



Research article

The performance of whole genome bisulfite sequencing on DNBSEQ-Tx platform examined by different library preparation strategies

Boyang Cao^{a,b,1}, Huijuan Luo^{a,b,1}, Tian Luo^{a,b}, Nannan Li^{a,b}, Kang Shao^{a,b},
 Kui Wu^{a,b}, Sunil Kumar Sahu^a, Fuqiang Li^{a,b,*}, Cong Lin^{a,b,*}

^a BGI-Shenzhen, Shenzhen, 518083, China

^b Guangdong Provincial Key Laboratory of Human Disease Genomics, Shenzhen Key Laboratory of Genomics, BGI-Shenzhen, Shenzhen 518083, China

ABSTRACT

Background: Whole-genome bisulfite sequencing (WGBS) technology can provide comprehensive DNA methylation at a single-base resolution on a genome-wide scale, and is considered to be the gold standard for the detection of 5-methylcytosine (5 mC). However, the International Human Epigenome Consortium propose a full DNA methylome should have at least 30 fold redundant coverage of the reference genome from a single biological replicate. Therefore, it remains cost prohibitive for large-scale studies. To find a solution, the DNBSEQ-Tx sequencing was developed that can generate up to 6 Tb data in a single run for projects involving large-scale sequencing.

Results: In this study, we provided two WGBS library construction methods DNB_PREBSseq and DNB_SPLATseq optimized for the DNBSEQ-Tx sequencer, and demonstrated the performance of these two methods on the DNBSEQ-Tx platform, using the DNA extracted from four different cell lines. We also compared the sequencing data from these two WGBS library construction methods with HeLa cell line data from ENCODE sequenced on Illumina HiSeq X Ten and WGBS data of two other cell lines sequenced on HiSeq2500. Various quality control (QC) analyses such as the base quality scores, methylation-bias (m-bias), and conversion efficiency indicated that the data sequenced on the DNBSEQ-Tx platform met the WGBS-required quality controls. Meanwhile, our data closely resembled the coverage shown by the data generated by the Illumina platform.

Conclusions: Our study showed that with our optimized methods, DNBSEQ-Tx could generate high-quality WGBS data with relatively good stability for large-scale WGBS sequencing applications. Thus, we conclude that DNBSEQ-Tx can be used for a wide range of WGBS research.

1. Background

As an important epigenetic modification, DNA methylation (5-methylcytosine, 5 mC) can regulate the function of the genome without changing the sequence of DNA molecules. It plays a key role in genome regulation, embryonic development, genomic imprinting, and X chromosome inactivation [1–5]. Aberrant DNA methylation is related to several diseases such as cancer, diabetes and autoimmune diseases [6–10]. We can now evaluate 5 mC at the genome-wide level because of the development of high-throughput sequencing (NGS). Currently, DNA methylation detection methods combined with NGS are commonly used in MeDIP-based antibody enrichment methods, simplified versions of restriction endonuclease sulfite sequencing RRBS, whole genome bisulfite sequencing, etc [11]. Among them, the whole-genome bisulfite sequencing is a technology that can provide comprehensive DNA methylation at single-base resolution on a genome-wide scale, and is considered to be the gold standard for the detection of 5 mC. The International

* Corresponding authors. BGI-Shenzhen, Shenzhen, 518083, China.

E-mail addresses: lifuqiang@genomics.cn (F. Li), lincong@genomics.cn (C. Lin).

¹ These authors contributed equally to this work.

<https://doi.org/10.1016/j.heliyon.2023.e16571>

Received 25 October 2022; Received in revised form 14 May 2023; Accepted 19 May 2023

Available online 22 May 2023

2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Human Epigenome Consortium proposed that a full DNA methylome should have at least 30 fold redundant coverage of the reference genome from a single biological replicate [12]. Therefore, it remains cost prohibitive for large-scale studies.

Recently, with the in-depth development of NGS technology, MGI Tech, a subsidiary of the Beijing Genomics Institute (BGI) Group, has invented DNBSEQ™, a DNA-based nanoball technology, which is mainly characterized by high accuracy, and low sequencing error rates [13,14]. DNBSEQ™ has been widely recognized and used on various sequencing platforms including BGISEQ-50, BGISEQ-500, DNBSEQ-G50, DNBSEQ-G400 and DNBSEQ-Tx. Particularly, DNBSEQ-Tx is a cost-effective method with a high sequencing throughput, which can support the simultaneous operation of 10 sequencing chips and generate up to 6 Tb of data per day (about 60 complete human genomes are sequenced at $30 \times$ depth) [15–17]. Currently, there have been numerous comparisons between MGI and Illumina sequencers' performance on multi-omics research, and the findings have shown that MGI sequencing can produce high-precision data at a reduced cost for studying whole genome sequencing (WGS) [16], whole-exome sequencing [18,19], RNA sequencing [20], single cell transcriptome [21,18], and Metagenome [23].

However, DNA methylation, as the most important epigenetic modification, has not been well characterized on the DNBSEQ-Tx platform for both library construction and sequencing. To exam the performance of WGBS sequencing on DNBSEQ-Tx platform, we used DNA from four different cell lines to construct two types of bisulfite-treated whole-genome sequencing libraries, DNB_PREBSseq and DNB_SPLATseq, and compared their sequencing data. The former was based on the conventional pre-bisulfite method, in which methylated adapters were added to double stranded sheared DNA fragments before the bisulfite treatment [24]. Based on this, to prevent the false methylation information during the end repair, we added C-free dNTP instead of a dNTP mix reagent. The latter was based on a post-bisulfite WGBS library preparation protocol, which skipped the fragmentation step and added splinted adaptor tagging (SPLAT) to the bisulfite-converted single-stranded DNA immediately after bisulfite treatment of genomic DNA to complete the construction of the library that had been optimized for the DNBSEQ-Tx sequencing platform [25]. At the same time, we compared our data with the HeLa cell line data (pre-bisulfite method obtained from the ENCODE), NA10860 and REH cell line data (SPLAT method obtained from SRA study SRP092113) sequenced on the Illumina platform. We demonstrated that DNBSEQ-Tx can generate high-quality WGBS data with excellent stability and reproducibility, and we believe that DNBSEQ-Tx is a promising platform for large-scale WGBS sequencing applications.

In addition, by systematically evaluating two WGBS library preparation techniques with DNBSEQ-Tx platform, we verified that DNB_SPLATseq outperforms DNB_PREBSseq in multiple ways. Due to the broken strands caused by the side effect of sodium bisulfite treatment, the majority of libraries constructed with DNB_PREBSseq could not be properly amplified in PCR [26]. Thus, coverage uniformity of the whole genome was greatly affected, especially in CpG islands (CGI) regions. On the other hand, data based on DNB_SPLATseq showed better coverage uniformity in each element of the genome. Meanwhile, the DNB_SPLATseq method required fewer amounts of DNA input. Most importantly, it can be used for automated library construction. Considering that DNBSEQ-Tx can produce 6 TB of sequence data in a single run but at a less expensive run cost than Illumina's platform [15,16], higher methylome coverage can be used to fulfill the scientific requirements of large-scale population-based WGBS research while it is still potentially more cost-effective when combined with DNBSEQ-Tx high-throughput sequencing.

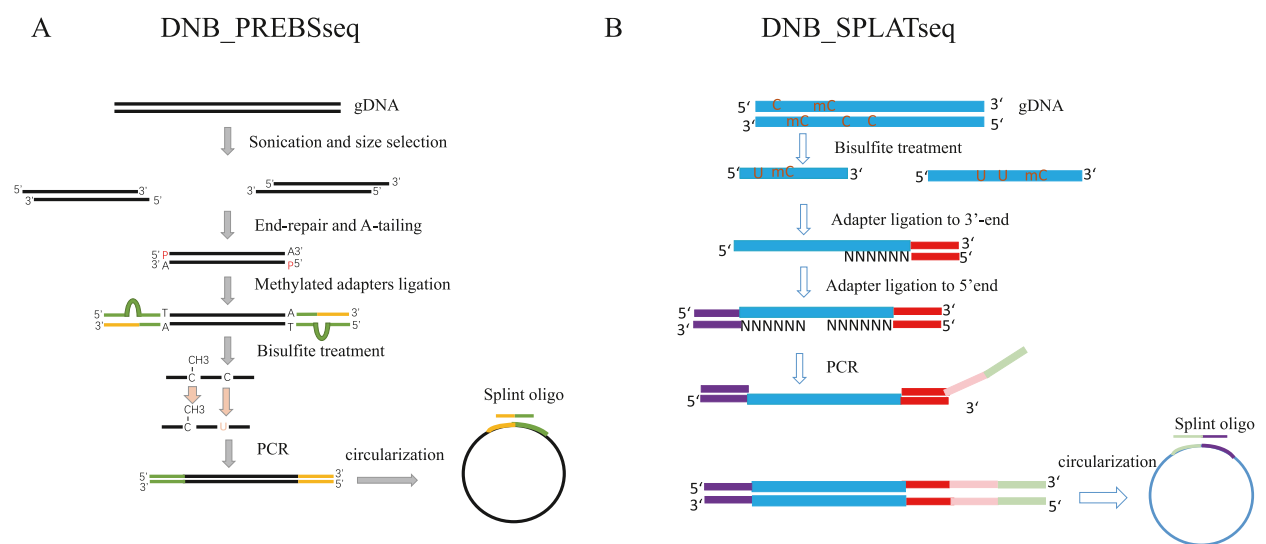


Fig. 1. Schematic diagram of library preparation methods for whole-genome bisulfite sequencing. (A). DNB_PREBSseq method. (B). DNB_SPLATseq method.

2. Materials and methods

2.1. Sample source and DNA preparation

Cell lines of HCT116 (Hereinafter referred to as HCT), RKO and HeLa were obtained from the American Type Culture Collection (ATCC). The lymphoblastic cell line (YH cell line) was established from an Asian genome donor [27]. Human genomic DNA from the cells was extracted with QIAmp DNA Mini kit (Qiagen, USA). Genomic DNA was quantified using the Qubit dsDNA HS Assay Kit (Invitrogen).

2.2. Library preparation with DNB_PREBSseq and DNB_SPLATseq

We performed two genome-wide methylation library preparation methods based on the different workflow of bisulfite treatment, DNB_PREBSseq and DNB_SPLATseq (Fig. 1 A, B).

DNB_PREBSseq library construction: Briefly, 1 µg of genomic DNA was fragmented using the Covaris ultrasonic system to get the 100-700bp fragmented DNA. The Ampure XP (Beckman) magnetic beads were used to do the size selection to obtain 200-300 bp fragmented DNA. Then the DNA concentration was measured with Qubit HS kit (Invitrogen) and 50 ng of fragmented DNA was mixed with 0.5 ng of unmethylated lambda DNA for each sample to do the end repair and A -tailing with 10X T4 Polynucleotide Kinase Buffer (700 mM Tris-HCl, 100 mM MgCl₂, 50 mM DTT, ENZYMATICS), 1.25 mM-dGTP/dATP/dTTP mix (ENZYMATICS), 6 units T4 Polynucleotide Kinase (ENZYMATICS), 20 units DNA Polymerase I (NEB), 0.5 units Klenow Fragment (ENZYMATICS), 1 unit rTaq (TAKARA) in a total volume 50 µL for 30 min at 37 °C and 15 min at 65 °C. Next, 1 mM ATP (NEB), 0.94 µM MGIEasy DNA methylation Adapters (Supplemental), 7.5% PEG 8000 and 600 units T4 DNA Ligase (ENZYMATICS) were added in 80 µL reaction volume and incubated for 30 min at 20 °C. The ligation product was purified with 80 µL Ampure XP and eluted with 20 µL nuclease-free water. The product was bisulfite-converted with an EZ DNA Methylation-Gold kit (Zymo Research) and eluted with 22 µL nuclease-free water. The bisulfite-converted DNA was PCR amplified with 25 µL KAPA HiFi HotStart Uracil + Ready Mix (2 ×) and 0.6 µM PCR 2_2 and 0.6 µM PCR 2_1 (Supplemental) using the following conditions: 98 °C for 30 s, a total of 10–13 cycles of 98 °C for 10 s/60 °C for 30 s/72 °C for 30 s, and 72 °C for 5 min. Subsequently, the PCR product was purified by 50 µL AMPure XP magnetic beads.

DNB_SPLATseq library construction: Briefly, 200 ng of genomic DNA mixed 1 ng of unmethylated lambda DNA was converted using the EZ DNA Methylation-Gold kit (Zymo Research) eluted with 12 µL nuclease-free water. The converted DNA product was first treated with 6 units T4 PNK (ENZYMATICS) in a total volume of 30 µL for 15 min at 37 °C. Then cooled on ice after reacting at 95 °C for 3 min. Afterwards 30 µM Adapter 1 (Supplemental) was ligated to the 3' end of DNA fragments with 600 units T4 DNA ligase (ENZYMATICS), 10X T4 DNA ligase buffer (40 mM Tris-HCl pH 7.8, 10 mM MgCl₂, 10 mM DTT, 0.5 mM ATP, Thermo Fisher Scientific), and PEG4000 (5% w/v, Thermo Fisher Scientific) in a total 50 µL for 60 min at 20 °C. The ligated product was purified using 50 µL Ampure XP (Beckman) magnetic beads and eluted with 20 µL nuclease-free water. After denaturing the adapter 1-ligated DNA for 3 min at 95 °C, next 20 µM Adapter 2 (Supplemental) was added to the 5' end of the DNA fragments with 600 units T4 DNA ligase (ENZYMATICS), 10X T4 DNA ligase buffer (40 mM Tris-HCl pH 7.8, 10 mM MgCl₂, 10 mM DTT, 0.5 mM ATP, Thermo Fisher Scientific), and PEG4000 (5% w/v, Thermo Fisher Scientific) in a total 40 µL for 60 min at 20 °C. The ligated product was then purified using 40 µL Ampure XP (Beckman) magnetic beads and eluted with 22 µL nuclease-free water. The purified product was amplified with 25 µL KAPA HiFi HotStart Uracil + Ready Mix (2 ×) and 0.6 µM universal oligo and 0.6 µM barcode oligo (Supplemental) using the following conditions: 98 °C for 3 min, a total of 10–13 cycles of 98 °C for 10 s/60 °C for 30 s/72 °C for 2 min, and 72 °C for 5 min, and the proper size of PCR product was selected by 0.7X+0.3X (of PCR mixture volume) AMPure XP magnetic beads volume.

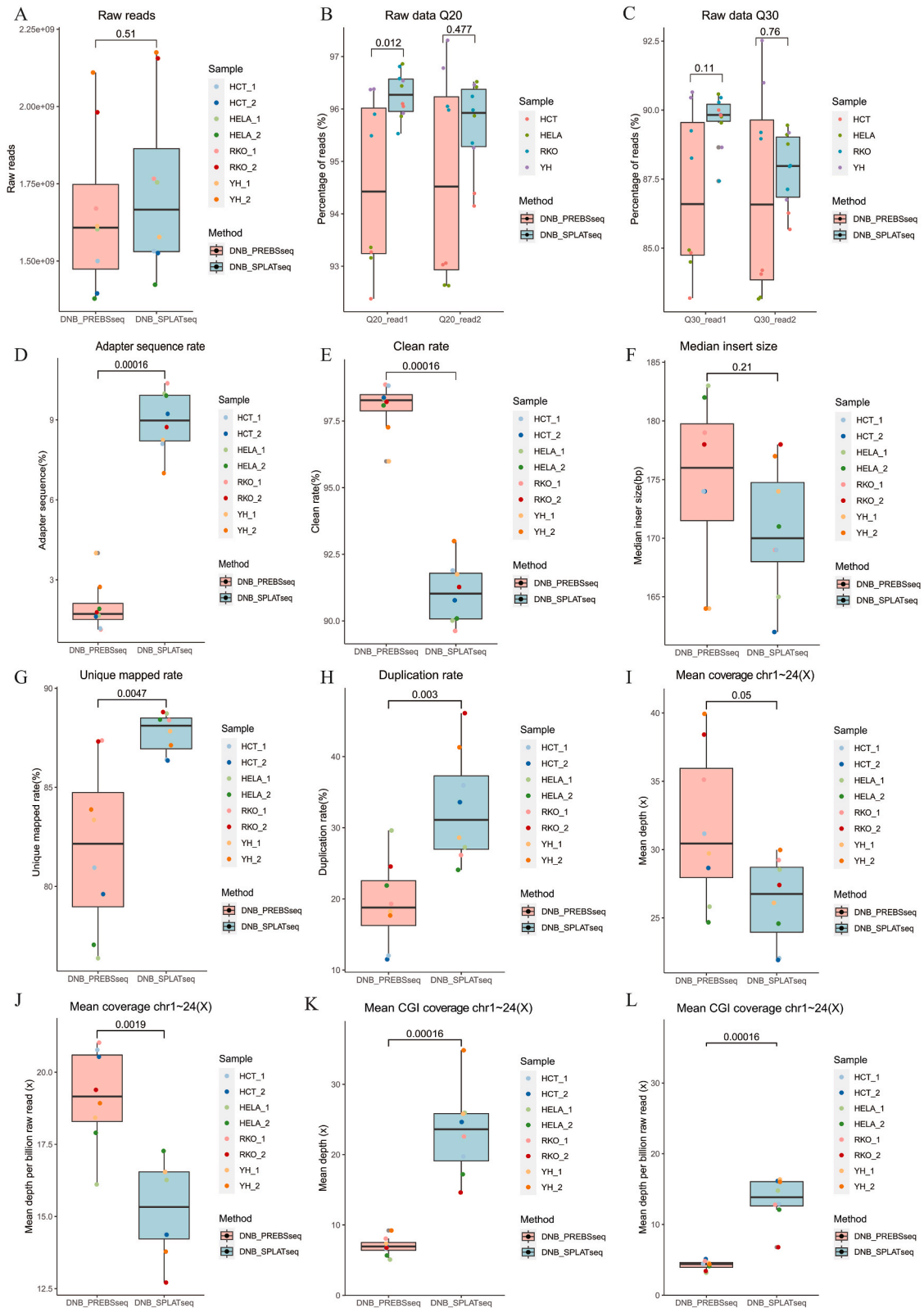
Library circularization and sequencing: PCR products were quantified with a Qubit ds DNA HS kit. Subsequently, the PCR product was normalized to 330 ng in a volume of 60 µL, annealed, and circularized with 1 mM Split oligo (Supplemental oligo) and 120 units T4 DNA ligase (ENZYMATICS). Linear DNA was then digested with 78 units Exo I and 26 units Exo III (NEB) and purified with 160 µL Ampure XP and eluted with 40 µL nuclease free water. After purification, the circularized library was quantified using the Qubit ssDNA kit, followed by rolling circle amplification (RCA) to obtain DNA nanospheres (DNB). Finally, DNB was quantified using the Qubit ssDNA kit and sequenced at PE100 + 10 on the DNBSEQ-Tx.

Processing and analysis of sequencing data: Sequencing quality was determined by FastQC software (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Raw sequencing reads were filtered by SOAPnuke (v2.0.7, <https://github.com/BGI-flexlab/SOAPnuke>) with the key parameter '-l 5 -q 0.5 -n 0.1 -f AAGTCGAGGCCAAGCGGTCTTAGGAAGACAA -r AAGTCGATCGTAGCCATGTCTGTTCTGTGAGCCAAGGAGTTG' for DNBseq-Tx platform and '-l 5 -q 0.5 -n 0.1 -f AGATCGGAAGAGCA-CACGTCTGAAGTCCAGTCAC -r AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTA' for Illumina platform (discard reads that containing adaptor sequence, more than 10% N bases, or more than 50% bases with quality less than 5). Then, the clean reads were mapped to the hg19 reference downloaded from the GATK resource bundle (<ftp://gsapubftp-anonymous@ftp.broadinstitute.org/bundle/hg19/ucsc.hg19.fasta.gz>) by BitMapperBS (v1.0.2.3, <https://github.com/chhy123/BitMapperBS>) with default parameters and sorted by samtools (v1.11, <http://www.htslib.org/>). Finally, the mapped reads were deduplicated by Sentieon [28] (-algo Dedup -rmdup) for further analysis. Coverage of read, coverage of CGI region and insert size were also analyzed by Sentieon (-algo WgsMetricsAlgo -min_map_qual 1 -min_base_qual 1 -algo InsertSizeMetricAlgo).

DNA methylated sites were identified, extracted and counted by MethylDackel (v0.5.1, <https://github.com/dpryan79/MethylDackel>), and annotated to genomic features by R package annotatr [29]. Methylation at the individual cytosines, instead of the dinucleotide level was computed by custom-made R scripts, so call sets from forward and reverse strands were not merged, and only mC (Cytosine DNA methylation site) sites with a depth of coverage >5X were considered for methylation analysis.

Table 1
Library construction and sequencing information.

sample	HCT cell line				HELA cell line				YH cell line				RKO cell line			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
library preparation methods	DNB_PREBSseq		DNB_SPLATseq		DNB_PREBSseq		DNB_SPLATseq		DNB_PREBSseq		DNB_SPLATseq		DNB_PREBSseq		DNB_SPLATseq	
DNA input (ng)	1000		200		1000		200		1000		200		1000		200	
DNA fragmentation before bisulfite treatment	YES		NO		YES		NO		YES		NO		YES		NO	
Bisulfite conversion kit									EZ DNA Methylation-Gold Kit							
Sequencing Platform									DNBSEQ-Tx							
Sequencing type									PE100 + 10							
Spike-in (%)									<i>Kineococcus radiotolerans</i> (15%)							
Data output per sample (Gb)	150	139.5	153.2	152.6	160.3	137.9	175.5	142.3	161.3	211	157.8	217.5	167	198.1	176.6	215.6



(caption on next page)

Fig. 2. The performance of quality metric of two library preparation methods. (A). Comparison of Raw reads number between data from two libraries (B) and (C). Raw reads with sequencing quality > Q20 or > Q30 (D). Reads discarded based on adapters (E). Clean rate (%) (F). Library insert size (bp) (G). Unique mapped rate (%) (H). Duplicates rate (%) (I). Average depth(X) chr1~chr24 (J). Average depth per billion raw reads(X) (chr1~chr24), calculated as mean depth divided by corresponding library size (K). CGI Average depth(X) chr1~chr24 (L). CGI Average depth per billion raw reads(X) (chr1~24). A two-sided Wilcoxon tests was used, and $P < 0.05$ was considered significant.

3. Results

To systematically investigate the WGBS performance using DNBSEQ-Tx platform, we selected four cell lines for library construction and sequencing using both the DNB_PREBSseq and the DNB_SPLATseq methods. Libraries were all sequenced at PE100 + 10 on the DNBSEQ-Tx platform (see Table 1). As shown in the table, Samples 1–4 were DNA samples extracted from the HCT cell line with technical replicates for both library construction methods, samples 5–8 were from the HeLa cell line, samples 9–12 were from the YH cell line, and sample 13–16 were from RKO cells line.

Considering the unbalanced base composition of the WGBS library, a substantial spike-in DNA with high GC content was used as an effective balance for the AT-rich bisulfite-converted [30] DNA. In our study, about 15% of *Kineococcus radiotolerans* were added to the library to balance the base ratio for WGBS sequencing on DNBSEQ-Tx platform (as shown in Table 1). Libraries with different barcodes were pooled together for sequencing on the DNBSEQ-Tx platform according to the required amount of sequencing data. After the data was released from the sequencer, it was split according to different barcodes to obtain the original data of different data samples.

4. Evaluation of quality metrics, alignment rate, and coverage depth

4.1. Performance of quality metrics

First, the data quality of the two types of libraries sequenced on the DNBSEQ-Tx platform was demonstrated. The sequencing data quality was assessed by the fqcheck program. Collectively, based on the same amount of data (Fig. 2 A), base quality scores 20 (Q20) (mean 94.6 for DNB_PREBSseq; mean 96.0 for DNB_SPLATseq) and Q30 (mean 87.0 for DNB_PREBSseq; mean 88.7 for DNB_SPLATseq) were similar between our two library preparation methods (Fig. 2 B and C). At the same time, two sequencing results of HeLa on Illumina HiSeq X Ten were also download from the ENCODE (<https://www.encodeproject.org/experiments/ENCSR550RTN/>). While two sequencing data of NA10860 and REH cell lines based on SPLAT method generated using Illumina HiSeq2500 were downloaded from the SRA study SRP092113 (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA350687&o=acc_s%3Aa&s=SRR4453306,SRR4453305,SRR4453297,SRR4453298) used as reference. Unexpectedly, there was a large difference between the Q20 and Q30 of the two samples from ENCODE, with 89.2 and 94.2 for one sample and 78.6 and 88.2 for the other sample, respectively. Of note, the average Q20 (91.7) and Q30 (83.4) were much lower than our DNB_SPLATseq results (Q20 96.0, Q30 88.7), while they were close to our DNB_PREBSseq results (Q20 94.6, Q30 87.0) (Supplementary Table 1). Moreover, the Illumina SPLAT data had a similar quality as that of ENCODE, with the average Q20 of 90.8 and Q30 of 85.1, and both showed lower Q30 of reads2.

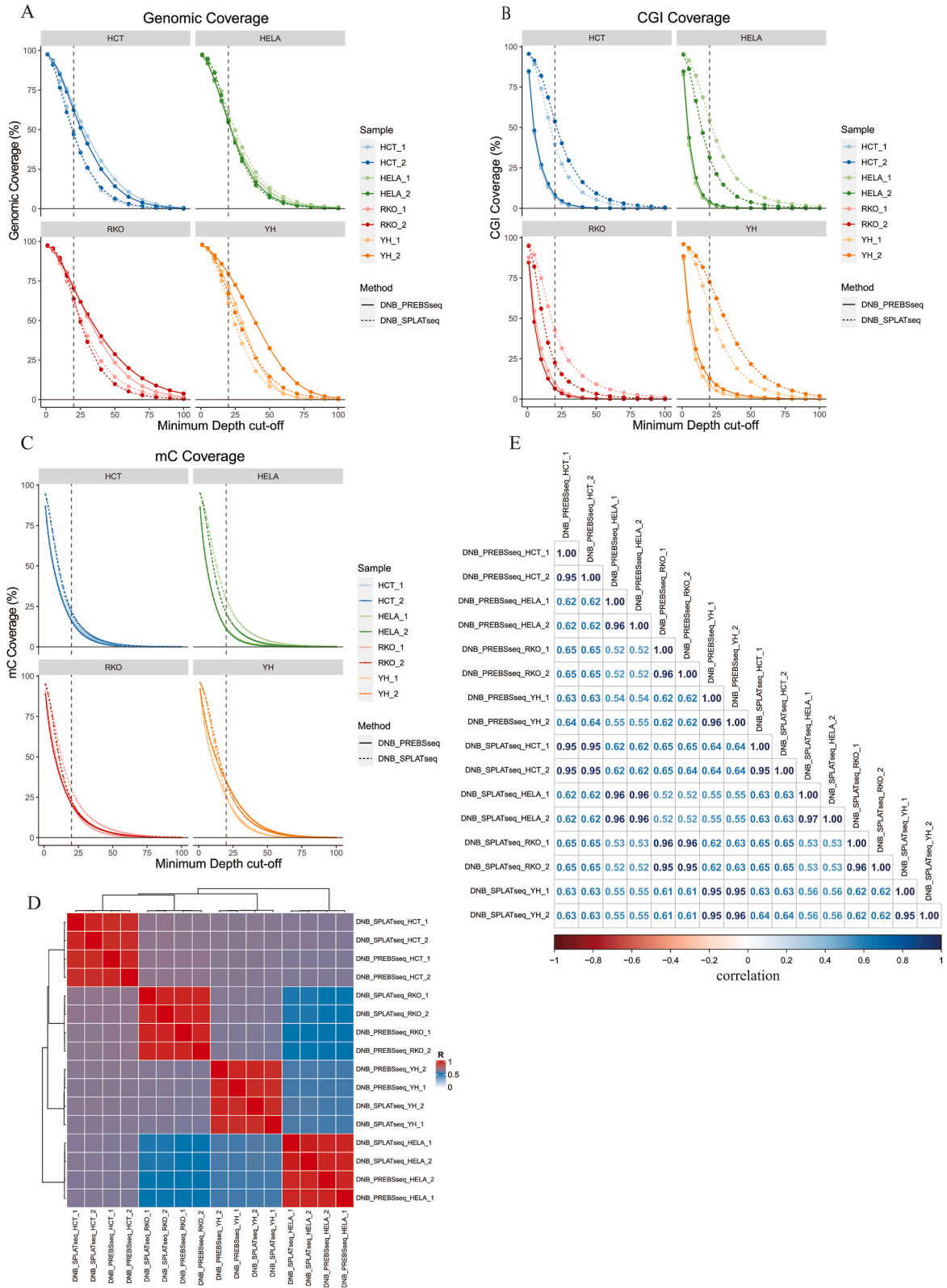
In terms of data clean rate, the DNB_PREBSseq data (mean clean rate 98.0%) was about 4–9% higher than the data from the DNB_SPLATseq method (mean clean rate 91.1%), which may be attributed to the adapter sequence rate in the DNB_SPLATseq method data (mean 8.9%) that was about 4–9% higher than the DNB_PREBSseq method data (mean 2.0%) (Fig. 2 D, E; Supplementary Table 2). As shown in Fig. 2 F, there was little difference in mean inserted fragment size between the data from the DNB_SPLATseq and the DNB_PREBSseq method. Correspondingly, though the ENCODE HeLa data had similar performance to DNB_PREBSseq data, it had a lower clean rate (mean 90.3%) and a higher adapter sequence rate (mean 9.7%). Similarly, the SPLAT data from SRA also had a lower clean rate (mean 86.7%), but a slightly higher adapter sequence rate (mean 11.1%) compared to the DNB_SPLATseq (Supplementary Table 2).

These results indicated that the relatively low clean rate of the DNB_SPLATseq library preparation method may be due to its higher percentage of adapter sequence rate.

4.2. Performance of alignment rate, coverage depth and bias

We examined the performance of the map rate and the duplication rate of the data from the two library preparation methods. The unique map rate of the DNB_SPLATseq data was about 6% significantly higher than the DNB_PREBSseq (DNB_SPLATseq:87.8%; DNB_PREBSseq:82.0%), while the duplication rate was about 14% higher than DNB_PREBSseq (DNB_SPLATseq:32.9%; DNB_PREBSseq:19.4%) (Fig. 2 G and H). The reason that DNB_PREBSseq had a lower unique map rate may be because some reads of DNB_PREBSseq were fragmented after bisulfite processing, which reduced the map rate. While the fact that DNB_SPLATseq had a small library diversity may be due to the low input of DNA and more PCR cycle for library construction, which led to a high duplication rate. In the actual experiment, this could be improved by increasing the starting amount DNA and reducing the PCR cycle. The average map rate of ENCODE HeLa data was 84.3%, and the average duplication rate was 26.9%, which was close to the performance of DNB_PREBSseq method, while the average map rate of SPLAT SRA data was 79.1%, and the average duplication rate was 1.3%, which suggested that the duplication rate was significantly lower than DNB_SPLATseq method (Supplementary Table 2).

We also looked at the coverage of the genome and CGI region respectively. We observed that the DNB_PREBSseq data had significantly higher mean genome coverage than that from the DNB_SPLATseq (26.2X for DNB_SPLATseq, and 31.7X for



(caption on next page)

Fig. 3. The performance of coverage at different minimum depth cutoffs between the two methods. (A) Genome coverage at each minimum depth cutoff across the two methods' data (B) CGI coverage at each minimum depth cutoff across the data from the two methods (C) Cytosine DNA methylation (mC) coverage at each minimum depth cutoff across the two methods data (D) Clustering of correlations across the two methods (5X mC cutoff) (E) sample correlations across the two methods (5X mC cutoff). Pearson correlation coefficient was used.

DNB_PREBSseq), while the average coverage of the CGI region for the DNB_PREBSseq data was much lower than that from the DNB_SPLATseq method (23.2X for DNB_SPLATseq, and 7.0X for DNB_PREBSseq, stat region chr1~chr24), which was consistent with the performance of the two methods on the Illumina sequencer [18] (Fig. 2 I, K). To avoid any potential bias caused by differences in raw sequencing output, we also normalized the effective depths by raw data size. Mean genome depth per billion raw reads and mean CGI region depth per billion raw reads were calculated with average effective genome reads and CGI region reads, respectively. As expected, the data from both methods still showed significant differences in normalized average read depth and normalized average CGI region effective read depth (13.5X for DNB_SPLATseq, and 4.2X for DNB_PREBSseq, stat region chr1 ~chr24) (Fig. 2 J, L; Supplementary Table 2). The dramatically reduced coverage of the CGI region in the DNB_PREBSseq data may be due to the fact that the reads in the CGI region with the adapters were more easily broken after being processed by bisulfite and the reads could not be sequenced, thus losing the information of the CGI region. While in the DNB_SPLATseq method, the adapter was linked after the bisulfite treatment, so the broken reads could still be sequenced by adding adapters, and the information of the CGI region was preserved. The results of ENCODE HeLa data (mean genomic coverage 33.8X, mean CGI coverage 10.7X) were very similar to the results of HELA data from DNB_PREBSseq method, indicating that this difference was an inherent defect of pre-bisulfite treatment method. In addition, there wasn't much of a difference in the m-bias of the read1 and read2 for the two approaches (Supplementary Fig. 1). The only difference was that the starting 5 bp of the reads of the DNB_PREBSseq method oscillates greatly, mainly because the start bases contained no C bases (dNTPs without dCTP were used for the end repair of the fragments), which led to the fluctuation in the base methylation rate. ENCODE HeLa's m-bias was consistent with previous reports [31,32], and reads2 displayed considerable decline at the beginning of the reads, for which end repair step used dNTP with unmethylated C (Supplementary Fig. 2A and B). The SRA data for SPLAT revealed a comparable m-bias to that of DNB_SPLATseq, however, reads2 did not exhibit a huge drop at the start of the reads (Supplementary Fig. 3A and B), and their mean genomic coverage was 15.5X, which were also similar to the results of DNB_SPLATseq method (SPLAT SRA data had 60% base of DNB_SPLATseq data, the mean genomic coverage was also 60%), but the mean CGI coverage was 8.3X, which was lower than DNB_SPLATseq method.

The bisulfite conversion efficiency was estimated using the non-methylated λ DNA introduced during library construction. The DNB_PREBSseq, DNB_SPLATseq and ENCODE HeLa data exhibited higher conversion efficiency (99.05%–99.58%), while due to the absence of λ DNA in SPLAT SRA data, its conversion efficiency could not be calculated.

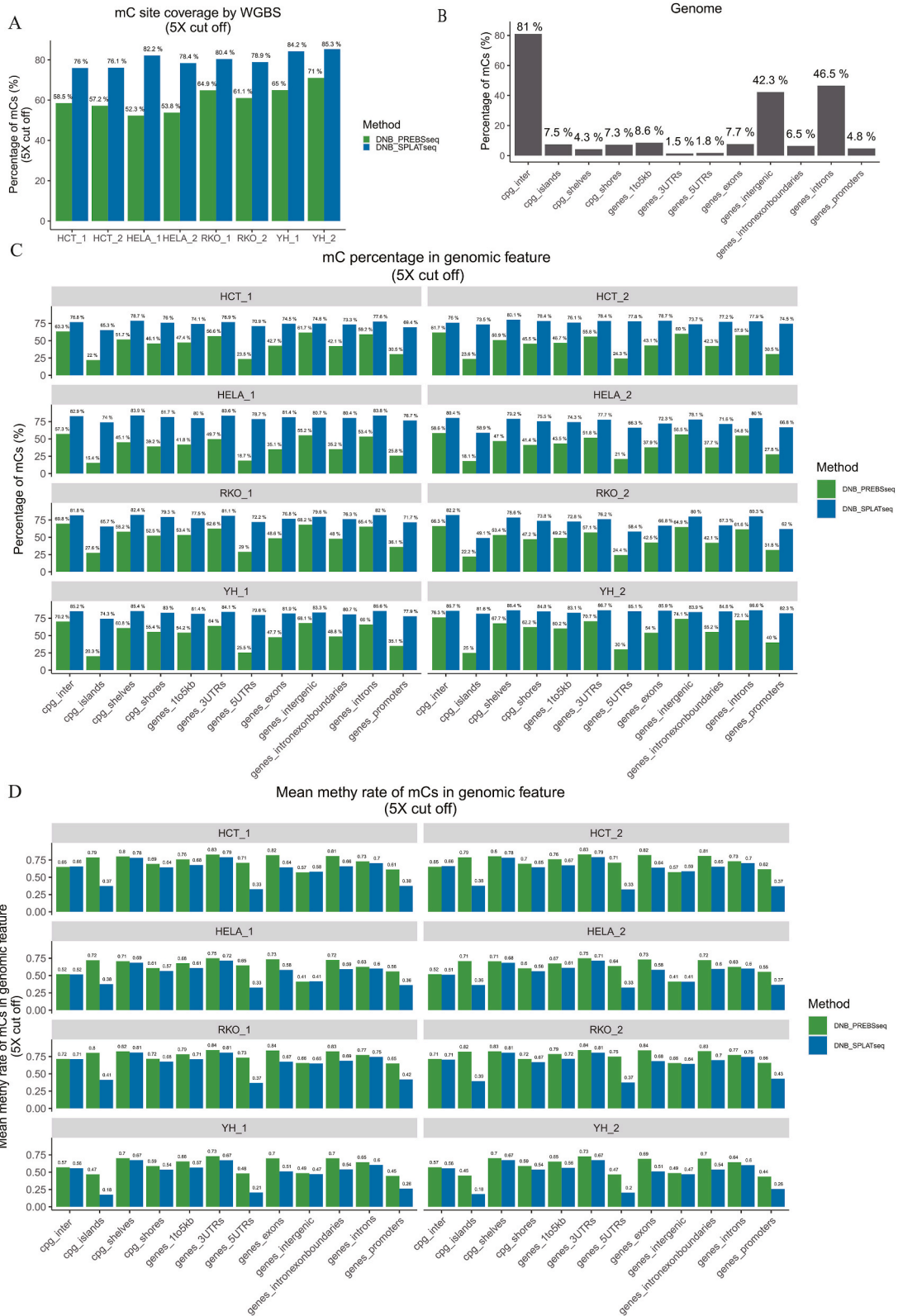
4.3. Performance of coverage at different minimum depths and downsampling

The performance of coverage at each minimum depth cutoff between the data from the two methods was also evaluated. At the same depth, the genome coverage of DNB_PREBSseq data was clearly higher than or equal to that from DNB_SPLATseq (Fig. 3 A), but the coverage of CGI region based on DNB_PREBSseq method was much lower than that of the DNB_SPLATseq method (Fig. 3 B). The results of ENCODE HeLa data were still similar to those of DNB_PREBSseq (Supplementary Fig. 2C and D), indicating that our modification of DNB_PREBSseq could not solve the problem of DNA damage caused by bisulfite treatment in traditional methods. The base number of the SPLAT SRA data was only 60% of DNB_SPLATseq, however it was still able to reach 60% mean genome coverage with slightly lower CGI coverage. This indicates that the performance of the original SPLAT method is roughly achieved by our DNB_SPLATseq approach (Supplementary Fig. 3C and D).

To further demonstrate the characteristics of the methylation site coverage of the data from DNB_PREBSseq and DNB_SPLATseq, we also showed the percentage of all mC sites (cytosine DNA methylation site) that were covered at each minimum depth cutoff between the two data sets. We found that the coverage of DNB_SPLATseq data was only slightly higher than that from DNB_PREBSseq at the same depth of the two methods (Fig. 3C), indicating that the performance of DNB_PREBSseq method may be insufficient with respect to the coverage of CGI region. Moreover, the methylation values obtained by the two library construction methods of the same sample were still highly correlated (Pearson's correlation coefficient: 0.96; Fig. 3 D, E). In addition, there were generally no differences between our data and the mC coverage of the two HeLa data from ENCODE (Supplementary Fig. 2 E).

To find out the distribution of mC sites on each element of the genome in data from the two methods, we also classified the mC sites and counted the distribution of mC sites in the whole genome at 5X coverage. The percentage of mC in DNB_SPLATseq data was much higher than that of DNB_PREBSseq for different samples under similar raw sequencing data (Fig. 4 A). Both of the methods could quantify the great majority of methylation at mC sites in the human genome (mean 80% at 5X mC coverage for DNB_SPLATseq and mean 60% at 5X mC coverage for DNB_PREBSseq). Although the mean 5X mC coverage of ENCODE data was as high as 78.4%, its data size was 1.5 times that of our sequencing data, so its mC coverage should be lower than that of DNB_SPLATseq method and nearly the same with DNB_PREBSseq under the same data size (Supplementary Fig. 2 F, Supplementary Table 3). The SPLAT SRA data base number was 60% of DNB_SPLATseq, but its 5X mC coverage could still be managed to reach a mean of 68.1% (Supplementary Fig. 3 F, Supplementary Table 3).

To further examine the mC coverage on different genomic regions, we also annotated the mC sites and calculated the ratio of the number of mC at 5X depth to the total number of mC in the certain genomic region. We found that the DNB_SPLATseq method had a more uniform performance on coverage of different regions (Fig. 4 B, C), while the DNB_PREBSseq method has much lower coverage of



(caption on next page)

Fig. 4. mC site coverage at different genomic features. The percentages were calculated by dividing the number of mC sites covered with a minimum depth of 5X for each genomic feature by the total number of mC sites in the genome for the corresponding genomic feature. (A). The ratio of mC in all genomes at 5X cutoff depth. (B). The ideal ratio of the total number of genome-wide mC in every genomic feature. (C). The ratio of the number of mC to the total number of mC in every genomic feature at 5X depth. (D). The average mC site methylation rate in every genomic feature at 5X depth.

the CGI, including 5'UTRs and promoter. From the statistics of the average mC site methylation rate in these regions, it could be seen that the average methylation rates of these three regions were significantly higher in DNB_PREBSseq data (Fig. 4 D), indicating that the methylation status of these three regions obtained by this method may be distorted. Although the 5X coverage data of the ENCODE HeLa samples was similar to that of DNB_PREBSseq data, the methylation rates of the two samples were similar to that of DNB_SPLATseq (Supplementary Fig. 2G and H). The SPLAT SRA data was supposed to be comparable to the DNB_SPLATseq approach, however, it showed poorer mC coverage in the CGI and 5'UTRs regions, suggesting that the original SPLAT method still had insufficient coverage in high GC regions (Supplementary Fig. 3 G, H).

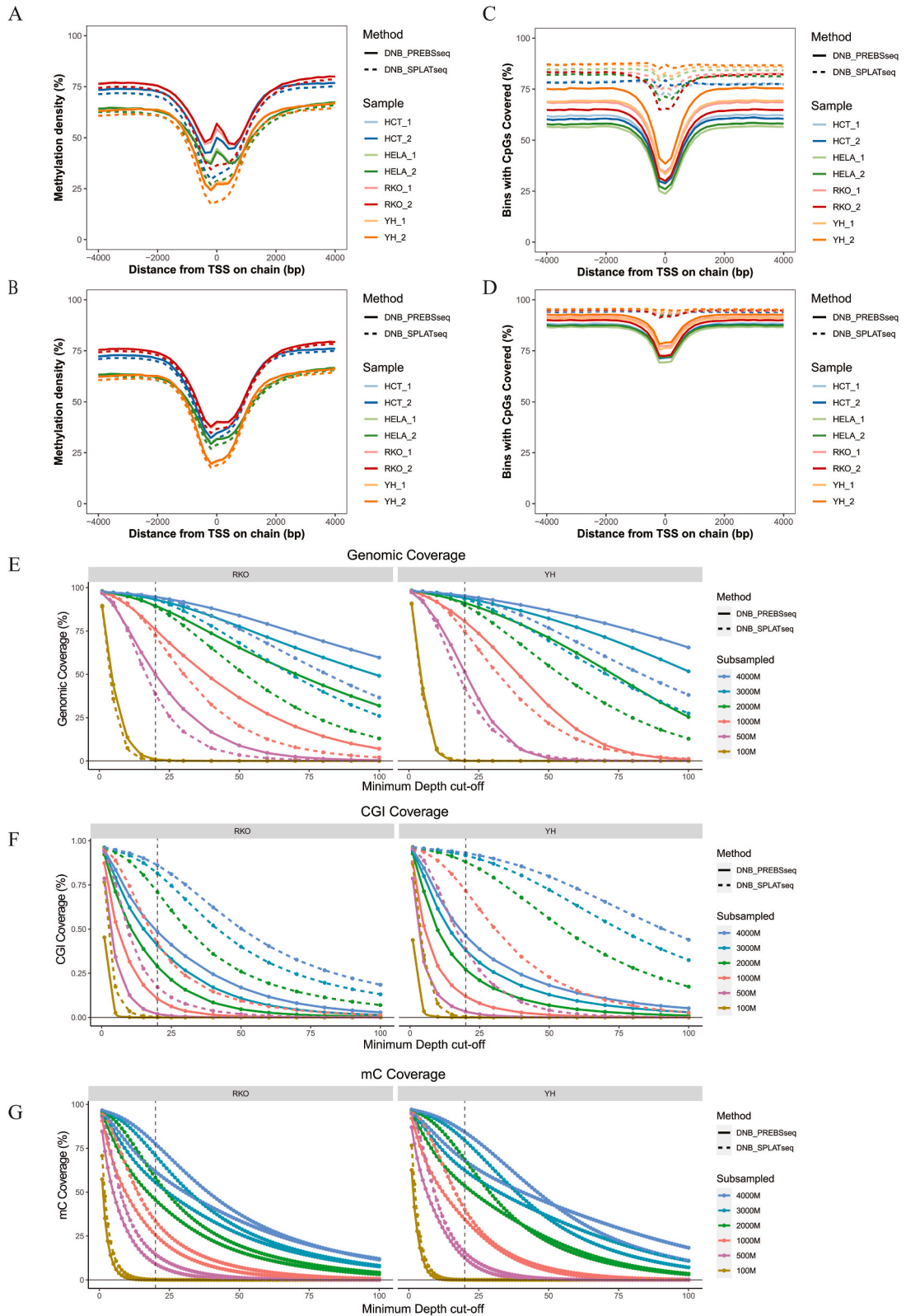
We also showed the genome-wide methylation profiles around the transcript start sites (TSS) (defined as ± 4 kb from the TSS) at 5X mC depth. As previously reported, the methylation levels dropped significantly at the TSS point and then rose again in DNB_SPLATseq data [33], while the DNB_PREBSseq data had a slight upward shift at the TSS point (Fig. 5 A), especially at 5X depth (Fig. 5 B). The changes in methylation density at the TSS may be reflected by the coverage of the mC at the TSS point in the DNB_PREBSseq data, which had a significant reduction, especially at 5X depth cutoff mC, whereas the coverage of the mC at the TSS point in the DNB_SPLATseq data was nearly unchanged between different depths (Fig. 5C, D). TSS analysis of ENCODE data were consistent with DNB_PREBSseq data, indicating that this was also an inherent flaw of the pre-bisulfite treatment method (Supplementary Fig. 4). The TSS results of SPLAT SRA data were in line with DNB_SPLATseq data, the 1X depth cutoff mC in TSS area were approximately horizontal, and there was no drop at TSS 0 position (Supplementary Fig. 5 D). To sufficiently support the accuracy of measured methylation levels of our platform, WGBS data from the same passages of YH cell line was used to evaluate the correlations between DNA methylation levels on DNBSEQ-Tx and NovaSeq 6000. Three replicate libraries were made using 200 ng of YH cell DNA in which two replicates DNB_SPLATseq libraries were applied to DNBSEQ-Tx, in order to show the stability of our platform. One Accel-NGS Methyl-Seq library was sequenced on Illumina NovaSeq 6000. Correlations between data from DNBSEQ-Tx and NovaSeq 6000 at coverage depths cutoff of 5X and 10X showed high correlations between the results from these two platforms (the Pearson's correlation coefficient was 0.94 for 5X and 0.95 for 10X) (Supplementary Fig. 6 A, B, C). In addition to this, the distribution plot of methylation rate over 20 kb bins across the genome had also been plotted, three libraries from the same source had very similar distribution of methylation rate (Supplementary Fig. 7).

Lastly, to determine the appropriate amount of sequencing data for methylation analysis, we performed downsampling analysis on YH and RKO cell line samples data of the two library construction methods, showing the final percentage of the genome and mC covered at different minimum depths of YH and RKO samples under different pairs of raw data. Although the DNB_SPLATseq method data generally had a low genome coverage due to its higher duplication rate compared with DNB_PREBSseq, which will eventually increase with the increase in the number of raw data. Therefore, the DNB_SPLATseq method could be recommended for the DNBseq-Tx platform because its CGI region and mC coverage is substantially higher than that of the DNB_PREBSseq data (Fig. 5 E, F, G).

5. Discussion

In the present study, we assessed the general performance of two methylation library preparation methods on the DNBSEQ-Tx platform. Comprehensive comparisons showed that DNB_SPLATseq was superior to DNB_PREBSseq in terms of unique alignment rate, coverage uniformity, and CGI coverage. Although the library prepared by DNB_SPLATseq had a higher adapter contamination ratio due to the shorter inserts, it could be improved by selecting longer inserts during library preparation. It could also be improved by making blocking modifications at all ends of the oligomer, which may reduce linker dimers. Simultaneously, the higher duplication of the DNB_SPLATseq library under the same amount of data may be enhanced by shortening the PCR cycle and increasing the DNA input, which was lower in our study than that of the DNB_PREBSseq library.

Our results further suggested that bisulfite-mediated DNA degradation is the underlying cause for biases in WGBS data. During bisulfite treatment, the DNA usually undergoes a strong degradation, which affects the GC-biasing and the estimation of methylation levels in some GC-rich regions. DNB_PREBSseq is a pre-bisulfite approach, the DNA is fragmented and adapter is tagged prior to bisulfite conversion. So, it shows poor coverage in G-rich regions and significantly higher coverage of C-poor regions that leads to an overestimation of the absolute methylation level in GC-rich regions like 5'UTR/CGI/promoter. DNB_SPLATseq is a post-bisulfite method, the DNA is adapter-tagged after bisulfite conversion. During this procedure, DNA with desired size were bisulfite converted and fragmented simultaneously from gDNA by bisulfite-induced fragmentation. Therefore, it allows less GC-biasing and more accurate estimation of methylation level. Additional comparisons of the ENCODE HeLa WGBS data showed highly comparable performance between DNBSEQ and HiSeq platforms, especially when compared with DNB_PREBSseq, which was also based on a pre-bisulfite treatment method. DNB_SPLATseq data not only had better performance than DNB_PREBSseq data but also had better performance than ENCODE HeLa WGBS data, it could still cover a higher percentage of the CGI region under the premise of less data than ENCODE. In the context of coverage of different elements, the coverage of the two repeats of the DNB_SPLATseq library were more uniform. In addition, on the basis of similar average depth of the genome, the DNB_SPLATseq library clearly had higher coverage and a more consistent mean methylation state in the CGI, gene 5'UTRs, and gene promoter regions. The whole genomic-wide methylation



(caption on next page)

Fig. 5. The genome-wide performance of methylation profiles around transcript start sites (TSS) at 5X and 1X mC depth and downsampling analysis on YH and RKO cell line samples based on two library construction methods. (A, B). Methylation profiles around TSS at 5X and 1X mC depth. (C, D). mC coverage around TSS at 5X and 1X mC cutoff. (E, F, G). Genome, CGI and mC site coverage at different minimum depths across different number of raw read pairs.

levels obtained by the DNB_SPLATseq and DNB_PREBSseq methods were consistent, however, differences in the methylation levels of certain genomic elements were revealed by further analysis. The DNB_PREBSseq method involved DNA fragmentation due to bisulfite treatment during library preparation, which made it lose part of the template in the GC-rich region, affecting the coverage of mC. In contrast, DNB_SPLATseq was constructed after bisulfite treatment, so its data could preserve a higher proportion of mC sites, and the coverage was more uniform across different elements.

According to downsampling analysis on YH and RKO cell line samples of the two library construction methods, our results suggest that at least 500 million raw library reads (pair-end reads) would be necessary to achieve 50% genome coverage at a minimum depth of 20X with the DNB_PREBSseq method, whereas more than 500 billion raw read pairs would be needed for DNB_SPLATseq (Fig. 5 E). However, in terms of CGI coverage, at least 4000 million raw read pairs would be required to achieve 50% CGI coverage at a minimum depth of 20X with the DNB_PREBSseq method, whereas only 1000 million raw read pairs are enough for DNB_SPLATseq (Fig. 5 F).

In conclusion, WGBS based on the DNBSEQ-Tx platform could accurately detect features of cytosine methylation modification. The DNB_SPLATseq library constructed after DNA bisulfite-treated could improve the CGI coverage compared with DNB_PREBSseq. Moreover, due to its low DNA demand and low library construction cost, the DNB_SPLATseq method is a primarily recommended method for large-scale genome-wide WGBS on the DNBSEQ-Tx platform.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repositories and accession number(s) can be found below: CNGB Sequence Archive (CNSA) of CNGBdb, CNP0003306.

Author contribution statement

Boyang Cao, Huijuan Luo, Fuqiang Li and Cong Lin conceived and designed the experiments;
 Boyang Cao and Huijuan Luo performed the experiments;
 Boyang Cao and Huijuan Luo analyzed and interpreted the data;
 Boyang Cao, Huijuan Luo, Fuqiang Li, Cong Lin, Tian Luo, Nannan Li, Kang Shao, Kui Wu and Sunil Kumar Sahu contributed reagents, materials, analysis tools or data;
 Boyang Cao, Huijuan Luo, Fuqiang Li, Cong Lin, Tian Luo, Nannan Li, Kang Shao, Kui Wu and Sunil Kumar Sahu wrote the paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgments

We thank China National GeneBank (CNGB) for assistance with sequencing and computational resources. This study was funded by the Guangdong Provincial Key Laboratory of Human Disease Genomics (2020B1212070028).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2023.e16571>.

References

- [1] K.D. Robertson, DNA methylation and human disease, *Nat. Rev. Genet.* 6 (8) (2005 Aug) 597–610, <https://doi.org/10.1038/nrg1655>. PMID: 16136652.
- [2] Z.D. Smith, A. Meissner, DNA methylation: roles in mammalian development, *Nat. Rev. Genet.* 14 (3) (2013 Mar) 204–220, <https://doi.org/10.1038/nrg3354>. Epub 2013 Feb 12. PMID: 23400093.
- [3] A.P. Bird, CpG-rich islands and the function of DNA methylation, *Nature* 321 (6067) (1986 May 15–21) 209–213, <https://doi.org/10.1038/321209a0>. PMID: 2423876.
- [4] R. Holliday, J.E. Pugh, DNA modification mechanisms and gene activity during development, *Science* 187 (4173) (1975 Jan 24) 226–232. PMID: 1111098.
- [5] R.L. Momparler, V. Bovenzi, DNA methylation and cancer, *J. Cell. Physiol.* 183 (2) (2000 May) 145–154, [https://doi.org/10.1002/\(SICI\)1097-4652\(200005\)183:2<145::AID-JCP1>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1097-4652(200005)183:2<145::AID-JCP1>3.0.CO;2-V). PMID: 10737890.
- [6] G. Egger, G. Liang, A. Aparicio, P.A. Jones, Epigenetics in human disease and prospects for epigenetic therapy, *Nature* 429 (6990) (2004 May 27) 457–463, <https://doi.org/10.1038/nature02625>. PMID: 15164071.
- [7] K.D. Robertson, DNA methylation and human disease, *Nat. Rev. Genet.* 6 (8) (2005 Aug) 597–610, <https://doi.org/10.1038/nrg1655>. PMID: 16136652.

- [8] S. Jayaraman, Epigenetics of autoimmune diabetes, *Epigenomics* 3 (5) (2011 Oct) 639–648, <https://doi.org/10.2217/epi.11.78>. PMID: 22126251.
- [9] C.J. Marsit, E.A. Houseman, B.C. Christensen, K. Eddy, R. Bueno, D.J. Sugarbaker, H.H. Nelson, M.R. Karagas, K.T. Kelsey, Examination of a CpG island methylator phenotype and implications of methylation profiles in solid tumors, *Cancer Res.* 66 (21) (2006 Nov 1) 10621–10629, <https://doi.org/10.1158/0008-5472.CAN-06-1687>. PMID: 17079487.
- [10] M. Jordà, A. Díez-Villanueva, I. Mallona, B. Martín, S. Lois, V. Barrera, M. Esteller, T. Vavouri, M.A. Peinado, The epigenetic landscape of Alu repeats delineates the structural and functional genomic architecture of colon cancer cells, *Genome Res.* 27 (1) (2017 Jan) 118–132, <https://doi.org/10.1101/gr.207522.116>. Epub 2016 Dec 20. PMID: 27999094; PMCID: PMC5204336.
- [11] D. Barros-Silva, C.J. Marques, R. Henrique, C. Jerónimo, Profiling DNA methylation based on next-generation sequencing approaches: new insights and clinical applications, *Genes* 9 (9) (2018) 429, <https://doi.org/10.3390/genes9090429>. PMID: 30142958; PMCID: PMC6162482.
- [12] <https://ihcc-epigenomes.org/research/reference-epigenome-standards/>.
- [13] MGI. INTRODUCTION TO MGI SEQUENCING TECHNOLOGY. <https://en.mgi-tech.com/products/>.
- [14] J. Foox, S.W. Tighe, C.M. Nicolet, J.M. Zook, M. Byrska-Bishop, W.E. Clarke, M.M. Khayat, M. Mahmoud, P.K. Laaguiby, Z.T. Herbert, D. Warner, G.S. Grills, J. Jen, S. Levy, J. Xiang, A. Alonso, X. Zhao, W. Zhang, F. Teng, Y. Zhao, H. Lu, G.P. Schroth, G. Narzisi, W. Farmerie, F.J. Sedlazeck, D.A. Baldwin, C.E. Mason, Performance assessment of DNA sequencing platforms in the ABRF next-generation sequencing study, *Nat. Biotechnol.* 39 (9) (2021 Sep) 1129–1140, <https://doi.org/10.1038/s41587-021-01049-5>. Epub 2021 Sep 9. Erratum in: *Nat Biotechnol.* 2021 Nov;39(11):1466. PMID: 34504351; PMCID: PMC8985210.
- [15] K.R. Kumar, M.J. Cowley, R.L. Davis, Next-generation sequencing and emerging technologies, *Semin. Thromb. Hemost.* 45 (7) (2019 Oct) 661–673, <https://doi.org/10.1055/s-0039-1688446>. Epub 2019 May 16. PMID: 31096307.
- [16] S.A. Jeon, J.L. Park, S.J. Park, J.H. Kim, S.H. Goh, J.Y. Han, S.Y. Kim, Comparison between MGI and Illumina sequencing platforms for whole genome sequencing, *Genes Genomics* 43 (7) (2021 Jul) 713–724, <https://doi.org/10.1007/s13258-021-01096-x>. Epub 2021 Apr 17. PMID: 33864614.
- [17] H.M. Kim, S. Jeon, O. Chung, J.H. Jun, H.S. Kim, A. Blazyte, H.Y. Lee, Y. Yu, Y.S. Cho, D.M. Bolser, J. Bhak, Comparative analysis of 7 short-read sequencing platforms using the Korean Reference Genome: MGI and Illumina sequencing benchmark for whole-genome sequencing, *GigaScience* 10 (3) (2021 Mar 12) giab014, <https://doi.org/10.1093/gigascience/giab014>. PMID: 33710328; PMCID: PMC7953489.
- [18] Y. Sun, J. Yuan, L. Wu, M. Li, X. Cui, C. Yan, L. Du, L. Mao, J. Man, W. Li, K. Kristiansen, X. Wu, W. Pan, Y. Yang, Panel-based NGS reveals disease-causing mutations in hearing loss patients using BGISEQ-500 platform, *Medicine (Baltim.)* 98 (12) (2019 Mar), e14860, <https://doi.org/10.1097/MD.00000000000014860>. PMID: 30896630; PMCID: PMC6709004.
- [19] Y. Xu, Z. Lin, C. Tang, Y. Tang, Y. Cai, H. Zhong, X. Wang, W. Zhang, C. Xu, J. Wang, J. Wang, H. Yang, L. Yang, Q. Gao, A new massively parallel nanoball sequencing platform for whole exome research, *BMC Bioinf.* 20 (1) (2019 Mar 25) 153, <https://doi.org/10.1186/s12859-019-2751-3>. PMID: 30909888; PMCID: PMC6434795.
- [20] J. Patterson, E.J. Carpenter, Z. Zhu, D. An, X. Liang, C. Geng, R. Drmanac, G.K. Wong, Impact of sequencing depth and technology on de novo RNA-Seq assembly, *BMC Genom.* 20 (1) (2019 Jul 23) 604, <https://doi.org/10.1186/s12864-019-5965-x>. PMID: 31337347; PMCID: PMC6651908.
- [21] K.N. Natarajan, Z. Miao, M. Jiang, X. Huang, H. Zhou, J. Xie, C. Wang, S. Qin, Z. Zhao, L. Wu, N. Yang, B. Li, Y. Hou, S. Liu, S.A. Teichmann, Comparative analysis of sequencing technologies for single-cell transcriptomics, *Genome Biol.* 20 (1) (2019 Apr 9) 70, <https://doi.org/10.1186/s13059-019-1676-5>. PMID: 30961669; PMCID: PMC6454680.
- [22] C. Fang, H. Zhong, Y. Lin, B. Chen, M. Han, H. Ren, H. Lu, J.M. Lubner, M. Xia, W. Li, S. Stein, X. Xu, W. Zhang, R. Drmanac, J. Wang, H. Yang, L. Hammarström, A.D. Kostic, K. Kristiansen, J. Li, Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing, *GigaScience* 7 (3) (2018 Mar 1) 1–8, <https://doi.org/10.1093/gigascience/gix133>. PMID: 29293960; PMCID: PMC5848809.
- [23] M.A. Urlich, J.R. Nery, R. Lister, R.J. Schmitz, J.R. Ecker, MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing, *Nat. Protoc.* 10 (3) (2015 Mar) 475–483, <https://doi.org/10.1038/nprot.2014.114>. Epub 2015 Feb 18. PMID: 25692984; PMCID: PMC4465251.
- [24] A. Raine, E. Manlig, P. Wahlberg, A.C. Syvänen, J. Nordlund, SPLinted Ligation Adapter Tagging (SPLAT), a novel library preparation method for whole genome bisulphite sequencing, *Nucleic Acids Res.* 45 (6) (2017 Apr 7) e36, <https://doi.org/10.1093/nar/gkw1110>. PMID: 27899585; PMCID: PMC5389478.
- [25] N. Olova, F. Krueger, S. Andrews, D. Oxley, R.V. Berrens, M.R. Branco, W. Reik, Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data, *Genome Biol.* 19 (1) (2018 Mar 15) 33, <https://doi.org/10.1186/s13059-018-1408-2>. Erratum in: *Genome Biol.* 2019 Feb 22;20(1):43. PMID: 29544553; PMCID: PMC5856372.
- [26] Y. Hou, K. Wu, X. Shi, F. Li, L. Song, H. Wu, M. Dean, G. Li, S. Tsang, R. Jiang, X. Zhang, B. Li, G. Liu, N. Bedekar, N. Lu, G. Xie, H. Liang, L. Chang, T. Wang, J. Chen, Y. Li, X. Zhang, H. Yang, X. Xu, L. Wang, J. Wang, Comparison of variations detection between whole-genome amplification methods used in single-cell resequencing, *GigaScience* 4 (2015 Aug 6) 37, <https://doi.org/10.1186/s13742-015-0068-3>. PMID: 26251698; PMCID: PMC4527218.
- [27] <https://www.sentieon.com/>.
- [28] R.G. Cavalcante, M.A. Sartor, annotatr: genomic regions in context, *Bioinformatics* 33 (15) (2017 Aug 1) 2381–2383, <https://doi.org/10.1093/bioinformatics/btx183>. PMID: 28369316; PMCID: PMC5860117.
- [29] M. Suzuki, W. Liao, F. Wos, A.D. Johnston, J. DeGrazia, J. Ishii, T. Bloom, M.C. Zody, S. Germer, J.M. Greally, Whole-genome bisulfite sequencing with improved accuracy and cost, *Genome Res.* 28 (9) (2018 Sep) 1364–1371, <https://doi.org/10.1101/gr.232587.117>. Epub 2018 Aug 9. PMID: 30093547; PMCID: PMC6120621.
- [30] K.D. Hansen, B. Langmead, R.A. Irizarry, BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions, *Genome Biol.* 13 (10) (2012 Oct 3) R83, <https://doi.org/10.1186/gb-2012-13-10-r83>. PMID: 23034175; PMCID: PMC3491411.
- [31] X. Lin, D. Sun, B. Rodriguez, Q. Zhao, H. Sun, Y. Zhang, W. Li, BSeQC: quality control of bisulfite sequencing experiments, *Bioinformatics* 29 (24) (2013 Dec 15) 3227–3229, <https://doi.org/10.1093/bioinformatics/btt548>. Epub 2013 Sep 23. PMID: 24064417; PMCID: PMC3842756.
- [32] R. Vaisvila, V.K.C. Ponnaluri, Z. Sun, B.W. Langhorst, L. Saleh, S. Guan, N. Dai, M.A. Campbell, B.S. Sexton, K. Marks, M. Samaranyake, J.C. Samuelson, H. E. Church, E. Tamana, I.R. Corréa Jr., S. Pradhan, E.T. Dimalanta, T.C. Evans Jr., L. Williams, T.B. Davis, Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA, *Genome Res.* 31 (7) (2021 Jun 17) 1280–1289, <https://doi.org/10.1101/gr.266551.120>. Epub ahead of print. PMID: 34140313; PMCID: PMC8256858.