ChemPubSoc Europe

DOI: 10.1002/cbic.201900035

CHEMBIOCHEM Minireviews

Compilation of Modern Technologies To Map Genome-Wide Cytosine Modifications in DNA

Hu Zeng[+], Bo He[+], and Chengqi Yi*[a]

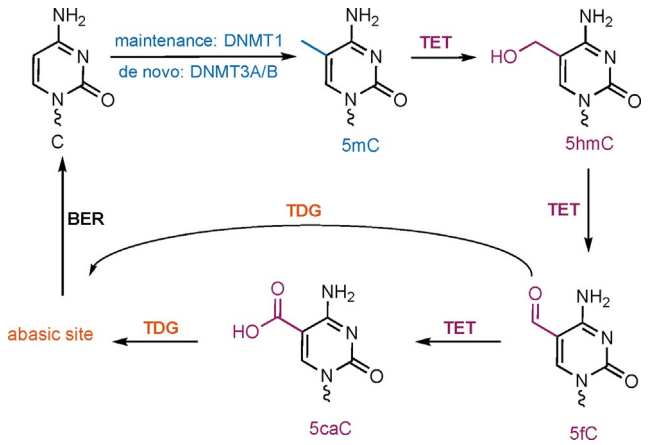
Over the past few decades, various DNA modification detection methods have been developed; many of the high-resolution methods are based on bisulfite treatment, which leads to DNA degradation, to a degree. Thus, novel bisulfite-free approaches have been developed in recent years and shown to be useful for epigenome analysis in otherwise difficult-to-handle, but important, DNA samples, such as hmC-seal and hmC-CATCH. Herein, an overview of advances in the development of epigenome sequencing methods for these important DNA modifications is provided.

1. Introduction

1.1. DNA methylation and its oxidized derivatives

DNA methylation is one of the most extensively characterized epigenetic modifications that plays an essential role in the regulation of gene expression in mammalian cells. Transfer of the methyl group to the fifth carbon of cytosine results in the generation of 5-methylcytosine (5mC). The reaction is catalyzed by DNA methyltransferases (DNMTs) and prefers to occur within CpG dinucleotides.[1] In the genome of mammalian somatic tissues, 70 to 80% of CpG cytosines are methylated.[2] The influence of DNA methylation on gene regulation was initially studied by analyzing CpG islands (CGIs), which are regions of DNA sequence enriched in CpG dinucleotides and found at the majority of mammalian promoters.[1b] The presence of 5mC at CGIs hinders the binding of transcriptional factors to DNA, and thus, is usually associated with gene repression.[3] In addition to the influence on transcription regulation, DNA methylation also plays important biological roles in genomic imprinting, X-chromosome inactivation (XCI), RNA splicing, and so forth.[4]

Apart from 5mC, other modifications also exist in mammalian genomic DNA. In the 1970s, two papers suggested that mammals contained high levels of 5-hydroxymethylcytosine (5hmC),[5] whereas others could not repeat these results and 5hmC was regarded as potential DNA damage as a result of oxidation of the 5-methyl group.[6] In 2009, Tahiliani et al. showed that the ten–eleven translocation 1 (TET1) catalyzed the conversion of 5mC into 5hmC,[7] and another paper published at the same time also demonstrated the presence of 5hmC in the mouse brain.[8] In 2011, two groups demonstrated that, similar to thymidine hydroxylase, TET was capable of iterative oxidation, producing 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC).[9] Once 5fC and 5caC are formed, they can be actively removed by thymine DNA glycosylase (TDG) through base excision repair (BER); thus resulting in active DNA demethylation in mammals[9b] (Scheme 1). Emerging evidence indicates that these DNA demethylation intermediates are stable and play distinct epigenetic roles.[10]



Scheme 1. The dynamic regulation of DNA methylation.

1.2. Bisulfite-dependent DNA modification analysis and challenges

Bisulfite sequencing (BS-seq), which is regarded as the gold standard for 5mC detection, has been widely used for genome-wide DNA methylation analysis.[11] Bisulfite treatment transforms cytosine (C) into uracil (U); this was first discovered independently by two groups in 1970.[12] Later studies in 1980 demonstrated much lower reactivity of 5mC in bisulfite treatment than that of C, due to inhibition of sulfonate adduct for-

[a] Dr. H. Zeng,[+] B. He,[+] Prof. C. Yi
State Key Laboratory of Protein and Plant Gene Research
School of Life Sciences, Department of Chemical Biology and
Synthetic and Functional Biomolecules Center
College of Chemistry and Molecular Engineering and
Peking-Tsinghua Center for Life Sciences, Peking University
Beijing 100871 (P. R. China)
E-mail: chengqi.yi@pku.edu.cn

[+] These authors contributed equally to this work.

The ORCID identification numbers for the authors of this article can be found under https://doi.org/10.1002/cbic.201900035.

This article is part of the young researchers' issue ChemBioTalents. To view the complete issue, visit http://chembiochem.org/chembiotalents

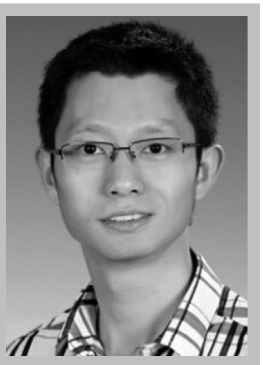
mation because of the electronic effect of the methyl group on the 5′-carbon of 5mC.[13] In subsequent PCR amplification and sequencing, all of the unmodified cytosine sites are read as T, whereas 5mC remains as C. This differential reactivity between 5mC and C during bisulfite treatment is the basis of BS-seq. The presence of 5mC oxidation derivatives complicates BS-seq, because 5hmC is read as C, similarly to 5mC, whereas 5fC and 5caC are read as T, similarly to C, in BS-seq.[9,14] To address this problem, additional chemical treatments have been coupled to BS-seq, leading to BS-dependent sequencing methods capable of distinguishing different types of cytosine modifications. For instance, oxidative bisulfite sequencing (oxBS-seq) and TET-assisted bisulfite sequencing (TAB-Seq) can distinguish 5hmC from 5mC and detect 5hmC at single-base resolution in the genome.[15] In addition, chemical-assisted bisulfite sequencing for 5fC (fCAB-seq) and reduced bisulfite sequencing (redBS-seq) can distinguish 5fC from C and 5caC and are of single-base resolution as well.[16] For 5caC, chemical-modification-assisted bisulfite sequencing (CAB-seq) was developed.[17] An additional bisulfite-dependent method, termed M.SssI methylase-assisted bisulfite sequencing (MAB-seq), can distinguish 5fC and 5caC from C within the context of CpG dinucleotides and provide a mapping method for 5fC/5caC simultaneously.[18] The readout of cytosine and modified cytosines in these bisulfite-dependent sequencing methods is summarized in Table 1.

A key limitation of these bisulfite-dependent sequencing methods is the high degree of DNA degradation caused by bisulfite treatment. A previous study reported that more than 99% of the input genomic DNA was degraded during a conventional 16 h long bisulfite treatment reaction.[19] Although recent efforts have been made to reduce such severe degradation,[20] large quantities of initial DNA input are generally required for the detection of 5mC oxidized derivatives, which significantly limits their applications to precious biological and clinical DNA samples. Moreover, because unmethylated cytosine is converted into thymine, bisulfite treatment reduces the sequence complexity of template DNA, which often complicates the design of efficient primers for PCR amplification. Therefore, bisulfite-free techniques for DNA modification detection are desired and indeed have been actively pursued in the community. Herein, we highlight developments and applications of these alternative bisulfite-free sequencing methods.

Table 1. Readout of cytosine modifications in bisulfite-dependent methods.

Method	C	5mC	5hmC	5fC	5caC
BS-seq	T	C	C	T	T
oxBS-seq	T	C	T	T	T
TAB-seq	T	T	C	T	T
fCAB-seq	T	C	C	C	T
redBS-seq	T	C	C	C	T
CAB-seq	T	C	C	T	C
MAB-seq	C	C	C	T	T

Chengqi Yi received his BS from the University of Science and Technology of China in 2005 and Ph.D. from the University of Chicago in 2010. After two years of post-doctoral training at the Department of Biochemistry and Molecular Biology at the University of Chicago, he joined the School of Life Sciences at Peking University in 2012, where he is now an assistant professor with tenure. His research is primarily focused on the pathways and mechanisms of DNA/RNA modification and de-modification.



2. Protein-Based Discrimination Methods

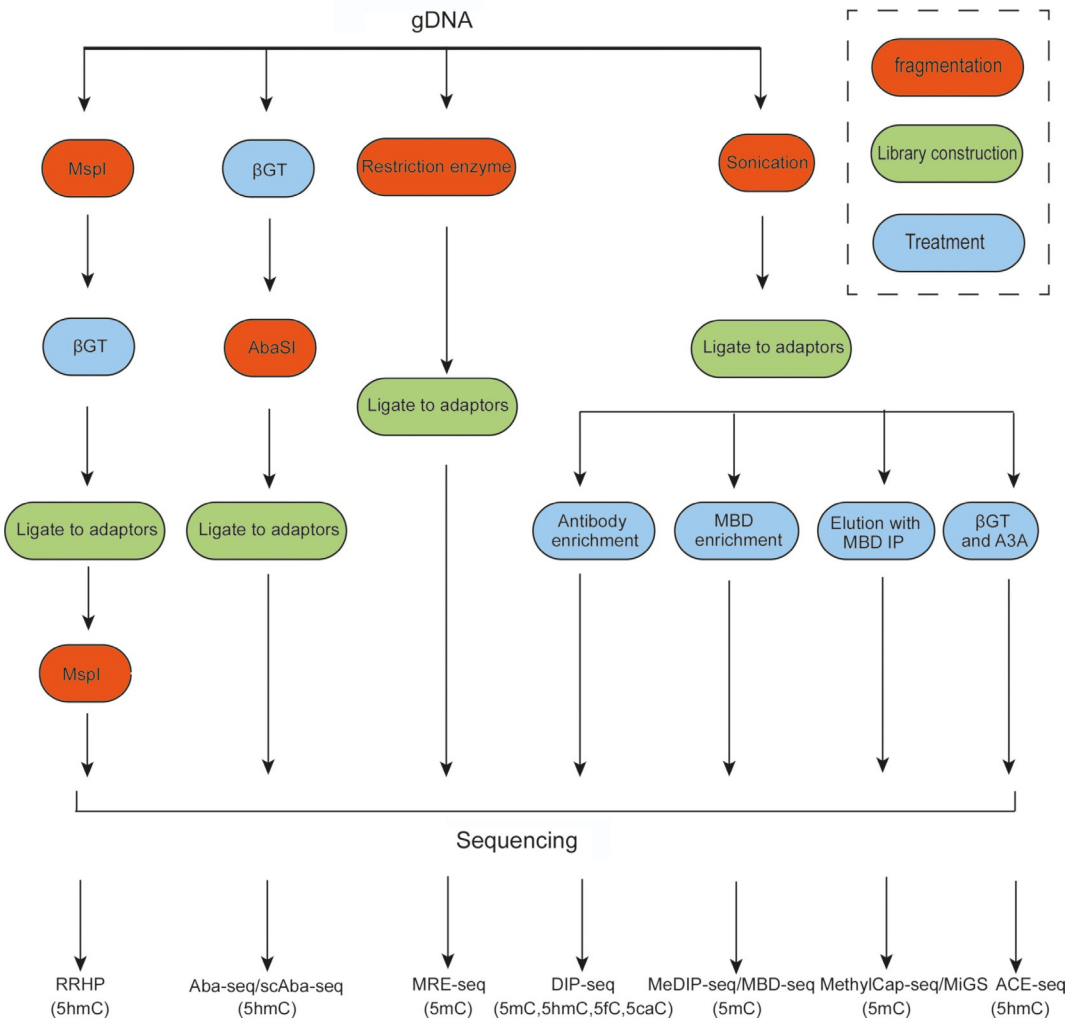
These methods involve several different kinds of DNA-binding proteins/enzymes and take advantage of their differential reactivity towards different DNA modifications to achieve selective detection.

2.1. Antibodies

Methylated DNA immunoprecipitation (MeDIP) is a widely used strategy for 5mC detection that uses antibodies to specifically enrich methylated DNA fragments for sequencing.[21] Through this method, genomic DNA is fragmented and denatured; the resulting single-stranded DNA (ssDNA) fragments containing 5mC are then immunoprecipitated by using 5mC antibodies (Scheme 2). The enriched fragments are decoded by high-throughput sequencing. Similar to MeDIP, oxidized derivatives of 5mC containing DNA can also be enriched by using commercial antibodies with well-characterized specificities.[22] Although immunoprecipitation-dependent methods are straightforward for the detection of DNA modifications, they are not quantitative, and their resolution is dependent on the size of the DNA fragments of the prepared DNA library, which are generally several hundreds of base pairs. In addition, antibody-based enrichment methods are highly dependent on the quality of the antibody. Low specificity of the antibody used or cross-reactivity with off-target sites leads to high background noise. The lot-to-lot inconsistency of the antibody is also a disadvantage.

2.2. Methyl CpG binding domains (MBDs)

The MBD has a strong affinity towards methylated DNA sequences[23] and can also be used for DNA methylation analysis.[24] Four proteins, namely, MBD1, MBD2, MBD3, and MeCP2, which share a similar structural domain, can recognize symmetric methyl groups in CpG dinucleotides of double-stranded DNA (dsDNA).[23c,e] Based on MBD proteins, some genome-wide detection DNA methylation methods, termed MBD-seq, MIGS, and MethylCap-seq, have been developed (Scheme 2).[25] Similar drawbacks exist for such methods in comparison with MeDIP.



Scheme 2. Protein-based discrimination techniques for DNA modification detection. The procedures involve fragmentation, treatment, library construction, and sequencing. MRE-seq, methylation restriction enzyme sequencing; RRHP, reduced representation 5hmC profiling; Aba-seq, DNA-modification-dependent restriction endonuclease AbaSI coupled with sequencing; DIP-seq, DNA immunoprecipitation and shotgun sequencing; MeDIP-seq, methylated DNA immunoprecipitation and sequencing; MBD-seq, methyl-CpG binding domain (MBD) protein-enriched genome sequencing; MethylCap-seq, methylation DNA capture sequencing; MiGS, MBD-isolated genome sequencing; ACE-seq, APOBEC-coupled epigenetic sequencing.

2.3. Restriction endonucleases

Restriction enzymes are useful tools in DNA modification detection, owing to their highly sequence-specific endonuclease activity.[26] The DNA cleavage activity of restriction endonucleases can be inhibited in the presence of DNA modifications in the recognition site.[27] By using restriction endonucleases that are sensitive to methylation states in their recognition sites, several DNA methylation detection methods have been developed (Scheme 2).[25e,28] Recently, combining restriction enzyme digestion with enzymatic glucosylation of 5hmC, base-resolution 5hmC sequencing methods for both bulk samples and single-cell samples have been developed (Scheme 2).[28e,29] Restriction enzymes provide a cost-effective way for accurately quantifying DNA modifications, whereas these methods are limited by the restriction sites of the endonucleases.

2.4. Cytosine deaminases

Members of the AID/APOBEC family are known as cytosine deaminases, which catalyze the deamination reaction in ssDNA and are involved in innate and adaptive immunity.[30] Kohli et al. showed that the catalytic activity of several members could discriminate between different cytosine modifications.[31] Recently, by using a human-specific deaminase, APOBEC3A (A3A), the authors further developed a bisulfite-free and base-resolution 5hmC sequencing method, termed ACE-seq[32] (Scheme 2). Owing to its non-destructive feature, it can use 1000-fold less DNA input than bisulfite treatment based methods and may be possible to detect 5hmC signals in long sequencing reads.

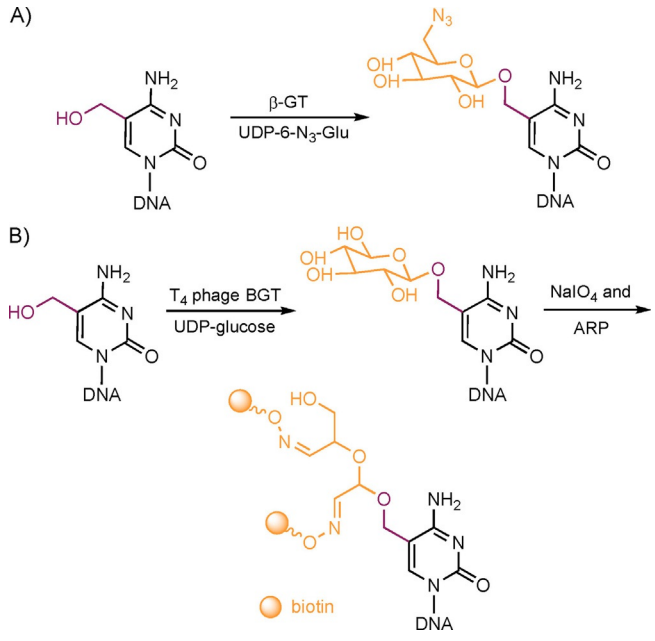
3. Bisulfite-Free Chemical Labeling Methods

Novel chemical reactions are invaluable for the development of innovative epigenome sequencing tools, and in this section

we focus on chemical approaches for the detection of DNA modifications.[33]

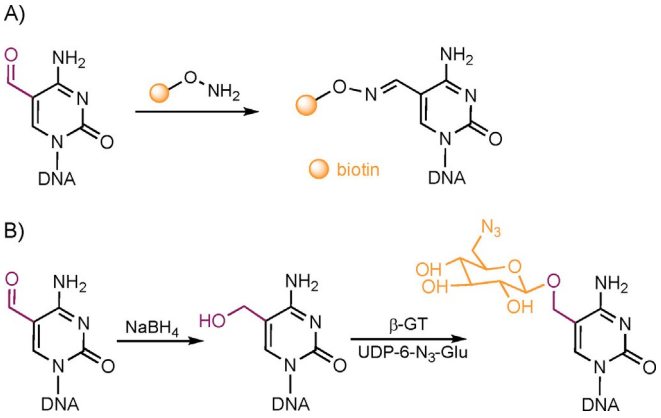
3.1. Chemical-based profiling methods

Many chemical profiling methods have been developed in recent years. For instance, Song et al. developed a chemical profiling 5hmC detection method termed hmC-seal.[34] The method took advantage of the β-glucosyltransferase (βGT)-catalyzed 5hmC glucosylation reaction that transferred an azide-modified glucose onto the hydroxyl moiety of 5hmC (Scheme 3A). A biotin can be subsequently installed onto the azido group through copper-free click chemistry. The biotin–streptavidin interaction, which is extremely tight and specific, can then be used to perform selective pull-down for genome-wide profiling of 5hmC distribution and has virtually no modification density bias. Using hmC-seal, researchers have performed genome-wide profiling of 5hmC in various cell lines and brain tissues, and found a distinct age-dependent distribution of 5hmC in brain tissues relative to embryonic stem (ES) cell lines.[34,35] Taking advantage of this highly sensitive and selective chemical labeling approach, several groups have performed genome-wide profiling of 5hmC in circulating cell-free DNA (cfDNA) to explore the diagnostic potential of 5hmC.[36] Their results showed that 5hmC signatures in cfDNA could be diagnostic biomarkers for human cancers. Nano-hmC-Seal, which combines the selective 5hmC chemical labeling of hmC-seal with an engineered Tn5 transposase-based library construction strategy, is able to generate genome-wide 5hmC maps from ultralow starting genomic DNA samples.[37] Another biotin-based 5hmC-profiling method, termed GLIB, used βGT to transfer an unmodified glucose to 5hmC, followed by sodium periodate treatment to generate reactive aldehyde groups.[38] The aldehydes can then react with an ARP, which can be used for further enrichment (Scheme 3B).



Scheme 3. Chemical labeling strategy for 5hmC genome-wide profiling. A) By using UDP-6-N3-Glu, an azide group can be installed onto the hydroxyl group of 5hmC by β-GT. B) The glucosylation, periodate oxidation, and biotinylation (GLIB) method. The 5hmC site conjugated two biotins after the GLIB reaction. T4 phage BGT: T4 phage β-glucosyltransferase, ARP: aldehyde-reactive probe.

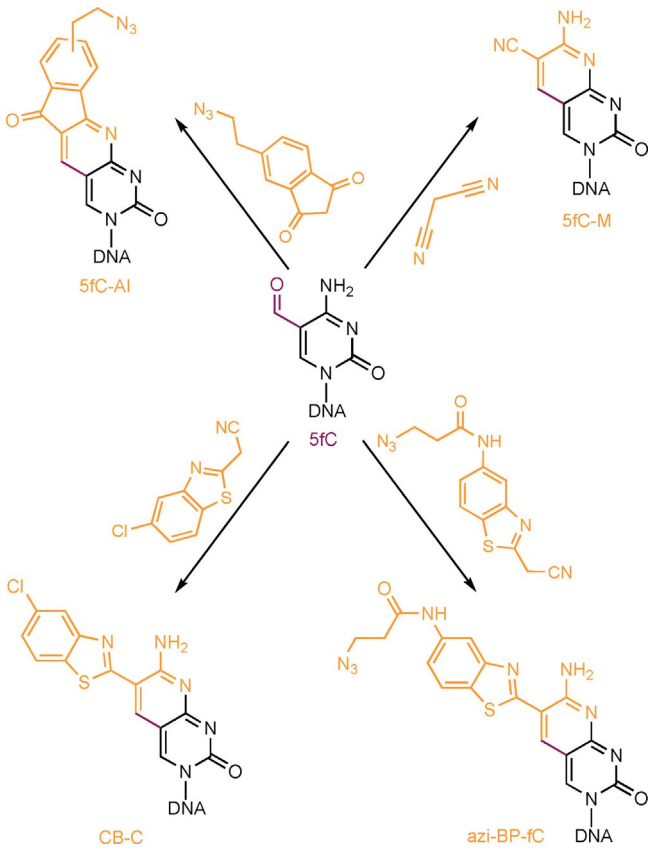
In the case of 5fC, the aldehyde group on the modified bases can be exploited to react with a hydroxylamine–biotin probe (Scheme 4A). Using this strategy, Raiber et al. performed genome-wide mapping of 5fC in mouse ES cells and their results suggested that 5fC was involved in epigenetic reprogramming within specific genomic regions.[39] Song et al. extended their hmC-seal method and developed another 5fC-profiling method termed fC-seal.[16b] To enrich 5fC-containing DNA, they first blocked endogenous 5hmC with unmodified UDP-Glc by using βGT, and then they reduced 5fC to 5hmC with $NaBH_4$ and glucosylated the newly generated 5hmC with an azide-modified glucose (Scheme 4B). The azide group was coupled to a biotin-containing probe through click chemistry, which allowed pull-down and enrichment of 5fC-containing DNA. Using this approach, they also performed genome-wide mapping of 5fC in mouse ES cells and their results revealed that 5fC preferentially occurred at poised enhancers.



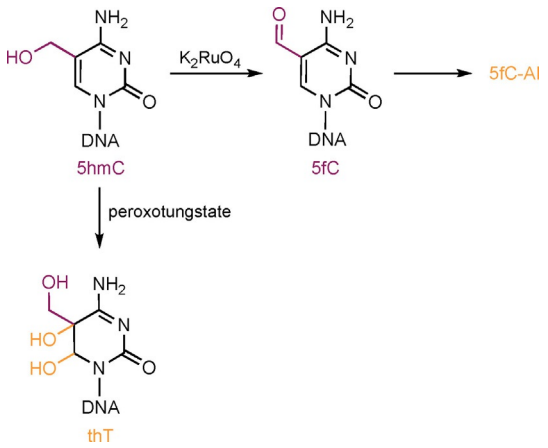
Scheme 4. Chemical labeling strategies for 5fC genome-wide profiling. A) Biotin-labeling of 5fC by using oxyamino-labeled biotin. B) 5fC was selectively reduced to 5hmC and then glucosylated with UDP-6-N3-Glu.

3.2. Chemical-labeling-enabled single-base-resolution sequencing methods

Bisulfite-free and single-base-resolution DNA modification sequencing methods have recently been developed. Xia et al. reported the first genome-wide bisulfite-free and base-resolution 5fC sequencing method termed cyclization-enabled C-to-T transition of 5fC (fC-CET).[40] This method is based on selective labeling of 5fC by using an azido derivative of 1,3-indandione (Scheme 5). The azide group in the labeling adduct enabled the enrichment of 5fC-containing DNA fragments before sequencing. This greatly reduced the sequencing cost for genome-wide 5fC detection compared with the two bisulfite-based methods redBS-seq and fCAB-seq, considering the limit-



Scheme 5. Chemical labeling methods for single-base-resolution detection of 5-5fC.



Scheme 6. Chemical labeling methods for single-base-resolution detection of 5-5hmC.

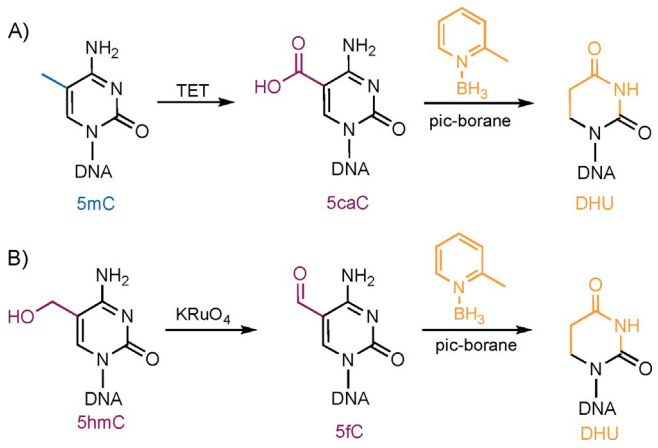
ed abundance of 5fC in the genome. Based on another biocompatible and selective chemical labeling of 5fC with malononitrile (Scheme 5), we further developed a single-cell and single-base-resolution 5fC sequencing method, termed chemical-labeling-enabled C-to-T conversion sequencing (CLEVER-seq).[41] Using this approach, we identified the dynamic 5fC landscapes in mouse gametes and early embryos at single-base resolution and the results revealed that promoter 5fC production preceded the upregulation of corresponding gene expression during mouse preimplantation development. Following a similar concept of fC-CET, Liu et al. reported a different chemical, namely, 2(5chlorobenzo[*d*]thiazol2yl)acetonitrile (CBAN), to label 5fC[42] (Scheme 5). The fluorescence of the labeling adduct enables the quantification of 5fC levels in DNA. Similarly, 5fC was labeled with another reagent termed aziBP (Scheme 5).[43]

Bisulfite-free and base-resolution 5hmC sequencing methods have also been reported. Hayashi et al. developed a chemical 5hmC detection method in which peroxotungstate was used to selectively oxidize and deaminate 5hmC.[44] The oxidized product, trihydroxylated thymine (thT), tolerated the incorporation of an adenine instead of a guanine in the process of DNA polymerase amplification (Scheme 6); thus enabling the detection of 5hmC at single-base resolution. However, its utility in detecting genome-wide 5hmC was not demonstrated. We recently developed a bisulfite-free, genome-wide, and base-resolution 5hmC sequencing method termed chemical-assisted C-to-T conversion of 5hmC sequencing (hmC-CATCH).[45] hmC-CATCH is based on the selective oxidation of 5hmC to 5fC, and subsequent labeling of newly generated 5fC by using derivatives of 1,3-indandione (Scheme 6); endogenous 5fC is first blocked before the oxidation reaction. Importantly, we showed that K_2RuO_4, but not $KRuO_4$, mediated very mild oxidation of DNA. The labeling adduct causes C-to-T transition during DNA amplification and enables base-resolution detection of 5hmC. In addition, the azido group allowed enrichment of 5hmC-containing DNA for detection. Compared with several existing base-resolution 5hmC methods (e.g., TAB-seq and oxBS-seq), hmC-CATCH causes minimal DNA degradation, is cost-effective, and is suitable for applications with limited starting DNA materials. The bisulfite-free nature of hmC-CATCH has enabled the first genome-wide detection of 5hmC in cfDNA isolated from the blood of healthy individuals and cancer patients.

Aside from BSseq, other novel chemical strategies for 5mC detection have been developed. Derivatization of DNA with *O*-allylhydroxylamine was able to discriminate between C and 5mC because the adducts had different basepairing selectivity.[46] Recently, Liu et al. reported a bisulfite-free and base-resolution cytosine modification sequencing method, termed TET assisted 2-picoline borane (pic-borane) sequencing (TAPS).[47] The approach was based on 5caC-to-T transition chemistry, which was enabled by the conversion of 5caC to dihydrouracil (DHU) using pic-borane. Combining the 5caC-to-T transition chemistry with TET oxidation of 5mC, the method realized whole-genome, base-resolution detection of 5mC (Scheme 7 A). The reaction can also be used to detect 5hmC because pic-borane can also convert 5fC into DHU and 5hmC can be oxidized to 5fC (Scheme 7 B).[47]

4. Direct Sequencing of Unamplified DNA

Currently, developments in third-generation sequencing technologies, such as single-molecule, real-time sequencing (SMRT) and nanopore-based sequencing, have allowed new opportu-



Scheme 7. Single-base-resolution detection of A) 5mC and B) 5hmC by using the pic-borane reaction.

nities to directly detect modifications at single-base resolution without prior amplification steps.

4.1. SMRT sequencing

The method exploits the pausing of DNA polymerases due to the presence of modifications.[48] The circularized DNA templates pass through a DNA polymerase and the instrument will obtain the sequencing information. Because the polymerase is immobilized in the zeromode waveguide (ZMW), the method can detect a signal from a single molecule by using a camera, which records the color and duration of emitted light during the incorporation of a fluorescently labeled deoxyribonucleoside triphosphate (dNTP). The system can detect not only the sequence information, but also the DNA modification status by monitoring changes of the duration of the fluorescent pulse and the time interval between successive pulses. Reports have shown the kinetics of the polymerase influenced by DNA modifications, such as 5mC, 5hmC, N^4-methylcytosine (4mC), N^6-methyladenine (6mA), and other types of DNA modifications.[48b,49] SMRT sequencing has different sensitivities for different modifications, and thus, each modification type has a specific signal-to-noise ratio. Detection of DNA methylation requires a high sequencing depth, and additional labeling steps can help to improve the signal confidence. By using Tet1 to convert 5mC into 5caC[50] and two T-even bacteriophage enzymes to convert 5hmC into a diglucosylated adduct,[51] the signal-to-noise ratios for the detection of these two important epigenetic markers can be improved.

4.2. Nanopore sequencing

Nanopore sequencing has been developed for decades, and the nanopore sequencer became commercially available in 2014 from Oxford Nanopore Technologies.[52] Recently, new instruments named GridION and PromethION were released. The approach also shows promise for DNA modification detection. With this technology, ssDNA travels through nanopores, which are proteins inserted in a synthetic polymer barrier.[52h] Shifts in voltage and the changing of these charges reflect the features of a specific DNA sequence, which is called a k-mer.[53]

In the early stages, nanopore sequencing attempted to detect methylation in eukaryotic cells by using different engineered protein nanopore constructs and experimental conditions. Accordingly, only 5mC and 5hmC could be detected.[52a,f,h,54] However, recently, the development of the MinION device opened up the way to characterize prokaryotic methylomes.[55] Meanwhile, researchers have reported a DNA methylation mapping workflow, which uses a variable-order hidden Markov model (HMM) with a hierarchical Dirichlet process (HDP) to decipher the ionic signal.[55b] This research has illustrated the probability of direct DNA methylation detection by using nanopores.[56] Engineering improved poreforming proteins or improving the statistical models will promote the accuracy of nanopore sequencing, and this may stimulate broad applications in epigenetics research. Aside from protein-based nanopores, pores formed by carbon nanotubes and solidstate nanopores have also been used in DNA modification detection.[57]

5. Summary and Outlook

The discovery of oxidized cytosine modifications in the mammalian genome stimulated new interest in epigenetics. The development of DNA modification detection methods has enabled a growth in knowledge of DNA base modifications and genome functions. Table 2 lists the features of these mapping techniques. Most of the earlier DNA modification sequencing methods are based on bisulfite treatment, which is very powerful, but also has unavoidable shortcomings. Herein, we mainly described epigenome sequencing technologies without bisulfite conversion. The prominent advantage of these bisulfite-free methods is minimal DNA degradation, and thus, is very helpful for DNA modification detection with low input samples. Some of these bisulfite-free methods have even realized genome-wide DNA modification detection for single-cell samples. The advantages of bisulfite-free approaches also enabled their application for limited clinical DNA samples. Despite the advantages, several challenges of bisulfite-free approaches still need to be addressed, such as low resolution and the loss of absolute-level information. The endonuclease-dependent bisulfite-free method can only detect DNA modifications in the recognition sites, which will lead to high type-I errors. For the direct sequencing of unamplified DNA approaches, the relatively high errors rates are also of major concern. We envision that further advances in the detection of DNA modifications will drive a more complete understanding of their roles in development and diseases.

Acknowledgements

We thank all members of the Yi laboratory for their insights and discussions. This work was supported by the Fok Ying Tung Education Foundation (no. 161018) and the National Natural Science Foundation of China (grants no. 21825701 and 91740112).

Table 2. A comparison of DNA modification sequencing methods.

	Methods	Input DNA[a]	Base resolution	Absolute level	Ref.
bisulfite-based detection methods	oxBS-seq	100 ng–1 μg	yes	yes	[15b]
	TAB-seq	1–3 μg	yes	yes	[15a]
	fCAB-seq	n.a.	yes	yes	[16b]
	redBS-seq	4 μg	yes	yes	[16a]
	MAB-seq	200–500 ng	yes	yes	[18]
	CAB-seq	n.a.	yes	yes	[17]
protein-based discrimination methods	MeDIP-seq	50 ng–5 μg	no	no	[21]
	DIP-seq	50 ng–5 μg	no	no	[22]
	MBD-seq	1 μg	no	no	[25a]
	MethylCap-seq	1 μg	no	no	[25c]
	MRE-seq	1–3 μg	no	no	[25e]
	Aba-seq	50 ng	yes	no	[28f]
	scAba-seq	one cell	yes	yes	[28e]
	RRHP	100 ng	yes	no	[28d]
	ACE-seq	10 ng	yes	yes	[32]
bisulfite-free chemical labeling methods	hmC-Seal	10 ng–10 μg	no	no	[34]
	nano-hmC-Seal	5 ng	no	no	[37]
	fC-Seal	50 μg	no	no	[16b]
	fC-CET	1–5 μg	yes	no	[40]
	CLEVER-seq	one cell	yes	yes	[41]
	CBAN	n.a.	yes	no	[42]
	hmC-CATCH	10 ng	yes	no	[45]
	TAPS	1 ng	yes	yes	[47]
direct sequencing of unamplified DNA	SMRT	50 ng	yes	yes	[51], [52]
	nanopore sequencing	1 μg	yes	yes	[55]

[a] n.a.: not available.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: chemical labeling · DNA · epigenetics · genomics · sequence determination

[1] a) G. Egger, G. Liang, A. Aparicio, P. A. Jones, *Nature* **2004**, *429*, 457–463; b) P. A. Jones, *Nat. Rev. Genet.* **2012**, *13*, 484–492.
[2] 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, *Nature* **2009**, *462*, 315–322.
[3] J. F. Costello, C. Plass, *J. Med. Genet.* **2001**, *38*, 285–303.
[4] V. R. Liyanage, J. S. Jarmasz, N. Murugeshan, M. R. Del Bigio, M. Rastegar, J. R. Davie, *Biology (Basel)* **2014**, *3*, 670–723.
[5] a) N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, *Biochem. J.* **1972**, *126*, 781–790; b) N. W. Penn, *Biochem. J.* **1976**, *155*, 709–712.
[6] a) R. M. Kothari, V. Shankar, *J. Mol. Evol.* **1976**, *7*, 325–329; b) E. Privat, L. C. Sowers, *Chem. Res. Toxicol.* **1996**, *9*, 745–750.
[7] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935.
[8] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930.
[9] a) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303; b) Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303–1307.
[10] a) M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 1049–1055; b) L. Shen, C. X. Song, C. He, Y. Zhang, *Annu. Rev. Biochem.* **2014**, *83*, 585–614; c) C. X. Song, C. He, *Trends Biochem. Sci.* **2013**, *38*, 480–484.
[11] a) R. Lister, R. C. O'Malley, J. Tonti-Filippini, B. D. Gregory, C. C. Berry, A. H. Millar, J. R. Ecker, *Cell* **2008**, *133*, 523–536; b) M. D. Schultz, Y. P. He, J. W. Whitaker, M. Hariharan, E. A. Mukamel, D. Leung, N. Rajagopal, J. R. Nery, M. A. Urich, H. M. Chen, S. Lin, Y. Lin, I. Jung, A. D. Schmitt, S. Selvaraj, B. Ren, T. J. Sejnowski, W. Wang, J. R. Ecker, *Nature* **2015**, *523*, 212-U189.
[12] a) H. Hayatsu, Y. Wataya, K. Kai, S. Iida, *Biochemistry* **1970**, *9*, 2858–2858; b) R. Shapiro, R. E. Servis, M. Welcher, *J. Am. Chem. Soc.* **1970**, *92*, 422–422.
[13] R. Y. H. Wang, M. Ehrlich, *Fed. Proc.* **1980**, *39*, 1882–1882.
[14] Y. Huang, W. A. Pastor, Y. H. Shen, M. Tahiliani, D. R. Liu, A. Rao, *PLoS One* **2010**, *5*, e8888.
[15] a) M. Yu, G. C. Hon, K. E. Szulwach, C. X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J. H. Min, P. Jin, B. Ren, C. He, *Cell* **2012**, *149*, 1368–1380; b) M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, *336*, 934–937.
[16] a) M. J. Booth, G. Marsico, M. Bachman, D. Beraldi, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 435–440; b) C. X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S. Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G. L. Xu, P. Jin, C. He, *Cell* **2013**, *153*, 678–691.
[17] X. Y. Lu, C. X. Song, K. Szulwach, Z. P. Wang, P. Weidenbacher, P. Jin, C. He, *J. Am. Chem. Soc.* **2013**, *135*, 9315–9317.
[18] a) F. Guo, X. Li, D. Liang, T. Li, P. Zhu, H. Guo, X. Wu, L. Wen, T. P. Gu, B. Hu, C. P. Walsh, J. Li, F. Tang, G. L. Xu, *Cell Stem Cell* **2014**, *15*, 447–459; b) H. Wu, X. Wu, L. Shen, Y. Zhang, *Nat. Biotechnol.* **2014**, *32*, 1231–1240.
[19] K. Tanaka, A. Okamoto, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1912–1915.
[20] a) S. A. Smallwood, H. J. Lee, C. Angermueller, F. Krueger, H. Saadeh, J. Peat, S. R. Andrews, O. Stegle, W. Reik, G. Kelsey, *Nat. Methods* **2014**, *11*, 817–820; b) S. Gravina, X. Dong, B. Yu, J. Vijg, *Genome Biol.* **2016**, *17*, 150; c) S. J. Clark, S. A. Smallwood, H. J. Lee, F. Krueger, W. Reik, G. Kelsey, *Nat. Protoc.* **2017**, *12*, 534-U159.
[21] a) M. Weber, I. Hellmann, M. B. Stadler, L. Ramos, S. Paabo, M. Rebhan, D. Schubeler, *Nat. Genet.* **2007**, *39*, 457–466; b) M. Weber, J. J. Davies, D. Wittig, E. J. Oakeley, M. Haase, W. L. Lam, D. Schubeler, *Nat. Genet.* **2005**, *37*, 853–862.
[22] a) L. Shen, H. Wu, D. Diep, S. Yamaguchi, A. C. D'Alessio, H. L. Fung, K. Zhang, Y. Zhang, *Cell* **2013**, *153*, 692–706; b) C. E. Nestor, R. Ottaviano, J. Reddington, D. Sproul, D. Reinhardt, D. Dunican, E. Katz, J. M. Dixon, D. J. Harrison, R. R. Meehan, *Genome Res.* **2012**, *22*, 467–477; c) K. Williams, J. Christensen, K. Helin, *EMBO Rep.* **2011**, *13*, 28–35; d) H. Stroud, S. Feng, S. Morey Kinney, S. Pradhan, S. E. Jacobsen, *Genome Biol.* **2011**, *12*, R54; e) G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature* **2011**, *473*, 398-U589.
[23] a) M. Wang, Z. Xu, L. Chen, H. Yin, S. Ai, *Anal. Chem.* **2012**, *84*, 9072–9078; b) Y. Yu, S. Blair, D. Gillespie, R. Jensen, D. Myszka, A. H. Badran, I. Ghosh, A. Chagovetz, *Anal. Chem.* **2010**, *82*, 5012–5019; c) R. R. Meehan, J. D. Lewis, S. McKay, E. L. Kleiner, A. P. Bird, *Cell* **1989**, *58*, 499–507; d) K. Tanaka, K. Tainaka, T. Kamei, A. Okamoto, *J. Am. Chem. Soc.* **2007**, *129*, 5612–5620; e) P. A. Wade, A. Gegonne, P. L. Jones, E. Ballestar, F. Aubry, A. P. Wolffe, *Nat. Genet.* **1999**, *23*, 62–66.
[24] L. G. Acevedo, A. Sanz, M. A. Jelinek, *Epigenomics* **2011**, *3*, 93–101.
[25] a) C. Clark, P. Palta, C. J. Joyce, C. Scott, E. Grundberg, P. Deloukas, A. Palotie, A. J. Coffey, *PLoS One* **2012**, *7*, e50233; b) K. A. Aberg, J. L. McClay, S. Nerella, L. Y. Xie, S. L. Clark, A. D. Hudson, J. Bukszar, D. Adkins, C. M. Hultman, P. F. Sullivan, P. K. E. Magnusson, E. J. C. G. van den Oord, S. S. Consortium, *Epigenomics* **2012**, *4*, 605–621; c) A. B. Brinkman, F. Simmer, K. Ma, A. Kaan, J. Zhu, H. G. Stunnenberg, *Methods* **2010**, *52*, 232–236; d) D. Serre, B. H. Lee, A. H. Ting, *Nucleic Acids Res.* **2010**, *38*, 391–399; e) A. K. Maunakea, R. P. Nagarajan, M. Bilenky, T. J. Ballinger, C. D'Souza, S. D. Fouse, B. E. Johnson, C. B. Hong, C. Nielsen, Y. J. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V. M. Heine, D. H. Rowitch, X. Y. Xing, C. Fiore, M. Schillebeeckx, S. J. M. Jones,

D. Haussler, M. A. Marra, M. Hirst, T. Wang, J. F. Costello, *Nature* **2010**, *466*, 253-U131.
[26] A. Pingoud, A. Jeltsch, *Nucleic Acids Res.* **2001**, *29*, 3705 – 3727.
[27] C. Waalwijk, R. A. Flavell, *Nucleic Acids Res.* **1978**, *5*, 3231 – 3236.
[28] a) S. Lindsay, A. P. Bird, *Nature* **1987**, *327*, 336 – 338; b) A. Bird, M. Taggart, M. Frommer, O. J. Miller, D. Macleod, *Cell* **1985**, *40*, 91 – 99; c) J. Kaput, T. W. Sneider, *Nucleic Acids Res.* **1979**, *7*, 2303 – 2322; d) A. Petterson, T. H. Chung, D. Tan, X. G. Sun, X. Y. Jia, *Genome Biol.* **2014**, *15*, 456; e) D. Mooijman, S. S. Dey, J. C. Boisset, N. Crosetto, A. van Oudenaarden, *Nat. Biotechnol.* **2016**, *34*, 852 – 856; f) Z. Sun, J. Terragni, J. G. Borgaro, Y. Liu, L. Yu, S. Guan, H. Wang, D. Sun, X. Cheng, Z. Zhu, S. Pradhan, Y. Zheng, *Cell Rep.* **2013**, *3*, 567 – 576.
[29] a) Z. Y. Sun, N. Dai, J. G. Borgaro, A. Quimby, D. P. Sun, I. R. Correa, Y. Zheng, Z. Y. Zhu, S. X. Guan, *Mol. Cell* **2015**, *57*, 750 – 761; b) A. A. Serandour, S. Avner, E. A. Mahe, T. Madigou, S. Guibert, M. Weber, G. Salbert, *Genome Biol.* **2016**, *17*, 56.
[30] S. U. Siriwardena, K. Chen, A. S. Bhagwat, *Chem. Rev.* **2016**, *116*, 12688 – 12710.
[31] E. K. Schutsky, C. S. Nabel, A. K. F. Davis, J. E. DeNizio, R. M. Kohli, *Nucleic Acids Res.* **2017**, *45*, 7655 – 7665.
[32] E. K. Schutsky, J. E. DeNizio, P. Hu, M. Y. Liu, C. S. Nabel, E. B. Fabyanic, Y. Hwang, F. D. Bushman, H. Wu, R. M. Kohli, *Nat. Biotechnol.* **2018**, *36*, 1083 – ##.
[33] M. J. Booth, E. A. Raiber, S. Balasubramanian, *Chem. Rev.* **2015**, *115*, 2240 – 2254.
[34] C. X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C. H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin, C. He, *Nat. Biotechnol.* **2011**, *29*, 68 – 72.
[35] K. E. Szulwach, X. K. Li, Y. J. Li, C. X. Song, H. Wu, Q. Dai, H. Irier, A. K. Upadhyay, M. Gearing, A. I. Levey, A. Vasanthakumar, L. A. Godley, Q. Chang, X. D. Cheng, C. He, P. Jin, *Nat. Neurosci.* **2011**, *14*, 1607-U1150.
[36] a) W. Li, X. Zhang, X. Lu, L. You, Y. Song, Z. Luo, J. Zhang, J. Nie, W. Zheng, D. Xu, Y. Wang, Y. Dong, S. Yu, J. Hong, J. Shi, H. Hao, F. Luo, L. Hua, P. Wang, X. Qian, F. Yuan, L. Wei, M. Cui, T. Zhang, Q. Liao, M. Dai, Z. Liu, G. Chen, K. Meckel, S. Adhikari, G. Jia, M. B. Bissonnette, X. Zhang, Y. Zhao, W. Zhang, C. He, J. Liu, *Cell Res.* **2017**, *27*, 1243 – 1257; b) C. X. Song, S. Yin, L. Ma, A. Wheeler, Y. Chen, Y. Zhang, B. Liu, J. Xiong, W. Zhang, J. Hu, Z. Zhou, B. Dong, Z. Tian, S. S. Jeffrey, M. S. Chua, S. So, W. Li, Y. Wei, J. Diao, D. Xie, S. R. Quake, *Cell Res.* **2017**, *27*, 1231 – 1242; c) X. Tian, B. Sun, C. Chen, C. Gao, J. Zhang, X. Lu, L. Wang, X. Li, Y. Xing, R. Liu, X. Han, Z. Qi, X. Zhang, C. He, D. Han, Y. G. Yang, Q. Kan, *Cell Res.* **2018**, *28*, 597 – 600.
[37] D. Han, X. Lu, A. H. Shih, J. Nie, Q. You, M. M. Xu, A. M. Melnick, R. L. Levine, C. He, *Mol. Cell* **2016**, *63*, 711 – 719.
[38] W. A. Pastor, U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal, A. Rao, *Nature* **2011**, *473*, 394 – 397.
[39] E. A. Raiber, D. Beraldi, G. Ficz, H. E. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik, S. Balasubramanian, *Genome Biol.* **2012**, *13*, R69.
[40] B. Xia, D. Han, X. Lu, Z. Sun, A. Zhou, Q. Yin, H. Zeng, M. Liu, X. Jiang, W. Xie, C. He, C. Yi, *Nat. Methods* **2015**, *12*, 1047 – 1050.
[41] C. Zhu, Y. Gao, H. Guo, B. Xia, J. Song, X. Wu, H. Zeng, K. Kee, F. Tang, C. Yi, *Cell Stem Cell* **2017**, *20*, 720 – 731, e725.
[42] C. Liu, Y. Wang, W. Yang, F. Wu, W. Zeng, Z. Chen, J. Huang, G. Zou, X. Zhang, S. Wang, X. Weng, Z. Wu, Y. Zhou, X. Zhou, *Chem. Sci.* **2017**, *8*, 7443 – 7447.
[43] Y. F. Wang, C. X. Liu, X. Zhang, W. Yang, F. Wu, G. R. Zou, X. C. Weng, X. Zhou, *Chem. Sci.* **2018**, *9*, 3723 – 3728.
[44] G. Hayashi, K. Koyama, H. Shiota, A. Kamio, T. Umeda, G. Nagae, H. Aburatani, A. Okamoto, *J. Am. Chem. Soc.* **2016**, *138*, 14178 – 14181.
[45] H. Zeng, B. He, B. Xia, D. Bai, X. Lu, J. Cai, L. Chen, A. Zhou, C. Zhu, H. Meng, Y. Gao, H. Guo, C. He, Q. Dai, C. Yi, *J. Am. Chem. Soc.* **2018**, *140*, 13190 – 13194.
[46] M. Münzel, L. Lercher, M. Muller, T. Carell, *Nucleic Acids Res.* **2010**, *38*, e192.
[47] P. S. Yi. Liu, G. Velikova, F. Yuan, M. Tomkova, C. Bai, L. Chen, Y. Bi, B. Schuster-Boeckler, C. Song, *BioRxiv* **2018**, https://doi.org/10.1101/307538.
[48] a) J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, S. Turner, *Science* **2009**, *323*, 133 – 138; b) B. A. Flusberg, D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, T. A. Clark, J. Korlach, S. W. Turner, *Nat. Methods* **2010**, *7*, 461-U472.
[49] a) E. E. Schadt, O. Banerjee, G. Fang, Z. Feng, W. H. Wong, X. Zhang, A. Kislyuk, T. A. Clark, K. Luong, A. Keren-Paz, A. Chess, V. Kumar, A. Chen-Plotkin, N. Sondheimer, J. Korlach, A. Kasarskis, *Genome Res.* **2013**, *23*, 129 – 141; b) T. A. Clark, K. E. Spittle, S. W. Turner, J. Korlach, *Genome Integr.* **2011**, *2*, 10.
[50] T. A. Clark, X. Lu, K. Luong, Q. Dai, M. Boitano, S. W. Turner, C. He, J. Korlach, *BMC Biol.* **2013**, *11*, 4.
[51] L. Chavez, Y. Huang, K. Luong, S. Agarwal, L. M. Iyer, W. A. Pastor, V. K. Hench, S. A. Frazier-Bowers, E. Korol, S. Liu, M. Tahiliani, Y. Wang, T. A. Clark, J. Korlach, P. J. Pukkila, L. Aravind, A. Rao, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E5149 – E5158.
[52] a) C. de Lannoy, D. de Ridder, J. Risse, *F1000Res* **2017**, *6*, 1083; b) M. Jain, H. E. Olsen, B. Paten, M. Akeson, *Genome Biol.* **2016**, *17*, 239; c) D. Deamer, M. Akeson, D. Branton, *Nat. Biotechnol.* **2016**, *34*, 518 – 524; d) C. L. C. Ip, M. Loose, J. R. Tyson, M. de Cesare, B. L. Brown, M. Jain, R. M. Leggett, D. A. Eccles, V. Zalunin, J. M. Urban, P. Piazza, R. J. Bowden, B. Paten, S. Mwaigwisya, E. M. Batty, J. T. Simpson, T. P. Snutch, E. Birney, D. Buck, S. Goodwin, H. J. Jansen, J. O'Grady, H. E. Olsen, I. O. N. A. Min, C. Reference, *F1000Res* **2015**, *4*, 1075; e) A. H. Laszlo, I. M. Derrington, B. C. Ross, H. Brinkerhoff, A. Adey, I. C. Nova, J. M. Craig, K. W. Langford, J. M. Samson, R. Daza, K. Doering, J. Shendure, J. H. Gundlach, *Nat. Biotechnol.* **2014**, *32*, 829 – 833; f) E. A. Manrao, I. M. Derrington, A. H. Laszlo, K. W. Langford, M. K. Hopper, N. Gillgren, M. Pavlenok, M. Niederweis, J. H. Gundlach, *Nat. Biotechnol.* **2012**, *30*, 349 – 353; g) E. A. Manrao, I. M. Derrington, M. Pavlenok, M. Niederweis, J. H. Gundlach, *PLoS One* **2011**, *6*, e25723; h) J. Clarke, H. C. Wu, L. Jayasinghe, A. Patel, S. Reid, H. Bayley, *Nat. Nanotechnol.* **2009**, *4*, 265 – 270.
[53] A. H. Laszlo, I. M. Derrington, H. Brinkerhoff, K. W. Langford, I. C. Nova, J. M. Samson, J. J. Bartlett, M. Pavlenok, J. H. Gundlach, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 18904 – 18909.
[54] Z. L. Wescoe, J. Schreiber, M. Akeson, *J. Am. Chem. Soc.* **2014**, *136*, 16582 – 16587.
[55] a) J. T. Simpson, R. E. Workman, P. C. Zuzarte, M. David, L. J. Dursi, W. Timp, *Nat. Methods* **2017**, *14*, 407 – 410; b) A. C. Rand, M. Jain, J. M. Eizenga, A. Musselman-Brown, H. E. Olsen, M. Akeson, B. Paten, *Nat. Methods* **2017**, *14*, 411 – 413.
[56] S. Goodwin, J. D. McPherson, W. R. McCombie, *Nat. Rev. Genet.* **2016**, *17*, 333 – 351.
[57] a) L. Liu, C. Yang, K. Zhao, J. Y. Li, H. C. Wu, *Nat. Commun.* **2013**, *4*, 2989; b) M. Wanunu, D. Cohen-Karni, R. R. Johnson, L. Fields, J. Benner, N. Peterman, Y. Zheng, M. L. Klein, M. Drndic, *J. Am. Chem. Soc.* **2011**, *133*, 486 – 492; c) J. Shim, G. I. Humphreys, B. M. Venkatesan, J. M. Munz, X. Q. Zou, C. Sathe, K. Schulten, F. Kosari, A. M. Nardulli, G. Vasmatzis, R. Bashir, *Sci. Rep.* **2013**, *3*, 1389.

Manuscript received: January 20, 2019
Accepted manuscript online: February 27, 2019
Version of record online: May 22, 2019