

Whole genome DNA methylation: beyond genes silencing

Roberto Tirado-Magallanes^{1,2}, Khadija Rebbani¹, Ricky Lim¹, Sriharsa Pradhan³, Touati Benoukraf¹

¹Cancer Science Institute of Singapore, National University of Singapore, 117599 Singapore, Singapore

²Computational Systems Biology Team, Institut de Biologie de l'École Normale Supérieure (IBENS), INSERM, École Normale Supérieure, PSL Research University, 75005 Paris, France

³New England Biolab Inc., Ipswich, MA 01938, USA

Correspondence to: Touati Benoukraf, **email:** benoukraf@nus.edu.sg

Keywords: DNA methylation, bisulfite sequencing, gene regulation, chromatin modeling, imprinting

Received: August 11, 2016 **Accepted:** November 07, 2016 **Published:** November 24, 2016

ABSTRACT

The combination of DNA bisulfite treatment with high-throughput sequencing technologies has enabled investigation of genome-wide DNA methylation at near base pair level resolution, far beyond that of the kilobase-long canonical CpG islands that initially revealed the biological relevance of this covalent DNA modification. The latest high-resolution studies have revealed a role for very punctual DNA methylation in chromatin plasticity, gene regulation and splicing. Here, we aim to outline the major biological consequences of DNA methylation recently discovered. We also discuss the necessity of tuning DNA methylation resolution into an adequate scale to ease the integration of the methylome information with other chromatin features and transcription events such as gene expression, nucleosome positioning, transcription factors binding dynamic, gene splicing and genomic imprinting. Finally, our review sheds light on DNA methylation heterogeneity in cell population and the different approaches used for its assessment, including the contribution of single cell DNA analysis technology.

INTRODUCTION

DNA methylation is a heritable covalent chemical modification of DNA, crucial for numerous biological processes such as gene regulation, cell fate decisions and disease development [1]. Of particular note is the use of DNA methylation inhibitors as powerful therapeutic agents in the treatment of myelodysplastic syndrome and, with lesser success, of solid tumors [2, 3]. The detailed molecular mechanisms underlying the effects of DNA methylation on chromatin folding and accessibility as well as on gene regulation remain poorly understood. The dogma portraying the function of DNA methylation as an inhibitor of gene expression is still pervasive. Global and unbiased DNA methylation analysis protocols such as whole genome bisulfite sequencing, empowered by the advent of high-throughput sequencing, revealed a more sophisticated role of DNA methylation, with gene silencing representing only one facet of its consequences. Increasing evidence suggests that DNA methylation is not only associated with gene repression, but also with gene activation [4], splicing

regulation [5], nucleosomes positioning [6–8], and the recruitment of transcription factors [9–12]. Together, this multiplicity of functions suggests that DNA methylation is more accurately described as a process akin to a cellular epigenetic memory [13–16]. DNA methylation is widely analyzed in the CpG context, due to the fact that 80 % of methylation events occur at CpG sites in mammalian genomes. However, in plants, only 24 % of CpG sites are methylated, while 6.7 % and 1.7 % of CHG and CHH (where H = A, T or G) are methylated, respectively [17]. The symmetry properties of CpG and CHG motifs imply a double-stranded DNA methylation, whereas methylation of the asymmetric CHH motif refers to single-stranded DNA methylation. The functions of CHG and CHH methylation are still unclear. Recent studies revealed that in *Arabidopsis*, CHH methylation occurs predominantly at the transposable elements, and has been involved in epigenetic inheritance [18], as well as in the prevention of transposon jumping during development [19]. In mammals, non-CpG methylation is detected at perceptible levels only in a few cell types, including neuronal cells during brain development and after neurogenesis [20]; in

embryonic stem cells and induced pluripotent stem cells [21]. Non-CpG methylation may thus also play a role in X-chromosome inactivation. Furthermore, the striking correlation between methylation patterns in CHG and CHH contexts and human cells suggests that methylation in these contexts might be maintained by a common machinery, in contrast to plants [21]. Furthermore, asymmetric DNA methylation is apparently enriched in the transposable elements SINEs and LINEs, but not for LTRs. So far, because of technological limitations and of the early dogma limiting DNA methylation to its sole gene silencing function, numerous studies have mainly focused on the promoter regions (particularly CpG island promoters). These biases are still evident in current microarray designs aiming at deciphering the state of genome-wide DNA methylation. When correlating DNA methylation status and gene transcription levels, CpG methylation scores are usually averaged throughout the promoters or CpG islands, resulting in a robust estimation of differential DNA methylation call on large regions at the expense of diluting the methylation signal on local loci. Sequencing-based technologies resulted in a dramatic

increase in resolution from CpG islands (CGI) level down to the level of single cytosine in the CpG, CHG or CHH contexts, shedding a new light on the various biological functions of DNA methylation. These discoveries were made possible by the development of computational strategies not restricted to promoter methylation/gene expression associations [22]. Methodologies used in the recent studies have revealed the necessity of tuning DNA methylation resolution into an adequate scale to ease the integration of DNA methylation information with other chromatin features. This review covers current major DNA methylation analytical schemes. In particular, we discuss optimal resolution choices associated with the study of each aspect of DNA methylation-related biological processes (Figure 1).

DNA methylation and gene expression

Early perception on the function of DNA methylation has been linked to gene expression [23, 24]. In the 80's, it was reported that promoters hypermethylation correlates with decreased expression levels of downstream

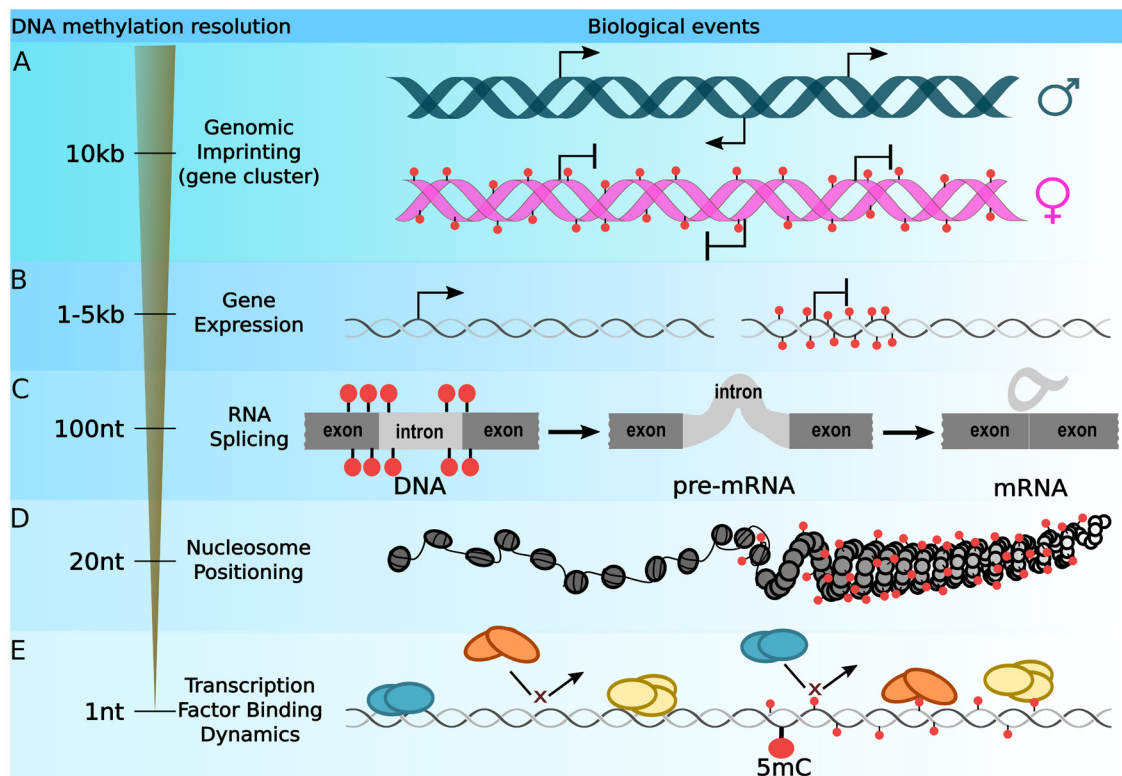


Figure 1: Optimizing the DNA methylation resolution according to a biological context. The most widely used strategy for integrating DNA methylation with gene regulation is to average CpG methylation signal throughout wide loci. **A.** The study of the imprinting regions can be achieved by averaging DNA methylation signal of loci ranging from 1 kb to 10 kb. **B.** Interplay between gene expression and DNA methylation are usually drawn by studying the DNA methylation level within the 1 kb to 5kb region surrounding the TSS. **C.** However, DNA methylation is involved in many other mechanisms. Extending the DNA methylation resolution to a 100 bp around splicing sites enables the investigation of exon inclusion. **D.** On the other hand, a 20 bp resolution was established to be optimal for studying the interplay between DNA methylation and nucleosome positioning. **E.** Finally, DNA methylation plays a key role in the recruitment of transcription factor and it has been shown that methylation of a single cytosine can affect protein/DNA binding affinity.

genes, such as γ -globin locus [25]. Most of the studies used Southern blot techniques then, which allowed measurement of DNA methylation at the resolution of about 1 kb. These promoter-centric studies substantiated the dogma associating DNA methylation with gene repression. Therefore, as approximately 70 % of annotated promoters overlap a CGI [26], biotechnology companies have developed microarrays containing probes that target preferentially gene promoters [27]. Recent genome-wide studies have demonstrated the soundness of the dogma and succeeded in characterizing genes directly inactivated by DNA methylation [28, 29]. Some studies focus mainly on the gene promoter analysis using various resolution levels. For instance, using the reduced representation of bisulfite sequencing (RRBS) method, which allows interrogation of about 30 % of CpG sites that overlap 65 % of human genome promoters [30], Amabile *et al.* have identified around 500 genes involved in chronic myeloid leukemia (CML) progression by analyzing DNA methylation on promoters [31]. This CML epigenetic signature is characterized by averaging methylation levels of CpG sites located in regions ranging from -1.5 kb to +0.5 kb (i.e. 2 kb-window) around the transcription start sites (TSS). This example shows how broad DNA methylation can change, particularly within the gene promoter regions, which determine the phenotype. With another strategy, Hodges *et al.* [28] studied DNA methylation dynamics during the hematopoietic development using the whole genome bisulfite sequencing method (WGBS) [32]. A 100 bp sliding window was applied to average out CpG methylation levels, which allows comparison of DNA methylation patterns across all TSS regions, regardless of the variations in CpG sites distributions throughout those loci. This strategy permitted the characterization of the typical TSS region (± 4 kb) methylation pattern. They reported that DNA methylation level within the ± 1 kb region surrounding the TSS showed the greatest correlation with gene repression. Notably, an analysis using GBSA (Genome Bisulfite Sequencing Analyser) [33] substantiated this observation. Interestingly, DNA methylation/gene repression correlation was evident at 1 kb downstream of the TSS of the genes. This observation corroborates another study where DNA methylation of the first exon is shown to be associated with transcriptional gene repression [29]. Refining DNA methylation analysis to 100 bp resolution demonstrated that DNA methylation patterns surrounding TSS are not homogeneous, suggesting that once methylated, some DNA regions, including the first exon, are subjected to gene repression more than the others, and might play a specific role in gene inactivation. In fact, the first exon-intron region of active tissue-specific genes was found to be enriched in the dimethylation of lysine 4 at histone 3 (H3K4me2), which is a predominant signature of gene regulatory elements [34].

Paradoxically, DNA methylation is also associated with gene activation, when it occurs within the transcribed

regions [35]. Apparently, gene body methylation enhances transcription. Similarly, in an *in vitro*-induced differentiation study of human embryonic stem cells, a large group of 3' CGI that underwent an increase in DNA methylation actually correlated with increased expression of these genes [36]. These relationships exhibit the multifaceted and complexity of DNA methylation roles in gene regulation and the importance of the genome structure integration.

DNA methylation dictates nucleosome positioning

An obvious mechanism in which DNA methylation participates in gene regulation is by nucleosome positioning, i.e. restricting the accessibility for protein complexes to DNA regulatory regions (such as gene promoters or TSS). An early depiction of the involvement of DNA methylation in nucleosome positioning was described in a study whereby DNA methylation in a 3x(CpG) element was targeted, leading to the depletion of a neighboring nucleosome [37]. Although it has been reported that methylation of DNA increases its stiffness which might alter nucleosomal formation [8, 38, 39], numerous studies based on whole-genome approaches have exhibited enrichment in methylated cytosines on the histone-bound DNA sequence. In one report, a 10 bp interval of methylated and unmethylated CpGs in the nucleosomal DNA was observed [6], showing that unmethylated CpG dinucleotides occur principally in the minor grooves facing away from the histone octamer, whereas the methylated counterpart is mostly seen in the minor groove in proximity to the octamer complex promoting nucleosome stability [6, 40]. A recent analysis has demonstrated that methylation of CpGs in the major groove of the DNA wrapped around the histone octamer greatly influences the nucleosome dynamics towards a more open structure, while the methylation state of CpGs located in the minor grooves has a negligible effect [8]. Interestingly, the nucleosome occupancy within exons correlates with the local CG density [41].

To date, NOME-Seq is the most robust method to study the relationship between nucleosome positions and DNA methylation status simultaneously in a genome [7]. This sequencing-based method was used to investigate nucleosome structure and DNA methylation at CGIs in oncogenes, by combining the usage of a GpC methyltransferase (M.CviPI) to obtain nucleosome positioning information based on enzyme accessibility to GpC sites, and bisulfite DNA treatment to determine the methylation status of cytosines [42]. A prominent anti-correlation between DNA methylation and nucleosome occupancy at CTCF binding regions was revealed in different cell lines by averaging NOME-Seq signal using a 20 bp sliding window. However, this anti-correlation was not observed at gene promoters. It was reported that

the 20 bp resolution is optimum as this is the average distance between two adjacent CpG dinucleotides. In subsequent studies, by using NOME-Seq and examining the profiles with 100 bp windows at 20 bp spacing, the interplay of DNA methylation and nucleosome occupancy at the enhancer regions of cancer cell lines was determined [43, 44]. An additional study observed the aggregate profile of 5mC coming from bisulfite-sequencing around CTCF binding sites as well as their MNase-Seq profiles on different cell lines using a single base pair resolution (1 bp) [45]. Their results were strikingly similar to the ones reported in the initial NOME-Seq study, further supporting that a 20 bp resolution is adequate to resolve the nucleosome positioning/DNA methylation interplay and that using a narrower resolution will yield similar results.

DNA methylation and transcription factor binding dynamics

Transcription factors (TFs) are key effectors in the activation of transcription. Their functions rely on binding to the DNA upon recognition of a particular nucleotide motif through steric interactions between the TF protein domains and the DNA molecule. It has been long known that chemical modifications to the DNA bases can either increase or restrain these interactions [46]. DNA methylation in the vicinity of TSS region is a common proxy for the transcriptional status of a gene. It has been reported that the methylation status of only 16 % of CpG sites surrounding the TSS correlates negatively with the expression of their corresponding genes [47]. Moreover, these CpG sites are generally avoided at predicted TF binding sites, especially at the binding sites of known TF with a repressive function. It is often assumed that the reason behind this negative correlation is the difficulty of TFs in binding to methylated cytosines. For instance, the abrogation of DNA methylation in murine stem cells by knocking out DNA methyltransferases DNMT1, DNMT3A and DNMT3B increases the amount of NRF1 binding events dramatically [48]. However, the relationship between DNA methylation and TF binding is complex and is often dependent on cell signaling and post-translational modifications of TF. By investigating the mechanism of interaction using a high-throughput array-based technology, Hu *et al.* evaluated the effect of single cytosine methylation on 154 TF DNA binding motifs containing at least one CpG site [11]. This study revealed that, depending on the TF, DNA methylation could either hamper or enhance TF/DNA interactions. Furthermore, some TFs showed the ability to bind different motifs depending on their methylation status. Likewise, previous studies have reported that DNA methylation in the close vicinity of TF binding sites, which do not contain CpG site, might also alter the strength of TF/DNA interaction. This is the case for AP-1, a TF complex composed of

cFOS and cJUN that recognizes the TGANTCA motif. In fact, this complex appears to lose its ability to bind DNA when a CpG site adjacent to the core-binding motif is methylated [10]. Similarly, although methylation within the consensus Sp1-binding site does affect Sp1/Sp3 binding, methylation adjacent to the core Sp1 motif induces a significant decrease in Sp1/Sp3 binding [9]. Interestingly, this phenomenon also occurs in an allele-specific manner. For instance, YY1 binding events are modulated by DNA methylation in a parent-of-origin-specific fashion, in such a way that only CpGs close to the binding site of the maternal allele are methylated, preventing YY1 to bind only the maternal allele [49].

Furthermore, it is worth mentioning that enzymatic modification of cytosine is a complex dynamic involving DNMT1, DNMT3A and DNMT3B methyltransferases, which methylates cytosines (5mC), and the TET family of cytosine oxygenase enzymes, which oxidizes 5mC to 5-hydroxymethylcytosine (5hmC), subsequently to 5-formylcytosine (5fC) and finally to 5-carboxycytosine (5caC) [50, 51]. These oxidized derivatives might also hinder TF binding. In principle, the presence of these derivatives can alter the way in which proteins bind to their recognition sequences in DNA by strengthening the interactions, weakening them, or by abolishing them completely. For example, Klf4 shows the strongest binding to fully methylated DNA, with slightly higher affinity (approximately 1.5-fold) than that of the unmodified DNA, and in each oxidation event, from 5mC to 5hmC to 5fC to 5caC, resulting in progressively weaker binding (by factors of ~2, 3, and 6, respectively) [52].

The function of these oxidized derivatives of 5mC is still under discussion and many consider these changes as a transitory process leading to DNA demethylation [53]. However, recent observations revealed that 5hmC is enriched in the short interspersed elements (SINEs) and long terminal repeat (LTR) regions, while 5fC and 5caC seemed to be more prevalent within the satellite repeats regions of the genome [54]; and that the DNA binding affinity of numerous proteins increases in the presence of these oxidized derivatives of 5mC [55, 56], suggesting that 5hmC, 5fC and 5caC possess functional epigenetic roles. Nonetheless, a significant drawback of WGBS is the inability to distinguish 5mC from 5hmC, because sodium bisulfite treatment is unable to convert both methylated states to uracil [57]. This frequently neglected limitation should be taken into account when interpreting bisulfite-converted data.

DNA methylation and gene splicing

The majority of eukaryotic genes give rise to several isoforms. Aberrant splicing can lead to extreme phenotypes such as spinal muscular atrophy [58], suggesting a tight regulation during exons selection. Studies of the splicing process began about 40 years ago, although many of the

mechanisms and signals controlling are still being actively investigated. It has been shown recently that gene body DNA methylation controls gene transcription and exon splicing [59]. Using methylation array technologies, a study showed that tissue-specific differentially methylated regions in mouse are preferentially located in exons and introns of protein coding genes with known alternative splicing variants [60]. Many of the splicing events in the human genome occur co-transcriptionally while the precursor mRNA remains associated with the chromatin until the introns are excised [61, 62]. This is related to findings suggesting that chromatin marks and structures provide the signals for exons selection [63]. Remarkably, genes with similar GC content within exons and introns exhibit significant decrease in CpG methylation levels of the 100 bp region towards the introns from the exon-intron junctions, compared to the rest of the intron; and a striking enrichment of DNA methylation levels in the 20 bp region surrounding the 5' and 3' splice sites [64].

Recent findings highlighted the essentiality of DNA methylation status in the recruitment of protein factors responsible for splicing signals such as CTCF and MeCP2 [59]. Depending on the context, DNA hypermethylation prevents CTCF from binding to the transcribed region, accelerating the processivity of Pol II which is translated into a higher frequency of exon exclusion [5]. In contrast, CpG methylation increases MeCP2 binding affinity to DNA, leading to the recruitment of histone deacetylases, decreasing Pol II activity and enhancing exon inclusion successively [65].

DNA methylation regulates genomic imprinting

Imprinting refers to the targeted inactivation of a genomic region in one of the parental chromosomes. The implications of imprinting in intergenerational inheritance are currently under intense investigation due to its medical implications, such as metabolic changes as a consequence of the diet of an ancestor. There is also emerging evidence showing DNA methylation as one of the major players in this biological process [66]. DNA methylation can establish differential parental methylation during gametogenesis *via de novo* methyltransferase activity (DNMT3A/B) in the testes or ovaries. DNA methylation can be inherited to the next generation by maintaining the methylation status with the help of DNMT1 after gametogenesis and also during subsequent embryonic cell divisions. Demethylation can occur passively after DNA replication if the mark is not maintained. Nonetheless, the active process of demethylation remains largely unveiled and is still being investigated.

Imprinted genes are generally clustered in DNA regions ranging from 100 kb to 3700 kb [67]. These imprinted loci can contain 3 to 12 genes. Strategies which consist of scanning methylC-seq datasets with a

resolution of 5 kb have succeeded in identifying known and several novel imprinting regions [68]. Most of the allele-specific DNA methylation (ASM) identification methods segregate alleles by genotypic variations (e.g. SNP). However, not all imprinting regions contain such variations. To overcome this limitation, Fang *et al.* have recently developed a statistical tool to identify ASM from bisulfite sequencing dataset based on the assumption that the methylation signal derived from sequenced reads can be de-convoluted into two distinct patterns [69]. In sequential order, this tool fits signal from the bisulfite sequencing data with a single-allele model, followed by an allele-specific model, and determines which model provides the best fit to the data using the Bayesian Information Criteria (BIC) method. Furthermore, they implemented an Expectation Maximum (EM) algorithm to infer the probabilities of the allele of origin for each methylation read. Using a resolution of 10 CpGs and 20 bp, this strategy allows identification of ASM from a wide range of human cell types. Although differentially methylated alleles in SNP regions in the genome have been effectively used to infer ASM, this strategy is based on the assumption that DNA methylation patterns remain consistent during cell division, and therefore is limited to the analysis of homogeneous cell populations.

DNA methylation heterogeneity patterns in cell population

Managing cellular heterogeneity information, particularly within tumor samples, represents one of the greatest challenges in (epi)genomic analysis. Promising statistical algorithms that can identify DMRs from individual tumor methylome samples without genomic variation information or prior knowledge from other datasets are becoming crucial tools for the efficient analysis of DNA methylation signal derived from heterogeneous cell populations. Software packages such as MethylPurify [70] allow segregation of DNA methylation patterns using complex deconvolution analysis in regions with bisulfite reads showing discordant methylation levels. Similarly, DMEAS [71, 72] uses a Shannon entropy model to assess the DNA methylation heterogeneity within DNA loci consisting of at least 4 CpG sites sequenced within the same reads. These methods require a DNA methylation signal at the nucleotide resolution and allow inferring of cell heterogeneity levels (i.e. tumor purity from tumor samples alone), which permit better characterization of DMR across samples. However, the increasing length of reads generated by the latest sequencers (> 20 kb) will allow the combination of DNA methylation patterns with SNP information and improve single molecule-level detection as well as phasing of DNA methylation.

Towards single cell methylome analysis

Methylome analysis of heterogeneous cell populations such as tumor samples, which are known to bear many different cell subtypes; or during developmental and differentiation processes, where cells are unlikely to differentiate in a synchronous manner, remains challenging due to the difficulty in deconvoluting DNA methylation signals of cell sub-populations. A strategy to overcome this obstacle is to measure the methylome at the single cell level [73]. However, bisulfite treatment degrades a non-negligible portion of DNA, and this situation worsens in single cell experiments due to the small quantity of DNA starting material. Indeed, bisulfite treatment in single cell protocols results in the loss of vast portions of the genome, which is translated into poor mapping efficiencies (of about 20 % or total library size) covering between 18 % and 50 % of all the CpGs in the human genome [74]. Nonetheless, despite a reduced mapping efficiency, a striking correlation was observed when looking at the methylation levels between single cell samples and population samples at 2-kb windows (average $R = 0.95$), suggesting that DNA methylation levels are consistent between different samples. Although single cell methylome analysis protocols have to be optimized, this method has already revealed interesting novel epigenetic mechanisms. For instance, by combining single cell genotyping, gene expression and DNA methylation profiles in a single assay, a research group was able to identify the DNA methylation changes driven by mutations in the EGFR gene from primary tumor samples that would have been otherwise masked by the heterogeneity of the tissue [75].

Conclusion and perspectives

In recent years, whole genome methylation assays have gained recognition in clinical and biomedical research. Working with DNA has more advantages over RNA assays, not only due to the importance of DNA methylation in genome regulation, but also because the former is more stable, thereby simplifying sample collection, processing, transport, and storage. A second advantage is that when performing assays in heterogeneous cell populations, the amount of DNA extracted is proportional to the cell count. On the contrary, for RNA, a small subpopulation of cells with high transcription rates can mask the profiles of the others. Such is the case of transcriptome assays on tumors infiltrated by immune cells, where the RNA profiles arise mostly from T-cells or macrophages rather than neoplastic cells. However, the spectrum of the roles of DNA methylation in biological processes is far from fully drawn. Indeed, as opposed to previous conceptions, DNA methylation is not a “molecular lock” that prevents gene expression, but rather a complex feature that affects many genomic features dynamically [13, 15, 16]. A

myriad of sequencing-based approaches to investigate the methylation status of the cells have been developed [76], but the bottle-neck remains in the downstream analysis and data interpretation. One of the major limitations lies in the sample preparation and the use of bisulfite reagent for deamination. Although sodium bisulfite-based sequencing accurately determines DNA methylation, it cannot distinguish between 5mC and 5hmC. Furthermore, degraded DNA and deaminated adducts are poor products for library construction and hence are not ideal for accurate quantification, particularly for single cell experiments. However, technologies for single molecule DNA sequencing for mammalian epigenome analysis are being developed.

During the recent years, DNA methylation in the CpG context has been shown to act on numerous biological processes, including tumorigenesis, depending on the 5mC density and genomic location. Nonetheless, various investigations have to be performed to explore the role of CHG and CHH methylation in mammalian cells.

To date, the main leverage to maximize biological knowledge extraction from methylome datasets derived from bisulfite sequencing methods is the optimization of DNA methylation resolution measurement according to a specific biological process of interest.

ACKNOWLEDGMENTS

The authors thank Denis Thieffry, Samuel Collombet and Nicolas Bertin for useful discussions and comments during the preparation of this manuscript, as well as Celestina Chin Ai Qi for her proofreading.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

GRANT SUPPORT

Work in T.B. laboratory is supported by the National Research Foundation, the Singapore Ministry of Education under its Centres of Excellence initiative, the National Medical Research Council of Singapore (Grant # NMRC/BNIG/2035/2015) and the Institut Français à Singapour (Merlion Project Grant # 6.10.14). S.P. laboratory is supported by basic research funding from New England Biolabs, Inc.

REFERENCES

1. Franchini D-M, Schmitz K-M, Petersen-Mahrt SK. 5-methylcytosine DNA demethylation: more than losing a methyl group. *Annu Rev Genet.* 2012; 46: 419–41. doi: 10.1146/annurev-genet-110711-155451.

2. Li H, Chiappinelli KB, Guzzetta AA, Easwaran H, Yen R-WC, Vataapalli R, Topper MJ, Luo J, Connolly RM, Azad NS, Stearns V, Pardoll DM, Davidson N, et al. Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers. *Oncotarget*. 2014; 5: 587–98. doi: 10.18632/oncotarget.1782.
3. Tsai H-C, Li H, Van Neste L, Cai Y, Robert C, Rassool F V, Shin JJ, Harbom KM, Beatty R, Pappou E, Harris J, Yen R-WC, Ahuja N, et al. Transient low doses of DNA-demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells. *Cancer Cell*. 2012; 21: 430–46. doi: 10.1016/j.ccr.2011.12.029.
4. Bahar Halpern K, Vana T, Walker MD. Paradoxical role of DNA methylation in activation of FoxA2 gene expression during endoderm development. *J Biol Chem*. 2014; 289: 23882–92. doi: 10.1074/jbc.M114.573469.
5. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*. 2011; 479: 74–9. doi: 10.1038/nature10442.
6. Chodavarapu RK, Feng S, Bernatavichute Y V, Chen P-Y, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, Casero D, Bernal M, Huijser P, et al. Relationship between nucleosome positioning and DNA methylation. *Nature*. 2010; 466: 388–92. doi: 10.1038/nature09147.
7. Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res*. 2012; 22: 2497–506. doi: 10.1101/gr.143008.112.
8. Jimenez-Useche I, Ke J, Tian Y, Shim D, Howell SC, Qiu X, Yuan C. DNA Methylation Regulated Nucleosome Dynamics. *Sci Rep*. 2013; 3: 1–5. doi: 10.1038/srep02121.
9. Zhu W-G, Srinivasan K, Dai Z, Duan W, Druhan LJ, Ding H, Yee L, Villalona-Calero MA, Plass C, Otterson GA. Methylation of adjacent CpG sites affects Sp1/Sp3 binding and activity in the p21(Cip1) promoter. *Mol Cell Biol*. 2003; 23: 4056–65. doi: 10.1128/MCB.23.12.4056.
10. Fujimoto M, Kitazawa R, Maeda S, Kitazawa S. Methylation adjacent to negatively regulating AP-1 site reactivates TrkA gene expression during cancer progression. *Oncogene*. 2005; 24: 5108–18. doi: 10.1038/sj.onc.1208697.
11. Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, Shin J, Cox E, Rho HS, Woodard C, Xia S, Liu S, Lyu H, et al. DNA methylation presents distinct binding sites for human transcription factors. *Elife*. 2013; 2: 1–16. doi: 10.7554/eLife.00726.
12. Breiling A, Lyko F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin*. *BioMed Central*; 2015; 8: 24. doi: 10.1186/s13072-015-0016-6.
13. Day JJ, Sweatt JD. DNA methylation and memory formation. *Nat Neurosci*. 2010; 13: 1319–23. doi: 10.1038/nn.2666.
14. Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, Kim J, Aryee MJ, Ji H, Ehrlich LIR, Yabuuchi A, Takeuchi A, Cunniff KC, et al. Epigenetic memory in induced pluripotent stem cells. *Nature*. 2010; 467: 285–90. doi: 10.1038/nature09342.
15. Raynal NJM, Si J, Taby RF, Gharibyan V, Ahmed S, Jelinek J, Estéicio MRH, Issa JPJ. DNA methylation does not stably lock gene expression but instead serves as a molecular mark for gene silencing memory. *Cancer Res*. 2012; 72: 1170–81. doi: 10.1158/0008-5472.CAN-11-3248.
16. Yoshida K, Maekawa T, Zhu Y, Renard-Guillet C, Chatton B, Inoue K, Uchiyama T, Ishibashi K-I, Yamada T, Ohno N, Shirahige K, Okada-Hatakeyama M, Ishii S. The transcription factor ATF7 mediates lipopolysaccharide-induced epigenetic changes in macrophages involved in innate immunological memory. *Nat Immunol*. 2015; 16: 1034–43. doi: 10.1038/ni.3257.
17. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*. 2010; 11: 204–20. doi: 10.1038/nrg2719.
18. Calarco JP, Borges F, Donoghue MTA, Van Ex F, Jullien PE, Lopes T, Gardner R, Berger F, Feijó JA, Becker JD, Martienssen RA. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell*. 2012; 151: 194–205. doi: 10.1016/j.cell.2012.09.001.
19. Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, Martienssen R a. miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. *Nature*. 2014; 508: 411–5. doi: 10.1038/nature13069.
20. Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, Yu M, Tonti-Filippini J, Heyn H, et al. Global epigenomic reconfiguration during mammalian brain development. *Science*. 2013; 341: 1237905. doi: 10.1126/science.1237905.
21. Guo W, Chung W-Y, Qian M, Pellegrini M, Zhang MQ. Characterizing the strand-specific distribution of non-CpG methylation in human pluripotent cells. *Nucleic Acids Res*. 2014; 42: 3009–16. doi: 10.1093/nar/gkt1306.
22. Adusumalli S, Omar MFM, Soong R, Benoukraf T. Methodological aspects of whole-genome bisulfite sequencing analysis. *Brief Bioinform*. 2014; 16: 369–79. doi: 10.1093/bib/bbu016.
23. Yagi M, Koshland ME. Expression of the J chain gene during B cell differentiation is inversely correlated with DNA methylation. *Proc Natl Acad Sci U S A*. 1981; 78: 4907–11. doi: 10.1073/pnas.78.8.4907.
24. Ehrlich M, Wang RY. 5-Methylcytosine in eukaryotic DNA. *Science*. 1981; 212: 1350–7. doi: 10.1126/science.6262918.
25. Busslinger M, Hurst J, Flavell RA. DNA methylation and the regulation of globin gene expression. *Cell*. 1983; 34: 197–206. doi: 0092-8674(83)90150-2.
26. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two

- distinct classes of promoters. *Proc Natl Acad Sci U S A*. 2006; 103: 1412–7. doi: 10.1073/pnas.0510310103.
27. Huang Y-W, Huang TH-M, Wang L-S. Profiling DNA methylomes from microarray to genome-scale sequencing. *Technol Cancer Res Treat*. 2010; 9: 139–47.
 28. Hodges E, Molaro A, Dos Santos CO, Thekkat P, Song Q, Uren PJ, Park J, Butler J, Rafii S, McCombie WR, Smith AD, Hannon GJ. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol Cell*. 2011; 44: 17–28. doi: 10.1016/j.molcel.2011.08.026.
 29. Brenet F, Moh M, Funk P, Feierstein E, Viale AJ, Socci ND, Scandura JM. DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS One*. 2011; 6: e14524. doi: 10.1371/journal.pone.0014524.
 30. Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc*. 2011; 6: 468–81. doi: 10.1038/nprot.2010.190.
 31. Amabile G, Di Ruscio A, Müller F, Welner RS, Yang H, Ebralidze AK, Zhang H, Levantini E, Qi L, Martinelli G, Brummelkamp T, Le Beau MM, Figueroa ME, et al. Dissecting the role of aberrant DNA methylation in human leukaemia. *Nat Commun*. 2015; 6: 7091. doi: 10.1038/ncomms8091.
 32. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, Edsall L, Antosiewicz-Bourget J, Stewart R, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009; 462: 315–22. doi: 10.1038/nature08514.
 33. Benoukraf T, Wongphayak S, Hadi LHA, Wu M, Soong R. GBSA: A comprehensive software for analysing whole genome bisulfite sequencing data. *Nucleic Acids Res*. 2013; 41. doi: 10.1093/nar/gks1281.
 34. Pekowska A, Benoukraf T, Ferrier P, Spicuglia S. A unique H3K4me2 profile marks tissue-specific gene regulation. *Genome Res*. 2010; 20: 1493–502. doi: 10.1101/gr.109389.110.
 35. Jones PA. The DNA methylation paradox. *Trends Genet*. 1999; 15: 34–7.
 36. Yu D-H, Ware C, Waterland RA, Zhang J, Chen M-H, Gadkari M, Kunde-Ramamoorthy G, Nosavanh LM, Shen L. Developmentally programmed 3' CpG island methylation confers tissue- and cell-type-specific transcriptional activation. *Mol Cell Biol*. 2013; 33: 1845–58. doi: 10.1128/MCB.01124-12.
 37. Davey C, Pennings S, Allan J. CpG methylation remodels chromatin structure *in vitro*. *J Mol Biol*. 1997; 267: 276–88. doi: 10.1006/jmbi.1997.0899.
 38. Nickol J, Behe M, Felsenfeld G. Effect of the B–Z transition in poly(dG-m5dC). poly(dG-m5dC) on nucleosome formation. *Proc Natl Acad Sci U S A*. 1982; 79: 1771–5.
 39. Choy JS, Wei S, Lee JY, Tan S, Chu S, Lee TH. DNA methylation increases nucleosome compaction and rigidity. *J Am Chem Soc*. 2010; 132: 1782–3. doi: 10.1021/ja910264z.
 40. Collings CK, Waddell PJ, Anderson JN. Effects of DNA methylation on nucleosome stability. *Nucleic Acids Res*. 2013; 41: 2918–31. doi: 10.1093/nar/gks893.
 41. Amit M, Donyo M, Hollander D, Goren A, Kim E, Gelfman S, Lev-Maor G, Burstein D, Schwartz S, Postolsky B, Pupko T, Ast G. Differential GC Content between Exons and Introns Establishes Distinct Strategies of Splice-Site Recognition. *Cell Rep*. 2012; 1: 543–56. doi: 10.1016/j.celrep.2012.03.013.
 42. Portela a, Liz J, Nogales V, Setién F, Villanueva A, Esteller M. DNA methylation determines nucleosome occupancy in the 5'-CpG islands of tumor suppressor genes. *Oncogene*. 2013; 32: 5421–8. doi: 10.1038/onc.2013.162.
 43. Taberlay PC, Statham AL, Kelly TK, Clark SJ, Jones PA. Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. *Genome Res*. 2014; 24: 1421–32. doi: 10.1101/gr.163485.113.
 44. Statham AL, Taberlay PC, Kelly TK, Jones PA, Clark SJ. Genome-wide nucleosome occupancy and DNA methylation profiling of four human cell lines. *Genomics Data*. 2015; 3: 94–6. doi: 10.1016/j.gdata.2014.11.012.
 45. Teif VB, Beshnova D a., Vainshtein Y, Marth C, Mallm JP, Rippe TH. Nucleosome repositioning links DNA (de) methylation and differential CTCF binding during stem cell development. *Genome Res*. 2014; 24: 1285–95. doi: 10.1101/gr.164418.113.
 46. Saluz HP, Wiebauer K, Wallace A. Studying DNA modifications and DNA-protein interactions *in vivo*: a window onto the native genome. *Trends Genet*. 1991; 7: 207–11.
 47. Medvedeva Y a, Khamis AM, Kulakovskiy I V, Ba-Alawi W, Bhuyan MSI, Kawaji H, Lassmann T, Harbers M, Forrest AR, Bajic VB. Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics*. 2014; 15: 119. doi: 10.1186/1471-2164-15-119.
 48. Domcke S, Bardet AF, Adrian Ginno P, Hartl D, Burger L, Schübeler D. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature*. 2015; 528: 575–9. doi: 10.1038/nature16462.
 49. Kim J Do, Hinz AK, Choo JH, Stubbs L, Kim J. YY1 as a controlling factor for the Peg3 and Gnas imprinted domains. *Genomics*. 2007; 89: 262–9. doi: 10.1016/j.ygeno.2006.09.009.
 50. Kubik G, Summerer D. Deciphering Epigenetic Cytosine Modifications by Direct Molecular Recognition. *ACS Chem Biol*. 2015; 10: 1580–9. doi: 10.1021/acscchembio.5b00158.
 51. Ulahannan N, Grealley JM. Genome-wide assays that identify and quantify modified cytosines in human

- disease studies. *Epigenetics Chromatin*. 2015; 8: 5. doi: 10.1186/1756-8935-8-5.
52. Liu Y, Olanrewaju YO, Zheng Y, Hashimoto H, Blumenthal RM, Zhang X, Cheng X. Structural basis for Klf4 recognition of methylated DNA. *Nucleic Acids Res*. 2014; 42: 4859–67. doi: 10.1093/nar/gku134.
 53. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet*. 2013; 14: 204–20. doi: 10.1038/nrg3354.
 54. Shen L, Wu H, Diep D, Yamaguchi S, D'Alessio AC, Fung HL, Zhang K, Zhang Y. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell*. 2013; 153: 692–706. doi: 10.1016/j.cell.2013.04.002.
 55. Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, M€ENZel M, Wagner M, M€Eller M, Khan F, Eberl HC, Mensinga A, Brinkman AB, et al. Dynamic readers for 5-(Hydroxy)methylcytosine and its oxidized derivatives. *Cell*. 2013; 152: 1146–59. doi: 10.1016/j.cell.2013.02.004.
 56. Iurlaro M, Ficiz G, Oxley D, Raiber E-A, Bachman M, Booth MJ, Andrews S, Balasubramanian S, Reik W. A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol*. 2013; 14: R119. doi: 10.1186/gb-2013-14-10-r119.
 57. Plongthongkum N, Diep DH, Zhang K. Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nat Rev Genet*. 2014; 15: 647–61. doi: 10.1038/nrg3772.
 58. Hua Y, Sahashi K, Rigo F, Hung G, Horev G, Bennett CF, Krainer AR. Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model. *Nature*. 2011; 478: 123–6. doi: 10.1038/nature10485.
 59. Lev Maor G, Yearim A, Ast G. The alternative role of DNA methylation in splicing regulation. *Trends Genet*. 2015; 31: 274–80. doi: 10.1016/j.tig.2015.03.002.
 60. Wan J, Oliver VF, Zhu H, Zack DJ, Qian J, Merbs SL. Integrative analysis of tissue-specific methylation and alternative splicing identifies conserved transcription factor binding motifs. *Nucleic Acids Res*. 2013; 41: 8503–14. doi: 10.1093/nar/gkt652.
 61. Tilgner H, Knowles DG, Johnson R, Davis C a., Chakraborty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigó R. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res*. 2012; 22: 1616–25. doi: 10.1101/gr.134445.111.
 62. Pandya-Jones A, Black DL. Co-transcriptional splicing of constitutive and alternative exons. *RNA*. 2009; 15: 1896–908. doi: 10.1261/rna.1714509.
 63. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of alternative splicing by histone modifications. *Science*. 2010; 327: 996–1000. doi: 10.1126/science.1184208.
 64. Gelfman S, Cohen N, Yearim A, Ast G. DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome Res*. 2013; 23: 789–99. doi: 10.1101/gr.143503.112.
 65. Maunakea AK, Chepelev I, Cui K, Zhao K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res*. 2013; 23: 1256–69. doi: 10.1038/cr.2013.110.
 66. Heard E, Martienssen RA. Transgenerational epigenetic inheritance: Myths and mechanisms. *Cell*. 2014. p. 95–109. doi: 10.1016/j.cell.2014.02.045.
 67. Barlow DP, Bartolomei MS. Genomic Imprinting in Mammals. *Cold Spring Harb Perspect Biol*. 2014; 6: a018382–a018382. doi: 10.1101/cshperspect.a018382.
 68. Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J, Dempster EL, Ren B. Base-resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. *Cell*. 2012; 148: 816–31. doi: 10.1016/j.cell.2011.12.035.
 69. Fang F, Hodges E, Molaro A, Dean M, Hannon GJ, Smith AD. Genomic landscape of human allele-specific DNA methylation. *Proc Natl Acad Sci U S A*. 2012; 109: 7332–7. doi: 10.1073/pnas.1201310109.
 70. Zheng X, Zhao Q, Wu H-J, Li W, Wang H, Meyer CA, Qin QA, Xu H, Zang C, Jiang P, Li F, Hou Y, He J, et al. MethylPurify: tumor purity deconvolution and differential methylation detection from single tumor DNA methylomes. *Genome Biol*. 2014; 15: 419. doi: 10.1186/s13059-014-0419-x.
 71. Xie H, Wang M, de Andrade A, Bonaldo M de F, Galat V, Arndt K, Rajaram V, Goldman S, Tomita T, Soares MB. Genome-wide quantitative assessment of variation in DNA methylation patterns. *Nucleic Acids Res*. 2011; 39: 4099–108. doi: 10.1093/nar/gkr017.
 72. He J, Sun X, Shao X, Liang L, Xie H. DMEAS: DNA methylation entropy analysis software. *Bioinformatics*. 2013; 29: 2044–5. doi: 10.1093/bioinformatics/btt332.
 73. Clark SJ, Lee HJ, Smallwood S a., Kelsey G, Reik W. Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol*. 2016; 17: 72. doi: 10.1186/s13059-016-0944-x.
 74. Smallwood S a, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, Andrews SR, Stegle O, Reik W, Kelsey G. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods*. 2014; 11: 817–20. doi: 10.1038/nmeth.3035.
 75. Courtois ET, Tan Y, Viswanathan R, Xing Q, Zhen Tan R, W Tan DS, Robson P, Loh Y, Feng Cheow L, Quake SR, Burkholder WF. Singlecell multimodal profiling reveals cellular epigenetic heterogeneity. *Nat Methods*. 2016; 13: 1–7. doi: 10.1038/nmeth.3961.
 76. Kurdyukov S, Bullock M. DNA Methylation Analysis: Choosing the Right Method. *Biology (Basel)*. 2016; 5: 3. doi: 10.3390/biology5010003.