



Recent advances in the detection of base modifications using the Nanopore sequencer

Liu Xu¹ · Masahide Seki¹

Received: 17 June 2019 / Revised: 22 September 2019 / Accepted: 26 September 2019 / Published online: 11 October 2019
© The Author(s) 2019. This article is published with open access

Abstract

DNA and RNA modifications have important functions, including the regulation of gene expression. Existing methods based on short-read sequencing for the detection of modifications show difficulty in determining the modification patterns of single chromosomes or an entire transcript sequence. Furthermore, the kinds of modifications for which detection methods are available are very limited. The Nanopore sequencer is a single-molecule, long-read sequencer that can directly sequence RNA as well as DNA. Moreover, the Nanopore sequencer detects modifications on long DNA and RNA molecules. In this review, we mainly focus on base modification detection in the DNA and RNA of mammals using the Nanopore sequencer. We summarize current studies of modifications using the Nanopore sequencer, detection tools using statistical tests or machine learning, and applications of this technology, such as analyses of open chromatin, DNA replication, and RNA metabolism.

Introduction

DNA and RNA modifications

More than 17 and 160 types of modified bases have been found in DNA and RNA, respectively [1, 2]. DNA modification plays roles in several biological processes, including development, aging, and cancer [3–5]. In vertebrates, m5C (5-methylcytosine) is frequently observed in CpG dinucleotides in DNA [3]. The methylation of cytosine is catalyzed by DNMTs (DNA methyltransferases). By methylating the CpGs of the newly synthesized strand complementary to methylated CpG, DNMT1 maintains the methylation of cytosine beyond DNA replication. DNMT3A and B are responsible for de novo methylation. CpG islands, where there is a high frequency of CpGs are predominately found in promoter regions. m5C at CpG islands in the promoters represses gene expression [3]. m5C is also important for the repression of transposable elements [6].

MBDs (Methyl-CpG-binding domain proteins) bind to methylated CpG sites, and recruit histone-modifying enzymes and chromatin remodeling complexes to organize the repressive chromatin structure [7]. hm5C (5-hydroxymethylcytosine) is generated by the oxidation of 5mC catalyzed by TET (Ten–eleven translocation) enzymes [8]. hm5C is predominantly observed in embryonic stem cells and neurons. hm5C is known as an intermediate in demethylation via the TET–TDG pathway. Although the underlying mechanism remains elusive, it is suggested that the function of hm5C is important for the maintenance of pluripotency and alternative splicing.

RNA modifications are also important for biological processes including development and cancer [9, 10]. Non-coding RNA, such as ribosomal RNA and transfer RNA, harbor several types of RNA modifications, and these modifications are indispensable for their proper function [11]. m6A (N6-methyladenine) is the most common modification found in the mRNA of eukaryotes and is catalyzed by METTL3, METTL14, and WTAP [11]. m6A is recognized by m6A-binding proteins, such as YTHDC1 and YTHDF3, and regulates mRNA processing, stability, translation, and transportation [12–15]. A-to-I (adenine to inosine) editing is also frequent in the RNA of higher eukaryotes [11]. A-to-I editing is catalyzed by ADAR (adenosine deaminase acting on RNA) enzymes. Due to the base pairing of inosine with cytosine, inosine is recognized

✉ Masahide Seki
mseki@edu.k.u-tokyo.ac.jp

¹ Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan

as guanosine during splicing and translation. Inosine is converted to guanosine via cDNA synthesis and is detected as A-to-G mismatch through sequencing.

Nanopore sequencing

The Nanopore sequencer released by Oxford Nanopore Technologies, performs DNA/RNA sequencing directly and in real time. The Nanopore sequencer carries out sequencing by predicting sequences from electric current patterns known as “squiggle”, which are affected by the bases inside the Nanopore. Oxford Nanopore Technologies has released several Nanopore sequencing instruments, including the MinION and PromethION sequencers [16, 17]. The MinION is a palm-top-sized sequencer. Taking advantage of the portability and real-time sequencing ability of MinION, this system has been applied in various situations, including field research and clinical diagnosis [18, 19]. The MinION flowcell has 512 pores in which sequencing can be performed at the same time [15]. The maximum throughput of MinION is 10–30 Gb per single MinION flowcell based on the manufacturer’s data. In a previous human genome study by our group, MinION stably generated over 3 Gb of DNA sequences in a single run [20]. The maximum length of its reads is greater than 2 Mb [21]. PromethION is a high-throughput stationary sequencer. PromethION can run up to 48 flowcells in parallel [16]. The PromethION flowcell has 3000 pores in which sequencing can be performed at the same time. For DNA sequencing, 158 Gb of data can be generated per single PromethION flowcell based on the

manufacturer’s data [16]. Therefore, PromethION can generate 7.6 Tb of sequence data. The average read length of human genome sequences without size selection is more than 10 kb for both the MinION and PromethION [20]. Although the throughput is becoming comparable with that of short-read sequencing, the accuracy of Nanopore sequencing (83–95%, depending on the library type) is lower than that of short-read sequencing using the Illumina platforms (99.9%) [22, 23].

Existing short-read sequencing-based methods for modified base detection have difficulty in determining the modification patterns of a chromosome or transcript and cannot detect modifications without an immunoprecipitation-grade antibody, distinguishable base conversion method, or the introduction of reverse transcription errors by modified bases [2, 24]. These detection methods are available for very limited kinds of modified bases. The Nanopore sequencer can read native single DNA and RNA molecules harboring base modifications without PCR amplification or cDNA conversion [25–27]. Nanopore sequencing can detect modified bases according to differences in squiggles between modified and unmodified base (Fig. 1a). The Nanopore sequencer is applied to detect several kinds of modified bases, such as m5C and m6A in DNA and m6A in RNA, in long single nucleic acid molecules [25, 26]. Modified bases can be detected based on significantly unique squiggle patterns. For the detection of modifications, some tools in which machine learning and statistical tests are applied are available at present. Recently, application methods have been developed in addition to simply detecting modifications. Here, we describe

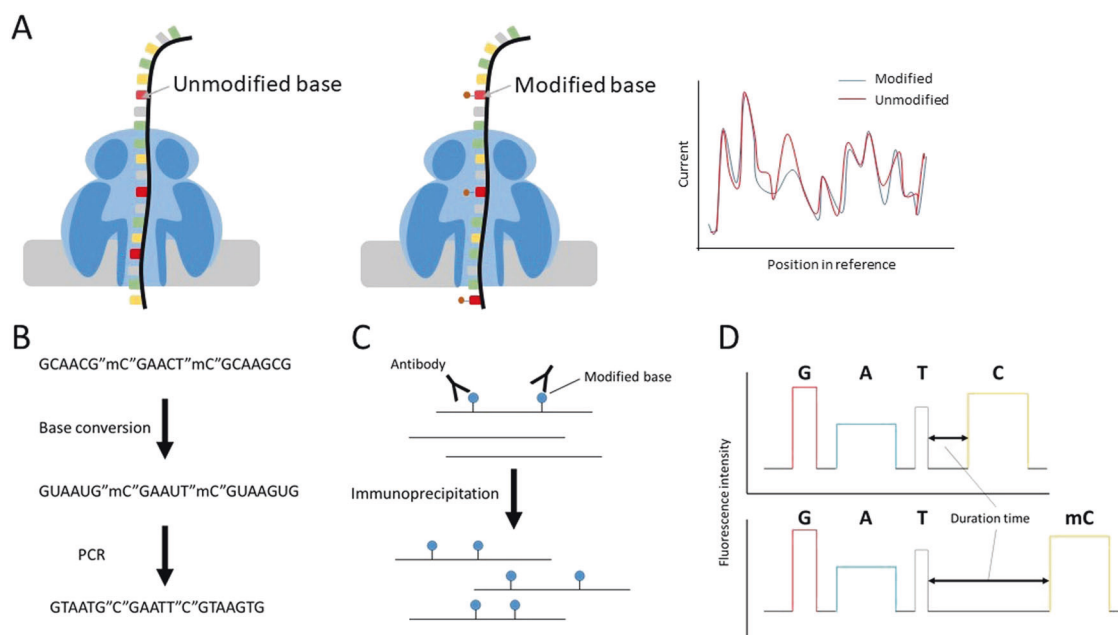


Fig. 1 Modified base detection using Nanopore sequencing and general methods. Schema of modified base detection using the Nanopore sequencer (**a**) and through bisulfite conversion (**b**), immunoprecipitation of nucleic acids (**c**), and SMRT sequencing (**d**)

recent advances in the detection of base modifications using the Nanopore sequencer.

General methods for base modification detection

Methods using short-read sequencing

For the detection of m5C and hm5C on DNA, BS-seq (bisulfite sequencing) approaches such as WGBS (whole genome bisulfite sequencing) and RRBS (reduced representation bisulfite sequencing), based on bisulfite conversion and subsequent short-read sequencing, are the most commonly applied methods [28] (Fig. 1b). While the whole genome is analyzed by WGBS, CpG-rich regions are mainly analyzed by RRBS [29]. In RRBS, a short-read sequencer preferentially reads CpG-rich short fragments generated by digestion using the restriction enzyme MspI, which recognizes the “CCGG” sequence motif. To discriminate m5C and hm5C from umC (unmethylated cytosine), umC is converted to uracil via bisulfite conversion, and uracils are replaced with thymines in the subsequent PCR step. Although BS-seq is unable to discriminate between m5C and hm5C, other bisulfite conversion-based methods, such as oxBS-seq (oxidative bisulfite sequencing) or TAB-seq (Tet-assisted bisulfite sequencing), are able to detect only m5C or hm5C, respectively [30, 31]. Bisulfite conversion is also used to detect m5C on RNA [32]. Because bisulfite conversion is a chemically severe reaction, a substantial proportion of nucleic acids are fragmented, and analyzing the methylation patterns of long nucleotides by combining this method and long-read sequencing is therefore difficult [33]. Recently, nondestructive methods for base conversion other than bisulfite conversion have been reported by some groups [34–36]. The EM-seq (enzymatic methyl-seq) method has been released by New England Biolabs, Inc., and ACE-seq (APOBEC-coupled epigenetic sequencing) employs the cytidine deaminase APOBEC. EM-seq is a detection method for m5C and hm5C [34]. After the oxidation of m5C and hm5C to ca5C (5-carboxylcytosine) for protection from deaminase, umC is converted to uracil by APOBEC. Similar to BS-seq, umC is sequenced as a T base. ACE-seq is a detection method for hm5C [35]. After the glucosylation of hm5C for protection, m5C and unmethylated cytosines are converted to uracil by APOBEC3A. In TAPS (TET-assisted pyridine borane sequencing), after the oxidation of m5C and hm5C to ca5C, ca5C is converted to dihydrouracil via pyridine borane reduction [36]. Therefore, in contrast to BS-seq and EM-seq, m5C and hm5C are sequenced as T bases in TAPS. Longer DNA can be obtained using these methods than by bisulfite conversion. Although it is unclear whether uracil or

dihydrouracil can be sequenced by long-read sequencers, these methods may be applicable for long-read sequencing.

For the detection of m6A, methods based on the short-read sequencing of RIP (RNA immunoprecipitation) samples, such as MeRIP-seq (methylated RIP sequencing) and m6A-seq, are the most prevalent [37, 38] (Fig. 1c). RIP-based methods rely on the specificity and affinity of an antibody that recognizes m6A. However, MeRIP-seq and m6A-seq are unable to reveal the specific position of the m6A. miCLIP (m6A individual nucleotide resolution crosslinking and immunoprecipitation) can identify the specific m6A position [39]. miCLIP involves the UV-crosslinking of an anti-m6A antibody to m6A sites and truncation or misincorporation of C bases at m6A sites during reverse transcription caused by crosslinking. This method requires short fragmented RNA as the input. Therefore, accurate estimation of expression and the full-length structure of RNA with m6A is difficult. In m6A-LAIC-seq (m6A-level and isoform-characterization sequencing), unfragmented RNA is used as the input for immunoprecipitation [40]. This method can detect structural isoforms of transcripts similar to typical RNA-seq. For the detection of DNA modifications, immunoprecipitation-based approaches, such as methylated DNA immunoprecipitation sequencing for m5C, are also used [41]. However, immunoprecipitation-based methods are limited by bias based on the immunoprecipitation efficiency.

SMRT (single molecule, real-time) sequencing

SMRT sequencing with long-read sequencers released by Pacific Bioscience can detect DNA modifications. In the PacBio Sequel sequencer, SMRT cell harboring a million wells is used for SMRT sequencing. The complex of a DNA polymerase and a single DNA molecule is fixed to each well of SMRT cell. SMRT sequencing is performed by continuously observing the incorporation of a fluorescently labeled nucleotide by DNA polymerase in each well. Approximately 365,000 reads with an average length of 10–14 kb are generated by the PacBio Sequel sequencer per an SMRT cell [42]. On the basis of differences in the duration of nucleotide incorporation between unmodified and modified bases, SMRT sequencing can detect the positions and kinds of base modifications (Fig. 1d). SMRT sequencing has been applied to detect various DNA modifications, such as m5C, m6A, and hm5C [43]. The detectability of a modification depends on the magnitude of the effect of the modification on the kinetics of the polymerase. Because the methyl group of m5C does not directly contribute to base pairing, m5C shows a subtler effect on nucleotide incorporation than m6A [44]. Therefore, relatively high-coverage sequencing or oxidation of m5C to ca5C using TET1 for signal enhancement is required for the

detection of m5C [45]. Minimum 25X and 250X coverage are recommended for the detection of m6A and m5C [46]. Although 6mA in DNA is responsible for biological functions in bacteria, its function is unclear and it is very rare in metazoan [47]. It is suggested that SMRT sequencing overestimates 6mA in DNA in which m6A is rare. SMRT-BS (SMRT sequencing of bisulfite-converted DNA) using gene-specific primers has also been developed [48]. However, the read length of SMRT-BS is up to ~1.5 kbp and SMRT-BS is not applied for genome-wide sequencing. For RNA, Iso-Seq, which is a SMRT sequencing method for full-length cDNA, can be used for the determination of entire sequences of transcript isoforms. However, due to the requirement for full-length cDNA synthesis, RNA modification information is lost. SMRT sequencing is also compatible with direct sequencing of RNA molecules using reverse transcription [49]. Vilfan et al. used this approach for the detection of m6A and RNA secondary structure. However, the detection of RNA modifications using SMRT sequencing has only rarely been performed in other studies.

Modified base detection using the Nanopore sequencer

Genome sequencing

Whole-genome sequencing and methylation calling using Nanopore sequencing has been performed in mammals, including humans and mice [26, 50, 51]. Jain et al. conducted whole-genome sequencing of a human cell line at approximately 30X coverage using MinION [50]. These authors compared CpG methylation frequencies on chromosome 20 called from Nanopore sequencing data using *Nanopolish* and short-read-based BS-seq. The data showed high correlations, with $r=0.895$. Gigante et al. conducted whole-genome sequencing of F1 B6 and Cast strain hybrid mice with ~10X coverage using MinION and PromethION [51]. They called the CpG methylation status using *Nanopolish* and compared the data with those obtained from RRBS with an Illumina sequencer. In addition, the authors conducted haplotyping of the Nanopore and Illumina reads utilizing SNPs between the B6 and Cast strains and showed that a larger fraction of the Nanopore reads were assigned compared with the Illumina reads. Known and novel imprinting control regions were detected by Nanopore sequencing.

To analyze the modification of target regions without performing whole-genome sequencing, native DNA must be enriched without PCR. Methods involving Cas9-mediated enrichment have been developed to enrich a target region [52, 53]. Gilpatrick et al. developed nCATS (nanopore Cas9-targeted sequencing), which is a simple target enrichment method [52]. In nCATS, all of the 5' ends of the DNA are

dephosphorylated to prevent adapter ligation to unwanted DNA. After dephosphorylation, the Cas9 enzyme and a guide RNA designed for both ends of the target region are applied to cut the target region. The Nanopore sequencing adapters are specifically ligated to the 5' phosphate of the target DNA generated by Cas9. Gilpatrick et al. applied nCATS to ten loci of ~18 kb in length from three human cell lines and obtained a median coverage of 165X at the ten loci from a MinION flowcell. They used the enriched reads for CpG methylation calling and the detection of structural variants.

Direct RNA-seq

Nanopore sequencing enables the direct sequencing of RNA molecules, which is referred to as direct RNA-seq [25]. Direct RNA-seq reveal full-length RNA structures and modifications simultaneously. Because direct RNA-seq is a PCR and reverse transcription-free method, it shows less bias than short-read-based RNA-seq and Nanopore cDNA-seq, which use reverse transcription and PCR amplification of cDNA. Garalde et al. showed differences in the squiggles associated with m6A and m5C compared with those of unmodified bases using synthetic RNA [25]. Smith et al. conducted direct RNA-seq of 16S rRNA in a wild type *Escherichia coli* strain, two mutant strains lacking a guanine methyltransferase or a pseudouridine synthases and a strain expressing another guanine methyltransferase [54]. They indicated that alternation of current signals and base calling error were observed around position where 7-methylguanosine and pseudouridine occurred. Workman et al. performed direct RNA-seq analysis of RNA from a human cell line, in vitro transcribed RNA from cDNA from the same cell line, and synthetic RNA [55] with a focus on the m6A methyltransferase-binding motif. They detected current differences for the motif and validated the differences in signals using data from the synthetic RNA. Utilizing the current difference, they detected m6A-modified motif in 57 genes. Furthermore, they attempted to detect A-to-I editing. They showed systematic miscalling and a change in the ion current for direct RNA-seq reads and A-to-G mismatch in cDNA-seq reads at A-to-I sites. They also measured polyA tail length using the polyA option of *Nanopolish* and showed differences in polyA tail length both between genes and between transcripts of the same gene. Finally, Viehweger et al. performed direct RNA-seq analysis of human coronavirus and detected m5C in the viral RNA using *Tombo* [56, 57].

Tools for modified base detection

Tools for the detection of modifications based on several principles have been released (Table 1). In most tools

Table 1 Tools for modified base detection

Tool	Target	URL
Nanopolish	DNA	https://nanopolish.readthedocs.io/en/latest/index.html
SignalAlign	DNA	https://github.com/ArtRand/signalAlign
mCaller	DNA	https://github.com/al-mcintyre/mCaller
DeepSignal	DNA	https://github.com/bioinformaticsCSU/depsignal
DeepMod	DNA	https://github.com/WGLab/DeepMod
NanoMod	DNA	https://github.com/WGLab/NanoMod
Tombo	DNA/RNA	https://github.com/nanoporetech/tombo
ELIGOS	RNA	https://bitbucket.org/piroonj/eligos.git
EpiNano	RNA	https://github.com/enovoa/EpiNano/issues
Flappy	DNA	https://github.com/nanoporetech/flappie
Taiyaki	DNA	https://github.com/nanoporetech/taiyaki

except for the base calling-based tools, the reads must first be aligned to the reference sequence using alignment software, such as *minimap2* [58]. For tools utilizing current signals, the current intensity is associated with the aligned base position using tools such as *resquiggle* in *Tombo* and *eventalign* in *Nanopolish* [26, 57] (Fig. 2a). A modified base is detected based on comparison with the training model or the squiggle of an unmodified nucleic acid. *Nanopolish*, which is based on the HMM (hidden Markov model), can call m5Cs on DNA in the CpG context [26]. *Nanopolish* is widely used for m5C calling in various species, including humans [50]. *SignalAlign* is based on the HMM with the hierarchical Dirichlet process [27]. Rand et al. detected m5C, hm5C, and m6A in *E. coli* DNA using *SignalAlign*. They showed that the detection of m5C was easier than that of m6A, contrary to the situation in SMRT sequencing. *mCaller* can use the four machine learning classifiers (neural network, random forest, logistic regression, and naive Bayes classifiers) to detect m6A on DNA [59]. McIntyre et al. showed that the predictor using the neural network was the most accurate. They also compared the detection of m6A on DNA using *mCaller* and *Tombo*. They showed that which tool is better depends on the sequence motif. *DeepSignal* and *DeepMod* detect m5C and m6A in DNA using neural networks [60, 61]. Ni et al. compared *DeepSignal* with *Nanopolish* and *SignalAlign* for the detection of m5C or m5C and m6A, respectively [60]. *DeepSignal* showed a higher performance, in terms of both accuracy and sensitivity than the other two tools. *DeepSignal* can detect m5C at CpGs with ~0.92 accuracy at 1X coverage and m6A at GATC motif with ~0.90 accuracy at 2X coverage. *DeepMod* showed precision up to 0.99 and 0.9 for m5C and m6A, respectively [61]. In the detection of m6A, *DeepMod* showed slightly higher performance than *mCaller*. Statistical test-based tools can detect de novo

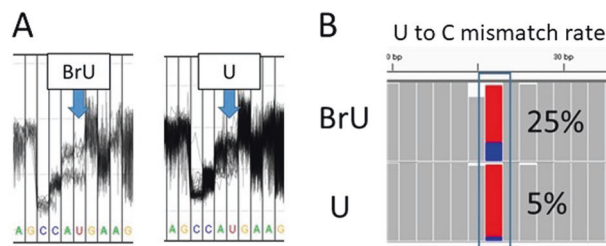


Fig. 2 Tools for modified base detection using Nanopore sequencing. Differences in the squiggles (a) and mismatch rates (b) of an artificial bromouridine base are shown. The direct RNA-seq reads of synthetic RNA were aligned using *minimap2* [58]. The squiggle plot was drawn using *Tombo* [57]. The aligned reads were visualized using *IGV* [79]

modifications without training using modified and unmodified samples. *Nanoraw* employs the Mann–Whitney U test [57]. *Tombo* is the successor of *Nanoraw* and supports the detection of m5C on DNA and RNA and m6A on DNA in all sequence contexts without a requirement for unmodified samples in addition to de novo detection. *NanoMod* uses the Kolmogorov–Smirnov test [62]. Liu et al. showed that *NanoMod* presented higher performance than *Nanoraw* for the detection of m5C in *E. coli*.

Due to differences in the current signal between modified and unmodified bases, there is a tendency for higher error rates to be observed at modified than at unmodified bases (Fig. 2b). Wongsurawat et al. developed *ELIGOS* (epi-transcriptional landscape inferring from glitches of ONT signals), which is a tool for the detection of RNA modifications [63]. *ELIGOS* detects RNA modifications by comparing the % error at a specific base between direct RNA-seq analyses of modified and unmodified RNA or direct cDNA-seq analysis, in which the modification information is erased during reverse transcription. They validated *ELIGOS* through the detection of known modifications in *Saccharomyces cerevisiae*, *E. coli*, and human cell line ribosomal RNA. The authors applied *ELIGOS* for the detection of modification from *S. cerevisiae* transcriptome data and to the detection of m6A and G-quadruplex (a type of secondary structure) from human cell line transcriptome data. *EpiNano* is a tool for the detection of the m6A modification of RNA [64]. *EpiNano* distinguishes modified bases by support vector classification using quality scores, mismatch patterns of base-called sequences, and raw current signals as input data. *EpiNano* detect the m6A modification in vivo with an accuracy of 0.87. Although *ELIGOS* and *EpiNano* can detect sites with modifications from bulk reads, they are not compatible with detection of modified bases in single reads at present.

Although the tools described thus far require base-called sequence data, some base callers can base call modified bases directly. *Flappie*, which was released by Oxford Nanopore Technologies, is a base caller based on

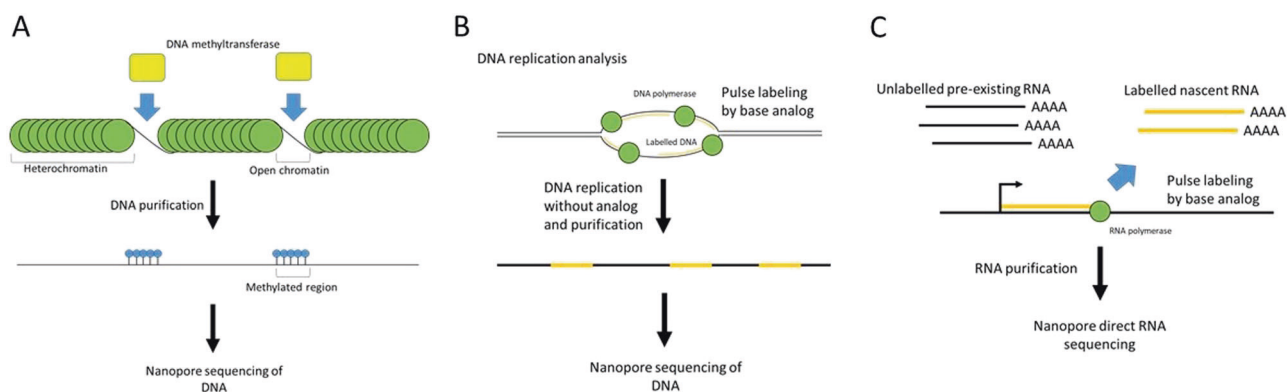


Fig. 3 Applied methods for modified base detection using Nanopore sequencing. **a** Detection of open chromatin regions using DNA methyltransferase accessibility. DNA in open chromatin regions is specifically methylated. Open chromatin regions are detected as highly methylated regions by Nanopore sequencing. **b** Detection of DNA replications using base analogs. The replicated regions are labeled by

pulse labeling during S phase. The replicated regions are detected as labeled regions by Nanopore sequencing. **c** Detection of nascent RNA using base analogs. Nascent RNAs are labeled by pulse labeling using base analogs. Nascent and preexisting RNAs are classified by the existence of base analogs

the ‘flip-flop’ algorithm [65]. m5C in the CpG context of DNA is called as the an extra base in addition to the four standard bases by *Flappie*. *Taiyaki* is a model training tool for the Nanopore base caller that can be used in the model training of modified bases [66]. The model trained on *Taiyaki* is usable in *Guppy*, which is a Nanopore base caller. These base callers calling modified bases have not been benchmarked thus far.

As described above, many detection tools for modified bases have become available recently. In the detection of m5C and m6A on DNA, the tools based on neural networks tend to show higher performance than other tools [60, 61]. For the detection of de novo modifications, statistical test-based tools is applicable [57, 62]. Because comparison results are highly dependent on comparison conditions, such as the sequence context and datasets, fair comparison by a third party is necessary for accurate assessment.

Application of modified base detection by the Nanopore sequencer

Analysis of open chromatin regions

DNase-seq, ATAC-seq, and NOMe-seq are short-read sequencing-based methods for the detection of open chromatin regions not protected by nucleosomes. Whereas DNase-seq and ATAC-seq utilize the accessibility of a DNase or transposase to open chromatin and DNA fragmentation by these enzymes, NOMe-seq utilizes the accessibility of a GpC-specific methyltransferase (M.CviPI) to open chromatin and detects open chromatin regions as methylated GpC regions, using BS-seq [67]. nanoNOMe, MeSMLR-seq, and SMAC-seq are methods that combine the accessibility of one or more methyltransferases and the

direct detection of methylated regions by the Nanopore sequencer instead of requiring BS-seq [68–70]. These methods can detect open chromatin patterns of single long DNA molecules (Fig. 3a). Although nanoNOMe and MeSMLR-seq employ a GpC methyltransferase similar to NOMe-seq, SMAC-seq employs a m6A methyltransferase (EcoGII) and CpG-specific methyltransferase (M.SssI) in addition to the GpC methyltransferase to increase the resolution [68]. Shipony et al. and Wang et al. applied SMAC-seq and MeSMLR-seq to *S. cerevisiae*, respectively [68, 70]. In SMAC-seq, *Tombo* is used for methylation detection [68]. Lee et al. applied nanoNOMe to four human cell lines and detected DNA methylation using *Nanopolish* [69]. Furthermore, they investigated not only open chromatin patterns but also CpG methylation from single reads and showed an anticorrelation between chromatin accessibility and DNA methylation.

Detection of DNA replication using base analogs

For genome-wide analysis of DNA replication, pulse labeling of replication sites by base analogs, such as BrdU (bromodeoxyuridine), is widely used [71]. Attempts to detect base analogs directly using a Nanopore sequencer have been reported by some groups [72–74] (Fig. 3b). Müller et al. developed D-Nascent (detecting nucleotide analog signal currents on extremely long Nanopore traces) using pulse labeling by BrdU with relatively low toxicity [72]. They applied D-Nascent to analyze replication fork dynamism in *S. cerevisiae* and achieved the detection of replication sites in reads 20–160 kb in length. They also developed a BrU detection tool based on the HMM for D-Nascent. Hennion et al. used BrU for the detection of DNA replication in *S. cerevisiae* [73]. They developed RepNano, which is a deep learning-based software, for BrU detection.

However, *RepNano* is not available at present. Georgieva et al. compared the differences in signals between thymidine and 11 species of thymidine analogs containing BrdU [74]. To detect base analogs in mouse embryonic stem cells, they employed IdU (iododeoxyuridine), which showed the greatest signal difference within a 5–6 bp k-mer used for base calling in Nanopore sequencing. They employed *NanoMod* for IdU detection.

Measurement of RNA metabolism using base analogs

RNA abundance measured via typical transcriptome analysis is the result of RNA synthesis and degradation together. Therefore, to understand the regulatory mechanism of gene expression, RNA synthesis and degradation must be measured. Measurements of RNA synthesis and degradation can be conducted through the metabolic labeling with RNA by base analogs, as in the GRO-seq, BRIC-seq, and SLAM-seq approaches [75–77]. Because these methods are based on the sequencing of short cDNA fragments using a short-read sequencer, determining the entire transcript structure is difficult. Combining RNA labeling with direct RNA-seq using the Nanopore sequencer, Maier et al. developed Nano-ID (nanopore sequencing-based isoform dynamics), which is a method for the measurement of RNA metabolism [78] (Fig. 3c). In Nano-ID, 5-EU (5-ethynyl uridine)-labeled newly synthesized RNA is detected by direct RNA-seq. The reads derived from labeled and unlabeled RNA are distinguished using a neural network. These authors conducted Nano-ID in a human cell line and identified labeled newly synthesized isoforms and unlabeled preexisting isoforms. Furthermore, they investigated RNA isoform metabolism following heat shock stress and the correlation between the polyA length and stability.

Conclusion

In this review, we summarized the approaches available for modified base detection using Nanopore sequencing. The emergence of Nanopore sequencers has enabled genome-wide and transcriptome-wide analyses of base modifications, including modifications for which detection methods are not available. For the detection of m6A and m5C of DNA, although high-coverage sequencing is required in SMRT sequencing, detection with a high accuracy is possible using low-coverage reads in nanopore sequencing. Many tools for the detection of modifications from Nanopore sequencing data have been developed. The systematic comparison of these tools is not sufficient. The accumulation of data on modified bases from Nanopore sequencing and the further development and maturation of the

associated software will allow the detection of various modifications simultaneously. Applications involving labeling with artificial bases or artificial modification are being developed. These methods reveal the relationships of natural modifications and signatures of labeled nucleotides. Nanopore sequencing will reveal detailed regulatory mechanisms via epigenome and epitranscriptome analyses.

Acknowledgements Our studies were supported by JSPS KAKENHI (Grant Numbers 19K16108 and 16H06279 (PAGS)).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Boccaletto P, Machnicka MA, Purta E, Piątkowski P, Bagiński B, Wirecki TK, et al. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* 2018;46:D303–7.
2. Raiber E-A, Hardisty R, van Delft P, Balasubramanian S. Mapping and elucidating the function of modified bases in DNA. *Nat Rev Chem.* 2017;1:0069.
3. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet.* 2013;14:204–20.
4. Field AE, Robertson NA, Wang T, Havas A, Ideker T, Adams PD. DNA methylation clocks in aging: categories, causes, and consequences. *Mol Cell.* 2018;71:882–95.
5. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet.* 2007;8:286–98.
6. Deniz Ö, Frost JM, Branco MR. Regulation of transposable elements by DNA modifications. *Nat Rev Genet.* 2019;20:417–31.
7. Du Q, Luu PL, Storzaker C, Clark SJ. Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics.* 2015;7:1051–73.
8. Shi DQ, Ali I, Tang J, Yang WC. New Insights into 5hmC DNA Modification: generation, distribution and function. *Front Genet.* 2017;8:100.
9. Delaunay S, Frye M. RNA modifications regulating cell fate in cancer. *Nat Cell Biol.* 2019;21:552–9.
10. Frye M, Harada BT, Behm M, He C. RNA modifications modulate gene expression during development. *Science.* 2018;361:1346–9.

11. Li X, Xiong X, Yi C. Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat Methods*. 2017;14:23.
12. Xiao W, Adhikari S, Dahal U, Chen Y-S, Hao Y-J, Sun B-F, et al. Nuclear m6A reader YTHDC1 regulates mRNA splicing. *Mol Cell*. 2016;61:507–19.
13. Shi H, Wang X, Lu Z, Zhao BS, Ma H, Hsu PJ, et al. YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. *Cell Res*. 2017;27:315–28.
14. Roundtree IA, Luo G-Z, Zhang Z, Wang X, Zhou T, Cui Y, et al. YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. *Elife*. 2017;6:e31311.
15. Li A, Chen Y-S, Ping X-L, Yang X, Xiao W, Yang Y, et al. Cytoplasmic m6A reader YTHDF3 promotes mRNA translation. *Cell Res*. 2017;27:444–7.
16. PromethION. Accessed 28 May 2019. <https://nanoporetech.com/products/promethion>.
17. MinION. Accessed 6 Jun 2019. <https://nanoporetech.com/products/minion>.
18. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature*. 2016;530:228–32.
19. Trotter AJ, Aydin A, Strinden MJ, O'Grady J. Recent and emerging technologies for the rapid diagnosis of infection and antimicrobial resistance. *Curr Opin Microbiol*. 2019;51:39–45.
20. Sakamoto Y, Xu L, Seki M, Yokoyama TT, Kasahara M, Kashima Y, et al. Long read sequencing reveals a novel class of structural aberrations in cancers: identification and characterization of cancerous local amplifications. *bioRxiv*. 2019;620047.
21. Payne A, Holmes N, Rakyen V, Loose M BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics*. 2019;35:2193–8.
22. Seki M, Katsumata E, Suzuki A, Sereewattanawoot S, Sakamoto Y, Mizushima-Sugano J, et al. Evaluation and application of RNA-Seq by MinION. *DNA Res*. 2019;26:55–65.
23. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17:333–51.
24. Jonkhout N, Tran J, Smith MA, Schonrock N, Mattick JS, Novoa EM. The RNA modification landscape in human disease. *RNA*. 2017;23:1754–69.
25. Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M, et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods*. 2018;15:201–6.
26. Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods*. 2017;14:407–10.
27. Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akesson M, et al. Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods*. 2017;14:411–3.
28. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009;462:315–22.
29. 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.
30. Booth MJ, Ost TWB, Beraldi D, Bell NM, Branco MR, Reik W, et al. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat Protoc*. 2013;8:1841–51.
31. Yu M, Hon GC, Szulwach KE, Song C-X, Zhang L, Kim A, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*. 2012;149:1368–80.
32. Schaefer M, Pollex T, Hanna K, Lyko F. RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res*. 2008;37:e12.
33. Tanaka K, Okamoto A. Degradation of DNA by bisulfite treatment. *Bioorg Med Chem Lett*. 2007;17:1912–5.
34. Sexton BS, Williams L, Ponnaluri VKC, Helbert Y, Zhang L, Sumner CJ, et al. Uncovering the Cannabis sativa methylome through enzymatic methyl-seq. Accessed 15 May 2019. https://www.neb.com/-/media/posters/document/mg_poster_011019_final.pdf.
35. Schutsyky EK, DeNizio JE, Hu P, Liu MY, Nabel CS, Fabyanic EB, et al. Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. *Nat Biotechnol*. 2018;36:1083–90.
36. Liu Y, Siejka-Zielińska P, Velikova G, Bi Y, Yuan F, Tomkova M, et al. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat Biotechnol*. 2019;37:424–9.
37. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*. 2012;485:201–6.
38. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*. 2012;149:1635–46.
39. Linder B, Grozhik AV, Olarerin-George AO, Meydan C, Mason CE, Jaffrey SR. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods*. 2015;12:767–72.
40. Molinie B, Wang J, Lim KS, Hillebrand R, Lu Z, Van Wittenberghe N, et al. m6A-LAIC-seq reveals the census and complexity of the m6A epitranscriptome. *Nat Methods*. 2016;13:692–8.
41. Down TA, Rakyen VK, Turner DJ, Flicek P, Li H, Kulesha E, et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol*. 2008;26:779–85.
42. Ardui S, Ameer A, Vermeesch JR, Hestand MS. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Res*. 2018;46:2159–68.
43. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*. 2010;7:461–5.
44. Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, et al. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res*. 2012;40:e29.
45. Clark TA, Lu X, Luong K, Dai Q, Boitano M, Turner SW, et al. Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. *BMC Biol*. 2013;11:4.
46. Pacific Bioscience. Detecting DNA base modifications using single molecule, real-time sequencing. https://www.pacb.com/wp-content/uploads/2015/09/WP_Detecting_DNA_Base_Modifications_Using_SMRT_Sequencing.pdf. Accessed 16 Sep 2019.
47. O'Brown ZK, Boulias K, Wang J, Wang SY, O'Brown NM, Hao Z, et al. Sources of artifact in measurements of 6mA and 4mC abundance in eukaryotic genomic DNA. *BMC Genom*. 2019;20:445.
48. Yang Y, Sebra R, Pullman BS, Qiao W, Peter I, Desnick RJ, et al. Quantitative and multiplexed DNA methylation analysis using long-read single-molecule real-time bisulfite sequencing (SMRT-BS). *BMC Genom*. 2015;16:350.
49. Vilfan ID, Tsai Y-C, Clark TA, Wegener J, Dai Q, Yi C, et al. Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *J Nanobiotechnol*. 2013;11:8.

50. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol*. 2018;36:338–45.
51. Gigante S, Gouil Q, Lucattini A, Keniry A, Beck T, Tinning M, et al. Using long-read sequencing to detect imprinted DNA methylation. *Nucleic Acids Res*. 2019;47:e46.
52. Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, et al. Targeted nanopore sequencing with Cas9 for studies of methylation, structural variants and mutations. *bioRxiv*. 2019;604173.
53. Gabrieli T, Sharim H, Fridman D, Arbib N, Michaeli Y, Ebenstein Y. Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic Acids Res*. 2018;46:e87.
54. Smith AM, Jain M, Mulrone L, Garalde DR, Akeson M. Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS ONE*. 2019;14:e0216709.
55. Workman RE, Tang AD, Tang PS, Jain M, Tyson JR, Zuzarte PC, et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *bioRxiv*. 2018;459529.
56. Viehweger A, Krautwurst S, Lamkiewicz K, Madhugiri R, Ziebuhr J, Hölzer M, et al. Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *bioRxiv*. 2019;483693.
57. Stoiber MH, Quick J, Egan R, Lee JE, Celniker SE, Neely R, et al. De novo Identification of DNA modifications enabled by genome-guided nanopore signal processing. *bioRxiv*. 2017;094672.
58. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–100.
59. McIntyre ABR, Alexander N, Grigorev K, Bezdán D, Sichtig H, Chiu CY, et al. Single-molecule sequencing detection of N6-methyladenine in microbial reference materials. *Nat Commun*. 2019;10:579.
60. Ni P, Huang N, Zhang Z, Wang D-P, Liang F, Miao Y, et al. DeepSignal: detecting DNA methylation state from Nanopore sequencing reads using deep-learning. *Bioinformatics*. 2019;btz276.
61. Liu Q, Fang L, Yu G, Wang D, Xiao C-L, Wang K. Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. *Nat Commun*. 2019;10:2449.
62. Liu Q, Georgieva DC, Egli D, Wang K. NanoMod: a computational tool to detect DNA modifications using Nanopore long-read sequencing data. *BMC Genomics*. 2019;20:78.
63. Wongsurawat T, Jenjaroenpun P, Wassenaar TM, Wadley TD, Wanchai V, Akel NS, et al. Decoding the epitranscriptional landscape from native RNA sequences. *bioRxiv*. 2018;487819.
64. Liu H, Begik O, Lucas MC, Mason CE, Schwartz S, Mattick JS, et al. Accurate detection of m6A RNA modifications in native RNA sequences. *bioRxiv*. 2019;525741.
65. Oxford Nanopore Technologies. Flappie. <https://github.com/nanoporetech/flappie>. Accessed 16 May 2019.
66. Oxford Nanopore Technologies. Taiyaki. <https://github.com/nanoporetech/taiyaki>. Accessed 16 May 2019.
67. 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.
68. Shipony Z, Marinov GK, Swaffner MP, Sinott-Armstrong NA, Skotheim JM, Kundaje A, et al. Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *bioRxiv*. 2018;504662.
69. Lee I, Razaghi R, Gilpatrick T, Molnar M, Sadowski N, Simpson JT, et al. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *bioRxiv*. 2019;504993.
70. Wang Y, Wang A, Liu Z, Thurman AL, Powers LS, Zou M, et al. Single-molecule long-read sequencing reveals the chromatin basis of gene expression. *Genome Res*. 2019;29:1329–42.
71. Gilbert DM. Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat Rev Genet*. 2010;11:673–84.
72. Müller CA, Boemo MA, Spingardi P, Kessler BM, Kriaucionis S, Simpson JT, et al. Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. *Nat Methods*. 2019;16:429–36.
73. Hennion M, Arbona J-M, Cruaud C, Proux F, Tallec B Le, Novikova E, et al. Mapping DNA replication with nanopore sequencing. *bioRxiv*. 2018;426858.
74. Georgieva D, Liu Q, Wang K, Egli D. Detection of base analogs incorporated during DNA replication by nanopore sequencing. *bioRxiv*. 2019;549220.
75. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*. 2008;322:1845–8.
76. Imamachi N, Tani H, Mizutani R, Imamura K, Irie T, Suzuki Y, et al. BRIC-seq: a genome-wide approach for determining RNA stability in mammalian cells. *Methods*. 2014;67:55–63.
77. Herzog VA, Reichholf B, Neumann T, Rescheneder P, Bhat P, Burkard TR, et al. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods*. 2017;14:1198–204.
78. Maier KC, Gressel S, Cramer P, Schwalb B. Native molecule sequencing by nano-ID reveals synthesis and stability of RNA isoforms. *bioRxiv*. 2019;601856.
79. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24–6.