMP4 promoter region using chimeric TALE DNA methyltransferase and demethylase. Oncotarget 6:10030–10044

Liao HK, Hatanaka F, Araoka T, Reddy P, Wu MZ, Sui Y, Yamauchi T, Sakurai M, O'Keefe DD, Nunez-Delicado E, Guillen P, Campistol JM, Wu CJ, Lu LF, Esteban CR, Izpisua Belmonte JC (2017) In vivo target gene activation via CRISPR/Cas9-mediated trans-epigenetic modulation. Cell 171:1495–1507 e15

Lin L, Liu Y, Xu F, Huang J, Daugaard TF, Petersen TS, Hansen B, Ye L, Zhou Q, Fang F, Yang L, Li S, Floe L, Jensen KT, Shrock E, Chen F, Yang H, Wang J, Liu X, Xu X, Bolund L, Nielsen AL, Luo Y (2018) Genome-wide determination of on-target and off-target characteristics for RNA-guided DNA methylation by dCas9 methyltransferases. Gigascience 7:1–19

Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, Shu J, Dadon D, Young RA, Jaenisch R (2016) Editing DNA

methylation in the mammalian genome. Cell 167:233–247 e17

Liu XS, Wu H, Krzisch M, Wu X, Graef J, Muffat J, Hnisz D, Li CH, Yuan B, Xu C, Li Y, Vershkov D, Cacace A, Young RA, Jaenisch R (2018) Rescue of fragile X syndrome neurons by DNA methylation editing of the FMR1 gene. Cell 172:979–992 e6

Lo CL, Choudhury SR, Irudayaraj J, Zhou FC (2017) Epigenetic editing of Ascl1 gene in neural stem cells by optogenetics. Sci Rep 7:42047

Lu Z, Liu Z, Mao W, Wang X, Zheng X, Chen S, Cao B, Huang S, Zhang X, Zhou T, Zhang Y, Huang X, Sun Q, Li JD (2020) Locus-specific DNA methylation of Mecp2 promoter leads to autism-like phenotypes in mice. Cell Death Dis 11:85

Luo N, Li J, Chen Y, Xu Y, Wei Y, Lu J, Dong R (2021) Hepatic stellate cell reprogramming via exosome-mediated CRISPR/dCas9-VP64 delivery. Drug Deliv 28:10–18

Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, Tsai SQ, Ho QH, Sander JD, Reyon D, Bernstein BE, Costello JF, Wilkinson MF, Joung JK (2013) Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat Biotechnol 31:1137–1142

Marx N, Grunwald-Gruber C, Bydlinski N, Dhiman H, Ngoc Nguyen L, Klanert G, Borth N (2018) CRISPR-based targeted epigenetic editing enables gene expression modulation of the silenced beta-galactoside alpha-2,6-sialyltransferase 1 in CHO cells. Biotechnol J 13:e1700217

Matharu N, Rattanasopha S, Tamura S, Maliskova L, Wang Y, Bernard A, Hardin A, Eckalbar WL, Vaisse C, Ahituv N (2019) CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. Science 363

McDonald JI, Celik H, Rois LE, Fishberger G, Fowler T, Rees R, Kramer A, Martens A, Edwards JR, Challen GA (2016) Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. Biol Open 5:866–874

Mlambo T, Nitsch S, Hildenbeutel M, Romito M, Muller M, Bossen C, Diederichs S, Cornu TI, Cathomen T, Mussolino C (2018) Designer epigenome modifiers enable robust and sustained gene silencing in clinically relevant human cells. Nucleic Acids Res 46:4456–4468

Morita S, Noguchi H, Horii T, Nakabayashi K, Kimura M, Okamura K, Sakai A, Nakashima H, Hata K, Nakashima K, Hatada I (2016) Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. Nat Biotechnol 34:1060–1065

Nakamura M, Ivec AE, Gao Y, Qi LS (2021a) Durable CRISPR-based epigenetic silencing. BioDesign Res 8

Nakamura M, Gao Y, Dominguez AA, Qi LS (2021b) CRISPR technologies for precise epigenome editing. Nat Cell Biol 23:11–22

Nicoglou A, Merlin F (2017) Epigenetics: A way to bridge the gap between biological fields. Stud Hist Philos Biol Biomed Sci 66:73–82

Nuñez JK, Chen J, Pommier GC, Cogan JZ, Replogle JM, Adriaens C, Ramadoss GN, Shi Q, Hung KL, Samelson AJ, Pogson AN, Kim JYS, Chung A, Leonetti MD, Chang HY, Kampmann M, Bernstein BE, Hovestadt V, Gilbert LA, Weissman JS (2021) Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. Cell 184:2503–2519 e17

O'Geen H, Bates SL, Carter SS, Nisson KA, Halmai J, Fink KD, Rhie SK, Farnham PJ, Segal DJ (2019) Ezh2-dCas9 and KRAB-dCas9 enable engineering of epigenetic memory in a context-dependent manner. Epigenetics Chromatin 12:26

Okada M, Kanamori M, Someya K, Nakatsukasa H, Yoshimura A (2017) Stabilization of Foxp3 expression by CRISPR-dCas9-based epigenome editing in mouse primary T cells. Epigenetics Chromatin 10:24

Onodera A, Gonzalez-Avalos E, Lio CJ, Georges RO, Bellacosa A, Nakayama T, Rao A (2021) Roles of TET and TDG in DNA demethylation in proliferating and non-proliferating immune cells. Genome Biol 22:186

Ou K, Yu M, Moss NG, Wang YJ, Wang AW, Nguyen SC, Jiang C, Feleke E, Kameswaran V, Joyce EF, Naji A, Glaser B, Avrahami D, Kaestner KH (2019) Targeted demethylation at the CDKN1C/p57 locus induces human beta cell replication. J Clin Invest 129:209–214

Petryk N, Bultmann S, Bartke T, Defossez PA (2021) Staying true to yourself: mechanisms of DNA methylation maintenance in mammals. Nucleic Acids Res 49:3020–3032

Pflueger C, Tan D, Swain T, Nguyen T, Pflueger J, Nefzger C, Polo JM, Ford E, Lister R (2018) A modular dCas9-SunTag DNMT3A epigenome editing system overcomes pervasive off-target activity of direct fusion dCas9-DNMT3A constructs. Genome Res 28:1193–1206

Policarpi C, Dabin J, Hackett JA (2021) Epigenetic editing: Dissecting chromatin function in context. Bioessays 43:e2000316

Qiu M, Glass Z, Chen J, Haas M, Jin X, Zhao X, Rui X, Ye Z, Li Y, Zhang F, Xu Q (2021) Lipid nanoparticle-mediated codelivery of Cas9 mRNA and single-guide RNA achieves liver-specific in vivo genome editing of Angptl3. Proc Natl Acad Sci U S A 118

Qu J, Zhu L, Zhou Z, Chen P, Liu S, Locy ML, Thannickal VJ, Zhou Y (2018) Reversing mechanoinductive DSP expression by CRISPR/dCas9-mediated epigenome editing. Am J Respir Crit Care Med 198:599–609

Rivenbark AG, Stolzenburg S, Beltran AS, Yuan X, Rots MG, Strahl BD, Blancafort P (2012) Epigenetic reprogramming of cancer cells via targeted DNA methylation. Epigenetics 7:350–360

Sapozhnikov DM, Szyf M (2021) Unraveling the functional role of DNA demethylation at specific promoters by targeted steric blockage of DNA methyltransferase with CRISPR/dCas9. Nat Commun 12:5711

Saunderson EA, Stepper P, Gomm JJ, Hoa L, Morgan A, Allen MD, Jones JL, Gribben JG, Jurkowski TP, Ficz G (2017) Hit-and-run epigenetic editing prevents

senescence entry in primary breast cells from healthy donors. Nat Commun 8:1450

Sgro A, Blancafort P (2020) Epigenome engineering: new technologies for precision medicine. Nucleic Acids Res 48:12453–12482

Shayevitch R, Askayo D, Keydar I, Ast G (2018) The importance of DNA methylation of exons on alternative splicing. RNA 24:1351–1362

Siddique AN, Nunna S, Rajavelu A, Zhang Y, Jurkowska RZ, Reinhardt R, Rots MG, Ragozin S, Jurkowski TP, Jeltsch A (2013) Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity. J Mol Biol 425: 479–491

Slaska-Kiss K, Zsibrita N, Koncz M, Albert P, Csabradi A, Szentes S, Kiss A (2021) Lowering DNA binding affinity of SssI DNA methyltransferase does not enhance the specificity of targeted DNA methylation in E. coli. Sci Rep 11:15226

Stepper P, Kungulovski G, Jurkowska RZ, Chandra T, Krueger F, Reinhardt R, Reik W, Jeltsch A, Jurkowski TP (2017) Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. Nucleic Acids Res 45:1703–1713

Stolzenburg S, Beltran AS, Swift-Scanlan T, Rivenbark AG, Rashwan R, Blancafort P (2015) Stable oncogenic silencing in vivo by programmable and targeted de novo DNA methylation in breast cancer. Oncogene 34:5427–5435

Stolzenburg S, Goubert D, Rots MG (2016) Rewriting DNA methylation signatures at will: the curable genome within reach? Adv Exp Med Biol 945:475–490

Suzuki T, Maeda S, Furuhata E, Shimizu Y, Nishimura H, Kishima M, Suzuki H (2017) A screening system to identify transcription factors that induce binding site-directed DNA demethylation. Epigenetics Chromatin 10:60

Tarjan DR, Flavahan WA, Bernstein BE (2019) Epigenome editing strategies for the functional annotation of CTCF insulators. Nat Commun 10:4258

Thakore PI, Kwon JB, Nelson CE, Rouse DC, Gemberling MP, Oliver ML, Gersbach CA (2018) RNA-guided transcriptional silencing in vivo with S. aureus CRISPR-Cas9 repressors. Nat Commun 9:1674

Tiane A, Schepers M, Riemens R, Rombaut B, Vandormael P, Somers V, Prickaerts J, Hellings N, Van Den Hove D, Vanmierlo T (2021) DNA methylation regulates the expression of the negative transcriptional regulators ID2 and ID4 during OPC differentiation. Cell Mol Life Sci 78:6631–6644

Van MV, Fujimori T, Bintu L (2021) Nanobody-mediated control of gene expression and epigenetic memory. Nat Commun 12:537

Vizoso M, Van Rheenen J (2021) Diverse transcriptional regulation and functional effects revealed by CRISPR/Cas9-directed epigenetic editing. Oncotarget 12:1651–1662

Vojta A, Dobrinic P, Tadic V, Bockor L, Korac P, Julg B, Klasic M, Zoldos V (2016) Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res 44:5615–5628

Wang Q, Dai L, Wang Y, Deng J, Lin Y, Wang Q, Fang C, Ma Z, Wang H, Shi G, Cheng L, Liu Y, Chen S, Li J, Dong Z, Su X, Yang L, Zhang S, Jiang M, Huang M, Yang Y, Yu D, Zhou Z, Wei Y, Deng H (2019) Targeted demethylation of the SARI promotor impairs colon tumour growth. Cancer Lett 448:132–143

Wang H, Han M, Qi LS (2021) Engineering 3D genome organization. Nat Rev Genet 22:343–360

Wei T, Cheng Q, Min YL, Olson EN, Siegwart DJ (2020) Systemic nanoparticle delivery of CRISPR-Cas9 ribonucleoproteins for effective tissue specific genome editing. Nat Commun 11:3232

Weichenhan D, Lipka DB, Lutsik P, Goyal A, Plass C (2020) Epigenomic technologies for precision oncology. Semin Cancer Biol 84:60–68

Wu X, Li G, Xie R (2018) Decoding the role of TET family dioxygenases in lineage specification. Epigenetics Chromatin 11:58

Wu H, Wang F, Jiang JH (2021) Inducible CRISPR-dCas9 transcriptional systems for sensing and genome regulation. Chembiochem 22:1894–1900

Xiong T, Meister GE, Workman RE, Kato NC, Spellberg MJ, Turker F, Timp W, Ostermeier M, Novina CD (2017) Targeted DNA methylation in human cells using engineered dCas9-methyltransferases. Sci Rep 7:6732

Xu GL, Bestor TH (1997) Cytosine methylation targetted to pre-determined sequences. Nat Genet 17:376–378

Xu SJ, Heller EA (2019) Recent advances in neuroepigenetic editing. Curr Opin Neurobiol 59:26–33

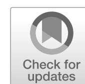
Xu X, Tao Y, Gao X, Zhang L, Li X, Zou W, Ruan K, Wang F, Xu GL, Hu R (2016) A CRISPR-based approach for targeted DNA demethylation. Cell Discov 2:16009

Xu X, Tan X, Tampe B, Wilhelmi T, Hulshoff MS, Saito S, Moser T, Kalluri R, Hasenfuss G, Zeisberg EM, Zeisberg M (2018) High-fidelity CRISPR/Cas9-based gene-specific hydroxymethylation rescues gene expression and attenuates renal fibrosis. Nat Commun 9:3509

Xu SJ, Lombroso SI, Fischer DK, Carpenter MD, Marchione DM, Hamilton PJ, Lim CJ, Neve RL, Garcia BA, Wimmer ME, Pierce RC, Heller EA (2021) Chromatin-mediated alternative splicing regulates cocaine-reward behavior. Neuron 109:2943–2966 e8

Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH (2015) Off-target effects in CRISPR/Cas9-mediated genome engineering. Mol Ther Nucleic Acids 4:e264

Zheng Y, Shen W, Zhang J, Yang B, Liu YN, Qi H, Yu X, Lu SY, Chen Y, Xu YZ, Li Y, Gage FH, Mi S, Yao J (2018) Author Correction: CRISPR interference-based specific and efficient gene inactivation in the brain. Nat Neurosci 21:894



DNA Labeling Using DNA Methyltransferases

19

Miglė Tomkuvienė, Edita Kriukienė, and Saulius Klimašauskas

Abstract

DNA methyltransferases (MTases) uniquely combine the ability to recognize and covalently modify specific target sequences in DNA using the ubiquitous cofactor *S*-Adenosyl-L-methionine (AdoMet). Although DNA methylation plays important roles in biological signaling, the transferred methyl group is a poor reporter and is highly inert to further biocompatible derivatization. To unlock the biotechnological power of these enzymes, extended cofactor AdoMet analogs have been developed that enable targeted MTase-directed attachment of larger moieties containing functional or reporter groups onto DNA. As the enlarged cofactors are not always compatible with the active sites of native MTases, steric engineering of the active site has been employed to optimize their alkyltransferase activity. In addition to the described cofactor analogs, recently discovered atypical reactions of DNA cytosine-5 MTases involving non-cofactor-like compounds can also be exploited for targeted derivatization and labeling of DNA. Altogether, these approaches offer new powerful tools for sequence-specific covalent DNA labeling, leading to a variety of useful techniques in DNA research, diagnostics and nanotechnologies, and have already proven practical utility for optical DNA mapping and high-throughput epigenome studies.

Keywords

DNA methyltransferase · *S*-Adenosyl-L-methionine · targeted DNA labeling · Synthetic cofactor AdoMet analog · Enzyme engineering · Optical mapping · Epigenomic mapping

List of Abbreviations

4mC	*N*4-methylcytosine
5mC	5-methylcytosine
6mA	*N*6-methyladenine
ACI	Activated cytosine intermediate
AdoHcy	*S*-Adenosyl-L-homocysteine
AdoMet	*S*-Adenosyl-L-methionine
caC	5-carboxycytosine
CuAAC	Cu(I)-catalyzed azide–alkyne cycloaddition
dSTORM	Direct stochastic optical reconstruction microscopy
ES	Embryonic stem (cells)
FRET	Förster resonance energy transfer
hmC	5-hydroxymethylcytosine

M. Tomkuvienė · E. Kriukienė · S. Klimašauskas (✉)
Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, Lithuania
e-mail: migle.tomkuviene@bti.vu.lt; edita.kriukiene@bti.vu.lt; saulius.klimasauskas@bti.vu.lt

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_19

IEDDA	Inverse electron demand Diels-Alder (reaction)
MAT	Methionine adenosyltransferase
mESC	Mouse embryonic stem cells
mTAG	Methyltransferase-directed transfer of activated groups
mTAG-chip	mTAG-enriched microarray analysis
mTAG-seq	mTAG-enriched sequencing
MTase	DNA methyltransferase
NHS	*N*-hydroxysuccinimide
ODN	Oligodeoxyribonucleotide
PDB	Protein Data Bank
SMILing	Sequence-specific methyltransferase-induced labeling
SPAAC	Strain-promoted azide–alkyne cycloaddition
TET	Ten-eleven translocation
TOP-seq	Tethered oligonucleotide-primed sequencing
uTOP-seq	TOP-seq analysis of unmethylated CG sites

19.1 Introduction

DNA is a large linear polymer comprised of aperiodic combinations of four major types of building blocks encoding the genetic blueprint of life. Since different loci of this largely uniform biomolecule rarely contain features distinct enough to permit their chemical or physical identification among other DNA loci or other biomolecules, a key task is to furnish them with suitable reporter tags for their selective visualization and isolation from biological samples. Among the variety of enzymes involved in DNA metabolism, DNA methyltransferases (MTases) uniquely combine two useful features required for targeted labeling: recognition of a vast repertoire of specific target sequences (2–8 nt long) and covalent modification of the target site. Although targeted DNA methylation can be "read" by specific cellular proteins and thus plays important roles in biological signaling, the naturally transferred methyl group is a poor reporter and is not readily amenable for further chemical derivatization. A general strategy to unlock the biotechnological potential of these highly specific enzymes is to make them transfer "pre-derivatized" (extended) versions of the methyl group. One such example, the carboxy-AdoMet cofactor, has been found in nature, however, the transferred functional group is not selectively reactive in the presence of other biomolecules (Wang and Kohli 2021). Therefore, a series of synthetic analogs of the AdoMet cofactor were developed that allowed MTases to tag DNA with a range of extended moieties, making sequence-specific MTase-directed labeling an attractive opportunity in various biotechnological applications (reviewed in Tomkuvienė et al. 2019). Two major types of cofactor analogs have been developed for MTase-catalyzed DNA labeling, which permit covalent deposition of either a whole cofactor molecule or its sulfonium-bound side chain (Klimašauskas and Weinhold 2007). Among the three known classes of DNA methyltransferases (cytosine-5, adenine-*N*6, and cytosine-*N*4 MTases), the first two have been largely utilized for the attachment of various reactive groups, biotin or fluorophores to DNA. Due to the universal nature of the AdoMet cofactor for biological methylations, the approach also proved applicable for labeling other biomolecules, such as RNA (Motorin et al. 2011; Tomkuvienė et al. 2012; Plotnikova et al. 2014; Schulz et al. 2013; Holstein et al. 2014), proteins (Peters et al. 2010; Islam et al. 2011; Willnow et al. 2012; Wang et al. 2013; Hymbaugh Bergman and Comstock 2015) and small molecules (Zhang et al. 2006; Stecher et al. 2009; Lee et al. 2010; Winter et al. 2013) using appropriate MTases.

Another recently developed, cofactor-independent DNA modification strategy is based on atypical reactions of DNA cytosine-5 MTases. Upon interaction with the target cytosine, these MTases use a covalent attack to transiently generate an activated cytosine intermediate (ACI) (see Chap. 1). In the absence of AdoMet or synthetic AdoMet analogs, the ACI can undergo a covalent addition of exogenous formaldehyde yielding 5-hydroxymethylcytosine (hmC)

(Liutkevičiūtė et al. 2009). Moreover, hmC residues at the target site can be dehydroxymethylated to yield cytosine or can undergo further addition of thiols or selenols to yield the corresponding 5-chalcogenomethyl derivatives in DNA in a 5mC-MTase-dependent manner (Liutkevičiūtė et al. 2011). These transformations open new possibilities for sequence-specific derivatization and analysis of epigenetic marks in mammalian DNA.

In the following sections, the DNA labeling approaches based on the two types of extended cofactor analogs and the reactions involving non-cofactor-like compounds are described and discussed in detail.

19.2 Synthetic Cofactor Analogs for MTase-Directed Modification of DNA

The first labeling strategy (named Sequence-specific Methyltransferase-Induced Labeling or SMILing) developed by the Weinhold group employed cofactor analogs in which the methionine moiety of AdoMet was synthetically replaced by an aziridine ring (*N*-adenosylaziridine cofactors) (Pignot et al. 1998). Upon reaction of *N*-adenosylaziridine with DNA in the presence of a DNA MTase, the “transfer” of an electrophilic carbon atom of the protonated aziridine ring to a nucleophilic target atom in DNA leads to ring opening, thereby turning the ring into an ethylamino linker that connects the cofactor molecule with the target nucleobase (Pignot et al. 1998) (Fig. 19.1). Although the attached ethylaminoadenosine moiety by itself is not a good reporter group, it can serve as a carrier to which desired chemical and reporter groups are attached (Pljevaljčić et al. 2003, 2004; Kunkel et al. 2015). Subsequently, the groups of Rajsky and Comstock expanded the chemical scope of this approach by introducing 2-haloethyl *N*-mustard analogs, which are converted into aziridines in situ and thus are presumed to work by a similar mechanism (Weller and Rajski 2005, 2006; Townsend et al. 2009; Mai and Comstock 2011; Du et al. 2012; Ramadan et al. 2014). In the *N*-mustard cofactors, the nitrogen atom, which is equivalent to the sulfur atom of the sulfonium group in AdoMet, can in addition be used to attach a reactive chemical group (alkyne) (Weller and Rajski 2005), a photocaging group (Townsend et al. 2009), or the original amino acid moiety (present in AdoMet but absent in the *N*-adenosylaziridine analogs) (Weller and Rajski 2006; Du et al. 2012; Ramadan et al. 2014). Both the aziridine and *N*-mustard cofactors (see Table 19.1) are obtained via multistep synthetic routes and can thus only be produced in specialized chemistry laboratories.

An inherent feature of the SMILing reaction is that the directing MTase remains tightly (although non-covalently) bound to the coupling product, which represents a chemically linked bisubstrate derivative entangling the enzyme. Therefore, stoichiometric amounts of an enzyme with respect to its target sites on target DNA are required for quantitative conversion, and additional steps may be necessary if the bound enzyme is to be removed from the DNA.

The second DNA labeling approach is based on AdoMet analogs in which the sulfonium-bound methyl group of AdoMet is replaced with an extended side chain, and only this part of the cofactor is transferred to the target nucleotide (Fig. 19.1). Replacement of the methyl group in AdoMet with larger aliphatic carbon chains leads to a drastic decline of MTase-catalyzed S_N2 reaction rates (Schlenk and Dainko 1975). In a joint effort, the Klimašauskas and Weinhold groups found that the deficiency of the reaction can be remedied by placing π-orbitals near the reaction center (Dalhoff et al. 2006a). This activation was observed with synthetic AdoMet analogs carrying a double bond (allylic system), a triple bond (propargylic system) or an aromatic ring next to the reactive carbon in the extended side chain (Fig. 19.1). Mechanistic considerations suggest that the π-orbitals in the unsaturated bond lower the energy barrier of the reaction via conjugative stabilization of a pentacoordinated S_N2 transition state. The discovery of the sidechain-activated AdoMet analogs paved the way to a rapid

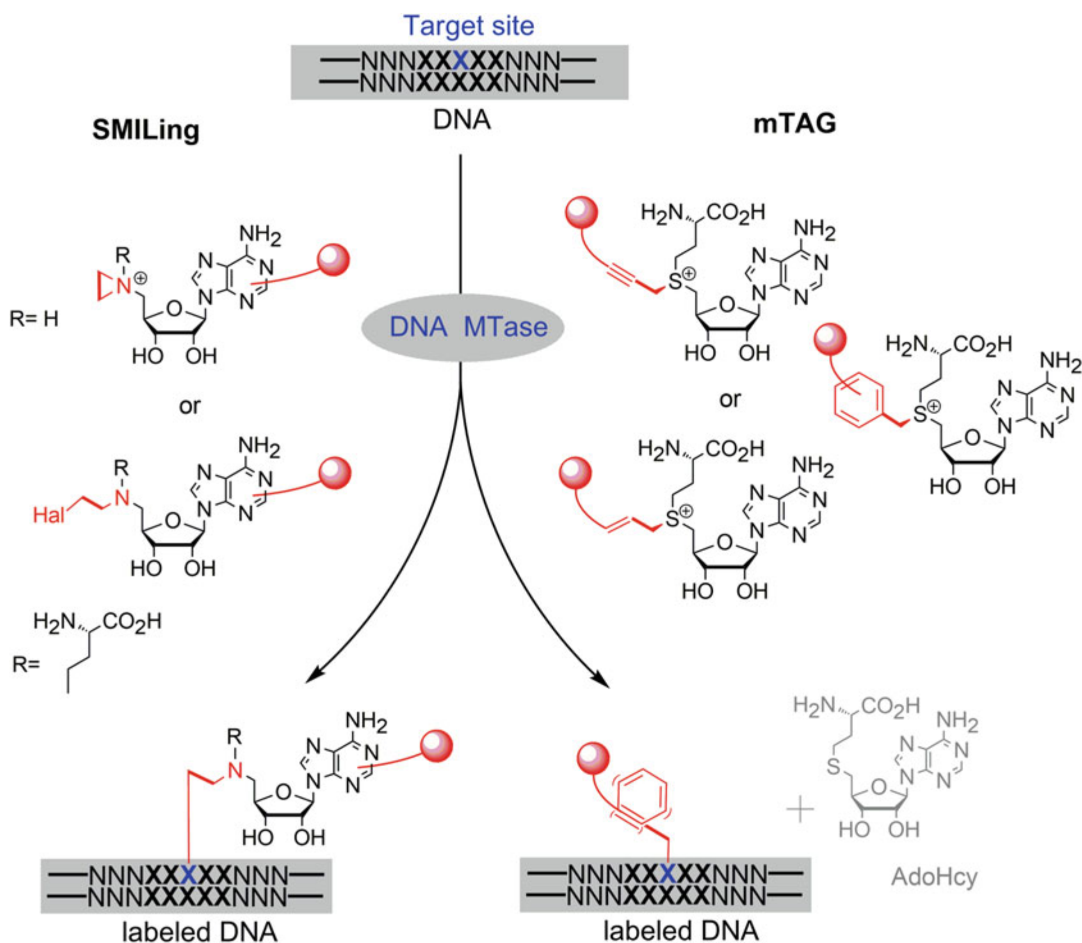


Fig. 19.1 Methyltransferase-directed sequence-specific labeling of DNA using synthetic analogs of the cofactor AdoMet. (Left) SMILing approach: covalent coupling of an aziridine (upper) or *N*-mustard (lower) cofactor carrying a functional or reporter group (red sphere) attached via a linker (red line) onto a target nucleobase (blue) in DNA. (Right) mTAG approach: transfer of a sulfonium-bound extended linear chain carrying an activating moiety (a triple or double bond or an aromatic ring), a linker and a functional or reporter group (red sphere) from a double-activated AdoMet analog onto a target nucleobase in DNA. N, random nucleotide; XXXXX, recognition sequence of the directing MTase

development of a new approach termed methyltransferase-directed Transfer of Activated Groups (mTAG).

Synthetic access to the mTAG cofactors appears somewhat easier as compared to the aziridine and N-mustard analogs, since they can be produced by chemical "recharging" of the cofactor product AdoHcy via regiospecific alkylation of its sulfur atom with a desired linear side chain (Fig. 19.2). Suitable electrophilic side chains can sometimes be obtained directly from commercial sources, but certain cases may require advanced synthetic skill (Lukinavičius et al. 2007, 2013; Dalhoff et al. 2006b). Chemical synthesis typically yields the cofactor analogs as diastereomeric mixtures of *R,S*- and *S,S*-isomers, which can be used directly for most purposes or can be chromatographically enriched in the enzymatically active *S,S*-isomer by reversed-phase chromatography (Lukinavičius et al. 2013). An important advantage of the mTAG approach is the possibility of chemo-enzymatic synthesis of enantiomerically pure mTAG cofactors from corresponding methionine analogs and ATP

Table 19.1 Cofactor analogs for MTase-directed derivatization and labeling of DNA

	Position of linker	Linker length	Functional or reporter group	Short name	MTases used	Applications	References
mTAG							
Propargyl analogs	S	1	Alkyne	2-Butynyl-SAM	M.TaqI	DNA labeling and extraction	Artyukhin and Woo (2012)
		1	Alkyne	AdoButyn	eM.HhaI	Studies of TET and glycosylase activity	Tomkuvienė et al. (2020)
		6	Alkyne	Ado-6-ethyne	eM.HhaI	Fluorescent labeling of plasmid DNA	Lukinavičius et al. (2013)
					M.EcoDam	Studies of TAL-effector's tolerance to large DNA modifications	Flade et al. (2017)
		6	Amine	Ado-6-amine	eM.HhaI	Fluorescent labeling of plasmid DNA	Lukinavičius et al. (2013)
					eM.SssI	Biotin labeling of DNA for epigenome profiling	Kriukienė et al. (2013)
					M.TaqI	Plasmid DNA labeling for a single-molecule fluorescence assay	Lauer et al. (2017)
		6+	Amine and Atto647N	Ado-6-amine–Atto647N	M.TaqI M.FokI eM.HhaI eM.MpeI	One-pot direct conjugation of functional compounds to DNA	Deen et al. (2019)
		6+	Amine and Atto565 TAMRA Atto520 Atto488 Alexa488 Biotin 5000 PEG	Ado-6-amine–Atto565 TAMRA Atto520 Atto488 Alexa488 Biotin 5000 PEG	M.TaqI	One-pot direct conjugation of functional compounds to DNA	Deen et al. (2019)
		9	Amine	Ado-9-amine	eM.HhaI M.TaqI	Fluorescent labeling of plasmid DNA	Lukinavičius et al. (2007)
		11	Amine	Ado-11-amine	eM2.Eco31I eM.HhaI eM.HpaII	Fluorescent labeling of plasmid and phage DNA Optical DNA mapping	Neely et al. (2010), Lukinavičius et al. (2012), Lukinavičius et al. (2013)
		6	Azide	Ado-6-azide	eM.HhaI	Fluorescent labeling of plasmid DNA ex vivo	Lukinavičius et al. (2013)
						Interrogation of TET and glycosylase activity on unnaturally modified cytosine	Tomkuvienė et al. (2020)

(continued)

Table 19.1 (continued)

	Position of linker	Linker length	Functional or reporter group	Short name	MTases used	Applications	References
					eM.SssI	Biotin labeling of DNA for epigenome profiling	Kriukienė et al. (2013)
						Oligonucleotide tethering to DNA for the high-resolution DNA epigenome profiling (TOP-seq)	Staševskij et al. (2017)
						Base resolution analysis of genomic 5-carboxylcytosine	Ličytė et al. (2020)
						Identifcation of fetal unmodified CG sites in maternal cell-free DNA for non-invasive prenatal testing	Gordevičius et al. (2020)
						Epigenetic analysis of neuroblastoma cells	Narmontė et al. (2021)
						Studies of nucleosome assembly	Tomkuvienė et al. (2022)
					M.TaqI	Labeling of DNA origami to create fluorescent beacons	Heck et al. (2020)
						Super-resolution optical genome mapping in nanochannels	Jeffet et al. (2016)
						Plasmid DNA labeling for a single-molecule fluorescence assay	Lauer et al. (2017)
						DNA optical mapping for whole genome, single-molecule analyses	Wand et al. (2019)
					eDnmt1	Dnmt1-specific tagging of genomic DNA in vivo	Stankevičius et al. (2022)
		8	Schiff-base and Azide		M.TaqI eM.MpeI	Reversible labeling of DNA	Wilkinson et al. (2020)
		8	Atto647N	AdoHcy-8-Atto647N	M.TaqI	FRET-based assay for methyltransferase activity	Long et al. (2021)
		18	Biotin	Ado-18-biotin	eM.HhaI	Labeling of plasmid DNA	Urbanavičiūtė et al. UO, Plotnikova et al. (2014)
		20	Biotin		M.TaqI	Labeling of short genome DNA for nanopore sensing	Chen et al. (2017)

		20	TAMRA	AdoYnTAMRA	M.TaqI	Fluorescent labeling of phage DNA for optical strain typing	Grunwald et al. (2015)
						Single-molecule sensing of unmethylated CpG sites with nanopores	Gilboa et al. (2016)
						Observation of epigenetic modulation during macrophage activation	Jain et al. (2019)
						Long-read single-molecule optical mapping of unmethylated CpG sites in disease samples	Sharim et al. (2019)
					M.EcoDam	Interrogation of TAL-effector's tolerance to large DNA modifications	Flade et al. (2017)
			CF640R	AdoYnCF640R	M.TaqI	Single-molecule sensing of unmethylated CpG sites in solid-state nanopores	Gilboa et al. (2016)
						Labeling of DNA origami to create fluorescent beacons	Heck et al. (2020)
		6, 14, 20	Cy2, Cy3 Cy5, sCy5 Cascade Blue Coumarin 343 Rhodamine B Rhodamine 6G Rhodamine 101 Pacific Blue	Cysteine-based AdoMet analogs	M.TaqI	Fluorescent labeling and optical mapping of DNA	Goyvaerts et al. (2020)
Allyl analogs	S	1	Alkene	AdoPropen	eM.HhaI	Studies of TET and glycosylase activity	Tomkuvienė et al. (2020)
		3	Alkyne	AdoEnYn	M.FokI M.TaqI M.XbaI	Fluorescent labeling of phage DNA for optical DNA mapping	Vranken et al. (2014)
Benzyl analogs	S	1	Substituted nitrobenzyl groups	AdoONB AdoPNB AdoANB AdoANT	M.TaqI	Reversible photo-cleavable modification of DNA	Anhäuser et al. (2018)
					eM.TaqI	Modification of plasmid DNA with a photocaging group in an AdoMet synthetase - MTase cascade reaction.	Michailidou et al. (2021)

(continued)

Table 19.1 (continued)

	Position of linker	Linker length	Functional or reporter group	Short name	MTases used	Applications	References
				Ado4MNB Ado4MNB AdoDMNB		Modification of DNA in the presence of AdoMet and photoregulation of gene expression in vitro	Heimes et al. (2018)
		Aminomethylbenzyl-	Norbornene	AdoNorb AdoNorc	M.TaqI	Two-step fluorescent labeling of plasmid DNA	Muttach et al. (2017)
SMILing							
N-adenosyl-aziridines	Ade-N6	5	Biotin	6BAz	M.BseCI, M.TaqI	Positioning of nanoparticles on bacteriophage DNA	Braun et al. (2008), Wilkinson et al. (2008)
						Engineering synaptic junctions in DNA Optical mapping of DNA-binding proteins	Kim et al. (2012)
						DNA modification on an origami-based enzyme nanofactory	Weinhold and Chakraborty (2021)
		11	Cy3	6Cy3Az	M.TaqI	Transfection with fluorescently labeled plasmid DNA	Schmidt et al. (2008)
	Ade-C7	4	Biotin		M.HhaI M.TaqI	Biotinylation and CpG methylation detection on plasmid DNA	Kunkel et al. (2015)
	Ade-C8	–	Azide		M.EcoRI M.HhaI M.SssI M.TaqI	Derivatization and biotinylation of ODNs MTase-directed DNA strand scission	Comstock and Rajski (2005a, b)
		5	Azide		M.EcoRI M.TaqI	ODN biotinylation	Comstock and Rajski (2005a)
		5	Biotin		M.TaqI	Biotinylation of plasmid DNA	Pljevaljčić et al. (2004, 2007)
		6	Dansyl		M.TaqI	Fluorescent labeling of plasmid DNA	Pljevaljčić et al. (2003, 2004)
N-adenosyl-*N*-mustards	Mustard-N	1	Alkyne		M.EcoRI M.TaqI	ODN derivatization	Weller and Rajski (2005)
	Ade-N6	1; 2; 4	Alkyne		M.HhaI M.TaqI	Derivatization of plasmid DNA	Ramadan et al. (2014)
		3; 4	Azide				
	Ade-C8	2	Alkyne		M.HhaI M.TaqI	Fluorescent labeling of plasmid DNA	Du et al. (2012)
		5	Azide				
		–	Azide		M.TaqI	Derivatization of plasmid DNA	Mai and Comstock (2011)

ODN, oligodeoxyribonucleotide; *UO*, unpublished observations

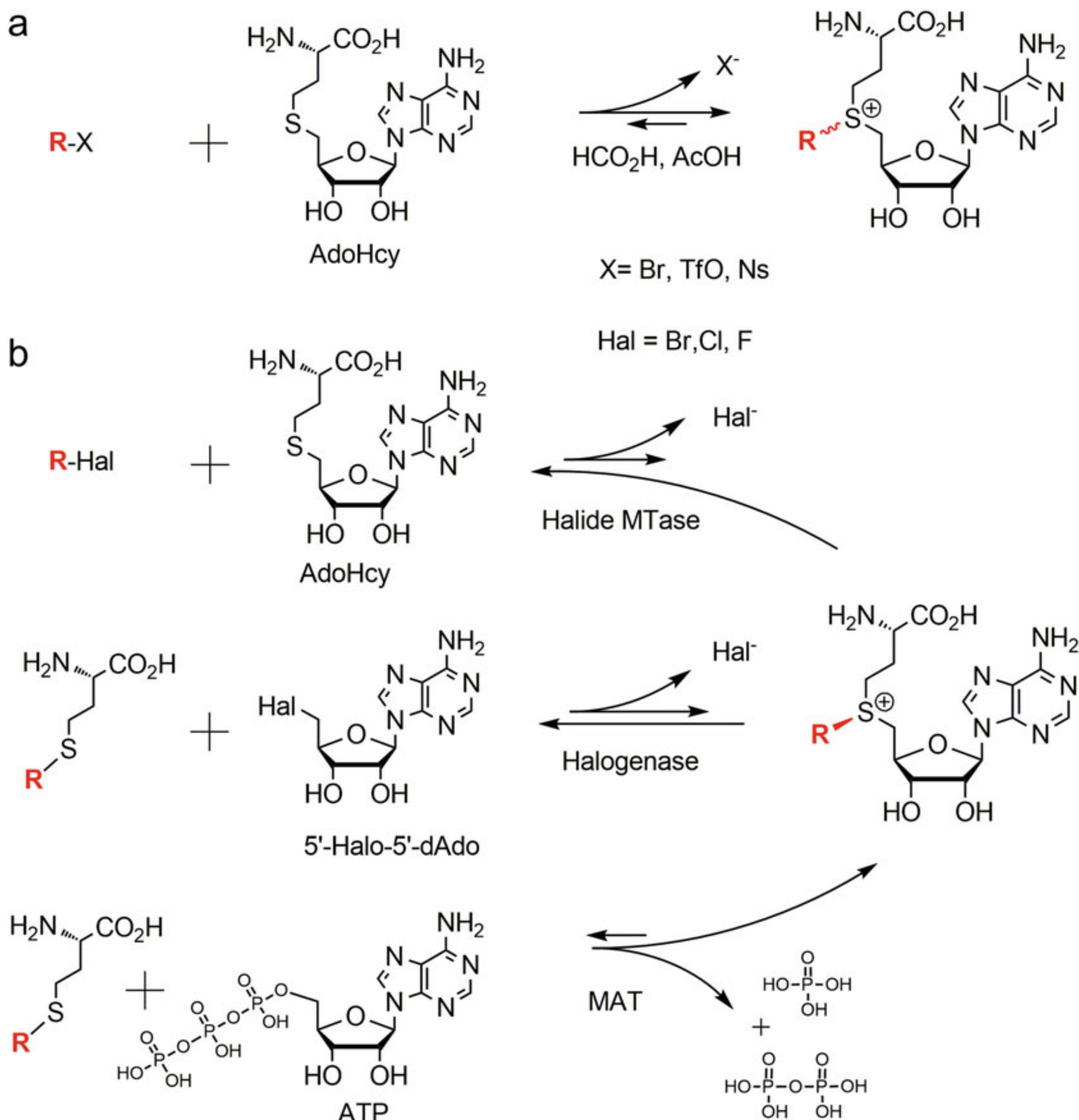


Fig. 19.2 Synthetic approaches to production of mTAG cofactor analogs. (**a**) Chemical synthesis of extended cofactors (as diastereomeric mixtures) by *S*-alkylation of AdoHcy under acidic conditions. (**b**) Enzymatic production of the biological enantiomer of extended cofactors by using Halide methyltransferases (reverse reaction), AdoMet-dependent Halogenases (reverse reaction), and methionine-adenosyl transferases (MAT). R, extended transferable group

using engineered methionine adenosyltransferases (Singh et al. 2014), which can in principle be performed in situ in cascade reactions (Michailidou et al. 2021), or even in living cells (Wang et al. 2013). Besides this natural reaction, two other "reverse" chemo-enzymatic routes exploited for the generation of enantiomerically pure cofactors include AdoMet-dependent Halogenases (from corresponding alkyl halides and AdoHcy) and Halide Methyltransferases (from extended methionine derivatives and 5′-halo-5′-deoxyadenosine) (Lipson et al. 2013, Tang et al. 2021a, b) (Fig. 19.2).

Since in mTAG reactions only the extended sulfonium-bound side chain is transferred from the cofactor analog to DNA, these AdoMet analogs circumvent the problem of catalytic product release. The produced AdoHcy, which may be inhibitory at higher concentrations, can be enzymatically removed by the addition of appropriate AdoHcy-degrading enzymes such as methylthioadenosine/ *S*-adenosyl homocysteine nucleosidase (Muttach and Rentmeister 2016; Gabrieli et al. 2021). A variety of allyl-based and propargyl-based analogs have been designed that carry unique chemical groups such as primary amine, alkyne and azide, or reporter groups (biotin, fluorophores) (see Table 19.1). Notably, although many MTases accept well both types of cofactors, some exhibit certain preferences with respect to the activating unsaturated bond (double or triple) or the side chain length. In particular, allylic cofactors have gained significant popularity with protein labeling (Peters et al. 2010; Islam et al. 2011, 2012; 2013; Wang et al. 2011, 2013; Blum et al. 2013; Bothwell and Luo 2014; Guo et al. 2014; Weiss et al. 2021), whereas propargylic side chains are more preferably transferred by the 5mC-DNA MTases (Table 19.2).

With some exceptions (Bothwell et al. 2012; Willnow et al. 2012), the life times of sulfonium-based propargylic cofactors in MTase buffers at 37 °C are in the order of 3 h (Lukinavičius et al. 2013). This affords reasonable reaction times for most practical in vitro and also in vivo applications taking into account that typical turnover rates of bacterial MTases are in the minutes range. Information on stability of other types of cofactors has not yet been reported.

Both the SMILing and mTAG cofactors can be used for two-step or one-step labeling. Single-step labeling by direct attachment of a desired reporter group may be beneficial in situations when minimal sample manipulations, simplicity, and speed are required. However, beside this potential advantage, the one-step approach entails an added synthetic complexity to the cofactor analog, as reporter groups are typically larger and more complex than functional groups. Moreover, an increased steric bulk of the transferable side chain may also lead to a partial or complete impairment of the directing MTase (Table 19.2). A key advantage of the two-step approach is the flexibility in manipulating the structure of the labeling product (linker length, conjugation chemistry, reporter group) for different downstream applications by combining different cofactors and chemoselective reporter compounds.

The spectrum of functional groups and applicable conjugation chemistries has been steadily increasing (Tomkuvienė et al. 2019, Goyvaerts et al. 2020, Wilkinson et al. 2020). Widely used conjugation chemistries involve *N*-hydroxysuccinimide ester - amino group conjugation, Cu(I)-catalyzed azide–alkyne cycloaddition (CuAAC) and Strain-promoted azide–alkyne cycloaddition (SPAAC). Among them, SPAAC has been found to be the most effective (Lauer et al. 2017), and the conjugation may proceed in one pot together with the enzymatic modification, reducing the number of handling steps needed for DNA labeling (Deen et al. 2019). To further broaden the spectrum of available modification schemes, a benzylic linker carrying a norbornene functionality has been demonstrated to permit a highly efficient inverse electron demand Diels-Alder (IEDDA) reaction with a tetrazine-bearing fluorophore (Muttach et al. 2017). Of special utility are linkers with built-in cleavable moieties whereby functional groups can be removed (disulfide or Schiff bases) or even exchanged (Schiff bases) with other functionalities on demand (Kriukienė et al. 2013, Wilkinson et al. 2020) permitting additional transformations of the labeled DNA. Some enzymatically deposited photo-sensitive aromatic groups can be completely removed by high wavelength UV light (360–400 nm) and have been shown to be applicable for manipulating transcription of model DNA fragments in vitro (Anhäuser et al. 2018; Heimes et al. 2018).

Table 19.2 Activity of DNA methyltransferases with AdoMet analogs

Enzyme[a]	Target sequence[b] 5′–3′	mTAG		SMILing	
		Reactions performed[c]	References	Reactions performed*	References
6mA-MTases					
M.TaqI wt	TCGA	M	Dalhoff et al. (2006a) Lukinavičius et al. (2007)	M	Pignot et al. (1998) Weller and Rajski (2006)
		F	Artyukhin and Woo (2012) Vranken et al. (2014) Jeffet et al. (2016) Lauer et al. (2017) Muttach et al. (2017) Anhäuser et al. (2018) Wand et al. (2019) Wilkinson et al. (2020) Heck et al. (2020) Michailidou et al. (2021)	F	Weller and Rajski (2005) Comstock and Rajski (2005a, b) Du et al. (2012) Mai and Comstock (2011) Ramadan et al. (2014)
		L	Grunwald et al. (2015) Deen et al. (2019) Gilboa et al. (2016) Goyvaerts et al. (2020) Heck et al. (2020) Jain et al. (2019) Long et al. (2021) Sharim et al. (2019) Chen et al. (2017)	L	Pljevaljčić et al. (2003, 2004, 2007) Braun et al. (2008) Schmidt et al. (2008) Wilkinson et al. (2008) Kunkel et al. (2015) Weinhold and Chakraborty (2021)
eM.TaqI		F	Heimes et al. (2018)		
M.BseCI	ATCGAT			L	Braun et al. (2008) Wilkinson et al. (2008) Kim et al. (2012)
M.EcoRI	GAATTC	N	Vranken et al. (2014)	M	Weller and Rajski (2006)
				F	Comstock and Rajski (2005a) Weller and Rajski (2005)
M.FokI	GGATG/	F	Vranken et al. (2014)		
	CATCC	L	Deen et al., (2019)		
M.XbaI	TCTAGA	F	Vranken et al. (2014)		
M.EcoDam	GATC	F, L	Flade et al. (2017)		
		N	Vranken et al. (2014)		
M.PstI	CTGCAG	N	Vranken et al. (2014)		
5mC-MTases					
M.HhaI wt	GCGC	M	Dalhoff et al. (2006a)	M F	Weller and Rajski (2006) Pljevaljčić et al. (2004) Comstock and Rajski (2005a, b) Ramadan et al. (2014) Du et al. (2012)
eM.HhaI		F	Lukinavičius et al. (2007, 2012, 2013) Neely et al. (2010) Tomkuvienė et al. (2020)	L	Kunkel et al. (2015)

(continued)

Table 19.2 (continued)

Enzyme[a]	Target sequence[b] 5′–3′	mTAG		SMILing	
		Reactions performed[c]	References	Reactions performed*	References
		L	Urbanavičiūtė et al. (unpublished) Deen et al. (2019)		
M.HhaI ΔL2–14 eM.HhaI ΔL2–14	GCG	M	Gerasimaitė et al. (2009)		
		F	Gerasimaitė et al. (unpublished)		
		L	Urbanavičiūtė et al. (unpublished)		
M.SssI wt	CG	N	Vranken et al. (2014)	M F	Weller and Rajski (2006) Comstock and Rajski (2005b)
eM.SssI		F	Kriukienė et al. (2013) Staševskij et al. (2017) Ličytė et al. (2020) Gordevičius et al. (2020) Narmontė et al. (2021)		
eM.MpeI	CG	F	Wilkinson et al. (2020)		
		L	Deen et al. (2019)		
M.HpaII wt eM.HpaII	CCGG	M F	Lukinavičius et al. (2012)	M	Comstock and Rajski (2005a)
M2.Eco31I wt eM2.Eco31I	GGTCTC	F	Lukinavičius et al. (2012)		
M.BsaHI wt eM.BsaHI	GRCGYC	N	Vranken et al. (2014) Deen et al. (2019)		
4mC-MTases					
M.BcnIB	CCSGG	M	Dalhoff et al. (2006a)	M	Pljevaljčić et al. (2004)
M.BamHI	GGATCC			M	Du et al. (2012)
M.PvuII	CAGCTG	N	Vranken et al. (2014) Deen et al. (2019)		

[a]eM, engineered MTase for improved mTAG reactions; [b]Target nucleotide is underlined; [c]N, none or low alkyltransferase activity; M, modification by transfer of a short non-functional moiety (mTAG) or a core unit (SMILing); F, derivatization with a functional group (2-step labeling possible); L, labeling with a reporter group in one step

19.3 MTase Activity with the Synthetic Cofactor Analogs

Bacterial and archaeal DNA MTases generally exhibit a clearly defined sequence and base specificity. Bacterial type-II DNA MTases (typically, single polypeptides of 250–400 residues) seem to be better suited for DNA labeling purposes as compared to the type I and III enzymes or mammalian DNA MTases, mostly due to their compact size and better enzymatic parameters (turnover rate, cofactor affinity, sequence fidelity, protein stability, etc.), although this general assumption does not preclude the existence of useful MTases derived from other than type-II cohorts. Current listings of type-II DNA MTases (REBASE, http://rebase.neb.com) count thousands of distinct recognition sequences ranging from two to eight base pairs in length. Therefore, a wide repertoire of DNA sequences can potentially be targeted.

Naturally, DNA MTases have evolved for optimal performance with the natural cofactor

AdoMet. The use of extended AdoMet analogs raises the question of steric limitations that may be imposed by the architectures of the active sites and cofactor binding pockets of MTases. The SMILing cofactors offer several potential anchoring points in the adenosine moiety (6, 7, and 8 positions of the adenine ring for the aziridines and additionally 5′-N for the *N*-mustards) that can be used for building a desired extension, thus offering several options for designing suitable cofactors for particular MTases. In the mTAG cofactors, the chemical variability of the side chain is foremost dictated by the activating allyl, propargyl, or aryl moieties, which demand quite distinct geometries of the cofactor pocket in both the ground and transition states. Of course, the length of the side chains and other chemical features can also influence the reaction to some extent, but such effects decline with increasing distance from the active site.

Representatives of all three classes of bacterial DNA MTases (6mA, 4mC, and 5mC forming enzymes) showed activity with certain types of extended cofactor analogs (see Table 19.2). On one end of the spectrum is M.TaqI, which demonstrated high tolerance with respect to a wide range of SMILing and mTAG cofactors examined (see Table 19.2 for references). More typically though, the efficiency of mTAG transalkylations with many wild-type enzymes is insufficient for routine applications. For 5mC-MTases, this issue was approached by engineering of the cofactor pocket of a well-characterized representative of the class, M.HhaI (Lukinavičius et al. 2012). The engineering effort was guided by a structure-based model of a M. HhaI-DNA-butynyl cofactor complex (Fig. 19.3), which suggested that the side chains of residues Gln82 and Asn304 (located in conserved sequence motives IV and X) and Tyr254 (located in a variable region) might sterically interfere with the extended transferable side chain, precluding cofactor binding or its proper orientation for catalysis. It turned out that double and triple Ala/Ser replacements at these positions conferred substantial improvements of the transalkylation activity in M.HhaI. The achieved turnover rates permitted complete derivatization of DNA in 15–30 min, which made the reaction suitable for routine laboratory applications. Detailed studies of the mutants showed that these replacements substantially enhanced the rate of alkyl transfer and also reduced the enzyme affinity towards the natural cofactor AdoMet and its product AdoHcy. Importantly, eM.HhaI can efficiently utilize extended synthetic analogs even in the presence of AdoMet (which is naturally abundant in cells and cell lysates), opening new possibilities for

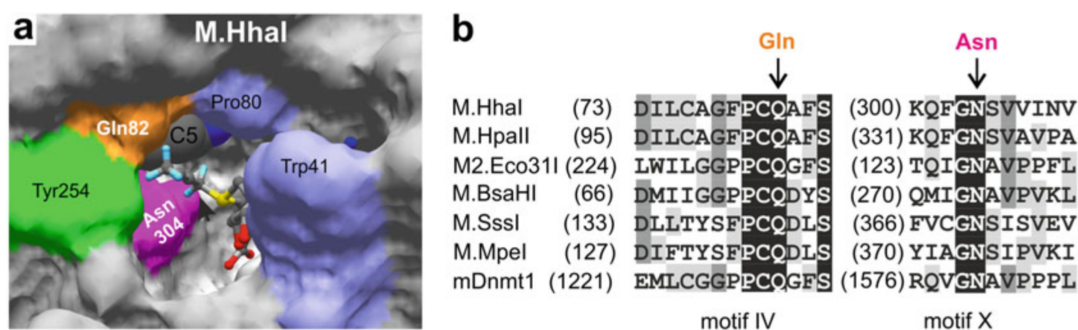


Fig. 19.3 Structure-guided engineering of DNA cytosine-C5 methyltransferases for the mTAG transalkylation reactions. (**a**) Model of an extended propargylic cofactor analog (AdoButyn, shown in ball and stick) bound in the active site of the HhaI MTase (based on M.HhaI-DNA-AdoMet ternary complex X-ray structure, PDB code 6mht, shown as space-fill). An arrow points at the transferable carbon atom. (**b**) Sequence alignment of regions corresponding to conserved motifs IV and X of sterically-engineered cytosine-5 MTases. Arrows indicate positions corresponding to Gln82 and Asn304 of M.HhaI. Adapted from Lukinavičius et al. (2012)

targeted covalent deposition of reporter groups onto DNA for a variety of ex vivo and in vivo applications (Lukinavičius et al. 2013).

The high structural conservation of 5mC-MTases suggested that other orthologs of M.HhaI can be similarly engineered based on sequence alignment even in the absence of crystal structures. Indeed, the double alanine mutants involving conserved motifs IV and X showed a significant improvement of the transalkylation activity with a wide range of propargyl-based cofactor analogs in the cases of M2.Eco31I and M.HpaII that recognize hexanucleotide and tetranucleotide target sites, respectively (Lukinavičius et al. 2012), as well as with M. SssI and M.MpeI acting on the 5′-CG-3′ dinucleotide (Kriukienė et al. 2013; Deen et al. 2019, Wilkinson et al. 2020). Most recently, a similar replacement of a single conserved residue in motif X of mouse Dnmt1 has been shown to confer a dramatic switch in cofactor selectivity in a much larger DNA 5mC-MTase (Stankevičius et al. 2022). On the other hand, analogous replacements in M.BsaHI showed no significant improvement in the transfer of allyl-based extended groups onto DNA (Vranken et al. 2014). This appears to agree with the observed weaker acceptance of double bond cofactors by the engineered M.HhaI (Lukinavičius, Lapinaitė, Klimašauskas, unpublished observations) and Dnmt1 (Stankevičius et al. 2022) variants, suggesting that the triple-bond cofactors are generally better compatible with the 5mC-MTases.

Interestingly, the engineering of adenine-*N*6 and cytosine *N*4 MTases for better cofactor acceptance did not gain much popularity. One reason is their lower sequence similarity, which complicates the identification of target residues in the absence of high-resolution structures. Another reason is that some of these enzymes are sufficiently active in their native form. For example, wild-type M.TaqI proved active with nearly all the known cofactor analogs, even very complex ones, in vitro, with no need for enzyme engineering. EcoDam wild type is also permissive (Flade et al. 2017). However, when it comes to applications in cell lysates or in vivo, competition between the natural and synthetic cofactors becomes a crucial factor. In this case, structure-based engineering of a single-residue V21A led to improved cofactor selectivity of M.TaqI not by enhancing the transalkylation, but rather by slowing down the methylation with AdoMet (Heimes et al. 2018).

DNA modification applications are increasingly moving towards the in vivo environment. However, this transition faces many more challenges besides MTase engineering to allow artificial modification in the presence of cellular AdoMet. Unfortunately, AdoMet and its analogs cannot travel through the cell membrane and are thus largely precluded from entering cells. The problem could be approached via enzymatic biosynthesis of AdoMet analogs inside the cell starting from corresponding methionine analogs by taking advantage of the natural metabolic pathway of the cell (Fig. 19.2). This has been shown to work to some extent for some small modifying groups (Hartstock et al. 2018; Shu et al. 2020), but larger transferable moieties require steric engineering of the MAT enzyme (Wang et al. 2013). However, the issue of cofactor specificity and competition with endogenous methionine and AdoMet still persists, and the solution may lie within a more elaborate engineering of the metabolic AdoMet synthesis pathway (Huber et al. 2020). A different solution to cofactor internalization has recently been proposed by Stankevičius et al., who used electroporation to achieve entry of a bulky exogenous cofactor, Ado-6-azide, into live mammalian cells (Stankevičius et al. 2022).

19.4 (Towards) Practical Implementation of MTase-Directed DNA Labeling in Genomic Research

Sequence-specific covalent derivatization and labeling of DNA potentially open new avenues in DNA research, diagnostics, and bionanotechnology. However, along with methodological developments of the MTase-directed labeling reactions, the properties and practical value of such covalently modified DNA

needed to be assessed. Many experimental demonstrations involving covalently tethered reporter and reactive groups have been performed at distinct levels ranging from oligonucleotides, DNA fragments, plasmids to whole genomes and epigenomes. These studies can be sorted into three main groups based on their popularity and level of development to those that exploited covalent derivatizations (1) for general manipulations and analysis of DNA; (2) for physical localization of individual target sites for submegabase-scale DNA genotyping and (3) for determination of their epigenetic modification states through label-specific next-generation sequencing analysis.

19.4.1 MTase-Directed Labeling for General Manipulations and Analysis of DNA

The exceptional specificity of DNA MTases towards DNA can firstly be used for selective covalent capture and extraction of DNA from complex biological mixtures (Artyukhin and Woo 2012). Optical tracking of a FRET (Förster resonance energy transfer) signal derived upon deposition of a second fluorophore onto DNA during the mTAG reaction, permits monitoring the catalytic activity of the MTase itself for applications such as inhibitor screening (Long et al. 2021). MTase-modified DNA endowed with unnatural chemical moieties offers a new toolbox for enzymology of DNA-interacting proteins. A range of extended groups installed at particular sites on DNA proved useful for probing the effects of structural variation on the activity of certain TET dioxygenases and glycosylases (Tomkuvienė et al. 2020), or for promoting assembly and repositioning of nucleosomes in vitro (Tomkuvienė et al. 2022). mTAG installation of photolabile nitrobenzylic groups on DNA permits their removal by UV illumination, thereby affording light-controlled manipulation of gene expression in vitro (Anhäuser et al. 2018; Heimes et al. 2018).

Another possible area of utility of the MTase-directed labeling is the construction of DNA-based nanostructures. In proof of principle studies, biotinylated aziridine cofactors together with M.TaqI and M.BseCI were used for targeted deposition of gold nanoparticles or for engineering three- and four-way junctions on model kilobase-sized DNA fragments, which were visualized using single-molecule AFM imaging (Braun et al. 2008; Wilkinson et al. 2008). Significant nanofabrication breakthroughs achieved in the DNA origami field (Wang et al. 2017) suggested that the functional capacity of DNA/RNA-based nanostructures could in principle be further expanded by their targeted decoration using MTase-directed labeling. However, first such attempts unveiled limitations of this approach related to the conformational rigidity and close packing of DNA helixes inside the DNA scaffolds interfering with the methyltransferase action (Heck et al. 2020). Subsequently, some of these limitations were overcome via covalent conjugation of a Cys-engineered variant of the M.TaqI MTase to a DNA origami which permitted templated covalent modification of a docked DNA fragment—a step towards “nano-manufacturing” devices (Weinhold and Chakraborty 2021).

An important branch of nanosciences deals with the analysis of biopolymers using nanopores. Installation of biotin, further appended with streptavidin, on DNA gave rise to distinct current blockade signals upon passage of the DNA through a solid-state nanopore. The signal allowed single-molecule genotyping of short genomes (Chen et al. 2017). Similarly, Gilboa et al. demonstrated a proof of principle of epigenetic analysis where M.TaqI was used to attach fluorophores at unmethylated TCGA sequences, in a mono-chromic or bi-color manner. Electrical and optical signals were then recorded simultaneously as individual labeled DNA molecules were passing through solid-state nanopores (Gilboa et al. 2016). The demonstrated ability to detect and quantify multiple colors/signals from single DNA molecules in nanopores holds promise of future analysis of distinct (epi)genetic features in clinical samples for potential diagnostic applications, either at high-throughput automated multiplex systems or/and, on contrary, simple portable devices.

Soon after the demonstration that both SMILing and mTAG techniques can achieve high sequence specificity of label incorporation into plasmid DNA (Pljevaljčić et al. 2007; Lukinavičius et al. 2007), the behavior of covalently labeled plasmid DNA was examined in cell lysates and in transfected cells. For example, an aziridine-based cofactor with a Cy3 fluorophore was used for labeling of pUC19 and pBR322 plasmids with M.TaqI; the plasmids were successfully transfected and optically tracked in mammalian cells (Schmidt et al. 2008). Independently, plasmids mTAG-derivatized with extended linear groups were shown to escape the McrBC restriction and confer transformation efficiencies similar to unmodified controls in *E. coli* cells (Lukinavičius et al. 2012). Moreover, robust sequence-specific two-step mTAG click labeling of endogenous plasmid DNA using eM.HhaI and Ado-6-azide cofactor was demonstrated in bacterial cell extracts (Lukinavičius et al. 2013). These experiments illustrated a rather high biological tolerance (bioorthogonality) of both types of covalent modifications pointing towards the potential suitability of this approach for in vivo studies. To this end, recent experiments indeed showed a highly selective azide tagging of the CG target sites in genomic DNA in vivo after electro-permiabilization of the Ado-6-azide cofactor into Dnmt1-engineered mouse ES cells (Stankevičius et al. 2022).

Altogether, for the purpose of general DNA labeling, the MTase-directed methods offer important advantages over other commonly used methods:

- Elimination of uncertainties related to loss of non-covalently bound labels in cells or in vitro experiments
- Control of the labeling density and positioning of reporter groups around (or away from) functional sites by selecting appropriate MTases
- High flexibility in selecting functional and reporter groups
- Covalent integrity of modified DNA strands (preserved supercoiling of plasmid DNA)
- Biological orthogonality of the underlying modifications

19.4.2 DNA Labeling for Analysis of Particular DNA Sites or Sequences

A genome or genomic region is uniquely identified by its nucleotide sequence (determined by sequencing); on the kilo-megabase scale, it can also be described by a characteristic distribution of specific shorter sequences (4–8 bp) along its length (Wand et al. 2019). The MTase-based approaches, owing to their unique combination of strict sequence specificity and covalent bonding, appear particularly suited for attaching detectable markers at specific short sequences in DNA. In the past decade, methods that use MTase-directed deposition of fluorophores for single-molecule optical mapping or tagging DNA with suitable primers for sequencing have undergone substantial progress. Certain modalities of these techniques also permit detection of naturally occurring epigenetic modifications. The utility of MTase-directed labeling for analysis of epigenomes lies in the selective covalent tagging of the unmodified fraction of target sites, whereas the naturally modified sites will remain untagged due to pre-existing modification of the target residue.

Below we describe recent advances and implementations of these techniques for analysis of genome structure and function and discuss their potential utility in scientific research and medical diagnostics.

19.4.2.1 Optical Mapping of DNA Sequences and Epigenetic States

Atomic force microscopy (Wilkinson et al. 2008) or electron microscopy (Kim et al. 2012, Kunkel et al. 2015) visualization of several bulky nano-objects along the irregular contour of a DNA molecule spotted on a mica surface gives a nice qualitative illustration, but is poorly suited for fast

parallel analysis of DNA molecules containing a large number of target sites. Direct determination of physical distances (positioning) between the specific sites becomes possible on stretched-out DNA molecules, leading to a visual pattern characteristic of that particular DNA. Such a linear representation of a DNA sequence, called optical map, can be read as a barcode and analyzed with a high degree of automation (Fig. 19.4a). However, implementation of various known methods for optical DNA map generation is dependent on many technical parameters related to the degree and accuracy of label incorporation, repertoire of available target sites, covalent continuity of labeled DNA strands, inhomogeneous stretching, chemical and physical stability of the fluorophores, resolution and speed of signal readout, etc. Other methods for specific visual pattern generation in optical DNA mapping include restriction map generation (Teague et al. 2010), nick-translation (Lam et al. 2012), or probe hybridization (Weier et al. 1995), all of which suffer from one or more of the above-listed limitations (discussed in Zohar and Muller 2011 and Levy-Sakin and Ebenstein 2013).

In a proof of principle study, two-step mTAG labeling was employed to attach fluorophores on 215 HhaI sites in bacteriophage lambda DNA (48.5 kb, see Fig. 19.4a) (Neely et al. 2010). The labeling employed engineered M.HhaI and a cofactor bearing a transferable linear side chain with a terminal amino group followed by a chemoselective attachment of an Atto647N dye. The DNA molecules were stretched by combing onto polymer-coated coverslips using an evaporating droplet technique. Positions of fluorophores along individual DNA strands were recorded at sub-diffraction resolution (10 nm, or just 20 bp) using dSTORM imaging, which utilized photobleaching of the fluorophores to ensure that single emitters are isolated and their positions accurately determined. An average density of localized sites of around one per 650 bases represented 34% of the 215 available HhaI sites on the DNA and thus a consensus "fluorocode" encompassing nearly 90% of the sites (density 1/270 bp) was generated from twenty automatically aligned molecules.

To further explore the technicalities of the labeling reaction for improved optical mapping of DNA, a cofactor carrying a short allylic side chain with a terminal alkyne group (AdoEnYn) was used along with CuAAC mediated attachment of a fluorophore (Vranken et al. 2014). Eleven MTases were screened for activity with this cofactor, of which three adenine-specific enzymes were found to be active. The CuAAC-based approach generated bacteriophage T7 fluorocodes with labeling efficiency reaching 70%, however, the authors noted substantial degradation of DNA, which precluded generation of full-length labeled DNA molecules. An inherent stability of the SPAAC reactants and the mild reaction conditions allowed long reaction times and a relatively high degree of labeling without inducing DNA damage or other topological abnormalities (Lauer et al. 2017). Wand et al. further mastered "DNA barcoding" to analyze and sort out complex mixtures of genomic material. For that purpose, they used M.TaqI with Ado-6-azide and two-step fluorophore labeling, involving SPAAC, with 90% efficiency of target site coverage. The potential applicability of the approach was demonstrated by exemplary detection of a model viral infection in human cells, identification of bacterial strains, and visualization of edited loci in a genome (Wand et al. 2019).

Further in pursuit of high-throughput routine analyses, an automated DNA stretching and imaging platform, which was previously applied for non-MTase-dependent optical mapping, appeared essential. To this end, one-step mTAG labeling with M.TaqI and a propargylic cofactor carrying a linker-bound TAMRA fluorophore was used to label over 200 target sites on lambda DNA (Grunwald et al. 2015), followed by physical stretching of the DNA molecules in commercial microchip-based nanochannel arrays. A fluorescent signature of DNA was generated by "conventional resolution" imaging, i.e., measuring the amplitude modulations of fluorescence intensity along its length rather than isolated fluorescent spots. However, unlike in case of immobilized DNA, the dynamic motion of markers due to thermal fluctuations of nanochannel-trapped DNA, sets a resolution

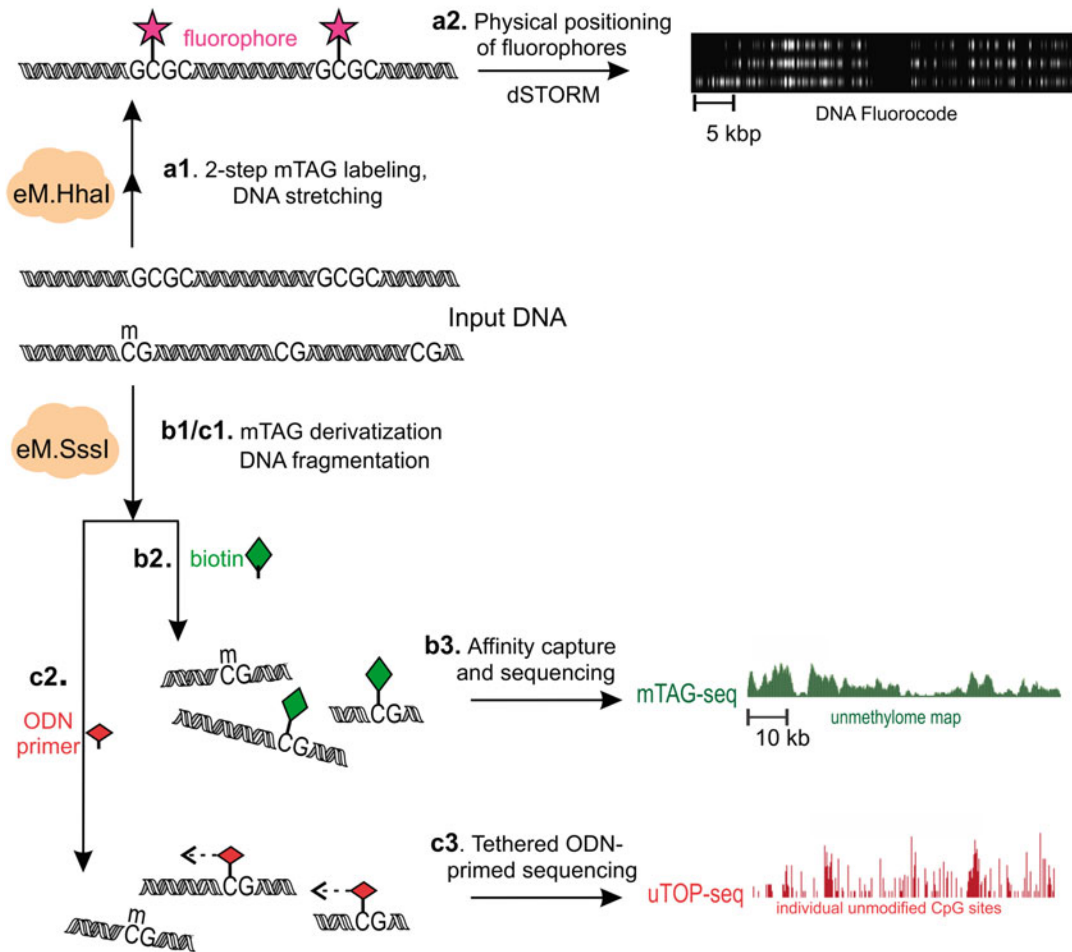


Fig. 19.4 Major applications of methyltransferase-directed labeling in genome studies. (a) Optical DNA mapping using fluorescent mTAG labeling. A two-step mTAG reaction involving an engineered version of HhaI methyltransferase (eM.HhaI) was used for fluorescent labeling of GCGC sites in bacteriophage lambda DNA. The labeled DNA molecules were stretched out by combing and positions of the fluorophores on individual DNA molecules were determined using super-resolution dSTORM imaging. Illustration on the right shows: (top) two lanes of experimental consensus fluorocodes derived by using different processing parameters; (bottom) a lane of in silico generated (theoretical) reference map (adapted from Neely et al. 2010). (b) DNA "unmethylome" profiling by covalent mTAG labeling of unmodified CG sites. Unmodified CG sites in fragmented genomic DNA from the human brain (50–300 bp fragments) are biotin-tagged in a two-step mTAG labeling reaction involving engineered eM.SssI MTase. Biotin-tagged fragments are affinity-enriched and sequenced to produce a genome-wide profile of unmodified CG sites. Illustration on the right shows an exemplary genome browser view of a mTAG-seq map (Kriukienė et al. 2013). (c) DNA "unmethylome" profiling by TOP-seq that enables genomic sequencing primed right from the MTase-tagged modification sites. This is achieved by tethering a DNA oligonucleotide (ODN) primer at the azide-derivatized unmodified genomic CG sites followed by in situ non-homologous priming of the DNA strand synthesis from the tethered ODN by DNA polymerase. C^m refers to a naturally modified cytosine. Illustration on the right shows an exemplary genome browser view of a uTOP-seq map (Kriukienė et al., unpublished data)

limit to ~1.5–3 kbp (which can be improved to ~700 bp by pairwise distance recording between the target and a close-by marker) (Jeffet et al. 2016). Nevertheless, the generality, rapidness and high-throughput capabilities make this concept promising for routine applications. Direct single-molecule analysis of large DNA fragments, which far exceed the read length of widely used sequencing technologies, provides valuable genomic information for the identification of structural or copy number variations and assists with DNA sequence assembly or rapid strain typing (reviewed in (Yuan et al. 2020; Jeffet et al. 2021), as well as analysis of difficult to sequence regions (Wight et al. 2020). With its ability to precisely visualize structural variations in megabase-long stretches of genomic DNA from patient samples, it is poised to replace classical cytogenetic assays as karyotyping, fluorescence in situ hybridization and copy number variation microarrays used routinely in medical diagnostics of genetic diseases (Mantere et al. 2021; Neveling et al. 2021). Several partially or fully automated optical DNA mapping platforms (BioNano Genomics, GenomicVision, Nabsys) are already available. A dramatic improvement in read length (50-fold longer on average, >2 Mbp native molecules often seen) was achieved by replacing a nicking enzyme-based labeling with "Direct labeling enzyme-1," presumably a MTase, that enables one-step CTTAAG-specific fluorescent labeling of genomic DNA under very mild conditions avoiding strand cleavage (Bionano Genomics https://bionanogenomics.com/technology/dls-technology).

Moreover, as noted above, DNA MTases (such as M.SssI and M.TaqI) are not only active at specific sequences, but are also sensitive to certain base modifications occurring at these sequences and thus can be exploited to interrogate the epigenetic states of CG target sites in mammalian genomes (reviewed in Heck et al. 2019). As an example, M.TaqI optical mapping was used for differentiating two cell types based on their methylation profiles, and showed that pairwise analysis encompassing both promoters and their distal enhancers yields a much better deconvolution than short-ranged approaches (Margalit et al. 2021). Furthermore, in combination with other methods, optical mapping with M.TaqI was used to detect dynamic changes of epigenetic DNA marks (methylated, hydroxymethylated, and unmethylated cytosine) related to macrophage response during pro-inflammatory activation (Jain et al. 2019). Finally, both DNA sequence and methylation status were simultaneously mapped on the same molecule, enabling recording of the haplotype, copy number, and methylation status of a highly repetitive locus associated with facioscapulohumeral muscular dystrophy (Sharim et al. 2019).

Altogether, these examples show that optical mapping can potentially deliver the benefits of single-molecule sensitivity, long reads, and high-throughput automation, but suffers from limited resolution, higher noise, and reduced analytical representation (M.TaqI can only access a fraction of CG sites). In the future, simultaneous characterization of both genome structural variation and methylation may lead to new generation research and diagnostics tools.

19.4.2.2 Applications of MTase-Directed Labeling in Epigenomics

Besides optical mapping described above, MTase-directed covalent DNA labeling offers an important entry point to mainstream sequencing-based genomic analysis of epigenetic DNA marks. As mentioned above, DNA cytosine-5 methylation in higher eukaryotes including mammals predominantly occurs at CG dinucleotides and is involved in many biological processes such as embryogenesis, establishment of cell identity and fate, and development of various pathological conditions, including cancer (see Chap. 15). Profiling the modification status of tens of millions of CG sites in the genome is a challenging task, and numerous epigenomic techniques have been developed that differ in their throughput, sensitivity, resolution, and cost (see Chap. 16). Covalent tagging of epigenetic modifications via incorporation of reactive azide, keto, or primary amine groups followed by chemoselective conjugation of biotin has

opened new avenues for highly sensitive and selective enrichment of modified DNA (Song et al. 2011; Pastor et al. 2011; Zhang et al. 2013).

The key concept of using MTase-directed labeling for analysis of epigenome lies in the selective covalent derivatization of the unmodified fraction of CG sites. An early attempt to analyze DNA methylation sites through targeted DNA scission (Comstock and Rajski 2005b) used derivatization of model oligodeoxynucleotides substrates with M.TaqI or M.HhaI and an azide-bearing aziridine cofactor. However, this chemistry leads to extensive DNA damage and proved incompatible with modern sequencing techniques. The first genome-wide chemo-enzymatic profiling of CG methylation was developed on the basis of selective two-step covalent biotin tagging directed by a CG-specific MTase (Kriukienė et al. 2013) (Fig. 19.4b). Covalent tagging of unmethylated target sites was performed using an engineered version of the M. SssI MTase, eM.SssI, and a synthetic AdoMet analog carrying a terminal amine or azide group. In the next step, conventional chemoselective coupling of the amine group with an NHS-biotin probe or, alternatively, strain-promoted alkyne-azide cycloaddition of the attached azide group with a dibenzocyclooctyne biotin reagent was employed, and the enriched biotin-labeled DNA fragments were analyzed on tiling DNA microarrays (mTAG-chip) or by next-generation sequencing (mTAG-seq). Due to covalent linkage of modified sites, the approach offered robust analysis of CG methylation and a resolution limit of 200–500 bp.

To push the resolution limit down to a single CG site, an advanced version of the latter approach, TOP-seq, was developed that enabled genomic sequencing primed right from the MTase-tagged modification sites. This was achieved by tethering a DNA oligonucleotide (instead of biotin) to the azide-derivatized unmodified genomic CG sites followed by in situ non-homologous priming of the DNA strand synthesis from the tethered oligonucleotide by DNA polymerase (Fig. 19.4c). This recently discovered priming reaction affords direct read-out and subsequent mapping of adjoining genomic regions. Owing to the robust and non-destructive nature of covalent tagging, the produced uTOP-seq epigenomic maps of unmethylated CG sites proved instrumental for discerning subtle tissue-specific methylation differences on a local or whole genome scale. To date, the general TOP-seq concept has been implemented for analysis of the unmodified (Staševskij et al. 2017; Gordevičius et al. 2020; Narmontė et al. 2021), hydroxymethylated (Hu et al. 2019; Gordevičius et al. 2020; Gibas et al. 2020) or carboxylated CG sites in DNA (Ličytė et al. 2020). The latter approach combines two distinct reactions of the eM.SssI MTase: removal of carboxyl groups from 5-carboxylated CG sites (Liutkevičiūtė et al. 2009, 2014; see also Sect. 19.4), yielding the unmodified CG sites, and their immediate tagging by an azide group upon addition of a synthetic cofactor Ado-6-azide. Notably, both mTAG-seq and TOP-seq-based approaches derive sequencing reads only from DNA fragments of interest (containing unmodified CG sites or epigenetic modifications) and thus demand less sequencing and analytical efforts (and cost), as compared to the gold standard bisulfite modification-based methods, which typically entail whole genome sequencing. For example, an adaptation of the uTOP-seq protocol for a fetal karyotype analysis using minuscule amounts of fetal cell-free DNA circulating in maternal blood enabled detection of a fetal trisomy of chromosome 21 from a few nanograms of maternal circulating DNA and 3 million sequencing reads (Gordevičius et al. 2020).

Altogether, TOP-seq has proven as a semi-quantitative laboratory platform that can infer relative modification of each CG in purified genomic DNA samples. As mentioned above, mammalian DNA methylation profiles are generated by collective action of three DNA methyltransferases (DNMT1, DNMT3A, and DNMT3B, see Chap. 4), whose catalytic interactions and temporal interplay in establishing and maintaining genomic methylation during cell growth and differentiation are poorly understood. The first example of selective tracking of the catalytic contribution of an individual DNMT enzyme has recently been demonstrated through

genomic installation of an engineered Dnmt1 variant in mouse embryonic stem cells (mESC) and pulse-internalization of the Ado-6-azide cofactor by electroporation. This permitted selective tagging of Dnmt1-specific genomic targets with azide groups in vivo (Stankevičius et al. 2022) which were exploited as click handles for reading adjoining sequences and precise mapping of the chemically tagged methylation sites in the genome (as described for TOP-seq above). This new general approach, named Dnmt1-TOP-seq, offers selective high-resolution genomic tracking of the Dnmt1 catalysis in live mammalian cells, paving the way to similar studies of other biologically important AdoMet-dependent methyltransferases in a wide range of eukaryotic model systems.

19.5 Cofactor-Independent MTase-Directed Labeling

In addition to their well-characterized catalytic activity, 5mC-MTases were found to catalyze atypical reactions involving non-cofactor-like substrates. As mentioned above (Sect. 19.3), the 5mC-MTases use a covalent mechanism for nucleophilic activation of their target cytosine residues. The transiently generated activated cytosine intermediate is not only active towards AdoMet or its synthetic analogs, but can also attack other exogenous electrophiles such as aliphatic aldehydes, yielding corresponding 5-α-hydroxyalkylcytosines (Liutkevičiūtė et al. 2009) (Fig. 19.5a). The reactions occur under fairly mild conditions and retain the high sequence and base specificity characteristic of bacterial DNA MTases. The coupling with formaldehyde yields 5-hydroxymethylcytosine (hmC), which is a naturally occurring cytosine modification in mammalian DNA (Kriaucionis and Heintz 2009; Tahiliani et al. 2009). Although the hydroxymethyl groups themselves are not good chemical reporters, they add a unique functionality to DNA (analogous to benzylic hydroxyl) that can be exploited for chemical or enzymatic derivatization (Jurkowski 2020). Alternatively, hmC residues can be enzymatically glucosylated (Gommers-Ampt and Borst 1995), thereby permitting selective DNA labeling through application of glycan modification/recognition techniques (Chittaboina et al. 2005; Song et al. 2011; Gibas et al. 2020).

Curiously, it was also found that the covalent activation of 5-substituted cytosine residues present at the target position of a 5mC-MTase, can lead to their conversion into unmodified cytosine (Fig. 19.5a). This reaction does occur with hmC and 5-carboxylcytosine, but was not observed with 5-formylcytosine (Liutkevičiūtė et al. 2009, 2014). The MTase-activated hmC in DNA can also undergo condensation with exogenous aliphatic thiols and selenols yielding corresponding 5-alkylchalcogenomethyl derivatives (Liutkevičiūtė et al. 2011) (Fig. 19.5a). Since this MTase-directed derivatization reaction is not possible at 5-methylated and unmodified cytosine residues, it appears well-suited for selective covalent capture 5-hydroxymethylated-CG sites in mammalian genomic DNA. As a proof of concept, 5mC-MTase-directed derivatization of hmC with cysteamine and subsequent amine-selective biotin labeling (Fig. 19.5b) was demonstrated on plasmid DNA and model DNA fragments (Liutkevičiūtė et al. 2011). M.HhaI and M.SssI have also been shown to render sequence-specific conjugation of short Cys-containing peptides, to hmC-containing DNA (Serva and Lagunavičius 2015). Moreover, cofactor-independent functionalization of hmC has been shown applicable for TOP-seq approach as described in Sect. 19.4.2.2 (Gibas et al. 2020).

Altogether, the presented variety of atypical reactions demonstrate a high catalytic plasticity of DNA 5mC-MTases and offer additional ways for sequence-specific derivatization of canonical and modified bases within DNA. As compared to the cofactor-based reactions, these reactions typically require simpler and less expensive compounds, thereby avoiding multistep syntheses of AdoMet analogs.

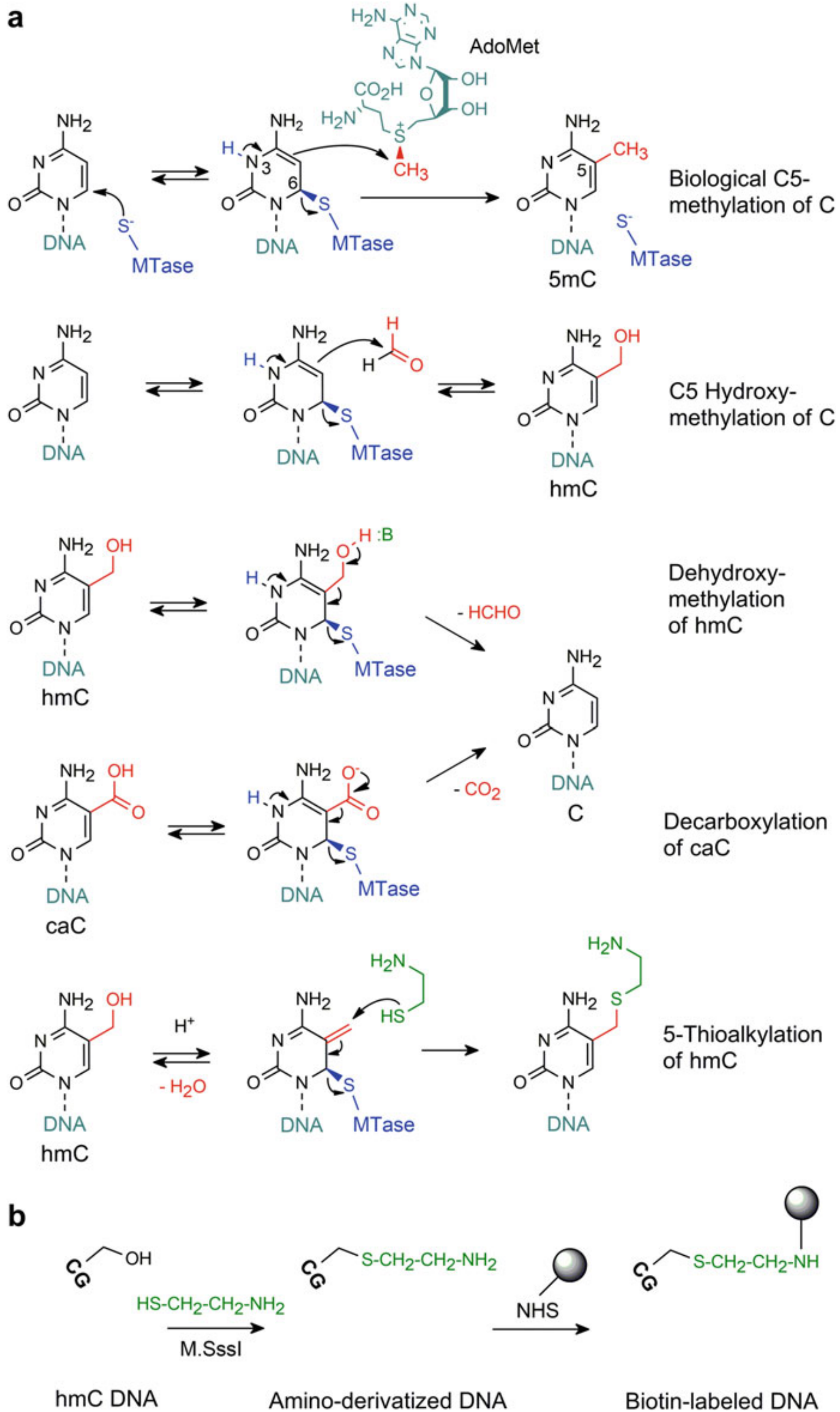


Fig. 19.5 Cofactor-independent methyltransferase-directed sequence-specific derivatization of DNA. (**a**) Transformations of a target cytosine catalyzed by 5mC-MTases. Biological methylation by 5mC-MTases occurs via an S_N2 reaction between an activated cytosine intermediate (ACI) and cofactor AdoMet, yielding

19.6 Conclusions/Outlook

MTase-directed labeling of DNA is an enabling technology with many unique demonstrated applications. Due to its relative simplicity, robustness, and wide-range applicability, this approach is becoming a method of choice where targeted covalent derivatization of DNA is required. Although certain technical questions still require attention, the rapidly growing popularity indicates that the field is approaching its maturity stage. The list of available and well-performing MTases and their variants is growing in response to demand from the scientific community. The repertoire of currently known synthetic cofactor analogs is impressive, although the lack of commercial sources is somewhat holding back the spread of applications. The two most advanced areas of MTase-directed labeling are optical DNA mapping and analysis of epigenetic states in mammalian DNA; both methods are now nearing/entering the phases of automation and commercial exploitations. Yet another important field of research that is starting to see a rapid growth is DNA labeling in biological systems and in living cells. Certain solutions have now been found to the two previously noted obstacles: (1) delivery of cofactor analogs into cells; and (2) design of orthogonal cofactor-MTases pairs for allele-specific labeling. The first issue is being addressed by harnessing enzymatic production of cofactor analogs in situ from corresponding methionine analogs, which show superior wall-penetration properties (Wang et al. 2013; Singh et al. 2014; Hartstock et al. 2018), or by controlled internalization of exogenous cofactors using electroporation (Stankevičius et al. 2022). Solutions to the second issue are being sought by designing MTase variants with preferential cofactor selectivity towards extended AdoMet analogs, which can provide substantial levels of DNA labeling in the presence of endogenous AdoMet (Lukinavičius et al. 2013, Heimes et al. 2018; Stankevičius et al. 2022).

Acknowledgments This work was supported by grants from the European Research Council (ERC-AdG-2016/742654 to S.K.) and by the European Social Fund according to the activity measure No. 09.3.3-LMT-K-712 “Improvement of researchers’ qualification by implementing world-class R & D projects” under grant agreement with the Research Council of Lithuania (grant number 09.3.3-LMT-K-712-01-0041 to E.K.)

References

Anhäuser L, Muttach F, Rentmeister A (2018) Reversible modification of DNA by methyltransferase-catalyzed transfer and light-triggered removal of photo-caging groups. Chem Commun 54:449–451

Artyukhin AB, Woo YH (2012) DNA extraction method with improved efficiency and specificity using DNA methyltransferase and “click” chemistry. Anal Biochem 425(2):169–174

Blum G, Islam K, Luo M (2013) Bioorthogonal profiling of protein methylation (BPPM) using an azido analog of S-adenosyl-L-methionine. Curr Protoc Chem Biol 5(1):45–66

Bothwell IR, Luo M (2014) Large-scale, protection-free synthesis of Se-adenosyl-L-selenomethionine analogues and their application as cofactor surrogates of methyltransferases. Org Lett 16(11):3056–3059

Bothwell IR, Islam K, Chen Y, Zheng W, Blum G, Deng H, Luo M (2012) Se-adenosyl-L-selenomethionine cofactor analogue as a reporter of protein methylation. J Am Chem Soc 134(36): 14905–14912

Braun G, Diechtierow M, Wilkinson S, Schmidt F, Huben M, Weinhold E, Reich NO (2008) Enzyme-

Fig. 19.5 (continued) 5-methylcytosine (5mC) (Biological C5-methylation). The ACI can undergo nucleophilic addition reaction with short aliphatic exogenous aldehydes, which in the case of formaldehyde yields hmC (C5-hydroxymethylation). In the reverse reaction, hmC residues can be converted to unmodified cytosines in DNA by the enzyme (Dehydroxymethylation of hmC). Similarly, 5-carboxycytosine (caC) can be converted to cytosine (Decarboxylation of caC). hmC residues, including those naturally occurring in DNA, can undergo further methyltransferase-directed condensation with thiol or selenol reagents to give stable 5-alkyl chalcogenomethyl derivatives (5-Thioalkylation of hmC). Modifying reagents are shown in red and green (thiol), 5mC-MTase and its catalytic moieties are shown in blue. (**b**) MTase-directed covalent amino-derivatization and labeling of hmCG dinucleotides in DNA with biotin (shown as a ball)

directed positioning of nanoparticles on large DNA templates. Bioconjug Chem 19(2):476–479

Chen K, Juhasz M, Gularek F, Weinhold E, Tian Y, Keyser UF, Bell NAW (2017) Ionic current-based mapping of short sequence motifs in single DNA molecules using solid-state nanopores. Nano Lett 17: 5199–5205

Chittaboina S, Xie F, Wang Q (2005) One-pot synthesis of triazole-linked glycoconjugates. Tetrahedron Lett 46(13):2331–2336

Comstock LR, Rajski SR (2005a) Conversion of DNA methyltransferases into azidonucleosidyl transferases via synthetic cofactors. Nucleic Acids Res 33(5): 1644–1652

Comstock LR, Rajski SR (2005b) Methyltransferase-directed DNA strand scission. J Am Chem Soc 127(41):14136–14137

Dalhoff C, Lukinavičius G, Klimašauskas S, Weinhold E (2006a) Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. Nat Chem Biol 2(1):31–32

Dalhoff C, Lukinavičius G, Klimašauskas S, Weinhold E (2006b) Synthesis of S-adenosyl-L-methionine analogs and their use for sequence-specific transalkylation of DNA by methyltransferases. Nat Protoc 1(4):1879–1886

Deen J, Wang S, Van Snick S, Leen V, Janssen K, Hofkens J, Neely RK (2019) A general strategy for direct, enzyme-catalyzed conjugation of functional compounds to DNA. Nucleic Acids Res 46:e64

Du Y, Hendrick CE, Frye KS, Comstock LR (2012) Fluorescent DNA labeling by N-mustard analogues of S-adenosyl-L-methionine. Chembiochem 13(15): 2225–2233

Flade S, Jasper J, Giess M, Juhasz M, Dankers A, Kubik G, Koch O, Weinhold E, Summerer D (2017) The N6-position of adenine is a blind spot for TAL-effectors that enables effective binding of methylated and fluorophore-labeled DNA. ACS Chem Biol 12:1719–1725

Gabrieli T, Michaeli Y, Avraham S, Torchinsky D, Juhasz M, Coruh C, Arbib N, Zhou ZS, Law JA, Weinhold E, Ebenstein Y (2021) Chemoenzymatic labeling of DNA methylation patterns for single-molecule epigenetic mapping. bioRxiv. https://doi.org/10.1093/nar/gkac460

Gerasimaitė R, Vilkaitis G, Klimašauskas S (2009) A directed evolution design of a GCG-specific DNA hemimethylase. Nucleic Acids Res 37(21):7332–7341

Gibas P, Narmontė M, Staševskij Z, Gordevičius J, Klimašauskas S, Kriukienė E (2020) Precise genomic mapping of 5-hydroxymethylcytosine via covalent tether-directed sequencing. PLoS Biol 18(4):e3000684

Gilboa T, Torfstein C, Juhasz M, Grunwald A, Ebenstein Y, Weinhold E, Meller A (2016) Single-molecule DNA methylation quantification using electro-optical sensing in solid-state nanopores. ACS Nano 10:8861–8870

Gommers-Ampt JH, Borst P (1995) Hypermodified bases in DNA. FASEB J 9(11):1034–1042

Gordevičius J, Narmontė M, Gibas P, Kvederavičiūtė K, Tomkutė V, Paluoja P, Krjutškov K, Salumets A, Kriukienė E (2020) Identification of fetal unmodified and 5-hydroxymethylated CG sites in maternal cell-free DNA for non-invasive prenatal testing. Clin Epigenet 12:153

Goyvaerts V, Van Snick S, D'Huys L, Vitale R, Helmer Lauer M, Wang S, Leen V, Dehaen W, Hofkens J (2020) Fluorescent SAM analogues for methyltransferase based DNA labeling. Chem Commun 56:3317–3320

Grunwald A, Dahan M, Giesbertz A, Nilsson A, Nyberg LK, Weinhold E, Ambjornsson T, Westerlund F, Ebenstein Y (2015) Bacteriophage strain typing by rapid single molecule analysis. Nucleic Acids Res. 43(18):e117

Guo H, Wang R, Zheng W, Chen Y, Blum G, Deng H, Luo M (2014) Profiling substrates of protein arginine N-methyltransferase 3 with S-adenosyl-L-methionine analogues. ACS Chem Biol 9(2):476–484

Hartstock K, Nilges BS, Ovcharenko A, Cornelissen NV, Püllen N, Lawrence-Dörner A-M, Leidel SA, Rentmeister A (2018) Enzymatic or in vivo installation of propargyl groups in combination with click chemistry for the enrichment and detection of methyltransferase target sites in RNA. Angewandte Chemie (International ed. in English) 57:6342–6346

Heck C, Michaeli Y, Bald I, Ebenstein Y (2019) Analytical epigenetics: single-molecule optical detection of DNA and histone modifications. Curr Opin Biotechnol 55:151–158

Heck C, Torchinsky D, Nifker G, Gularek F, Michaeli Y, Weinhold E, Ebenstein Y (2020) Label as you fold: methyltransferase-assisted functionalization of DNA nanostructures. Nanoscale 12:20287–20291

Heimes M, Kolmar L, Brieke C (2018) Efficient cosubstrate enzyme pairs for sequence-specific methyltransferase-directed photolabile caging of DNA. Chem Commun 54:12718–12721

Holstein JM, Schulz D, Rentmeister A (2014) Bioorthogonal site-specific labeling of the 5′-cap structure in eukaryotic mRNAs. Chem Commun 50(34): 4478–4481

Hu L, Liu Y, Han S, Yang L, Cui X, Gao Y, Dai Q, Lu X, Kou X, Zhao Y, Sheng W, Gao S, He X, He C (2019) Jump-seq: genome-wide capture and amplification of 5-hydroxymethylcytosine sites. J Am Chem Soc 141: 8694–8697

Huber TD, Clinger JA, Liu Y, Xu W, Miller MD, Phillips GNJ, Thorson JS (2020) Methionine adenosyltransferase engineering to enable bioorthogonal platforms for AdoMet-utilizing enzymes. ACS Chem Biol 15:695–705

Hymbaugh Bergman SJ, Comstock LR (2015) N-mustard analogs of S-adenosyl-L-methionine as biochemical probes of protein arginine methylation. Bioorg Med Chem 23(15):5050–5055

Islam K, Zheng W, Yu H, Deng H, Luo M (2011) Expanding cofactor repertoire of protein lysine methyltransferase for substrate labeling. ACS Chem Biol 6(7):679–684

Islam K, Bothwell I, Chen Y, Sengelaub C, Wang R, Deng H, Luo M (2012) Bioorthogonal profiling of protein methylation using azido derivative of S-adenosyl-L-methionine. J Am Chem Soc 134(13): 5909–5915

Islam K, Chen Y, Wu H, Bothwell IR, Blum GJ, Zeng H, Dong A, Zheng W, Min J, Deng H, Luo M (2013) Defining efficient enzyme-cofactor pairs for bioorthogonal profiling of protein methylation. Proc Natl Acad Sci USA 110(42):16778–16783

Jain N, Shahal T, Gabrieli T, Gilat N, Torchinsky D, Michaeli Y, Vogel V, Ebenstein Y (2019) Global modulation in DNA epigenetics during pro-inflammatory macrophage activation. Epigenetics 14:1183–1193

Jeffet J, Kobo A, Su T, Grunwald A, Green O, Nilsson AN, Eisenberg E, Ambjörnsson T, Westerlund F, Weinhold E, Shabat D, Purohit PK, Ebenstein Y (2016) Super-resolution genome mapping in silicon nanochannels. ACS Nano 10:9823–9830

Jeffet J, Margalit S, Michaeli Y, Ebenstein Y (2021) Single-molecule optical genome mapping in nanochannels: multidisciplinarity at the nanoscale. Ess Biochem 65:51–66

Jurkowski TP (2020) Technologies and applications for the assessment of 5-hydroxymethylcytosine. In: Epigenetics methods, Vol.18 in translational epigenetics, vol 13, pp 261–278. https://doi.org/10.1016/B978-0-12-819414-0.00013-6

Kim S, Gottfried A, Lin RR, Dertinger T, Kim AS, Chung S, Colyer RA, Weinhold E, Weiss S, Ebenstein Y (2012) Enzymatically incorporated genomic tags for optical mapping of DNA-binding proteins. Angewandte Chemie 51(15):3578–3581

Klimašauskas S, Weinhold E (2007) A new tool for biotechnology: AdoMet-dependent methyltransferases. Trends Biotechnol 25(3):99–104

Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324(5929):929–930

Kriukienė E, Labrie V, Khare T, Urbanavičiūtė G, Lapinaitė A, Koncevičius K, Li D, Wang T, Pai S, Ptak C, Gordevičius J, Wang SC, Petronis A, Klimašauskas S (2013) DNA unmethylome profiling by covalent capture of CpG sites. Nat Commun 4:2190

Kunkel F, Lurz R, Weinhold E (2015) A 7-deazaadenosylaziridine cofactor for sequence-specific labeling of DNA by the DNA cytosine-C5 methyltransferase M.HhaI. Molecules 20(11): 20805–20822

Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, Deshpande P, Cao H, Nagarajan N, Xiao M, Kwok PY (2012) Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol 30(8):771–776

Lauer MH, Vranken C, Deen J, Frederickx W, Vanderlinden W, Wand N, Leen V, Gehlen MH, Hofkens J, Neely RK (2017) Methyltransferase-directed covalent coupling of fluorophores to DNA. Chem Sci 8:3804–3811

Lee BW, Sun HG, Zang T, Kim BJ, Alfaro JF, Zhou ZS (2010) Enzyme-catalyzed transfer of a ketone group from an S-adenosylmethionine analogue: a tool for the functional analysis of methyltransferases. J Am Chem Soc 132(11):3642–3643

Levy-Sakin M, Ebenstein Y (2013) Beyond sequencing: optical mapping of DNA in the age of nanotechnology and nanoscopy. Curr Opin Biotechnol 24(4):690–698

Ličytė J, Gibas P, Skardžiūtė K, Stankevičius V, Rukšėnaitė A, Kriukienė E (2020) A bisulfite-free approach for base-resolution analysis of genomic 5-carboxylcytosine. Cell Rep 32

Lipson JM, Thomsen M, Moore BS, Clausen RP, La Clair JJ, Burkart MD (2013) A tandem chemoenzymatic methylation by S-adenosyl-L-methionine. ChemBioChem 14:950–953

Liutkevičiūtė Z, Lukinavičius G, Masevičius V, Daujotytė D, Klimašauskas S (2009) Cytosine-5-methyltransferases add aldehydes to DNA. Nat Chem Biol 5(6):400–402

Liutkevičiūtė Z, Kriukienė E, Grigaitytė I, Masevičius V, Klimašauskas S (2011) Methyltransferase-directed derivatization of 5-hydroxymethylcytosine in DNA. Angew Chem Int Ed Engl 50(9):2090–2093

Liutkevičiūtė Z, Kriukienė E, Ličytė J, Rudytė M, Urbanavičiūtė G, Klimašauskas S (2014) Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. J Am Chem Soc 136(16): 5884–5887

Long Y, Ubych K, Jagu E, Neely RK (2021) FRET-based method for direct, real-time measurement of DNA methyltransferase activity. Bioconjug Chem 32:192–198

Lukinavičius G, Lapienė V, Staševskij Z, Dalhoff C, Weinhold E, Klimašauskas S (2007) Targeted labeling of DNA by methyltransferase-directed transfer of activated groups (mTAG). J Am Chem Soc 129(10): 2758–2759

Lukinavičius G, Lapinaitė A, Urbanavičiūtė G, Gerasimaitė R, Klimašauskas S (2012) Engineering the DNA cytosine-5 methyltransferase reaction for sequence-specific labeling of DNA. Nucleic Acids Res 40(22):11594–11602

Lukinavičius G, Tomkuvienė M, Masevičius V, Klimašauskas S (2013) Enhanced chemical stability of adomet analogues for improved methyltransferase-directed labeling of DNA. ACS Chem Biol 8(6): 1134–1139

Mai V, Comstock LR (2011) Synthesis of an azide-bearing N-mustard analogue of S-adenosyl-L-methionine. J Org Chem 76(24):10319–10324

Mantere T, Neveling K, Pebrel-Richard C, Benoist M, van der Zande G, Kater-Baats E, Baatout I, van Beek R, Yammine T, Oorsprong M, Hsoumi F, Olde-

Weghuis D, Majdali W, Vermeulen S, Pauper M, Lebbar A, Stevens-Kroef M, Sanlaville D, Dupont JM, Smeets D, Hoischen A, Schluth-Bolard C, El Khattabi L (2021) Optical genome mapping enables constitutional chromosomal aberration detection. Am J Hum Genet 108:1409–1422

Margalit S, Abramson Y, Sharim H, Manber Z, Bhattacharya S, Chen YW, Vilain E, Barseghyan H, Elkon R, Sharan R, Ebenstein Y (2021) Long reads capture simultaneous enhancer–promoter methylation status for cell-type deconvolution. Bioinformatics 37: I327–I333

Michailidou F, Klöcker N, Cornelissen NV, Singh RK, Peters A, Ovcharenko A, Kümmel D, Rentmeister A (2021) Engineered SAM synthetases for enzymatic generation of AdoMet analogs with photocaging groups and reversible DNA modification in cascade reactions. Angew Chem Int Ed Eng 60:480–485

Motorin Y, Burhenne J, Teimer R, Koynov K, Willnow S, Weinhold E, Helm M (2011) Expanding the chemical scope of RNA:methyltransferases to site-specific alkynylation of RNA for click labeling. Nucleic Acids Res 39(5):1943–1952

Muttach F, Rentmeister A (2016) One-pot modification of 5′-capped RNA based on methionine analogs. Methods 107:3–9

Muttach F, Muthmann N, Reichert D, Anhäuser L, Rentmeister A (2017) A benzylic linker promotes methyltransferase catalyzed norbornene transfer for rapid bioorthogonal tetrazine ligation. Chem Sci 8: 7947–7953

Narmontė M, Gibas P, Daniūnaitė K, Gordevičius J, Kriukienė E (2021) Multiomics analysis of neuroblastoma cells reveals a diversity of malignant transformations. Front Cell Dev Biol 9

Neely RK, Dedecker P, Hotta J-i, Urbanavičiūtė G, Klimašauskas S, Hofkens J (2010) DNA fluorocode: A single molecule, optical map of DNA with nanometre resolution. Chem Sci 1(4):453–460

Neveling K, Mantere T, Vermeulen S, Oorsprong M, van Beek R, Kater-Baats E, Pauper M, van der Zande G, Smeets D, Weghuis DO, Stevens-Kroef MJPL, Hoischen A (2021) Next-generation cytogenetics: Comprehensive assessment of 52 hematological malignancy genomes by optical genome mapping. Am J Hum Genet 108:1423–1435

Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, Tahiliani M, Daley GQ, Liu XS, Ecker JR, Milos PM, Agarwal S, Rao A (2011) Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature 473(7347):394–397

Peters W, Willnow S, Duisken M, Kleine H, Macherey T, Duncan KE, Litchfield DW, Luscher B, Weinhold E (2010) Enzymatic site-specific functionalization of protein methyltransferase substrates with alkynes for click labeling. Angew Chem Int Ed Engl 49(30): 5170–5173

Pignot M, Siethoff C, Linscheid M, Weinhold E (1998) Coupling of a Nucleoside with DNA by a Methyltransferase. Angew Chem Int Ed Engl 37(20): 2888–2891

Pljevaljčić G, Pignot M, Weinhold E (2003) Design of a new fluorescent cofactor for DNA methyltransferases and sequence-specific labeling of DNA. J Am Chem Soc 125(12):3486–3492

Pljevaljčić G, Schmidt F, Weinhold E (2004) Sequence-specific methyltransferase-induced labeling of DNA (SMILing DNA). Chembiochem 5(3):265–269

Pljevaljčić G, Schmidt F, Scheidig AJ, Lurz R, Weinhold E (2007) Quantitative labeling of long plasmid DNA with nanometer precision. Chembiochem 8(13): 1516–1519

Plotnikova A, Osipenko A, Masevičius V, Vilkaitis G, Klimašauskas S (2014) Selective covalent labeling of miRNA and siRNA duplexes using HEN1 methyltransferase. J Am Chem Soc 136(39): 13550–13553

Ramadan M, Bremner-Hay NK, Carlson SA, Comstock LR (2014) Synthesis and evaluation of N6-substituted azide- and alkyne-bearing N-mustard analogs of S-adenosyl-l-methionine. Tetrahedron 70(34): 5291–5297

Schlenk F, Dainko JL (1975) The S-n-propyl analogue of S-adenosylmethionine. Biochim Biophys Acta 385(2): 312–323

Schmidt FH, Huben M, Gider B, Renault F, Teulade-Fichou MP, Weinhold E (2008) Sequence-specific methyltransferase-induced labelling (SMILing) of plasmid DNA for studying cell transfection. Bioorg Med Chem 16(1):40–48

Schulz D, Holstein JM, Rentmeister A (2013) A chemo-enzymatic approach for site-specific modification of the RNA cap. Angew Chem Int Ed Engl 52(30): 7874–7878

Serva S, Lagunavičius A (2015) Direct conjugation of peptides and 5-hydroxymethylcytosine in DNA. Bioconjug Chem 26(6):1008–1012

Sharim H, Grunwald A, Gabrieli T, Michaeli Y, Margalit S, Torchinsky D, Arielly R, Nifker G, Juhasz M, Gularek F, Almalvez M, Dufault B, Sen CS, Liu A, Bhattacharya S, Chen YW, Vilain E, Wagner KR, Pevsner J, Reifenberger J, Lam ET, Hastie AR, Cao H, Barseghyan H, Weinhold E, Ebenstein Y (2019) Long-read single-molecule maps of the functional methylome. Genome Res 29:646–656

Shu X, Cao J, Cheng M, Xiang S, Gao M, Li T, Ying X, Wang F, Yue Y, Lu Z, Dai Q, Cui X, Ma L, Wang Y, He C, Feng X, Liu J (2020) A metabolic labeling method detects m(6)A transcriptome-wide at single base resolution. Nat Chem Biol 16:887–895

Singh S, Zhang J, Huber TD, Sunkara M, Hurley K, Goff RD, Wang G, Zhang W, Liu C, Rohr J, Van Lanen SG, Morris AJ, Thorson JS (2014) Facile chemoenzymatic strategies for the synthesis and utilization of S-adenosyl-(L)-methionine analogues. Angew Chem Int Ed Engl 53(15):3965–3969

Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang B, Godley LA, Hicks LM, Lahn BT, Jin P, He C (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol 29(1): 68–72

Stankevičius V, Gibas P, Masiulionytė B, Gasiulė L, Masevičius V, Klimašauskas S, Vilkaitis G (2022) Selective chemical tracking of Dnmt1 catalytic activity in live cells. Mol Cell 82:1053–1065.e8

Staševskij Z, Gibas P, Gordevičius J, Kriukienė E, Klimašauskas S (2017) Tethered oligonucleotide-primed sequencing, TOP-Seq: a high-resolution economical approach for DNA epigenome profiling. Mol Cell 65:554–564.e6

Stecher H, Tengg M, Ueberbacher BJ, Remler P, Schwab H, Griengl H, Gruber-Khadjawi M (2009) Biocatalytic Friedel-Crafts alkylation using non-natural cofactors. Angew Chem Int Ed Engl 48(50):9546–9548

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324(5929):930–935

Tang Q, Pavlidis IV, Badenhorst CPS, Bornscheuer UT (2021a) From natural methylation to versatile alkylations using halide methyltransferases. ChemBioChem 22:2584–2590

Tang Q, Grathwol CW, Aslan-Üzel AS, Wu S, Link A, Pavlidis IV, Badenhorst CPS, Bornscheuer UT (2021b) Directed evolution of a halide methyltransferase enables biocatalytic synthesis of diverse SAM analogs. Angew Chem Int Ed 60:1524–1527

Teague B, Waterman MS, Goldstein S, Potamousis K, Zhou S, Reslewic S, Sarkar D, Valouev A, Churas C, Kidd JM, Kohn S, Runnheim R, Lamers C, Forrest D, Newton MA, Eichler EE, Kent-First M, Surti U, Livny M, Schwartz DC (2010) High-resolution human genome structure by single-molecule analysis. Proc Natl Acad Sci USA 107(24):10848–10853

Tomkuvienė M, Clouet-d'Orval B, Černiauskas I, Weinhold E, Klimašauskas S (2012) Programmable sequence-specific click-labeling of RNA using archaeal box C/D RNP methyltransferases. Nucleic Acids Res 40(14):6765–6773

Tomkuvienė M, Mickutė M, Vilkaitis G, Klimašauskas S (2019) Repurposing enzymatic transferase reactions for targeted labeling and analysis of DNA and RNA. Curr Opin Biotechnol 55:114–123

Tomkuvienė M, Ikasalaitė D, Slyvka A, Rukšėnaitė A, Ravichandran M, Jurkowski TP, Bochtler M, Klimašauskas S (2020) Enzymatic hydroxylation and excision of extended 5-methylcytosine analogues. J Mol Biol 432

Tomkuvienė M, Meier M, Ikasalaitė D, Wildenauer J, Kairys V, Klimašauskas S, Manelytė L (2022) Enhanced nucleosome assembly at CpG sites containing an extended 5-methylcytosine analogue. Nucleic Acids Res 50(11):6549–6561

Townsend AP, Roth S, Williams HEL, Stylianou E, Thomas NR (2009) New *S*-adenosyl-L-methionine analogues: synthesis and reactivity studies. Org Lett 11(14):2976–2979

Vranken C, Deen J, Dirix L, Stakenborg T, Dehaen W, Leen V, Hofkens J, Neely RK (2014) Super-resolution optical DNA mapping via DNA methyltransferase-directed click chemistry. Nucleic Acids Res 42(7):e50

Wand NO, Smith DA, Wilkinson AA, Rushton AE, Busby SJW, Styles IB, Neely RK (2019) DNA barcodes for rapid, whole genome, single-molecule analyses. Nucleic Acids Res 47:e68

Wang T, Kohli RM (2021) Discovery of an unnatural DNA modification derived from a natural secondary metabolite. Cell Chem Biol 28(1):97–104.e4

Wang R, Zheng W, Yu H, Deng H, Luo M (2011) Labeling substrates of protein arginine methyltransferase with engineered enzymes and matched S-adenosyl-L-methionine analogues. J Am Chem Soc 133(20): 7648–7651

Wang R, Islam K, Liu Y, Zheng W, Tang H, Lailler N, Blum G, Deng H, Luo M (2013) Profiling genome-wide chromatin methylation with engineered posttranslation apparatus within living cells. J Am Chem Soc 135(3):1048–1056

Wang P, Meyer TA, Pan V, Dutta PK, Ke Y (2017) The beauty and utility of DNA. Origami Chem 2:359–382

Weier HU, Wang M, Mullikin JC, Zhu Y, Cheng JF, Greulich KM, Bensimon A, Gray JW (1995) Quantitative DNA fiber mapping. Hum Mol Genet 4(10): 1903–1910

Weinhold E, Chakraborty B (2021) DNA modification and visualization on an origami-based enzyme nanofactory. Nanoscale 13:2465–2471

Weiss N, Seneviranthe C, Jiang M, Wang K, Luo M (2021) Profiling and validation of live-cell protein methylation with engineered enzymes and methionine analogues. Curr Protoc 1:e213

Weller RL, Rajski SR (2005) DNA methyltransferase-moderated click chemistry. Org Lett 7(11):2141–2144

Weller RL, Rajski SR (2006) Design, synthesis, and preliminary biological evaluation of a DNA methyltransferase-directed alkylating agent. Chembiochem 7(2):243–245

Wight DJ, Aimola G, Aswad A, Lai CYJ, Bahamon C, Hong K, Hill JA, Kaufer BB (2020) Unbiased optical mapping of telomere-integrated endogenous human herpesvirus 6. Proc Natl Acad Sci USA 117:31410–31416

Wilkinson S, Diechtierow M, Estabrook RA, Schmidt F, Hüben M, Weinhold E, Reich NO (2008) Molecular scale architecture: engineered three- and four-way junctions. Bioconjug Chem 19:470–475

Wilkinson AA, Jagu E, Ubych K, Coulthard S, Rushton AE, Kennefick J, Su Q, Neely RK, Fernandez-Trillo P (2020) Site-selective and rewritable labeling of DNA

through enzymatic, reversible, and click chemistries. ACS Cent Sci 6:525–534

Willnow S, Martin M, Luscher B, Weinhold E (2012) A selenium-based click AdoMet analogue for versatile substrate labeling with wild-type protein methyltransferases. Chembiochem 13(8):1167–1173

Winter JM, Chiou G, Bothwell IR, Xu W, Garg NK, Luo M, Tang Y (2013) Expanding the structural diversity of polyketides by exploring the cofactor tolerance of an inline methyltransferase domain. Org Lett 15(14):3774–3777

Yuan Y, Chung CYL, Chan TF (2020) Advances in optical mapping for genomic research. Comput Struct Biotechnol J 18:2051–2062

Zhang C, Weller RL, Thorson JS, Rajski SR (2006) Natural product diversification using a non-natural cofactor analogue of S-adenosyl-L-methionine. J Am Chem Soc 128(9):2760–2761

Zhang L, Szulwach KE, Hon GC, Song CX, Park B, Yu M, Lu X, Dai Q, Wang X, Street CR, Tan H, Min JH, Ren B, Jin P, He C (2013) Tet-mediated covalent labelling of 5-methylcytosine for its genome-wide detection and sequencing. Nat Commun 4:1517

Zohar H, Muller SJ (2011) Labeling DNA for single-molecule experiments: methods of labeling internal specific sequences on double-stranded DNA. Nanoscale 3(8):3027–3039