MEASURING DNA METHYLATION WITH 5-BASE HIFI SEQUENCING

Genome-wide detection and phasing of genetic and epigenetic variants from a single library prep

HiFi sequencing produces long, accurate reads of the 4 DNA bases — A, C, G, and T — that deliver the most comprehensive characterization of genomes.[1,2]

But HiFi sequencing is not limited to characterizing the genome. It simultaneously measures the epigenome by detecting a fifth base — 5mC at CpG sites.

- Detects distinct regional epigenetic patterns
- Accesses methylation in the full genome
- Identifies allele-specific methylation

	Methylation microarrays	Short-read sequencing	Nanopore sequencing	HiFi 5-base sequencing
SNVs	⊗	✓	✓	✓
Indels	⊗	✓	⊗	✓
SVs	⊗	⊗	✓	✓
Haplotype phasing	⊗	⊗	Limited	✓
Genomewide	⊗	✓	✓	✓
5mC in CpG contexts	Limited	Requires special library preparation	Requires special data processing. Conflated with basecalling.	✓

SMRTbell® library **PacBio® long-read systems**

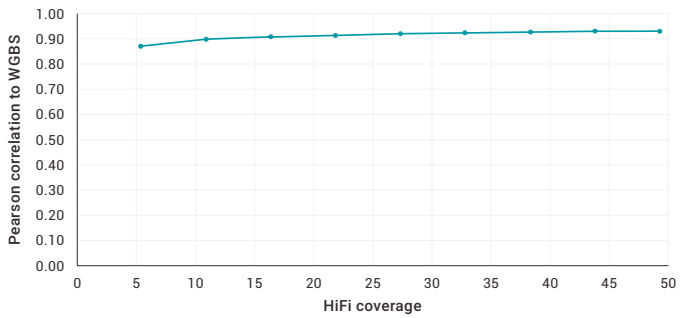
Nucleotide incorporation kinetics are measured in real time

5-base HiFi sequencing with A, C, G, T, +5mC

5mC encoded with standard BAM tags[3]
MM:Z:C+m,4,12,16,4,16,19,44,10
ML:B:C,249,4,247,177,210,228,245,244

The PacBio long-read systems directly output long, highly accurate HiFi reads with annotation of 5mC methylation at all CpG sites. No special library preparation like bisulfite treatment is required.

Coverage



Correlation of methylation calling in HiFi reads to whole-genome bisulfite sequencing (WGBS) of the human sample HG002.[4,5,6]

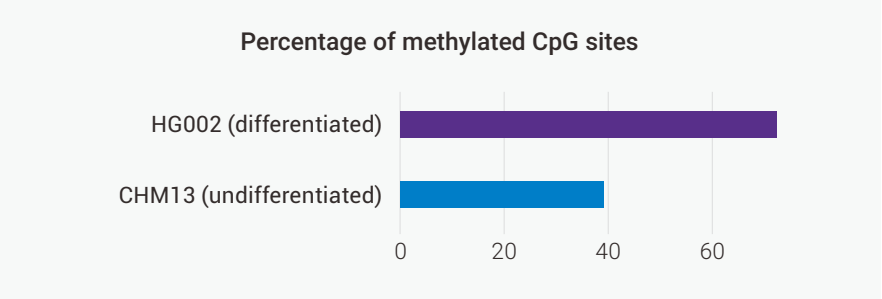
Applicability

Methylation	Species	5-base HiFi sequencing
5mC at CpG sites	Human and other vertebrates	✓
5mC at various motifs	Other eukaryotes, including plants	✓ Useful though partial view
4mC and 6mA	Microbes	Enabled through SMRT® Link microbial genome analysis

What 5-base sequencing reveals

Methylation across space and time

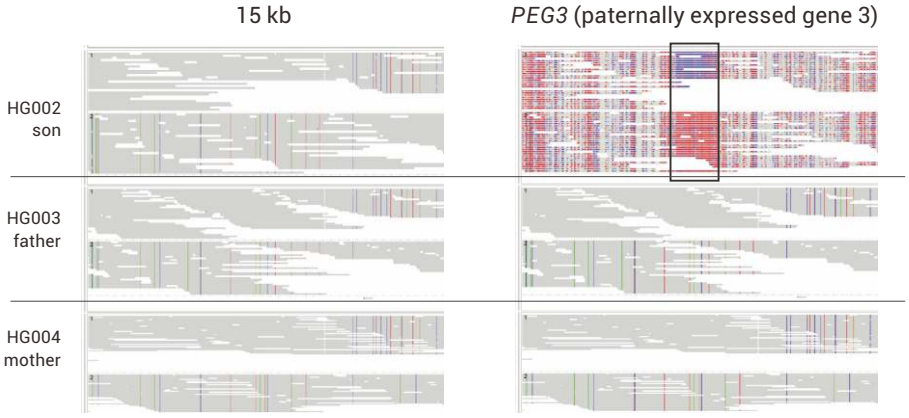
HiFi sequencing of a single sample detects methylation patterns across the genome, such as hypomethylation at transcription start sites. Sequencing multiple samples identifies differential methylation.



Most CpG sites are methylated in differentiated human cells like HG002 lymphoblastoids.[1] The genome shows lower levels of methylation in CHM13, a less differentiated human cell line.[6]

Parental imprinting

HiFi sequencing enables simultaneous phasing of reads into maternal and paternal haplotypes and detection of methylation. This directly reveals patterns of imprinting, where methylation status of a chromosome depends on whether it was inherited from a mother or father.

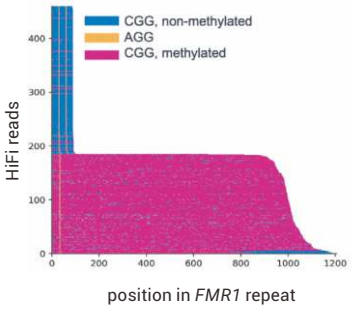


Maternal imprinting at PEG3 in a human trio. HiFi sequencing allows phasing of the haplotypes per sample; the trio identifies which allele is transmitted from which parent.

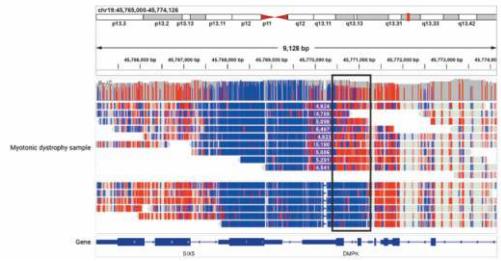
Atypical methylation patterns in rare disease

Atypical methylation patterns contribute to rare diseases like Prader-Willi syndrome and are important factors in pathogenic repeat expansions, such as the CGG expansion at the FMR1 locus that causes Fragile X syndrome. With high accuracy, long reads, and methylation detection, HiFi sequencing is ideal for characterizing these repeat expansions.

Methylation of expanded *FMR1* allele



DMPK repeat expansion induces hypermethylation



HiFi sequencing phases and identifies hypermethylation of expanded FMR1 repeats in NA07537 and the region adjacent to a DMPK expansion in a sample with myotonic dystrophy (Children's Mercy Kansas City).

Learn about 5-base sequencing:
pacb.com/epigenetics

KEY REFERENCES

1. Nurk et al. (2022) **The complete sequence of a human genome** *Science. 44–53*
2. Olson et al. (2021) **precisionFDA *Truth Challenge V2*: Calling variants from short- and long-reads in difficult-to-map regions** *biorxiv*
3. **Sequence alignment/map optional fields specification**
4. **pb-CpG-tools** PacBio GitHub
5. Foox et al. (2021) **The SEQC2 epigenomics quality control (EpiQC) study Genome Biology**
6. HG002 data release **https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA586863**

102-193-697 REV02 10OCT2022