bio-protocol

Published: Mar 05, 2022

Whole-genome Methylation Analysis of APOBEC Enzyme-converted DNA (~5 kb) by Nanopore Sequencing

Suzuko Zaha[1], Yoshitaka Sakamoto[1], Satoi Nagasawa[1], Sumio Sugano[2, 3], Ayako Suzuki[1], Yutaka Suzuki[1, *] and Masahide Seki[1, *]

[1]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan
[2]Institute of Kashiwa-no-ha Omics Gate, Kashiwa, Chiba, Japan
[3]Future Medicine Education and Research Organization, Chiba University, Chuo, Chiba-city, Chiba, Japan
*For correspondence: mseki@edu.k.u-tokyo.ac.jp; ysuzuki@hgc.jp

Abstract

In recent years, DNA methylation research has been accelerated by the advent of nanopore sequencers. However, read length has been limited by the constraints of base conversion using the bisulfite method, making analysis of chromatin content difficult. The read length of the previous method combining bisulfite conversion and long-read sequencing was ~1.5 kb, even using targeted PCR. In this study, we have improved read length (~5 kb), by converting unmethylated cytosines to uracils with APOBEC enzymes, to reduce DNA fragmentation. The converted DNA was then sequenced using a PromethION nanopore sequencer. We have also developed a new analysis pipeline that accounts for base conversions, which are not present in conventional nanopore sequencing, as well as errors produced by nanopore sequencing.

Keywords: DNA methylation, Epigenetics, Nanopore sequencing, Long-read sequencing, Next generation sequencing

This protocol was validated in: Nucleic Acids Res (2021), DOI:10.1093/nar/gkab397

Background

DNA methylation is an important mechanism for epigenetic regulation of gene expression (Greenberg and Bourc'his, 2019). It has a wide range of effects on genes via several biological processes. DNA methylation is usually detected and analyzed using bisulfite sequencing short reads. However, it is difficult to align these short reads (~150 bp) to some chromosomal regions, such as repetitive sequences and structural variants (Goerner-Potvin and Bourque, 2018). Similarly, short reads also constrain detection of chromosome-specific methylation patterns, such as imprinted regions, in polyploid organisms (Akbari *et al*., 2021). A comprehensive understanding of epigenetic regulation by DNA methylation will therefore require complementary methods.

The bisulfite method distinguishes between unmethylated and methylated cytosine (C vs. mC), by chemically converting unmethylated C to uracil (U) and to thymine (T), by subsequent amplification (Lister *et al*., 2009). However, since this reaction is carried out under chemically severe conditions, a large proportion of the DNA in the reaction is fragmented and degraded. The genomic regions that undergo this degradation show biased representation (Olova *et al*., 2018), further limiting the experimental conclusions this method can provide. The read length of the previous method combining bisulfite conversion and long-read sequencing was only ~1.5 kb, even using targeted PCR (Yan *et al*., 2015). Recently, enzymatic methyl sequencing (EMseq) was developed as an alternative to the bisulfite method of base conversion (Vaisvila *et al*., 2021). EMseq involves the oxidation of mC by ten-eleven translocation (TET) enzymes to protect them, followed by base conversion of unmethylated C to U, by APOBEC enzymes. Via the amplification process, U is converted to T, as in the bisulfite method. Because this reaction is performed under milder chemical conditions than with the bisulfite method, longer DNA fragments are obtainable. In fact, a previous study showed that DNA fragments over 5 kb long can be obtained using EMseq and target-specific PCR, and that these fragments can be successfully sequenced in a long-read sequencer (Sun *et al*., 2021).

Nanopore sequencers read nucleic acid sequences by measuring the change in electric current while the nucleic acids are passing through the nanopore. The maximum read length of nanopore sequencing is over 100 kb (Sakamoto *et al*., 2020). By recognizing specific electrical patterns for modified bases, base modifications can also be detected (Rand *et al*., 2017; Simpson *et al*., 2017). However, while the base-reading accuracy of nanopore sequencers is currently up to 90%, this is not quite high enough to accurately infer methylation patterns (Sakamoto *et al*., 2020). Furthermore, it requires about 500 ng–1 µg of DNA input, reducing its practical utility for rare samples, such as clinical specimens and biopsies. Although several methods combining base-conversion and long-read sequencing have been developed, thus far all have employed gene-specific amplification (Yang *et al*., 2015; Liu *et al*., 2020; Sun *et al*., 2021). A method for whole-genome methylation analysis by this method, and a bioinformatic pipeline to process the sequence data it generates, have not heretofore been developed.

Here, we report a method for whole-genome long-read methylation sequencing, using a relatively small amount of input DNA, for nanopore sequencing of base-converted DNA by APOBEC enzymes (Figure 1) (Sakamoto *et al*., 2021). Our method, which we designate nanoEM, allows for whole-genome long-read methylation analysis with 10–100 ng of DNA. In addition, we have developed a data analysis pipeline for nanoEM reads by adopting a three-letter alignment approach to long-read alignment. NanoEM is an useful approach for detecting methylation status of structural variants (SVs), repetitive regions, and imprinting regions, which are difficult to analyze using short read sequencing (Sakamoto *et al*., 2021).

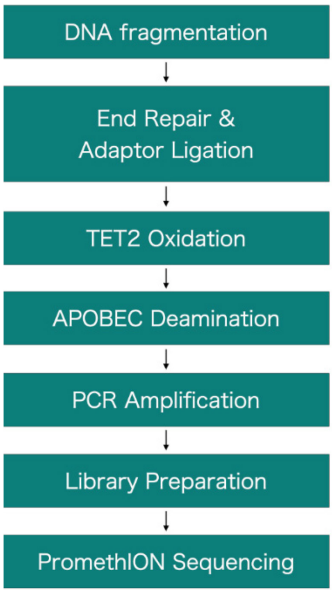

Figure 1. Flow chart of the experimental procedure.

Materials and Reagents

1. Filter pipette tips 10, 20, 200, and 1,000 µL [*e.g.*, Pipette Tips RT UNV F (RAININ, catalog numbers: 30389172, 30389189, 30389186, and 30389165)]
2. 1.5 mL tubes [*e.g.*, DNA LoBind Tube 1.5 mL (Eppendorf, catalog number: 0030108051)]
3. PCR tubes [*e.g.*, Temp Assure 0.2 mL PCR 8-Tube Strips, Att. Optical Caps (USA Scientific, catalog number: 1402-4700)]
4. Mag Attract HMW DNA Kit (QIAGEN, catalog number: 67563)
5. g-TUBE (Covaris, catalog number: 520079)
6. Ethanol (*e.g.*, FUJIFILM WAKO Pure Chemical Corporation, catalog number: 057-00456)
7. Nuclease-free water (Thermo Fisher Scientific, catalog number: AM9930)
8. Formamide (FUJIFILM Wako Pure Chemical Corporation, catalog number: 064-00423)
9. NEBNext Enzymatic Methyl-seq kit (New England Biolabs, catalog number: E7120S)
10. KOD One PCR Master Mix (TOYOBO, catalog number: KMM-101)
11. Ligation Sequencing kit (Oxford Nanopore Technologies, catalog number: SQK-LSK110)
12. PromethION flowcell (Oxford Nanopore Technologies, catalog number: FLO-PRO002)
13. NEBNext Ultra II End Repair/dA-Tailing Module (New England Biolabs, catalog number: E7546)
14. NEBNext FFPE DNA Repair Mix (New England Biolabs, catalog number: M6630)
15. NEBNext Quick Ligation Module (New England Biolabs, catalog number: E6056)
16. Qubit ds DNA HS Assay kit (Thermo Fisher Scientific, catalog number: Q32854)
17. Agilent DNA 12000 kit (Agilent Technologies, catalog number: 5067-1508)
18. Agencourt AMPure XP (Beckman Coulter, catalog number: BC-A63880)
19. DNA Clean & Concentrator-5 (Zymo Research, catalog number: D4013)
20. ProNex Size-Selective DNA Purification System (Promega, catalog number: NG2001)
21. TET2 Reaction Buffer with supplement (see Recipes)
22. 70% and 80% (v/v) ethanol (see Recipes)
23. Wash Buffer of ProNex Size-Selective DNA Purification System (NG2001) (see Recipes)

Equipment

1. PromethION sequencing device (Oxford Nanopore Technologies, catalog number: PRM48BasicSP)
2. 2100 Bioanalyzer Instrument (Agilent Technologies, catalog number: G2939BA)
3. Thermal cycler [*e.g.*, T100 thermal cycler (Bio-Rad, catalog number: 1861096)]
4. Qubit 4 Fluorometer (Thermo Fisher Scientific, catalog number: Q33238)
5. Vortex Mixer [*e.g.*, Vortex-Genie 2 (Scientific Industries, catalog number: SI-0236)]
6. High speed centrifuge (*e.g.*, MDX-310 with rack for 2 mL × 24 tubes, TOMY SEIKO)
7. Tabletop centrifuge for 1.5 and 0.2 mL tubes [*e.g.*, MyFuge mini centrifuge (Benchmark Scientific, catalog number: C1008-B)]
8. Magnetic stand for 1.5 and 2 mL tubes [*e.g.*, DynaMag-2 (Thermo Fisher Scientific, catalog number: 12321D)]
9. Magnetic stand for 0.2 mL tubes [*e.g.*, 10× Magnetic Separator (10× Genomics, catalog number: 120250)]
10. Pipettes for 10, 20, 200, and 1,000 µL tips
11. Racks for 0.2 mL PCR tubes and 1.5 mL tubes

Software

1. Python3 (version 3.8.6, https://www.python.org/downloads/)
2. pysam (version 0.17.0, https://github.com/pysam-developers/pysam)
3. minimap2 (version 2.22) (Li, 2018)
4. sambamba (version 0.7.1) (Tarasov *et al*., 2015)
5. samtools (version 1.9) (Li *et al*., 2009)
6. Integrated Genome Viewer (IGV) (version 2.5.3) (Thorvaldsdóttir *et al*., 2013)

Procedure

A. DNA Extraction

The MagAttract HMW DNA Kit is used for DNA extraction from cultured cells (<2 × 10^9 cells) and/or clinical specimens (<25 mg tissues), in accordance with the manufacturer's instructions without modification (Note 1).

B. DNA Fragmentation

For fragmentation of genomic DNA, add 150 µL of DNA (<4 µg) diluted with nuclease-free water (NFW) to a g-TUBE, and centrifuge twice at 4,700 × *g* and room temperature (RT) for 1 min. Using the Bioanalyzer with the Agilent DNA 12000 kit, measure the concentration and the length distribution of the fragmented DNA following the manufacturer's protocol (Note 2). Apply 10–50 ng of fragmented DNA to the next step, diluted with NFW to a 50 µL volume.

C. End Repair and Adaptor Ligation

1. Set up the programs of Steps 3, 5, and 12 in a thermal cycler.
2. Combine 50 µL of the fragmented DNA, 7 µL of NEBNext Ultra II End-Prep Reaction Buffer, and 3 µL of NEBNext Ultra II End-Prep Enzyme Mix in a PCR tube. Mix by pipetting.
3. Incubate at 20°C for 30 min, at 65°C for 30 min, then hold at 4°C in a thermal cycler with the heated lid set to 75°C.
4. Add 2.5 µL of NEBNext EMseq Adaptor, 1 µL of NEBNext Ligation Enhancer, and 30 µL of NEBNext Ultra II Ligation Master Mix to the sample. Mix by pipetting.

5. Incubate at 20°C for 15 min, then hold at 4°C in a thermal cycler with the heated lid off.
6. Add 110 µL of NEBNext Sample Purification Beads to the sample. Mix by pipetting. Incubate at RT for 5 min.
7. Place the tube on a magnetic stand for 0.2 mL tubes until it becomes clear (it takes ~2 min) (Figure 2). Remove and discard the supernatant.
8. Add 200 µL of 80% ethanol to the tube. After 30 s, remove and discard the supernatant.
9. Repeat Step 8 once.
10. Remove the tube from the magnetic stand and spin down on a tabletop centrifuge (~2,000 × *g* at RT). Place the tube on the magnetic stand. Remove and discard the remaining supernatant completely.
11. Air dry the pellet for 1 min.
12. Remove the tube from the magnetic stand. Elute the DNA from the beads by adding 29 µL of Elution Buffer from the EMseq kit and incubating at 37°C in a thermal cycler for 10 min.
13. Place the tube on the magnetic stand until it becomes clear (it takes ~2 min) (Figure 2). Transfer 28 µL of the supernatant to a new PCR tube.

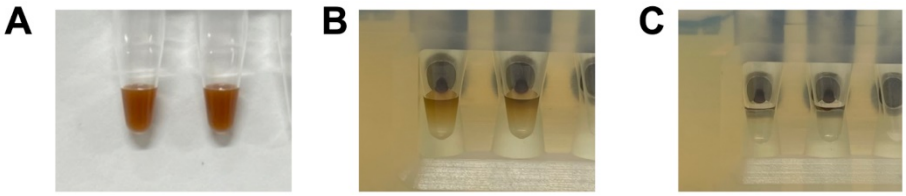

Figure 2. Bead separation by magnetic stand.
(A) Suspended beads before magnetic separation. (B) Insufficient separation of beads. The solutions are slightly cloudy. (C) Sufficient separation of beads. The solutions are clear.

D. Oxidation of 5 mC's/5 hmC's

1. Set up the programs of Steps 5, 7, and 14 in a thermal cycler.
2. Add 10 µL of TET2 Reaction Buffer with supplement, 1 µL of Oxidation Supplement, 1 µL of DTT, 1 µL of Oxidation Enhancer, and 4 µL of TET2 from the EMseq kit. Mix by pipetting.
3. Dilute 1 µL of 500 mM Fe(II) Solution from the EMseq kit in 1,249 µL of NFW in a new 1.5 mL tube.
4. Add 5 µL of the diluted Fe (II) Solution to the sample. Mix by pipetting.
5. Incubate at 37°C in a thermal cycler with the heated lid set to 45°C for 1 h.
6. Add 1 µL of Stop Reagent to the sample. Mix by pipetting.
7. Incubate at 37°C in the thermal cycler with the heated lid set to 45°C for 30 min.
8. Add 90 µL of NEBNext Sample Purification Beads from the EMseq kit to the sample. Mix by pipetting. Incubate at RT for 5 min.
9. Place the tube on a magnetic stand for 0.2 mL tube until it becomes clear (it takes ~2 min) (Figure 2). Remove and discard the supernatant.
10. Add 200 µL of 80% ethanol to the tube. After 30 s, remove and discard the supernatant.
11. Repeat Step 10 once.
12. Remove the tube from the magnetic stand and spin down on a tabletop centrifuge. Place the tube on the magnetic stand. Remove and discard the remaining supernatant completely.
13. Air dry the pellet for 1 min.
14. Remove the tube from the magnetic stand until it becomes clear (it takes ~2 min) (Figure 2). Elute the target DNA from the beads by adding 17 µL of Elution Buffer from the EMseq kit and incubating at 37°C in the thermal cycler for 10 min.
15. Place the tube on the magnetic stand until it becomes clear (it takes ~2 min) (Figure 2). Transfer 16 µL of the supernatant to a new PCR tube.

E. Denaturation of Cytosines

1. Set up the program of Step 3 in a thermal cycler.
2. Add 4 µL of formamide to the sample. Mix by pipetting.
3. Incubate at 85°C in a thermal cycler with the heated lid set to 95°C for 10 min, then place on ice immediately.

F. Deamination of Cytosines

1. Set up the programs of Steps 3 and 10 in a thermal cycler.
2. Add 68 µL of NFW, 10 µL of APOBEC Reaction Buffer, 1 µL of BSA, and 1 µL of APOBEC from the EMseq kit to the sample. Mix by pipetting.
3. Incubate at 37°C for 3 h, then hold at 4°C, in a thermal cycler with the heated lid set to 45°C.
4. Add 100 µL of NEBNext Sample Purification Beads from the EMseq kit to the sample. Mix by pipetting. Incubate for 5 min at RT.
5. Place the tube on the magnetic stand for 0.2 mL tubes until it becomes clear (it takes ~2 min) (Figure 2). Remove and discard the supernatant.
6. Add 200 µL of 80% ethanol to the tube. After 30 s, remove and discard the supernatant.
7. Repeat Step 6 once.
8. Remove the tube from the magnetic stand and spin down on a tabletop centrifuge. Place the tube on the magnetic stand. Remove and discard the remaining supernatant completely.
9. Air dry the pellet for 1 min.
10. Remove the tube from the magnetic stand. Elute the target DNA from the beads by adding 41 µL of NFW and incubating at 37°C in a thermal cycler with the heated lid set to 45°C for 10 min (Note 3).
11. Place the tube on the magnetic stand until it becomes clear (it takes ~2 min) (Figure 2). Transfer 20 µL of the supernatant to two PCR tubes (each tube contains of the supernatant 10 µL).

G. PCR Amplification

1. Set up the PCR program of Step 3 in a thermal cycler.
2. Add 5 µL of the custom primer mix (10 µM each, described in Table 1), and 25 µL of KOD ONE PCR Master Mix to each tube. Mix by pipetting.
3. Perform PCR amplification of both tubes using the following PCR program: 13–16 cycles of 94°C for 15 s, at 57°C for 5 s, 68°C for 15 min, then hold at 4°C. The number of PCR cycles depend on the amount of input DNA (16 cycles for 10 ng DNA input, 13 cycles for 50 ng DNA input) and the quality of DNA.

Table 1. PCR program

Temperature	Time	Cycles
94°C	15 s	13–16
57°C	5 s	
68°C	15 min	
4°C	Hold	1

4. Combine the separately amplified samples into one tube. Purify the sample by using a purification column of DNA Clean & Concentrator-5, according to the manufacturer's instructions. Elute the DNA from the column by adding 52 µL of NFW, pre-incubated at 70°C. Repeat the elution step by adding the 52 µL of flowthrough back to the column. The quality of the purified DNA is measured using the Agilent DNA 12000 kit (Figure 3A).

Table 2. Custom primer sequences

Primer	
Forward primer	CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT
Reverse primer	AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCTTTC CCTACACGACGCTCTTCCGATCT

H. Size Selection

1. Set up the programs of Step 8 in a thermal cycler.
2. Add 41–45 µL of ProNEX Chemistry (0.82–0.9×) to 50 µL of the DNA. Incubate at RT for 10 min.
3. Place the tube on a magnetic stand for 0.2 mL tubes until it becomes clear (it takes ~2 min) (Figure 2). Remove and discard the supernatant.
4. Add 200 µL of wash buffer to the tube. After 30 s, remove and discard the supernatant.
5. Repeat Step 4 once.
6. Remove the tube from the magnetic stand and spin down on a tabletop centrifuge. Place the tube on the magnetic stand. Remove and discard the remaining supernatant completely.
7. Air dry the pellet for 1 min.
8. Remove the tube from the magnetic stand. Elute the target DNA from the beads by adding 51 µL of NFW and incubating at 37°C in the thermal cycler with the heated lid set to 45°C for 10 min.
9. Place the tube on a magnetic stand until it becomes clear (it takes ~2 min) (Figure 2). Transfer 50 µL of the supernatant to a new PCR tube. The quality and quantity of the purified DNA are assessed using the Agilent DNA 12000 kit (Note 2) and the Qubit ds DNA HS Assay kit (Figure 3B) (Note 4).

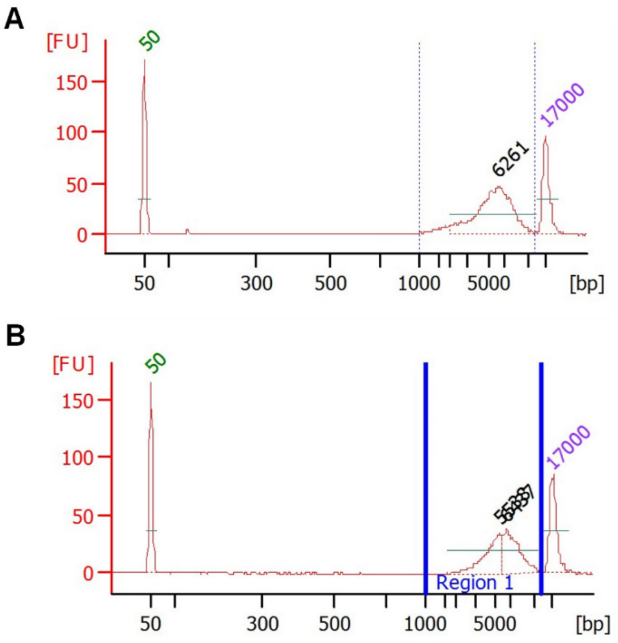

Figure 3. Amplicon of base-converted DNA.
(A and B) The amplicon distribution of base-converted DNA (A) before and (B) after size selection. DNA was quantified with the Agilent DNA 12000 kit. The reaction was performed with 50 ng of fragmented DNA from a breast cancer cell line BT-474 (Lasfargues *et al*., 1978). The amplification was performed with 13 cycles of polymerase chain reaction, using the KOD ONE PCR Master Mix and the primers described in Table 1. Size selection of the amplified DNA was performed using the 0.82× volume of ProNEX Chemistry.

I. Library Preparation for Nanopore Sequencing

1. Set up the programs of Steps 3 and 22 in a thermal cycler.
2. Combine 48 µL of the sample, 3.5 µL of NEBNext FFPE DNA Repair Buffer, 3.5 µL of Ultra II End-Prep Reaction Buffer, 2 µL of NEBNext FFPE DNA Repair Mix, and 3 µL of Ultra II End-Prep Enzyme Mix in a new PCR tube. Mix by flicking the tube and spin down on a tabletop centrifuge.
3. Incubate at 20°C for 5 min, then 65°C for 5 min, in a thermal cycler with the heated lid set to 75°C.
4. Add 60 µL of AMPure XP beads to the sample. Mix by flicking the tube and spin down on a tabletop centrifuge.
5. Incubate at RT for 5 min.
6. Place the tube on a magnetic stand for 0.2 mL tubes until it becomes clear (it takes ~2 min) (Figure 2). Remove and discard the supernatant.
7. Add 200 µL of 70% ethanol to the tube. After 30 s, remove and discard the supernatant.
8. Repeat Step 7 once.
9. Remove the tube from the magnetic stand and spin down on a tabletop centrifuge. Place the tube on the magnetic stand. Remove and discard the remaining supernatant completely.
10. Air dry the pellet for 1 min.
11. Remove the tube from the magnetic stand. Elute the DNA from the beads by adding 61 µL of NFW and incubating at 37°C in a thermal cycler with the heated lid set to 45°C for 10 min.
12. Place the tube on the magnetic stand until it becomes clear (it takes ~2 min) (Figure 2). Transfer 61 µL of the supernatant to a new PCR tube. Use 1 µL of the sample for quantification by Qubit ds DNA HS Assay kit.
13. Add 25 µL of ligation buffer, 10 µL of NEBNext Quick T4 DNA Ligase, and 5 µL of Adapter Mix F to the sample. Mix by flicking the tube and spin down on a tabletop centrifuge.
14. Incubate at RT for 10 min.
15. Add 40 µL of AMPure XP beads to the sample. Mix by flicking the tube and spin down on a tabletop centrifuge.
16. Incubate at RT for 5 min.
17. Place the tube on a magnetic stand for 0.2 mL tubes until it becomes clear (it takes ~2 min) (Figure 2). Remove and discard the supernatant.
18. Remove the tube from the magnetic stand. Wash the beads by adding 250 µL of long fragment buffer to the tube. After flicking the beads to resuspend, return the sample to the magnetic stand. Once the solution is clear, remove and discard the supernatant.
19. Repeat Step 18 once.
20. Remove the tube from the magnetic stand and spin down on a tabletop centrifuge. Place the tube on the magnetic stand. Remove and discard the remaining supernatant completely.
21. Air dry the pellet for 30 s.
22. Remove the tube from the magnetic stand. Elute the DNA from the beads by adding 25 µL of elution buffer from the Ligation Sequencing kit and incubating at 37°C in a thermal cycler with the heated lid set to 45°C for 10 min.
23. Place the tube on a magnetic stand until it becomes clear (it takes ~2 min) (Figure 2). Transfer 25 µL of the supernatant to a new PCR tube. Use 1 µL of the sample for quantification by Qubit ds DNA HS Assay kit. Estimate the molarity of the prepared library by correcting the mass concentration of the library with the mole and mass concentrations of the DNA before library preparation. Apply 5–50 fmol of the eluted library to the next step (Note 5). If more than 50 fmol of DNA is contained in 24 µL of the library, diluted 5–50 fmol of the library up to 24 µL by elution buffer from the Ligation Sequencing kit

J. Priming and Loading the PromethION Flowcell

1. Add 30 µL of Flush Tether (FLT) to one tube of Flush Buffer (FB). Mix by vortexing.
2. Set the flowcell to the PromethION sequencer. Remove air from the inlet port of the flowcell by pipetting, to avoid the introduction of air bubbles.
3. Prime the flowcell with 500 µL of FB/FLT mix. After incubation for 5 min, re-prime with 500 µL of

FB/FLT mix.

4. Add 75 µL of Sequencing Buffer II and 51 µL of re-suspended Loading Beads II to the library.
5. Mix by gently pipetting. Immediately load the 150 µL of library to the flowcell and run the program. A video for priming and loading the flow cell of PromethION is available on the community site of Oxford Nanopore Technologies (https://community.nanoporetech.com/protocols/genomic-dna-by-ligation-sqk-lsk110/v/gde_9108_v110_revl_10nov2020/priming-and-loading-the-flow-cell?devices=promethion).

Data analysis

Two fastq files containing 1d pass reads or 1d fail reads are generated via the real-time basecalling of Guppy, a basecaller integrated into MinKNOW software for the PromethION sequencer. A fastq file contains the base sequence and the quality of each of the bases for sequence reads. We recommend using the fastq file of 1d pass reads, which passed the filter of base quality, for the data analysis. DNA sequence after base-conversion by bisulfite or EMseq consists of A, G, T (original T and unmethylated C), and C (methylated C) in the original strand, or A (complementary of original T and unmethylated C), G (complementary of methylated C), T, and C in the complementary strand generated by PCR. Therefore, it is difficult to align the sequence to the normal reference sequence. To map nanoEM data to reference genome data, we adopted a three-letter alignment approach—which is also used for Bismark (Krueger and Andrews, 2011)—to long-read alignment. In the three-letter approach, to enable alignment of the base-converted reads, two types of reads are computationally prepared, where all the C are converted to T or all the G are converted to A, and two types of the reference genome sequence, where all the C are converted to T or all the G are converted to A. After alignment of the reads to the reference genomes, it is possible to determine whether each read is derived from the original or complementary strand, by choosing the best alignment combination with the best alignment score for each read, and to detect the methylation status of each C, by referring to the original sequence of reads and reference genome. A flow chart of the data analysis is shown in Figure 4. To perform the operations correctly, all information, including bioinformatics scripts and explanation, is available in a GitHub repository at this link: https://github.com/yos-sk/nanoEM. Software used in this protocol can be easily installed via the conda command of miniconda (https://docs.conda.io/en/latest/miniconda.html) or anaconda (https://www.anaconda.com/products/individual).

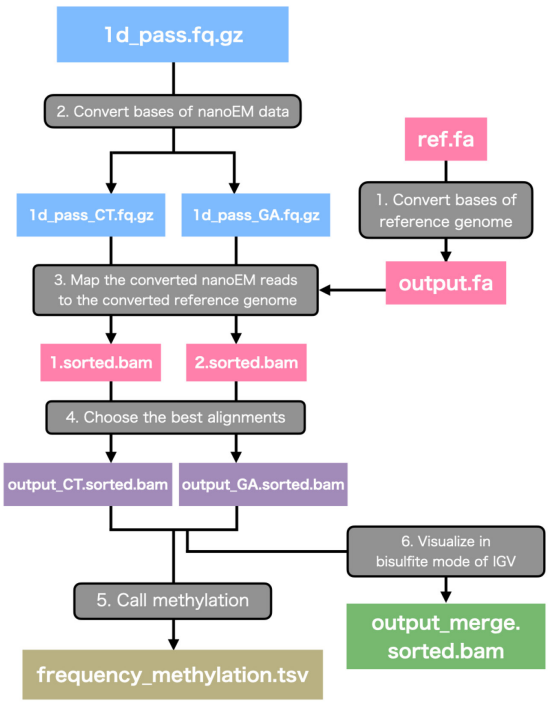

Figure 4. Flow chart of data analysis.

1. Convert bases of reference genome
   From a fasta file (ref.fa) of the reference genome (such as human genome hg38), generate a fasta file (output.fa) for a modified reference genome, representing the reference genome with the Cs converted to Ts, and with the Gs converted to As on the reverse strand.

   ```
   $ python src/convert_ref.py ref.fa > output.fa
   ```

2. Convert bases of nanoEM data
   From the compressed fastq file of nanoEM reads (1d_pass.fq.gz), generate two modified fastq files: one with the Cs converted to Ts (1d_pass_CT.fq.gz), the other with Gs converted to As (1d_pass_GA.fq.gz).

   ```
   $ python src/convert_reads.py 1d_pass.fq.gz
   ```

3. Map the converted nanoEM reads (1d_pass_CT.fq.gz and 1d_pass_GA.fq.gz) to the converted reference genome (output.fa). Map the processed nanoEM reads to the processed reference genome, using minimap2 with the "map-ont" option. Then, two bam files (1.sorted.bam and 2.sorted.bam) and two respective index files (1.sorted.bam.bai and 2.sorted.bam.bai) are generated.

   ```
   $ minimap2 -t 8 -split-prefix temp_sam1 -ax map-ont output.fa
1d_pass_CT.fq.gz -eqx | samtools view -b | samtools sort -@ 8 -o
1.sorted.bam
   $ samtools index 1.sorted.bam
   $ minimap2 -t 8 -split-prefix temp_sam2 -ax map-ont output.fa
1d_pass_GA.fq.gz -eqx | samtools view -b | samtools sort -@ 8 -o
2.sorted.bam
   $ samtools index 2.sorted.bam
   ```

4. Choose the best alignments
   From the alignment results (1.sorted.bam and 2.sorted.bam), select the most appropriate alignment combination by the alignment score. Then, two bam files (output_CT.sorted.bam and output_GA.sorted.bam) and two corresponding index files (output_CT.sorted.bam.bai and output_GA.sorted.bam.bai) are generated.

   ```
   $ python src/best_align.py --bam1 1.sorted.bam --bam2 2.sorted.bam  --
fastq nanoEM_read.fq.gz
   $ samtools view -b output_CT.sam | samtools sort -o
output_CT.sorted.bam
   $ samtools view -b output_GA.sam | samtools sort -o
output_GA.sorted.bam
   $ rm output_*.sam
   $ samtools index output_CT.sorted.bam
   $ samtools index output_GA.sorted.bam
   ```

5. Call methylation
   Using the sambamba mpileup command, detect the methylation frequencies of the cytosines in the CpG sites of the reference genome (ref.fa). After processing by a python script (src/call_methylation.py), a tsv file of methylation frequency (frequency_methylation.tsv) is generated.

   ```
   $ sambamba mpileup output_CT.sorted.bam -L cpg_sites.bed -o
pileup_CT.tsv -t 8 --samtools -f ref.fa
   $ sambamba mpileup output_GA.sorted.bam -L cpg_sites.bed -o
pileup_GA.tsv -t 8 --samtools -f ref.fa
   ```

```
$ python src/call_methylation.py pileup_CT.tsv pileup_GA.tsv >
frequency_methylation.tsv
```

6. Visualize in bisulfite mode of IGV
   To visualize in bisulfite mode of IGV, correct the sequence of G-to-A-converted reads (output_GA.sorted.bam) to that of the complementary strand and merge it (output_GA_vis.sorted.bam) with the bam file of the C-to-T-converted reads (output_CT.sorted.bam). After sorting, the merged bam file (output_merge.sorted.bam) can be visualized in the bisulfite mode of IGV. The bisulfite mode option can be activated from the right-click pop-up menu. Select “Color alignments by”, ”bisulfite mode”, then “CG”. Visualization of a typical nanoEM result is shown in Figure 5.

```
$ python script/vis_GA_utilities.py -b output_GA.sorted.bam | samtools
view -b | samtools sort -@ 4 -o output_GA_vis.sorted.bam
$ samtools index output_GA_vis.sorted.bam
$ samtools merge output_merge.bam output_CT.sorted.bam
output_GA_vis.sorted.bam
$ samtools sort -@ 4 -o output_merge.sorted.bam output_merge.bam
$ samtools index output_merge.sorted.bam
```

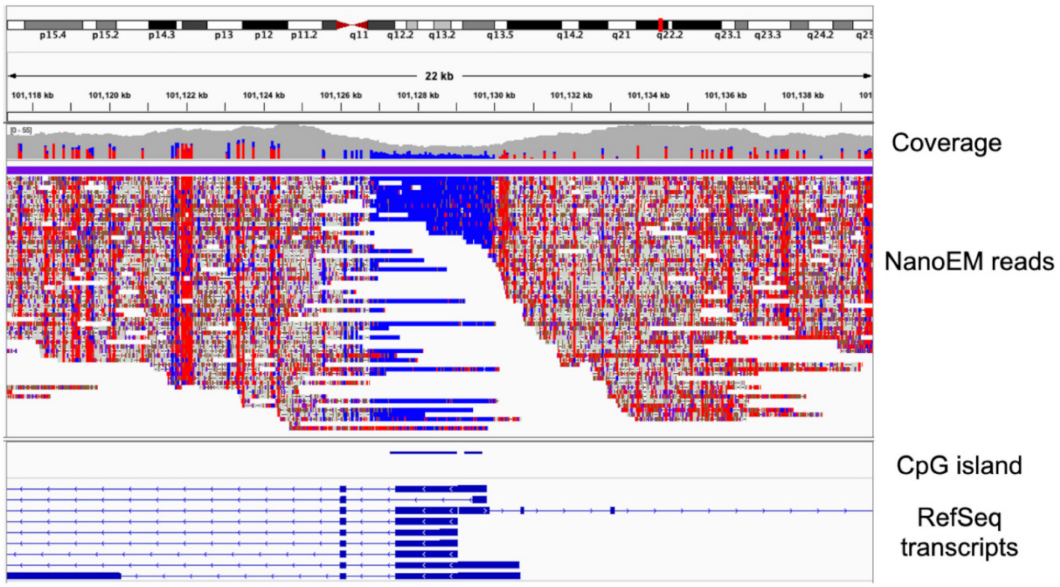

Figure 5. Visualization of nanoEM reads.
Visualization of a representative nanoEM result in bisulfite mode of IGV, in the region surrounding the promoter of the gene PGR. Methylated and unmethylated CpGs are shown in red and blue, respectively. The annotations of CpG islands were obtained from the UCSC table browser (Karolchik *et al*., 2004).

Notes

1. We have successfully used this protocol with mammalian cell lines and human clinical specimens of lung and breast.
2. After filling the DNA chip with the pre-filtrated Gel-Dye mix by the Chip Priming Station, an accessory of the 2100 Bioanalyzer Instrument, add 9 µL of the Gel-Dye mix and 5 µL of the Marker to wells of the chip following the manufacturer’s instructions. Add 1 µL of the Ladder and 1 µL of the samples to the ladder well and the sample wells, respectively. After vortexing for 1 min, set the chip to 2100 Bioanalyzer Instrument and

start measurement.
3. In this step, DO NOT use Elution Buffer from the EMseq kit, as it is detrimental to the subsequent PCR reaction.
4. We recommend using at least 200 ng of DNA for the subsequent library preparation of nanopore sequencing.
5. When the amount of library loaded is too high or too low, the yield of the sequencing data will be reduced.

Recipes

1. **TET2 Reaction Buffer with supplement**
   Add 100 µL of TET2 Reaction Buffer to a tube of TET2 Reaction Buffer Supplement and mix by vortexing. The TET2 Reaction Buffer with supplement can be stored at -20°C for 4 months.

2. **70% and 80% (v/v) ethanol**
   Mix ethanol and NFW. These reagents were freshly prepared at the time of use.

3. **Wash Buffer of ProNex Size-Selective DNA Purification System (NG2001)**
   Add 75 mL of ethanol to a Bottle of Wash Buffer.

Acknowledgments

This protocol is based on our previous publication (Sakamoto *et al*., 2021). We are supported by JSPS KAKENHI [JP21K15074, JP19K16108]; MEXT KAKENHI [JP16H06279 (PAGS), JP17H06306, JP20H05906]; JSPS Fujita Memorial Fund for Medical Research; National Cancer Center Research and Development Fund (29-A-6).

Competing interests

There are no conflicts of interest or competing interests.

References

Akbari, V., Garant, J. M., O'Neill, K., Pandoh, P., Moore, R., Marra, M. A., Hirst, M. and Jones, S. J. M. (2021). Megabase-scale methylation phasing using nanopore long reads and NanoMethPhase. *Genome Biol* 22(1): 68.

Goerner-Potvin, P. and Bourque, G. (2018). Computational tools to unmask transposable elements. *Nat Rev Genet* 19(11): 688-704.

Greenberg, M. V. C. and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* 20(10): 590-607.

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D. and Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32(Database issue): D493-496.

Krueger, F. and Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27:1571-1572.

Lasfargues, E. Y., Coutinho, W. G. and Redfield, E. S. (1978). Isolation of two human tumor epithelial cell lines from solid breast carcinomas. *J Natl Cancer Inst* 61(4): 967-978.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18): 3094-3100.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16): 2078-2079.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., *et al*. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462: 315-322.

Liu, Y., Cheng, J., Siejka-Zielinska, P., Weldon, C., Roberts, H., Lopopolo, M., Magri, A., D'Arienzo, V., Harris, J. M., McKeating, J. A., *et al*. (2020). Accurate targeted long-read DNA methylation and hydroxymethylation sequencing with TAPS. *Genome Biol* 21(1): 54.

Olova, N., Krueger, F., Andrews, S., Oxley, D., Berrens, R. V., Branco, M. R. and Reik, W. (2018). Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biol* 19(1): 33.

Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M. and Paten, B. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods* 14: 411-413.

Sakamoto, Y., Xu, L., Seki, M., Yokoyama, T. T., Kasahara, M., Kashima, Y., Ohashi, A., Shimada, Y., Motoi, N., Tsuchihara, K., Kobayashi, S. S., *et al*. (2020). Long-read sequencing for non-small-cell lung cancer genomes. *Genome Res* 30(9): 1243-1257.

Sakamoto, Y., Zaha, S., Nagasawa, S., Miyake, S., Kojima, Y., Suzuki, A., Suzuki, Y. and Seki, M. (2021). Long-read whole-genome methylation patterning using enzymatic base conversion and nanopore sequencing. *Nucleic Acids Res* 49(14): e81.

Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J. and Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* 14: 407-410.

Sun, Z., Vaisvila, R., Hussong, L. M., Yan, B., Baum, C., Saleh, L., Samaranayake, M., Guan, S., Dai, N., Correa, I. R., Jr., *et al*. (2021). Nondestructive enzymatic deamination enables single-molecule long-read amplicon sequencing for the determination of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Genome Res* 31: 291-300.

Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 31(12): 2032-2034.

Thorvaldsdóttir, H., Robinson, J. T. and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14(2): 178-192.

Vaisvila, R., Ponnaluri, V. K. C., Sun, Z., Langhorst, B. W., Saleh, L., Guan, S., Dai, N., Campbell, M. A., Sexton, B. S., Marks, K., *et al*. (2021). Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Res* 31: 1280-1289.

Yang, Y., Sebra, R., Pullman, B. S., Qiao, W., Peter, I., Desnick, R. J., Geyer, C. R., DeCoteau, J. F. and Scott, S. A. (2015) Quantitative and multiplexed DNA methylation analysis using long-read single-molecule real-time bisulfite sequencing (SMRT-BS). *BMC Genomics* 16: 350.