Review

Identifying Readers for (hydroxy) methylated DNA Using Quantitative Interaction Proteomics: Advances and Challenges Ahead

Velin Marita Sequeira and Michiel Vermeulen

Department of Molecular Biology, *Faculty of Science, Radboud Institute for Molecular Life Sciences, Oncode Institute, Radboud University Nijmegen, 6525 GA, Nijmegen, the Netherlands*

Correspondence to Michiel Vermeulen: *Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Oncode Institute, Radboud University Nijmegen, 6525 GA, Nijmegen, the Netherlands.* *Michiel.vermeulen@science.ru.nl*
https://doi.org/10.1016/j.jmb.2019.12.014
Edited by Tuncay Baubec

Abstract

DNA methylation is an epigenetic modification, which regulates gene expression during cellular differentiation. This important function is thought to be carried out by transcriptional regulators, which are "readers" and effectors of this mark. In recent years, quantitative mass spectrometry-based interaction proteomics technology has emerged as a powerful tool to identify readers for methylated and unmethylated DNA in different cellular contexts. Furthermore, recent technology enables proteome-wide quantification of absolute affinities between proteins and methylated and unmethylated DNA in the context of crude nuclear extracts. Finally, recently developed locus-specific interaction proteomics approaches and modifications thereof facilitate an unbiased proteome characterization of methylated and unmethylated genomic loci *in vivo*. We summarize these recent findings in this review, and we argue that the integration of all these technologies, with also genomic sequencing-based approaches, will eventually result in a more detailed understanding of the link between DNA methylation and the regulation of transcription in health and disease.

Introduction

One of the enduring questions in molecular biology is how an adult eukaryotic organism can consist of a multitude of phenotypically different cell types, although all these cells harbor the same genome. This phenotypic diversity based on a single-genomic sequence is observed since each cell type only expresses a subset of the ~20,000 genes for which the human genome encodes. Cell-type specific gene expression is regulated by modifications of DNA and histones, which are actively altered, and therefore, providing a framework for directing gene activity. The most extensively studied epigenetic modification is DNA methylation, which plays a crucial role in cell type and tissue-specific gene expression, genomic imprinting, X chromosome inactivation, and maintenance of chromosomal instability. Human diseases, most notably cancer, show aberrant methylation patterns that contribute to the disease phenotype. DNA methylation entails the covalent addition of a methyl moiety to the 5th carbon of cytosine (5mC). DNA methylation dynamics involve *de novo* synthesis, maintenance, and removal of methylation marks, which are added and removed by two sets of enzymes, DNA methyltransferases (DNMTs) and DNA demethylases, respectively. The "'*de novo*" methyltransferases DNMT3A and DNMT3B, in cooperation with DNMT3L, establish a pattern of DNA methylation, which is faithfully transferred during cell division by DNMT1 that methylates the newly synthesized strand of these Hemimethylated duplexes after DNA replication. During the last decade, it became apparent that DNA methylation is actively removed from the genome [1–4]. Various DNA demethylation mechanisms have been described, including passive loss due to imperfect maintenance and active

demethylation catalyzed by ten-eleven translocation (TET) pathways. With iterative oxidations, TET enzymes convert mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) followed by glycosylase-driven excision of fC and caC from DNA [5].

Advances in sequencing-based technologies make it possible to generate genome-wide, base-resolution maps of DNA methylation. The golden standard for DNA methylation profiling is bisulfite sequencing. Bisulfite sequencing entails the treatment of DNA with bisulfite, which converts unmethylated cytosine to uracil while leaving methylated cytosine undisturbed [4,6]. However, bisulfite conversion involves the treatment of DNA at high temperatures resulting in substantial DNA degradation. Therefore, several bisulfite-free techniques have also been developed, including the recently developed TET-assisted pyridine borane sequencing (TAPS), which enables simultaneous detection of 5mC and 5hmC [7]. For exploring cellular heterogeneity at the epigenetic level, single-cell methylation profiling methods are also continuously being developed [8–10] (also reviewed in Ref. [4]). Collectively, these studies have revealed that in general, mammalian genomes are globally CpG-depleted, and approximately 60–80% of the 28 million CpGs in the human genome are methylated. Less than 10% of CpGs occur in CG dense regions known as CpG islands. These islands are predominantly observed at transcription start sites of housekeeping genes and developmental regulators. Although DNA methylation mostly occurs in a CG context, high throughput sequencing has also revealed cytosine methylation at non-CpG sites. However, the function and mechanisms of non-CpG methylation have still not been elucidated. Non-CpG methylation is mostly observed in a CpA context and is particularly abundant in oocytes, embryonic stem cells, a subset of neurons [1,2,6].

Despite our ability to identify DNA methylation patterns at the single-cell level with unprecedented resolution, our current understanding of the functional consequences of DNA methylation at a given locus is contrastingly quite limited. For instance, active regulatory elements are mostly unmethylated, and therefore, DNA methylation is typically associated with transcriptional silencing. This DNA methylation-mediated gene silencing is thought to involve direct interference of transcription factor binding and/or recruitment of transcriptionally repressive methyl CpG binding proteins (MBPs), often referred to as DNA methylation "readers." However, contrasting evidence indicates that DNA methylation mediated transcriptional regulation is more nuanced than previously perceived. The biological effects of DNA methylation are exerted by transcription factors that differentially interact with methylated and unmethylated CpG regions, as illustrated by targeted studies on DNA methylation "readers" [1,2]. For example, ZFP57 is a protein crucial for transcriptional silencing of imprinted alleles and can bind to its consensus binding site only when it is methylated [11]. On the other hand, C/EBPα binds to methylated *cis*-regulatory elements to activate gene programs essential for keratinocyte differentiation [12]. These findings highlight the importance of identifying readers for (hydroxy) methylated and unmethylated sequences in different cellular contexts in a high throughput manner. Such knowledge will provide valuable information that will help to elucidate how gene expression programs are controlled by DNA methylation during development and how perturbations thereof can contribute to malignant transformation and genomic imprinting associated diseases, for example. In recent years, multiple technologies have been developed to identify proteins that interact with (hydroxy)methylated and nonmethylated DNA sequences *in vitro* and *in vivo*. In this review, we summarize these methods and argue that the integration of these various technologies is needed to decipher how DNA methylation regulates gene expression and cell fate.

Strategies to study protein-methylated DNA interactions

Protein microarray and SELEX based approaches

Interactions between transcription factors and genomic sequences are typically studied using next-generation sequencing-based techniques, such as ChIP-sequencing or biochemical essays, such as EMSA, both of which are biased essays that focus on one protein of interest at a time.

In the past decade, numerous *in vitro* approaches have been employed to comprehensively profile the influence of CpG methylation on DNA binding for transcription factors. For example, using a protein microarray, Hu, and coworkers investigated the ability of 1531 proteins, including 1321 transcription factors to bind to 154 human mCpG containing promoters [13]. They identified 47 novel methylated DNA binding proteins, of which most proteins bound to methylated promoters in a sequence-dependent manner. This indicates that a subset of methylated proteins binds in a sequence-independent manner. Furthermore, they confirmed the *in vivo* interaction between methylated DNA and the Yamanaka factor KLF4 (which was previously identified to be an mCpG reader in a quantitative interaction proteomics study [14]) using ChIP-bisulfite sequencing [13]. Another *in vitro* method developed was methyl sensitive SELEX (Systematic evolution of ligands by exponential enrichment), which involved incubating an array of 542 purified transcription factors with a pool of random unmethylated or methylated oligos

[15]. The bound oligos were PCR-amplified and reincubated with the TF. Thus, repeated selection rounds enrich a set of motifs, which are correlated to bind with a high affinity to the transcription factor. Methyl sensitive SELEX revealed that 60% of the studied transcription factors were affected by the presence or absence of DNA methylation, underscoring the regulatory impact of DNA methylation. Transcription factors from the bHLH, bZIP, and ETS family were repelled by DNA methylation. Notably, these proteins have pivotal roles in embryonic development like homeodomain, and POU proteins demonstrated a binding preference to certain methylated DNA sequences. The authors further validated the *in vivo* binding of the pluripotency factor Oct4 to a methylated motif using ChIP-seq in mouse embryonic stem cells. However, in both these methods, the caveat is that the transcription factor binding is studied using a methylated or an unmethylated sequence and not in competition, as would occur *in vivo*. To circumvent this, EpiSELEX-seq was developed, which allows simultaneous probing of protein binding to methylated and unmethylated DNA sequences [16]. Thus, protein microarrays and SELEX-based approaches enable systematic identification of methylation-dependent or methylation-sensitive binding of transcription factors to mCpG containing DNA sequences.

Quantitative interaction proteomics

With the advent of high-resolution mass spectrometry (MS), DNA affinity purifications combined with quantitative mass spectrometry has emerged as a powerful tool to identify specific protein-DNA interactions in an unbiased, proteome-wide manner. In a typical DNA affinity purification experiment, biotinylated DNA baits of interest (including modifications such as CpG methylation) and a negative control sequence (for example a mutated DNA sequence or a DNA sequence lacking CpG methylation), immobilized on streptavidin-conjugated beads, are separately incubated with crude nuclear extract from a relevant cell line or tissue. After incubation, the biotinylated baits are then washed to remove background proteins and to enrich for proteins that specifically interact with the (modified) bait sequence relative to the control pull-down. Affinity enriched proteins are finally on-bead digested with trypsin and then subjected to LC-MS analysis. In these affinity purifications from crude lysates, specific interactions are masked by a large number of high abundant background proteins, necessitating a quantitative filter to discriminate specific interaction partners from background proteins [14,17,18]. This can be achieved by stable isotope labeling-based or label-free approaches (reviewed in Refs. [19–21]). A limitation of these DNA affinity purifications is that the modified DNA baits lack nucleosomal structure. Thus, to better recapitulate the *in vivo* situation in an *in vitro* setting, an affinity purification strategy employing recombinant nucleosomes containing methylated DNA and certain histone modifications has been developed [22]. Moreover, the DNA affinity purification workflow has been established in a high-throughput 96-well format that allows screening a large number of bait sequences with a reduced amount of input protein lysate in a semi-automated and reproducible manner [23].

Readers of DNA methylation: insights from quantitative proteomics

The canonical view of DNA methylation-driven protein–DNA interactions was that only proteins with a methyl-CpG (mCpG)-binding domain (MBD) or a SET- and RING-associated (SRA) domain interact with methylated CpG dinucleotides, mostly in a DNA sequence context-independent manner. The MBD protein family consists of seven proteins in mammals, namely MeCP2 (methyl-CpG-binding protein 2), MBD1, MBD2, MBD3, MBD4, MBD5, and MBD6 [24]. MeCP2, MBD1, MBD2, and MBD4 interact with mCpG mostly in a DNA sequence–independent manner and genome-wide binding these proteins correlates with mCpG density, although DNA methylation independent binding events are also observed for these proteins [25]. Mammalian MBD3, however, does not interact with methylated DNA. Two additional mammalian MBD domain-containing proteins, MBD5 and MBD6, also do not interact with methylated DNA. Instead, their MBD domain evolved as a protein-protein interaction module [26]. UHRF1 and UHRF2 carry an SRA domain, which interacts with Hemimethylated DNA, also in a DNA sequence context-independent manner. Quantitative MS-based proteomics and other biochemical and structural approaches have revealed that in addition to the MBD, several other domains, including homeobox, winged-helix domains (including forkhead boxes) and BTB/POZ-Zn-fingers can also interact with methylated DNA sequences. Many of the proteins carrying these domains interact with methylated DNA in a DNA sequence context-dependent manner, and some of these interactions may be explained by the fact that mC structurally resembles a thymidine when viewed from the major helix of the DNA (for review see Refs. [2,3]). In the first study that attempted to comprehensively characterize dynamic readers for 5mC and its oxidized derivatives (5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)), DNA affinity purifications coupled to LC-MS were performed in nuclear lysates from mouse embryonic stem cells, neural precursor cells (NPCs) and brain tissue [14]. In this study, 19 proteins were identified that preferentially bind to the methylated DNA probe rather than the nonmethylated counterpart in mouse

embryonic stem (ES) cell nuclear extracts. Apart from the well-known methyl binding, such as MeCP2, MBD1, and MBD4, several novel mCpG transcription factors were identified, such as MHC class II regulatory factor RFX1, zinc-finger and BTB domain-containing protein 44 (ZBTB44), zinc-finger homeobox 3 (ZFHX3), lysine-specific histone demethylase 1A (LSD1), and Krüppel-like factors (for example, KLF2, KLF4, and KLF5). When DNA affinity purifications using the same probes were performed using nuclear extracts from neuronal precursor cells, a distinct set of methyl CpG proteins were identified. For example, the transcriptionally repressive MBD2−NuRD complex interacts with methylated DNA in NPCs and adult mouse brain but does not interact with methylated DNA in nuclear extracts derived from mouse embryonic stem cells. This indicates that protein interactions with methylated and hydroxymethylated DNA are dynamic during cellular differentiation. These observations are relevant, given a reported partial uncoupling between DNA methylation and repression of transcription in early vertebrate development [27]. Strikingly, several DNA glycosylases (including TDG and the Neil glycosylases), as well as other DNA repair proteins were found to preferentially interact with hmC, fC, and caC compared to mC, further illustrating the involvement of these oxidized mC derivatives in base excision repair driven DNA demethylation pathways. Interestingly, DNMT1 was also identified as a specific fC/caC reading protein, which was confirmed by an electrophoretic mobility shift assay. mC and its oxidized derivatives not only recruit but also repel proteins from binding DNA. As is observed for mC, these repelled proteins can be classified into proteins that are repelled in a DNA sequence context-dependent and independent manner [2,14]. For example, CXXC domain-containing proteins interact with all CpG containing DNA sequences, irrespective of the DNA sequence surrounding the CpG, and methylation inhibits this binding. Genome-wide binding of CXXC domain-containing proteins thus correlates with CpG density. Other proteins not carrying a CXXC domain also can interact with certain CpG containing sequences in a methylation-sensitive manner, but in this case, the DNA sequence surrounding the CpG dinucleotide is also necessary for high-affinity binding.

A similar screen in mouse embryonic stem cells was performed using 5hmC and 5fC probes covering the Pax6 and Fgf15 promoters, which are known to be enriched for 5hmC in mouse embryonic stem cells [28]. This study revealed that forkhead box-containing transcription factors and the NuRD complex bind to 5fC-containing DNA, thus indicating that 5fC might be involved in gene expression regulation, in addition to its role in active DNA demethylation. This finding is particularly relevant given the fact that it has been reported that at least a subfraction of cellular fC is rather stable and does not turn over very rapidly [29,30]. Apart from using immobilized methylated oligonucleotides, *in vitro* reconstituted nucleosomes containing methylated DNA and various histone modifications have also been used as affinity baits for interaction proteomics purposes to uncover how different epigenetic modifications cooperate or antagonize each other to regulate protein interactions with nucleosomes. For instance, nucleosome pulldowns revealed that UHRF1 had increased affinity to H3K9me3-modified nucleosomes in the presence of CpG methylation. Furthermore, nucleosome pulldowns demonstrated that CpG methylation disrupts the Polycomb repressive complex 2 enrichment on H3K27me3-marked nucleosomes [22]. With constant improvements in experimental approaches including multiplexed labeling methods and data analysis tools and given the ever-increasing sensitivity and sequencing speed of modern mass spectrometers, we anticipate that quantitative mass spectrometry-based proteomics will continue to be useful to uncover DNA (hydroxy)methylation readers in early embryonic vertebrate development [31], as well as in adult life and in the context of various diseases.

Perspectives: new avenues and future challenges

Beyond sequence specificity: affinities of methylated DNA readers

Mass spectrometry-based interaction proteomics has contributed significantly to expand the catalog of proteins which specifically interact with (hydroxy) methylated DNA sequences *in vitro*. It has, however, on numerous occasions, proven quite difficult to confirm methylation-dependent binding for these proteins in the genome *in vivo* [3]. This is at least partially due to the fact that the affinity purification-based strategies described above are semi-quantitative and only provide information regarding binding specificity. However, transcription factor occupancy in the genome is dictated by (local) transcription factor concentration, and the affinity with which these transcription factors interact with (hydroxy)methylated genomic DNA sequences [32]. Determining affinities between proteins and (hydroxy)methylated DNA sequences is, therefore, important to better predict the *in vitro* detected (hydroxy)methylation-specific interactions that may also occur *in vivo*. Affinities between proteins and nucleic acids are typically determined on a case-by-case basis using conventional protein biochemistry methods such as isothermal calorimetry or fluorescence polarization measurements. Hashimoto and colleagues used fluorescence polarization experiments to investigate the affinity between MBD proteins and methylated and hydroxymethylated DNA, which provided

important insights. For example, they observed that MBD1 binds to fully methylated CpG-containing DNA with a higher affinity than the other MBD proteins. Furthermore, while MBD1 and MBD2 preferentially interact with methylated DNA compared to a hydroxymethylated DNA, MBD4 shows little discrimination. Finally, MBD3 binds with a low affinity to DNA and does not specifically interact with hydroxymethylated DNA [33], in contrast to a previous report [34]. Conventional protein biochemistry measurements, such as the ones described above, are very informative and provide important information. However, these techniques are low throughput in nature and necessitate the laborious expression and purification of recombinant proteins. For circumventing these drawbacks, an assay was recently developed, which provides apparent K_d values for many proteins in parallel in a single experiment. In this assay, a series of affinity purifications with increasing bait concentrations and fixed amounts of crude nuclear extracts are combined with isobaric stable isotope (TMT) labeling and quantitative mass spectrometry to generate Hill-like K_d curves for dozens of proteins in a single experiment [35]. In the future, it will be of interest to implement this workflow to interrogate the proteome-wide binding affinities of readers for (un)methylated and hydroxymethylated DNA sequences of interest. These experiments, combined with mass spectrometry-based measurements to determine cellular protein copy numbers in a global, proteome-wide manner [36–38], will provide more accurate predictions regarding *in vitro* detected protein-(hydroxy) methylated DNA interactions which are likely biologically relevant *in vivo*.

Profiling interactions with (hydroxy)methylated DNA *in vivo*: integrating genomics and proteomics approaches

In vivo interactions between proteins and genomic loci are usually profiled using next-generation sequencing-based technologies. The most well-known method for this is Chromatin Immunoprecipitation combined with DNA sequencing or ChIP-seq. In a ChIP-seq experiment, antibodies against bait proteins are used to immune-precipitate cross-linked chromatin fragments from cells of interest. Protein-DNA complexes are then de-crosslinked, and specially enriched genomic fragments are identified using next-generation DNA sequencing. ChIP-seq thus provides genome-wide binding profiles for proteins of interest (reviewed in Ref. [39]). These profiles can subsequently be compared with genome-wide (hydroxy)methylation profiles, which may or may not reveal a substantial genome-wide overlap between the epigenetic modification and the protein of interest. For confirming that a protein of interest is indeed binding to genomic loci containing methylated DNA, a method called ChIP-bisulfite sequencing was developed [40]. In this method, immunoprecipitated DNA obtained in a ChIP experiment is subjected to bisulfite sequencing to investigate whether this DNA is methylated or not. In the future, such assays could also be combined with sequencing methods, which allow discriminating methyl and hydroxymethylcytosine or other oxidized mC derivatives. ChIP-seq or ChIP-bisulfite sequencing experiments can also be complemented with perturbation experiments in which cellular DNA (hydroxy)methylation levels are altered, to investigate whether genomic binding for readers of interest is affected as a consequence. While several genomic sequencing-based methods for improved mapping of protein and DNA interactions have been developed like CUT&RUN sequencing (Cleavage Under Targets and Release Using Nuclease) [41], these methods are biased in the sense that bait proteins of interest are profiled genome-wide in a case-by-case manner. In recent years, chromatin immunoprecipitation workflows have been combined with mass spectrometry-based proteomics measurements (ChIP-MS) to provide a global view of the *in vivo* proteomic environment for a chromatin-associated protein of interest. ChIP-MS approaches were first described in yeast [42], but later also for human cells. ChIP-MS for H3K4me3 and H3K9me3 in HeLa cells enriched for proteins associated with active and silent chromatin, respectively. Moreover, the H3K9me3 interactome included H2A.X and the WICH complex, indicating heterochromatin specific DNA damage response mechanisms [43]. ChIP-MS for multiple histone modifications in mouse embryonic stem cells (mESCs) identified 114 proteins that were not previously known to be associated with histone modifications [44].

ChIP-MS or related approached are very useful since they provide an unbiased, proteome-wide overview of the cellular environment for proteins or epigenetic modifications of interest. These experiments thus provide clues regarding proteins, which may cooperate with the bait protein to regulate gene expression, for example. However, these methods still depend on prior knowledge of the factor to be studied and the availability of highly specific antibodies. The most crucial disadvantage of these methods, however, is that an average proteomic profile is generated over hundreds or thousands of binding sites in the genome [45,46]. Ideally, one would like to generate a proteomics-based inventory of individual genomic loci, such as a CpG island or CpG rich enhancer in a methylated versus non-methylated state. During the last decade, there has been a huge focus on developing robust technologies to achieve this objective. In a pioneering locus-specific proteomics study, a hybridizing desthiobiotinylated locked nucleic acid (LNA) containing oligonucleotide probe complementary to telomere

repeats was employed to target and capture telomeric regions of chromatin in multiple cell lines. This approach is called proteomics of isolated chromatin segments (PICh) [45]. To obtain a quantitative readout, PICh has been combined with SILAC labeling, and this has enabled analyzing the impact of perturbing DNA or histone-modifying enzymes on the proteome of pericentromeres and telomeres, respectively [47,48]. A similar method, which employs DNA probes called Hybridization Capture of Chromatin Associated Proteins for Proteomics (HyCAPP) was developed [49]. While this method does not require LNA probes, which are difficult to synthesize, DNA probes bind to genomic loci with lower affinity, which may hamper efficient protein enrichment. The large number of telomeres present in a cell, coupled with the repetitive nature of the telomere, proved to be advantageous for the PiCh technique, allowing a comprehensive identification of the telomere-bound proteome. However, for a single-copy locus, such as a promoter or an enhancer, there are several technical hurdles to overcome, including efficient affinity enrichment of that single-locus against a vast background signal represented by the rest of the genome, which is very difficult to achieve with single-affinity enrichment. Furthermore, since proteins cannot be amplified; a large amount (hundreds of millions to billions) of input cells are required to purify enough material to allow comprehensive single-locus proteomics. Given the ever-increasing sensitivity of modern mass spectrometers and continuously improving sample preparation methods, hybridization-based enrichment workflows may eventually also be suited for single-copy loci applications.

The adaptation of the CRISPR system utilizing an inactive dCas9 nuclease potentially offers attractive tools for improving locus-specific proteomics workflows [50,51]. A recently developed method called CAPTURE [46] makes use of a guide RNA to target a dCas9 protein that can be biotinylated to a genomic locus of interest. The biotinylated target engaged dCas9 protein, and associated locus-bound proteins can subsequently be affinity enriched

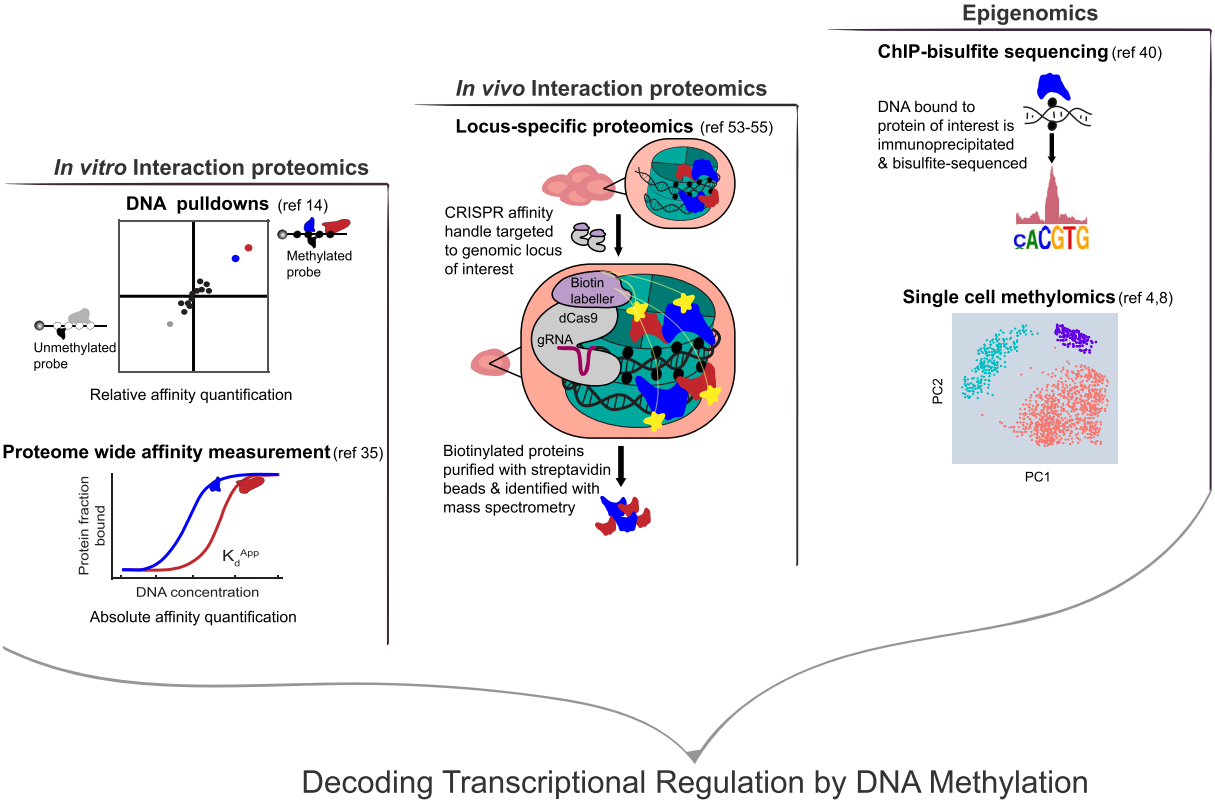

Fig. 1. Integrative omic approaches to decipher DNA methylation regulatory mechanisms. (a) *In vitro* interaction proteomics methods include DNA pulldown [14] and proteome-wide binding affinity assay [35], which determines specific binders to a methylated sequence and apparent binding affinities to the methylated motif respectively. (b) Locus-specific proteomics involves incorporating an affinity handle with a gene targeting module (like CRISPR-Cas9) and a proximity labeler into the nucleus followed by subsequent enrichment of labeled proteins [53–55]. (c) ChIP-bisulfite sequencing [40] is an essential tool to validate *in vivo* binding of a protein to methylated DNA. Single-cell methylation profiling [4,8] can identify differentially methylated regions in a heterogeneous cell population. Thus, proteomic and genomic approaches complement each other to fill the missing pieces in the DNA methylation regulation puzzle.

on streptavidin beads. Alternative strategies make use of proximity labeling, in which a promiscuous mutated BirA* ligase [52]), or ascorbate peroxidase enzymes [53–55] are fused to dCas9 to facilitate locus-specific biotinylation followed by streptavidin based affinity enrichment of locus-bound proteins. Potential drawbacks of these CRISPR-based systems coupled to proximity labeling and AP-MS include off-target binding of the dCas9 protein [56] and the presence of a large amount of free-floating, nontarget engaged dCas9 molecules in the nucleoplasm, which generates a substantial amount of background signal [51]. Finally, the actual targeting of dCas9 to genomic loci may alter the locus-specific proteome and may perturb the binding of biologically relevant proteins. Overcoming these challenges is expected to make CRISPR-based methods versatile and powerful tools for locus-specific proteomics (see Fig. 1). In the future, we foresee that robust, adaptable locus-specific proteome purification methods, which are scalable for high-throughput screenings, will be developed, which will allow asking fundamentally important questions. For example, transcription factors are known to bind only to a subset of their motifs *in vivo* [3,57]. For elucidating the unresolved molecular mechanisms behind this selectivity, locus-specific proteomics could provide important new insights. Furthermore, these methods can also be combined with perturbation experiments in which a genomic locus of interest will be methylated or hydroxymethylated on purpose (refer for approaches [58]), after which the perturbed locus can be characterized by mass spectrometry-based proteomics. Such methods will be very informative to uncover the molecular mechanisms of gene expression regulation by DNA (hydroxy)methylation.

Concluding remarks

The impact of DNA methylation on transcriptional regulation has been extensively investigated using genomic sequencing-based techniques. While DNA methylation profiling has facilitated discriminating methylated versus unmethylated CpG-containing regions in promoters, transcriptome profiling, ATAC sequencing, and ChIP-sequencing of histone markers have provided correlative information on how these different regulatory layers orchestrate gene expression. Moreover, to unravel cellular heterogeneity, single-cell multiomic strategies are constantly being devised to simultaneously profile mRNA, chromatin state, and DNA methylation to infer transcriptional regulatory mechanisms in a heterogeneous sample. However, as we have argued here, quantitative proteomics provides missing puzzle pieces by identifying transcriptional regulators that interpret this information to direct gene expression or repression. DNA affinity purifications, combined with quantitative mass spectrometry, allow the identification of transcriptional regulators and estimation of their binding affinities for (hydroxy)methylated genomic regions of interest. This knowledge can be complemented with absolute global protein levels to predict whether an *in vitro* identified reader is indeed bound at a region of interest *in vivo*. Locus-specific interaction proteomics, when developed robustly, will be a widely used technique to confirm the binding of these interactors *in vivo*. These proteomics applications will ideally always be complemented with genomic sequencing-based approaches, such as ChIP-seq, ChIP-bisulfite sequencing and single-cell genomics and single-cell multiomics profiling methods in an iterative manner (see Fig. 1). Ultimately, such studies will elucidate the dynamics of protein complexes assembling on (hydroxy)methylated and nonmethylated regulatory regions during cellular differentiation and pathogenesis and will thus uncover how DNA (hydroxy)methylation coordinates gene expression regulation in health and disease.

Acknowledgements

We apologize to authors whose work could be not cited due to space constraints. M.V. and V.M.S are part of the Oncode Institute, which is partly funded by the Dutch Cancer Society (KWF). Funding for V.M.S is provided by a grant from the Marie Curie Initial Training Network (ITN) EpiSyStem (EU, grant number 765966).

Declaration of competing interest

The authors declare no conflict of interest.

Received 11 October 2019;
Received in revised form 4 December 2019;
Accepted 5 December 2019
Available online 19 December 2019

Keywords:
DNA methylation;
Quantitative interaction proteomics;
Single-locus proteomics;
(hydroxy)methylation readers

Abbreviations used:
DNMT, DNA methyltransferases; TET, Ten eleven translocation enzymes; MS, Mass spectrometry; mESC, Mouse embryonic stem cells; ChIP, Chromatin immunoprecipitation; MBD, Methyl-binding proteins; PICh, proteomics of isolated chromatin segments; LNA, Locked nucleic acid.

References

[1] D. Schübeler, Function and information content of DNA methylation, Nature 517 (2015) 321–326.

[2] C.G. Spruijt, M. Vermeulen, DNA methylation: old dog, new tricks? Nat. Struct. Mol. Biol. 21 (2014) 949–954.

[3] H. Zhu, G. Wang, J. Qian, Transcription factors as readers and effectors of DNA methylation, Nat. Rev. Genet. 17 (2016) 551–565.

[4] I.D. Karemaker, M. Vermeulen, Single-cell DNA methylation profiling: technologies and biological applications, Trends Biotechnol. 36 (2018) 952–965.

[5] R.M. Kohli, Y. Zhang, TET enzymes, TDG and the dynamics of DNA demethylation, Nature 502 (2013) 472–479.

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

[7] Y. Liu, P. Siejka-Zielińska, G. Velikova, Y. Bi, F. Yuan, M. Tomkova, C. Bai, L. Chen, B. Schuster-Böckler, C.-X. Song, Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution, Nat. Biotechnol. 37 (2019) 424–429.

[8] H. Guo, P. Zhu, X. Wu, X. Li, L. Wen, F. Tang, Single-Cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing, Genome Res. 23 (2013) 2126–2135.

[9] S.J. Clark, S.A. Smallwood, H.J. Lee, F. Krueger, W. Reik, G. Kelsey, Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq), Nat. Protoc. 12 (2017) 534–547.

[10] S.M. Linker, L. Urban, S.J. Clark, M. Chhatriwala, S. Amatya, D.J. McCarthy, I. Ebersberger, L. Vallier, W. Reik, O. Stegle, M.J. Bonder, Combined single-cell profiling of expression and DNA methylation reveals splicing regulation and heterogeneity, Genome Biol. 20 (2019) 30.

[11] S. Quenneville, G. Verde, A. Corsinotti, A. Kapopoulou, J. Jakobsson, S. Offner, I. Baglivo, P.V. Pedone, G. Grimaldi, A. Riccio, D. Trono, In Embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions, Mol. Cell 44 (2011) 361–372.

[12] V. Rishi, P. Bhattacharya, R. Chatterjee, J. Rozenberg, J. Zhao, K. Glass, P. Fitzgerald, C. Vinson, CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes, Proc. Natl. Acad. Sci. U. S. A. 107 (2010), 20311–6.

[13] S. Hu, J. Wan, Y. Su, Q. Song, Y. Zeng, H.N. Nguyen, J. Shin, E. Cox, H.S. Rho, C. Woodard, S. Xia, S. Liu, H. Lyu, G.-L. Ming, H. Wade, H. Song, J. Qian, H. Zhu, DNA methylation presents distinct binding sites for human transcription factors, Elife 2 (2013).

[14] C.G. Spruijt, F. Gnerlich, A.H. Smits, T. Pfaffeneder, P.W.T.C. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, H.C. Eberl, A. Mensinga, A.B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, Dynamic readers for 5-(Hydroxy)Methylcytosine and its oxidized derivatives, Cell 152 (2013) 1146–1159.

[15] Y. Yin, E. Morgunova, A. Jolma, E. Kaasinen, B. Sahu, S. Khund-Sayeed, P.K. Das, T. Kivioja, K. Dave, F. Zhong, K.R. Nitta, M. Taipale, A. Popov, P.A. Ginno, S. Domcke, J. Yan, D. Schübeler, C. Vinson, J. Taipale, Impact of cytosine methylation on DNA binding specificities of human transcription factors, Science (80-.) 356 (2017) 2239.

[16] J.F. Kribelbauer, O. Laptenko, S. Chen, G.D. Martini, W.A. Freed-Pastor, C. Prives, R.S. Mann, H.J. Bussemaker, Quantitative analysis of the DNA methylation sensitivity of transcription factor complexes, Cell Rep. 19 (2017) 2383–2395.

[17] G. Mittler, F. Butter, M. Mann, A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements, Genome Res. 19 (2009) 284–293.

[18] S.J.J. Bartels, C.G. Spruijt, A.B. Brinkman, P.W.T.C. Jansen, M. Vermeulen, H.G. Stunnenberg, A SILAC-based screen for methyl-CpG binding proteins identifies RBP-J as a DNA methylation and sequence-specific binding protein, PLoS One 6 (2011), e25884.

[19] M. Vermeulen, N.C. Hubner, M. Mann, High confidence determination of specific protein–protein interactions using quantitative mass spectrometry, Curr. Opin. Biotechnol. 19 (2008) 331–337.

[20] K.A. Neilson, N.A. Ali, S. Muralidharan, M. Mirzaei, M. Mariani, G. Assadourian, A. Lee, S.C. van Sluyter, P.A. Haynes, Less label, more free: approaches in label-free quantitative mass spectrometry, Proteomics 11 (2011) 535–553.

[21] O. Chahrour, D. Cobice, J. Malone, Stable isotope labelling methods in mass spectrometry-based quantitative proteomics, J. Pharm. Biomed. Anal. 113 (2015) 2–20.

[22] T. Bartke, M. Vermeulen, B. Xhemalce, S.C. Robson, M. Mann, T. Kouzarides, Nucleosome-interacting proteins regulated by DNA and histone methylation, Cell 143 (2010) 470–484.

[23] N.C. Hubner, L.N. Nguyen, N.C. Hornig, H.G. Stunnenberg, A quantitative proteomics tool to identify DNA–protein interactions in primary cells or blood, J. Proteome Res. 14 (2015) 1315–1329.

[24] P.A. Defossez, I. Stancheva, Biological functions of methyl-CpG-binding proteins, Prog. Mol. Biol. Transl. Sci. (2011) 377–398.

[25] T. Baubec, R. Ivánek, F. Lienert, D. Schübeler, Methylation-dependent and -independent genomic targeting principles of the mbd protein family, Cell 153 (2013) 480–492.

[26] H.I. Baymaz, A. Fournier, S. Laget, Z. Ji, P.W.T.C. Jansen, A.H. Smits, L. Ferry, A. Mensinga, I. Poser, A. Sharrocks, P.-A. Defossez, M. Vermeulen, MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain, Proteomics 14 (2014) 2179–2189.

[27] O. Bogdanovic, S.W. Long, S.J. van Heeringen, A.B. Brinkman, J.L. Gómez-Skarmeta, H.G. Stunnenberg, P.L. Jones, G.J.C. Veenstra, Temporal uncoupling of the DNA methylome and transcriptional repression during embryogenesis, Genome Res. 21 (2011) 1313–1327.

[28] M. Iurlaro, G. Ficz, D. Oxley, E.-A. Raiber, M. Bachman, M.J. Booth, S. Andrews, S. Balasubramanian, W. Reik, A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation, Genome Biol. 14 (2013) R119.

[29] M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, 5-Formylcytosine could Be a semipermanent base in specific genome sites, Angew. Chem. Int. Ed. 55 (2016) 11797–11800.

[30] M. Bachman, S. Uribe-Lewis, X. Yang, H.E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, 5-

Formylcytosine can be a stable DNA modification in mammals, Nat. Chem. Biol. 11 (2015) 555–557.
[31] O. Bogdanović, A.H. Smits, E. de la Calle Mustienes, J.J. Tena, E. Ford, R. Williams, U. Senanayake, M.D. Schultz, S. Hontelez, I. van Kruijsbergen, T. Rayon, F. Gnerlich, T. Carell, G.J.C. Veenstra, M. Manzanares, T. Sauka-Spengler, J.R. Ecker, M. Vermeulen, J.L. Gómez-Skarmeta, R. Lister, Active DNA demethylation at enhancers during the vertebrate phylotypic period, Nat. Genet. 48 (2016) 417–426.
[32] E. Segal, J. Widom, From DNA sequence to transcriptional behaviour: a quantitative approach, Nat. Rev. Genet. 10 (2009) 443–456.
[33] H. Hashimoto, Y. Liu, A.K. Upadhyay, Y. Chang, S.B. Howerton, P.M. Vertino, X. Zhang, X. Cheng, Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation, Nucleic Acids Res. 40 (2012) 4841–4849.
[34] O. Yildirim, R. Li, J.-H. Hung, P.B. Chen, X. Dong, L.-S. Ee, Z. Weng, O.J. Rando, T.G. Fazzio, Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells, Cell 147 (2011) 1498–1510.
[35] M.M. Makowski, C. Gräwe, B.M. Foster, N.V. Nguyen, T. Bartke, M. Vermeulen, Global profiling of protein–DNA and protein–nucleosome binding affinities using quantitative mass spectrometry, Nat. Commun. 9 (2018) 1653.
[36] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M. Selbach, Global quantification of mammalian gene expression control, Nature 473 (2011) 337–342.
[37] A.H. Smits, R.G.H. Lindeboom, M. Perino, S.J. van Heeringen, G.J.C. Veenstra, M. Vermeulen, Global absolute quantification reveals tight regulation of protein expression in single Xenopus eggs, Nucleic Acids Res. 42 (2014) 9880–9891.
[38] J. Bonnet, R.G.H. Lindeboom, D. Pokrovsky, G. Stricker, M.H. Çelik, R.A.W. Rupp, J. Gagneur, M. Vermeulen, A. Imhof, J. Müller, Quantification of proteins and histone marks in Drosophila embryos reveals stoichiometric relationships impacting chromatin regulation, Dev. Cell 51 (5) (2019 Dec 2) 632–644.e6, https://doi.org/10.1016/j.devcel.2019.09.011.
[39] D.S. Johnson, A. Mortazavi, R.M. Myers, B. Wold, Genome-wide mapping of in vivo protein-DNA interactions, Science 316 (2007) 1497–1502.
[40] A.B. Brinkman, H. Gu, S.J.J. Bartels, Y. Zhang, F. Matarese, F. Simmer, H. Marks, C. Bock, A. Gnirke, A. Meissner, H.G. Stunnenberg, Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk, Genome Res. 22 (2012) 1128–1138.
[41] P.J. Skene, S. Henikoff, An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites, Elife 6 (2017).
[42] J.-P. Lambert, L. Mitchell, A. Rudner, K. Baetz, D. Figeys, A novel proteomics approach for the discovery of chromatin-associated protein networks, Mol. Cell. Proteom. 8 (2009) 870–882.
[43] M. Soldi, T. Bonaldi, The proteomic investigation of chromatin functional domains reveals novel synergisms among distinct heterochromatin components, Mol. Cell. Proteom. 12 (2013) 764–780.
[44] X. Ji, D.B. Dadon, B.J. Abraham, T.I. Lee, R. Jaenisch, J.E. Bradner, R.A. Young, Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 3841–3846.
[45] J. Déjardin, R.E. Kingston, Purification of proteins associated with specific genomic loci, Cell 136 (2009) 175–186.
[46] X. Liu, Y. Zhang, Y. Chen, M. Li, F. Zhou, K. Li, H. Cao, M. Ni, Y. Liu, Z. Gu, K.E. Dickerson, S. Xie, G.C. Hon, Z. Xuan, M.Q. Zhang, Z. Shao, J. Xu, In situ capture of chromatin interactions by biotinylated dCas9, Cell 170 (2017) 1028–1043.e19.
[47] N. Saksouk, T.K. Barth, C. Ziegler-Birling, N. Olova, A. Nowak, E. Rey, J. Mateos-Langerak, S. Urbach, W. Reik, M.E. Torres-Padilla, A. Imhof, J. Déjardin, Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation, Mol. Cell 56 (2014) 580–594.
[48] M. Gauchier, S. Kan, A. Barral, S. Sauzet, E. Agirre, E. Bonnell, N. Saksouk, T.K. Barth, S. Ide, S. Urbach, R.J. Wellinger, R.F. Luco, A. Imhof, J. Déjardin, SETDB1-dependent heterochromatin stimulates alternative lengthening of telomeres, Sci. Adv. 5 (2019).
[49] H. Guillen-Ahlers, P.K. Rao, M.E. Levenstein, J. Kennedy-Darling, D.S. Perumalla, A.Y.L. Jadhav, J.P. Glenn, A. Ludwig-Kubinski, E. Drigalenko, M.J. Montoya, H.H. Göring, C.D. Anderson, M. Scalf, H.I.S. Gildersleeve, R. Cole, A.M. Greene, A.K. Oduro, K. Lazarova, A.J. Cesnik, J. Barfknecht, L.A. Cirillo, A.P. Gasch, M.R. Shortreed, L.M. Smith, M. Olivier, HyCCAPP as a tool to characterize promoter DNA-protein interactions in Saccharomyces cerevisiae, Genomics 107 (2016) 267–273.
[50] J.D. Sander, J.K. Joung, CRISPR-Cas systems for editing, regulating and targeting genomes, Nat. Biotechnol. 32 (2014) 347–355.
[51] M. Adli, The CRISPR tool kit for genome editing and beyond, Nat. Commun. 9 (2018) 1911.
[52] E. Schmidtmann, T. Anton, P. Rombaut, F. Herzog, H. Leonhardt, Determination of local chromatin composition by CasID, Nucleus 7 (2016) 476–484.
[53] X.D. Gao, L.C. Tu, A. Mir, T. Rodriguez, Y. Ding, J. Leszyk, J. Dekker, S.A. Shaffer, L.J. Zhu, S.A. Wolfe, E.J. Sontheimer, C.- BERST, Defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2, Nat. Methods 15 (2018) 433–436.
[54] W. Qiu, Z. Xu, M. Zhang, D. Zhang, H. Fan, T. Li, Q. Wang, P. Liu, Z. Zhu, D. Du, M. Tan, B. Wen, Y. Liu, Determination of local chromatin interactions using a combined CRISPR and peroxidase APEX2 system, Nucleic Acids Res. 47 (2019) e52–e52.
[55] S.A. Myers, J. Wright, R. Peckner, B.T. Kalish, F. Zhang, S.A. Carr, Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling, Nat. Methods 15 (2018) 437–439.
[56] M. Wierer, M. Mann, Proteomics to study DNA-bound and chromatin-associated gene regulatory complexes, Hum. Mol. Genet. 25 (2016) R106–R114.
[57] M. Slattery, T. Zhou, L. Yang, A.C. Dantas Machado, R. Gordân, R. Rohs, Absence of a simple code: how transcription factors read the genome, Trends Biochem. Sci. 39 (2014) 381–399.
[58] C.B. Waryah, C. Moses, M. Arooj, P. Blancafort, Zinc fingers, TALEs, and CRISPR systems: a comparison of tools for epigenome editing, in: A. Jeltsch, M.G. Rots (Eds.), Epigenome Ed. Methods Protoc., Springer New York, New York, NY, 2018, pp. 19–63.