

RESEARCH

Open Access



Comparison of current methods for genome-wide DNA methylation profiling

Ana Regina de Abreu^{1,2†}, Joe Ibrahim^{1,2†}, Vasileios Lemonidis^{1,2}, Ligia Mateiu¹, Guy Van Camp¹ and Ken Op de Beeck^{1,2*}

Abstract

Background DNA methylation is an epigenetic mechanism involved in gene regulation and cellular differentiation. Accurate and comprehensive assessment of DNA methylation patterns is thus essential for understanding their role in various biological processes and disease mechanisms. Bisulfite sequencing has long been the default method for analyzing methylation marks due to its single-base resolution, but the associated DNA degradation poses a concern. Although several methods have been proposed to circumvent this issue, there is no clear consensus on which method might be better suited for specific study designs.

Results We conducted a comparative evaluation of four DNA methylation detection approaches: whole-genome bisulfite sequencing (WGBS), Illumina methylation microarray (EPIC), enzymatic methyl-sequencing (EM-seq) and third-generation sequencing by Oxford Nanopore Technologies (ONT). DNA methylation profiles were assessed across three human genome samples derived from tissue, cell line, and whole blood. We systematically compared these methods in terms of resolution, genomic coverage, methylation calling accuracy, cost, time, and practical implementation. EM-seq showed the highest concordance with WGBS, indicating strong reliability due to their similar sequencing chemistry. ONT sequencing, while showing lower agreement with WGBS and EM-seq, captured certain loci uniquely and enabled methylation detection in challenging genomic regions. Despite a substantial overlap in CpG detection among methods, each method identified unique CpG sites, emphasizing their complementary nature.

Conclusions Our findings underscore the strengths and limitations of current DNA methylation detection methods. EM-seq and ONT emerge as robust alternatives to WGBS and EPIC, offering unique advantages: EM-seq delivers consistent and uniform coverage, while ONT excels in long-range methylation profiling and access to challenging genomic regions. These insights provide practical guidance for method selection based on specific experimental goals.

Keywords DNA methylation, Cytosine, Whole-genome bisulfite sequencing (WGBS), Enzymatic methyl-sequencing (EM-seq), Oxford Nanopore Technologies (ONT), Illumina EPIC array, Coverage, Epigenomics

[†]Ana Regina de Abreu and Joe Ibrahim have contributed equally to this work.

*Correspondence:

Ken Op de Beeck

ken.opdebeeck@uantwerpen.be

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

DNA methylation, a fundamental epigenetic modification, regulates gene expression without altering the DNA sequence. Epigenetic modifications, including DNA methylation, orchestrate essential biological processes such as genomic imprinting, X-chromosome inactivation, gene stability and expression regulation, embryonic development, and aging [1–3]. Epigenetic changes have attracted significant attention in recent years as they are known to be involved in the initiation and progression of several human diseases, including cancer.

While DNA methylation predominantly occurs at cytosine–phosphate–guanine (CpG) dinucleotide sites, it also extends to non-CpG sites, albeit to a lesser extent, exerting distinct effects on gene structure and function [4, 5]. Notably, the impact of DNA methylation on gene expression varies depending on the genomic location of the gene. Methylation within promoter regions typically suppresses gene expression, whereas methylation of gene bodies involves complex regulatory mechanisms that influence gene expression and maintain genomic stability [6]. It can suppress gene expression by promoting chromatin densification [7] and interacting with functional elements, such as repetitive elements [8]. However, it can also increase transcription by regulating splicing processes [9]. While much attention has been given to promoter methylation, the role of methylation at regulatory elements beyond promoters remains underexplored. Studies have begun to reveal the more complex relationship between methylation changes across the genome and gene expression, shedding light on the broader regulatory role of DNA methylation [10]. Considering the impact of DNA methylation on gene regulation, precise and reliable detection methods are paramount for understanding its functional significance and potential diagnostic and therapeutic applications.

Bisulfite conversion is a reliable method for determining the methylation status of cytosines within a DNA sequence. Various generations of bisulfite-based microarrays, particularly those developed by Illumina, have been used to profile the DNA methylation of CpG sites in thousands of human samples because of their low cost and easy, standardized data processing and analysis. The Infinium BeadChip platforms, such as the Infinium MethylationEPIC array and the Infinium HumanMethylation450K array (HM450K), utilize the same technology, with the EPIC array assessing nearly double the number of sites as the 450 K array. The first version of EPIC interrogates >850,000 methylation sites covering 99% of the RefSeq genes [11]. In comparison, the improved second version covers over 935,000 sites, 77.63% of which are homologous to the first version and more than 200,000

of which are new CpGs covering regions located in open chromatin and enhancer regions [12–15].

The next significant advancement in genome-wide DNA methylation profiling was the integration of bisulfite conversion into next-generation sequencing (NGS), a methodology known as whole-genome bisulfite sequencing (WGBS). The major advantage of WGBS lies in its ability to assess the methylation state of nearly every CpG site across the genome, achieving whole-genome coverage of approximately 80% of all CpG sites. Moreover, it can determine absolute DNA methylation levels and reveal the context of methylation sequences. Despite its strengths, WGBS has certain limitations. The cost considerations, challenges associated with analyzing NGS data and biases introduced during library preparation are noteworthy [16].

Both EPIC and WGBS are potent tools for assessing genome-wide DNA methylation, although certain pitfalls should be considered [17]. Bisulfite treatment is a harsh method involving extreme temperatures and strong basic conditions, introducing single-strand breaks and substantial fragmentation of DNA [16, 18]. However, if milder conditions are applied to mitigate the risk of DNA degradation, such as a lower conversion temperature, low sodium bisulfite molarity, and alkaline denaturation, incomplete conversion is a concern regarding the methylation status of specific cytosines. The incomplete cytosine conversion to uracil, which is inherent to bisulfite treatment, may yield false-positive results because unconverted unmethylated CpG sites are misinterpreted as methylated [16, 18]. Because bisulfite reacts only with cytosines that are not involved in base pairing, incomplete denaturation of the DNA template or its partial renaturation during bisulfite treatment are the most obvious explanations for the incomplete conversion of cytosine. This is particularly problematic for the biological interpretation of the methylation state of GC-rich regions, such as CpG islands.

Two new technologies aim to overcome the limitations associated with bisulfite conversion: enzymatic methyl-seq (EM-seq) [19] and third-generation sequencing [20, 21]. EM-seq uses the TET2 enzyme for conversion and protection of 5-methylcytosine (5mC) to 5-carboxylcytosine (5caC). In addition, T4 β -glucosyltransferase (T4-BGT) is included in this reaction to specifically glucosylate any 5-hydroxymethylcytosine (5hmC), which protects 5hmC from further oxidation and deamination. After this conversion, APOBEC selectively deaminates unmodified cytosines, while all modified cytosines—including 5mC, 5hmC, 5caC, and 5-formylcytosine (5fC)—are protected from deamination. Unlike bisulfite treatment, enzymatic conversion does not further fragment the DNA after

adapter ligation, thereby preserving DNA integrity and reducing sequencing bias while also improving CpG detection [19]. Moreover, compared with WGBS, EM-seq can handle lower amounts of DNA input. The primary strength of third-generation sequencing techniques lies in their ability to directly detect DNA methylation without requiring chemical or enzymatic treatments. Direct sequencing is based on electrical readouts for nanopore sequencing (Oxford Nanopore Technologies). The process consists of threading DNA through protein nanopores embedded in synthetic membranes, where changes in electrical current are measured as individual bases pass through the pore [22]. Each nucleotide is uniquely characterized by different structural and geometrical properties, which alter its resistance to electrical current. For example, 5C, 5mC and 5hmC can be distinguished from each other by electric signal deviations [20, 21]. In addition, nanopore sequencing benefits from long-read sequencing, enabling efficient resolution of highly dense CG genomic regions. A downside, however, lies in the inability to amplify DNA for methylation detection, necessitating relatively high amounts of DNA for successful sequencing (approximately one μg of 8 kb fragments) [23].

While each technology has distinct strengths and limitations for methylation detection, direct comparisons using biological samples different from cell lines are scarce. This study addresses this gap by analyzing genome-wide methylation profiles across various sample sources, including a cell line, blood and tissue. Beyond evaluating their resolution and performance, we also assess these methods in terms of cost-effectiveness and practicality, offering insights tailored to different research scenarios. By considering both technical and practical aspects, our findings can guide researchers in selecting the most suitable approach for their needs.

Materials and methods

Human ethics approval

Two samples in the study were of human origin. Biopsies of colorectal cancer were stored as fresh frozen tissue embedded in optimal cutting temperature (OCT) compound at -80°C . Freshly frozen clinical specimens were obtained from the Antwerp University Hospital. The breast cancer (MCF7, breast cancer) cell line was purchased from ATCC. The cells were cultivated in Dulbecco's modified Eagle's medium (DMEM) (Gibco, 11960044) supplemented with 10% fetal bovine serum (FBS) (Gibco, 26140079). This study was approved by the Clinical Ethics Committee of the Antwerp University

Hospital. Informed consent was obtained from a healthy volunteer from which fresh blood was obtained.

DNA extraction

DNA from fresh frozen tissue was extracted using the Nanobind Tissue Big DNA Kit (Circulomics). For the extraction of DNA from the cell line, the DNeasy Blood & Tissue Kit (Qiagen) was used. The salting-out method was used for whole-blood DNA extraction [24]. DNA was extracted from the samples following the manufacturer's instructions. Following extraction, the purity of the DNA was assessed using a NanoDrop (Thermo Scientific) 260/280 and 260/230 ratio measurement, which was quantified using an Invitrogen Qubit 3.0 fluorometer.

Illumina MethylationEPIC array

Five hundred nanograms of DNA were bisulfite treated using the EZ DNA Methylation Kit (Zymo Research, USA) following the manufacturer's recommendations for Infinium assays. To assess the methylation status of the CpG sites, the Infinium MethylationEPIC v1.0 BeadChip array was used. The hybridization volume of the processed sample used to load the microarray was 26 μl .

The *minfi* (v1.48.0) package [25] was used to perform initial quality checks and preprocessing. Methylation is reported as a β -value, which is the ratio of the methylated probe intensity to the sum of the methylated and unmethylated probe intensities and ranges from 0 for unmethylated probes to 1 for fully methylated probes. β -values were calculated and normalized using the beta-mixture quantile normalization method [26]. Further analyses were carried out using the *ChAMP* package (v2.12.2) [27], where underperforming and control probes were removed from the downstream analysis. This included probes with a detection p -value > 0.01 , control probes, multihit probes, and probes with known single nucleotide polymorphisms (SNPs).

WGBS

High-molecular-weight DNA (1 μg) was sent to BGI Genomics (Shenzhen, China), following their recommendations. The DNA samples were transferred to a Covaris microTUBE and sheared into 250 bp fragments using a CovarisTM focused-ultrasonicator (Covaris, MA, USA). Sonicated DNA was then size-selected with $0.8\times + 0.2\times$ AMPure XP beads (Agencourt). Lambda DNA (200 ng/ μl) underwent the same fragmentation and size selection procedures as the DNA samples. The DNA samples were subsequently quantified using the Qubit[®] dsDNA HS Assay Kit. DNA samples (100 ng) and lambda DNA (2 ng) were transferred to a PCR strip tube, and the MGIEasy Whole Genome Bisulfite Sequencing Library Prep Kit was used to construct the libraries, whereas

the EZ DNA Methylation Gold Kit (Zymo Research) was used for bisulfite treatment and cleanup of adapter-ligated products. PCR amplification was performed with an initial denaturation at 95 °C for 2 min, followed by 13 cycles of 98 °C for 20 s, 62 °C for 20 s, and 72 °C for 30 s, and a final extension at 72 °C for 3 min. Afterwards, the libraries were 150 bp paired-end sequenced on the MGISEQ-2000.

After sequencing, the resulting paired-end reads from the FASTQ files were passed through the *fastp* tool (v0.23.4) for quality control (QC) filtering and trimming [28]. Reads with a PHRED quality of less than 15 in more than 40% of the bases were removed. Read adapters and polyG tails were trimmed. The resulting reads with a length of less than 15 base pairs were discarded. The *bwa-meth* (v0.2.7) aligner subsequently aligned the QC-passing reads against the GRCh38/hg38 reference genome. The average sequencing depth was computed, and the binary alignment map (BAM) file was downsampled to the lowest observed depth among samples and WGS methods, which was evaluated as 29x, covered by any read in a pair. Coordinate sorting of the BAM files was performed with *samtools* (v1.20), followed by duplicate marking by *samblaster* (v0.1.26) [29]. *MethylDackel* (v0.6.1) was used to produce a browser-extensible data (BED) file containing all detected CpG regions in both strands, along with the corresponding numbers of methylated and unmethylated occurrences. In our bisulfite sequencing data, methylated cytosines represent both 5mC and 5hmC, as both modifications cannot be distinguished by this method. We did not perform additional experimental steps to separate 5mC from 5hmC.

EM-seq

Two hundred nanograms of genomic DNA were diluted to 50 µL of 10 mM Tris 0.1 mM EDTA, pH 8.0, following the manufacturer's recommendations. The DNA was then transferred to a Covaris microTUBE and sheared to an average size of 240–290 bp using a CovarisTM focused-ultrasonicator (Covaris, MA, USA). A total of 0.1 ng of CpG methylated pUC19, and two nanograms of unmethylated lambda control DNA underwent the same fragmentation procedure as the DNA samples. Sonicated DNA was used to construct libraries with the NEB-Next Enzymatic Methyl-seq Kit (New England Biolabs, NEB) per the manufacturer's instructions. The denaturation method used was 0.1 N sodium hydroxide. The resulting libraries were analyzed and quantified using a D1000 High Sensitivity ScreenTape for TapeStation (Agilent Technologies). Library qualities were checked on an Illumina MiSeq platform with a 10% PhiX spike-in. The whole-genome libraries were 150 bp paired-end sequenced using the Illumina NovaSeq 6000 sequencer

on an S1 flow cell (Illumina) with 5% PhiX. Base calling was performed using Illumina RTA3, and the output of the NGS control software (NCS) was demultiplexed and converted to FASTQ format with Illumina *Bcl2fastq* (v1.9.0). Further bioinformatic analyses were conducted using the workflow described above to minimise any variation introduced in the computational steps. EM-seq can distinguish 5mC from 5hmC but it involves two distinct enzymatic conversion reactions with different enzyme uses. We did not perform these additional steps, as our study was not designed to differentiate between these two modifications.

Nanopore sequencing (PromethION)

DNA samples were sequenced at the VIB Center for Molecular Neurology (Antwerp, Belgium). The DNA concentration was measured using the Qubit dsDNA High Sensitivity Assay Kit (Q33231, Thermo Fisher), the purity with Little Lunatic (Unchained Labs), and the integrity on a Fragment Analyzer using the DNF-464 High Sensitivity Large Fragment 50-Kb Kit (Agilent). The DNA was sheared with the Megaruptor3 (Diagenode) to approximately ~30 kb (21.5–26 kb after shearing). Short fragments were eliminated via SRE XS (Circulomics) with a cutoff of 10 kb. Final checks on the Fragment Analyzer confirmed the size distribution. Sheared and size-selected DNA was prepared from the SQK-LSK109 gDNA via a Ligation Sequencing Kit (Oxford Nanopore Technologies) with minor adaptations. Approximately 3000 ng of DNA was repaired and dA-tailed using NEBNext FFPE DNA repair mix and NEBNext Ultra II End repair/dA Tailing Module (M6630, E7546, NEB), followed by AMPure XP bead clean-up for increased adaptor ligation efficiency at a ratio of 1:1 (vol/vol) and extended incubation (10 min on Hulamixer). Sequencing adaptors were ligated onto DNA fragments using SQK-LSK109 reagents under optimized ligation conditions. Final clean-up was performed using 0.4×vol/vol AMPure XP with extended incubation (10 min on a Hulamixer), followed by washing with large fragment buffer (LFB, SQK-LSK109, ONT) and subsequent elution in 48 µL of elution buffer (EB, SQK-LSK109, ONT). Fifty femtomoles of adaptor-ligated DNA were loaded onto three distinct flow cells: Colorectal Cancer Tissue (727 ng), Breast Cancer Cell Line (705 ng) and Healthy Blood (785 ng). After 24 h and 48 h, nuclease flushes were performed with DNase I (M0303, NEB), followed by a library reload of 50 fmol. A total of 150 fmol of sample was added per flow cell.

Simultaneous basecalling and alignment were conducted using *Dorado*.

(v0.5.3) with the *dna_r9.4.1_e8_hac@v3.3* and 5mCG models to detect modified bases. The resulting BAM files were downsampled to 29×depth, as described earlier.

Modkit (v0.3.1) converted BAM files into BED files containing CpG site methylation data referenced to GRCh38/h38. As Nanopore sequencing directly detects and distinguishes between multiple DNA modifications, including 4mC, 5mC, 5hmC, and 6-methyladenine (6 mA) [30]; in our analysis, only 5mC calls were retained, ensuring that the ONT-derived methylation profiles exclusively represent 5mC.

Data analysis and comparison

The produced BED files were processed using in-house developed Python 3.10 pipelines. The following packages were used: *pandas* [31] for data frames manipulation; *scikit-learn* (v1.6.1) [32] and *SciPy* (v1.14.1) [33] for statistical processing; and *Matplotlib* (v3.9.2) [34], *seaborn* (v0.13.2) [35], and *adjustText* (commit aab1e19 on package's GitHub) [36] for visualization. Only CpG sites on chromosomes 1–22, X, and Y were considered. Coverage was computed as the sum of methylated and unmethylated counts per CpG. For the subsequent analysis, we define “NGS aggregates on X” as 16 data frames representing individual NGS-produced bed files (9), the mean aggregations of the concatenated NGS bed files by CpG and method on X (3), the mean aggregations of the concatenated NGS bed files by CpG and tissue type on X (3), and the total mean aggregation of all NGS bed files on X by CpG (1), where X is coverage, methylation, or a data-analysis specific value. Similarly, we define “total aggregates on X” as the aforementioned group, equipped with the corresponding microarray bed files, totaling 20 data frames. BED files were annotated with an in-house tool. Specifically, the structural annotations were formed by querying the UCSC EPD database whether a site belongs to a promoter, CpG island, shore (2 k base pairs upstream (south) and downstream (north) to an island) or shelf (2 k base pairs upstream of a south shore and 2 k base pairs downstream to a north shore); the ENSEMBL database whether a site belongs to a gene, intergenic region (complementary to the gene regions), coding sequence, exon, or 5' or a 3' untranslated region (5' UTR, 3' UTR). Additionally, six cancer-related genomic regions were utilized to annotate the CpG sites, producing cancer-guided annotations. The following is a description of each computational analysis performed on the generated files. (1) The empirical cumulative coverage distribution was computed by identifying the frequencies of each coverage occurrence and calculating the cumulative sum of those frequencies from highest to lowest coverage. NGS aggregates on cumulative coverage were produced. (2) The coverage and methylation distribution of the annotations were estimated from the BED files. For the methylation-specific analyses, in the case of an NGS BED file, a minimum coverage of 10× was used to remove

under-represented sites. Subsequently, violin plots were generated with *seaborn*. For the structural annotation, due to computational restraints, the input annotated bed files were randomly downsampled to 1%. No down-sampling took place for the cancer-guided annotation. Violin plots were produced for each total aggregate on methylation and each NGS aggregate on coverage. Analysis of variance (ANOVA) on the CpGs of each annotation group, with groups dictated by the aggregation, was applied using the *f_classif* scikit-learn function. (3) Uniquely missed and uniquely covered CpG sites were identified per method based on a 10× minimum coverage across all tissues. Annotated sets underwent the same process with additional grouping by annotation. (4) Density estimations were performed globally and for pairwise comparisons of coverage and methylation distributions across all total aggregates. (5) Inter-method agreement was assessed via Fleiss' Kappa. CpG sites with coverage $\geq 10\times$ were classified into five methylation levels: no methylation ($<20\%$), low methylation ($\geq 20\%$ and $<30\%$), medium methylation ($\geq 30\%$ and $<70\%$) and high methylation ($\geq 70\%$). For each tissue type, the common CpG sites were compared across methods. For each pairwise method combination, a frequency pivot table was generated based on the methylation levels, and Fleiss' kappa statistics with 95% confidence intervals were computed. The resulting confidence lower and upper bounds were omitted when the absolute difference from the estimated statistic was <0.01 .

Results

General overview and comparison of the different methods

This study included cell line, tissue, and blood samples to compare the performance of different methylation methods across various sample types. Table 1 presents a general comparison of the different approaches, including the average results of all samples. Generally, the largest DNA inputs are required for ONT. The current DNA input range for WGBS varies from 1 to 500 ng depending on the library preparation method, with traditional pre-bisulfite adapter ligation protocols requiring higher input due to bisulfite-induced degradation, while post-bisulfite adapter tagging (PBAT) protocols minimize DNA loss and enable effective library preparation from as little as one ng of DNA [37]. EM-seq can be performed with a DNA input ranging from 10 to 200 ng. We used 200 ng because testing the effectiveness of low DNA inputs was not our primary interest. Another metric is the turnaround time (TAT), which includes library preparation, sequencing and bioinformatic analyses. The relative TAT for EPIC is significantly lower compared to ONT, WGBS and EM-seq, as the readout takes only approximately

Table 1 Overview of the characteristics of each whole genome and targeted sequencing technology

	Whole genome			Targeted
	ONT	WGBS	EM-seq	EPIC
DNA Input	1–5 µg	1 – 500 ng	10–200 ng	250 ng–500 ng
Single-base Resolution	Yes	Yes	Yes	No
Approximate Run Time	80–84 h	20–24 h	20–24 h	30 min
Yield [Gb]	139	163	137	NA
Sequencing Coverage (x)	34	46	41	NA
Total Reads (M)	7.5	1132.5	986	NA
Number of QC-Passed Reads (M)	NA*	1041.7	976	NA
Percentage of Mapped Reads	90.8%	99.87%	99.99%	NA
Percentage of Mapped Duplicates	0	9.5%	7.0%	NA
Mean Read Length (bp)	16,922	150	150	NA
Longest Read (bp)	856,100	150	151	NA
Number of Called CpGs	56,715,299	53,912,145	54,178,937	865,596
Computational Run Time	Very high	High	High	Low
Complexity of Analytic Pipeline**	High	Medium	Medium	Low
Generated Data Size (GB)	~1200	~120	~70	~150mb
Turnaround Time (TAT)	7–12 days	6–10 days	6–10 days	3–4 days

For each parameter, the value represents the mean value of all samples

* Unlike WGBS and EM-seq, ONT incorporates QC at raw read levels. Instead of traditional quality control (QC) filtering, guppy uses multiple long reads to correct sequencing errors rather than removing reads. As ONT sequencing produces long reads, aggressive QC filtering could disproportionately remove long reads, severely affecting genome coverage

** ONT, WGBS and EM-seq require an analytic pipeline for sequencing data, including alignment, base calling, and QC criteria on sequencing depth. The high complexity of the ONT pipeline's is due to the need for software (e.g., Dorado) to detect nucleotides from signal-level data, whereas WGBS and EM-seq directly output nucleotide sequences. The EPIC array requires QC criteria that remove CpGs that could be affected by poor hybridization, such as CpGs close to known SNPs, and converts intensity signals to methylation values

30 min per BeadChip of eight samples, and the bioinformatic pipeline requires fewer resources and is less time-consuming. The library prep, on the other hand, takes two days. Conversely, ONT has the longest TAT because of its longer runtime and the greater number of required resources; consequently, ONT has the longest computing time. Notably, the sequencing coverage between WGBS and EM-seq is comparable, as both methods follow an identical experimental and bioinformatic workflow, differing only in the DNA conversion step. Nanopore sequencing, on the other hand, involves an entirely different chemistry and analysis pipeline, resulting in longer reads and lower sequencing coverage. However, the efficiency of long-read mapping is negatively impacted, as shown in Table 1 (90.8%). Longer reads require a more comprehensive match to align correctly, and if they contain large insertions, deletions, or rearrangements compared with the reference, their alignment can become more challenging. The N1 tier of Google Cloud machines was used throughout all computations. The machines spanned from n1-high CPU-8 to n1-standard-64, depending on the computational needs of each process, to strike a balance between cost and performance. For nanopore basecalling, an Nvidia Tesla T4 graphics

processing unit (GPU) accelerator was appended to the virtual machine.

Sample and sequencing quality assessment

We performed quality control of all sequence data generated in this study using the tools mentioned in the Materials and Methods section. The first set of metrics compared between the three whole-genome methods encompasses the quality of raw reads, including base quality and adapter contamination, as well as the effect of trimming on these reads. On average, the libraries generated approximately 1,132.5 million reads for WGBS and 986 million reads for EM-seq across the three samples, with no replicates considered (Table 1). After adapter trimming, quality trimming, and read length filtering, over 91.9% of all reads passed the quality thresholds for WGBS, and 98.9% of all the reads passed the quality thresholds for EM-seq. Generally, the raw base qualities, percentage of reads with adapter contamination, and percentage of bases trimmed are comparable between WGBS and EM-seq. However, this is not applicable for ONT since no filtering steps are performed. Following adapter and quality trimming of the raw reads, except for ONT, the reads were mapped to the human genome

(reference genome hg38), and the fraction of aligned read fragments was calculated. WGBS and EM-seq showed optimally aligned reads (99.87% and 99.99%, respectively). ONT constituted a relatively low percentage (90.8%). WGBS and EM-seq, on average, had 9.5% and 7.0% duplicate reads, respectively. For EPIC, the ratio of median intensities was well above 10.5, indicating a good early-quality output. On average, 8.87% of all probes had an arbitrary fluorescence unit (AFU) of ≤ 2000 . A total of 3.41% of all probes, on average per sample, had a detection p -value > 0.01 . In total, 4,120 probes had a p -value > 0.01 in more than 50% of all samples and were removed from the final probe list. After QC, normalization and filtering of the final β value matrix resulted in 865,455 CpG sites across all samples.

Genomic coverage and annotation of CpG sites

After alignment, we assessed the coverage distribution across different genomic regions for each method by mapping the reads to the GRCh38/h38 reference genome and annotating them. WGBS and EM-seq had an average coverage value of $46\times$ and $41\times$, respectively, whereas ONT had an average coverage of $34\times$ (Table 1). In total, ONT called the most CpG sites at ~ 56 million (for both strands), whereas both WGBS and EM-seq called ~ 54 million CpGs (for both strands), encompassing almost all the CpGs found in the reference human genome, considering both strands [38] (Table 1).

Each method exhibited different genomic coverage (Figure S1). To control for the effect of uneven sequencing depth across the different methods, which is particularly pronounced in WGBS, we downsampled the methylation call sets to a given mean coverage value (Fig. 1), as mentioned in the Materials and Methods. Generally, a minimum of $20\times$ is considered sufficiently deep to characterize the methylation status of a genomic region accurately. We normalized the methylation call sets to a mean of $29\times$ coverage per site per strand, as this was the lowest observed coverage, more specifically for the ONT Breast Cancer Cell Line (Figure S1 & S2). This downsampling leads to an expected $14.5\times$ coverage per strand-specific CpG site, presuming a homogeneous distribution of reads (methylation plots per method calculated on each sample are provided in Figure S5).

We also examined the coverage of CpG islands, CpG shores, CpG shelves, intergenic regions, genes, promoters, 5' untranslated regions (UTRs), coding sequences (CDs), introns and 3' UTRs, as well as GC-rich regions (Fig. 2). Overall, WGBS shows a larger spread in overall coverage depth for these regions compared to EM-seq and ONT. This is likely due to the biases introduced during PCR sequencing, which result in differential representation of GC-rich regions. Sequences with high GC content can form secondary structures that are difficult to denature, leading to reduced amplification efficiency compared to sequences with lower GC content [39]. Our experiment involved 13 PCR cycles for WGBS and only 9

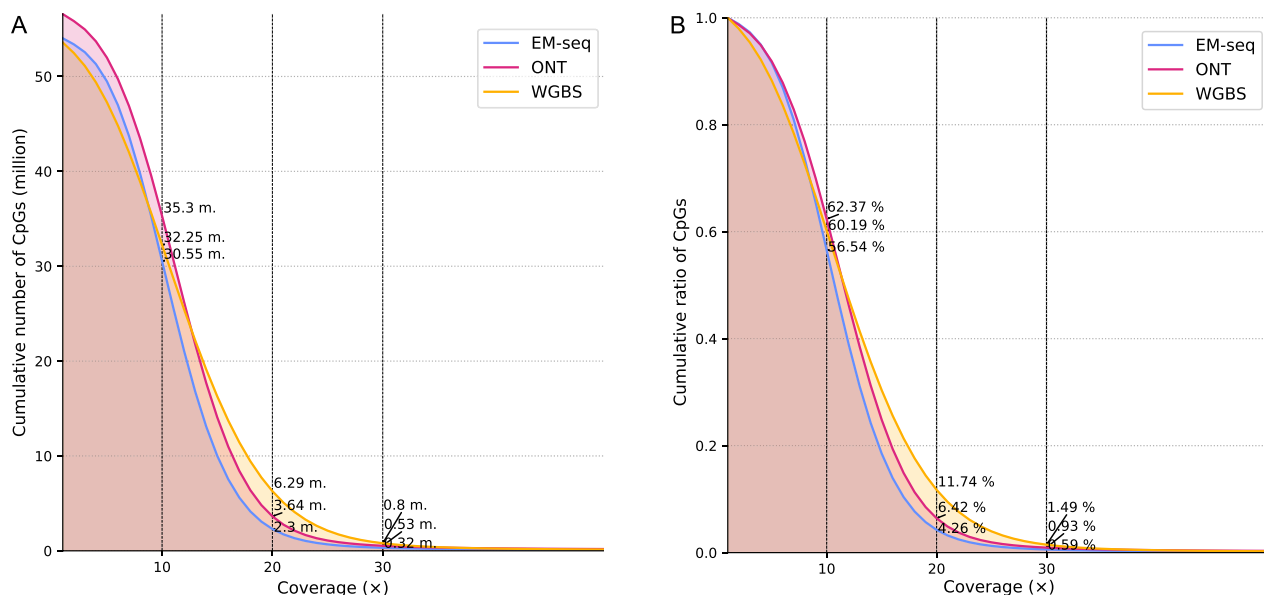


Fig. 1 Empirical cumulative distribution function (ECDF) of CpG reads calculated among the average of all samples. Overall, the cumulative number of CpGs drops with increasing coverage. **A** Analysis performed after downsampling to $29\times$ coverage. All methods show a similar trend, with WGBS exhibiting an overall higher relative coverage than ONT and EM-seq. **B** Analysis performed to show the relative coverage, taking into account the total number of CpGs for each method

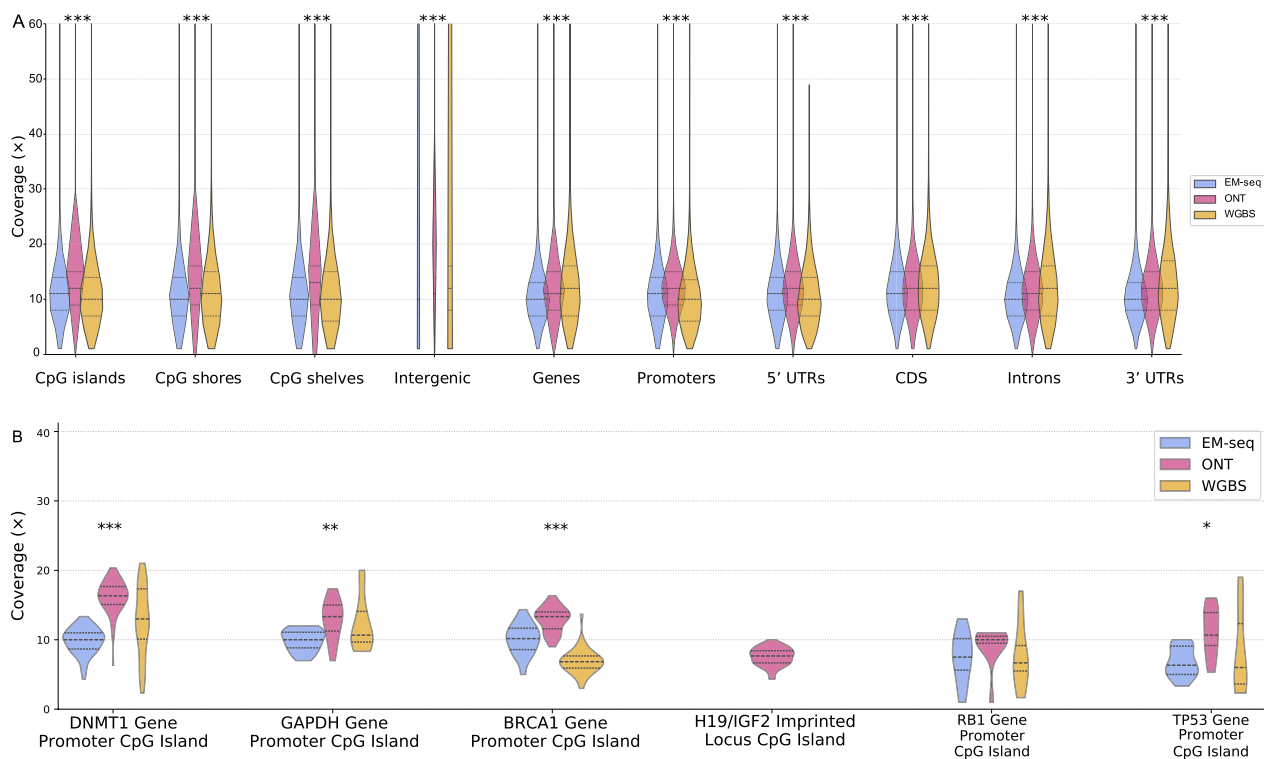


Fig. 2 Violin plot of coverage distribution across different genomic regions. **A** The coverage is variably distributed in different genomic regions, while ONT appears to cover most of the regions with higher depth than the other methods. **B** Coverage was observed in specific cancer-related genomic regions. Significance is calculated using one-way ANOVA, indicated with '*' for $p=0.01-0.05$, '**' for $p=0.001-0.01$, and '***' for $p=0-0.001$. It showcases that the assessed methods are significantly different across all considered annotation entities

PCR cycles for EM-seq. Furthermore, ONT shows higher mean coverage rates for most regions compared to EM-seq and WGBS, with consistent even coverage across all regions. More specifically, at the H19/IGF2 imprinted locus CpG island, only ONT appears suitable for cytosine detection (Fig. 2B). A correlation analysis among the general coverage data across the different methods is shown in Figure S6.

For a coverage depth greater than 10 $\times$, the total number of CpG sites covered by ONT was 51 million, WGBS covered 44.4 million, EM-seq covered 46.7 million, and EPIC microarrays covered 866 thousand (Figure S3). A total of 5.31 M sites (9.5%) were covered by ONT alone. WGBS and EM-seq detected 622 K (1.1%) and 850 K (1.5%) unique CpG sites, respectively, and even EPIC detected 34.7 K (0.1%) CpG sites that were not captured by the other three methods (Figure S3). A significant overlap is shown between the three whole-genome sequencing methods with 36.2 million common sites. The EPIC microarray results overlapped greatly with those of the other methods, with 722 K sites common to WGBS and 735 K sites common to EM-seq (Figure S3). Interestingly, EPIC has the highest overlap with ONT, with 758 K common CpGs. These results demonstrate that

while there is considerable overlap in CpG site detection among the different methods, each method also captures a unique set of CpG sites.

Notably, when intersecting the whole-genome sequencing methods only (Fig. 3), WGBS missed the highest number of sites captured by EM-seq and ONT (5.4 million). Most of these are located in intergenic regions, most likely involving long repetitive regions and tandem repeats (e.g. telomeres). More specifically, WGBS suffers from bisulfite-induced degradation of unmethylated C-rich sequences (e.g., CCCTAA telomeric repeats) [16]. EM-seq and ONT, on the other hand, missed a comparably lower number of sites, at 3.35 and 3.42 million, respectively (Fig. 3).

Among the uniquely covered sites, ONT led with 4.17 million sites, followed by EM-seq with 833 K sites and WGBS with 594 K sites (Fig. 3). Most of the sites are located in intergenic regions that can be easily spanned by long-read sequencing without the need for assembly, thereby capturing more CpGs that may be missed or fragmented by short-read sequencing methods. These results emphasize the extensive genomic coverage provided by ONT compared to EM-seq and WGBS. Although ONT covers intergenic regions extensively, it may also be

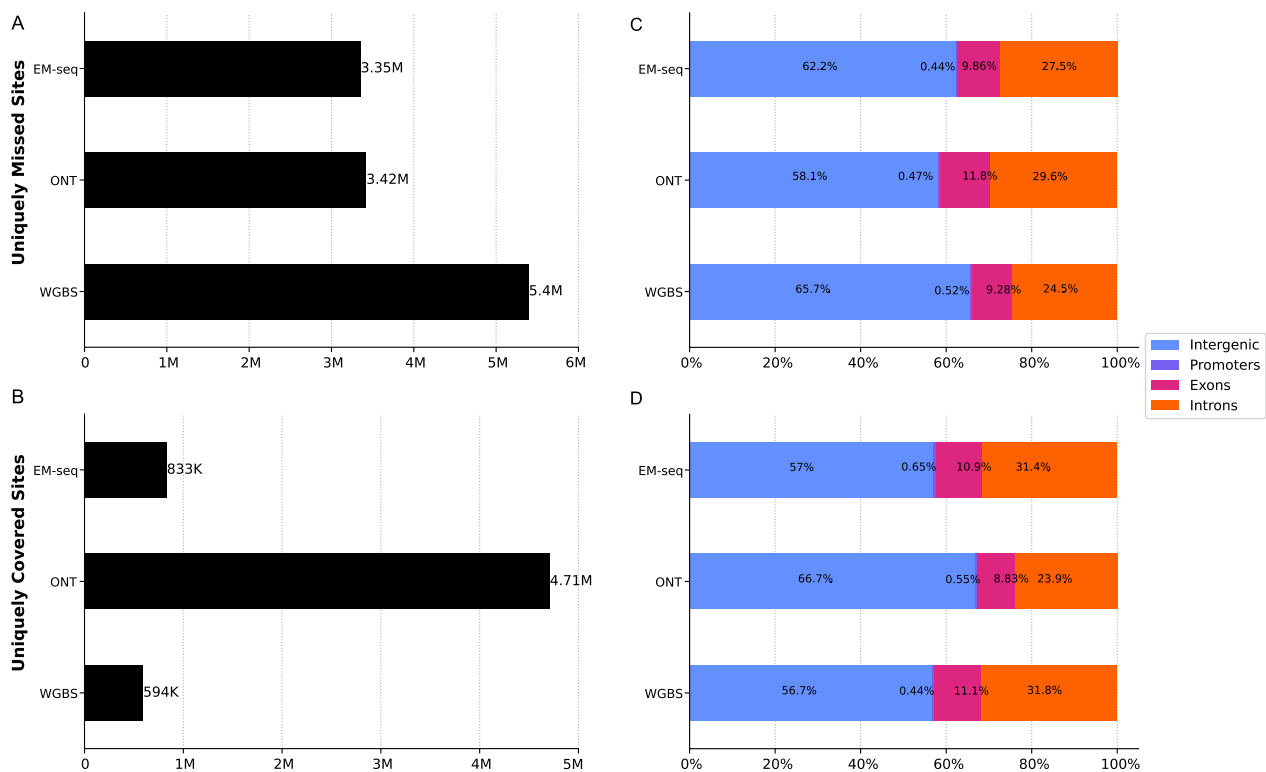


Fig. 3 Counts and genomic annotations for uniquely missed and covered sites at 10× coverage. **A** Sites that have been missed by one method but covered by the others. **B** Sites that have been covered by one method but missed by the other methods. **C** Percentage of missed sites in corresponding genomic regions. **D** Percentage of covered sites in corresponding genomic regions

biased toward regions that are easier to sequence, such as those with fewer repetitive elements or secondary structures. This could explain the higher percentage of missed sites in exons and introns compared to EM-seq and WGBS.

Methylation calling

Unlike nanopore sequencing, the bisulfite- and enzyme-based methods distinguish DNA methylation by converting unmethylated cytosines into uracil and thymine during PCR. Therefore, the effective conversion of unmethylated cytosines and the preservation of methylated cytosines determine the accuracy of the methods. WGBS and EM-seq use spike-in controls, such as unmethylated lambda phage DNA and CpG-methylated pUC19 DNA, which aid in determining the effectiveness of the conversion. For pUC19 DNA, NEB claims a CpG methylation conversion efficiency of 96–98%. Our EM-seq analysis demonstrated an efficiency of 99.46% for unmethylated lambda, whereas 97.65% CpG methylation was detected for pUC19. Both values fall within NEB's range. On the other hand, our WGBS analysis only included unmethylated lambda, with a conversion efficiency of 99.47%, similar to that of EM-seq. Typically, up

to 0.5% methylation is detected in unmethylated lambda, indicating an average conversion efficiency of 99.5%.

We assessed the performance of the methods in evaluating methylation calling across different genomic annotations (Fig. 4 and S4), using one-way ANOVA to determine the statistical significance across inter-method variability for each distinct genomic region. Generally, all methods showed higher methylation levels in intergenic regions, gene bodies, coding sequences, and introns and lower methylation levels in CpG islands and promoters, which is consistent with known methylation dynamics (Fig. 4). CpG islands, typically located at gene promoters, usually exhibit relatively low methylation levels, reflecting their active role in gene regulation. However, for the breast cancer cell line, the methylation patterns slightly shifted toward hypermethylation, as opposed to those in colorectal cancer tissue and healthy blood.

Correlation analysis of methylation status of CpG sites between methods

We conducted a correlation analysis among the methylation data to confirm the reliability of methylation calling across the different methods (Fig. 5, S6, S7 and S8). For all samples, very high correlations were observed

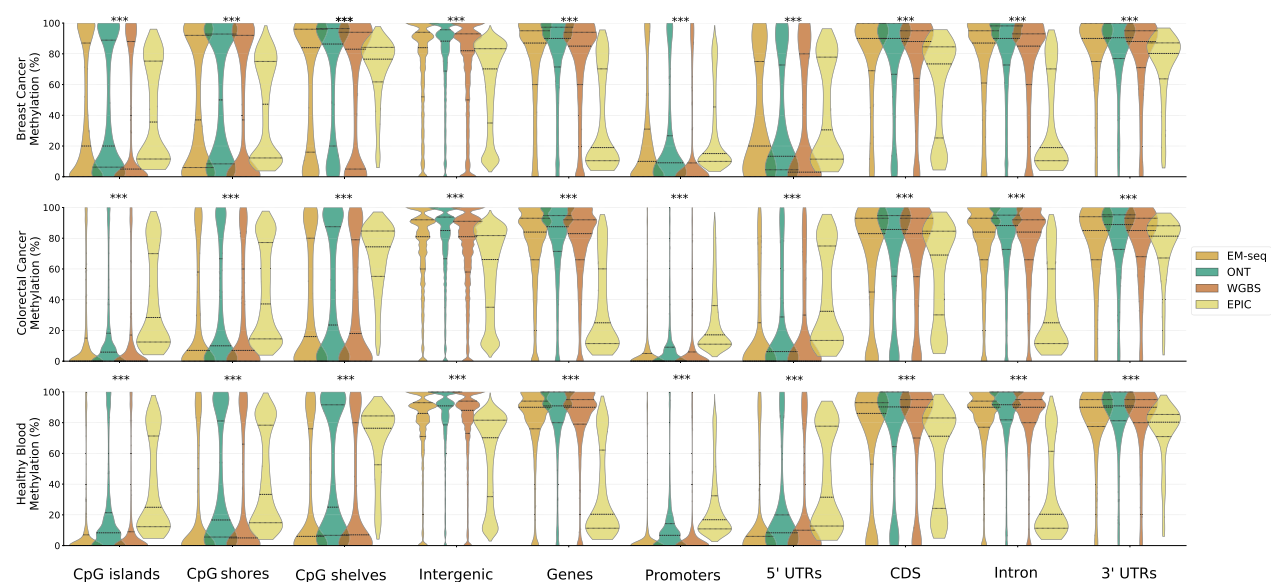


Fig. 4 Violin plot showing the methylation value distribution across different genomic regions for each method. Statistical significance is calculated using ANOVA to evaluate inter-method variability within each genomic region, indicated with '*' for $p=0.01-0.05$, '**' for $p=0.001-0.01$, and '***' for $p<0.001$ above the corresponding plot

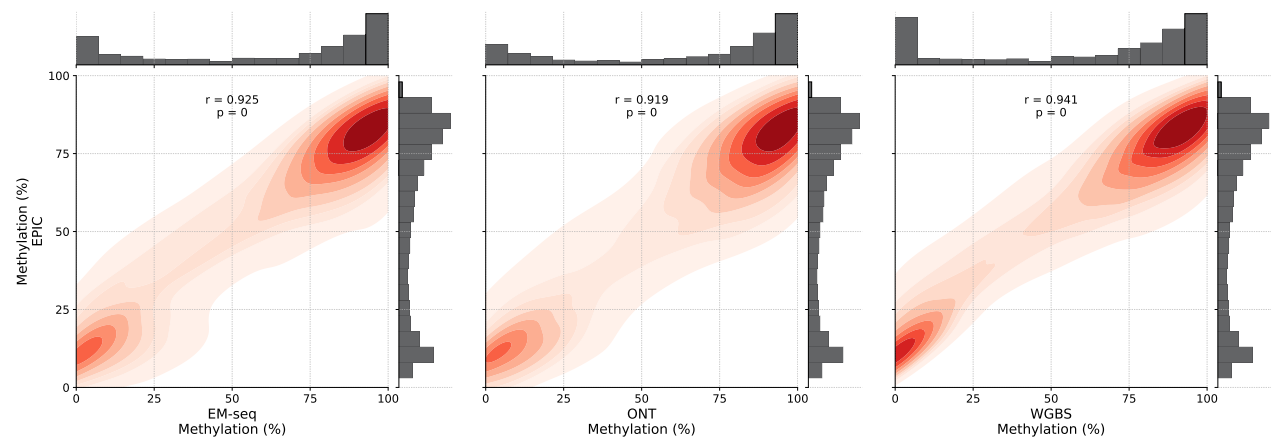


Fig. 5 Pairwise correlation in methylation levels between EPIC and other methods, illustrated with density heatmaps and Pearson correlation values for a breast cancer cell line at 10x coverage. The mean methylation values are displayed on the x- and y-axes, with corresponding distribution histograms shown above the axes. The color intensity is proportional to the correlation value

between the methylation values of EPIC and EM-seq ($r=0.925$, Breast Cancer cell line; $r=0.926$, Colorectal Cancer tissue; $r=0.935$, healthy blood), ONT ($r=0.919$, Breast Cancer cell line; $r=0.924$, Colorectal Cancer tissue; $r=0.920$, healthy blood) and WGBS ($r=0.941$, Breast Cancer cell line; $r=0.931$, Colorectal Cancer tissue; $r=0.939$, healthy blood), respectively (Fig. 5 and S7). EPIC showed a slightly weaker correlation with ONT than EM-seq and WGBS. Interestingly, although only 7,890 CpGs overlap between EPIC and EM-seq

(Figure S3), the correlation remains high and similar to that observed between EPIC and WGBS, which share 7,948,614 common CpG sites. This suggests that the Pearson correlation value does not heavily depend on the number of interrogated CpG sites but instead on the consistency of methylation detection across these sites. A few thousand sites might be sufficient to determine a correlation, eliminating the need for millions of sites for correlation analyses. Taken together, these data indicate that methylation calling with ONT and EM-seq produces results comparable to those of WGBS and

EPIC. The high correlation underscores their potential as robust methods for analyzing DNA methylation.

The density plot (Fig. 6) reveals a bimodal distribution of methylation values, with prominent peaks at the unmethylated (0%) and fully methylated (100%) states. This bimodal pattern is characteristic of DNA methylation data, reflecting regions that are heavily methylated or predominantly unmethylated. The distribution indicates a high level of agreement among EM-seq, WGBS, and ONT in identifying fully unmethylated and methylated CpGs. The EPIC microarray shows a more balanced distribution of methylation values across the scale. This is likely due to the predefined set of probes targeting CpGs, which may include a broader range of methylation levels and result in a less pronounced bimodal distribution.

Agreement between the methods

To better understand the agreement between the methods in their ability to detect hyper- and hypomethylated CpGs across the entire genome, the profiles of methylated CpGs were visualized by circos heatmaps, as shown in Figure S9. This visualization provides an overview of the methylation patterns at each chromosome. The figure illustrates the concordance between the methods over the entire genome. Specifically, there was a consistent overlap in the identification of hyper- and hypomethylated

sites. The agreement in the methylation patterns suggests that when a CpG site is detected as methylated (either hyper- or hypomethylated) by one method, it is generally confirmed by other methods. Conversely, CpGs not detected by one method are typically also undetected by the others, indicating a high level of consistency in the methods' performance.

Furthermore, to determine the reproducibility of DNA methylation values, the inter-method variability was assessed using Fleiss' kappa (inter-rater reliability). This metric reflects the degree of agreement between methods, accounting for agreement that would occur by chance. The kappa value was interpreted according to Landis and Koch [40]. Overall, the highest agreement was observed between EM-seq and WGBS (Fig. 7), indicating a strong concordance between methods that use the same sequencing chemistry. Comparisons involving ONT, WGBS and EM-seq showed the second-highest agreement. ONT and EPIC, on the other hand, show the lowest agreement in all samples yet still exhibit a good level of concordance despite employing completely different techniques. All method comparisons fall within the range of 0.65–0.78, indicating a consistent pattern of high concordance among the methods. Moreover, DNA methylation profiles across gene bodies (± 10 kb from TSS to TTS) showed similar trends for all platforms for

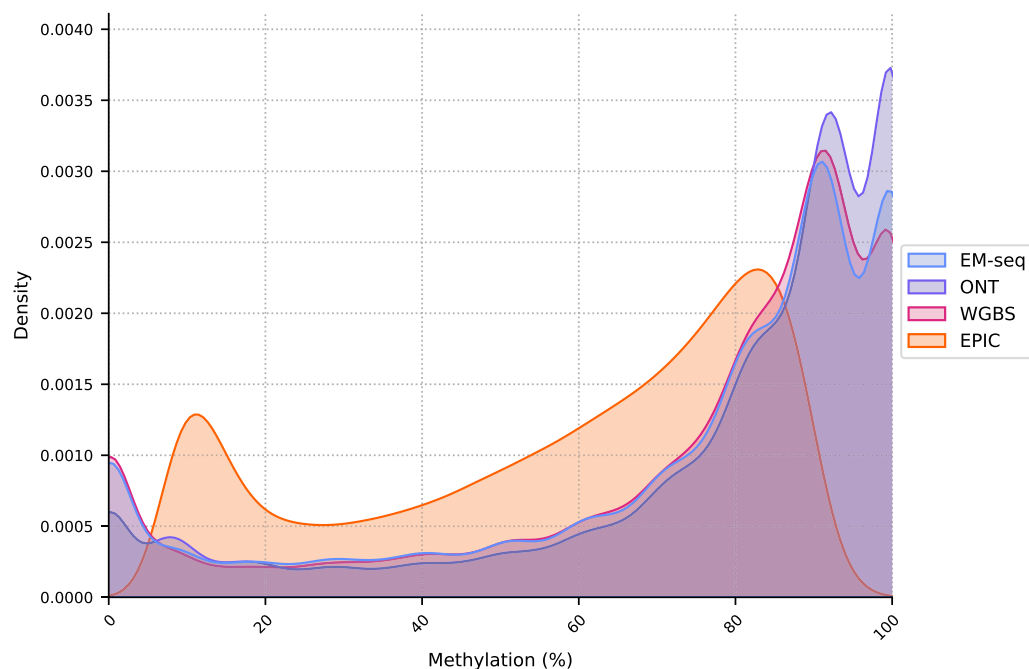


Fig. 6 Density plot outlining the distribution of methylation values across all methods using the average counts of all samples at 10× coverage. All methods exhibit the characteristic bimodal methylation distribution, with ONT, EM-seq and WGBS having similar densities. The scale on the x-axis ranges from 0 (no methylation) to 1.0 (100% methylation). EM-seq, WGBS and ONT show a predominance of fully unmethylated and fully methylated cytosines compared to EPIC, where the distribution of methylation values is more balanced along the scale

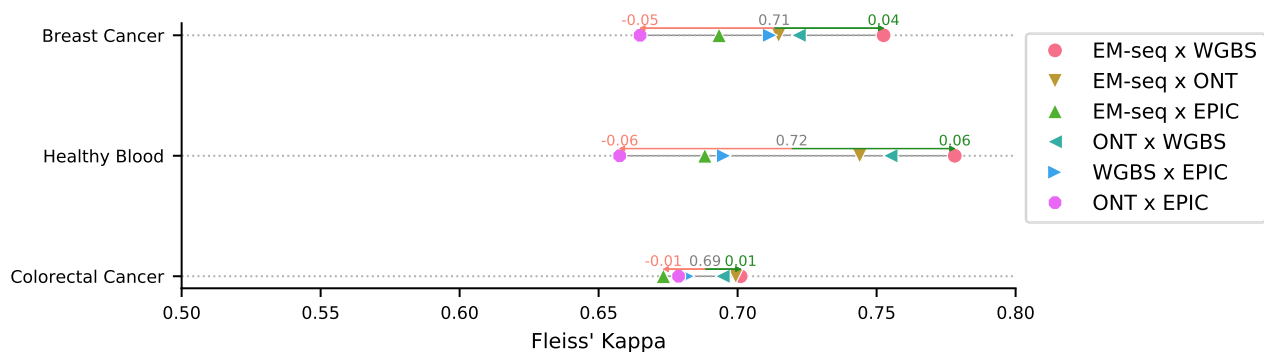


Fig. 7 Fleiss' kappa plot showing inter-method agreement analysis. A kappa between 0.01 and 0.20 represents "slight" agreement, a value between 0.21 and 0.40 represents "fair" agreement, a value between 0.41 and 0.60 represents "moderate" agreement, a value between 0.61 and 0.80 represents "substantial" agreement, and a value greater than 0.81 represents "nearly perfect" agreement. All comparisons showed a substantial agreement, with EM-seq and WGBS agreeing the most in paired comparisons, and ONT added to them for the highest agreement in tripled comparisons

all samples (Figure S10) suggesting cross-platform concordance in capturing methylation dynamics. This supports the technical comparability of these methods for genome-wide methylation profiling, despite their differing chemistries and resolution.

Discussion

DNA methylation (5mC) is a fundamental epigenetic modification with regulatory effects on gene expression [2]. A variety of methods for detecting 5mC exist, hinging primarily on the type and quality of the sample, as well as the specific research objectives. WGBS and EPIC have a long tradition and excellent reputation for accuracy in methylation calling [41]. However, emerging methods such as EM-seq and ONT sequencing show great potential to complement or replace established methods [19, 42]. While EPIC methylation arrays may have several advantages for differential methylation analysis in cohort studies, such as high reproducibility across technical and biological replicates [43], NGS-based technologies can detect additional methylated sites not covered by the array and provide high-coverage information for specific genes of interest. Hence, the use of EPIC for very deep and comprehensive characterization of the methylome is limited, underscoring the need for NGS methods to supplement or replace targeted EPIC data. We evaluated WGBS, EM-seq, ONT and EPIC to provide a comparative analysis and assess the performance of these methods. Our approach profiled tumor tissue, cell line and a blood sample, in contrast to other studies where only cell lines were employed [44, 45].

As reported in the literature, WGBS is considered the leading method for whole epigenome sequencing due to its numerous advantages, including its ability to map DNA methylation at single-base resolution and

offering comprehensive coverage of CpG sites across the genome, theoretically encompassing 28 million CpG sites per strand [16, 46]. However, WGBS is no longer the only method for single-base resolution DNA methylation profiling. EM-seq offers an alternative and requires fewer PCR cycles for amplification, resulting in fewer duplicates. Moreover, EM-seq requires less input DNA, and our results show that it yields uniform coverage, which was also reported by Han et al. [37]. In addition to conversion-based methods, ONT offers the indisputable advantage of natively reading the cytosine methylation status without any external changes to the DNA through a direct sequencing approach. Moreover, ONT's long-read capability is advantageous for covering regions that are traditionally challenging for short-read methods, such as repetitive regions and regions with secondary structures, resulting in an overall higher number of CpGs detection (~56 M compared to ~54 M for WGBS and EM-seq, described earlier in Table 1), even with a lower sequencing yield. On the other hand, arrays will never be as comprehensive as sequencing-based methods in interrogating the whole genome, as shown in Figure S3. However, they still offer an accessible and representative method for profiling large sample cohorts. EPIC arrays (and more recently EPIC V2) represent a significant improvement in genomic coverage compared with the previous HM450 array, particularly in enhancers regions [41, 47]. This platform has also shown high inter-array accuracy and single-locus reproducibility across technical and biological replicates (Fig. 7). A persistent limitation is their limited interrogation of distal regulatory elements and that the methylation level of one CpG probe per element is not always reflective of adjacent sites [41]. A limitation of the study involves a potential biological source of variation across the different techniques, as the

WGBS and EM-seq datasets contain a mixture of 5mC and 5hmC signals, whereas ONT exclusively reflects 5mC. This discrepancy may confound the interpretation of technical biases between methods. While this is a recognized limitation, the impact is likely minimal in our case, as 5hmC is most abundantly found in neural tissues, which are not included in this study. Furthermore, selectively quantifying 5mC alone from WGBS or EM-seq data would require additional experimental steps, such as oxidative bisulfite or specialized enzymatic treatments [19, 48], which are more labor-intensive, costly, and beyond the scope of this study. Nonetheless, the potential contribution of 5hmC to the WGBS and EM-seq signals should be considered when interpreting cross-platform comparisons.

To ensure a fair and unbiased comparison between the sequencing methods, downsampling was performed to reach a common coverage of $29\times$ (Fig. 1). This step was necessary to create an equal baseline across the three methods and ensure that the observed differences in methylation profiling were due to inherent methodological differences rather than variations in sequencing depth or data volume. ONT sequencing typically produces long reads, albeit at a lower depth, whereas WGBS and EM-seq are optimized for high-depth, short-read sequencing. If one dataset has significantly higher coverage than another, it will inherently detect more methylation sites. However, in the context of methylation value comparisons, differences in coverage can bias the results, as shown by Guanzon et al. [49]. A method with higher sequencing depth may estimate the methylation fraction more precisely, whereas a lower-depth method could introduce increased variability [49, 50]. By equalizing coverage, we mitigate the risk of inflating the performance of a technique that generates more data. Hence, the comparative analysis remains robust, allowing for an accurate assessment of how each method quantifies DNA methylation levels.

Our results show a high correlation among all methods (Fig. 5, S6, S7 and S8), with Pearson correlation values ($r > 0.91$). This strong concordance suggests robust agreement in methylation calling despite the different underlying chemistries and sequencing approaches. Notably, ONT sequencing showed a slightly lower correlation than EM-seq and WGBS did, likely due to differences in its detection chemistry and basecalling software. As observed in other studies, EM-seq and WGBS exhibited the highest concordance, reflecting their shared sequencing technology [19, 44]. Additionally, EPIC readouts were strongly correlated with WGBS, EM-seq and ONT, which is in line with other studies [41, 44]. We also observed that the average EPIC-WGBS correlation ($r = 0.937$) was consistently

higher than the EPIC-EM-seq correlation ($r = 0.929$) (Fig. 5 and S7), unlike observations in other studies [44, 49]. Although EPIC and WGBS use bisulfite-converted DNA, differences in assay chemistry and mathematical approaches for methylation quantification remain. WGBS quantifies the level of methylation based on read counts, whereas EPIC measures fluorescence intensity. We showed that there was no significant technical variability between WGBS, EM-seq and ONT at a sequencing depth of $29\times$ (Fig. 5). In this context, the chosen coverage might involve sufficient reads for reliable quantification. While some researchers suggest that a coverage of approximately $10\times$ or greater per sample and CpG is adequate for reliable methylation detection [51], other studies highlight critical limitations [16, 49, 52, 53]. Specifically, filtering low-coverage CpGs can disproportionately remove poorly amplified regions (e.g., CG-rich areas), thereby introducing biases and compromising the accuracy of global methylation estimation, particularly in WGBS workflows [54]. Importantly, the tradeoff in coverage filtering should be made depending on the study objective: for instance, for the identification of differentially methylated regions (DMRs), low-coverage filtering ($1\text{--}2\times$) may be suitable when identifying DMRs with large methylation differences. As such, resources can be invested in increasing the number of biological replicates to increase statistical power [52]. High-coverage filtering may be more suitable when focusing on specific CpG sites. Indeed, the precision of each sequencing technology improves with increasing sequencing depth, making each analysis considerably costly. We also noted that ONT data are shifted toward hypermethylation compared with WGBS and EM-seq data (Fig. 6). Most ONT tools (e.g. Guppy) correctly identify unmethylated sites but fail to accurately identify fully methylated sites, as demonstrated by Yuen et al. [50]. This is because the electrical current changes caused by a methylated cytosine can be subtle and complex to distinguish from background noise, resulting in an overrepresentation of methylated sites. Improvements in basecalling algorithms may help mitigate this issue in the future [50]. On the other hand, WGBS also tends to overrepresent methylated reads, which is consistent with previous work that observed a greater recovery of fully methylated reads compared to fully unmethylated ones after bisulfite treatment [16]. This discrepancy can be attributed to an artefact of biased coverage, which favors retention of highly methylated CpGs over their neighboring regions. Notably, this bias is more common in WGBS than in EM-seq because the bisulfite treatment causes degradation and selective loss of unmethylated sequences, whereas EM-seq largely avoids this effect by employing mild

enzymatic steps. Additionally, an incomplete conversion of unmethylated cytosines may also contribute to the overrepresentation of methylated cytosines [18].

The original WGBS protocol requires high amounts of high-quality DNA, which is not ideal for fragmented or low-input samples such as cell-free DNA (cfDNA). However, several variants, including Accel-NGS Methyl-Seq and SPLinted Ligation Adapter Tagging (SPLAT), have been developed to reduce the required DNA input to as little as 1–100 ng [55–59]. These protocols are optimized for use with fragmented and low-input samples and achieve quite even coverages, while acknowledging a coverage decline in CpG islands and in very GC rich promoter regions, remaining a common characteristic of WGBS data [59]. With the classical pre-bisulfite adapter ligation protocol, we also observed that WGBS has a lower mean coverage in GC-rich regions compared to EM-seq and ONT (Fig. 2A), supporting previous research [19, 60]. In contrast, EM-seq generally offers consistent yet slightly lower coverage than ONT and WGBS. However, we opted to investigate GC-rich regions (*TP53*, *DNMT1*, *BRCA1*, *RBI*, and *GAPDH* gene promoters, and the *H19/IGF2* imprinted locus), suggesting that the methylation status of these regions is not well assessed by WGBS and EM-seq (Fig. 2B). Furthermore, lower library complexity and, thus, a greater proportion of duplicate reads are observed in our samples with WGBS (9.5%) than EM-seq (7.0%) [55, 56, 61]. In contrast, EM-seq has been successfully applied to <1 ng of DNA but also at the cost of a higher number of PCR cycles, resulting in a greater number of duplicates and the need for deeper sequencing to obtain the same depth of data [19]. Thus, the recommended starting input is at least >10 ng. Due to the absence of an amplification step, ONT sequencing often requires greater amounts of input DNA (1–5 µg), making it more suitable for samples with abundant starting material [51]. ONT, however, offers considerable advantages over WGBS and EM-seq, most significantly the ability to determine the methylation status without needing any treatment. Furthermore, nanopore sequencing enables rapid methylation profiling of cfDNA, albeit at low coverages [62]. However, in a research or clinical context, it still lags. Long-read sequencing is significantly more expensive than WGBS and EM-seq. However, it has proven to benefit fragmentomic research or specific clinical scenarios, e.g. neurosurgical decision-making [63, 64].

The turnaround time (TAT) is a crucial factor to consider when selecting a particular method, especially in clinical settings. The TAT for EPIC is significantly shorter than that for NGS-based methods, making it highly suitable for situations where quick results are needed. Typically, the entire EPIC process can be completed within 3 to 4 days. In contrast, WGBS and EM-seq are more

time-consuming procedures, often taking up to 1.5 weeks to complete the entire workflow and obtain the results. Due to its ability to directly sequence native DNA, ONT sequencing can offer faster TAT (60–90 min) in terms of library preparations and obtain very limited sequencing data (100–400 Mb) [63, 65]. However, the sequencing run when aiming for high-coverage data can take several days and, depending on the selected data analysis workflow, requires significant computing time, potentially extending the TAT.

Computationally, the different methods also vary in hardware requirements and computational time (Table 1). On the simpler end of the spectrum is processing EPIC methylation data, which can be run on a standard modern computer relatively quickly when processing a few samples [66]. Next are EM-seq and WGBS, which require running on a computing cluster but now have standardized software and workflows [53]. The most challenging computational method is ONT. First, the large data size makes it cumbersome to wrangle and requires much more memory to store, as well as resources and time to process. Second, its base callers require GPUs [67, 68], which are not readily available in every research laboratory or clinic. Nanopore also includes a compute tower in their sequencers, which performs basecalling in real-time. Moreover, its software landscape is still relatively young and rapidly evolving, making it challenging to achieve repeatable results consistently [69]. Despite the varying computational requirements, the emerging widespread adoption of cloud services makes accessing the proper hardware for processing data from any of the discussed platforms easier.

Each method offers unique advantages and has limitations for DNA methylation profiling. The choice of method primarily depends on specific experimental conditions and research objectives, including DNA input limitations, interest in particular genomic regions, sample type considerations, and budget limitations. The overall costs of the three genome-wide sequencing methods are currently comparable in practice. Moreover, the ongoing decline and variability in sequencing prices make cost-effectiveness analyses difficult to generalize. As such, detailed cost comparisons were not included in this manuscript, as they may become outdated soon. Instead, the focus remains on selecting the appropriate method based on experimental needs, available infrastructure, and technical expertise rather than relying solely on cost considerations. Considering the current “traditional method”, WGBS remains a widely used platform that combines robustness and practicality. In contrast, EM-seq obviates the need for harsh bisulfite treatment, preserving DNA integrity and requiring less DNA input. ONT sequencing offers real-time sequencing

and portability, making it ideal for field-based or point-of-care applications. Its ability to produce long reads with quick turnaround times facilitates rapid detection of methylation. Despite a drop in sequencing costs, the EPIC platform retains several key features, including ease of analysis and cost-effectiveness, ensuring its continued relevance in research.

Accurate methylation profiling could be improved by integrating different methodologies. For example, combining EM-seq with ONT sequencing (nanoEM) [70] could leverage the strengths of both methods, providing native methylation data with reduced biases. Another example of the integration of different approaches is FRAGmentomics-based methylation analysis (FRAGMA). This technique is based on the distinct cleavage patterns of methylated and unmethylated CpGs resulting from differential nuclease activity without the need for bisulfite or enzymatic treatment. By analyzing these fragmentation signatures, FRAGMA enables the assessment of methylation at specific genomic regions [71]. Combining FRAGMA with existing sequencing technologies could increase the accuracy of methylation analysis. Recent advancements in ONT sequencing, known as adaptive sequencing, have opened new possibilities for targeted enrichment of specific loci. This technology enables the real-time selection of DNA fragments, providing increased coverage of regions of interest and thereby improving the accuracy of the methylation read-out [72]. Without the need for additional sample preparation, adaptive sequencing offers a more affordable approach to analyzing target regions intensively [73, 74]. Additionally, in a clinical context, NGS of cfDNA opens new avenues for early cancer detection and other diagnostic applications. cfMeDIP-seq [75] has been developed to avoid the use of bisulfite treatment, yet room for improvement remains due to limitations inherent to its affinity-based nature. NGS is rapidly becoming a more affordable option with increased feasibility in genome-wide methylation profiling and will inevitably become the standard technology upon which all global epigenetic profiling will be based.

Conclusion

All evaluated methods pose opportunities and challenges for assessing DNA methylation. Each technique has merit, and selecting the most appropriate technology depends on the research question or intended application.

For researchers with limited DNA input, EM-seq is recommended because of its ability to handle low-input and fragmented samples with considerable cost efficiency. WGBS, while comprehensive, is less suitable for low-input samples due to higher input requirements. EPIC is

ideal for studies requiring standardized and cost-effective analysis but is limited by its predefined probe set. ONT is particularly interesting in resolving complex regions that are not covered by the other two techniques but at a considerably higher cost. Additionally, complex computational processes need to be considered.

Given the cost and accuracy considerations, researchers should weigh their choice based on the trade-offs between comprehensive coverage, resolution, and budget constraints. As NGS technologies continue to evolve, integrating different methodologies or optimizing current protocols will improve the accuracy and cost-effectiveness of DNA methylation profiling. Future advancements, such as the development of hybrid methods combining the strengths of different approaches, hold promise for further improving the field of epigenetic research.

Abbreviations

AFU	Arbitrary fluorescence unit
BAM	Binary alignment map
BED	Browser extensible data
CDs	Coding sequences
CpG	Cytosine-phosphate-Guanine
EM-seq	Enzymatic methyl-seq
GPU	Graphics processing unit
NGS	Next-generation sequencing
NCS	Next-generation sequencing control software
OCT	Optimal cutting temperature
ONT	Oxford Nanopore Technologies
QC	Quality control
SNPs	Single nucleotide polymorphisms
TAT	Turnaround time
UTR	Untranslated region
WGBS	Whole-genome bisulfite sequencing

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13072-025-00616-3>.

Supplementary material 1.

Acknowledgements

Not applicable.

Author contributions

ARdA and JI contributed to the conception and design of the work and writing the initial and final draft. JI, VL and LM contributed to the data analysis and revising the final draft. GVC and KODB contributed to the conception of the draft and revising the final draft. All authors have read and approved the published version of the manuscript.

Funding

Ana Regina de Abreu is supported by a strategic basic PhD fellowship of the Research Foundation Flanders (Belgium) (FWO; 1SD3722N). Vasileios Lemonidis is supported by a scholarship provided by Stichting Tegen Kanker (Belgium) (C/2022/2056). Research performed by Dr. Ligia Mateiu is supported by the Methusalem-OEC grant 40790. Research performed by Prof. Guy Van Camp and Prof. Ken Op de Beeck is supported by grants awarded by the University of Antwerp (BOF/Methusalem grant 40790, BOF/TOP 39705).

Data availability

The datasets generated and analysed during this study are available on request from the European Genome-Phenome Archive (EGAS00001008014).

Declarations**Ethics approval and consent to participate**

This study was approved by the Clinical Ethics Committee of the Antwerp University Hospital.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Center of Medical Genetics, University of Antwerp and Antwerp University Hospital, Prins Boudewijnlaan 43, 2650 Edegem, Belgium. ²Integrated Personalized and Precision Oncology Network (IPPON), Center for Oncological Research (CORE), University of Antwerp and Antwerp University Hospital, Universiteitsplein 1, 2610 Wilrijk, Belgium.

Received: 13 March 2025 Accepted: 29 July 2025

Published online: 25 August 2025

References

- Lister R, Pelizzola M, Downen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009;462:315–22. <https://doi.org/10.1038/nature08514>.
- He X-J, Chen T, Zhu J-K. Regulation and function of DNA methylation in plants and animals. *Cell Res*. 2011;21:442–65. <https://doi.org/10.1038/cr.2011.23>.
- Plongthongkum N, Diep DH, Zhang K. Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nat Rev Genet*. 2014;15:647–61. <https://doi.org/10.1038/nrg3772>.
- Jang HS, Shin WJ, Lee JE, Do JT. CpG and non-CpG methylation in epigenetic gene regulation and brain function. *Genes*. 2017. <https://doi.org/10.3390/genes8060148>.
- Guo JU, Su Y, Shin JH, et al. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci*. 2014;17:215–22. <https://doi.org/10.1038/nn.3607>.
- Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*. 2012;13:484–92. <https://doi.org/10.1038/nrg3230>.
- Deaton AM, Webb S, Kerr AR, et al. Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res*. 2011;21:1074–86. <https://doi.org/10.1101/gr.118703.110>.
- Yang X, Han H, De Carvalho DD, et al. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell*. 2014;26:577–90. <https://doi.org/10.1016/j.ccr.2014.07.028>.
- Wang Q, Xiong F, Wu G, et al. Gene body methylation in cancer: molecular mechanisms and clinical applications. *Clin Epigenetics*. 2022;14:154. <https://doi.org/10.1186/s13148-022-01382-9>.
- Heyn H, Vidal E, Ferreira HJ, et al. Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. *Genome Biol*. 2016;17:11. <https://doi.org/10.1186/s13059-016-0879-2>.
- Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics*. 2016;8:389–99. <https://doi.org/10.2217/epi.15.114>.
- Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74. <https://doi.org/10.1038/nature11247>.
- Moore JE, Purcaro MJ, Pratt HE, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. 2020;583:699–710. <https://doi.org/10.1038/s41586-020-2493-4>.
- Lizio M, Harshbarger J, Shimoji H, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol*. 2015;16:22. <https://doi.org/10.1186/s13059-014-0560-6>.
- Noguera-Castells A, García-Prieto CA, Álvarez-Errico D, Esteller M. Validation of the new EPIC DNA methylation microarray (900k EPIC v2) for high-throughput profiling of the human DNA methylome. *Epigenetics*. 2023;18:2185742. <https://doi.org/10.1080/15592294.2023.2185742>.
- Olova N, Krueger F, Andrews S, et al. Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biol*. 2018;19:33. <https://doi.org/10.1186/s13059-018-1408-2>.
- Frommer M, McDonald LE, Millar DS, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A*. 1992;89:1827–31. <https://doi.org/10.1073/pnas.89.5.1827>.
- Dai Q, Ye C, Irklyenko I, et al. Ultrafast bisulfite sequencing detection of 5-methylcytosine in DNA and RNA. *Nat Biotechnol*. 2024;42:1559–70. <https://doi.org/10.1038/s41587-023-02034-w>.
- Vaisvila R, Ponnaluri VKC, Sun Z, et al. Enzymatic methyl sequencing detects DNA methylation at single-base resolution with picograms of DNA. *Genome Res*. 2021;31:1280–9. <https://doi.org/10.1101/gr.266551.120>.
- Laszlo AH, Derrington IM, Brinkerhoff H, et al. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc Natl Acad Sci U S A*. 2013;110:18904–9. <https://doi.org/10.1073/pnas.1310240110>.
- Schreiber J, Wescoe ZL, Abu-Shumays R, et al. Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. *Proc Natl Acad Sci U S A*. 2013;110:18910–5. <https://doi.org/10.1073/pnas.1310615110>.
- De Roeck A, De Coster W, Bossaerts L, et al. Nanosatellite: accurate characterization of expanded tandem repeat length and sequence through whole genome long-read sequencing on PromethION. *Genome Biol*. 2019;20:239. <https://doi.org/10.1186/s13059-019-1856-3>.
- Gouil Q, Keniry A. Latest techniques to study DNA methylation. *Essays Biochem*. 2019;63:639–48. <https://doi.org/10.1042/ebc20190027>.
- Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 1988;16:1215. <https://doi.org/10.1093/nar/16.3.1215>.
- Fortin J-P, Triche TJ Jr, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics*. 2016;33:558–60. <https://doi.org/10.1093/bioinformatics/btw691>.
- Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013;29:189–96. <https://doi.org/10.1093/bioinformatics/bts680>.
- Tian Y, Morris TJ, Webster AP, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics*. 2017;33:3982–4. <https://doi.org/10.1093/bioinformatics/btx513>.
- Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*. 2023;2: e107. <https://doi.org/10.1002/imt2.107>.
- Faust GG, Hall IM. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics*. 2014;30:2503–5. <https://doi.org/10.1093/bioinformatics/btu314>.
- Kolmogorov M, Billingsley KJ, Mastorosa M, et al. Scalable nanopore sequencing of human genomes provides a comprehensive view of haplotype-resolved variation and methylation. *Nat Methods*. 2023;20:1483–92. <https://doi.org/10.1038/s41592-023-01993-x>.
- The-pandas-development-team. pandas-dev/pandas:Pandas (v2.2.3). Zenodo, 2024. 10.5281/zenodo.13819579
- Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–30.
- Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17:261–72. <https://doi.org/10.1038/s41592-019-0686-2>.
- Hunter JD. Matplotlib: A 2D graphics environment. *Comput Sci Eng*. 2007;9:90–5. <https://doi.org/10.1109/MCSE.2007.55>.
- Waskom LM. Seaborn: statistical data visualization. *J Open Source Softw*. 2021;6:3021. <https://doi.org/10.21105/joss.03021>.

36. Flyamer I, Xue Z, Colin et al. Phyla/adjustText: 1.3.0 Zenodo, 2024
37. Han Y, Zheleznyakova GY, Marincevic-Zuniga Y, et al. Comparison of EM-seq and PBAT methylome library methods for low-input DNA. *Epigenetics*. 2022;17:1195–204. <https://doi.org/10.1080/15592294.2021.1997406>.
38. Youk J, An Y, Park S, et al. The genome-wide landscape of C:G > T:A polymorphism at the CpG contexts in the human population. *BMC Genom*. 2020;21:270. <https://doi.org/10.1186/s12864-020-6674-1>.
39. Kalle E, Kubista M, Rensing C. Multi-template polymerase chain reaction. *Biomol Detect Quantif*. 2014;2:11–29. <https://doi.org/10.1016/j.bdq.2014.11.002>.
40. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–74.
41. Pidsley R, Zotenko E, Peters TJ, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol*. 2016;17:208. <https://doi.org/10.1186/s13059-016-1066-1>.
42. Katsman E, Orlanski S, Martignano F, et al. Detecting cell-of-origin and cancer-specific methylation features of cell-free DNA from Nanopore sequencing. *Genome Biol*. 2022;23:158. <https://doi.org/10.1186/s13059-022-02710-1>.
43. Khodasevich D, Smith AR, Huen K, et al. Comparison of DNA methylation measurements from EPIC beadchip and seqcap targeted bisulphite sequencing in PON1 and nine additional candidate genes. *Epigenetics*. 2022;17:1944–55. <https://doi.org/10.1080/15592294.2022.2091818>.
44. Foox J, Nordlund J, Lalancette C, et al. The SEQC2 epigenomics quality control (EpiQC) study. *Genome Biol*. 2021;22:332. <https://doi.org/10.1186/s13059-021-02529-2>.
45. Davenport CF, Scheithauer T, Dunst A, et al. Genome-wide methylation mapping using nanopore sequencing technology identifies novel tumor suppressor genes in hepatocellular carcinoma. *Int J Mol Sci*. 2021;22:3937.
46. Loyer N, Magenheimer J, Peretz A, et al. A DNA methylation atlas of normal human cell types. *Nature*. 2023;613:355–64. <https://doi.org/10.1038/s41586-022-05580-6>.
47. Peters TJ, Meyer B, Ryan L, et al. Characterisation and reproducibility of the HumanMethylationEPIC v2.0 BeadChip for DNA methylation profiling. *BMC Genomics*. 2024;25(1):251. <https://doi.org/10.1186/s12864-024-10027-5>.
48. Booth MJ, Branco MR, Ficiz G, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Sci*. 2012;336:934–7. <https://doi.org/10.1126/science.1220671>.
49. Guanzon D, Ross JP, Ma C, et al. Comparing methylation levels assayed in GC-rich regions with current and emerging methods. *BMC Genomics*. 2024;25:741. <https://doi.org/10.1186/s12864-024-10605-7>.
50. Yuen ZW, Srivastava A, Daniel R, et al. Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. *Nat Commun*. 2021;12:3438. <https://doi.org/10.1038/s41467-021-23778-6>.
51. Sigurpalsdottir BD, Stefansson OA, Holley G, et al. A comparison of methods for detecting DNA methylation from long-read sequencing of human genomes. *Genome Biol*. 2024;25:69. <https://doi.org/10.1186/s13059-024-03207-9>.
52. Ziller MJ, Hansen KD, Meissner A, Aryee MJ. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nat Methods*. 2015;12(3):230–2. <https://doi.org/10.1038/nmeth.3152>.
53. Bock C. Analysing and interpreting DNA methylation data. *Nat Rev Genet*. 2012;13:705–19. <https://doi.org/10.1038/nrg3273>.
54. Stuart T, Buckberry S, Nguyen TV, Lister R. Approaches for the analysis and interpretation of whole-genome bisulfite sequencing data. *Methods Mol Biol*. 2024;2842:391–403. https://doi.org/10.1007/978-1-0716-4051-7_20.
55. Krepelova A, Neri F. Low-input whole-genome bisulfite sequencing. *Methods Mol Biol*. 2021;2351:353–68. https://doi.org/10.1007/978-1-0716-1597-3_20.
56. Zhou L, Ng HK, Drautz-Moses DI, et al. Systematic evaluation of library preparation methods and sequencing platforms for high-throughput whole genome bisulfite sequencing. *Sci Rep*. 2019;9:10383. <https://doi.org/10.1038/s41598-019-46875-5>.
57. Miura F, Enomoto Y, Dairiki R, Ito T. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res*. 2012;40: e136. <https://doi.org/10.1093/nar/gks454>.
58. Wang Q, Gu L, Adey A, et al. Tagmentation-based whole-genome bisulfite sequencing. *Nat Protoc*. 2013;8:2022–32. <https://doi.org/10.1038/nprot.2013.118>.
59. Raine A, Manlig E, Wahlberg P, et al. SPLINTED ligation adapter tagging (SPLAT), a novel library preparation method for whole genome bisulphite sequencing. *Nucleic Acids Res*. 2016;45:e36–e36. <https://doi.org/10.1093/nar/gkw1110>.
60. Feng S, Zhong Z, Wang M, Jacobsen SE. Efficient and accurate determination of genome-wide DNA methylation patterns in *Arabidopsis thaliana* with enzymatic methyl sequencing. *Epigenetics Chromatin*. 2020;13:42. <https://doi.org/10.1186/s13072-020-00361-9>.
61. Morrison J, Koeman JM, Johnson BK, et al. Evaluation of whole-genome DNA methylation sequencing library preparation protocols. *Epigenetics Chromatin*. 2021;14:28. <https://doi.org/10.1186/s13072-021-00401-y>.
62. Afflerbach AK, Albers A, Appelt A, et al. Nanopore sequencing from formalin-fixed paraffin-embedded specimens for copy-number profiling and methylation-based CNS tumor classification. *Acta Neuropathol*. 2024;147:74. <https://doi.org/10.1007/s00401-024-02731-z>.
63. Vermeulen C, Pagès-Gallego M, Kester L, et al. Ultra-fast deep-learned CNS tumour classification during surgery. *Nature*. 2023;622:842–9. <https://doi.org/10.1038/s41586-023-06615-2>.
64. van der Pol Y, Tanyo NA, Evander N, et al. Real-time analysis of the cancer genome and fragmentome from plasma and urine cell-free DNA using nanopore sequencing. *EMBO Mol Med*. 2023;15: e17282. <https://doi.org/10.15252/emmm.202217282>.
65. Oehler JB, Wright H, Stark Z, et al. The application of long-read sequencing in clinical settings. *Hum Genom*. 2023;17:73. <https://doi.org/10.1186/s40246-023-00522-3>.
66. Sahoo K, Sundararajan V. Methods in DNA methylation array dataset analysis: a review. *Comput Struct Biotechnol J*. 2024;23:2304–25. <https://doi.org/10.1016/j.csbj.2024.05.015>.
67. AWS. Benchmarking the Oxford Nanopore Technologies basecallers on AWS. 2022
68. ONT. Dorado: Oxford Nanopore Technologies' Basecalling Software. 2023
69. Amarasinghe SL, Su S, Dong X, et al. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol*. 2020;21:30. <https://doi.org/10.1186/s13059-020-1935-5>.
70. Sakamoto Y, Zaha S, Nagasawa S, et al. Long-read whole-genome methylation patterning using enzymatic base conversion and nanopore sequencing. *Nucleic Acids Res*. 2021;49: e81. <https://doi.org/10.1093/nar/gkab397>.
71. Zhou Q, Kang G, Jiang P, et al. Epigenetic analysis of cell-free DNA by fragmentomic profiling. *Proc Natl Acad Sci U S A*. 2022;119: e2209852119. <https://doi.org/10.1073/pnas.2209852119>.
72. Payne A, Holmes N, Clarke T, et al. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nat Biotechnol*. 2021;39:442–50. <https://doi.org/10.1038/s41587-020-00746-x>.
73. Nakamura W, Hirata M, Oda S, et al. Assessing the efficacy of target adaptive sampling long-read sequencing through hereditary cancer patient genomes. *NPJ Genom Med*. 2024;9: 11. <https://doi.org/10.1038/s41525-024-00394-z>.
74. Martin S, Heavens D, Lan Y, et al. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biol*. 2022;23:11. <https://doi.org/10.1186/s13059-021-02582-x>.
75. Shen SY, Singhania R, Fehringer G, et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature*. 2018;563:579–83. <https://doi.org/10.1038/s41586-018-0703-0>.

Publisher's Note

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