1. Data QC

Sequencing run summary:

- run duration: ~ 5.17 hours
- active channels: 491 pores/channels were active during the run
- barcodes: no

Basecall and yield metrics:

- total throughput: the run generated a total of 316,470 reads comprising 413.32 million bases (Mb)
- pass filter efficiency: data passed the quality filters: 310,919 reads (406.50 Mb) were classified as "Pass Reads"; ~98.2% of the generated reads met the sequencing quality threshold.

Read length and quality profiles:

- median read length: the median read length is 1,148 bp for all reads and 1,149 bp for pass reads
- N50 metric: the N50 length is 1,326 bp (1,327 bp for pass reads); half of the total sequenced bases are contained in reads longer than this value.
- PHRED Quality Scores: the median PHRED score is 10.42 for all reads and 10.46 for pass reads, reflecting an overall accuracy rate above 90%.

PycoQC report - generated with pycoQC 2.5.2

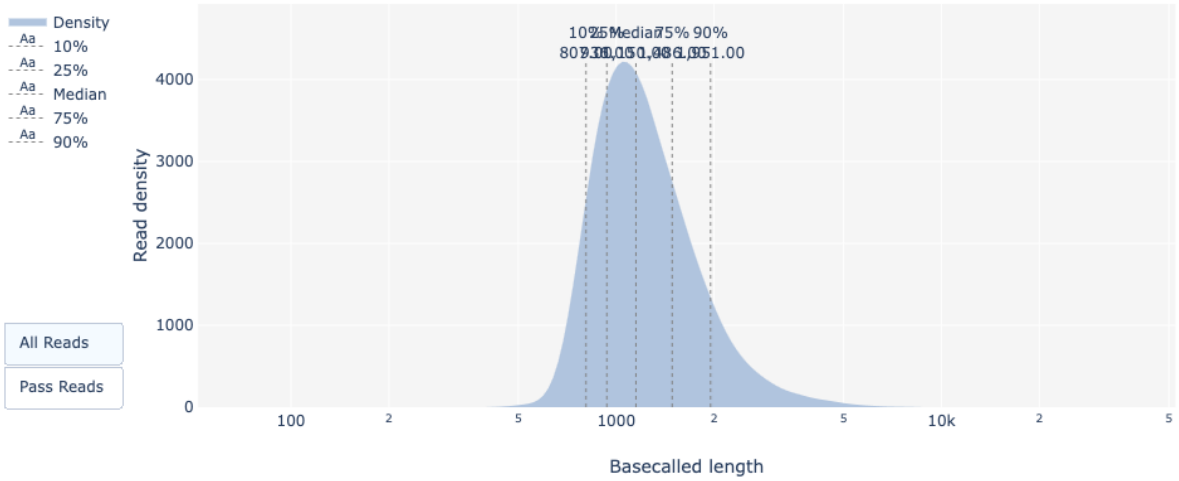
General run summary

Status	Run Duration (h)	Active Channels	Number of Runids	Number of Barcodes
All Reads	5.17	491	1	1
Pass Reads	5.17	491	1	1

Basecall summary

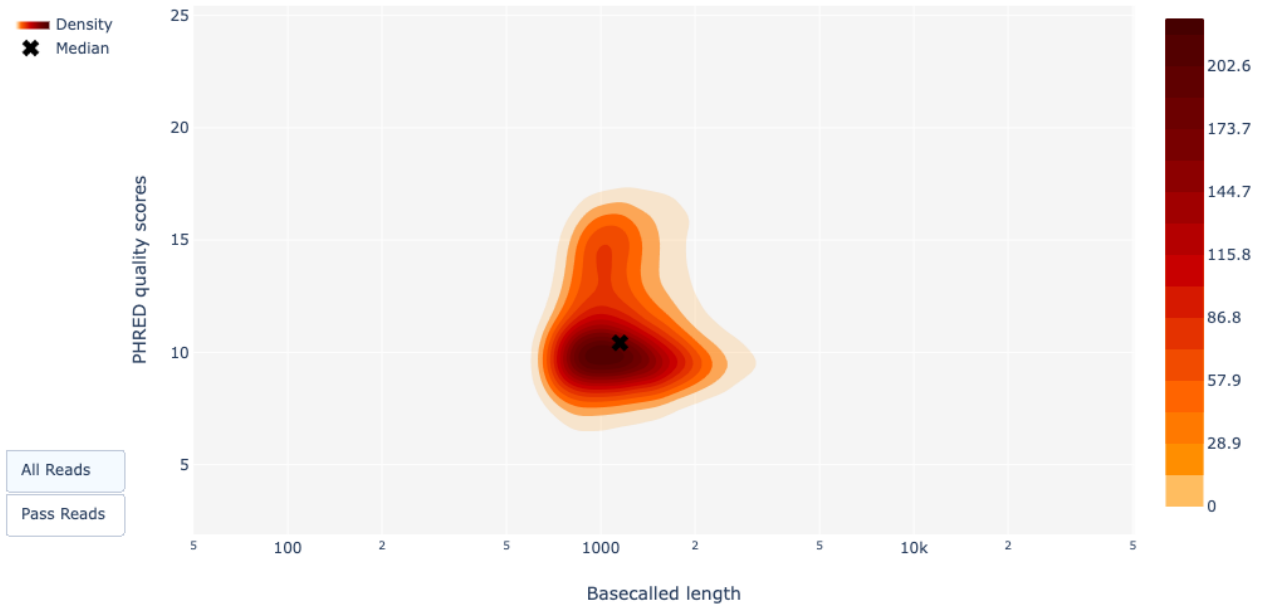
Status	Reads	Bases	N50	Median Read Length	Median PHRED score
All Reads	3.164700e+5	4.133204e+8	1330	1150	10.419
Pass Reads	3.109190e+5	4.064976e+8	1330	1150	10.461

Basecalled reads length

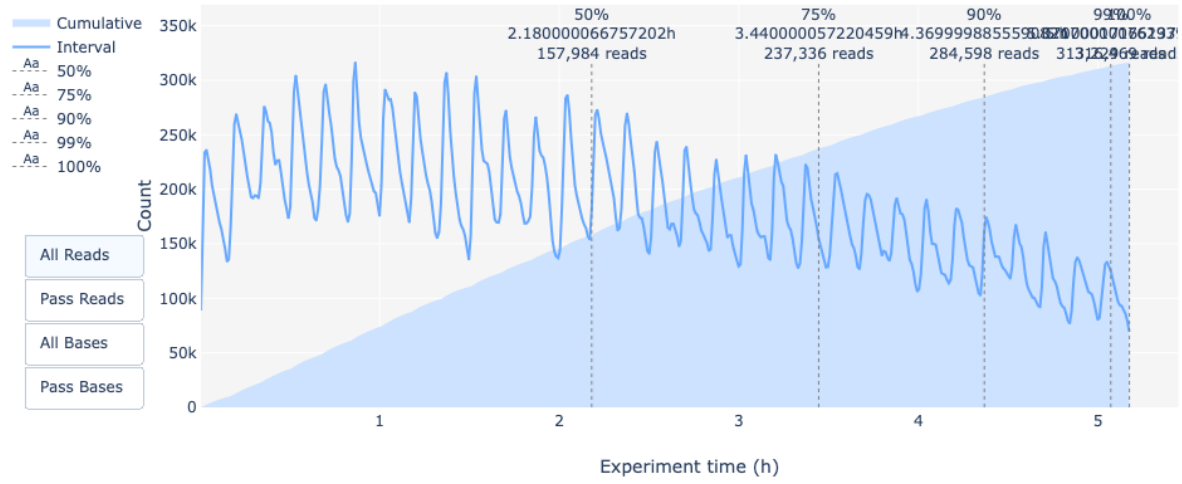


Basecalled reads PHRED quality

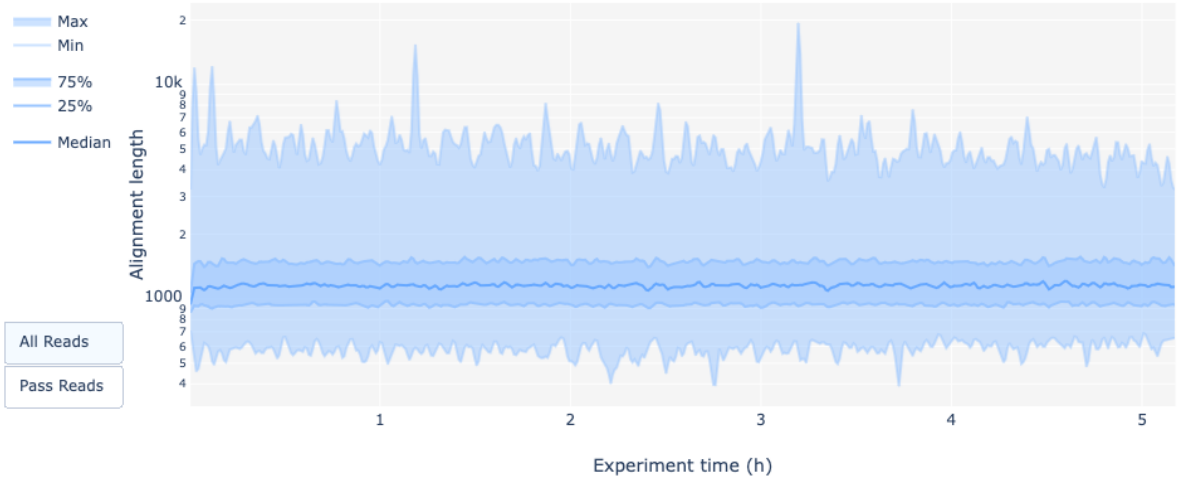
Basecalled reads length vs reads PHRED quality



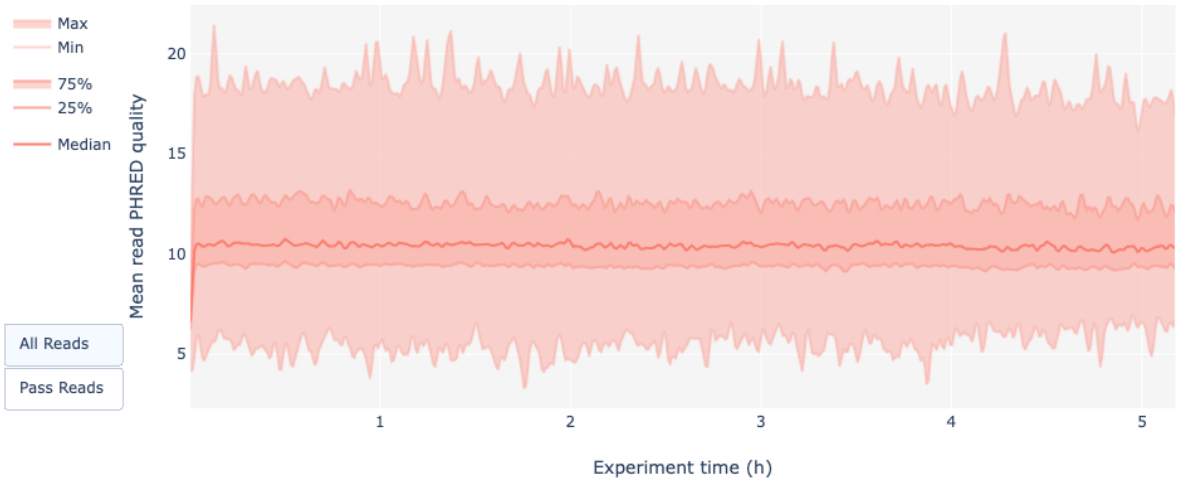
Output over experiment time



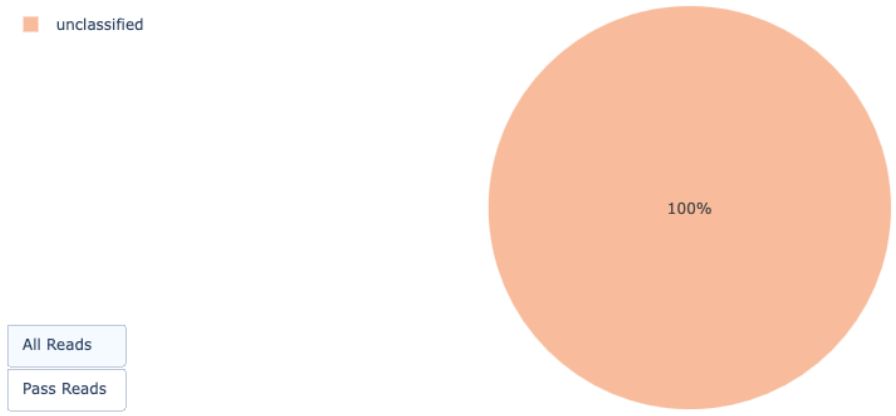
Read length over experiment time



Read quality over experiment time



Number of reads per barcode



2. Confidence summary

The table summarises metrics extracted from the modkit summary reports across two stages, step 5 and step 7, basecalling and mapping to the genome, respectively. The confidence threshold ranging from 0.75 to 0.95. The total number of identified epigenetic modifications refers to 5mC and 5hmC.

Threshold	Modification	Basecalling (step 5)	Mapping on human genome (step 7)	Human signal retention [%]
0.75	**5hmC** (h) **5mC** (m)	2,719,526 619,947	7,886 1,550	~0.29% ~0.25%
0.80	**5hmC** (h) **5mC** (m)	2,395,283 523,198	7,038 1,306	~0.29% ~0.25%
0.85	**5hmC** (h) **5mC** (m)	2,022,218 420,511	6,049 1,073	~0.30% ~0.26%
0.90	**5hmC** (h) **5mC** (m)	1,601,375 315,032	4,909 794	~0.31% ~0.25%
0.95	**5hmC** (h) **5mC** (m)	1,078,801 196,467	3,487 517	~0.32% ~0.26%

3. Protocol

This protocol was run and tested on MacBook Pro M5 with 24 Gb RAM with 500Gb free space. The protocol can be run on any other platform if required software is available. The original code

developed for this project is available through GitHub at https://github.com/urniaz/Epigenetic-Wastewater-Surveillance or can be provided on request.

Step 0. Set up environment

The protocol bases on the software versions provided. Required software need to be installed prior to the run.

Software	Version	Instructions how to install
Python	3.14.6	https://www.python.org/downloads/
pod5	0.0.43	https://pypi.org/project/pod5/0.0.43/
dorado	0.9.6	https://github.com/nanoporetech/dorado/releases/tag/v0.9.6 [1]
ont_fast5_api	4.1.3	https://github.com/nanoporetech/ont_fast5_api
wget	1.25.0	https://formulae.brew.sh/formula/wget
minimap2	2.31	https://formulae.brew.sh/formula/minimap2
bedtools	2.31.1	https://formulae.brew.sh/formula/bedtools
pycoqc	2.5.2	https://github.com/a-slide/pycoQC

[1] pre-compiled files from the official distribution are available here:
[dorado-0.9.6-linux-x64](https://cdn.oxfordnanoportal.com/software/analysis/dorado-0.9.6-linux-x64.tar.gz),
[dorado-0.9.6-linux-arm64](https://cdn.oxfordnanoportal.com/software/analysis/dorado-0.9.6-linux-arm64.tar.gz),
[dorado-0.9.6-osx-arm64](https://cdn.oxfordnanoportal.com/software/analysis/dorado-0.9.6-osx-arm64.zip),
[dorado-0.9.6-win64](https://cdn.oxfordnanoportal.com/software/analysis/dorado-0.9.6-win64.zip)

Step 1

Automated retrieval of compressed raw nanopore sequencing data from AWS S3 cloud storage for five wastewater samples and followed by extraction local directory.

Step 2

Quality filtering to purge instrument calibration files (mux_scan) and conversion from single-read to multi-read format.

Step 3

Conversion from FAST5 to POD5.

Step 4

Pooling POD5 into a single archive.

Step 5

Run Dorado basecaller using a high-accuracy to call 5mC/5hmC epigenetic modifications.

Step 6

Download of the human reference genome assembly (GRCh38.p14), sequence alignment via minimap2, and sorting and indexing.

Step 7

The methylations were quantified using modkit at set confidence threshold and followed by genomic annotation of the sites using bedtools intersect.