

## Bioinformatic Note



# HiFi Methylation Sequencing

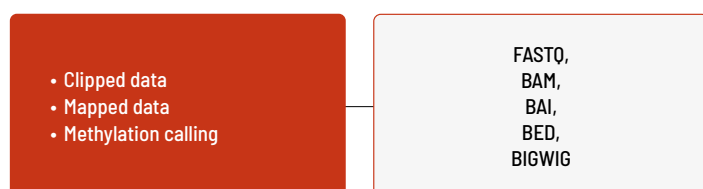
DNA methylation is one of the most common epigenetic modifications that fundamentally influence gene expression, cellular differentiation, and genomic imprinting. Without any change in the DNA sequence itself, gene activity and function can be regulated by DNA methylation. DNA methyltransferases mediate the regulation of gene activity and function by transferring a methyl group to the fifth carbon of the cytosine ring. In mammals, the resulting 5-methylcytosines (5-mC) and 5-hydroxymethylcytosines (5-hmC) occur mainly in cytosine-phosphate-guanine (CpG) dinucleotides. However, methylation can also be found in non-CpG contexts in other organisms.

Changes in the epigenetic signature, especially in DNA methylation, have been reported to happen in normal cell development and aging. However, alterations in DNA methylation are also closely associated with diseases like cancer, metabolic disorders, and neurological diseases. Global hypomethylation and locus-specific hypermethylation of CpG islands have been shown to increase genomic instability and promote tumor progression.

High-quality methylation data analysis can be used for:

- ✗ Biomarker discovery
- ✗ Clinical studies with methylation-associated treatments or other clinical and scientific applications
- ✗ Exploration of cell differentiation mechanisms, characteristic methylation profiles, and specific tissue development

We offer the following bioinformatic analysis for HiFi Whole Genome Methylation Sequencing (HiFi WGM):



## Clipped data

The sequencing reads are demultiplexed. HiFi reads (CCS reads with a predicted accuracy  $\geq Q20$ ) are clipped and provided in FASTQ format. Additionally, the quality of the FASTQ files is analyzed.

## Mapped data

Next, we map the demultiplexed and clipped sequencing data to a reference genome. In this example, the reads are mapped to the reference genome GRCh38. You receive the mapped BAM files, along with their corresponding index files in BAI format.

With PacBio's HiFi sequencing technology, DNA methylation can be detected without treating the DNA by using kinetic data. During sequencing, the kinetics of the polymerase is measured by capturing the inter-pulse duration and the pulse width of nucleotide incorporation. Thus, 5-methylcytosine (5mC) and 6-methyladenine (6mA) can be directly detected during the sequencing run. With the on-instrument secondary analysis workflow, methylation information is directly available. It is stored in the methylation tags **MM** (base modification / methylation) and **ML** (base modification probabilities) of the BAM file. BAM files are best opened and visualized in genome browsers, such as the Integrative Genomics Viewer (IGV), or processed with appropriate software.

## Methylation calling

Last, appropriate software tools are used to analyze the methylation information provided in the clipped BAM files. The software generates site methylation probabilities from mapped HiFi reads using distributions of modification scores and a machine-learning model to estimate modification probabilities across CpG sites. These site probabilities are reported as BED and bigwig (BW) outputs.

All BED files include a header with run and analysis metadata. The columns in the BED file contain the chromosome name (chrom), start coordinates (begin), end coordinates (end), modification scores (mod\_score), haplotypes (type), coverage (cov), estimated modified site counts extrapolated from the model modification score (est\_mod\_count), estimated unmodified site counts extrapolated from the model modification score (est\_unmod\_count), and discretized modification scores calculated from the estimated modified/unmodified site counts (discretized\_mod\_score). An excerpt of a BED file can be found in table 1. In addition to the BED file, a corresponding tabix index file is provided.

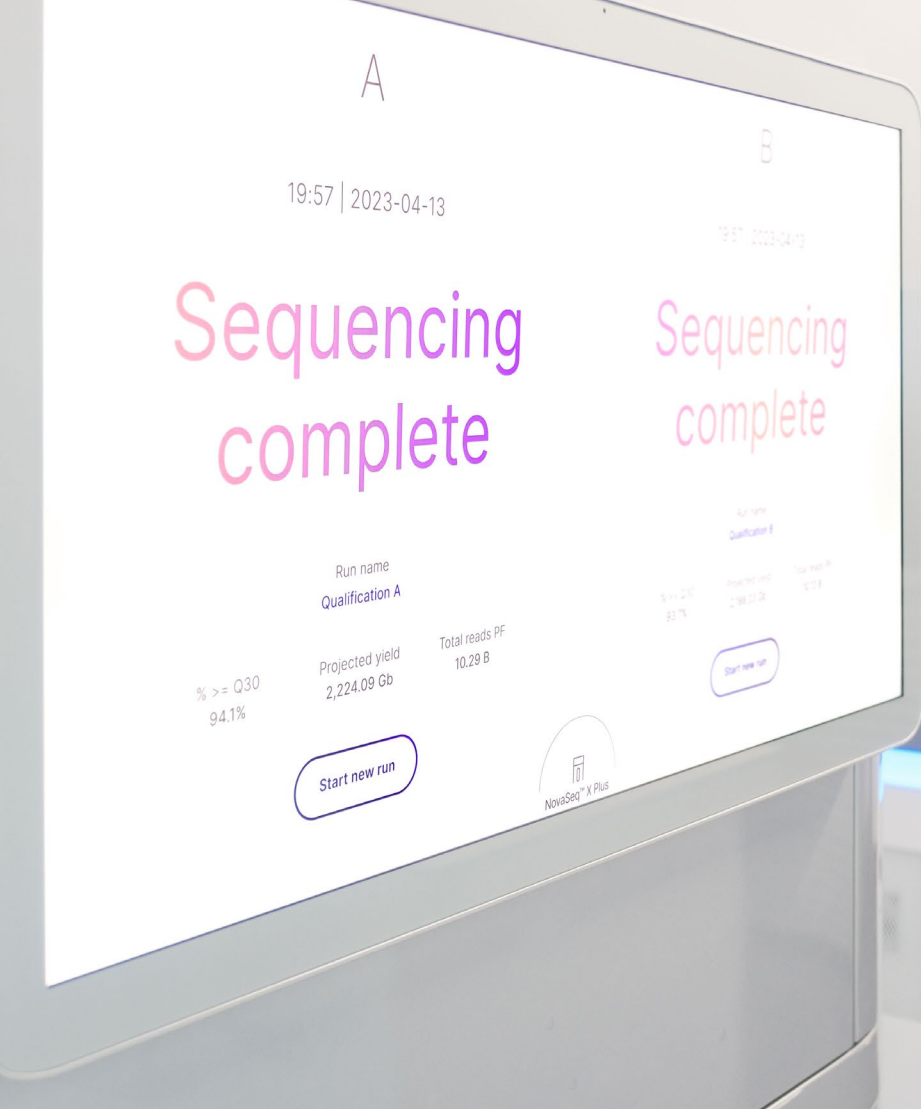
The indexed binary bigwig (BW) file can be used to visualize the methylation probabilities, e.g., in IGV.

Table 1 | Excerpt of the methylation probability BED file.

#chrom	begin	end	mod_score	type	cov	est_mod_count	est_unmod_count	discretized_mod_score
chr1	13643	13644	66.3	Total	4	3	1	75
chr1	21501	21502	65.4	Total	14	10	4	71.4
chr1	21537	21538	76.1	Total	13	10	3	76.9
chr1	63964	63965	73	Total	21	16	5	76.9
chr1	112067	112068	10.3	Total	32	3	29	9.4

The additionally generated project report provides information for every sample about the laboratory protocol, including data about quality control of the starting material, library preparation, sequencing platform, and the median QV for the HiFi reads. For the clipped data, the number of sequenced reads and bases is reported. The sequencing length and the GC content of

the clipped FASTQ reads are illustrated in bar plots for all samples. Additional statistics for the mapping step are provided, including the number of mapped reads, the proportion of the sequenced reads, the median insert size, and the average coverage.



## About Us

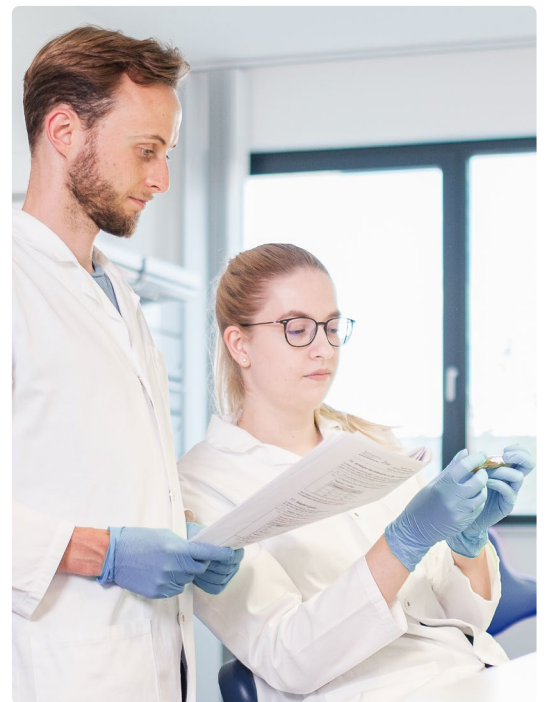
CeGaT was founded in 2009 in Tübingen, Germany. Our scientists are specialized in next-generation sequencing (NGS) for genetic diagnostics, and we also provide a variety of sequencing services for research purposes and pharma solutions. Our sequencing service portfolio is complemented by analyses suited for microbiome, immunology, and translational oncology studies.

Our dedicated project management team of scientists and bioinformaticians works closely with you to develop the best strategy to realize your project. Depending on its scope, we select the most suitable library preparation and conditions on our sequencing platforms.

**We would be pleased to provide you with our excellent service.  
Contact us today to start planning your next project.**



For more details please visit  
[www.cegat.com/methylation-sequencing](http://www.cegat.com/methylation-sequencing)



CeGaT GmbH  
Research & Pharma Solutions  
Paul-Ehrlich-Str. 23  
72076 Tübingen  
Germany

Phone: +49 7071 56544-333  
Fax: +49 7071 56544-56  
Email: [rps@cegat.com](mailto:rps@cegat.com)



CLIA CERTIFIED ID: 99D2130225



Accredited according to  
DIN EN ISO/IEC 17025