

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael



# Oxford Nanopore Technologies Sequencing for Tuberculosis

## Handbook



EDCTP



ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

## Contents

1	Introduction .....	3
1.1	Abbreviations .....	3
2	Culturing <i>M. tuberculosis</i> .....	4
2.1	Purpose .....	4
2.2	Principle .....	4
2.3	A note on growth times for greatest bacterial numbers and therefore optimal DNA extraction yield .....	5
3	DNA Extraction .....	5
3.1	CTAB method .....	5
3.1.1	Purpose .....	5
3.1.2	Principle .....	5
3.1.3	Obtaining the required quantity and MW of DNA .....	6
3.1.4	Precipitation CTAB DNA extraction .....	6
3.1.5	Spin column CTAB DNA extraction .....	9
3.1.6	Further options for improving your DNA concentration and quality .....	11
4	DNA Quality Analysis .....	12
4.1	DNA input requirements .....	12
4.1.1	Concentration .....	12
4.1.2	Quality .....	12
4.1.3	Integrity .....	13
4.2	Quality Control of the DNA Extraction .....	14
4.2.1	Estimation of DNA Concentration Using the Qubit Fluorometer .....	14
4.2.2	Estimating concentration and measuring the purity of DNA extracts using Nanodrop15 .....	16
4.2.3	DNA size analysis using the TapeStation 2200 .....	16
4.2.4	DNA size analysis using Bioanalyzer .....	17
4.2.5	DNA size analysis using gel electrophoresis .....	18
5	Sequencing .....	19
5.1	Purpose .....	19
5.2	Principle .....	19
5.2.1	Anatomy of a flow cell .....	19
5.2.2	Rapid Barcoding kit 96 (SQK-RBK110.96) .....	20
5.2.3	Using ONT to sequence MTB .....	21
5.3	Procedure .....	21
5.3.1	Equipment, consumables and reagents .....	21
5.3.2	Library preparation .....	23
5.3.3	Flow cell quality control (QC) .....	23
5.3.4	Sequencing preparation .....	26
5.3.5	Flow cell preparation .....	28
5.3.6	Library preparation .....	28
5.3.7	Starting your run (using MinkNOW software) .....	29
6	Sequencing data analysis .....	35
6.1	Calculating sequencing coverage and depth .....	35
6.2	Completing base-calling if your run is interrupted .....	36
6.2.1	Fast vs HAC vs SUP base-calling .....	36
6.3	Analysis of data – TB-Profiler .....	37
6.3.1	Preparing your data for TB-Profiler .....	37
6.3.2	Using TB-Profiler .....	37
6.3.3	Interpreting TB-Profiler reports .....	38

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

## 1 Introduction

### 1.1 Abbreviations

ASIC	Application-Specific Integrated Circuit
BSL3	Biological safety level 3 (laboratory)
CLI	Command line interface
CPU	Central processing unit
CTAB	Cetyl trimethyl ammonium bromide
DST	Drug sensitivity testing
FB	Flush buffer
FLT	Flush tether
GPU	Graphical processing unit
GUI	Graphical user interface
HAC	High accuracy basecalling
LB	Loading beads
LJ	Löwenstein–Jensen (media)
MDR	Multi drug resistant
MSC	Microbiological safety cabinet
MTB	<i>Mycobacterium tuberculosis</i>
NFW	Nuclease free water
nt	Nucleotide
NTM	Non-tuberculous <i>Mycobacteria</i>
ONT	Oxford Nanopore Technologies
PPE	Personal Protective Equipment
QC	Quality control
RAP	Rapid adaptor
RT	Room temperature
SDS	Sodium dodecyl sulphate
SNP	Single nucleotide polymorphism
SQB	Sequencing buffer
SUP	Super accuracy basecalling
TBE buffer	Tris/Borate/EDTA Buffer
WCP	Windows Command Prompt
WGS	Whole genome sequencing
XDR	Extensively drug resistant

This handbook is designed to cover a pragmatic whole genome sequencing (WGS) system for *Mycobacterium tuberculosis* for laboratories who want to utilise long read technology or with little/no access to larger scale Illumina sequencing platforms. It is designed to identify lineages (which can be of both clinical and epidemiological use), and antimicrobial resistances (e.g. for hard-to-treat cases, or to ensure early treatment is appropriate).

*M. tuberculosis* (MTB) drug resistance can be determined phenotypically, by carrying out drug sensitivity testing (DST), but (a) this is slow, (b) it is more difficult for some drugs, and (c) there are occasions when a patient's response to a drug differs from what is expected. In addition, rapid genomic testing, e.g. with GeneXpert are not designed to detect all mutations, so will miss some resistant cases. WGS can support the decision-making for initial and continuing drug regimens. WGS can also be used to establish strain relatedness and whether disease after treatment is due to

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

reinfection or relapse, by comparing the initial and post-treatment cultures. Table 1 shows the approximate timeline for MTB diagnostics and where WGS might fit in when using cultured organisms.

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Sputum AFB						
GeneXpert						
Hain test						
Culture						
DST						
WGS						

Table 1. Gantt chart showing an MTB diagnostics timeline. WGS, using cultured isolates, can be undertaken around week 3 after start of treatment, if a patient is not responding as expected.

WGS is the process of determining the complete DNA sequence of an organism's genome. This can be done using a number of platforms (Illumina, PacBio, Ion Torrent, Oxford Nanopore Technologies). A small-scale approach is to use the Oxford Nanopore Technologies (ONT) MinION device. The nanopores are biological structures – protein complexes – that contain a nano-scale hole through which a DNA linear molecule can pass, like a string passing through a small hole. In its devices, ONT passes an ionic current across a membrane where the nanopores are fixed and measures the changes in current as biological molecules (the DNA) pass through the nanopores. The information about the change in current can be used to identify the different DNA nucleotides within the DNA molecule (e.g. A, T, G or C). Computer software is then used to convert this into the DNA sequence. [ONT have a very useful YouTube Channel](#), that shows videos explaining the technology.

ONT have a useful section on their website for anyone wishing to [set up ONT sequencing in their laboratories](#).

An overview of the practicalities of WGS laboratory set up can be found on the [PANDORA-ID-NET TGHN sequencing set up hub page](#).

Please note that MTB culturing must be done in a biological safety level 3 (BSL3) laboratory, as per the [WHO's Laboratory Biosafety Manual](#). Once the DNA has been heat killed (and extracted, depending on local biosafety rules), it can be quantified and sequenced outside of the BSL3 facilities.

## 2 Culturing *M. tuberculosis*

### 2.1 Purpose

Oxford Nanopore (ONT) sequencing, unlike Illumina, can utilise native DNA in most of their kits (i.e. there are no PCR amplification steps). This is because it's difficult to successfully amplify the long strands needed for ONT. Whilst there are some papers that have used straight-from-sputum DNA extractions for MTB, the minimum input for the rapid barcoding kit (SBK110.96) is 50 ng (which is around 5 ng/ul, when the input volume is taken into consideration). Unless the patient has a very high number of bacteria in their sputum, it is extremely unlikely DNA extraction from sputum will yield this amount. Therefore, culturing the bacteria is necessary to obtain the concentration of DNA needed.

### 2.2 Principle

The *Mycobacteria* can be cultured in a number of different ways, but it is important to note that it is vital to do all culturing in the appropriate biosafety laboratory (some Mycobacterial species, such as *M. tuberculosis*, are classed as a hazard group 3 (HG3) organism, whilst others are classed as hazard

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

group 2). You can read more about [biosafety levels on our health and safety page](#) (whilst the page relates to SARS-CoV-2, the same biosafety rules will apply to all HG3 organisms).

There are a number of methods for TB culture, including LJ slopes, 7H9 (liquid) or 7H11 (solid) medias and the BD's automated system Mycobacterial Growth Indicator Tubes (MGIT). [More about the BACTEC MGIT system can be found here.](#)

Below we have outlined two different methods for CTAB DNA extraction, the first utilising DNA precipitation, the second using spin columns.

### 2.3 A note on growth times for greatest bacterial numbers and therefore optimal DNA extraction yield

We have found that growing bacteria for approximately 2-3 weeks after they have flagged in a MGIT tube provides the greatest DNA yield. However, whilst liquid culture provides an appropriate yield using the precipitation CTAB method, we have found that solid culture (and therefore a greater mass of bacteria) are needed for the spin-column CTAB protocol.

## 3 DNA Extraction

### 3.1 CTAB method

#### 3.1.1 Purpose

The methods described here is designed to yield microgram quantities of DNA suitable for genotyping. Note that while high molecular weight DNA is not essential for Illumina sequencing (which utilises very short reads of ~100 bp, DNA for direct ONT sequencing (as used in this protocol) should ideally be ~30kb or above size.

#### 3.1.2 Principle

Although DNA can be extracted from MTB bacilli by a variety of methods, with a range of complexity, the method described here yields high quality, large fragment DNA from a colony scrape off solid media, or from a liquid culture such as a MGIT tube. Using a combination of enzymatic digestion and organic partition, a loop of colonies picked from an LJ slope yield 0.1-50 µg of DNA. Following resuspension and heat-killing of the colonies, bacteria are digested first with lysozyme to breakdown the cell wall then with proteinase K, which has further action on the cell wall but, importantly, digests any enzymes released by the lysed bacterium, including DNases.

MTB has a lipid-rich cell wall, and so two rounds of detergent are used, first Sodium dodecyl sulphate (SDS) and then Cetyl trimethylammonium bromide (CTAB). These detergents have action on molecules with different charges thus affecting different cell wall components. EDTA is used to chelate  $Mg^{+2}$  and  $Ca^{+2}$  ions, inhibiting DNase activity; similarly, high salt concentrations inhibit DNA-enzyme binding. Finally, organic solvents are used to partition the DNA to an aqueous phase, leaving lipids and proteins in the organic phase. The aqueous phase is then concentrated using isopropanol, this concentrates the DNA and removes excess salt. Isopropanol is used in preference to ethanol as a lower volume for precipitation can be used (1:1 rather than 2:1). Yeast tRNA is used during the alcohol precipitation step as it is an effective co-precipitant to aid in recovery of small amounts of nucleic acids and improve the DNA yield.

Vigorous mixing by Vortex is not recommended at any stage of the extraction as this causes DNA shearing.

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

### 3.1.3 Obtaining the required quantity and MW of DNA

Due to the properties of MTB described in section 3.1.2, it can be difficult to obtain the desired quantity of DNA (~5-44 ng/ $\mu$ l) needed for ONT sequencing. Cultures can be obtained from either MGITs or LJ slopes and the highest yield of DNA was obtained approximately 2 weeks after MGIT flagging. Culturing for DNA extraction using Middlebrook 7H11 slopes is not recommended as it does not yield as high MW as MGIT or LJ slopes.

Typically from a loop of cells we obtain 10 $\mu$ l of DNA at 1-500ng/ $\mu$ l (the variation may reflect the strain, medium, or age of cells). If a single extraction from a loopful of cells, or of 400  $\mu$ l of liquid culture does not provide enough DNA (~53.3 ng/ $\mu$ l), then it is recommended that multiple extractions per isolate be done, then the eluted DNA combined. Keeping the elution volume low (<15  $\mu$ l) will also help to create high concentration DNA. If there is enough DNA (e.g. ~400 ng), but it is too diluted, then ethanol precipitation is recommended. Ethanol precipitation is beneficial for removal of salts etc. to improve DNA purity, which will improve sequencing efficiency and flow cell pore longevity.

### 3.1.4 Precipitation CTAB DNA extraction

#### 3.1.4.1 Procedure

##### 3.1.4.1.1 Equipment, consumables and reagents

Below is the list of equipment, consumables and reagents needed for this procedure. We have provided some example product codes, but alternatives may be used.

Equipment	Product Number	Manufacturer
Microfuge (e.g. Eppendorf max speed 15,000 rpm)	15881635	Fisher Scientific
Centrifuge (for spinning down liquid cultures)		
1000 $\mu$ l pipette and aerosol resistant tips		
200 $\mu$ l pipette and aerosol resistant tips		
20 $\mu$ l pipette and aerosol resistant tips		
10 $\mu$ l pipette and aerosol resistant tips		
Vortex (e.g. Fisher mini vortexer)	16343196	Fisher Scientific
-20°C freezer		
4°C refrigerator		
Microbiological safety cabinet (BSL3)		
Water bath or heating blocks (80-95°C, 60-65°C and 37°C)		

Consumables	Product Number	Manufacturer
2 ml screw-cap tubes with rubber 'O' ring seal	11519934	Fisher Scientific
Sterile DNAase-free 1.5 ml tubes	S1615-5510	Star Labs
Empty FastPrep 2ml lysing matrix tubes	115076200	MPBio
4mm glass beads	116914801	MPBio
FastPrep lysing tube lids	115067002-CF	MPBio
Plastic discard containers		
10 $\mu$ l loops (if solid cultures are used)		

Reagents	Product Number	Manufacturer
Nuclease-Free H <sub>2</sub> O	W4502-10x50ml	Sigma Aldrich

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

Ethanol, absolute	16606002	Fisher Scientific
Tris-EDTA (TE) buffer	93283	Sigma
Sodium dodecyl sulphate (SDS)	11667289001	Sigma
Lysozyme	L1667	Sigma
Proteinase K	P5665	Sigma
Cetyl trimethyl ammonium bromide (CTAB)	H6269-100G	Sigma
5M NaCl	71386	Sigma
Chloroform: Isoamylalcohol	C0549-1PT	Sigma
Yeast tRNA	AM7119	Thermo Fisher
Isopropanol/2-propanol	I9516	Sigma

### 3.1.4.1.2 Preparation of extraction reagents

Reagent	Working concentration	Preparation/storage conditions
Lysozyme solution	10 mg/ml	Store in small aliquots at -20°C. Use a new aliquot each time, do not re-freeze. Any remaining reagent should be discarded.
SDS	10%	Add distilled water to 10 g of SDS to make up 100 ml of total solution. Dissolve by heating at 65 °C for 20 min. Do not autoclave. Store at room temperature for no longer than 1 month.
Proteinase K	20 mg/ml	Store in small aliquots at -20 °C. Use a new aliquot each time, do not re-freeze. Any remaining reagent should be discarded. Unless stated otherwise on manufacturing guidelines.
NaCl	5M	29.2 g NaCl/100 ml distilled water. Heat to 65°C and mix until dissolved (this may take hours). Autoclave at 121°C for 20 minutes. Store at room temperature for no longer than 1 year. This can be purchased commercially (Sigma Aldrich, S5150-1L).
CTAB/NaCl	(10% CTAB in 0.7 M NaCl)	Dissolve 4.1 g NaCl in 80 ml distilled water. While stirring, add 10 g CTAB. If necessary, heat solution to 65 °C Adjust the volume to 100 ml with distilled water and autoclave at 121°C for 20 minutes. Store at room temperature for no longer than 6 months.
Chloroform/ isoamylalcohol	(24:1)	If purchased separately, mix 1 part of isoamylalcohol with 24 parts of chloroform in a fume hood. Mix thoroughly by shaking vigorously for 5 seconds. Store in cool, dark, ventilated place, use within 6 months or by expiry date indicated. This can be purchased commercially (Sigma Aldrich, C0549-1QT).
Ethanol	70%	70 ml 100% ethanol in 30 ml distilled water, store at -20°C.
Yeast tRNA	10 mg/ml	Store in small aliquots at -20°C. Use a new aliquot each time, do not refreeze. Any remaining reagent should be discarded. (Catalog number: AM7119)
Ultrapure molecular grade DNase/RNase free water	n/a	Prepare 50 ml aliquots (or sufficient for the sample number in the batch) inside a Class 2 BSC to avoid frequent opening of stock bottle. Record the date the aliquot was prepared in the tube and use a new aliquot for each batch

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

Tris-EDTA (TE) buffer	1x	100x TE buffer are purchased commercially (Sigma Aldrich T9285) and diluted to 1x in molecular grade DNase/RNase free water (e.g. 500 $\mu$ l of 100x TE buffer and 49.5 ml of water).
-----------------------	----	--

### 3.1.4.1.3 Process

We recommend that all steps until after the addition of chloroform/isoamylalcohol (step 15) be carried out in an appropriate microbiological safety cabinet (MSC) inside a BSL3 Laboratory, as the heat kill temperature was not validated to kill all bacteria. Heat killing at greater than 80°C can damage the DNA and long strands are required for ONT sequencing. Use tubes with 'o' rings in the lid to no spillage or aerosolization (e.g. from opening a flip lid microcentrifuge tube).

1. Switch on the heating block (set at 80°C).
2. Label sufficient 2.0 ml clear screw-capped tubes containing 'O' rings. It is important to use clear tubes for ease of collection of aqueous phase after phase separation from chloroform/isoamyl alcohol (Step 18). Label both lid and side, as some of the solvents could delete the writing.
3. Aliquot 400  $\mu$ L volumes of 1x Tris-EDTA (TE) buffer into the tubes using aerosol resistant tips.
4. From solid media (LJ/7H11 slopes) with good growth, take all organisms using a 10  $\mu$ l loop and emulsify them in the appropriate tubes containing the TE buffer taking care not to create splashes or aerosols.
5. Pulse down the tubes in the microfuge using the aerosol-containing rotor for 5 seconds to ensure that all organisms are at the bottom of the tube and unload the rotor in the BSC.
6. Place the tubes in a suitable rack and heat-kill in the heating block at 80°C for 30 minutes.
7. Sonicate in a water bath for 15 minutes at room temperature.
8. Pulse down the tubes, as above.
9. Place the tubes in a suitable rack and add 50  $\mu$ L 10 mg/ml lysozyme and vortex briefly. Incubate at 37°C in the incubator or heating block overnight (if overnight is not possible at least one hour is required).
10. Switch on 60°C heating block. Pre-warm the CTAB/NaCl to 60°C. Ensure a sufficient volume of Isopropanol (450  $\mu$ L per sample) and 70% Ethanol (1ml per sample) are chilled at -20°C. Take aliquot of tRNA out of the -20°C freezer, vortex for a few seconds when fully thawed, and place on ice.
11. Add 70  $\mu$ L 10% SDS and 5  $\mu$ L 20 mg/ml proteinase K. Vortex briefly and incubate at 60°C for 10 minutes.
12. Add 100  $\mu$ L 5M NaCl to each tube.
13. Add 100  $\mu$ L CTAB/NaCl (pre-warmed to 60°C). Vortex briefly and incubate at 60°C for 10 minutes.
14. Ensure that the chloroform/isoamylalcohol (24:1 v/v) is well mixed and add 750  $\mu$ L to each tube inside a ventilated cabinet. Vortex.
15. Wipe tubes externally with disinfectant (e.g. Tristel, but must be validated to kill MTB) and place inside a zip bag.

***Tubes can now be removed from the BSL3 Laboratory.***

16. Set the heating block temperature to 30°C.
17. Microfuge at 10,000 x g for 5 minutes.
18. Label lid and side of sterile DNAase-free 1.5ml microfuge tubes and transfer the aqueous supernatants into these. Take care not to disturb the interface.
19. Add 3  $\mu$ L tRNA to each tube
20. Aliquot 450  $\mu$ L of ice-cold isopropanol to each tube. Vortex briefly and place at -20°C for at least 30 minutes. Keep isopropanol and ethanol (step 22) on ice while in use.
21. Microfuge at 10,000 x g for 15 minutes at room temperature (it can be helpful to draw a line on

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

the outside of the tube, where the pellet will form, as the pellet can be hard to see).

22. Remove the supernatants and wash the pellets with 1ml 70% ethanol at -20°C. Invert gently.
23. Microfuge at 10,000 x *g* for 5 minutes at room temperature. Remove and discard as much of the ethanol as possible. Place the tubes with open lids in a heating block at 30°C to allow the pellets to air-dry (at least 15 min).
24. Rehydrate the pellets in 10-15 µL molecular grade water (preferred) or TE buffer with a maximum concentration of 0.1 mM EDTA overnight at 4°C (or 1 hour at 65 °C).

### 3.1.5 Spin column CTAB DNA extraction

#### 3.1.5.1 Procedure

##### 3.1.5.1.1 Equipment, consumables and reagents

Below is the list of equipment, consumables and reagents needed for this procedure. We have provided some example product codes, but alternatives may be used.

Equipment	Product Number	Manufacturer
Microfuge (e.g. Eppendorf max speed 15,000 rpm)	15881635	Fisher Scientific
Thermomixer (e.g. Eppendorf F2.0 Model)	15356551	Fisher Scientific
<b>Optional:</b> Hula mixer	10548425	Fisher Scientific
Centrifuge		
1000 µL pipette and aerosol resistant tips		
200 µL pipette and aerosol resistant tips		
20 µL pipette and aerosol resistant tips		
10 µL pipette and aerosol resistant tips		
Vortex (e.g. Fisherbrand mini vortexer)	16343196	Fisher Scientific
-20°C freezer		
4°C refrigerator		
Microbiological safety cabinet (BSL3)		
Water bath or heating blocks (80-95°C, 60-65°C and 37°C)		

Consumables	Product Number	Manufacturer
2 ml screw-cap tubes with rubber 'O' ring seal	11519934	Fisher Scientific
Empty FastPrep 2ml lysing matrix tubes	115076200	MPBio
4mm glass beads (e.g. MPBio)	116914801	MPBio
FastPrep lysing tube lids	115067002-CF	MPBio
Plastic discard containers		
10 µL loops (if solid cultures are used)		

Reagents	Product Number	Manufacturer
Nuclease-Free H <sub>2</sub> O	W4502-10x50ml	Sigma Aldrich
Ethanol, absolute	16606002	Fisher Scientific
PBS pH 7.4	11503387	Fisher Scientific
TE buffer	8890-100ML	Sigma
Lysozyme	L1667	Sigma
CTAB buffer:		

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

CTAB	H6269-100G	Sigma
EDTA	11568896	Fisher Scientific
NaCl	S6546	Sigma
Trizma hydrochloride solution	T2819	Sigma
5M NaCl	71386	Sigma
NEB Monarch Genomic DNA purification kit	T3010S	New England Biolabs

### 3.1.5.1.2 Preparation of extraction reagents

#### 3.1.5.1.3 Process

We recommend that all steps be carried out in an appropriate microbiological safety cabinet (MSC) inside a BSL3 Laboratory, as the heat kill temperature was not validated to kill all bacteria. Heat killing at greater than 80°C can damage the DNA and long strands are required for ONT sequencing. Use tubes with 'o' rings in the lid to no spillage or aerosolization (e.g. from opening a flip lid microcentrifuge tube).

Note that for the spin column CTAB method, the high salt concentrations used to purify the DNA and reduce potential inhibitors mean a generally lower yield than the precipitation CTAB. We therefore recommend using solid agar to grow MTB. You need a lot of culture to get enough DNA from this protocol.

1. In the MPBio tube, aliquot 500 µl PBS
2. Loop up ¼ to ½ of the cultures on the plate (note: if the cultures are very dry, dip the loop in the PBS to moisten them, they will come off more easily). Add the cultures to the PBS in the tubes.
3. Incubate the sample at 80°C for 30 min (to heat kill the organism)
4. Centrifuge at 16,000 x g for 5 min to pellet the culture
5. Remove the supernatant and resuspend in 100 µl TE buffer (10 mM Tris pH8 and 1mM EDTA)
6. Add 2 x 4mm glass beads to the tube (it's easier to decant beads into a 15 ml falcon lid and then use the fat end of a pipette tip to pick them up)
7. Vortex for 45 sec (ensure complete homogenisation)
8. Add 10 µl lysozyme (10 mg/ml) and mix by pulse vortexing
9. Incubate at 37°C for 15 min at 800 RPM
10. Add 10 µl proteinase K and 3 µl RNase A (provided in NEB Monarch Genomic Purification kit) then vortex to mix
11. Add 200 µl CTAB buffer (2% CTAB, 40mM EDTA, 1.4M NaCl and 100mM Tris pH8.0)
12. Transfer the sample (110 µl) into a 1.5 ml screw cap tube (use the CTAB to loosen culture from the sides of the MPBio tube) and vortex to mix
13. Incubate at 56°C for 30 min at 1000 RPM
14. Add 100 µl 5M NaCl and centrifuge at 17,000 x g for 5 min and transfer ~200-400 µl of supernatant to a new screw cap tube (it is crucial to avoid the pellet)

#### Genomic DNA extraction

15. Manual extraction recommendation: NEB Monarch Genomic Purification Kit, part 2: Genomic DNA Binding and Elution. Manufacturer instructions followed with the following alterations:
  - a. Add 400 µl gDNA binding buffer and mix thoroughly by pulse-vortexing for 5-10 secs (thorough mixing is essential)

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

- b. Transfer the lysate/binding buffer mix (~600 µl) to a gDNA purification column pre-inserted into a collection tube, without touching the upper column area
- c. Close the cap and centrifuge: first for 3 min at 1,000 x g to bind the gDNA (no need to empty the collection tube or remove from centrifuge), and then for 1 min at maximum speed (>12,000 x g) to clear the membrane. Discard the flow-through and the collection tube
- d. Transfer column to a new collection tube and add 500 µl gDNA wash buffer. Close the cap and centrifuge immediately for 1 min at maximum speed, then discard the flow through and collection tube.
- e. Transfer the column into a new collection tube. Add 500 µl gDNA wash buffer and close the cap. Centrifuge for 1 min at maximum speed
- f. Discard the collection tube and flow through then spin again for 2 mins at maximum speed to completely remove all the remaining wash buffer
- g. Place the gDNA purification column in a sterile DNase-free 1.5 ml microfuge tube. Add **50-100 µl** preheated (60°C) gDNA elution buffer, close the cap and incubate at RT for 1 min
- h. Centrifuge for 1 min at maximum speed (>12,000 x g) to elute the gDNA

### 3.1.6 Further options for improving your DNA concentration and quality

Whilst we found that the following protocols did not improve sequencing data yield, they may be useful, for instance if your samples are high volume but lower concentration than needed.

#### 3.1.6.1 Ethanol precipitation

Ethanol precipitation will increase the concentration of your DNA, as well as helping to remove impurities. A more [detailed protocol for ethanol precipitation can be found here](#).

#### Reagents and equipment needed:

- Sodium or ammonium acetate
- 100% ethanol
- Freshly made 70% ethanol
- Microcentrifuge tubes capable of refrigeration
- Nuclease free water (NFW)

#### Protocol:

1. Adjust the salt concentration of your sample by adding 1/10 volume of 3 M sodium acetate (pH 5.2) (0.3 M final concentration) - you can use ammonium acetate (2 M final concentration) if you don't have Na-acetate
2. Add 2-3 volumes of ice cold 100% ethanol
3. Incubate at -20°C for at least 1 h (leaving overnight is recommended for best results)
4. Centrifuge at maximum speed at 4°C centrifuge for 20 minutes
5. Wash pellet with ice cold 70% ethanol by centrifuging (4°C) at maximum speed again for 20 minutes
6. Redissolve DNA in 10-15 µl NFW

#### 3.1.6.2 AMPure bead wash

Washing the extracted DNA using magnetic beads also helps to purify the DNA and ensure nothing inhibits the library preparation. It will also remove smaller fragments of DNA to help ensure only longer strands get sequenced.

#### Reagents and equipment needed:

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

- AMPure XP beads (Beckman Coulter Life Sciences, (Product No: A63880))
- 1.5 ml LoBind tubes
- Vortex
- Microcentrifuge
- Magnetic rack
- Nuclease free water (NFW)
- Freshly prepared 70% ethanol

#### Protocol:

1. Add 1X (by volume) of AMPure beads and your DNA to a LoBind 1.5 ml tube
2. Gently pipette up and down until the solution has visually mixed
3. Invert and lightly shake the tube for 5 minutes
4. Pulse spin down the tube and place the tube on the magnetic rack for 5 minutes to separate the beads from the solution
5. Slowly aspirate the cleared solution and discard
6. Slowly add 500  $\mu$ L of 70% ethanol without disturbing the beads. Wait 30 seconds and aspirate out
7. Repeat the 70% ethanol wash for a total of 2 washes
8. Pipette off residual ethanol as much as possible and let the tube air dry on the magnetic rack for 10 minutes with the lid open. Visually inspect it at 8 minutes to see if it has dried. Don't let the beads over-dry to the point of cracking
9. Elute in 12  $\mu$ L of NFW. Remove the tube from the magnetic rack, mix by pipetting up and down slowly 6 times. Let it stand for 2 minutes off the magnetic rack
10. Return tubes to the magnetic rack for 2 minutes and pipette out 12  $\mu$ L of the clear solution which contain the purified DNA, avoid carry over of the beads

## 4 DNA Quality Analysis

The concentration and integrity of the extracted DNA is important for WGS. There are a number of methods to do this, but Qubit is recommended for determining quantity. The use of Nanodrop or other spectrophotometric methods for estimating nucleic acid concentrations is not recommended, as they can be unreliable; in addition, we add tRNA as a carrier, which is detected using such methods and will confound the results. For checking the integrity (average molecular weight, and an estimate of how much degradation there is, the Agilent TapeStation is recommended, but Agilent Bioanalyzer or agarose gel electrophoresis is also suitable.

### 4.1 DNA input requirements

#### 4.1.1 Concentration

Overloading or underloading DNA onto the ONT flow cell can result in suboptimal sequencing. The ONT Rapid Barcoding kit (SQK-RBK110.96) used in this pipeline requires a total of 50-100 ng input DNA, in a maximum of 9  $\mu$ L. More or less can impact the efficiency of the sequencing and is likely to result in lower yield. It is also important, when multiplexing (e.g. using multiple barcoded samples), that the input quantity for each sample is as even as possible, to avoid uneven reading across the samples. The Rapid Barcoding Kit requires an input of 9  $\mu$ L for each sample, which means obtaining at 5.5 - 11 ng/ $\mu$ L is necessary.

#### 4.1.2 Quality

Chemical contaminants e.g., detergents, denaturants, chelating agents and high concentration of salts which could be carried over from the DNA method of extraction should be avoided as it can affect the efficiency of enzymatic steps downstream the process. Also, contaminants like single stranded DNA,

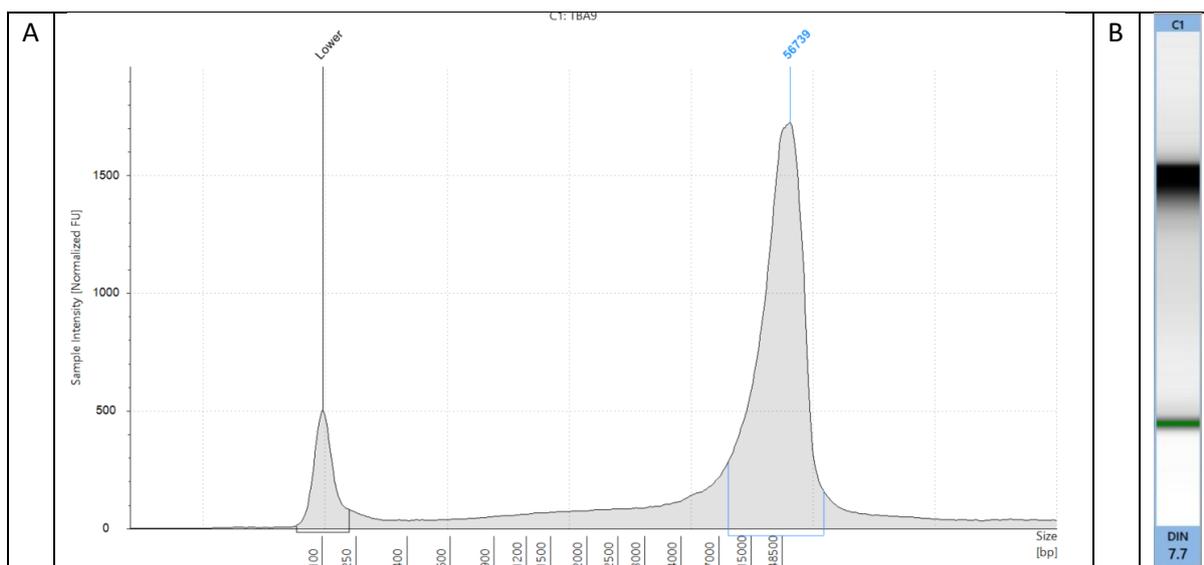
ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

RNA and proteins from the sample in your DNA extracts reduce the efficiency of steps in the library preparation.

### 4.1.3 Integrity

ONT sequencing utilises long read technology, and sequencing is much more efficient with longer strands of nucleic acid. The ONT Rapid Barcoding kit (SQK-RBK110.96) used in this pipeline requires an average DNA size of at least 30 kb for each sample, however, for the Mycobacteria, aiming for 20-30 kb may be better, as the complex, repetitive structure of mycobacterial genomes seem to block pores faster than other genomes, such as the gram negatives. When multiplexing, it is important to ensure the DNA sizes are all similar to ensure even sequencing across all samples. If there are large differences in size between samples (e.g. one is 30 kb and one is 50 kb), normalising by [calculating the fmoI](#) for each sample can help to ensure even sequencing.

gDNA should have a high size band (>12kb) with little/no evidence of degradation (smearing). Degraded genomic DNA may result in over-fragmentation of DNA to below the desired size range and poor-quality libraries. A good library should be free of any smaller or larger products than the target size for the library preparation technique. Figure 1 shows an example of good (acceptable quality) DNA and poor (not acceptable quality) DNA.



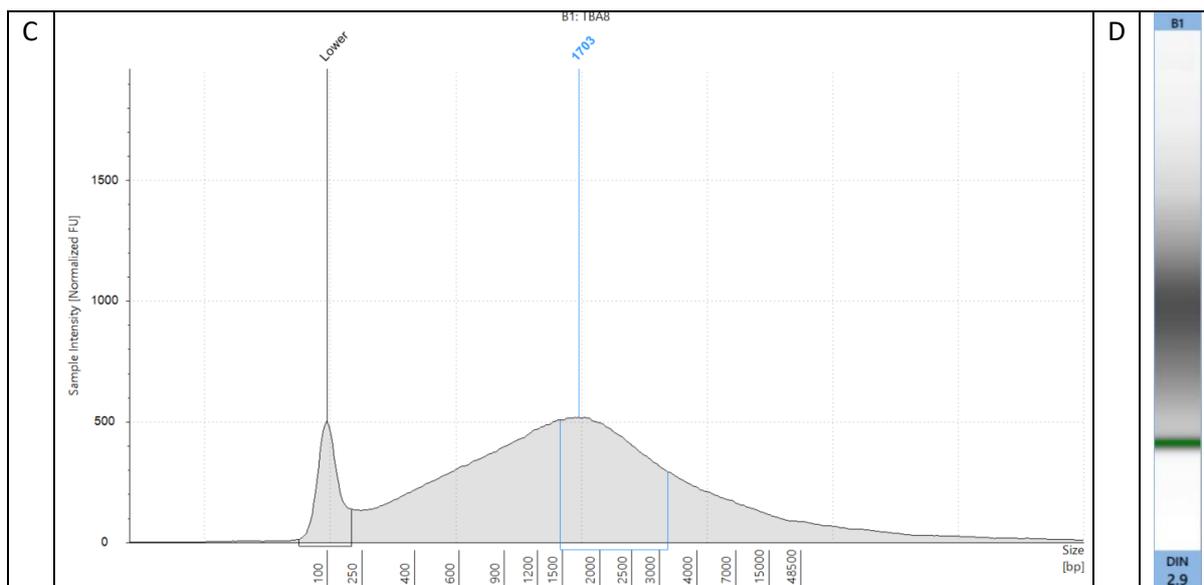


Figure 1. Images from the Agilent TapeStation showing the peaks and gel image for high quality and low-quality DNA extractions. A) (peak) and B) (gel image) shows high quality DNA that would be suitable for use in this MTB ONT sequencing pipeline. It is high quality (note the single peak), high size (56,739 kb) and high concentration (97.8 ng/ $\mu$ l), which were all greater than the minimum ONT input requirements. C) (peak) and D) (gel image) show poor quality DNA, that would not be suitable for use in this MTB ONT sequencing pipeline. It is low quality (note the 'smear' rather than a specific peak, suggesting large amounts of DNA shearing) and low size (1,703 kb). It did meet the concentration requirements (98.8 ng/ $\mu$ l), but as it is not high size, it would not produce efficient sequencing, as ONT requires long reads (ideally 10-100's of kb) to work optimally.

## 4.2 Quality Control of the DNA Extraction

The quantity, quality and integrity of the extracted DNA must be measured, using appropriate methods detailed below to ensure it is sufficient for the success of WGS analysis. The recommended method for DNA quantification is the Qubit, for purity check of the DNA extract is the Nanodrop and for DNA integrity check the TapeStation is preferred but gel electrophoresis could be used to identify the range and distribution of the sizes of the DNA fragments (whilst the TapeStation and Nanodrop also quantify the DNA concentration, the Qubit is considered more accurate).

### 4.2.1 Estimation of DNA Concentration Using the Qubit Fluorometer

The Qubit dsDNA Assay Kits are designed for accurate DNA quantification and are highly selective for double-stranded DNA (dsDNA) over RNA.

#### 4.2.1.1 Equipment/Reagents

- Qubit<sup>®</sup> dsDNA Assay HS Kit. This is the high sensitivity (HS) kit with quantitation range of 0.2-100ng. Note: The broad range (BR) kit (quantitation range of 2-1000ng) can also be used if required.
  - Qubit<sup>®</sup> dsDNA HS Reagent (Component A)
  - Qubit<sup>®</sup> dsDNA HS Buffer (Component B)
  - Qubit<sup>®</sup> dsDNA HS Standard #1 (Component C)
  - Qubit<sup>®</sup> dsDNA HS Standard #2 (Component D)
- Disposable plastic container for mixing the Qubit<sup>®</sup> working solution

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

- Thin-wall, clear, 0.5ml PCR tubes. Acceptable tubes include Qubit® assay tubes (Life Technologies Cat No. Q32856) or Axygen® PCR-05-C tubes (VWR Cat No. 10011-830).
- Qubit Fluorometer

Storage of reagents: The Qubit® dsDNA HS Reagent and Buffer must be stored at room temperature (22-28°C) and the Qubit® DNA standards must be stored at 4°C. The Qubit Reagent must be protected from light. When stored as directed kits are stable for 6 months.

#### 4.2.1.2 Process

1. Set up the required number of 0.5 ml tubes for the samples and standards (2 tubes) and label the lids with the sample name.
2. Prepare Qubit® working solution by diluting the Qubit® dsDNA HS Reagent 1:200 in the Qubit® dsDNA HS Buffer in a clean plastic tube. The final volume in each tube is 200 µL, so prepare enough working solution for all the samples and the two standards. For example, for 8 samples and 2 standards make 2 ml of working solution by diluting 1 µL of reagent into 1990 µL of buffer.
3. Add 190 µL of Qubit® working solution to each of the standard tubes, and then add 10 µL of each Qubit® standard to the appropriate tube and mix by vortex for 2-3 seconds. Be careful not to create bubbles.
4. Add Qubit® working solution to individual sample tubes so that the final volume including the sample is 200 µL. The volume of the sample can be between 1-20 µL, so the volume of working solution can be adjusted accordingly e.g. 1:199 or 2:198 or 10:190
5. Add each sample to the appropriate assay tubes and mix by vortex for 2-3 seconds.
6. Allow all tubes to incubate at room temperature for 2 minutes – the samples are now ready to be read on the fluorometer.
7. On the home screen of the Qubit® Fluorometer, press DNA and then select dsDNA High Sensitivity as the assay type. The standards screen is displayed.
8. Press Yes to read the standards. Insert the tube containing Standard #1 into the sample chamber, close the lid and press Read. When the reading is complete (~3 seconds) remove the standard. Insert the tube containing Standard #2 into the sample chamber, close the lid and press Read. When the reading is completed remove the standard. When the calibration is complete the instrument displays the Sample screen.
9. Insert a sample tube into the sample chamber, close the lid and press Read. When the reading is complete (approx. 3 seconds) remove the tube. The instrument displays the results on the screen. The first value displayed is the concentration in the assay tube.
10. To find out the concentration in the original sample, press Calculate Stock Conc. The Dilution Calculator Screen is displayed. Select the volume of the sample that you added to the assay tube, once selected the Qubit® Fluorometer calculates the original sample concentration using the volume and the measured assay concentration.
11. Press Read Next Sample and repeat steps 9 and 10 for all remaining samples.

NOTE: if the DNA concentration obtained is above the upper limit of the test kit, dilute the sample 1:10 and repeat testing to obtain an absolute concentration value.

#### 4.2.2 Estimating concentration and measuring the purity of DNA extracts using Nanodrop Manual: [Thermo Scientific NanoDrop 2000/2000c Spectrophotometer V1.0 User Manual](#)

NanoDrop 2000 is a spectrophotometer that can be used to quantify the DNA and protein content in a 1-2 µl sample. Nucleic acids (such as DNA) and proteins absorb light at the 260 nm and 280 nm

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

respectively. RNA is a common contaminant in genomic DNA extracts and nanodrop cannot distinguish between DNA and RNA very well, hence less accurate for quantification of DNA for library preparation purposes. It also can be used to measure the purity of DNA extracts through measuring the A260/A280 ratio (recommended values 1.8 – 2) and A260/A230 (recommended value (2.0 – 2.2)).

Process:

1. Launch the NanoDrop 2000 software interface.
2. Raise the sampling arm, apply 2  $\mu$ L blank solution onto the lower pedestal, carefully lower down the sampling arm and initiate blank measurement using the software on the PC.
3. When the measurement is complete, raise the sampling arm and wipe the blank from both the upper and lower pedestals using a clean lint-free laboratory wipe.
4. Apply 2  $\mu$ L sample solution onto the lower pedestal and initiate a spectral measurement.
5. Wipe the sample from both the upper and lower pedestals using a clean lint-free laboratory wipe before you apply the next sample.

#### 4.2.2.1 *Cleaning of the instrument*

##### **Between measurements**

Upon completion of each sample measurement, wipe the sample from both the upper and lower pedestals with a clean, dry, lint-free lab wipe, to prevent sample carryover and avoid residue build-up. It is not necessary to clean with distilled water (dH<sub>2</sub>O) between samples.

##### **Between users**

A final cleaning of both measurement surfaces with dH<sub>2</sub>O is recommended before the first sample measurement and after the last sample measurement.

##### **Additional cleaning**

When a more rigorous cleaning protocol is required (i.e. dried proteins) substitute 0.5M HCl for the dH<sub>2</sub>O in the procedure above. Follow with 3-5  $\mu$ L of dH<sub>2</sub>O.

##### **Decontamination**

A sanitizing solution, such as a 0.5% solution of sodium hypochlorite (1:10 dilution of common commercial bleach solutions – freshly prepared), may be used to ensure that no biological material is present.

#### 4.2.3 *DNA size analysis using the TapeStation 2200*

##### 4.2.3.1 *Equipment/reagents*

- Genomic DNA Ladder (store at 2 – 8°C)
- Genomic DNA Sample Buffer (store at 2 – 8°C)
- Genomic DNA ScreenTape (store at 2 – 8°C. If you run less than 16 lanes, store used ScreenTape device upright at 2 – 8°C for a maximum of 2 weeks. Never freeze).
- Loading tips (5067- 5152 or 5067- 5153)
- Optical Tube 8x Strip (401428)
- Optical Cap 8x Strip (401425) or 96-well Sample Plates (5067- 5150) and 96-well Plate Foil Seal (5067- 5154).
- Vortex mixer
- Volumetric pipette
- Microcentrifuge

##### 4.2.3.2 *Process*

1. Launch the 2200 TapeStation Controller Software.

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

2. Load Genomic DNA ScreenTape device and loading tips into the 2200 TapeStation instrument.
3. Allow reagents to equilibrate at room temperature for 30 min.
4. Vortex mix reagents before use.
5. In the first well in the tube (A1) Prepare ladder by mixing 10  $\mu\text{L}$  Genomic DNA Sample Buffer with 1  $\mu\text{L}$  Genomic DNA Ladder (Use a fresh ladder for each run. No electronic ladder is available for the Genomic DNA assay)
6. Prepare sample(s) by mixing 10  $\mu\text{L}$  Genomic DNA Sample Buffer with 1  $\mu\text{L}$  genomic DNA sample (10 – 100 ng/ $\mu\text{L}$ ) in the wells following ladder (B1, C1, D1 etc.).
7. Spin down, then vortex using IKA vortex and adaptor at 2000 rpm for 1 min.
8. Spin down to position the sample at the bottom of the tube.
9. Load samples into the 2200 TapeStation instrument.
10. Select the required samples on the 2200 TapeStation Controller Software.
11. Click Start.

#### 4.2.4 DNA size analysis using Bioanalyzer

Manual: [Agilent Bioanalyser](#)

- DNA 12000 kits are designed for the sizing and quantitation of double-stranded DNA fragments from 100 to 12000 bp.
- For accurate determination of DNA concentration, the total amount of DNA in sample must be between 0.5–50 ng/ $\mu\text{L}$ . If concentration is excessively high, dilute to 0.5–50 ng/ $\mu\text{L}$ .
- Can analyse 12 samples/chip, chip cannot be reused.

##### 4.2.4.1 Equipment/reagents

- Agilent 2100 Bioanalyzer instrument
- Chip priming station (065-4401)
- IKA vortex mixer
- Agilent DNA 12000 Kit (5067-1508)
- Pipettes (10  $\mu\text{L}$ , 100  $\mu\text{L}$  and 1000  $\mu\text{L}$ ) with compatible tips
- 0.5 mL microcentrifuge tubes for sample preparation
- Microcentrifuge

N.B. Keep all reagents and reagent mixes refrigerated at 4 °C when not in use to avoid poor results caused by reagent decomposition. Protect dye and dye mixtures from light. Remove light covers only when pipetting. Dye decomposes when exposed to light.

##### 4.2.4.2 Process

###### Preparing the Gel-Dye Mix

- 1) Allow DNA dye concentrate (blue) and DNA gel matrix (red) to equilibrate to room temperature for 30 min.
- 2) Vortex DNA dye concentrate (blue) and add 25  $\mu\text{L}$  of the dye to a DNA gel matrix vial (red).
- 3) Vortex solution well and spin down. Transfer to spin filter tube.
- 4) Centrifuge at 1500 x g  $\pm$  20 % for 10 min. Protect solution from light. Store at 4°C.

###### Loading the Gel-Dye Mix

- 1) Allow the gel-dye mix to equilibrate to room temperature for 30 min before use.
- 2) Clean the Bioanalyzer electrodes before using them
- 3) Put a new DNA chip on the chip priming station
- 4) Pipette 9 $\mu\text{L}$  of gel-dye mix in the well marked with a shaded G.

*Note: Pipette in the bottom corner of the wells. Make sure the gel-dye mix is at RT. Wait if not.*

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

- 5) Make sure the plunger is set to 1mL and close the chip priming station.
- 6) Press the plunger until it is held by the clip.
- 7) Wait for exactly 30 seconds and release the clip.
- 8) Wait for 5 seconds and then slowly pull up the plunger back to the 1 ml position.
- 9) Open the chip priming station and pipette 9µL into the other two wells marked G.
- 10) Pipette 5µL of marker (green tube) in all 12 sample and ladder wells.
- 11) Pipette 1µL of DNA ladder (yellow tube) in the ladder well.
- 12) Load 1µL of the DNA sample in the first well and 1µL of MinION buffer in the second.
- 13) In the remaining 10 sample wells, load 1µL of NFW. Do not leave any well empty.
- 14) Put the chip in the Agilent Vortex and vortex for 1 minute.
- 15) Load the chip into the Bioanalyzer and start the run within 5 minutes.

#### 4.2.5 DNA size analysis using gel electrophoresis

Conventional agarose gel electrophoresis employs a static field and can resolve DNA fragments up to 50 kilobases (kb).

##### 4.2.5.1 Equipment/reagents

- Agarose powder
- Tris/Borate/EDTA buffer (TBE buffer)
- GelRed® (Biotium) or Ethidium Bromide
- High MW DNA ladder (Size range: 0.5 kb to 48.5 kb)
- Loading/tracking dye mixture
- Paraffin film or 0.2 mL PCR tubes
- Casting tray/combs
- 250 ml glass flasks
- Balance
- Microwave
- Electrophoresis battery
- Power supply
- UV lamp/camera

##### 4.2.5.2 Process

- 1) Prepare 0.5x Tris/Borate/EDTA buffer (TBE buffer) by diluting 50 mL TBE buffer (often supplied as 10x concentrate) in 950 mL distilled water.
- 2) Prepare 0.7% w/v agarose solution in an appropriate volume of 0.5x TBE: 50 or 100 mL according to the capacity of the casting tray in a flask that is several times larger than the volume required. Do not cover or close the flask/bottle
- 3) Boil the solution for 1-2 minutes to dissolve the agarose, swirl the solution periodically until the solution turns clear, note the flask is hot use appropriate heat insulating gloves.
- 4) Cool down by swirling the flask under running tap water
- 5) Once the gel has reached about 55°C, add the appropriate volume of the DNA staining dye: Ethidium bromide (final concentration 0.3 µg/mL in the agarose gel) or the safer GelRed® (10,000x concentrate in DMSO), swirl gently. Note that ethidium bromide is a mutagenic agent so handle with care and use appropriate PPE.
- 6) Pour the agarose gel/dye mixture slowly in the casting tray with the appropriate comb, avoid creating bubbles, leave the gel to cast ~30 minutes.
- 7) Place the tray with the gel in the electrophoresis tank and fill it with 0.5x TBE buffer (electrophoresis buffer) to the maximum level mark.
- 8) Carefully pull up the comb to create the wells in the agarose gel.

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

- 9) Mix 5µL of each DNA sample with 1µL of sample loading dye/tracking dye mixture (6x) by pipetting up and down on a paraffin film or in PCR tubes.
- 10) Load the samples into the wells, note the order of samples, reserve one well for the DNA Ladder.
- 11) Load high MW DNA Ladder in the gel
- 12) Cover the electrophoresis tank with the lid, make sure that the electrical leads are connected to the battery's electrodes (power supply) correctly.
- 13) Switch on the battery note the current is running through the buffer in the electrophoresis chamber by observing bubbles at the electrodes. Electrophoresis is usually done at 100V for 0.5-1 hour at room temperature.
- 14) At the end of the run switch off the power supply, remove the gel tray from the electrophoresis chamber, wipe the outside to get rid of excess buffer
- 15) Image the gel under UV lamp.

## 5 Sequencing

### 5.1 Purpose

The primary purposes of sequencing in this protocol are to identify mutations within an MTB isolate's genome which may contribute to its drug resistance profile. Sequencing may be especially useful when the patient is not responding to the drug regimen assigned to them. This is a pragmatic pipeline, developed to aid clinical decision making on treatment, especially for MDR and XDR patients.

### 5.2 Principle

WGS can be done using a number of platforms, but a small-scale approach is to use the ONT MinION device. A MinION is a pocket-sized device which applies nanopore sequencing technology to nucleic acid analyses, with far reaching applications. A nanopore is a nano-scale hole. In its devices, ONT passes an ionic current through nanopores and measures the changes in current as biological molecules (the DNA) pass through the nanopore. The information about the change in current can be used to identify that molecule. The overlapping identical sections of these strings of bases can then be lined up, so that the whole genome can be predicted. This can then be compared to a reference genome and differences mapped. To watch ONT's explanatory videos, follow this link: <https://nanoporetech.com/how-it-works>.

#### 5.2.1 Anatomy of a flow cell

A MinION Flow Cell is a consumable used with MinION, Mk1C and GridION devices. It contains the Application-Specific Integrated Circuit (ASIC), where the nanopores are embedded into a membrane and which controls and measures the experiments. Priming buffer is loaded into the priming port, whilst the library containing the barcoded samples are loaded into the SpotON port. A flow cell and its important features is shown in Figure 2.

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael



Figure 2. Diagram of an ONT flow cell and the important features. Image reproduced from [Street Science Community](#).

Note that Nanopore give a 3 month guarantee with flow cells when kept at 4°C, but even when stored correctly, the number of active pores will decline over time, so it is better to use them as soon as possible. As delivery may take weeks, consider the time frame when ordering flow cells.

Generally, a flow cell with ~1,500 pores is capable of sequencing 12 MTB isolates to a depth required for downstream analysis; therefore a flow cell needs >100 active pores per sample you run. Remember that the lower the number of pores, the less data will result from the run and therefore if you have a flow cell with 900 active pores, you should ideally reduce the number of samples you run to e.g. 9. Note that this is not an exact science, but can be used as a guideline, as the amount of data generated will depend on multiple factors such as DNA quality, library preparation, computing power etc.

### 5.2.2 Rapid Barcoding kit 96 (SQK-RBK110.96)

More information about this specific kit can be found here: <https://store.nanoporetech.com/uk/rapid-barcoding-kit-1.html>. The kit is optimised for simplicity and speed, rather than for obtaining maximum throughput. Due to the simple nature of the workflow and the fact that little sample manipulation is required (e.g. minimal pipetting steps and no clean-ups), some very long reads (up to 100 kb; however, the average usually around 10-15 kb) can be achieved with this kit, despite the required transposase fragmentation. However, in order for long reads to be observed in sequencing, long fragments need to be present in the sample in the first place.

This kit and protocol is recommended for users who:

- wish to sequence purified (bacterial) pathogen DNA
- wish to multiplex samples to reduce price per sample
- need a PCR-free method of multiplexing to preserve additional information such as base modifications
- require a short preparation time
- have limited access to laboratory equipment
- have 50-100 ng of DNA in each sample

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

Using high quality native DNA is required for MTB ONT sequencing, as PCR amplifying of MTB does not provide even coverage and results in poor or no depth in some areas, which would result in missing data and potentially incorrect results.

### 5.2.3 Using ONT to sequence MTB

Sequencing can be done using a number of platforms, but a small-scale approach is to use the ONT MinION device. A nanopore is a nano-scale hole. In its devices, ONT passes an ionic current through nanopores and measures the changes in current as biological molecules (the DNA) pass through the nanopore or near it. The information about the change in current can be used to identify that molecule (e.g. A, T, G or C). The overlapping identical sections of these strings of bases can then be lined up, so that the whole genome can be predicted. This can then be compared to a reference genome and differences mapped. For more information on the requirements of ONT sequencing, please refer to the [PANDORA-ID-NET TGHN bioinformatics hub page](#).

## 5.3 Procedure

Tutorial videos for ONT procedures are available on [the PANDORA-ID-NET TGHN sequencing hub page](#). Familiarisation with the methods is recommended before starting this pipeline.

### 5.3.1 Equipment, consumables and reagents

The majority of the reagents needed for this protocol are provided in the SQK-RBK110.96 kit (see Figure 3). Purchase of an SQK-RBK110.96 kit includes Flow Cell Priming flush buffer and flush tether, so this does not need to be purchased additionally. The Rapid Barcoding Kit contains twelve unique barcodes (RB01-12, called ‘fragmentation mix’ in ONT literature). Each kit has sufficient reagents to generate 12 libraries containing 24 barcodes, 6 libraries containing 48 barcodes, 3 libraries containing 96 barcodes.

Note that if you are using a MinION Mk1B, you will need a computer to run it. ONT have provided a list of [minimum IT requirements](#) for a sequencing computer. We recommend a computer with a GPU, as it will be able to real-time basecall using the high accuracy (HAC) algorithm, which is recommended for accuracy.

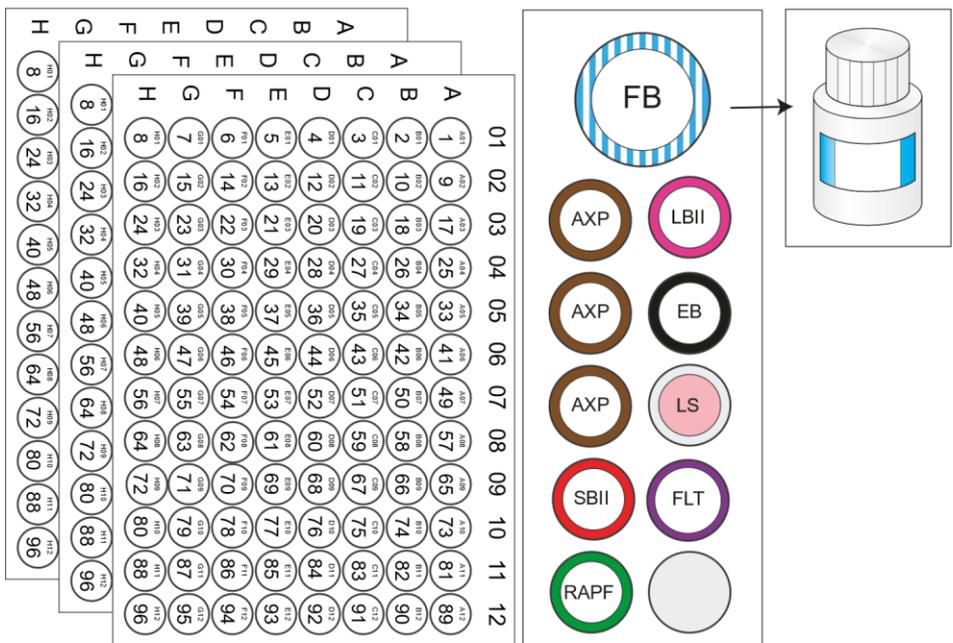
Equipment	Product Number	Manufacturer
MinION Mk1B, Mk1C or GridION sequencer		ONT
Microfuge	15881635	Fisher Scientific
PCR machine (e.g. T100 Thermal Cycler)	T100	BioRad
Magnetic rack Invitrogen™ DynaMag™-2 Magnet	10723874	Fisher Scientific
<b>Optional:</b> Hula mixer	10548425	Fisher Scientific
Computer (if using a MinION Mk1B)		
Qubit 4 fluorometer	16223001	Fisher Scientific
Vortex (e.g. Fisherbrand mini vortexer)	16343196	Fisher Scientific
Timer		
Ice box		
Appropriate storage for your generated data (either the computer, a hard drive or cloud-based)		

Consumables	Product Number	Manufacturer
-------------	----------------	--------------

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

Eppendorf DNA lo bind tubes 1.5ml (250)	EP0030108051-250EA	Sigma Aldrich
Qubit Assay Tubes (500)	12037609	Fisher Scientific
R9.4.1 Flowcell	FLO-MIN106D	ONT
1000 µL pipette and aerosol resistant tips		
200 µL pipette and aerosol resistant tips		
20 µL pipette and aerosol resistant tips		
10 µL pipette and aerosol resistant tips		
PCR tubes	I1402-8100	Starlabs

Reagents	Product Number	Manufacturer
Nuclease-Free H2O	W4502-10x50ml	Sigma Aldrich
Qubit 1x dsDNA HS Assay Kit (100)	Q33230	Fisher Scientific
Rapid Barcoding Kit 96	SQK-RBK110.96	ONT
Ethanol, absolute (e.g. Fisher Bioreagents)	16606002	Fisher Scientific
<b>Optional:</b> Flow cell wash kit	EXP-WSH004	ONT



FB : Flush Buffer  
 AXP : AMPure XPBeads  
 SBII : Sequencing Buffer II  
 RAPF : Rapid Adapter F  
 LBII : Loading Beads II  
 EB : Elution Buffer  
 LS : Loading Solution  
 FLT : Flush Tether

Figure 3. Contents of ONT's Rapid Barcoding 96 Kit (SQK-RBK110.96). Image taken from the [ONT store website](https://www.ont.com).

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

### 5.3.2 Library preparation

ONT's quick version of [this protocol can be found here](#) (you will have to log in to your ONT account). Thaw kit components at room temperature and mix by pipetting as indicated by the table below. Then return reagents on ice. The reagents should all be stored at -20°C. The flow cells should be stored at 4°C (NEVER freeze them).

Reagent	Thaw at room temperature	Briefly spin down	Mix well by pipetting
Fragmentation Mix plates RB01-96	Doesn't freeze at -20°C	✓	✓
Rapid Adapter F (RAPF)	Doesn't freeze at -20°C	✓	✓
Sequencing Buffer (SBII)	✓	✓	✓*
Loading Beads (LBII)	✓	✓	✓**
AMPure XPBeads (AXP)	✓	✓	✓**
Elution buffer (EB)	✓	✓	✓
Loading Solution (LS)***	✓	✓	✓
Flush Buffer (FB)	✓	✓	✓*
Flush Tether (FLT)	✓	✓	✓

\* Vortexing, followed by a brief spin in a microfuge, is recommended for Sequencing Buffer (SBII) and Flush Buffer (FB)

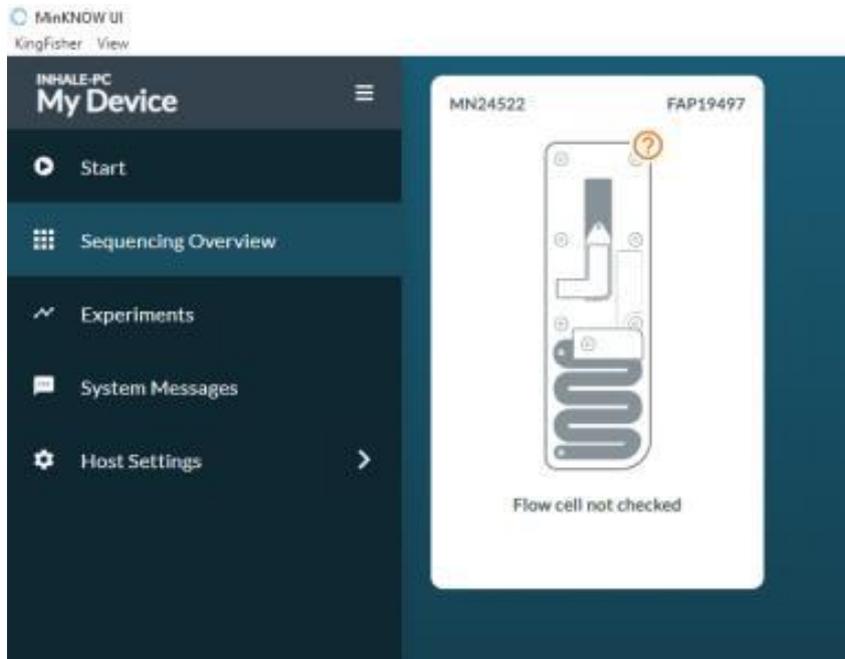
\*\* Mix by pipetting or vortexing immediately before use

\*\*\* Not all protocols require this (this protocol does not use it)

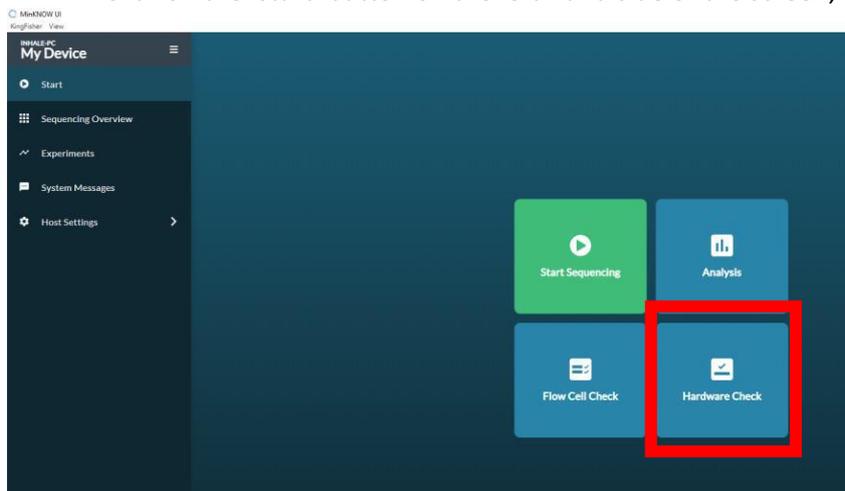
### 5.3.3 Flow cell quality control (QC)

You should check the quality (QC) of your flow cells within a few days of receiving them to ensure they have sufficient active pores to run sequencing to acceptable standards. A tutorial video explaining [this protocol can be found here](#). Note that depending on the computer and version of MinKNOW that you use, the screen may look slightly different.

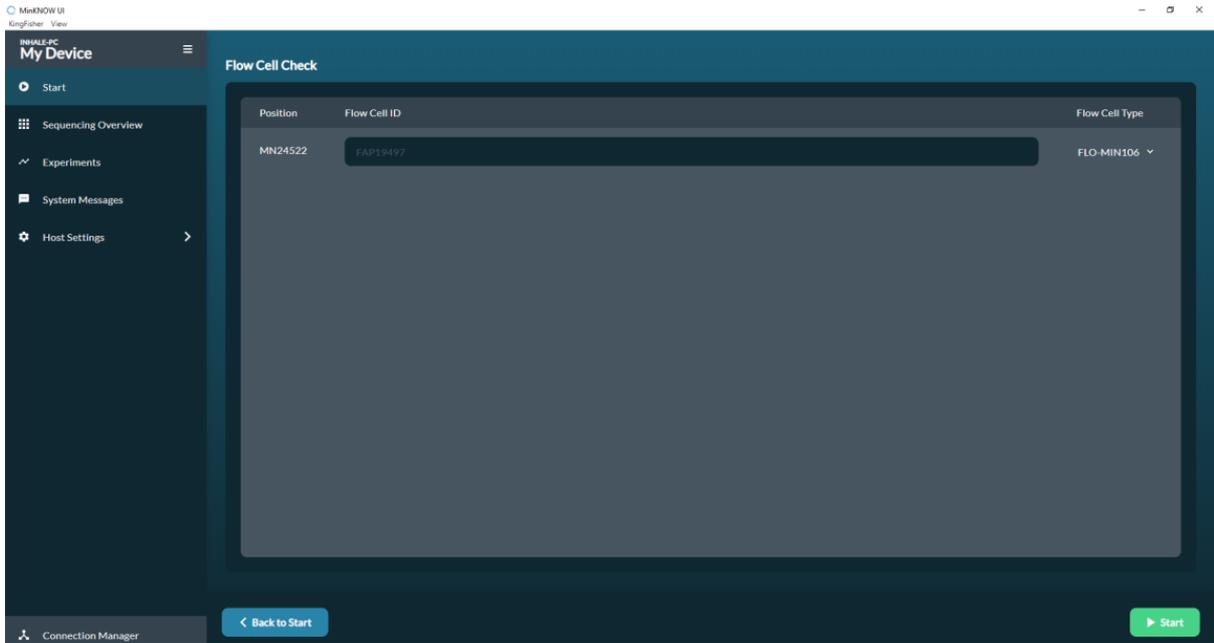
1. Open the sequencing programme MinKNOW on your computer (unless you are using the Mk1C, MinKNOW is integrated into the device).
2. Take a MinION flow cell out of the fridge and insert it into a MinION. Once you have loaded the flow cell, you will see that MinKNOW will recognise it. It will say 'flow cell not checked'.
3. Perform Platform QC to determine number of active pores in your flow cell. The steps to do this depend on the version of MinKNOW software installed on your computer (you might see a different screen to the one in this protocol).



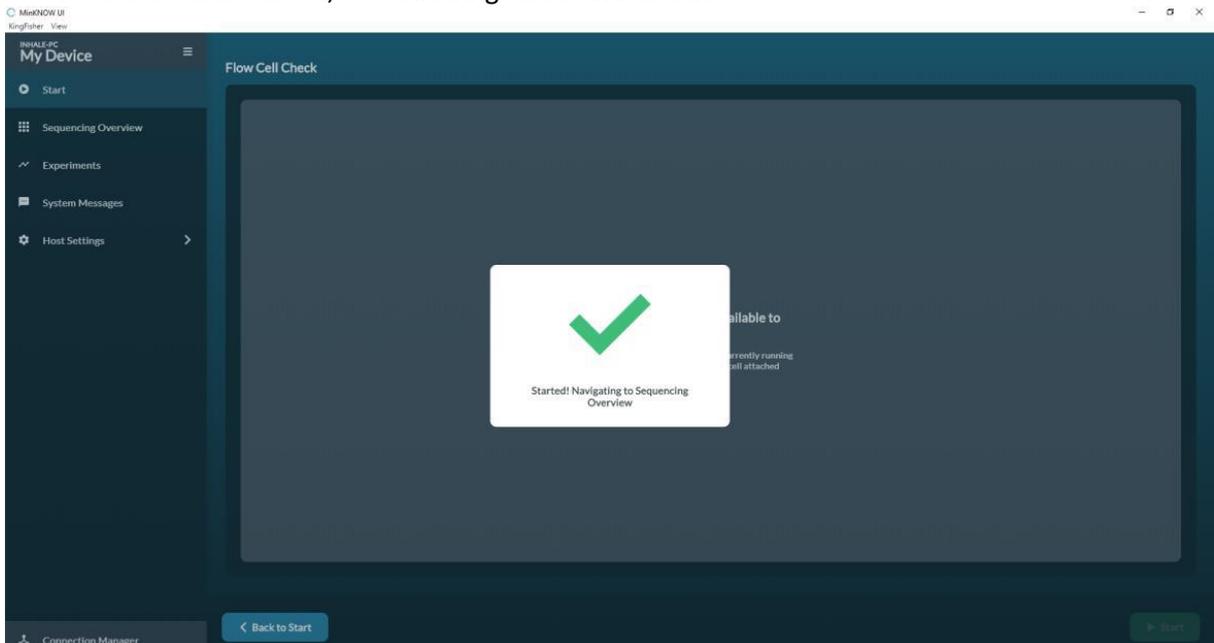
4. Click on the 'start' button on the left-hand side of the screen, then click on 'flow cell check'



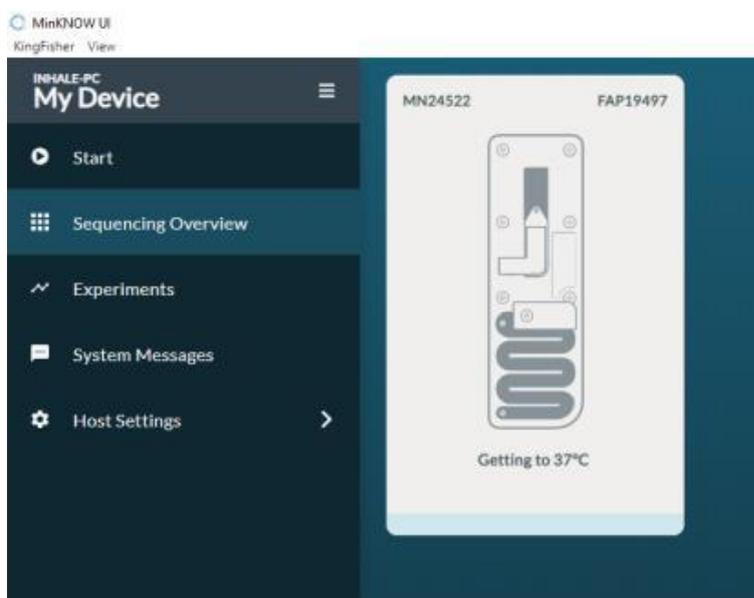
5. This will bring up the following screen:



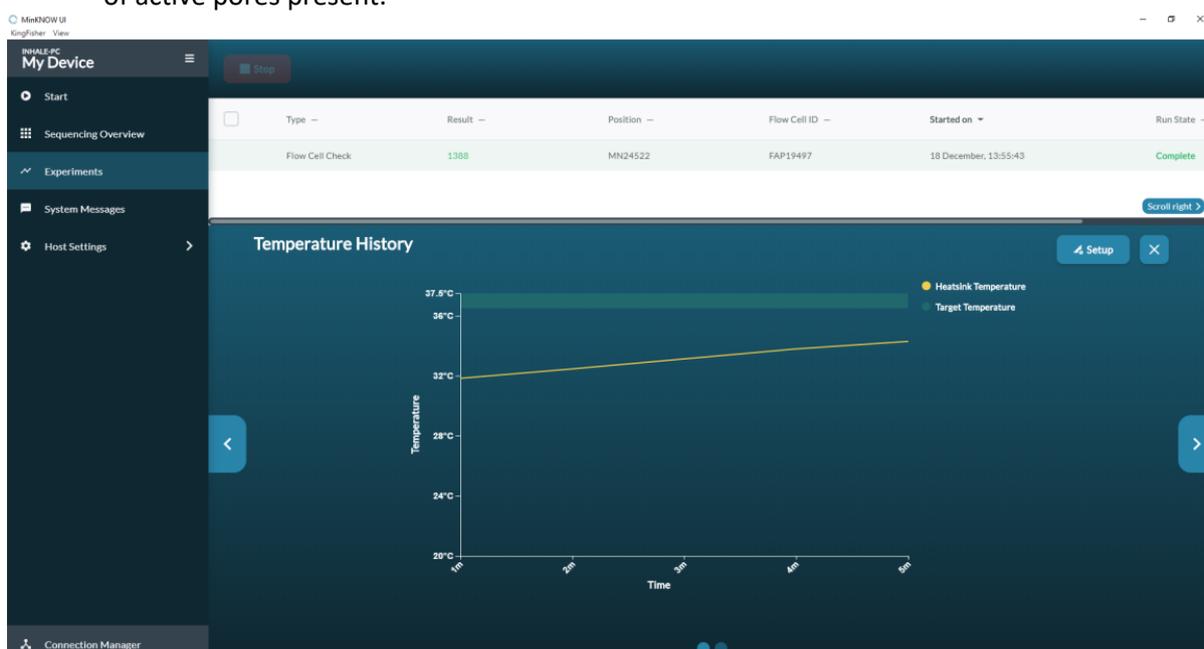
6. Once it has started, the following screen will show:



7. It will then return to the sequencing overview page. Click on the MinION box (it will say 'getting to 37°C').



- You will then see information about the QC run, including (once it has finished) the number of active pores present.



- Note down the flow cell number and number of active pores on your sequencing documentation

NB. You should get around 1300-1500 active pores. If you get less than 800, [follow these steps](#), if the flow cell still has less than 800 pores, then contact ONT ([support@nanoporetech.com](mailto:support@nanoporetech.com)) and they will send you a replacement.

### 5.3.4 Sequencing preparation

As multiple samples will be run on a single flow cell, it is important to input the same amount of DNA for each sample, so that all has a similar number of reads (if you have one or more with greater amounts, then it can be preferentially sequenced and others will have less reads, which may mean re-sequencing is required for low-read samples).

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

If all of the samples have a similar DNA sizes, then normalising to 50-100 ng should be enough to ensure even sequencing. If your samples have varying DNA sizes (e.g. some that are 30 kb and others that are 50 kb), then we recommend normalising the samples by fmol instead. 50-100 ng of 30 kb DNA is approximately 27 – 5.4 fmol, so you should aim for this when normalising your samples. This [New England Biolabs DNA mass/moles convertor](#) may be helpful. Adding less than the recommended amount of DNA means less will be available for sequencing and the sequencing pores remain idle for some time until finds the next DNA strand to sequence which reduces the efficiency and the throughput, making the run less efficient. However, adding (a lot) more can also cause issues, as it may start to block the nanopores.

Note that this kit suggests 50-100 ng of input DNA, but it's recommended to start with 100 ng (as more DNA strands make sequencing more efficient). If you can't achieve 100 ng, then add the same amount for all samples (e.g. if your lowest sample is 75 ng, then normalise all the other sample to 75 ng).

#### 5.3.4.1 DNA prep and barcoding

1. In a 0.2 ml thin-walled PCR tube mix the following:
  - 9  $\mu$ L (50-100 ng) of sample DNA
  - 1  $\mu$ L Fragmentation Mix (RB01-96: one for each sample)
2. Mix gently by flicking the tube, and spin down
3. Incubate the tube at 30°C for 2 minutes and then at 80°C for 2 minutes using a thermocycler. Briefly put the tube on ice to cool it down
4. Pool all barcoded samples in your desired ratio, noting the total volume

#### 5.3.4.2 Bead wash

Note that a bead wash is done to reduce the volume of the library prep

5. Resuspend the AMPure XP Beads (AXP, or SPRI) by vortexing
6. To the entire pooled barcoded sample from Step 4, add an equal volume of resuspended AMPure XP Beads (AXP, or SPRI) and mix by flicking the tube
7. Incubate on a Hula mixer (rotator mixer) for 5 minutes at RT
8. Prepare at least 3 mL of fresh 80% ethanol in Nuclease-free water
9. Spin down the sample and pellet on a magnet. Keep the tube on the magnet, and pipette off the supernatant
10. Keep the tube on the magnet and wash the beads with 1.5 ml of freshly-prepared 80% ethanol without disturbing the pellet. Remove the ethanol using a pipette and discard
11. Repeat the previous step
12. Briefly spin down and place the tube back on the magnet. Pipette off any residual ethanol. Allow to dry for 30 seconds, but do not dry the pellet to the point of cracking
13. Remove the tube from the magnetic rack and resuspend the pellet in 15  $\mu$ L Elution Buffer (EB). Incubate for 10 minutes at RT
14. Pellet the beads on a magnet until the eluate is clear and colourless, for at least 1 minute
15. Remove and retain 12  $\mu$ L of eluate (1  $\mu$ L for Qubit, 11  $\mu$ L for sequencing) into a clean 1.5 ml Eppendorf DNA LoBind tube (avoid the beads. Check the pipette tip against a white piece of paper to see if any brown beads are in the tip. If it is stained brown, put back in the tube on the rack and allow the beads to go back to the magnet, then try again)

#### 5.3.4.3 Qubit

16. Quantify 1  $\mu$ L of eluted sample using a Qubit fluorometer, as described in Section 4.2.1

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

#### 5.3.4.4 Addition of RAP F

17. Add 1  $\mu\text{L}$  of Rapid Adaptor F (RAP F) to the 11  $\mu\text{L}$  barcoded DNA in your tube. Mix gently by flicking the tube, and spin down
18. Incubate at RT for 5 minutes then keep on ice until you are ready to prepare the library (section 5.3.6)

#### 5.3.5 Flow cell preparation

Whilst ONT sequencing is relatively forgiving, the one thing you should be very careful about is not to introduce any air bubbles to the ASIC window. If air touches any pores, they will no longer work. Therefore, if you introduce an air bubble, then push it across the ASIC window when adding more priming mix/sample, you will most likely destroy all of the pores that the air bubble has passed over. You can massively reduce the sequencing capacity this way, so it is imperative to maintain a liquid surface across the ASIC window at all times. There are a few things you can do to try and ensure this:

- Whenever you are introducing more liquid to the priming port (e.g. priming mix, wash buffer or storage buffer) always pipette up a small volume of liquid from the flow cell before you introduce more priming mix.
- When introducing liquid to the priming port, do not use the plunger on the pipette, as you risk accidentally pressing too hard and expelling air. Instead, twist the plunger instead. This gives you much more control.
- Just before introducing the priming mix, gently form a small drip at the end of the pipette tip by twisting the plunger. This can be placed on top of the priming port. This ensures that you are definitely expelling liquid, rather than sucking it up and also ensures there is a liquid-liquid interface between the priming port and the pipette tip.
- It also helps to leave around 10-20  $\mu\text{L}$  of the priming mix in the pipette tip. This will ensure you haven't introduced bubbles.

##### 5.3.5.1 To prepare the flow cell

1. Prepare the priming mix: In a 1.5 mL microcentrifuge tube, add 30  $\mu\text{L}$  of thawed and mixed Flush Tether (FLT) and 1,170  $\mu\text{L}$  Flush Buffer (FB), and mix by pipetting up and down
2. Open the MinION lid and slide the flow cell under the clip. Press down firmly on the flow cell to ensure correct thermal and electrical contact
3. Open priming port cover by sliding clockwise to 90 degrees
4. Using a P1000 pipette, draw back a small volume (no more than 20  $\mu\text{L}$ ) of buffer from the flow cell (this will remove any air pockets found directly under priming port)
5. Using a P1000 pipette load 800  $\mu\text{L}$  of priming mix through priming port. Wait 5 minutes. During this time, prepare the library for loading by following the steps below

#### 5.3.6 Library preparation

1. In a new tube, prepare the library for loading by adding the following:
  - 37.5  $\mu\text{L}$  Sequencing Buffer (SBII)
  - 25.5  $\mu\text{L}$  Loading Beads (LBII), mixed immediately before use
  - 12  $\mu\text{L}$  DNA library

Note that step 1 is your prepared library. The following steps are to complete the priming of the flow cell and then loading of the library onto the flow cell.

2. Gently open SpotON sample port cover
3. Using a P1000 pipette, draw back a small volume (no more than 20  $\mu\text{L}$ ) of buffer from the flow cell (this will remove any air pockets found directly under priming port)
4. Using a P1000 pipette load 200  $\mu\text{L}$  of priming mix through priming port (with SpotON port open). Proceed promptly to next step

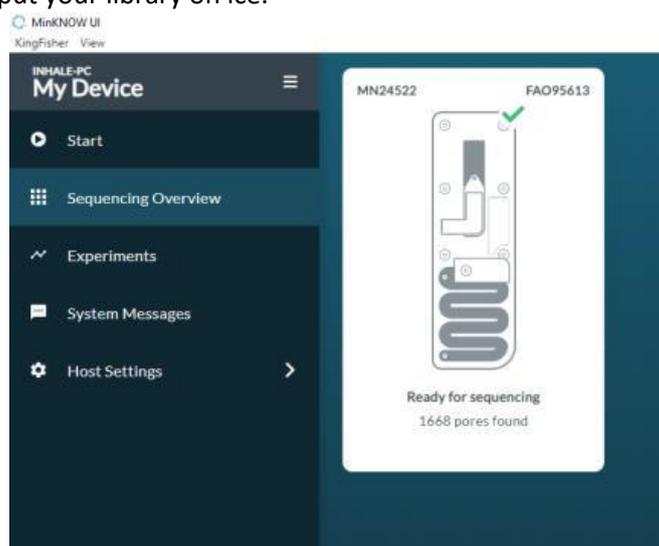
ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

- Using a P200 pipette slowly add 75  $\mu$ L of sequencing mix in a dropwise fashion through the SpotON port. Do not insert pipette tip into SpotON port (you risk forcing air into the window)
- Close both ports (note that you will need both ports open when you load the sample, as it enables the sample to be sucked into the SpotON port and across the ASIC window. If the Priming port is closed, you will find that the sample will sit in the SpotON port and not get sucked across the ASIC window)

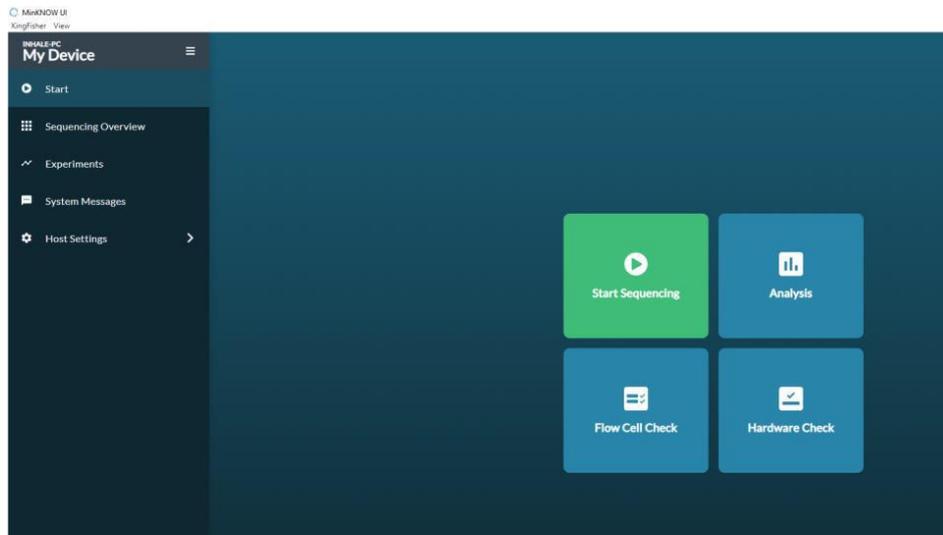
### 5.3.7 Starting your run (using MinKNOW software)

The version of MinKNOW will affect how you navigate around the programme, but the principle will be the same. Your MinION and flow cell should still be connected. Check your MinION is plugged in to your computer's USB 3.0 port, and the flow cell is clipped firmly into the MinION device. On the left-hand side of the window is the option bar. These buttons provide access to the functions in MinKNOW. A tutorial video explaining [how to start a sequencing run can be found here](#). Note that MinKNOW is updated often, so check the ONT software downloads webpage to make sure you have the most up to date version (it can make a big difference to speed and basecalling accuracy).

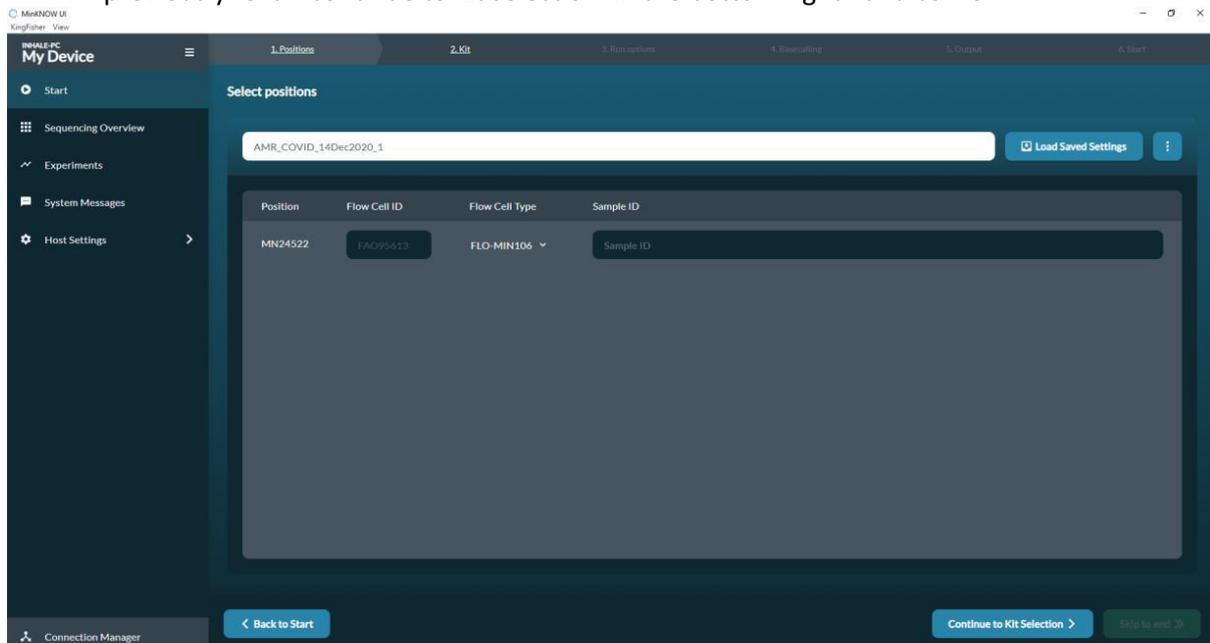
- Each MinION device will be visible as a white rectangle in the main panel. If you have a flow cell in place, you will see the flow cell number and it should say 'ready for sequencing' and the number of active pores (if not you will need to QC the flow cell again, see section 5.3.3). You should ensure you QC the flow cell at the previous stage, but if you must repeat it at this stage, make sure to put your library on ice.



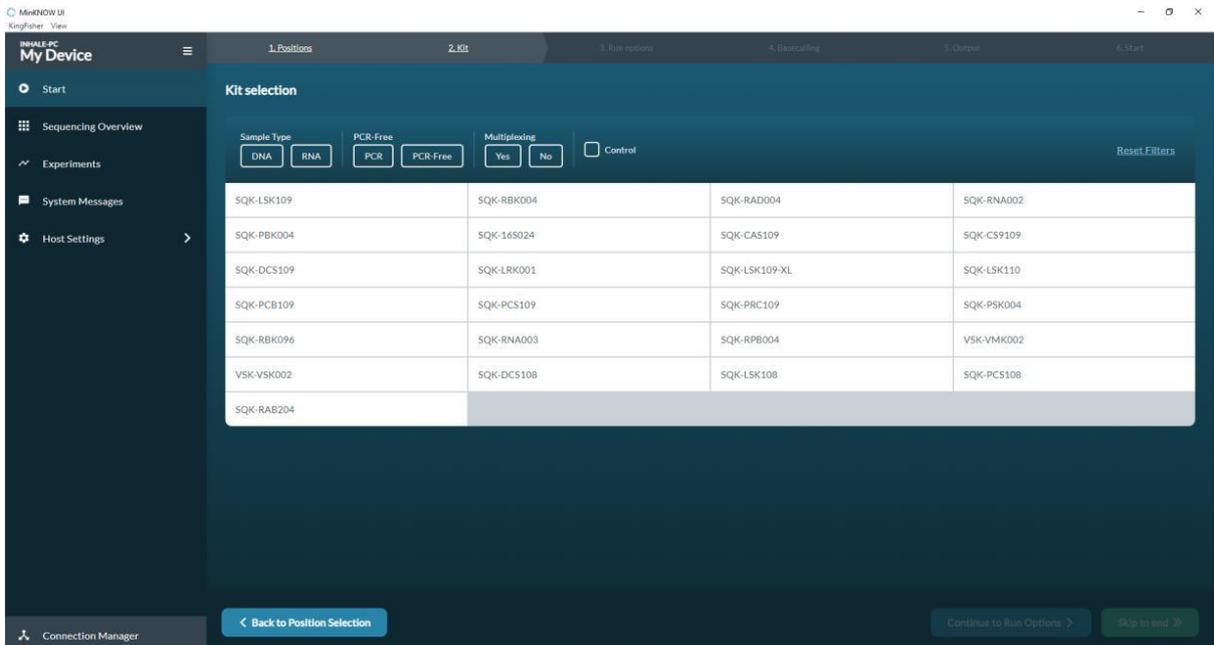
- In the option bar on the left-hand side of the screen, click on 'Start'. This will bring up the screen below. Click on 'start sequencing'



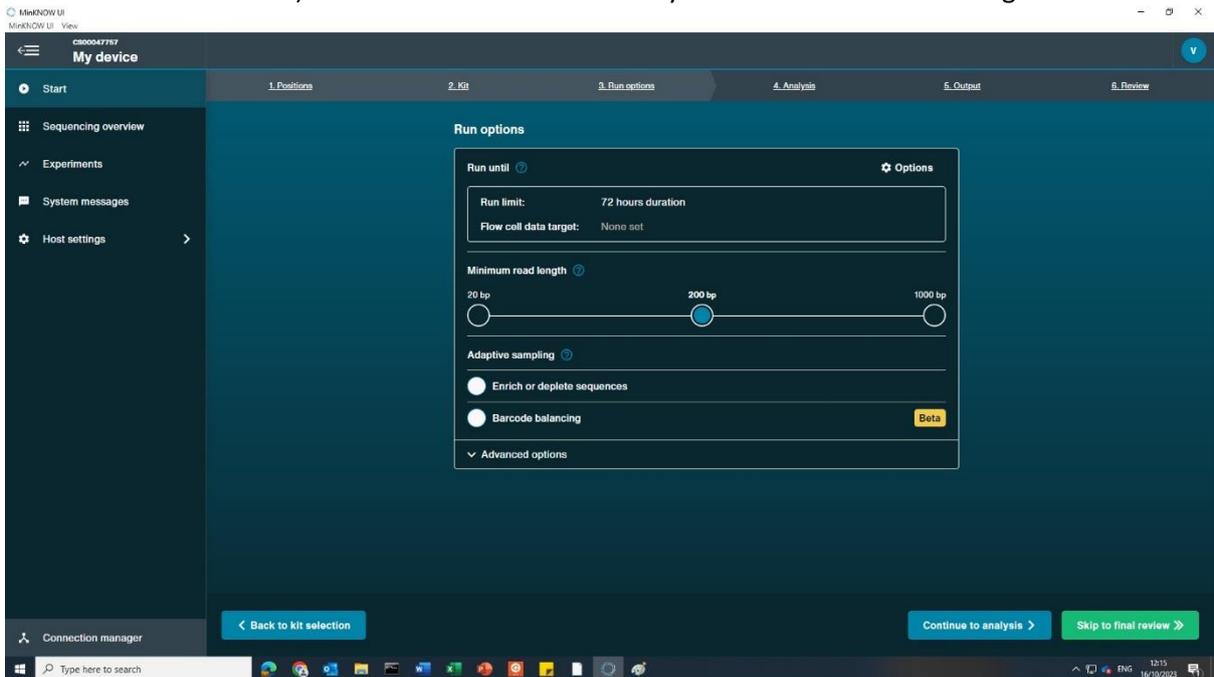
- You will be taken to the 'select positions' screen. In the white box, type in the name of your experiment. It also gives you the option to load a sequencing run if you have saved one previously. Click 'continue to kit selection' in the bottom right hand corner.



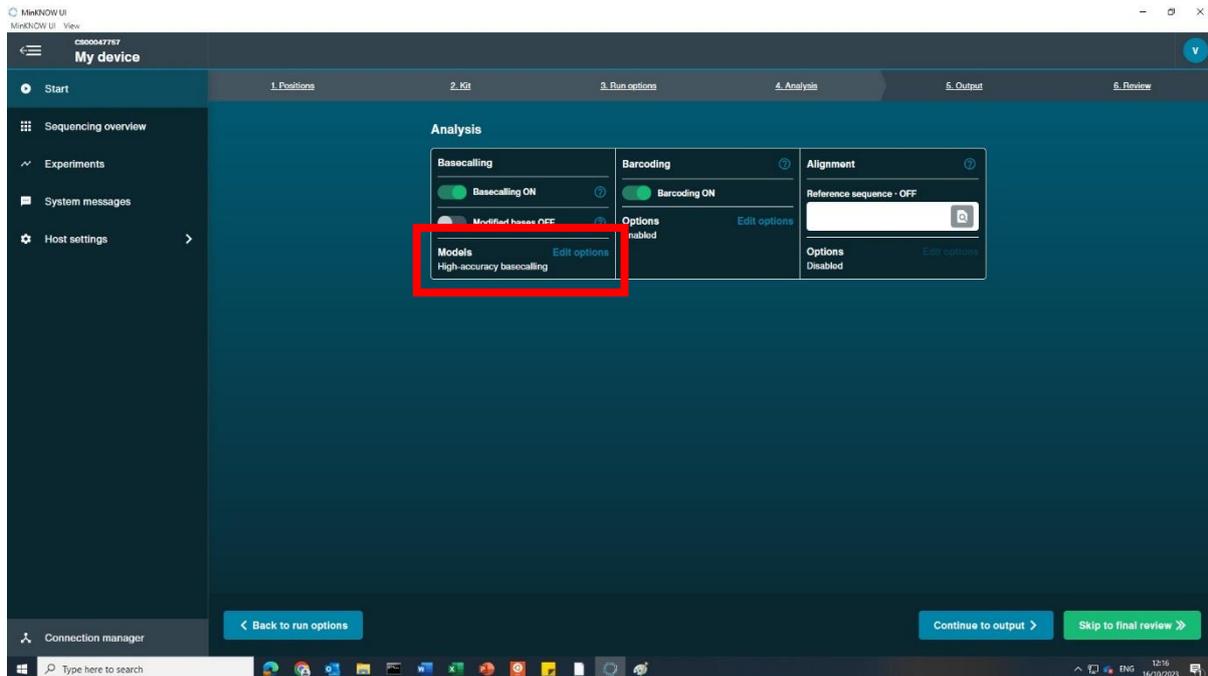
- Click on the kit that you have used. In this protocol, the kit number is SQK-RBK004. You can also narrow down the search by choosing the options 'DNA', 'PCR-free' and 'Yes' (to multiplexing). Once you have selected the right kit, you can click 'continue to run options' in the bottom right corner.



- The run length can be edited on this page (keep the bias voltage at -180 mV). 48 hours is usually sufficient (you will find many of the pores will have died after 48 hours, but it can be left for 72 hours). Click on the 'continue to analysis' button in the bottom right corner.

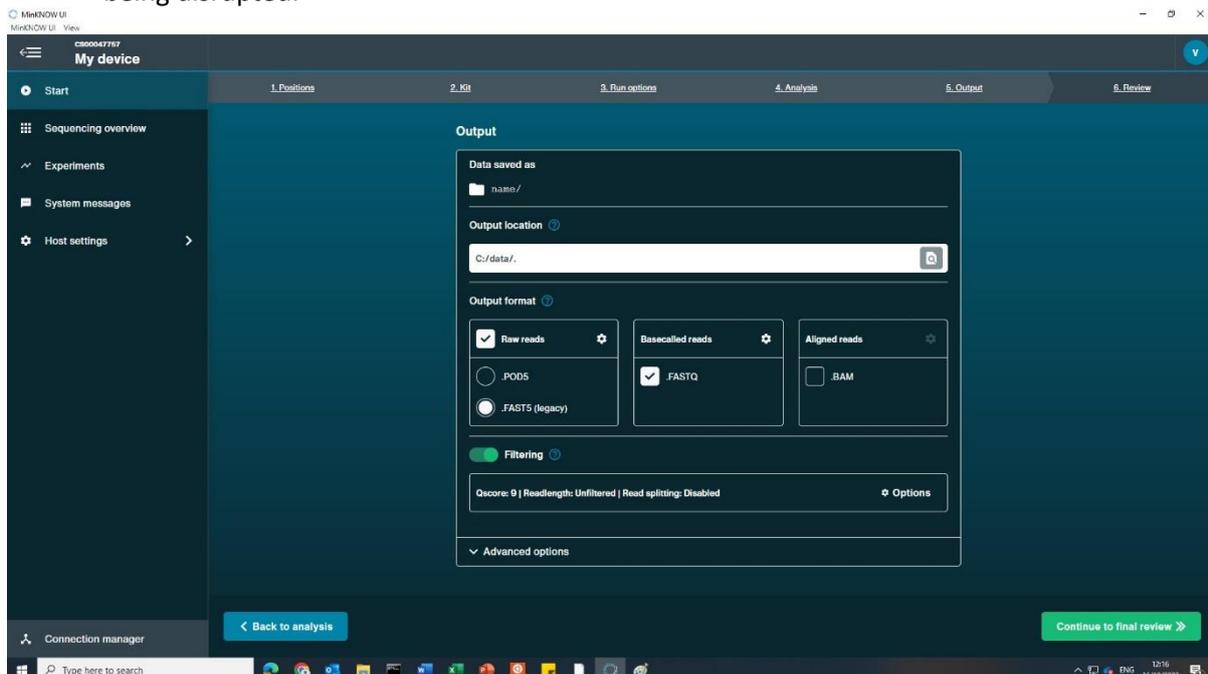


- Ensure that the basecalling and barcoding boxes are both checked. High accuracy or super high accuracy basecalling is recommended for this pipeline, so change it from fast (automatic option) to HAC (see the red box in the image below). Click on 'continue to output' button in the bottom right corner.

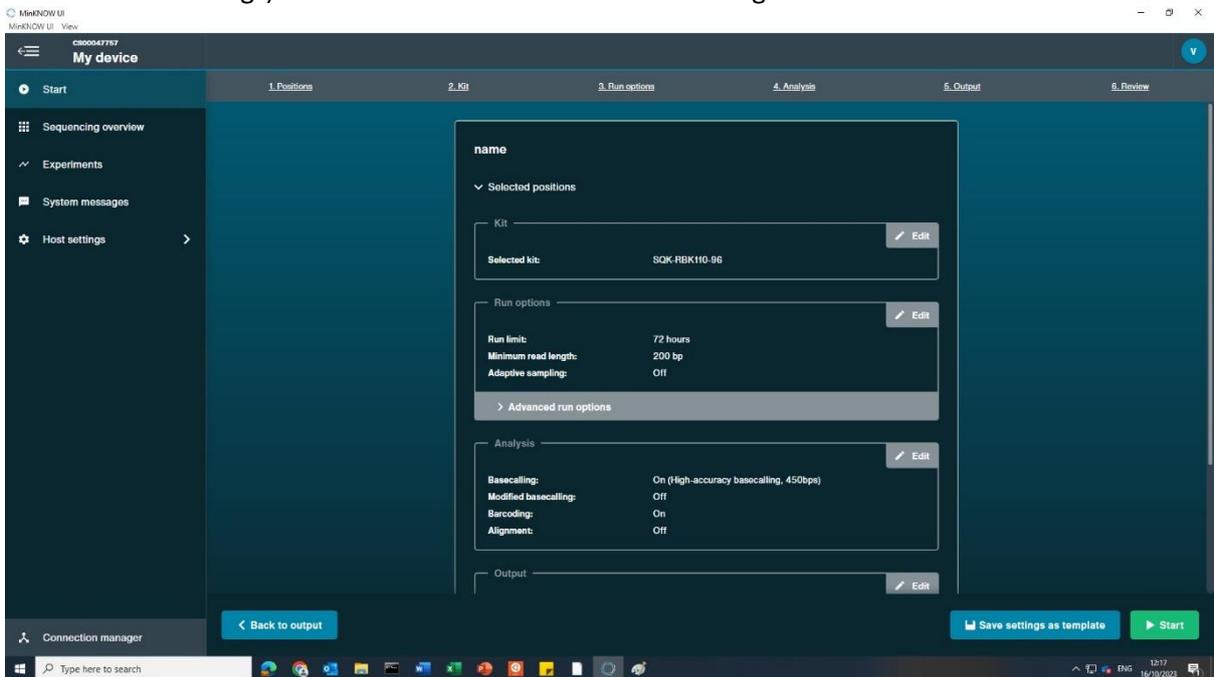


- Click on the browse button in the white text bar to select the output folder (it's useful to create a new folder for each new sequencing run). Be aware that a good sequencing run will generate many GB (potentially up to 500 GB) of data, so make sure you have enough space to store it all. Sequencing will stop if the computer runs out of disc space. Ensure fast5 and fastq output format boxes are checked and leave the Q score at 9 (the threshold number defining which reads pass (>9) and which fail (<9). Nine is generally used as the threshold (this can be changed depending on the amount of reads you want to include), although it depends which version of MinKNOW you are using. Click the 'continue to run setup' button in the bottom right corner.

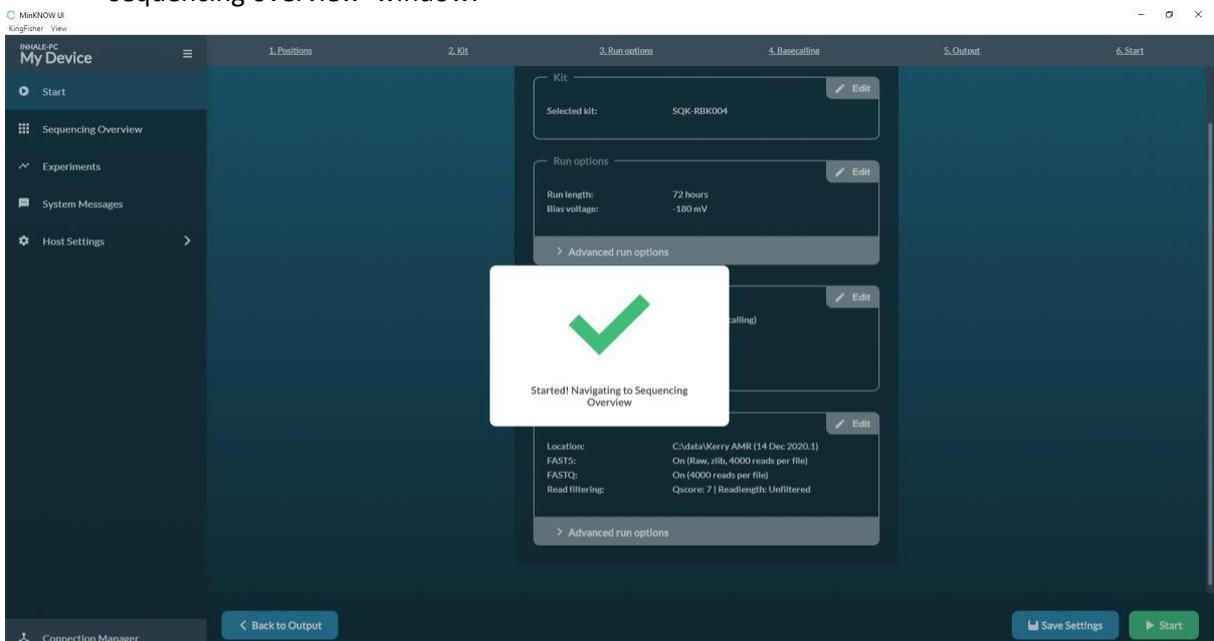
NOTE that you should always save your sequencing data to the computer's internal storage and NOT to an external hard drive or remote drive, otherwise you risk the sequencing run being disrupted.



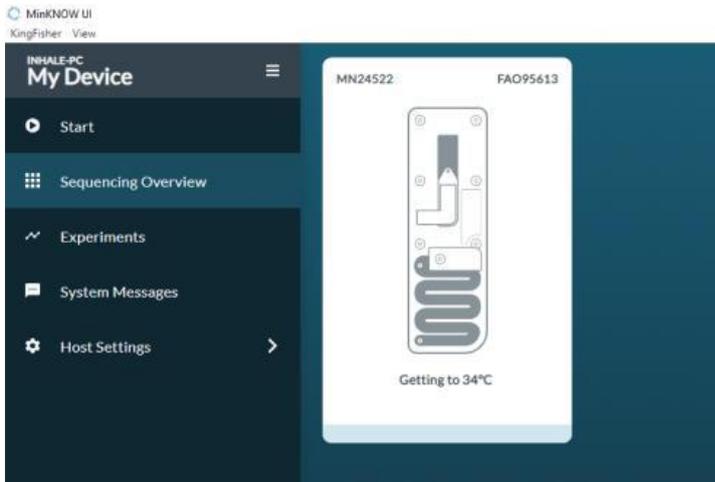
8. This is an overview screen to ensure the correct settings have been selected. At this point the sequencing set up can be saved (useful if you are running a number of experiments using the same settings). Click the 'start' button in the bottom right corner.



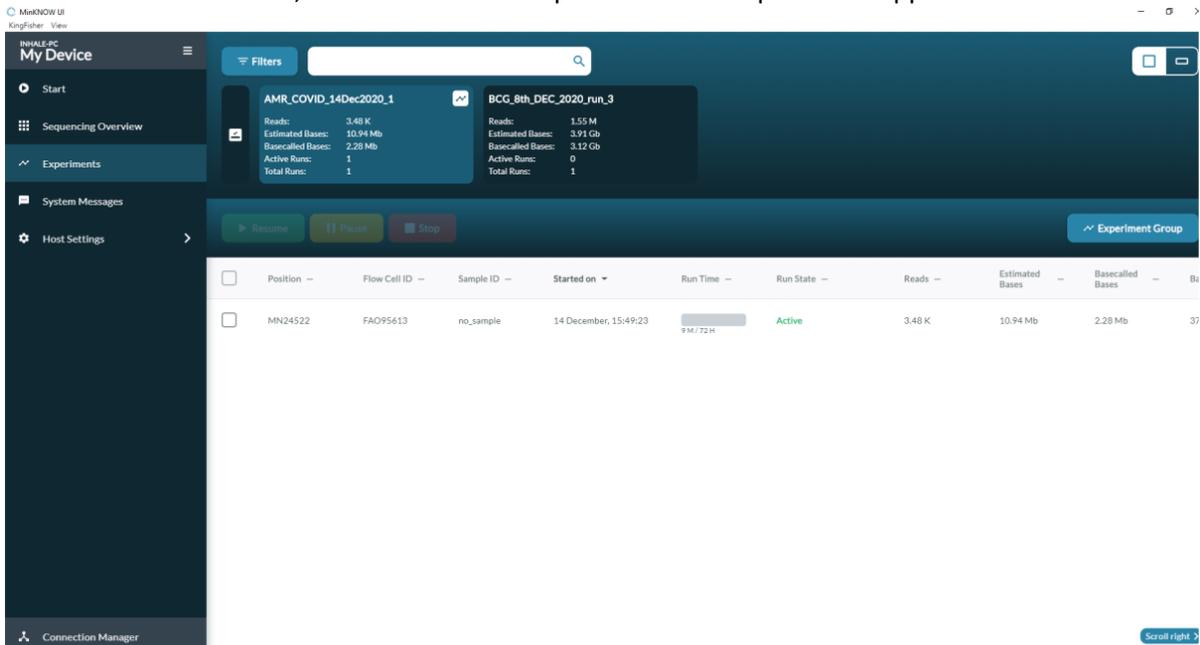
9. The following message box will appear and the software will then be directed to the 'sequencing overview' window.



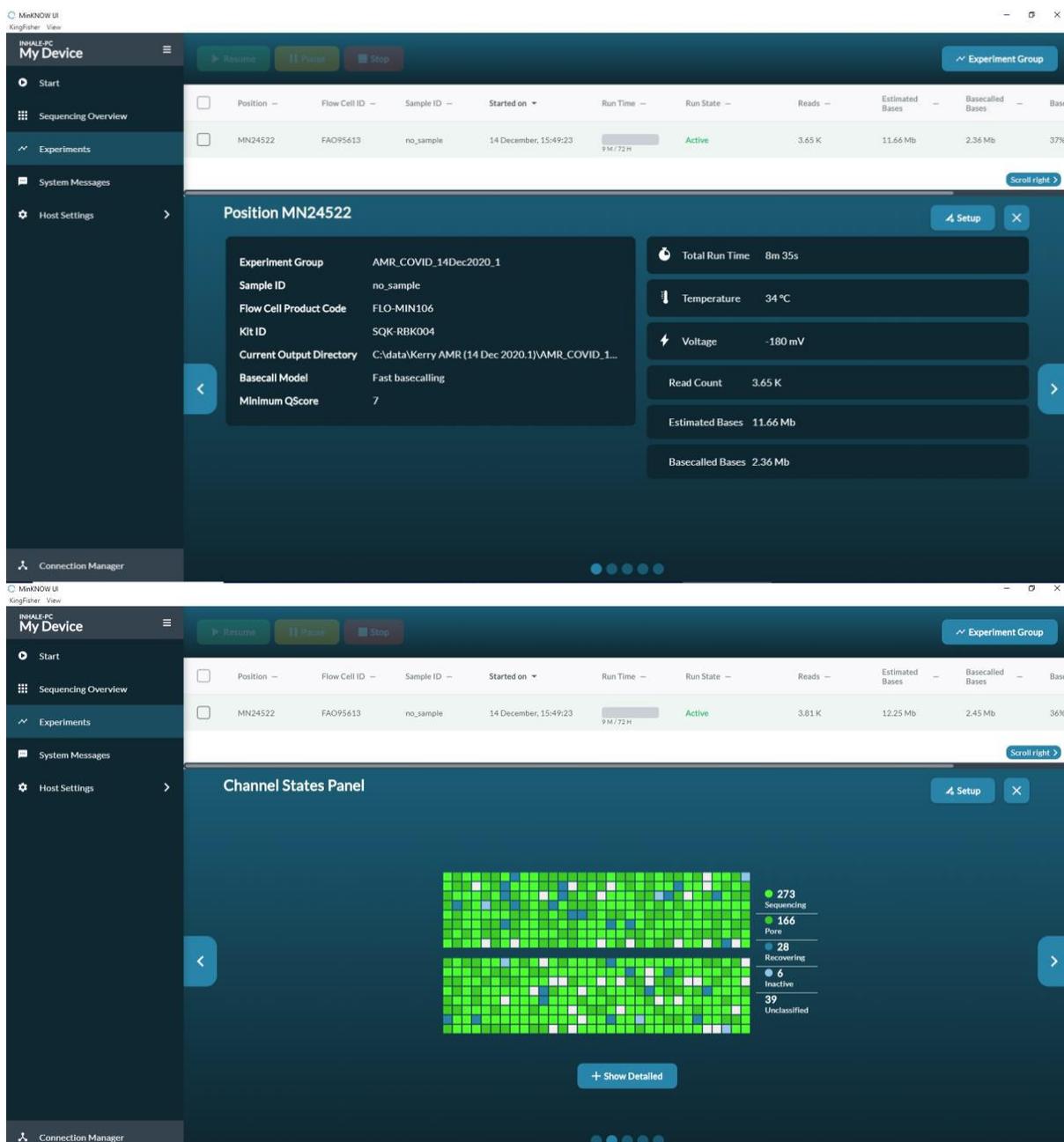
10. In the sequencing overview window, the white MinION box will show that the temperature is ramping up to 34°C. Double click on the white MinION box or the experiments button in the option bar to show the current experiment.



11. It will take a few minutes for the green 'active' button to appear once sequencing has started, but once it does, click on it and the experiment details panel will appear.



12. By clicking on the left and right arrows on the experiment details panel you can scroll through the different information windows, including the overview window and channel states panel. You will also find the cumulative reads, read length and barcode reads windows useful.



## 6 Sequencing data analysis

To familiarise yourself with bioinformatic analysis of sequencing data, recommended reading can be [found on PANDORA-ID-NET's TGHN bioinformatics hub page](#). You will need to be familiar with basic Linux command line interface (CLI). If you use a Windows machine, you can create a Linux environment using the Windows Subsystem for Linux (WSL). To learn how to do this, [we recommend reading this PDF](#), which explains the basics of ONT sequencing data analysis. Whichever programmes you use, ensure you regularly check for updates, as this technology is rapidly developing.

### 6.1 Calculating sequencing coverage and depth

Once the sequencing has begun, it can be left to run for the required length of time. However, you need to ensure that you allow it to sequence long enough to ensure you get an adequate read coverage, or depth. This is defined as the number of unique reads that include a given nucleotide in

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

the reconstructed sequence. The more times a nucleotide is “read” (sequenced), the more confident you will be that it is the correct base and not just a sequencing error. For more information on depth and coverage, please refer to [the PANDORA-ID-NET ONT bioinformatics document](#).

Because MTB has a slow mutation rate, the drug resistance is usually conferred by mutations, known as single nucleotide polymorphisms (SNPs), so to be sure that the bases we see after sequencing are correct, we need to have a good ‘depth’. For instance, if we had a depth of ‘3’, and the reads for a base were A, G and T, we could not be sure which was the correct, original base. For this reason, a coverage of at least 40 is generally suggested to ensure confidence in the base that is called. To calculate how long you will need to run the sequencing for to get the required coverage/depth, the below calculation can be used.

$$N = CG/L$$

- N is the number of reads
- C stands for coverage/depth (usually set at 40)
- G is the haploid genome length (4,400,000 base pairs for MTB)
- L is the average read length (once you begin an experiment there will be a graph on the active channel page that tells you the average read length)

For instance, to get a coverage/depth of 40 for the MTB genome, with an average read length of 10 kb, the number of reads needed to achieve this is 17,600. However, we will need to overestimate this number, as we must take into account loss of reads:

- 20% might be lost to low quality of DNA
- 10-20% might be lost to unclassified barcodes
- There will be uneven coverage, some bases will be underrepresented and only get a few x coverage if we aim exactly for 40x
- Reads might be fragmented and not be 10,000 kb depending on the pre-processing. Read length is often overestimated (it’s perhaps closer to 5-7000)

To achieve e.g. 40x coverage across the entire genome (taking into account the above considerations), it is advisable to multiply the calculated number of reads by 10 (17,600 x 10 = 176,000 reads needed for 40x coverage of the whole genome).

## 6.2 Completing base-calling if your run is interrupted

If you have turned base-calling on when you started the run, MinKNOW will base-call for you, using standard parameters. This can take many hours even once the sequencing has finished. If this is interrupted, you can use the programme ‘Guppy’ to complete the base-calling yourself, using simple line commands. Refer to the bioinformatics guide on [the PANDORA-ID-NET TGHN bioinformatics hub page](#). The PDF attachment explains the process.

The output files you will have once it has finished base-calling are .fast5 (the ‘raw’ files) and .fastq (the processed, base-called files which can then be inputted into a database). FAST5 files take up more space than fastq files, so the decision to delete them can be made (this does mean that you won’t be able to re-base-call them at a future date).

### 6.2.1 Fast vs HAC vs SUP base-calling

The latest versions of Guppy support three different types of base-calling, each of which have their own pros and cons. Whilst the minimum ONT computing requirements will enable users to run fast base-calling, it will realistically not be in ‘real time’ and may take days to complete once the actual

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

sequencing has finished (basecalling speed will depend on many factors such as, number of files, total GB of data to be basecalled and the power of the computer). High (HAC) and super high (SUP) accuracy basecalling requires computers with graphical processing units (GPU) and/or access to a high throughput computing cluster.

### 6.3 Analysis of data – TB-Profiler

There are a number of online databases available to identify the variants that are associated with drug resistance in MTB. We recommend TB-Profiler tool which maintains a database of variants curated by Dr Jody Phelan at the London School of Hygiene and Tropical Medicine. Also, this tool identifies lineages, sub-lineages of the sequenced MTB isolates and predicts the drug susceptibility from the genotype. TBProfiler comes as a browser-based graphical user interface (GUI), or a command line interface (CLI), depending on user familiarity with the command line, and the size of the data files being analysed. To learn more about how TBProfiler can be used:

Detailed Manual: <https://jodyphelan.gitbook.io/tb-profiler>

#### 6.3.1 Preparing your data for TB-Profiler

The fastq files are what is inputted into the TBProfiler database so that drug resistance mutations and strain can be identified. TB-Profiler requires one fastq file per sample. To concatenate the fastq files into one file using the Windows Command Prompt (WCP) first navigate to the folder location:

```
cd <folder address>
```

For example:

```
cd C:\Users\rekglel\Documents\Sequencing\MTB\isolate01
```

Once in the folder with your fastq files, type the following command to concatenate the files into one single file:

```
type *.fastq > isolate01.fastq.gz
```

Note that \* tells the WCP to include all of the files with that file ending.

Note that TBProfiler works with gzipped files (this compresses them to make them smaller), so we need to add the file extension .gz to the output file.

**Note that if you are using Linux or MacOS, these commands, and moving to the folder/directory will be slightly different.**

#### 6.3.2 Using TB-Profiler

##### 6.3.2.1 TB-Profiler (GUI)

The GUI version of TB-Profiler can be found here: <https://tbdr.lshtm.ac.uk/>

Using this version is faster and simpler than the CLI version, but only files up to 4GB can be uploaded. The ONT option must be checked, then the .fastq.gz file is uploaded and a reference number will be provided. Take note of the number, as it will be required to retrieve the results report otherwise you may have to re-upload your data. Note that you can batch upload files.

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

### 6.3.2.2 TB-Profiler (CLI)

The TB-Profiler pipeline is available on GitHub (from <https://github.com/jodyphelan/TBProfiler>). You will need to install TB-Profiler. The easiest way of installing it is by using Conda ([for information about how to use and install Conda, refer to the website](#)). To install TB-Profiler, follow the instructions on the GitHub page (link above). To run a sample through TB-Profiler CLI, navigate to the folder where the file is located (as in section 6.3.1). **Note that to use the CLI version, you must use Linux or MacOS. It does not work with Windows.**

You can then type the following command to create a report:

```
tb-profiler profile -m nanopore -1 isolate01.fastq -p isolate01 --txt
```

Note that you will need to change both the 'isolate01' references to whatever the name of your isolate being tested is. This will create a .txt file and a .json file that can be opened in FireFox as well as text report. To collate the TB-profiler results and generate summary files; run the following collate command:

```
tb-profiler collate
```

Batch analysis is possible in the CLI version, but first you have to create a list of all the samples' names and then run the following command:

```
ls fastq/ | grep .fastq | sed 's/.fastq//' > samples.txt
cat samples.txt | parallel --bar -j 4 tb-profiler profile -1
fastq/{}.fastq -p {} -m nanopore --txt
```

### 6.3.3 Interpreting TB-Profiler reports

The main parts to each report are:

- **Lineage report:** which MTB strain your isolate belongs to. The resolution to which TB-Profiler can identify the lineage depends on the depth and quality of DNA
- **Resistance report:** A prediction of which drugs your MTB isolate will be resistant to. Displayed as a list of drugs tested for, those with resistance have 'R' next to the drug name. It will also have the gene and SNP

#### 6.3.3.1 GUI report

The GUI browser version of the report is easy to read and interpret, and gives you multiple file choice saving options, including PDF, csv, txt and .json files.

The below example is of lineage 2.2.2 and is pre-XDR. It is resistant to rifampicin, ethambutol, pyrazinamide, streptomycin, fluoroquinolones, moxifloxacin, ofloxacin, levofloxacin and ciprofloxacin.

ONT sequencing for TB (v2.0)	University College London
16 October 2023	Authors: Dr Linzy Elton, Professor Neil Stoker, Dr Sylvia Rofael

[TB-Profiler](#)
[Home](#)
[Upload](#)
[SRA data](#)

### TB-Profiler result

Run: 768345cd-b4b1-4d22-a41b-2881dcdf3af3

Summary

**Run ID:** 768345cd-b4b1-4d22-a41b-2881dcdf3af3

**Sample name:**

**Date:** 26-01-2023 15:39:34

**Number of reads:** 30203

**Percentage reads mapped:** 95.41

**Strain:** lineage2.2.2

**Spoligotype:** 0000000000000000

**Drug-resistance:** Pre-XDR-TB

Drug	Resistance	Supporting mutations
Rifampicin	R	rpoB p.Ser450Leu (0.91)
Isoniazid		
Ethambutol	R	embB p.Met306Ile (1.00)
Pyrazinamide	R	pncA p.Cys14Arg (1.00)
Streptomycin	R	gid p.Leu79Ser (1.00)
Fluoroquinolones	R	gyrA p.Asp94Gly (1.00)
Moxifloxacin	R	gyrA p.Asp94Gly (1.00)
Ofloxacin	R	gyrA p.Asp94Gly (1.00)
Levofloxacin	R	gyrA p.Asp94Gly (1.00)

### 6.3.3.2 CLI report

If you run the CLI version of TB-Profiler, it will output the .json and .txt files, which will provide the same information but are likely to look different. A .json file can be opened using a web browser e.g. FireFox and has multiple 'drop down' options, which can be confusing. A .txt file can be opened using a text reader e.g. Notepad. The .txt file is generally easier to interpret, we recommend using this to analyse your isolate data.

Examples of a .json report (left) and .txt report (right) can be seen below. The example isolate is of lineage 4.1.2.1, is MDR and is resistant to rifampicin, isoniazid, ethambutol, pyrazinamide and streptomycin. It is susceptible to fluoroquinolones, amikacin etc.

```

▼ lineage:
  ▶ 0:      (...)
  ▶ 1:      (...)
  ▶ 2:      (...)
  ▼ 3:
    lin:    "lineage4.1.2.1"
    family: "Euro-American (Haarlem)"
    spoligo:"T1;H1"
    rd:     "RD182"
    frac:   0.9984939759036144
  main_lin:"lineage4"
  sublin:  "lineage4.1.2.1"
▼ dr_variants:
  ▼ 0:
    chrom:  "Chromosome"
    genome_pos: 761155
    ref:     "C"
    alt:    "T"
    freq:   1
    feature_id: "CCP43410.1"
    type:    "missense_variant"
    nucleotide_change: "c.1349C>T"
    protein_change:    "p.Ser450Leu"
  ▼ annotation:
    ▼ 0:
      type: "resistance_association_confidence"
      drug: "rifampicin"
    
```

Example of a .json file output

```

TBProfiler report
=====

The following report has been generated by TBProfiler.

Summary
-----
ID: pan346
Date: Fri Feb 4 14:16:27 2022
Strain: lineage4.1.2.1
Drug-resistance: MDR

Lineage report
-----
Lineage Estimated Fraction Family Spoligo type Rd
lineage4 1.000 Euro-American LAM;T;S;X;H None
lineage4.1 1.000 Euro-American T;X;H None
lineage4.1.2 0.998 Euro-American T;H None
lineage4.1.2.1 0.998 Euro-American (Haarlem) T1;H1 RD182

Resistance report
-----
Drug Genotypic Resistance Mutations
Rifampicin R rpoB p.Ser450Leu (1.00)
Isoniazid R katG p.Ser315Thr (1.00)
Ethambutol R embB p.Met306Ile (1.00)
Pyrazinamide R pncA p.Leu85Pro (1.00)
Streptomycin R rpsL p.Lys43Arg (1.00)
Fluoroquinolones
Amikacin
Capreomycin
Kanamycin
    
```

Example of a .txt file output