Guidelines for Genomic DNA Sequencing in the PromethION (ONT_gDNA)

Introduction

Nanopore sequencing developed by Oxford Nanopore Technologies (ONT) uses molecular sensors or *nanopores* capable of measuring changes in electric current when a single DNA or RNA molecule passes through them. Different sequences going through a nanopore give different electrical signals, producing an electrical trace or 'squiggle.' These squiggles can be decoded into basecalls in real-time. Read lengths produced by Nanopore sequencing are only limited by the size of the library, which enables obtaining DNA reads spanning up to 2Mb. Nanopore technology can also sequence native DNA, allowing for the detection of base modifications. The PromethION sequencer is the highest throughput sequencer from ONT. Each PromethION flow cell is advertised to produce up to 180Gb[1] of data.
Genomic DNA (gDNA) sequencing applications supported by the PromethION include:

- ✓ Structural variation
- ✓ Assembly
- ✓ DNA base modifications
- ✓ Multiplex sequencing

Technical specifications

Advertised[1] flow cell performance for currently available flow cell types:
Table 1 ONT advertised flow cell specifications

Flow cell type	**R9.4.1[2]**
Typical per flow cell[1] DNA	80-100Gbp
1D accuracy	Up to 98%
Q score (consensus)	~ Q42
Recommended for	Higher output Consistent performance

[1]Performance may vary greatly depending on input material (DNA/cDNA/RNA), sample quality, library size, library purity, preparation methods. For details, consult the Nanopore website. **Please note:** these specifications are for guidance purposes only. The BRF does not guarantee these performance metrics will be met.
[2] ONT offers a higher accuracy flow cell (R10.3), which is a less stablished product than the R9.4 flow cell. The BRF will only offer these when flow cells are pre-purchased by the client.

Genomic DNA processing at the BRF

The ONT_gDNA service offers sample QC, library preparation with standard ligation methods and sequencing. Clients must supply extracted HMW DNA. This workflow does not apply for ultra-long DNA sequencing.

The workflow for this service is outlined below:

Sample submission

Please complete the gDNA-ONT sample submission form in our online system BRF (see our website for details on electronic submissions). The sample submission form asks for:

- ✓ Amount and quality of input material
- ✓ Research goal of your experiment
- ✓ Any QC data you have on your library (e.g. concentration readings and size traces)
- ✓ Additional procedures such as shearing, size selection, multiplexing.
- ✓ Acceptance of terms and conditions of service.

Once your samples have been submitted electronically, bring your samples in person or by courier to the BRF.
Ensure samples are:

- ✓ Labelled appropriately: Write the *Sample Submission Id* and the *Request Id* from the electronic form. Label cap and side of the tube.
- ✓ Transported in conditions that preserve sample integrity (e.g tubes are in good condition (low retention preferred), properly sealed, transported cold, proper padding to avoid crushing, etc)

Initial quality control checks (Initial QC)

Once received, we will check your sample against the following:

Table 2 gDNA QC parameters

QC Parameter	Specification	Measurement method
Appearance	Correct labelling and container integrity Clear viscous liquid with no visible particles of precipitates.	Visual check
Volume check	At least 50 ul per flow cell	Pipetting
Average fragment size	50Kb mean fragment size	Pulse field electrophoresis (Femto Pulse)
Amount of material[3]	5-10ug (100-300fmol) per flow cell	Qubit or equivalent
Purity	Absorbance ratios: 260/280: 1.8-2.0, 260/230: 2.0-2.2	Nanodrop

[3] For maximum flow cell output, multiple flushes and reloads may be necessary. In this case, 2-3 library preparations are advisable and higher amounts of DNA may be required.

Project acceptance:

Upon QC completion, we will get in touch to discuss QC results and how to best proceed with sample processing. Your project will be marked as ***accepted*** when:

- ✓ Your samples meet QC parameters* as per table 2. Genomic DNA samples that fail QC are likely to perform poorly on a Nanopore flow cell.
- ✓ There is an agreement between you and the BRF on what services will be included (e.g. size selection, multiplexing, etc) and their price.
- ✓ You have arranged enough electronic storage to transfer your data files to. We recommend ~2TB of disk per flow cell sequenced.

If samples in your project do not meet QC parameters, you can:

- ✓ Withdraw the project: you will only be charged the for the QC fees. Reagents/consumables purchased for your project may also be charged at the BRF’s discretion.
- ✓ Resubmit your sample: you can reuse the sample submission form. QC for failed samples will still carry a fee.
- ✓ Proceed with the prep: Recommended only if it is apparent that your project goals can be met despite sample not meeting QC. The PI must provide explicit written authorisation for the project to proceed.

Library preparation and sequencing:

gDNA libraries will be prepared using the ONT ligation methods and protocols (SQK-LSK110 or SQK-LSK109 with native barcoding). We also offer the following additional services:

Table 3 Additional services for gDNA preps

Additional service	Purpose
Sample clean-up with Ampure XP beads	May improve sample purity and ligation performance during prep.
Shearing using the Covaris gTube	To maximise flow cell output. N50s will be reduced
Short read elimination using the Circulomics SRE	Will improve N50.
Size selection using the Sage Blue Pippin	Will improve N50 and may reduce contaminants and impurities in the sample.
Barcoding using the EXP-NBD104, EXP-NBD114 kits	Will allow multiplexing and or running consecutive samples in the same flowcell
Flow cell wash and re-use	Necessary to maximise flow cell output for HMW DNA and/or blocky samples. Allows multiple uses of a flow cell.

Whenever reasonable, we will test the library using a Flongle or MinION flowcell to rule out library performance issues prior loading the PromethION flow cell.

Your library will be sequenced in the PromethION using ONT standard methods.

Data collection:

We will perform real-time base calling for each flow cell as per ONT recommendations. Customers can choose to receive fast5 or fast5 *and* fastQ files. Boutique basecalling/analysis methods are not offered as part of the standard service.
Basecalled data provided by the BRF uses the pre-installed basecaller in the instrument control software, which is often a few versions behind the latest ONT release. It is highly encouraged that our clients retain fast5s files as basecallers are continually improved and raw files can be re-analysed.

Important: ONT sequencing is a fast growing and developing technology, and as such, hardware/reagents/consumables occasionally may not perform as advertised. ONT is also extremely sensitive to the quality and purity of the input material you provide. At the BRF, we will strive to give you the best possible output and data quality as per your project goals, but due to the above mentioned limitations, we are unable to guarantee data outputs or read quality for your runs.

Data management and delivery

PromethION run outputs often exceed 1TB, which hinders out ability to provide long term data storage/back up for this service. To ensure a smooth and error-free data hand-over, we ask that data storage arrangements are made prior your sample is due to be sequenced.

- For clients with NCI accounts (ANU only): Transfer of data to an NCI account is the fastest and most reliable way to deliver data to you. If you are an ANU client, it is highly recommended you secure enough storage though an NCI project (nci.org.au). Please detail your NCI project and username in the electronic sample submission.
- For external clients: We will generate a transfer link using CloudStor or out in hour data sender system. It is the client's responsibility to download the data and check its integrity.
- If you work in collaboration with an ANU bioinformatics team (e.g. the ABC or EMBL Australia), you are encouraged to negotiate NCI storage through your collaborators.

We require acknowledgement that you have received your data within 5 working days after hand over. Data will be archived by the BRF for a maximum of 3 months, and a you may be charged a weekly storage fee until you confirm your data has been received successfully. We reserve the right to permanently delete data after we receive confirmation of reception and/or after 3 months from the date of data hand over.

Storing and backing up your data is solely your responsibility. We strongly advise you back up your data securely upon reception. Options include:

- ✓ Mass Data Storage System (MSDD) through NCI can be suitable for data archiving.
- ✓ Cloud services such as AARNET CloudStor or AWS
- ✓ Online public data repositories (data can be under embargo for a period of time):
    - The Sequence Read Archive, SRA
    - European Nucleotide Archive, ENA
- ✓ Controlled-access data repositories:
    - European Genome-phenome Archive, EGA
    - Database of Genotype and Phenotype dbGAP
- ✓ *Note*: Storing data in hard drives is discouraged as data transfer is slow and error prone.