Oxford NANOPORE Technologies

Investigating methylation in the human genome with Oxford Nanopore

GETTING STARTED

nanoporetech.com

Introduction

Epigenetics, the study of chemical modifications that can alter phenotype without altering nucleotide sequence, is an ever-growing field of research. The study of epigenetic modifications in humans — encompassing modifications to DNA, RNA, and histone proteins — is significant in fields ranging from developmental biology[1] to clinical research into diseases, including cancers[2] and neurological disorders[3].

The most well-characterised and widely studied epigenetic modification in mammalian genomes, and especially in humans, is 5-methylcytosine (5mC) DNA methylation — the addition of a methyl group to a cytosine nucleotide. Methylation plays an important role in regulating gene expression, with aberrant methylation in gene promoters being associated with disease.

There are many methods available for the analysis of methylation in human genomes. However, traditional techniques of detection present some limitations. For example, short-read sequencing technology involves PCR, during which epigenetic modifications are lost; as methylation cannot be sequenced from PCR products, the presence of methylation is instead inferred via chemical treatment of DNA samples prior to sequencing.

Using nanopore sequencing, it is possible to prepare and sequence native DNA and RNA molecules, without the need for PCR. This enables the direct detection of intact methylation alongside nucleotide sequence, without chemical conversion or additional library preparation steps. Providing unprecedented resolution of methylation in targeted regions or across the whole human genome, the end-to-end nanopore workflows combine simple library preparation with flexible sequencing options to suit your experimental goals.

This guide provides an introduction to the direct sequencing of DNA methylation in human genomes. Nanopore sequencing is also the only technology that enables the direct sequencing and analysis of epigenetic modifications in native RNA. To find out more about native RNA sequencing, visit **nanoporetech.com/rna-and-cdna-sequencing**

1. Greenberg, M.V.C. and Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Cell Bio.* 20: 590–607 (2019). DOI: https://doi.org/10.1038/s41580-019-0159-6
2. Simpson, J.T. et al. Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* 14: 407–410 (2017). DOI: https://doi.org/10.1038/nmeth.4184
3. Weng, Y.L., et al. DNA modifications and neurological disorders. *Neurotherapeutics* 10(4): 556-567 (2013). DOI: https://doi.org/10.1007/s13311-013-0223-4

Gold-standard methylation calling with Oxford Nanopore: benchmarking performance

To benchmark nanopore methylation calling performance, we sequenced two replicate native DNA libraries of the well-characterised human genome HG002 on a PromethION™ device, generating 20x depth of coverage per sample (ONT_1 and ONT_2); these datasets were also merged to produce a third with 40x coverage (ONT_3). Basecalling and modification calling was performed using Remora: an algorithm that is now integrated into the basecaller Dorado, which can be used as a standalone tool or through the software onboard nanopore sequencers. These methylation calls were compared to those from two short-read bisulfite sequencing datasets (BS-Seq_1[4] and BS-Seq_2[5,6]).

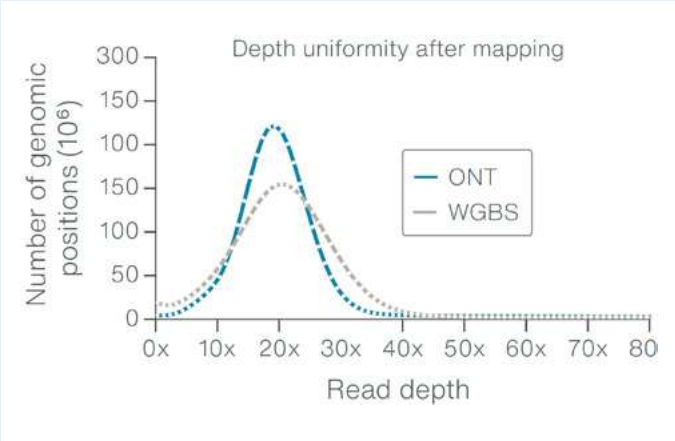


Figure 1. Nanopore methylation data shows greater evenness of coverage than whole-genome bisulfite sequencing (WGBS)

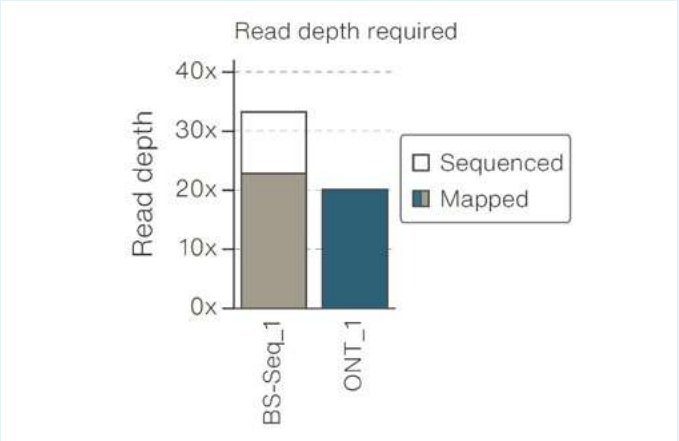


Figure 2. Whilst short-read bisulfite data shows incomplete mapping, nanopore methylation reads map fully to the reference genome

Bisulfite conversion followed by short-read sequencing is a widely used method of methylation detection. Bisulfite converts unmethylated cytosines to uracil, which is read as thymine during sequencing. 5mC is not converted, although the modification is lost during PCR. Comparison to a reference, or ideally a paired, untreated dataset, is then used to infer the presence of 5mC as cytosine. However, bisulfite treatment is a time-consuming process requiring the use of hazardous chemicals, and considerably fragments the sample, causing degradation of up to 90% of the DNA[7]. Chemical conversion can also be incomplete, with variability seen from sample to sample, which may impact the validity of results; furthermore, bisulfite treatment does not distinguish between 5mC and 5-hydroxymethyl cytosine (5hmC).

In 5mC benchmarking studies, the nanopore sequencing data showed greater uniformity of coverage than the bisulfite data **(Figure 1)**. Not all DNA sequences are amenable to PCR, meaning that areas of the genome, such as GC-rich regions, may be poorly represented in short-read sequencing datasets, or missed entirely. With PCR-free nanopore sequencing, it is possible to generate more even coverage across the genome, including regions that are inaccessible to other sequencing methods[8]. Bisulfite conversion reduces sequence complexity which, combined with short-read sequencing, makes mapping to a reference genome challenging. By contrast, the long, PCR-free nanopore reads mapped to the reference genome (hg38) fully **(Figure 2)**.

In the nanopore datasets, the combined analysis time for both basecalling and modified base calling was over 2.5x faster than that for bisulfite modification calling alone. A high correlation (0.97) was seen between the methods for 5mC calling, and the nanopore datasets showed high reproducibility (0.95). With nanopore methylation detection, there is also no need to sequence a second, paired sample to determine the presence or absence of methylation — or for a separate run to be performed to enable the analysis of other variants of interest. Instead, structural variants (SVs), single nucleotide variants (SNVs), copy number variants (CNVs), short tandem repeats (STRs), and methylation can all be called from a single dataset.

Long and ultra-long nanopore reads retain long-range information for methylation and genomic variants, and also enable effective phasing. An HG002 library prepared with the Ultra-Long DNA Sequencing Kit produced a read N50 of 100 kb with a mapped depth of coverage of 40x. SNP-based phasing of the data enabled haplotype resolution of 90% of methylation calls, representing >80% of CpGs across the human genome.

4. Publicly available BS-Seq open dataset. Available at: https://labs.epi2me.io/category/data-releases/ [Accessed: 11 Jan 2024].
5. NCBI. Sample GSM5649436. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM5649436 [Accessed: 11 Jan 2024].
6. NCBI. Sample GSM5649437. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM5649437 [Accessed: 11 Jan 2024].
7. Grunau, C. et al. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res.* Jul 1;29(13):E65-5 (2001). DOI: https://doi.org/10.1093/nar/29.13.e65
8. Ebbert, M.T.W. et al. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biol.* 20,97 (2019). DOI: https://doi.org/10.1186/s13059-019-1707-2

Gold-standard methylation calling with Oxford Nanopore: benchmarking performance

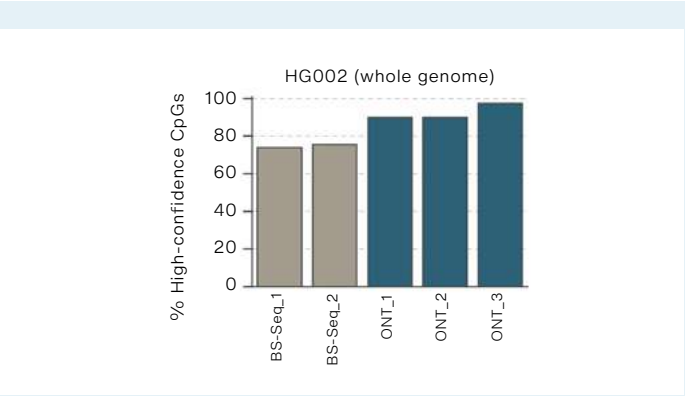


Figure 3. Percentage of called high-confidence CpGs in 20x bisulfite datasets compared with 20x and 40x nanopore methylation datasets

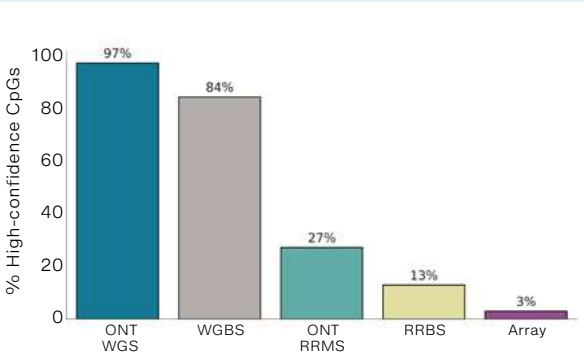


Figure 4. Percentage of CpGs detected in the human genome using different sequencing approaches (40x depth) and methylation microarray

The uniform sequencing coverage seen in nanopore data allows for methylation detection across the human genome. 5mC is most commonly seen in CpG dinucleotides. Comparing CpG methylation calling between the datasets revealed that 20x bisulfite data enabled only ~75% of CpGs to be called in the human genome, whilst at the same depth of coverage, nanopore sequencing enabled ~90% of CpGs to be called **(Figure 3)**. Increasing nanopore sequencing depth to 40x enabled detection of >97% of CpGs. Though this dataset focuses on 5mC detection, a 5hmC Remora model is also available. We observe positive benchmarks for the 5hmC model validation metrics on a range of sample types. It is also possible to train your own Remora model to detect any epigenetic modification of interest.

Compared with whole-genome bisulfite sequencing, nanopore methylation detection demonstrates:

- Strong correlation
- A higher number of CpG positions called
- Lower data requirements
- Faster analysis
- A simpler workflow, with no toxic components
- Better reproducibility and consistency run-to-run
- More even coverage, less effect of GC bias
- The option to phase methylation

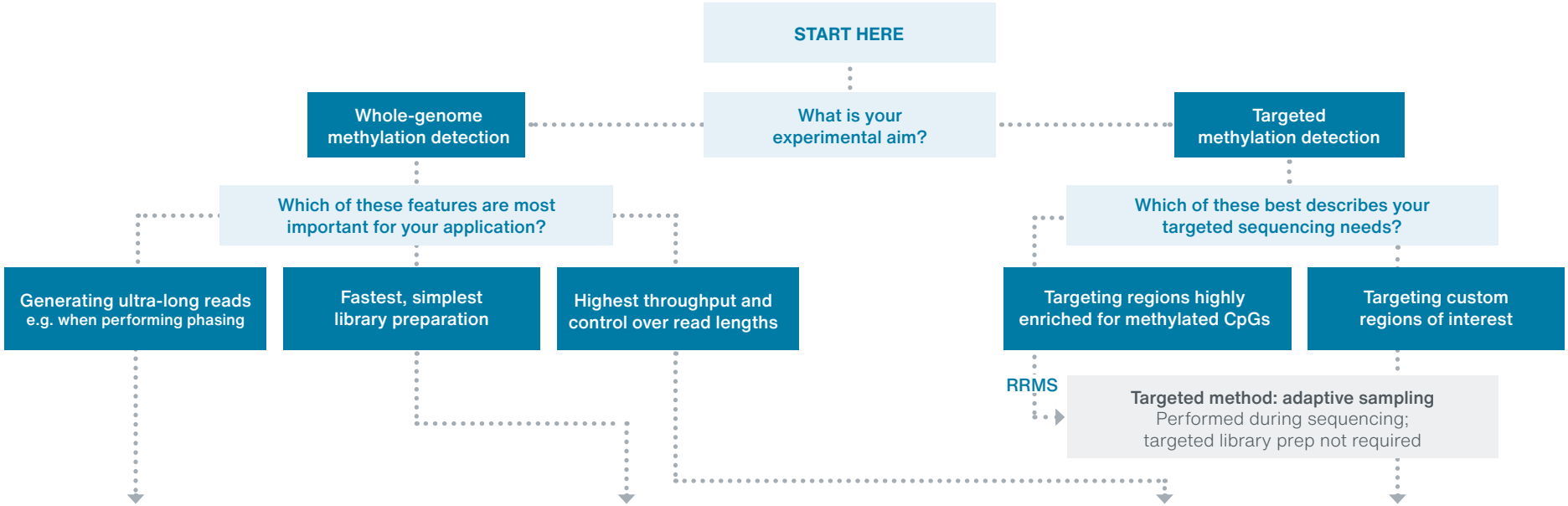
Targeted methods offer a cost-efficient way to perform methylation analysis without the need to perform whole-genome sequencing. Methylation microarrays and reduced-representation bisulfite sequencing (RRBS) are frequently used to target methylated CpGs; however, both rely on bisulfite conversion. Methylation arrays have been used in large-scale 5mC profiling studies, despite being limited to only thousands of CpG sites, representing ~3% of the ~28 million CpGs present in the human genome **(Figure 4)**. RRBS focuses on a larger subset of CpG islands; however, the method is expensive and time consuming.

To address these challenges, Oxford Nanopore has developed Reduced-Representation Methylation Sequencing (RRMS): a cost-effective method for characterisation of methylation patterns across regions of interest, such as CpG islands. The simple, end-to-end workflow utilises adaptive sampling **(see page 7)**, which requires no special library prep, offering flexible, real-time enrichment of native DNA during sequencing. Canonical base calling is performed alongside the detection of base modifications, so that nucleotide sequence and epigenetic changes, such as 5mC and 5hmC, can be detected simultaneously in a single sequencing experiment. RRMS displays more even coverage and higher reproducibility than RRBS, while capturing a significantly larger proportion of CpGs **(Figure 4)** than RRBS and methylation arrays — with near-perfect correlation (0.94) observed between the methods.

View the methylation benchmarking results in full:
nanoporetech.com/methylation-benchmarking-poster

Read more about RRMS:
nanoporetech.com/rrms-poster

Which approach do I choose?



Ultra-Long DNA Sequencing Kit	
Preparation time	~200 mins + overnight elution
Input recommendation	6 million tissue culture cells or 1–2 ml blood
Fragmentation	Transposase-based
Read length	N50 >50 kb, reads up to 4+ Mb

Rapid Sequencing Kit	
Preparation time	10 minutes
Input recommendation	100 ng HMW gDNA (>30 kb)
Fragmentation	Transposase-based
Read length	Random distribution, dependent on input fragment length

Ligation Sequencing Kit	
Preparation time	~60 minutes
Input recommendation	1 µg dsDNA
Fragmentation	Optional; recommended for inputs of 100–500 ng
Read length	Equal to fragment length

Find out more about Oxford Nanopore library preparation solutions:
nanoporetech.com/prepare

From sample to answer

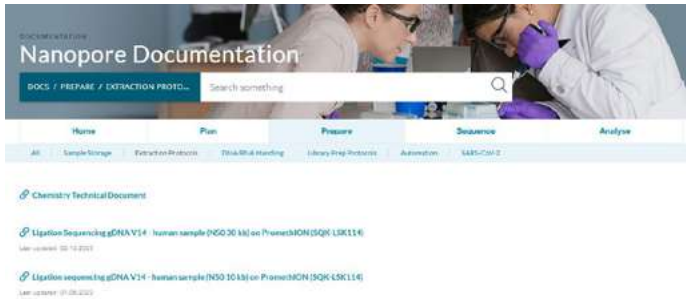
EXTRACTION

Can I extract high-quality DNA from my sample?

The Nanopore Documentation section of the Nanopore Community features recommended DNA extraction protocols for a wide range of sample types, including from cell lines, blood, and saliva. In this section you'll also find info sheets covering effective sample storage, DNA (and RNA) handling, size selection, and more.

View extraction protocols:
community.nanoporetech.com/docs/prepare/extraction_protocols

Nanopore Community Documentation



LIBRARY PREP

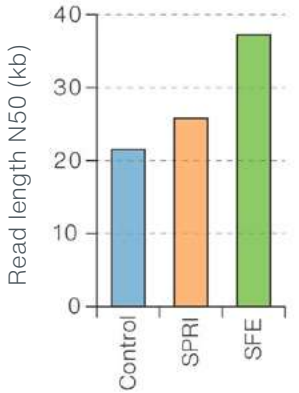
Should I perform fragmentation or size selection?

As read length is unrestricted in nanopore sequencing, fragmentation of your DNA sample is optional. However, unless you plan to phase your data, maximising read length is generally not required for methylation detection. We have observed that performing light shearing can help improve sequencing output, to help maximise depth of coverage of methylated regions of interest.

If you do wish to perform phasing, you can optimise for long reads by avoiding fragmentation, whilst size selection can be used to minimise unwanted shorter fragments. For longer reads still — N50s over 50 kb and reaches reaching the megabase scale — the Ultra-Long DNA Sequencing Kit can be used **(see page 5)**. Guidance on both fragmentation and size selection is available on the Community.

Find out more about optional fragmentation and size selection methods:
community.nanoporetech.com/docs/prepare

Read length N50s from three sequencing libraries, prepared without size selection (control), with SPRI size selection (SPRI), and with the Oxford Nanopore Short Fragment Eliminator Expansion (SFE)



From sample to answer

LIBRARY PREP

Can I perform targeted sequencing without losing methylation information?

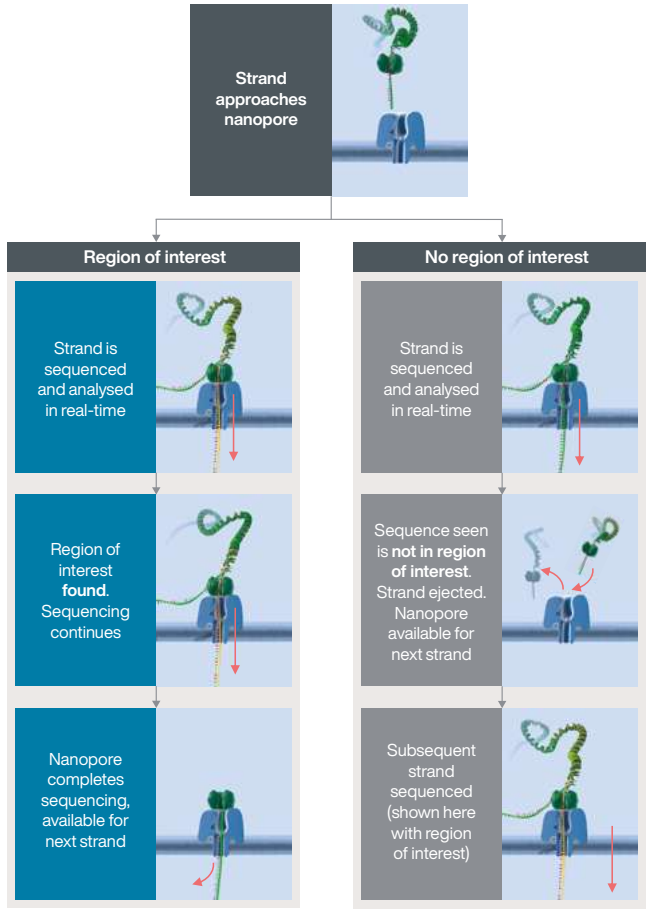
Opting for a targeted approach is ideal for efficiently assessing methylation in specific regions of interest. To ensure that methylated bases are preserved through library preparation for detection in sequencing, PCR-free target enrichment is essential.

Adaptive sampling is an innovative, PCR-free method of target enrichment that is performed entirely during a nanopore sequencing run, with no special library prep needed; it is also the method underpinning Reduced-Representation Methylation Sequencing (RRMS). Sequences to be enriched (or alternatively, depleted) are supplied to MinKNOW™, the software on board nanopore sequencing devices, in a BED file. DNA molecules representing regions of interest are allowed to pass through nanopores and are sequenced in full. When an off-target sequence enters a nanopore, it is recognised in real time and ejected from the pore, freeing up more sequencing time for regions of interest. With no limit to the length of target regions, adaptive sampling is ideal for enrichment of very large regions of interest from native DNA samples, such as megabase-scale SVs or whole chromosomes.

Learn about RRMS, targeted methylation detection with adaptive sampling:
nanoporetech.com/methylation-rrms

View the Targeted sequencing Getting started guide:
nanoporetech.com/targeted-sequencing-guide

The adaptive sampling workflow



From sample to answer

LIBRARY PREP

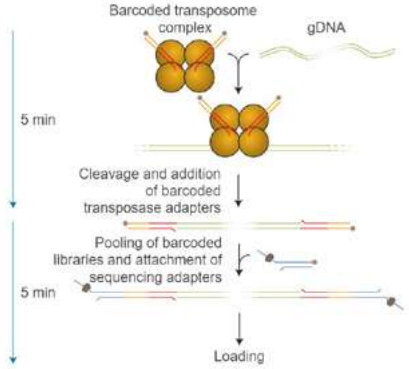
How can I sequence multiple samples at once?

Our **Native Barcoding Kits** enable sample barcoding without the use of PCR, for multiplexed sequencing of up to 96 native DNA libraries. This is achieved via the ligation of barcode adapters to each sample to be sequenced in the same run. These are then pooled and prepared for sequencing as a single sample.

Where rapid library preparation is key, the **Rapid Barcoding Kits** enable the preparation of up to 96 barcoded libraries in one hour. Here, a transposase-based method is used to shear the DNA and attach barcoded sequencing adapters in a single step.

Find out more about nanopore sequencing kits:
nanoporetech.com/products/prepare

The rapid barcoding workflow



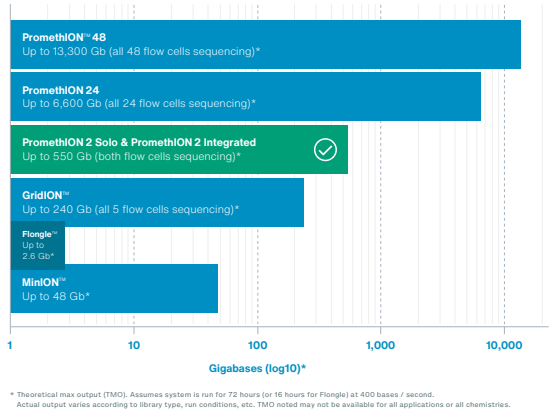
SEQUENCING

Which device should I choose?

A wide range of nanopore sequencing devices is available, with each suiting different experimental requirements. For targeted analysis of methylation in specific regions, the MinION™, a portable device that can be operated from a laptop, is ideal. Providing the flexibility to scale up or down, the GridION™ allows sequencing on up to five MinION Flow Cells at once, for on-demand sequencing of target panels or whole genomes. Flongle™ adapts MinION or GridION for use with smaller Flongle Flow Cells. Flongle provides the output to sequence a small number of targets of interest and can also be used to QC libraries prior to sequencing on higher output devices.

For analysis of methylation across the whole human genome, we recommend sequencing on a PromethION device. The ultra-high throughput, benchtop PromethION 24 and PromethION 48 devices, with powerful onboard compute, enable sequencing on up to 24 or 48 PromethION Flow Cells respectively. The compact PromethION 2 Integrated, with integrated compute, and PromethION 2 Solo, which connects to a GridION or existing compute infrastructure, provide a simple plug-and-play setup, offering the flexibility of two independent, high-output PromethION Flow Cells for lower sample throughput requirements.

Maximum outputs of nanopore sequencing devices*



* Theoretical max output (TMO). Assumes system is run for 72 hours (or 16 hours for Flongle) at 400 bases / second.
Actual output varies according to library type, run conditions, etc. TMO noted may not be available for all applications or all chemistries.

Find out more about nanopore sequencing devices:
nanoporetech.com/products

From sample to answer

SEQUENCING

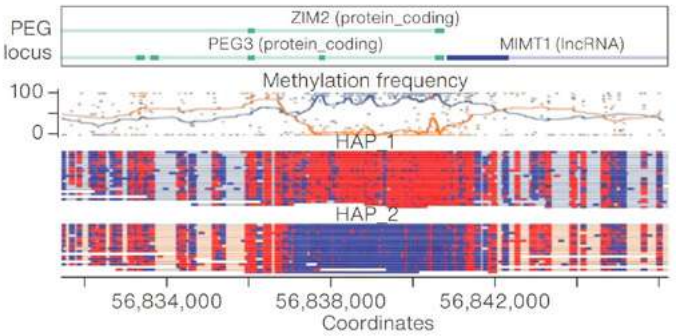
How much data do I need?

For good metrics in human genome-wide calling of 5mC and 5hmC methylation, we recommend sequencing to 20x depth of coverage; this was the read depth used in our nanopore methylation calling performance benchmarking dataset **(see pages 3–4)**.

To phase your data, we recommend increasing read depth to around 30x. These depths can be achieved by sequencing on one PromethION Flow Cell for 72 hours. Long and ultra-long, native nanopore reads allow for easy and accurate phasing of data, with uniform coverage enabling haplotype-resolved methylation calling across large regions of interest.

View best practice guidance on epigenetics, SNV calling and phasing, and more:
nanoporetech.com/resource-centre/workflows

A differentially methylated region overlapping two imprinted genes in the human genome was observed with haplotype-resolved nanopore methylation data



DATA ANALYSIS

How can I detect methylation in my data?

For the highest accuracy canonical base and methylation calls from a single run, we recommend using the basecaller Dorado, which delivers gold-standard methylation calling alongside canonical basecalling, with minimal impact to basecalling speed. Dorado is integrated into MinKNOW — the software onboard nanopore sequencing devices — enabling live basecalling; it can also be used as a standalone tool. We recommend using the high accuracy (HAC) basecalling model. Using Dorado in HAC mode on a PromethION 24 device, live methylation calling can be performed in real-time in all positions concurrently.

For a deeper analysis of variants within the human genome, we recommend the workflow **wf-human-variation**. As well as providing methylation annotations and enabling haplotype phasing, this workflow — an EPI2ME™ solution — allows for the analysis of SNVs, CNVs, SVs, and STRs. The workflow can be run with simple, point-and-click implementation, or via the command line.

Find out more about live basecalling capabilities on our devices:
nanoporetech.com/products/sequence

Find out more about nanopore data analysis:
nanoporetech.com/analyse

From sample to answer

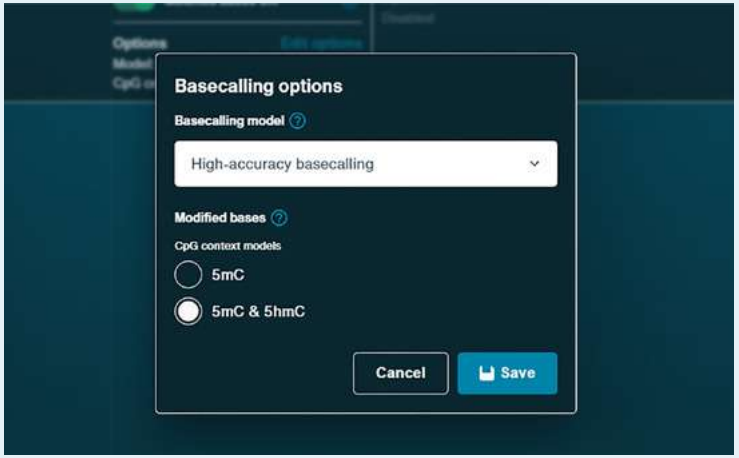
DATA ANALYSIS

Can I distinguish between 5mC and 5hmC?

Unlike bisulfite sequencing and other available sequencing technologies, which are unable to distinguish between 5mC and 5hmC, nanopore technology can accurately detect and distinguish between both modifications in the same dataset.

Models for the detection of both 5mC and 5hmC are available. This will be expanded to include models for further modifications.

Read more about the latest modification models integrated into Dorado:
github.com/nanoporetech/dorado



DATA ANALYSIS

Can I detect methylation at single-base resolution?

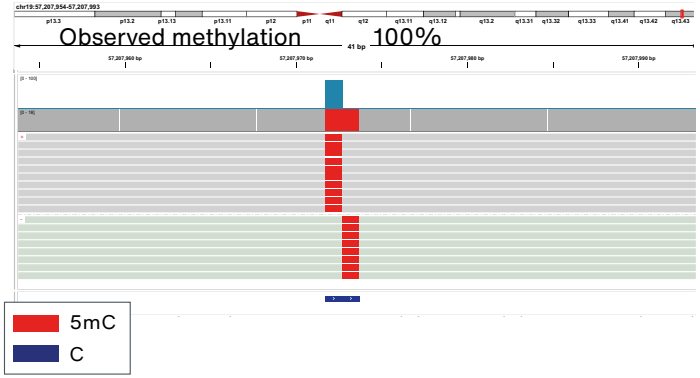
Traditional analysis of CpG methylation typically involves characterising averaged levels of CpG methylation at a site of interest; however, this precludes resolution of heterogeneity. With nanopore sequencing, it is possible to perform high-accuracy methylation calling at the single-base level, allowing for exploration of epigenetic heterogeneity with high confidence.

Single-molecule DNA methylation patterns allow:

- Investigation of heterogenous cell populations
- Identification of epialleles
- Sample comparisons: e.g. matched tumour-normal sequencing in cancer research

Find out more about epigenetics and methylation analysis using nanopore sequencing:
nanoporetech.com/epigenetics

Nanopore sequencing enables single-nucleotide resolution of base modifications with high accuracy



Case studies

Case study 1: Exploring the human epigenome with ultra-long nanopore reads

Studying the epigenetic profile of the human genome has, until recently, proven challenging — firstly due to the lack of a complete human reference genome assembly, and secondly because of the reduced mappability of short-read sequencing-based methods of methylation detection. Making use of the release of the first complete human genome reference sequence by the Telomere-to-Telomere consortium, CHM13–T2T[9], Gershman *et al.* used ultra-long nanopore reads to investigate the human epigenome[10]. The data enabled them to call CpG methylation in newly complete regions of the human genome, such as repeat-rich areas that could not previously be resolved. Sequencing DNA from the CHM13 cell line and HG002, they were able to assess 5mC frequency at 99.7% and 99.9% of CpGs within the human genome respectively. Combining the data with that obtained from a nanoNOMe experiment — a nanopore sequencing-based chromatin accessibility method — they investigated the epigenetic patterns within centromeric repetitive arrays. The group were also able to use the methylation data alone to phase repetitive regions of the genome that were devoid of heterozygosity and explore epigenetic heterogeneity within the centromere. Their analysis, revealing the epigenetic profiles of the most challenging regions of the genome, represents *'the most complete human methylome ever produced'*.[10]

Read the publication (Apr 2022):
nanoporetech.com/resource-centre/epigenetic-patterns-complete-human-genome

Case study 2: Characterising methylation markers in imprinting disorders with adaptive sampling

Prader-Willi syndrome (PWS) and Angelman syndrome (AS) are both imprinting disorders, resulting from aberrant genomic imprinting on chromosome 15q11.2. Both syndromes involve methylation of a CpG island within the promoter region of the gene *SNRPN*; the conditions are identified by determining the methylation state of this region. In healthy individuals, the maternally derived allele is methylated, whilst the paternal allele is unmethylated. In subjects with PWS, both alleles are methylated; in subjects with AS, both are unmethylated.

Currently, special library preparation steps, such as bisulfite treatment followed by short-read sequencing, are required to assess methylation. However, bisulfite treatment is a time-consuming process that considerably fragments DNA, whilst the use of PCR can result in bias, leading to poor resolution of areas such as GC-rich regions. In a proof-of-concept study, Yamada *et al.* used PCR-free enrichment via adaptive sampling to investigate methylation within the relevant region of the *SNRPN* gene[11].

Adaptive sampling enabled sequencing of the target region to ~17.8x depth of coverage and calling of CpG methylation within the region. The nanopore data showed that *'virtually all of the CpGs within the differentially methylated region of the promoter region of the* SNRPN *locus were methylated'* in the PWS research samples, whereas *'virtually none'* were methylated in the AS samples, demonstrating the potential of the technology to clearly characterise the epigenomic markers of each disease. The team highlighted the potential of nanopore sequencing with adaptive sampling as *'a very efficient one step assay'* for identifying the molecular markers and underlying mechanisms of PWS and AS.

Read the publication (Feb 2023):
nanoporetech.com/resource-centre/pws-as

9. Nurk, S. et. al. The complete sequence of a human genome. *Science* 376(6588): 44-53 (2022). DOI: https://doi.org/10.1126/science.abj6987
10. Gershman, A. et. al. Epigenetic patterns in a complete human genome. *Science* 376(6588) (2022). DOI: https://doi.org/10.1126/science.abj5089
11. Yamada, M. et al. Diagnosis of Prader-Willi syndrome and Angelman syndrome by targeted nanopore long-read sequencing. *Eur. J. Med. Genet.* 66(2), 104690 (2023). DOI: https://doi.org/10.1016/j.ejmg.2022.104690

Oxford Nanopore Technologies
Phone: +44 (0)845 034 7900
Email: support@nanoporetech.com
𝕏 @nanopore

www.nanoporetech.com

GS_1131(EN)_V4_22Jan2024