Deciphering bacterial epigenomes using modern sequencing technologies

John Beaulaurier, **Eric E. Schadt**, and **Gang Fang**[*]

Department of Genetics and Genomic Sciences and Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Abstract

Although 5-methylcytosine (5mC) is the most widely studied form of DNA methylation in eukaryotes, together, N6-methyladenine (6mA), N4-methylcytosine (4mC) and 5mC form the prokaryotic methylome. Recent advances in DNA sequencing technology have enabled systematic detection of DNA methylations in bacteria at genome-wide scale. In the past six years, >1,900 bacterial methylomes have been mapped, which has led to the discovery of several novel insights into the complexity and functions of bacterial DNA methylation. Increasing evidence suggests that epigenetic regulation of gene expression, virulence and pathogen–host interaction are prevalent in bacteria. We review the currently available technologies for studying bacterial methylomes and the novel biological insights they have provided. We also offer perspectives on the unprecedented opportunities and challenges for achieving a more complete understanding of bacterial epigenomes.

Introduction

Technological advances during the past several decades have led to dramatic decreases in the per-base cost of DNA sequencing[1], allowing researchers to access the genomic content from a wide variety of organisms. In this 'age of genomics', the majority of research has focused on decoding the sequences of the four canonical DNA bases: adenine, cytosine, guanine, and thymine. There is, however, considerable epigenetic information encoded in covalent modifications to these canonical bases. These modified bases effectively expand the DNA alphabet beyond four nucleotides[2].

[*] gang.fang@mssm.edu.

Related links
NEB: link
REBASE: link
PacBio white book: link
SMALR: link
mBin: link
SMRTER: link
nanopolish: link
signalAlign: link
mCaller: link
nanoraw: link

Despite DNA methylation being discovered in bacteria more than a half century ago[3], the most widely studied form of DNA methylation today is still 5-methylcytosine (5mC; Fig. 1a) in eukaryotes. 5mC is formed when a methyltransferase (MTase) transfers a methyl group from S-adenosyl-L-methionine (SAM) to the C5 of an unmodified cytosine[4]. This type of DNA methylation has been found to be essential for a variety of critical processes during growth and development in eukaryotes, including gene expression, genome maintenance, and parental imprinting[5,6]. 5mC has also been widely implicated in the etiology of various diseases, including fragile X syndrome[7], immunodeficiency[8], and cancer[9], among many others[10]. Furthermore, oxidation of 5mC bases by the ten-eleven translocation (TET) class of enzymes can lead to additional diversification of the epigenetic alphabet[11]. These oxidative derivatives of 5mC, including 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), might exist both as demethylation intermediates and as distinct epigenetic regulators, although their broader and deeper functional significance continues to be actively studied[2,12].

In the bacterial world, 5mC is not the dominant DNA modification type[13]. Rather, it exists in the genome alongside N6-methyladenine (6mA; Fig. 1b), the most prevalent form, and N4-methylcytosine (4mC; Fig. 1c), an alternative form of cytosine methylation. Recent advances in the field of methylation detection have only begun to shed light on the extent and function of these two forms of DNA modification. While 4mC and 6mA are structurally distinct from 5mC, all three forms of DNA methylation (5mC, 4mC, and 6mA) are sequence-specific in bacteria. In most bacteria, a small set of sequence motifs (three on average in each genome) are targeted by MTases for methylation (e.g. 5'-G**A**TC-3' by Dam or 5'-**C**CWGG-3' by Dcm in *Escherichia coli)* at nearly all their occurrences in the genome, with only a small fraction of these motif sites remaining non-methylated[13].

Bacterial DNA methylation has primarily been studied in the context of restriction-modification (RM) systems whose primary function is to protect cells from invading DNA by distinguishing the endogenous, methylated DNA from the foreign, non-methylated DNA[13,14]. In these systems (detailed in **Box 1**), a MTase is encoded in the vicinity of a cognate restriction enzyme (RE) that restricts any DNA it encounters lacking a protective methylation at its target sequence. Other so-called orphan MTases, such as Dam in gammaproteobacteria, lack a cognate RE and are thought to serve as regulators of DNA replication and gene expression among other functions[13]. DNA methylation, by both orphan MTases and those belonging to RM systems, has been found to play important regulatory roles in bacteria[13–23]. Beyond basic regulatory mechanisms, there is also emerging evidence that heterogeneously methylated bacterial populations can drive heterogeneity in gene expression and cellular phenotypes, thereby serving as units of adaptive selection beyond simple genetic variation[24,25]. This heterogeneity of methylation can stem from either local interactions at target motifs between MTases and DNA binding proteins or at a more global level involving phase-variable MTases[26,27]. An impressive diversity of MTases has been uncovered in recent years, with evidence suggesting that DNA methylation is present in the vast majority of the >6,000 sequenced bacterial genomes[28,29]. Given that the precise sequence targets and biological roles of most MTases remain largely unknown, the potential scope for gaining new and important biological insights is vast[30].

A major limitation that has hampered the study of bacterial DNA methylomes is the lack of high-throughput tools for detecting the most common methylation types in bacteria. Bisulfite sequencing is commonly used to achieve nucleotide resolution and genome-wide scale for 5mC detection in eukaryotes[31–35], but cannot effectively resolve the other major methylation types in bacteria (4mC and 6mA). Until very recently, most researchers relied on methyl-sensitive or methyl-dependent restriction digests to probe the methylation status of bacterial DNA[36,37]. This approach, however, depends on a finite, and fairly small, set of well-characterized restriction enzymes with known target specificities, making them not well suited for exploring the massive diversity of bacterial methylation target sequences.

In the past six years, efforts to map over 1,900 methylomes of a diverse collection of bacterial and archaeal species have led to the discovery of several novel insights into the complexity and functions of bacterial DNA methylation. This methylome information is centralized in the REBASE database[28] (http://rebase.neb.com), which has historically served as a repository of information about bacterial RM systems. The methylomes profiled to date belong to a wide variety of isolates from over 750 distinct species. Unsurprisingly, isolates of common human pathogens such as *Salmonella enterica* (n=150), *Escherichia coli* (n=123), *Klebsiella pneumoniae* (n=93) and *Staphylococcus aureus* (n=47) comprise a significant fraction of the total mapped methylomes. Much of the progress in building this extensive repository of methylomes is due to the introduction of new technologies for high-throughput detection of bacterial DNA methylation. Specifically, these include single-molecule, real-time (SMRT) sequencing, which enabled the first simultaneous detection of the three major forms of DNA methylation and contributed to most of the currently mapped bacterial methylomes[16,29,38–43], and Oxford Nanopore sequencing[44], which also holds potential to detect different forms of DNA methylation. Using these new technologies, compelling questions are now being asked and bacterial methylomes are being investigated at a level of detail that was impossible only a decade ago (Fig. 1d).

The many recent technological advances and their applications in numerous contemporary studies make it a good time to review the methods for surveying bacterial methylomes and biological insights they have produced. Because the history and foundations of bacterial epigenetics have previously been thoroughly reviewed[13,14,30,45,46], this article will focus on the current landscape of cutting-edge technologies and the insights into bacterial epigenomes that these technologies have afforded. In addition, we will provide a perspective on the unprecedented opportunities and challenges for achieving a more complete understanding of bacterial epigenomes and the complex roles they play in defining their interactions with host organisms.

Modern technologies for mapping bacterial methylomes

The bulk of methodological development for DNA methylation detection has been devoted to characterizing 5mC in higher eukaryotes, largely because the biological significance of 5mC in mammalian cells has been recognized for over half a century[11,47,48]. This effort has led to the development of a variety of approaches that rely on digestion by 5mC-sensitive restriction enzymes, affinity enrichment of methylated DNA fragments, or chemical conversion of 5mC bases using sodium bisulfite or certain enzymes[2]. Such methods for 5mC

detection in eukaryotes have been thoroughly reviewed elsewhere[2,49–51]. However, owing to the different forms of methylation they contain, characterization of bacterial methylomes requires alternative detection methods beyond these legacy methods.

Recent advances in sequencing technologies have now made it possible to obtain sequences and many associated base modifications directly from individual DNA molecules. So-called "third-generation" sequencing technologies, including SMRT and nanopore sequencing, present the most significant opportunities yet for comprehensively characterizing bacterial methylomes. A fully characterized methylome contains not just the full set of methylated positions and targeted motifs, but also a complete mapping of the MTases and RM systems responsible for each methylated motif (**Box 2**). The notable features of the various methodologies available for detection of DNA methylation in bacteria are summarized in **Table 1.**

Legacy methods

While method development for detecting eukaryotic 5mC has flourished over the past few decades, this same period has seen only modest development of new approaches for detecting the principle forms of DNA methylation in bacterial genomes. Because prokaryotic MTases are known to primarily target specific sequence motifs for methylation, researchers have historically digested genomic DNA with one or more methyl-sensitive restriction enzymes of known specificities[52]. Analysis of the resulting restriction sites sometime makes it possible to assess methylation status and deduce the methylated motif[37]. However, restriction enzymes-based studies are limited to methylation motifs that perfectly or partially match the known specificities of available restriction enzymes, making the approach not generally suited for *de novo* discovery of novel methylation motifs. An alternative approach uses modified traces in dye-terminator Sanger sequencing to identify methylated bases. The presence of 4mC, 5mC, and 6mA in the DNA template affects the amplitude of peaks in the sequencing trace, theoretically enabling the detection of the most common forms of bacterial DNA methylation as a byproduct of Sanger sequencing. While several studies used this method to investigate the methylomes of pathogenic bacteria[53–57], technical limitations, including subtle peak signatures and the low throughput of Sanger sequencing, have prevented it from achieving wider usage[58]. Perhaps owing to the fact that 5mC seems to play a more minor role in bacterial genomes, bisulfite sequencing was only recently applied to the study of 5mC in bacteria[59,60]. An additional treatment with ten-eleven translocation (TET) enzymes makes it possible to characterize both 5mC and 4mC bacterial methylomes using bisulfite sequencing, although 6mA is left uncharacterized with this approach[61].

Direct detection using single-molecule, real-time sequencing

SMRT sequencing, available in the commercialized RS II and Sequel instruments manufactured by Pacific Biosciences Inc., is the first third-generation sequencing technology with a record of successfully characterizing bacterial methylomes. SMRT sequencing can simultaneously report both nucleotide sequence and all three major types of DNA methylation in bacteria (4mC, 5mC, and 6mA), albeit at different sensitivities due to the

signal-to-noise ratios specific to each modification type (6mA: high confidence; 4mC: moderate; 5mC: low)[16,25,29,39,40,62].

In SMRT sequencing, each molecule consists of a double-stranded native DNA fragment that has been circularized by ligating hairpin adapters to each end[38] (Fig. 2a). Real-time observation of sequencing by synthesis occurs in a zeptoliter-scale observation chamber called a zero-mode waveguide (ZMW) that limits background fluorescence originating from outside of this small observation chamber[38,63] (Fig. 2b). Immobilized DNA polymerase enzymes are bound to adapter-bound template DNA molecules in the ZMWs and DNA synthesis is initiated. The DNA polymerase proceeds around the circularized DNA template multiple times, generating a number of subreads. During each base incorporation event, one of four fluorescently labeled deoxyribonucleoside triphosphates (dNTPs), one color for each of the canonical four bases, is briefly immobilized in the ZMW observation window by the polymerase. A camera captures the resulting fluorescent pulse and the series of observed pulses are used to construct the read sequence. While the base calling is accomplished by monitoring the order of dNTP incorporation events, DNA modifications are detected by identifying changes in the kinetics of the polymerase (Fig. 2c). Specifically, in SMRT sequencing, the time interval between the pulses that signal incorporation events is referred to as the inter-pulse duration (IPD) and these values describe the polymerase kinetics as it translocates along the DNA template.

After observing that primary and secondary structure of the DNA template molecules could perturb IPD values through interactions with the sequencing polymerase[38], it was also shown that changes in IPD values could reveal the presence of covalent DNA modifications, including 4mC, 6mA, 5mC, 5hmC, and other types of DNA methylation and damage[39,62,64,65] (Box 3 describes technical considerations of methylation detection using SMRT sequencing). Among the epigenetic marks most relevant in bacterial methylomes, 4mC and 6mA can be directly detected in native DNA given their relatively high signal-to-noise ratio[16,25,62]. 5mC and 5hmC, however, require either high sequencing coverage or separate conversion steps to 5-formylcytosine and 5-carboxylcytosine, both of which have enhanced signal-to-noise ratioes in SMRT-based sequencing assays[62,65].

Direct detection using nanopore sequencing

Nanopore sequencing has been under active development for decades, but recent progress has led to the release of the first commercially available sequencing platform by Oxford Nanopore Technologies (ONT)[44,66–72]. The underlying technology leverages variations in ionic current induced by the passage of different nucleotides through genetically engineered protein nanopores. Although the vast majority of research applications to date have focused on using MinION to call the four canonical bases, ionic current has been shown to differ between canonical bases and covalently modified nucleotides, enabling the technology to detect chemical DNA modifications[72].

Library construction for nanopore sequencing, for which multiple protocols are available, involves the ligation of adapter sequences and the addition of a motor protein to double-stranded DNA fragments (Fig. 2d). The adapter sequences help to concentrate the DNA fragments near the nanopore-containing lipid membrane, while the motor protein facilitates

the processive ratcheting of ssDNA through the protein nanopore at a fixed rate during sequencing. Sensors monitor each nanopore during this process and detect the variations in ionic current through the nanopore caused by the set of nucleotides obstructing the channel (Fig. 2e). The picoamp current fluctuations are a function of the precise 4–6mer occupying the nanopore channel at a given moment (Fig. 2f) and are processed by a recursive neural network to construct the sequence of nucleotides in the read.

Importantly, chemical modification to bases in the native library can induce current signals on top of canonical bases[72]. While this provides an opportunity to detect DNA modifications, it can potentially complicate the base calling process, which relies on the characteristic current levels produced as each k-mer (a DNA "word" comprised of combinations of nucleotides of length k) passes through the nanopore. The presence of multiple types of base modifications greatly expands the set of possible k-mers beyond those constructed exclusively from the four canonical bases, which introduces significant computational challenges. Early attempts to detect methylation during nanopore sequencing, using a variety of protein nanopore configurations and experimental conditions, focused on eukaryotic applications and therefore were limited to 5mC and 5hmC detection[66,67,72–74]. However, the introduction of the MinION device has recently broadened the development focus to also include methods capable of characterizing prokaryotic methylomes[75–77]. While several recent studies have demonstrated the feasibility of nanopore-based methylation detection, some challenges remain in this relatively early-stage yet very active field of research.

Rand *et al.* developed a variable order hidden Markov model (HMM) trained to identify methylation events in bacterial genomes[75]. By pairing the HMM with a hierarchical Dirichlet process (HDP) to learn current distributions from the MinION, the method can detect both 5mC and 6mA at the specific motifs included in the training data. This approach can detect these modifications in individual reads, albeit at significantly lower sensitivities than when the HMM- HDP is applied to consensus current signals from multiple aligned reads. The model, however, is constrained by the contents of its training data, limiting its ability to identify novel modification types or methylated motifs. By implementing a neural network classifier trained on an expanded sequence context spanning known 6mA positions, McIntyre *et al.* showed increased sensitivity in their read-level 6mA detection[76]. While encouraging, such a model-based approach remains limited in its ability to *de novo* identify diverse modification types at various sequence motifs. More recently, a preprint from Stoiber *et al.* described nanopore-based methylation detection using a statistical comparison of ionic current signals from native and methylation-free whole genome amplified (WGA) DNA[77]. Following the design first proposed by Flusberg *et al.* for SMRT sequencing, this method has the advantage of not requiring any *a priori* knowledge of modification types or their distinct current signatures[39]. Stoiber *et al.* were able to correctly identify several expected 4mC, 5mC, and 6mA motifs in bacterial genomes carrying MTases of known specificity, although their detection accuracy fluctuates with different modification types and motif specificities. Although encouraging, detection is not possible at the level of single molecules and methods like this one that do not require any a priori knowledge may not be able to distinguish among diverse forms of DNA methylation and DNA damage events, especially in eukaryotic genomes[78].

To date, none of these nanopore sequencing based methods have been applied for the biological characterization of an unknown bacterial methylome. But the rapid pace of methods development in this field and the ongoing technological advancements in the underlying sequencing technology make nanopore sequencing an interesting and dynamic field to watch for methylome researchers going forward.

Methylation types and their motif specificities

The recent advances in methylation detection technologies have precipitated a surge of studies devoted to the characterization and functional examination of bacterial methylomes. These studies have built upon decades of previous work, most of which has relied on experimental approaches focused on a handful of loci in a relatively small number of well-characterized genomes[14,22,37,79–81]. The hard-won insights produced by these foundational studies have long hinted at an unappreciated level of complexity and regulatory potential present in modifications to the four canonical bases. The application of modern methylation detection technologies is shedding new light on this epigenetic realm by delivering what was impossible until very recently: genome-wide mapping of the three primary forms of DNA methylation in bacteria.

N4-methyl cytosine.

The extent of 4mC in bacterial genomes is not well known and its function remains largely a mystery. While occurring less frequently than 6mA in bacteria, this form of methylation has been observed more often in thermophilic bacteria, potentially due to its substantially higher resistance (compared to 5mC) to heat-induced deamination to thymine[82,83]. Digestion of genomic DNA with methyl-sensitive restriction enzymes has recently proved useful in conclusively identifying 4mC methylation in bacteria[84]. However, this approach is limited by the availability of such restriction enzymes. Two recently developed methods for 4mC detection make use of bisulfite sequencing protocols that were modified to extend detection capabilities beyond just 5mC positions in the genome. The first approach, applied by Huo *et al.* to a strain of *Enterococcus faecalis,* relies on a previous observation that 5mC sites are fully protected from bisulfite conversion to uracil, while 4mC sites are only partially protected[85,86]. The second approach, which requires both TET enzymes and bisulfite treatment in a two-step conversion process, has been used to identify the methylated cytosine motifs in the hyperthermophilic species *Caldicellulosiruptor kristjanssonii*[61].

SMRT sequencing is currently the most broadly applied method for 4mC detection. A variety of 4mC motif specificities have been identified in species such as *Bacillus* cereus[40], *Helicobacter pylori*[41,42,87], *Campylobacter jejuni*[88], and *Salmonella enterica*[89]. Aside from its known involvement in multiple RM systems and a single study that suggests its link to gene expression in *H. pylori*[90], the biological functions of 4mC remain largely unclear.

5-methyl cytosine.

Although the orphan cytosine MTase Dcm has been the subject of study for several decades and its target specificity of 5'-C**C**WGG-3' has long been known[91], insights into the biological role of 5mC in bacteria has remained somewhat elusive. Beyond its known role in

protecting bacteria against parasitism by the EcoRII RM system[92], methylation by Dcm has been associated with Tn3 transposition[93], lambda phage recombination[94], and the expression of ribosomal proteins during stationary phase[95].

Two recent studies have taken advantage of the genome-wide and single-nucleotide resolution of bisulfite sequencing to conduct thorough investigations of 5mC functions in Gammaproteobacteria. In the first study, Kahramanoglou *et al.* linked Dcm methylation of 5'-C**C**WGG-3' in *E. coli* to the expression of RNA polymerase sigma factor *rpoS* and many of its target genes in stationary phase[59]. Subsequent work in *Vibrio cholerae* revealed that methylation of 5'-R**C**CGGY-3' by the cytosine MTase VchM is required for optimal growth and affects the cell envelope stress response, potentially by downregulating genes required for modifying the lipopolysaccharide inner core of the cell envelope[60]. In both studies, however, direct regulation of transcription by the 5mC methylation was not proven.

SMRT sequencing has also yielded insights into bacterial 5mC methylomes, revealing the 5mC motif specificities of active cytosine MTases in a variety of species and strains[41,96–99]. Identification of these 5mC methylated motifs has revealed a higher frequency of point mutations at 5mC bases[98] and has facilitated the design of plasmids capable of overcoming barriers to transformation in an important strain of *Bifidobacterium animalis*[99]. By enabling the design of efficient shuttle vectors, this latter finding provides an important new tool to researchers examining the precise molecular mechanisms underlying the observed correlations between bifidobacteria and gut health[100].

N6-methyl adenine.

While some work in recent years has applied more traditional digestion-based approaches using methyl-sensitive or methyl-dependent restriction enzymes[101], the majority of modern 6mA studies have used SMRT sequencing to identify genome-wide 6mA events in bacteria. The abundance of 6mA MTases in the bacterial world and the robust IPD signature generated by 6mA during SMRT sequencing have led to the discovery of a vast diversity of 6mA MTases and methylated motifs in bacteria. These include many previously unknown orphan MTases and a multitude of previously uncharacterized type I, II, and III RM systems[28,29].

SMRT sequencing has elucidated the 6mA methylated motifs in a wide variety of organisms across multiple phyla, including Bacteroidetes *(Bacteroides dorei*[102]), Firmicutes *(Enterococcus faecalis*[85], *Listeria monocytogenes*[103], *Streptococcus pneumoniae*[104,105]), Actinobacteria *(Bifidobacterium animalis*[99], *Mycobacterium tuberculosis*[43,106]), and Proteobacteria *(Bibersteinia trehalosi*[97], *Campylobacter jejuni*[88,107,108], *Caulobacter crescentus*[96], *Chromohalobacter salexigens*[40], *Escherichia coli*[16,109], *Geobacter metallireducens*[40], *Haemophilus infuenzae*[110], *Helicobacter pylori*[41,42,87], *Moraxella catarrhalis*[111], *Neisseria meningitidis*[98,112], *Salmonella enterica*[89], *Shewanella oneidensis*[113], *Vibrio breoganii*[40]).

Functional knockout studies in many of these organism highlight the ability of certain 6mA MTases to induce widespread transcriptional changes[16,25,43,104,114,115], while other work

has revealed differentially methylated 6mA positions in response to varied growth stages and environmental conditions[43,96,113].

Researchers have also taken advantage of modern methylation detection techniques to explore mechanisms of bacteriophage invasion and host defense, revealing the presence of multiple 6mA MTases encoded by the 936-type bacteriophages that commonly infect *Lactococcus lactis* starters used in cheese production[116]. These MTases likely provide the bacteriophage with protective methylation, allowing it to circumvent host RM systems. On the other side of this pitched microbial battle, Goldfarb *et al.* describe a gene cassette, termed the bacteriophage exclusion (BREX) system, conferring bacteriophage resistance in a wide range of host bacteria. Interestingly, although activity of a 6mA MTase in the cassette is required for successful host defense, phage DNA does not appear to be targeted for restriction, suggesting a novel mechanism of methylation-based host defense[117].

Diversity of MTases and target specificities

DNA methylation events in prokaryotic genomes are highly motif driven for all three of the primary methylation types in bacteria. If a methylation motif is targeted by an MTase, typically >95% of occurrences of the motif sites are methylated[13,25,29,118]. The accumulating modern methylome surveys have contributed to the rapidly growing catalog of known bacterial RM systems documented in the REBASE database[28,41,42,87–89]. RM systems often represent a significant obstacle to genetic manipulation of an organism through transformation, leading to low transformation efficiencies. The design of effective shuttle vectors must therefore either include compatible methylation pattern to provide protective methylation or else limit the number of motif sites in the vector that are subject to restriction by the host RM system[85,99]. Both of these approaches require a thorough understanding of the host RM repertoire and benefit from a comprehensive catalog of known RM systems and specificities.

Historically, novel type II RM systems have been identified through restriction digest approaches, as restriction in these systems occurs at precisely the same motif that is methylated. However, the restriction site in type I and III RM systems cannot serve as a proxy for the site of methylation; restriction in these type I and III RM systems occurs at a variable distance from the site of methylation[119]. As a result, there was until recently a notable paucity of known type I and III RM systems contained in REBASE. Fortunately, the introduction of methylation detection by SMRT sequencing in 2012 has dramatically increased the number of known type I and III RM systems in recent years (Supplementary Figure 1).

Perhaps the most surprising observation made in recent years by the multitude of prokaryotic methylome studies is the remarkable diversity of MTase genes and target specificities. A recent survey of 230 diverse bacterial and archaeal epigenomes, enabled by SMRT sequencing, found DNA methylation in 93% of genomes across a wide diversity of methylated motifs (834 distinct motifs; averaging three motifs per organism)[29]. The primary driver behind this diversity is the spread of MTase-containing mobile genetic elements through horizontal gene transfer (HGT)[29,120,121]. Mutation events can also occur in the target recognition domain (TRD) of MTase genes and thereby modify the sequence motif

targeted for methylation, providing a route to further methylome diversification[42]. As a consequence of such diversification, researchers commonly find significantly diverging methylomes among not only species, but even different strains of the same species[41,89,97,98,102,103,106,107].

Insights into epigenetic regulation

Understanding of epigenetic regulation in bacteria is largely pursued along two closely related dimensions: methylation as a cellular regulatory signal and cell-to-cell heterogeneity. The advancement of sequencing technologies has played an important role in multiple studies that revealed novel biological insights into epigenetic regulation in bacteria.

Methylation as a cellular regulatory signal

Several MTases have been shown to be capable of inducing dramatic shifts in global gene expression[16,59,60,114,122]. The epigenetic toggling of the pap pili and *agn43* expression via methylation status at 5'-G**A**TC-3' sites in *E. coli* serve as the canonical examples of the local competition model, where competitive binding between Dam and other DNA-binding proteins (e.g. transcription factors) at specific motif sites affects transcription of a nearby gene, leading to phenotypic variation[18,123–125] (Fig. 3a & b). In the case of the pap pili, the binding competition at the 5'-G**A**TC-3' in the *pap* promoter is partially skewed by sequence contexts that hinder the local processivity of Dam, providing more time for the DNA binding proteins to access their target sites[126]. Several similar mechanisms of transcriptional regulation that rely on the competition between Dam and various DNA binding proteins have also been described using legacy methylation detection approaches[13,14,127–130]. In such a model, the methylating action of the MTase serves as the initiator of phase variation by activating or repressing the transcription of the downstream gene in some fraction of the cells in the population. This is not to be confused with the phase variable MTases (to be described shortly), where the expression of a MTase itself is subject to reversible toggling.

Apart from its roles in local transcriptional switches involving competitive binding, such as *agn43* and *pap,* DNA methylation also exerts critical regulatory signals through other means. For instance, both *E. coli* and *C. crescentus* cannot initiate replication without methylation of specific motif sites (5'-G**A**TC-3' and 5'-G**A**NTC-3', respectively) within their replication origin[18]. In addition, transient hemimethylation in *E. coli* following passage of the replication fork permits the transcription of the transposase gene of IS10. Methylation by Dam of a hemimethylated 5'-G**A**TC-3' site in the promoter quickly represses that transcription, presumably to limit potential transposition to the moment where a cell contains more than a single copy of the chromosome[22].

Systematic mapping of non-methylated motif sites in bacterial genomes only became feasible with the introduction of SMRT sequencing. Specifically, several hundreds of non-methylated motif sites have been reported across various bacteria, suggesting that the competition between MTases and DNA binding proteins is prevalent in bacteria[43,96,106,109,114]. Because non-methylated motif sites tend to occur in regulatory regions[13,14,29], they also suggest the prevalence of epigenetic switches to regulate bacterial gene expression. While detailed mechanisms in most cases remain to be identified, Cota *et*

al. recently SMRT sequencing to show that site-specific patterns of 5-G**A**TC-3' methylation and non-methylation by Dam in the regulatory region of the *opvAB* operon of *Salmonella enterica* are responsible for determining the O-antigen chain length[131].

Clues to the biological function of the MTase can occasionally be found by identifying genomic regions that are enriched for the methylation target motifs. For instance, enrichment of the Dam 5'-G**A**TC-3' target motif near the origin of replication in *E. coli* and other Gammaproteobacteria has been well documented and linked to roles in the initiation of replication[81,113]. The application of SMRT sequencing has led to the observation of additional examples of 6mA motif site enrichment near the origins of replication in *Arthrobacter* and *Nocardia,* indicating that this phenomenon may not be limited to Gammaproteobacteria[29]. Outside of the origin of replication, regions of over- and under-enriched motif sites have been identified by SMRT in a wide variety of bacteria[16,29,43,60,96], which provide important clues towards further understanding the biological purpose of such local enrichments.

Phase variation and epigenetic heterogeneity

Although phase variation of bacterial surface proteins, caused by reversible mutations at a hypervariable locus[132–134], has long been recognized as a mediator of antigenic variation and immune evasion, the significance and extent of phase-variable MTase took longer to emerge. The expression of these MTases is subject to hypervariable mutations that induce on/off or target specificity switching, resulting in heterogeneous methylating activity of the enzyme. Consequently, heterogeneous methylation patterns can develop within a clonally expanded population, often with dramatic and genome-wide regulatory consequences[27,119,135–137] (Fig. 3c).

Example of phase-variable MTases were first observed almost two decades ago due to observed hypervariable inversion events in the type I RM system-encoding *hsd* genes of *Mycoplasma pulmonis*[138] and *S. pneumoniae*[139]. Further examples were subsequently uncovered in *Pasteurella haemolytica*[140], *Moraxella catarrhalis*[141], *Haemophilus influenza*[27,142–144], *Helicobacter pylori*[145–147], and *Neisseria meningitidis*[144,148–150]. When Srikhanta *et al.* demonstrated that the variable methylation states could affect the expression of multiple genes throughout the genome, termed a phase-variable regulon (or *phasevarion),* the biological significance of this mechanism was quickly appreciated[27,119].

As the aforementioned studies lacked modern methods for comprehensive methylome mapping, the phase-variable behavior of MTases could only be inferred indirectly based on laborious experimentation and various observations, including differential restriction digest results in closely related strains or the presence of variable length tandem repeat tracts upstream of MTase genes. However, without details about the MTases activity (i.e. individual methylated positions and targeted motifs), the precise mechanisms by which variable MTase activity affects gene expression remained unknown.

SMRT sequencing has been used to broaden our understanding of previously identified phase-variable MTases. Seib *et al.* identified the methylated motif sites for three common alleles of the phase-variable ModA and ModD MTases in *N. meningitidis*[112,150]. Blakeway

et al. used SMRT sequencing to identify the methylated motifs of the most prevalent allele of ModM (ModM2) in the human respiratory pathogen *M. catarrhalis*[111]. Building on previous studies of the phase-variable ModA MTase in *H. influenzae*[27,142–144], Atack *et al.* used SMRT sequencing to identify the target motifs of a set of ModA alleles commonly found in otitis media isolates[110]. Two further studies focused on characterization of phase-variable MTases in *H. pylori* and how they contribute to the highly complex methylome of that organism. In addition to multiple phase variable MTases driven by slipped-strand mispairing in homopolymer tracts, both Krebes *et al.* and Furuta *et al.* used SMRT sequencing to identify an unusual type I MTase that achieves multiple bipartite motif specificities through interaction with several TRD elements, a process that can generate methylome diversification through recombination within the S subunit[41,42]. Srikhanta et al. had previously demonstrated that phase-variable MTases in *H. pylori* are capable of regulating phasevarions[147] and recently used SMRT sequencing to demonstrate the importance of the ModH5 allele of the phase-variable MTase ModH in regulating virulence genes in *H.* pylori[151]. Manso *et al.* surveyed the methylation landscape in S. *pneumoniae* and found that previously observed phase variation of the type I *hsd* system[139] is capable of inducing dramatic consequences in the bacterial methylome. Rearrangements in the configuration of five TRDs in the S subunit lead to six possible alleles, each with its own target specificity, representing one of the most complex form of phasevarion characterized to date[104].

While the above findings helped deepen our understanding of previously identified phase-variable MTases, other studies have taken advantage of the hypothesis-free nature of methylome analysis via SMRT sequencing to uncover novel phase-variable MTases in other pathogenic bacteria. For example, SMRT sequencing led to the recent discovery of MTase phase variation in the human gastric pathogen *C. jejuni*[108] and the bovine respiratory pathogen *B. trehalosi*[97]. This phase variation in *C. jejuni* was shown to affect cell adherence, invasion, and biofilm formation, while additional study is required to determine the functional consequences of MTase phase variation in *B. trehalosi.* In addition to phase variation, a software package named SMALR was developed to enable single molecule level analysis of methylation status using SMRT sequencing and it revealed another type of epigenetic heterogeneity in the marine bacterium *Chromohalobacter salexigens*[25], wherein methylation is dispersed across some, but not all, instances of a target motif. The biological reason for this observed pattern of incomplete methylation is unknown.

There is now a wealth of evidence, much of it derived from recent technological advances, implicating MTase phase variation as a crucial survival mechanism for host-adapted bacteria. Variability in methylation patterns, gene expression phenotypes has been observed, but future work will be required to clarify the precise mechanisms through which methylation regulate gene expression.

Epigenetic regulation of clinically important phenotypes

Among the many molecular and cellular phenotypes regulated by DNA methylation, clinically important phenotypes are of particular interest. Previous studies using legacy methods hinted at the clinical relevance of bacterial methylation, finding that methylation by

Dam (the MTase methylating 5'-G**A**TC-3') in *Salmonella typhimurium* is essential for virulence[152,153]. More recently, additional clinically important phenotypes have been linked to bacterial DNA methylation, many of which used SMRT sequencing to precisely associate specific methylation motifs targeted by phase-variable MTases with particular phenotypes.

Seib *et al.* linked two alleles of the phase-variable MTases ModA, ModA11 and ModA12, in *N. meningitidis* to sensitivity to several antibiotics typically prescribed for meningococcal disease, while also linking the phase-variable ModD MTase to hypervirulent strains of the pathogen[149,150]. Blakeway *et al.* found that the phase-variable MTase ModM in *M. catarrhalis* has potential roles in colonization, infection, and immune evasion[111]. Furthermore, their observation of the enrichment for ModM3 over the ModM2 allele in middle ear isolates points to a potentially significant role for ModM methylation in the colonization and infection of *M. catarrhalis* in human hosts. Atack *et al.* studied ModA alleles *of H. influenza* and observed a selection for specific ModA alleles was observed *in vivo* during progression of otitis media in chinchillas, suggesting a role for DNA methylation in *H. influenzae* colonization and infection[110]. Additionally, experiments using locked variants of these phase-variable ModA alleles demonstrated regulation of a variety of clinically important pathways, such as immune evasion, biofilm formation, antibiotic susceptibility, virulence, and niche adaptation. These results corroborate orthogonal studies by Brockman *et al.* and VanWagnoner *et al.* supporting ModA phase variation as an important regulator of virulence and immune evasion[154–156]. For S. *pneumoniae*, Manso *et al.* found the six-phase MTase have different virulence phenotypes and are selected at various stages of colonization and infection[104].

Collectively, these studies have important implications that many other bacterial pathogens may also exploit epigenetic switches as a flexible mechanism to regulate gene expression during host colonization and infection. Some of these mechanisms may serve as targets of potential therapeutic intervention strategies.

Towards deeper mechanistic insights

A very first step to study the functional impacts of bacterial DNA methylation is to compare global gene expression between wildtype and MTase mutant strains. As demonstrated by a number of studies that employed RNA-seq for such comparisons, perturbation of a single DNA MTase often induces the differential expression of tens and hundreds of genes, and as many as a thousand in some cases[16,25,43,104,114,115]. These data highlight the under-estimated impact of DNA methylation in the regulation of gene expression, but also reveal some unexpected findings. In some cases the regulation can be conclusively traced to methylation in the gene promoter. For instance, the MTase ModH5 in *H. pylori* has been shown to regulate the activity of the gene flagellin A *(flaA)* via methylation in the *flaA* promoter [151]. In general, however, only a small proportion (e.g. <10%) of the differentially expressed genes have methylated sites in their promoter regions[16,59,60,114]. This implies that most differentially expressed genes cannot be explained by the local competition model between a DNA MTase and other DNA binding proteins at the promoter of a gene (Fig. 3a). One possibility is that methylation status at individual motif sites might regulate the expression of a transcription factor, causing a broad downstream shift in the expression of

genes targeted by the transcription factor (Fig. 3b). In order to obtain mechanistic insights, specific methylation sites must be mutated individually using genetic tools such as site-directed mutagenesis[123–125]. Multiple studies have observed a positive correlation between the number of methylation sites in a gene and the fold change of expression between wildtype and MTase mutants[16,60], suggesting that epigenetic regulation of the expression may in fact be driven by multiple methylation sites in both the promoter region and gene body. Another intriguing hypothesis relates to the effect of DNA methylation on the chromosomal topology in the cell[157–159], whereby methylation induces structural changes that expose specific genes to the cellular transcriptional machinery (Fig. 3d).

Relationship with eukaryotic methylomes

DNA methylation studies in eukaryotic genomes have been focused on 5mC. Because it is much less prevalent in the bacterial kingdom, functional studies of 5mC have been rare, even with the advent of second generation and third-generation sequencing technologies. Therefore, the relationship between bacterial and eukaryotic methylomes has not been examined. However, the recent discovery of 6mA in a number of eukaryotes[160], including algae[161], fungi[162], worms[163], insects[164], and mammals[165,166], makes such a comparison possible. Specifically, as these recent studies have revealed diverse functions impacted by 6mA events in eukaryotes, including the regulation of gene expression[163,164,166], transposons[164,166], and cross talk with histone variants and modifications[163,166], a fundamental question remains: what are the similarities and differences in 6mA function between bacterial and eukaryotic genomes?

To approach this question, we must first recognize that these different kingdoms display different patterns of 6mA deposition in the genome. First, in contrast to its relatively high abundance in prokaryotes, 6mA frequency (as a fraction of the total number of adenine residues in the genome) is orders of magnitudes lower in most eukaryotes[78,160]. Second, in contrast to the highly motif driven 6mA deposition in prokaryotes, eukaryotic 6mA events are much less motif- driven. This is likely due to the fact that 6mA modified sites in eukaryotes are not targeted by cognate restriction enzymes and therefore do not need to be located at specific sequence motifs. Another reason may be that DNA MTases have limited access to DNA due to the existence of nucleosomes in eukaryotes. For example, 6mA motifs have been identified in *Chlamydomonas reinhardtii, Caenorhabditis elegans, Plasmodium falciparum* and mouse embryonic stem cells (mESCs), where very few occurrence (often <3%) of the motif across the genome sites are methylated, making them *weakly motif-driven*[78,161,163,166,167]. These differences have important implications in the use of third-generation sequencing for methylation mapping, as discussed in a recent work[78].

Despite these fundamental differences, some commonalities do exist. For example, 6mA events are known to repress a form of transposon (insertion elements) in bacteria[22,168], which is an analog of the observed enrichment of 6mA events at transposons in both *C. elegans* and mESCs[163,166]. More fundamentally, the intrinsic properties of 6mA and its impact on DNA conformation is expected to be consistent between bacteria and eukaryotes[157], although different organisms may exploit these properties in different molecular and cellular contexts. Importantly, high-resolution, complete maps of 6mA events

are the foundation of future studies comparing bacterial and eukaryotic 6mA methylomes. Although SMRT sequencing and Oxford nanopore sequencing hold great promise in the mapping of DNA methylation in bacteria, their successful application to eukaryotic genomes for DNA methylation detection face critical challenges. As recent work has suggested, 6mA detection in eukaryotes requires crossvalidation and integration between complementary sequencing and molecular technologies[78].

Conclusions and future directions

The study of bacterial methylomes has been revolutionized by the introduction of technologies capable of detecting 4mC, 5mC, and 6mA at genome-wide scale and single-nucleotide resolution. Application of these new technologies has led to a greater appreciation for the sheer quantity and diversity of methylation systems and their target specificities in bacteria. Deposition of newly discovered MTase genes and their target motifs to community databases like REBASE[28] has created a powerful resource for researchers, providing a catalog of the restriction enzymes that can act as barriers to efficient transformation and opening many opportunities to study epigenetic regulation in bacteria.

Beyond revealing stable MTase genes and their methylated motifs, technological advances have also helped to highlight hypervariable MTases and their consequences on genome-wide methylation, gene expression, and phenotypic plasticity. Phase-variable MTases have now been observed in a wide range of host-adapted pathogenic bacteria, with apparent roles in immune evasion, virulence, colonization and infection. Hypervariable switching of methylation activity, both ON/OFF and across multiple target specificities, has been shown to be capable of modulating expression of a large set of genes (a 'phasevarion'), but the mechanism of modulation remains unclear.

The situation is similar with stably expressed MTases. Although there are examples of gene expression being governed by methylation status in promoter regulatory regions, evidence suggests that this does not seem to account for the significant gene expression changes observed in MTases knockout studies. Comprehensive studies are necessary and would benefit from a richer collection of functional genomics data (e.g. transcription factor binding assays) of many bacterial species across different genetic background (wild type vs. MTase mutants) and growth/stress conditions, and must be followed by genetic experiments that mutate and characterize specific methylation sites. In addition, perhaps the thermodynamic effect of DNA methylation induces conformational changes to a bacterial chromosome, rendering previously inaccessible genes accessible to the transcriptional machinery[157,169] (Fig. 3d). Chromatin conformation capture (Hi-C) sequencing and experimental characterization is needed to elucidate the precise mechanisms at work[158].

The ability of phase-variable MTases to activate antigenic diversity in host-adapted pathogens (Fig. 3e) makes them very relevant from the perspective of vaccine development. Antigens that are known to possess diversity and variability do not make good vaccine candidates and are typically avoided. However, genes for outer membrane proteins or other antigens that lack simple tandem repeats (common indicators of phase variation) might still be subject to variable expression if they are part of a phasevarion[170]. It has been shown that

multiple vaccine candidates are likely subject to this epigenetic means of antigenic variation, highlighting the importance of identifying phase-variable MTases and their phasevarions in host-adapted pathogens[110] for the more effective development of vaccines.

While SMRT sequencing has been instrumental in enabling the study of bacterial methylomes, additional novel sequencing technologies, such as those commercialized by Oxford Nanopore Technologies, have the potential to make significant contributions to the field of bacterial epigenetics in the near future. Assuming continued maturation of the technology and improvements in the modification detection algorithms, the very long read lengths offered by nanopore sequencing devices should be able to provide single-molecule, phased detection of bacterial DNA methylation in samples from a variety of environments. This will be especially significant for the epigenetic study of heterogeneous bacterial samples, including metagenomic populations, where the study of methylation has so far been limited[25,102,171]. The recent use of methylation signatures as discriminative features for metagenomic binning suggests that the applications for methylation detection in long reads extend beyond identifying methylated motifs in bacteria.[118]

These advances come at a time when researchers are increasingly exploring the presence and importance in eukaryotes of some DNA methylation types, such as 6mA, that have traditionally only been recognized in prokaryotes[162,163,166]. Detection of these modifications in eukaryotes presents additional challenges stemming from the modification scarcity and lack of clear target motifs. However, as the significance and function of these epigenetic marks become better understood, it will be interesting to see whether these eukaryotic modifications share any functional traits with those found in their prokaryotic ancestors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledegements

References

1. Shendure J & Ji H Next-generation DNA sequencing. Nat. Biotechnol. 26, 1135–1145 (2008). [PubMed: 18846087]

2. Plongthongkum N, Diep DH & Zhang K Advances in the profiling of DNA modifications: cytosine methylation and beyond. Nat. Rev. Genet. 15, 647–661 (2014). [PubMed: 25159599]

3. Boyer H Genetic control of restriction and modification in Escherichi coli. J. Bacteriol. 88, 1652–60 (1964). [PubMed: 14240953]

4. Kumar S et al. The DNA (cytosine-5) methyltransferases. Nucleic Acids Res. 22, 1–10 (1994). [PubMed: 8127644]

5. Arand J et al. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. PLoS Genet. 8, e1002750 (2012).

6. Jones PA Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13, 484–92 (2012). [PubMed: 22641018]

7. Bell MV et al. Physical mapping across the fragile X: Hypermethylation and clinical expression of the fragile X syndrome. Cell 64, 861–866 (1991). [PubMed: 1997211]

8. Xu GL et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 402, 187–91 (1999). [PubMed: 10647011]

9. Baylin SB & Jones PA A decade of exploring the cancer epigenome - biological and translational implications. Nat. Rev. Cancer 11, 726–34 (2011). [PubMed: 21941284]

10. Robertson KD & Wolffe AP DNA methylation in health and disease. Nat. Rev. Genet. 1, 11–19 (2000). [PubMed: 11262868]

11. Ito S et al. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 466, 1129–1133 (2010). [PubMed: 20639862]

12. Smith ZD & Meissner A DNA methylation: roles in mammalian development. Nat. Rev. Genet. 14, 204–20 (2013). [PubMed: 23400093]

13. Casadesús J & Low D Epigenetic gene regulation in the bacterial world. Microbiol. Mol. Biol. Rev. 70, 830–56 (2006). [PubMed: 16959970]

14. Wion D & Casadesus J N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. Nat. Rev. Microbiol. 4, 183–92 (2006). [PubMed: 16489347]

15. Waldron DE, Owen P & Dorman CJ Competitive interaction of the OxyR DNA-binding protein and the Dam methylase at the antigen 43 gene regulatory region in Escherichia coli. Mol. Microbiol. 44, 509–520 (2002). [PubMed: 11972787]

16. Fang G et al. Genome-wide mapping of methylated adenine residues in pathogenic Escherichia coli using single-molecule real-time sequencing. Nat. Biotechnol. 30, 1232–9 (2012). [PubMed: 23138224]

17. LØbner-Olesen A, Marinus MG & Hansen FG Role of SeqA and Dam in Escherichia coli gene expression: a global/microarray analysis. Proc. Natl. Acad. Sci. U. S. A. 100, 4672–7 (2003). [PubMed: 12682301]

18. Low D. a & Casadesus J Clocks and switches: bacterial gene regulation by DNA adenine methylation. Curr. Opin. Microbiol. 11, 106–12 (2008). [PubMed: 18396448]

19. Boye E, L0bner-Olesen A & Skarstad K Limiting DNA replication to once and only once. EMBO Rep. 1, 479–83 (2000). [PubMed: 11263490]

20. Boye E, Stokke T, Kleckner N & Skarstad K Coordinating DNA replication initiation with cell growth: differential roles for DnaA and SeqA proteins. Proc. Natl. Acad. Sci. U. S. A. 93, 12206–11 (1996). [PubMed: 8901558]

21. Hsieh P Molecular mechanisms of DNA mismatch repair. Mutat. Res. 486, 71–87 (2001). [PubMed: 11425513]

22. Roberts D, Hoopes BC, McClure WR & Kleckner N IS10 transposition is regulated by DNA adenine methylation. Cell 43, 117–30 (1985). [PubMed: 3000598]

23. Hernday A, Krabbe M, Braaten B & Low D Self-perpetuating epigenetic pili switches in bacteria. Proc. Natl. Acad. Sci. U. S. A. 99 Suppl 4, 16470–6 (2002). [PubMed: 12202745]

24. Furuta Y & Kobayashi I Mobility of DNA sequence recognition domains in DNA methyltransferases suggests epigenetics-driven adaptive evolution. Mob. Genet. Elements 2, 292–296 (2012). [PubMed: 23481556]

25. Beaulaurier J et al. Single molecule-level detection and long read-based phasing of epigenetic variations in bacterial methylomes. Nat. Commun. 6, 7438 (2015). [PubMed: 26074426]

26. Casadesús J & Low D a. Programmed heterogeneity: epigenetic mechanisms in bacteria. J. Biol. Chem. 288, 13929–35 (2013). [PubMed: 23592777]

27. Srikhanta YN, Maguire TL, Stacey KJ, Grimmond SM & Jennings MP The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. Proc. Natl. Acad. Sci. U. S. A. 102, 5547–51 (2005). [PubMed: 15802471]

28. Roberts RJ, Vincze T, Posfai J & Macelis D REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. Nucleic Acids Res. 43, D298–D299 (2015). [PubMed: 25378308]

29. Blow MJ et al. The Epigenomic Landscape of Prokaryotes. PLOS Genet. 12, e1005854 (2016).

30. Davis BM, Chao MC & Waldor MK Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. Curr. Opin. Microbiol. 16, 192–8 (2013). [PubMed: 23434113]

31. Hayatsu H Discovery of bisulfite-mediated cytosine conversion to uracil, the key reaction for DNA methylation analysis — A personal account. Proc. Japan Acad. Ser. B, Phys. Biol. Sci. 84, 2–11 (2008).

32. Frommer M et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc. Natl. Acad. Sci. 89, 1827–1831 (1992). [PubMed: 1542678]

33. Cokus SJ et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452, 215–219 (2008). [PubMed: 18278030]

34. Lister R et al. Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. Cell 133, 523–536 (2008). [PubMed: 18423832]

35. Lister R et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462, 315–322 (2009). [PubMed: 19829295]

36. Nelson M, Christ C & Schildkraut I Alteration of apparent restriction endonuclease recognition specificities by DNA methylases. Nucleic Acids Res. 12, 5165–5173 (1984). [PubMed: 6087274]

37. Zweiger G, Marczynski G & Shapiro L A Caulobacter DNA Methyltransferase that Functions only in the Predivisional Cell. J. Mol. Biol. 235, 472–485 (1994). [PubMed: 8289276]

38. Eid J et al. Real-time DNA sequencing from single polymerase molecules. Science (80-.). 323, 133–138 (2009).

39. Flusberg B a et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat. Methods 7, 461–5 (2010). [PubMed: 20453866]

40. Murray I a et al. The methylomes of six bacteria. Nucleic Acids Res. 40, 11450–62 (2012). [PubMed: 23034806]

41. Krebes J et al. The complex methylome of the human gastric pathogen Helicobacter pylori. Nucleic Acids Res. 1–18 (2013). doi:10.1093/nar/gkt1201

42. Furuta Y et al. Methylome diversification through changes in DNA methyltransferase sequence specificity. PLoS Genet. 10, e1004272 (2014).

43. Lluch-Senar M et al. Comprehensive methylome characterization of Mycoplasma genitalium and Mycoplasma pneumoniae at single-base resolution. PLoS Genet. 9, e1003191 (2013).

44. Jain M, Olsen HE, Paten B & Akeson M The Oxford Nanopore MinlON: delivery of nanopore sequencing to the genomics community. Genome Biol. 17, 239 (2016). [PubMed: 27887629]

45. Sánchez-Romero M.a, Cota I & Casadesus J DNA methylation in bacteria: from the methyl group to the methylome. Curr. Opin. Microbiol. 25, 9–16 (2015). [PubMed: 25818841]

46. Casadesús J in DNA Methyltransferases - Role and Function (eds. Jeltsch A & Jurkowska RZ) 35–61 (Springer International Publishing, 2016). doi:10.1007/978-3-319-43624-1_3

47. Razin A & Riggs AD DNA methylation and gene function. Science (80-.). 210, 604–10 (1980).

48. Robertson KD DNA methylation and human disease. Nat Rev Genet 6, 597–610 (2005). [PubMed: 16136652]

49. Bock C Analysing and interpreting DNA methylation data. Nat. Rev. Genet. 13, 705–719 (2012). [PubMed: 22986265]

50. Laird PW Principles and challenges of genomewide DNA methylation analysis. Nat. Rev. Genet. 11, 191–203 (2010). [PubMed: 20125086]

51. Hirst M & Marra M. a. Next generation sequencing based approaches to epigenomics. Brief. Funct. Genomics 9, 455–65 (2010). [PubMed: 21266347]

52. Nelson M, Raschke E & McClelland M Effect of site-specific methylation on restriction endonucleases and DNA modification methyltransferases. Nucleic Acids Res. 21, 3139–54 (1993). [PubMed: 8392715]

53. Rao BS & Buckler-White A Direct visualization of site-specific and strand-specific DNA methylation patterns in automated DNA sequencing data. Nucleic Acids Res. 26, 2505–7 (1998). [PubMed: 9580708]

54. Bart a, van Passel MWJ, van Amsterdam K & van der Ende a. Direct detection of methylation in genomic DNA. Nucleic Acids Res. 33, e124 (2005). [PubMed: 16091626]

55. Broadbent SE, Balbontin R, Casadesus J, Marinus MG & van der Woude M YhdJ, a nonessential CcrM-like DNA methyltransferase of Escherichia coli and Salmonella enterica. J. Bacteriol. 189, 4325–7 (2007). [PubMed: 17400740]

56. Shell SS et al. DNA methylation impacts gene expression and ensures hypoxic survival of Mycobacterium tuberculosis. PLoS Pathog. 9, e1003419 (2013).

57. Bart A, Pannekoek Y, Dankert J & Van Der Ende A NmeSI restriction-modification system identified by representational difference analysis of a hypervirulent Neisseria meningitidis strain. Infect. Immun. 69, 1816–1820 (2001). [PubMed: 11179359]

58. Korlach J & Turner SW Going beyond five bases in DNA sequencing. Curr. Opin. Struct. Biol. 22, 251–61 (2012). [PubMed: 22575758]

59. Kahramanoglou C et al. Genomics of DNA cytosine methylation in Escherichia coli reveals its role in stationary phase transcription. Nat. Commun. 3, 886 (2012). [PubMed: 22673913]

60. Chao MC et al. A Cytosine Methyltransferase Modulates the Cell Envelope Stress Response in the Cholera Pathogen. PLoS Genet. 11, 1–24 (2015).

61. Yu M et al. Base-resolution detection of N4-methylcytosine in genomic DNA using 4mC-Tet-assisted-bisulfite sequencing. Nucleic Acids Res. 43, 1–10 (2015). [PubMed: 25505162]

62. Schadt EE et al. Modeling kinetic rate variation in third generation DNA sequencing data to detect putative modifications to DNA bases. Genome Res. 23, 129–41 (2013). [PubMed: 23093720]

63. Levene MJ et al. Zero-mode waveguides for single-molecule analysis at high concentrations. Science (80-.). 299, 682–6 (2003).

64. Clark T. a, Spittle KE, Turner SW & Korlach J Direct detection and sequencing of damaged DNA bases. Genome Integr. 2, 10 (2011). [PubMed: 22185597]

65. Clark T a et al. Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. BMC Biol. 11, 4 (2013). [PubMed: 23339471]

66. Clarke J et al. Continuous base identification for single-molecule nanopore DNA sequencing. Nat. Nanotechnol. 4, 265–270 (2009). [PubMed: 19350039]

67. Manrao E a et al. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. Nat. Biotechnol. 30, 349–53 (2012). [PubMed: 22446694]

68. Laszlo AH et al. Decoding long nanopore sequencing reads of natural DNA. Nat. Biotechnol. 32, 829–834 (2014). [PubMed: 24964173]

69. Manrao EA, Derrington IM, Pavlenok M, Niederweis M & Gundlach JH Nucleotide discrimination with DNA immobilized in the MSPA nanopore. PLoS One 6, 1–7 (2011).

70. Deamer D, Akeson M & Branton D Three decades of nanopore sequencing. Nat. Biotechnol. 34, 518–524 (2016). [PubMed: 27153285]

71. Ip CLC et al. MinION Analysis and Reference Consortium: Phase 1 data release and analysis. F1000Research (2015). doi:10.12688/f1000research.7201.1

72. Laszlo AH et al. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. Proc. Natl. Acad. Sci. U. S. A. 110, 18904–9 (2013). [PubMed: 24167255]

73. Wescoe ZL, Schreiber J & Akeson M Nanopores discriminate among five C5-cytosine variants in DNA. J. Am. Chem. Soc. 136, 16582–7 (2014). [PubMed: 25347819]

74. Simpson JT et al. Detecting DNA cytosine methylation using nanopore sequencing. Nat. Methods 14, 407–410 (2017). [PubMed: 28218898]

75. Rand AC et al. Mapping DNA methylation with high-throughput nanopore sequencing. Nat. Methods 14, 411–413 (2017). [PubMed: 28218897]

76. McIntyre ABR et al. Nanopore detection of bacterial DNA base modifications. bioRxiv 0–10 (2017).

77. Stoiber MH et al. De novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing. bioRxiv 094672 (2016). doi:10.1101/094672

78. Zhu S et al. Mapping and characterizing N6-methyladenine in eukaryotic genomes using single-molecule real-time sequencing. Genome Res. gr.231068.117 (2018). doi:10.1101/gr.231068.117

79. Blyn LB, Braaten BA & Low DA Regulation of pap pilin phase variation by a mechanism involving differential dam methylation states. EMBO J. 9, 4045–4054 (1990). [PubMed: 2147413]

80. Boyer HW DNA restriction and modification mechanisms in bacteria. Annu. Rev. Microbiol. 25, 153–76 (1971). [PubMed: 4949033]

81. LØbner-Olesen A, Skovgaard O & Marinus MG Dam methylation: Coordinating cellular processes. Curr. Opin. Microbiol. 8, 154–160 (2005). [PubMed: 15802246]

82. Ehrlich M, Wilson GG, Kuo KC & Gehrke CW N4-methylcytosine as a minor base in bacterial DNA. J. Bacteriol. 169, 939–43 (1987). [PubMed: 3029036]

83. Ehrlich M et al. DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N5methyladenine. Nucleic Acids Res. 13, 1399–1412 (1985). [PubMed: 4000939]

84. Chung D, Farkas J, Huddleston JR, Olivar E & Westpheling J Methylation by a unique a-class N4-Cytosine methyltransferase is required for DNA transformation of caldicellulosiruptor bescii DSM6725. PLoS One 7, 1–9 (2012).

85. Huo W, Adams HM, Zhang MQ & Palmer KL Genome modification in Enterococcus faecalis OG1RF assessed by bisulfite sequencing and single-molecule real-time sequencing. J. Bacteriol. 197, 1939–1951 (2015). [PubMed: 25825433]

86. Vilkaitis G & Klimasauskas S Bisulfite sequencing protocol displays both 5-methylcytosine and N4-methylcytosine. Anal. Biochem. 271, 116–9 (1999). [PubMed: 10361019]

87. Lee WC et al. The complete methylome of Helicobacter pylori UM032. BMC Genomics 16, 424 (2015). [PubMed: 26031894]

88. O'Loughlin JL et al. Analysis of the Campylobacter jejuni genome by SMRT DNA sequencing identifies restriction-modification motifs. PLoS One 10, 1–18 (2015).

89. Pirone-Davies C et al. Genome-wide methylation patterns in Salmonella enterica subsp. enterica Serovars. PLoS One 10, 1–13 (2015).

90. Kumar S et al. N4-cytosine DNA methylation regulates transcription and pathogenesis in Helicobacter pylori. Nucleic Acids Res. 1–17 (2018). doi:10.1093/nar/gky126

91. Boyer HW, Chow LT, Dugaiczyk A, Hedgpeth J & Goodman HM DNA substrate site for the EcoRII restriction endonuclease and modification methylase. Nat. New Biol. 244, 40–3 (1973). [PubMed: 4124385]

92. Takahashi N, Naito Y, Handa N & Kobayashi I A DNA methyltransferase can protect the genome from postdisturbance attack by a restriction-modification gene complex. J. Bacteriol. 184, 6100–8 (2002). [PubMed: 12399478]

93. Yang MK, Ser SC & Lee CH Involvement of E. coli dcm methylase in Tn3 transposition. Proc. Natl. Sci. Counc. Repub. China. B. 13, 276–83 (1989). [PubMed: 2561572]

94. Korba BE & Hays JB Partially deficient methylation of cytosine in DNA at CCATGG sites stimulates genetic recombination of bacteriophage lambda. Cell 28, 531–41 (1982). [PubMed: 6280874]

95. Militello KT et al. Conservation of Dcm-mediated cytosine DNA methylation in Escherichia coli. FEMS Microbiol. Lett. 328, 78–85 (2012). [PubMed: 22150247]

96. Kozdon JB et al. Global methylation state at base-pair resolution of the Caulobacter genome throughout the cell cycle. Proc. Natl. Acad. Sci. U. S. A. 110, E4658–67 (2013). [PubMed: 24218615]

97. Anton BP, Harhay GP, Smith TPL, Blom J & Roberts RJ Comparative methylome analysis of the occasional ruminant respiratory pathogen bibersteinia trehalosi. PLoS One 11, 1–17 (2016).

98. Sater MRA et al. DNA Methylation assessed by SMRT sequencing is linked to mutations in neisseria meningitidis isolates. PLoS One 10, 1–19 (2015).

99. O'Connell Motherway M. et al. Identification of restriction-modification systems of Bifidobacterium animalis subsp. lactis CNCM I-2494 by SMRT sequencing and associated methylome analysis. PLoS One 9, e94875 (2014).

100. O'Callaghan A & van Sinderen D Bifidobacteria and Their Role as Members of the Human Gut Microbiota. Front. Microbiol. 7, 925 (2016).

101. Dalia AB, Lazinski DW & Camilli A Characterization of undermethylated sites in Vibrio cholerae. J. Bacteriol. 195, 2389–99 (2013). [PubMed: 23504020]

102. Leonard MT et al. The methylome of the gut microbiome: disparate Dam methylation patterns in intestinal Bacteroides dorei. Front. Microbiol. 5, 361 (2014). [PubMed: 25101067]

103. Chen P et al. Comparative Genomics Reveals the Diversity of Restriction-Modification Systems and DNA Methylation Sites in Listeria monocytogenes. Appl. Environ. Microbiol. 83, 1–16 (2017).

104. Manso AS et al. A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. Nat. Commun. 5, 5055 (2014). [PubMed: 25268848]

105. Li J et al. Epigenetic Switch Driven by DNA Inversions Dictates Phase Variation in Streptococcus pneumoniae. PLoS Pathog. 12, 1–36 (2016).

106. Zhu L et al. Precision methylome characterization of Mycobacterium tuberculosis complex (MTBC) using PacBio single-molecule real-time (SMRT) technology. Nucleic Acids Res. 44, 730–743 (2016). [PubMed: 26704977]

107. Mou KT et al. A comparative analysis of methylome profiles of Campylobacter jejuni sheep abortion isolate and gastroenteric strains using PacBio data. Front. Microbiol. 5, 1–15 (2014). [PubMed: 24478763]

108. Anjum A et al. 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, 4581–94 (2016). [PubMed: 26786317]

109. Cohen NR et al. A role for the bacterial GATC methylome in antibiotic stress survival. Nat. Genet. 48, 581–586 (2016). [PubMed: 26998690]

110. Atack JM et al. A biphasic epigenetic switch controls immunoevasion, virulence and niche adaptation in non-typeable Haemophilus influenzae. Nat. Commun. 6, 7828 (2015). [PubMed: 26215614]

111. Blakeway LV et al. ModM DNA methyltransferase methylome analysis reveals a potential role for Moraxella catarrhalis phasevarions in otitis media. FASEB J. 28, 5197–5207 (2014). [PubMed: 25183669]

112. Seib KL et al. Specificity of the ModA11, ModA12 and ModD1 epigenetic regulator N6- adenine DNA methyltransferases of Neisseria meningitidis. Nucleic Acids Res. 1–13 (2015). doi: 10.1093/nar/gkv219

113. Bendall ML et al. Exploring the roles of DNA methylation in the metal-reducing bacterium Shewanella oneidensis MR-1. J. Bacteriol. 195, 4966–74 (2013). [PubMed: 23995632]

114. Gonzalez D, Kozdon JB, McAdams HH, Shapiro L & Collier J The functions of DNA methylation by CcrM in Caulobacter crescentus: a global approach. Nucleic Acids Res. 1–16 (2014). doi:10.1093/nar/gkt1352

115. Zhou B et al. The Global Regulatory Architecture of Transcription during the Caulobacter Cell Cycle. PLoS Genet. 11, e1004831 (2015).

116. Murphy J et al. Methyltransferases acquired by lactococcal 936-type phage provide protection against restriction endonuclease activity. BMC Genomics 15, 831 (2014). [PubMed: 25269955]

117. Goldfarb T et al. BREX is a novel phage resistance system widespread in microbial genomes. EMBO 1–16 (2014).

118. Beaulaurier J et al. Metagenomic binning and association of plasmids with bacterial host genomes using DNA methylation. Nat. Biotechnol. 36, 61–69 (2018). [PubMed: 29227468]

119. Srikhanta YN, Fox KL & Jennings MP The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. Nat. Rev. Microbiol. 8, 196–206 (2010). [PubMed: 20140025]

120. Kobayashi I, Nobusato a, Kobayashi-Takahashi N & Uchiyama I Shaping the genome-- restriction-modification systems as mobile genetic elements. Curr. Opin. Genet. Dev. 9, 649–656 (1999). [PubMed: 10607611]

121. Conlan S et al. Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. Sci. Transl. Med. 6, 254–126 (2014).

122. Balbontin R et al. DNA adenine methylation regulates virulence gene expression in Salmonella enterica serovar Typhimurium. J. Bacteriol. 188, 8160–8168 (2006). [PubMed: 16997949]

123. Woude M Van der, Braaten B & Low D Epigenetic phase variation of the pap operon in Escherichia coli. Trends Microbiol. 4, 5–9 (1996). [PubMed: 8824788]

124. Wallecha A, Munster V, Correnti J, Chan T & van der Woude M Dam-and OxyR-dependent phase variation of agn43: essential elements and evidence for a new role of DNA methylation. J. Bacteriol. 184, 3338–3347 (2002). [PubMed: 12029051]

125. Lim HN & van Oudenaarden A A multistep epigenetic switch enables the stable inheritance of DNA methylation states. Nat. Genet. 39, 269–75 (2007). [PubMed: 17220888]

126. Peterson SN & Reich NO GATC flanking sequences regulate Dam activity: evidence for how Dam specificity may influence pap expression. J. Mol. Biol. 355, 459–72 (2006). [PubMed: 16321401]

127. Davies MR, Broadbent SE, Harris SR, Thomson NR & van der Woude MW Horizontally Acquired Glycosyltransferase Operons Drive Salmonellae Lipopolysaccharide Diversity. PLoS Genet. 9, (2013).

128. Broadbent SE, Davies MR & Van Der Woude MW Phase variation controls expression of Salmonella lipopolysaccharide modification genes by a DNA methylation- dependent mechanism. Mol. Microbiol. 77, 337–353 (2010). [PubMed: 20487280]

129. Cota I, Blanc-Potard AB & Casadesus J STM2209-STM2208 (opvAB): A phase variation locus of salmonella enterica involved in control of O-antigen chain length. PLoS One 7, (2012).

130. Camacho EM & Casadesus J Regulation of traJ transcription in the Salmonella virulence plasmid by strand-specific DNA adenine hemimethylation. Mol. Microbiol. 57, 1700–1718 (2005). [PubMed: 16135235]

131. Cota I et al. OxyR-dependent formation of DNA methylation patterns in OpvAB [OFF] and OpvAB [ON] cell lineages of *Salmonella enterica*. Nucleic Acids Res. 44, 3595–3609 (2016). [PubMed: 26687718]

132. Jennings MP, Hood DW, Peak IR, Virji M & Moxon ER Molecular analysis of a locus for the biosynthesis and phase-variable expression of the lacto-N-neotetraose terminal lipopolysaccharide structure in *Neisseria meningitidis*. Mol. Microbiol. 18, 729–740 (1995).

133. van der Ende a et al. Variable expression of class 1 outer membrane protein in Neisseria meningitidis is caused by variation in the spacing between the −10 and −35 regions of the promoter. J. Bacteriol. 177, 2475–80 (1995). [PubMed: 7730280]

134. Cerdeno-Tarraga A & Patrick S Extensive DNA Inversions in the B. fragilis Genome Control Variable Gene Expression. Science (80-.). 1463–1466 (2005).

135. Henderson IR, Owen P & Nataro JP Molecular switches - the ON and OFF of bacterial phase variation. Mol. Microbiol. 33, 919–932 (1999). [PubMed: 10476027]

136. Atack JM, Tan A, Bakaletz LO, Jennings MP & Seib KL Phasevarions of Bacterial Pathogens: Methylomics Sheds New Light on Old Enemies. Trends Microbiol. (2018). doi:10.1016/j.tim.2018.01.008

137. Atack JM, Yang Y, Seib KL, Zhou Y & Jennings MP A survey of Type III restriction-modification systems reveals numerous, novel epigenetic regulators controlling phase-variable regulons; phasevarions. Nucleic Acids Res. 46, 3532–3542 (2018). [PubMed: 29554328]

138. Dybvig K, Sitaraman R & French CT A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. Proc. Natl. Acad. Sci. U. S. A. 95, 13923–8 (1998). [PubMed: 9811902]

139. Tettelin H et al. Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. Science (80-.). 293, 498–506 (2001).

140. Ryan KA & Lo RY Characterization of a CACAG pentanucleotide repeat in Pasteurella haemolytica and its possible role in modulation of a novel type Ill restriction- modification system. Nucleic Acids Res. 27, 1505–11 (1999). [PubMed: 10037813]

141. Seib KL, Peak IRA & Jennings MP Phase variable restriction-modification systems in Moraxella catarrhalis. FEMS Immunol. Med. Microbiol. 32, 159–65 (2002). [PubMed: 11821238]

142. Zaleski P, Wojciechowski M & Piekarowicz A The role of Dam methylation in phase variation of Haemophilus influenzae genes involved in defence against phage infection. Microbiology 151, 3361–3369 (2005). [PubMed: 16207918]

143. Fox KL et al. Haemophilus influenzae phasevarions have evolved from type III DNA restriction systems into epigenetic regulators of gene expression. Nucleic Acids Res. 35, 5242–5252 (2007). [PubMed: 17675301]

144. Gawthorne JA, Beatson SA, Srikhanta YN, Fox KL & Jennings MP Origin of the diversity in DNA recognition domains in phasevarion associated modA genes of pathogenic Neisseria and Haemophilus influenzae. PLoS One 7, 1–10 (2012).

145. De Vries N et al. Transcriptional phase variation of a type III restriction-modification system in Helicobacter pylori. J. Bacteriol. 184, 6615–6623 (2002). [PubMed: 12426350]

146. Skoglund A et al. Functional analysis of the M.HpyAIV DNA methyltransferase of Helicobacter pylori. J. Bacteriol. 189, 8914–21 (2007). [PubMed: 17921292]

147. Srikhanta YN et al. Phasevarion mediated epigenetic gene regulation in Helicobacter pylori. PLoS One 6, e27569 (2011).

148. Srikhanta YN et al. Phasevarions mediate random switching of gene expression in pathogenic Neisseria. PLoS Pathog. 5, e1000400 (2009).

149. Seib KL et al. A novel epigenetic regulator associated with the hypervirulent Neisseria meningitidis clonal complex 4¼4. FASEB J. 25, 3622–3633 (2011). [PubMed: 21680891]

150. Jen FEC, Seib KL & Jennings MP Phasevarions mediate epigenetic regulation of antimicrobial susceptibility in Neisseria meningitidis. Antimicrob. Agents Chemother. 58, 4219–4221 (2014). [PubMed: 24777094]

151. Srikhanta YN et al. Methylomic and phenotypic analysis of the ModH5 phasevarion of Helicobacter pylori. Sci. Rep. 7, 1–14 (2017). [PubMed: 28127051]

152. Heithoff DM, Sinsheimer RL, Low DA & Mahan MJ An essential role for DNA adenine methylation in bacterial virulence. Science (80-.). 284, 967–970 (1999).

153. Garcia-Del Portillo F., Pucciarelli MG & Casadesus J DNA adenine methylase mutants of Salmonella typhimurium show defects in protein secretion, cell invasion, and M cell cytotoxicity. Proc. Natl. Acad. Sci. 96, 11578–11583 (1999). [PubMed: 10500219]

154. Brockman KL et al. ModA2 phasevarion switching in nontypeable haemophilus influenzae increases the severity of experimental otitis media. J. Infect. Dis. 214, 817–824 (2016). [PubMed: 27288538]

155. Brockman KL et al. The ModA2 Phasevarion of nontypeable Haemophilus influenzae Regulates Resistance to Oxidative Stress and Killing by Human Neutrophils. Sci. Rep. 7, 1–11 (2017). [PubMed: 28127051]

156. VanWagoner TM et al. The modA10 phasevarion of nontypeable Haemophilus influenzae R2866 regulates multiple virulence-associated traits. Microb. Pathog. 92, 60–67 (2016). [PubMed: 26718097]

157. Polaczek P, Kwan K & Campbell JL GATC motifs may alter the conformation of DNA depending on sequence context and N 6-adenine methylation status: possible implications for DNA-protein recognition. Mol. Gen. Genet. MGG 258, 488–493 (1998). [PubMed: 9669330]

158. Le T, Imakaev M, Mirny L, Science ML- & 2013, undefined. High-resolution mapping of the spatial organization of a bacterial chromosome. science.sciencemag.org

159. Diekmann S DNA methylation can enhance or induce DNA curvature. EMBO J. 6, 4213–4217 (1987). [PubMed: 2832157]

160. Luo G-Z & He C DNA N6-methyladenine in metazoans: functional epigenetic mark or bystander? Nat. Struct. Mol. Biol. 24, 503–506 (2017). [PubMed: 28586322]

161. Fu Y et al. N 6-Methyldeoxyadenosine Marks Active Transcription Start Sites in Chlamydomonas. Cell 161, 879–892 (2015). [PubMed: 25936837]

162. Mondo SJ et al. Widespread adenine N6-methylation of active genes in fungi. Nat. Genet. 49, 964–968 (2017). [PubMed: 28481340]

163. Greer EL et al. DNA methylation on N6-adenine in C. elegans. Cell 161, 868–878 (2015). [PubMed: 25936839]

164. Zhang G et al. N 6-Methyladenine DNA Modification in Drosophila. Cell 161, 893–906 (2015). [PubMed: 25936838]

165. Koziol MJ et al. Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. Nat. Struct. Mol. Biol. 23, 24–30 (2016). [PubMed: 26689968]

166. Wu TP et al. DNA methylation on N6-adenine in mammalian embryonic stem cells. Nature 532, 1–18 (2016).

167. Luo G-Z et al. Characterization of eukaryotic DNA N6-methyladenine by a highly sensitive restriction enzyme-assisted sequencing. Nat. Commun. 7, (2016).

168. Yin JC, Krebs MP & Reznikoff WS Effect of dam methylation on Tn5 transposition. J. Mol. Biol. 199, 35–45 (1988). [PubMed: 2451025]

169. Ngo TTM et al. Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. Nat. Commun. 7, 1–9 (2016).

170. Tan A, Atack JM, Jennings MP & Seib KL The capricious nature of bacterial pathogens: Phasevarions and vaccine development. Front. Immunol. 7, 1–9 (2016). [PubMed: 26834743]

171. Beckmann ND, Karri S, Fang G & Bashir A Detecting epigenetic motifs in low coverage and metagenomics settings. BMC Bioinformatics 15, S16 (2014).

172. Bickle TA & Kruger DH Biology of DNA restriction. Microbiol Rev 57, 434–450 (1993). [PubMed: 8336674]

173. Murray NE. 2001 Fred Griffith review lecture. Immigration control of DNA in bacteria: self versus non-self. Microbiology 148, 3–20 (2002). [PubMed: 11782494]

174. Loenen WAM, Dryden DTF, Raleigh EA & Wilson GG Type i restriction enzymes and their relatives. Nucleic Acids Res. 42, 20–44 (2014). [PubMed: 24068554]

175. Pingoud A, Wilson GG & Wende W Type II restriction endonucleases--a historical perspective and more. Nucleic Acids Res. 42, 7489–7527 (2014). [PubMed: 24878924]

176. Rao DN, Dryden DTF & Bheemanaik S Type III restriction-modification enzymes: A historical perspective. Nucleic Acids Res. 42, 45–55 (2014). [PubMed: 23863841]

177. Lacks S & Greenberg B Complementary specificity of restriction endonucleases of Diplococcus pneumoniae with respect to DNA methylation. J. Mol. Biol. 114, 153–168 (1977). [PubMed: 20509]

178. Loenen WAM & Raleigh EA The other face of restriction: Modification-dependent enzymes. Nucleic Acids Res. 42, 56–69 (2014). [PubMed: 23990325]

179. O'Connor BD, Merriman B & Nelson SF SeqWare Query Engine: storing and searching sequence data in the cloud. BMC Bioinformatics 11, S2 (2010).

180. Clark T a et al. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. Nucleic Acids Res. 40, e29 (2012). [PubMed: 22156058]

181. Banerjee S & Chowdhury R An orphan DNA (cytosine-5-)-methyltransferase in Vibrio cholerae. Microbiology 152, 1055–1062 (2006). [PubMed: 16549669]

Box 1:

Restriction-modification (RM) systems

RM systems, found in bacterial and archaeal genomes, serve an innate immune function by identifying and digesting foreign DNA. Complementary, and often proximally encoded, RM system components target a specific sequence motif on endogenous DNA for protective methylation, while any exogenous DNA lacking the protective methylation at that motif is targeted for restriction[14,172,173]. RM systems are divided into four categories based on the subunits involved and the precise site of DNA restriction.

Type I systems are comprised of a single enzyme containing restriction (R), modification (M), and specificity (S) subunits. They target bipartite motifs, where two short, specific sequence sub-motifs are separated by a fixed number of non-specific nucleotides (e.g. 5'-**ACG**NNNNNNGTT-3', methylated position in bold), and cleavage can occur several kilobases (kb) away from the non-methylated motif site[174].

Type II RM systems are comprised of separately transcribed MTase and restriction enzymes that target short, palindromic motifs (e.g. 5'-**GATC**-3') and cleave DNA close to the non- methylated motif sites. An exception to this are the Type IIG RM systems, where the MTase and restriction enzyme activities are encoded in a single polypeptide chain and the target motifs are short, non-palindromic sequences[175].

Type III RM systems also target short, non-palindromic sequences, but the specificity element is contained in the MTase enzyme. Non-methylated motif sites are targeted by a separate restriction enzyme, which must bind with the MTase to achieve sequence specificity[176].

Type IV RM systems are not strictly RM systems, as the restriction enzymes target methylated, rather than non-methylated, motifs for cleavage and lack a cognate MTase[177,178].

Orphan MTases are those that do not belong to an RM system and therefore lack a cognate restriction enzyme. Because non-methylated target motifs are not targeted by a cognate restriction enzyme, it is hypothesized that orphan MTases are not involved in defense against foreign DNA and rather more likely to serve in a regulatory capacity.

Box 2:

Matching methylated motifs to MTases

Comprehensive mapping of a bacterial methylome requires more than just detection of the methylated nucleotides and subsequent identification of the methylation motifs; it also requires identification of the MTase responsible for the observed methylation. Gene prediction and homology searching tools like SEQWARE[179] are often used to identify genes likely to be components of an RM system, including subunits responsible for restriction (R), specificity (S), and MTase (M) activity. These components are typically encoded by genes proximal to each other in the genome and can be classified by RM system type (**Box 1**) based on type-specific functional domains. Once classified by type, the characteristic methylation properties of the different RM system types can be leveraged to narrow the list of putative MTases responsible for an observed methylation motif. For instance, type I MTases target complementary bipartite motifs on both strands, while type IIG and III MTases target contiguous, non-palindromic motifs on a single strand. After narrowing the set of candidate MTase genes, the sequences of these candidates can be queried against MTases sequences with known motif specificities in REBASE[28], where a high-quality sequence match is often sufficient for a confident mapping[29,40].

Absent a high-quality MTase match in REBASE, two experimental approaches can be used to identify the MTase gene responsible for an observed methylated motif. The first relies on heterologous expression of the putative MTase gene in an otherwise non-methylated host, such as *E. coli* ER2796[16,40,110,116,180]. Alternatively, the putative MTases can be subject to an inactivating mutation, where the mutation is either carried out experimentally[41,113] or occurs naturally in a related strain[111,149]. If heterologous expression of the MTase results in methylation of the motif in question, or if inactivation of the MTase abolishes methylation at that motif, the causal role of that MTase is confirmed.

Box 3:

SMRT sequencing for detecting modified bases

Single-molecule, real-time (SMRT) sequencing describes the process of real-time monitoring of the incorporation of fluorescently labeled deoxyribonucleoside triphosphates (dNTPs) during replication of a template DNA molecule (Fig. 2a)[38]. Variations in the speed of nucleotide incorporation events (i.e. the observed fluorescent pulses), termed the *polymerase kinetics* and quantified using inter-pulse duration (IPD) metrics, provide an additional data dimension for interrogating the template DNA. Base modification events disrupt the observed polymerase kinetics and can be detected through comparisons with control IPD values (Fig. 2b).

Control values

IPD values from sequencing of native DNA can be compared against control IPD values from either methylation-free whole-genome amplified (WGA) DNA or pre-computed *in silico* IPD models. The *in silico* model is trained using large amounts of sequencing data from unmodified DNA and consists of predicted IPD values for a given local sequence context[40].

Local sequence context

The processivity of the polymerase is highly dependent on the precise nucleotide sequence surrounding the site of nucleotide incorporation. This causes fluctuations in IPD values that must be accounted for when looking for IPD deviations indicative of a base modification event[39],[62].

Modification type

Due to contact between the polymerase and the template DNA molecule extending over multiple nucleotides, a modified base can affect the IPD values both upstream and downstream of the modified position. The resulting IPD signatures across multiple positions surrounding each modified position follow patterns that are correlated with the type of modification present (6mA, 4mC, 5mC). By comparing the local IPD signature to known IPD signatures for various modification types, it is usually possible to assign a modification type to an observed methylation motif[39].

Detection statistics

The vast majority of SMRT methylome studies have utilized consensus IPD values assessed across deep coverage of aligned reads to identify methylated positions and motifs using standard statistical tests such as student's t-test [16],[39],[180]. Alternative methods and statistical models have been proposed for methylation detection in low coverage SMRT $_{sequencing}$[25],[118],[171].
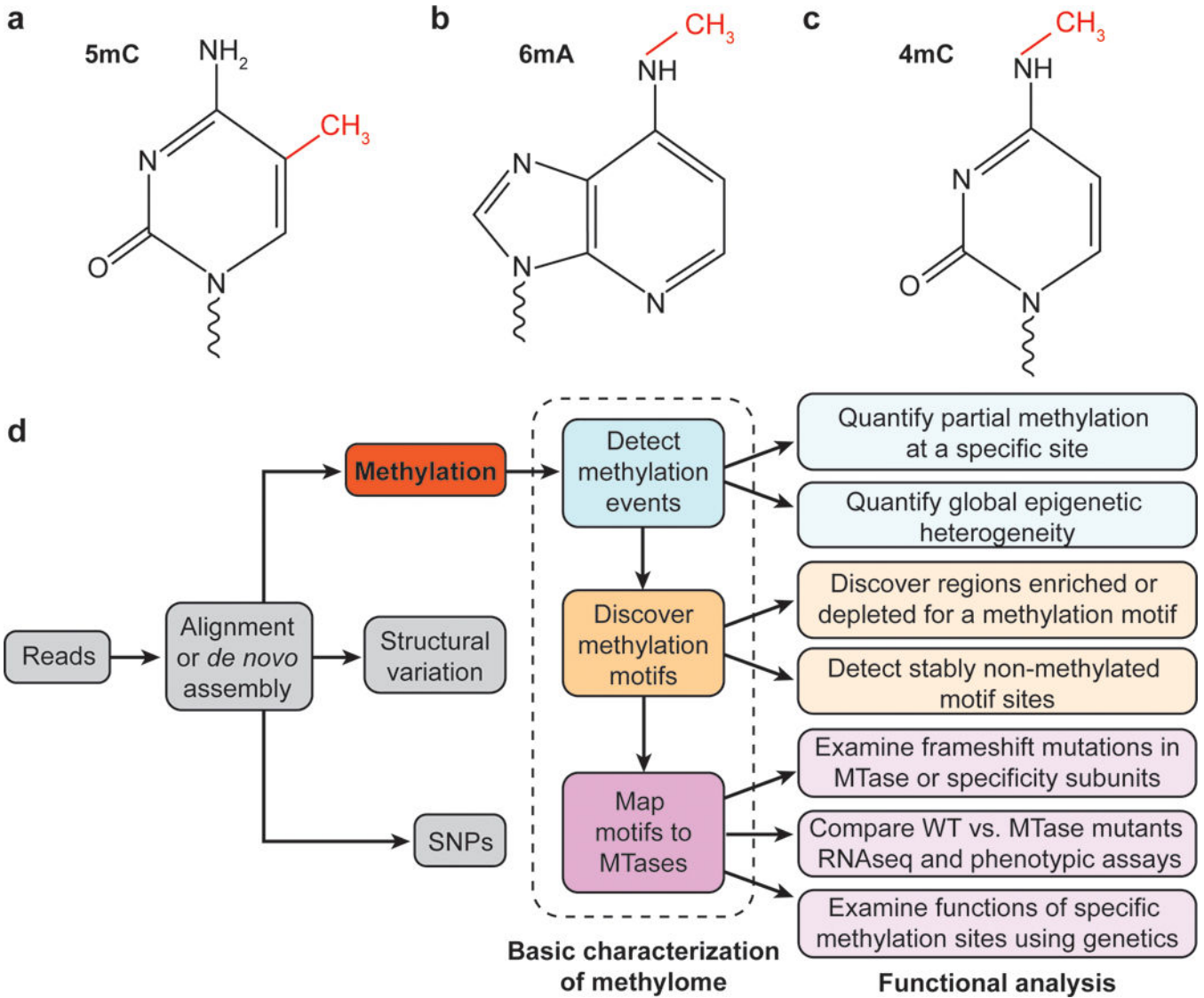
Figure 1: DNA methylation in bacteria as a mechanism of phenotypic plasticity.
Chemical structures of the most common forms of DNA methylation in bacteria, including
(a) 5-methylcytosine, **(b)** N6-methyladenine, and **(c)** N4-methylcytosine. **(d)** Methylome
characterization is increasingly becoming a standard component of bacterial genomic
research. The detection of methylated positions can lead to the identification of precise
methylated sequence motifs. A methylated motif can then be assigned to the responsible
MTase based on either querying a database of MTases with known target motifs or through
experimental means, involving comparisons with strains where the MTase is inactivated
(Box 2). Multiple lines of functional investigation can lead from this basic characterization
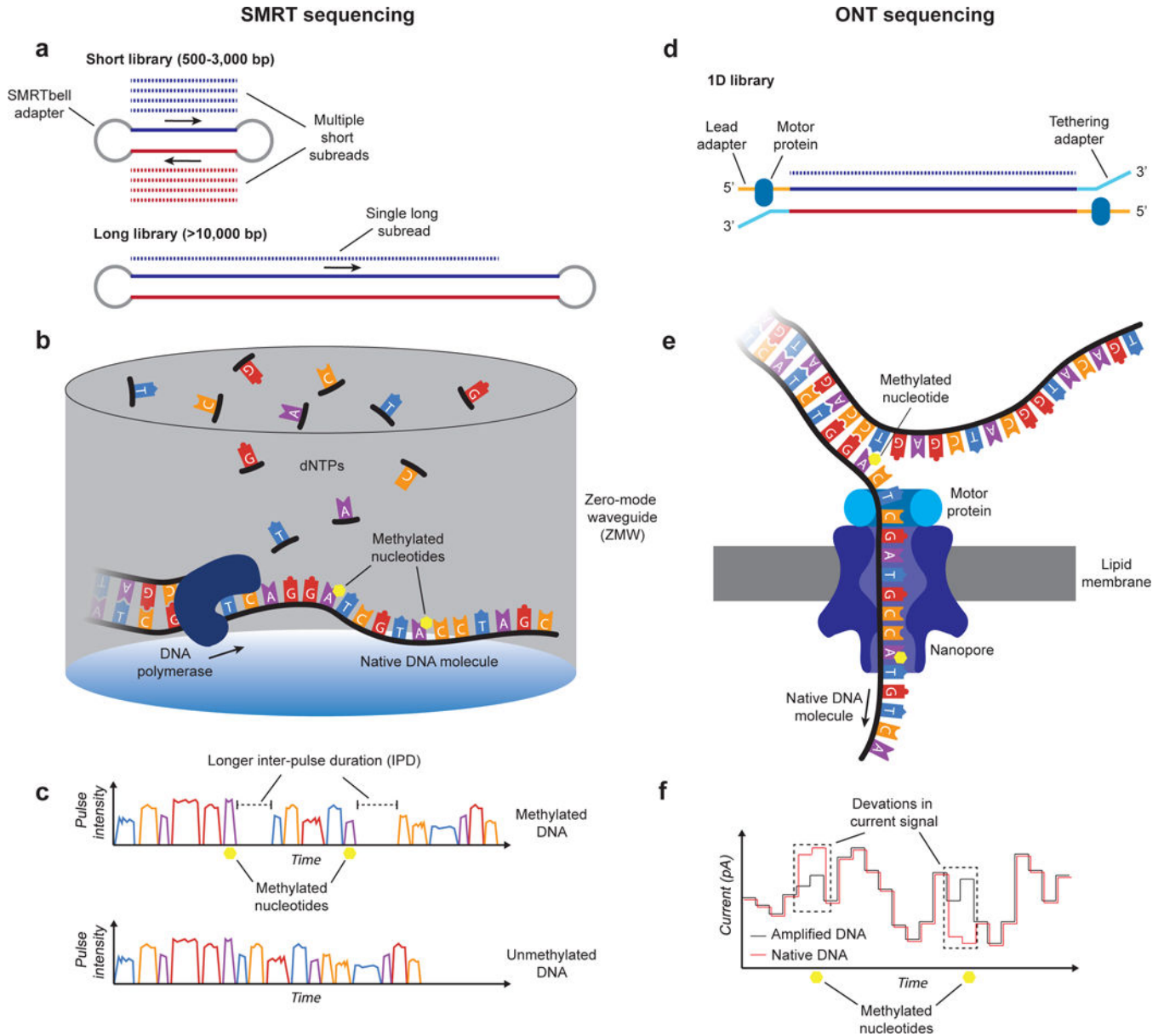of the primary features of a bacterial methylome.

Figure 2: Technologies for detection of DNA methylation through direct sequencing of native DNA molecules.

(a-c) SMRT sequencing. **(d-f)** ONT sequencing. **(a)** Sequencing libraries for single-molecule, real-time (SMRT) sequencing from Pacific Biosciences libraries consist of double-stranded DNA fragments flanked by the hairpin SMRTbell adapters that permit the polymerase to process through both strands of the template. The libraries can take on various configurations depending upon the requirements of the application. Short insert libraries generate multiple subreads from both strands of the template molecule (useful for generating higher accuracy consensus subreads), while long insert libraries are used to generate the longest subread lengths (critical for *de novo* assembly and detection of structural variants). **(b)** SMRT sequencing relies on a sequencing-by-synthesis approach. A DNA polymerase is bound within a zeptoliter-scale observation chamber (called a zero-mode waveguide, or

ZMW) and uses a strand from the native sequencing library as a template for the read, incorporating fluorescently labeled deoxyribonucleoside triphosphates (dNTPs) as they diffuse into the ZMW. Each incorporated dNTP is briefly immobilized at the polymerase active site, emitting a fluorescent pulse in the corresponding color channel. **(c)** When observing the fluorescent traces produced by each ZMW, which are highly multiplexed on a chip, the order of pulses provides the read sequence, while pauses between pulses indicate the presence of a covalent modification in the template DNA. **(d)** The 1D library preparation from Oxford Nanopore Technologies (ONT) use a lead adapter (loaded with a motor protein) and a tethering adapter, which helps co-locate the molecule near the nanopore, to enable the sequencing of a single DNA strand from the molecule[44]. **(e)** ONT sequencing instruments that rely on engineered biological nanopores embedded in a lipid membrane to sequence ssDNA. While a voltage potential is applied across the membrane, ssDNA is ratcheted through the nanopore by a motor protein bound to the DNA library molecule **(f)**. The ionic current flowing through the nanopore depends on the precise set of nucleotides (k = 4 or 5) occupying the constriction point. Methylated nucleotides in the ssDNA introduce distinct current patterns, making it possible to distinguish modified bases relative to amplified (methylation-free) DNA or precomputed models.
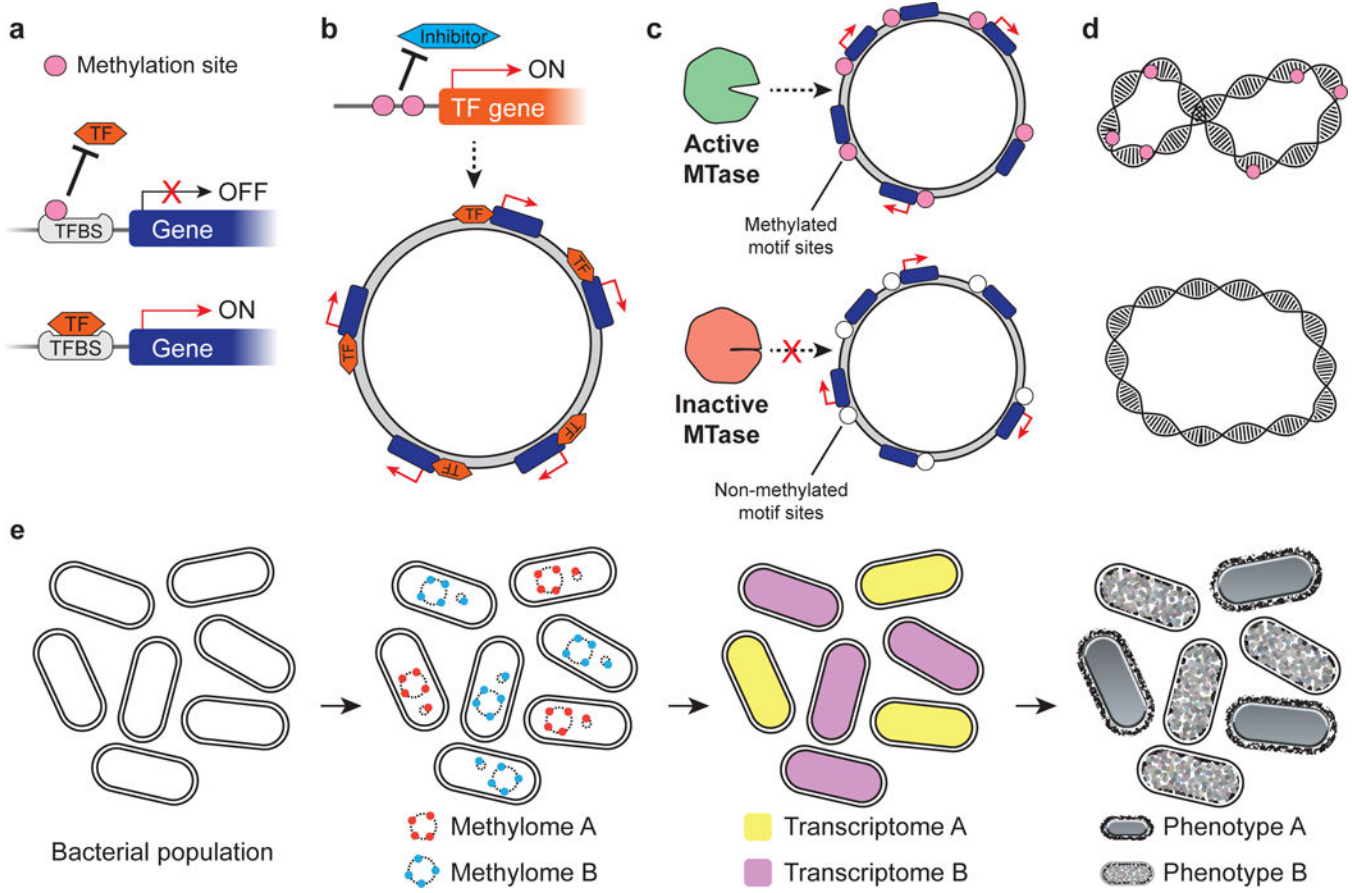
Figure 3: Epigenetic mechanisms of gene regulation and its consequences.

(a) Transcription of the *agn43* gene and *pap* gene cluster in *E. coli* serve as canonical examples of gene expression being regulated according to the methylation status at motif sites within its upstream regulatory sequence. The presence of methylated bases in this region can interfere with the binding of regulatory proteins, leading to either up- or down-regulation of the gene. For instance, methylation can prevent a transcription factor (TF) from binding to its transcription factor binding site (TFBS), thereby preventing transcription of the downstream gene. **(b)** If the gene affected by methylation status encodes a transcription factor, or another protein with promiscuous DNA-binding specificity, the local methylation status can potentially trigger a cascade of downstream changes on gene expression. **(c)** Some bacteria are capable of inducing genome-wide changes in methylation status and gene expression through phase- variable MTases. Spontaneous and reversible frameshift mutations in the MTase gene lead to a clonally expanded bacterial population with divergent methylation activity and distinct gene expression regimes. **(d)** DNA methylation is likely to be involved in alternative mechanisms of gene regulation. For example, methylation is known to affect the curvature of DNA molecules, which could potentially control which regions of a chromosome are exposed to the transcriptional machinery of the cell. **(e)** The presence of phase-variable methyltransferases can introduce heterogeneous methylation patterns in a clonally expanded bacterial population, leading to subpopulations with distinct gene expression regimes and phenotypes.

Table 1:

Summary of available methods for detecting modified DNA in prokaryotes at single-nucleotide resolution.

Technique	Modification types	AdvAdantages	Limitations	Notes	Refs
Bisulfite sequencing	5mC, 5hmC, 4mC	Gold standard for detection of 5mC modifications	• Bisulfite treatment fragments library DNA molecules • 4mC detection requires additional TET conversion • Cannot detect 6mA	-	59,60,181
Restriction enzyme digest followed by next generation sequencing	4mC, 5mC, 6 mA	• Gold standard approach for experimental validation of methylated sites • Highly sensitive • Works well with limited input DNA	Limited choice of restriction enzymes with known target sequence specificity	-	101
SMRT sequencing	4mC, 5mC, 5hmC, 6mA	• LoLong reads enable phasing of multiple methylated positions • Reveals both methylated motifs and methylation sites at single nucleotide and single molecule resolution	5mC detection requires TET conversion or very deep sequencing coverage	Enabled the mapping of the first complete bacterial epigenomes (>1,200 mapped to date)	16,25,39,40,62
Nanopore sequencing	4mC, 5mC, 5hmC, 6mA	Long reads hold promise of phasing multiple methylated positions	Accurate modification detection complicated by noisy current signal	Modification detection algorithms under active development	72,74–77