

cDNA sequencing on the ONT MinION platform

Protocol for cDNA synthesis and cDNA sequencing on the Oxford Nanopore Technologies (ONT) MinION platform

Overall purpose: Generate a cDNA library of poly-A transcripts from purified total RNA and sequence it on the ONT MinION platform.

Overall process overview: Poly-A tailed mRNA is reverse transcribed to cDNA. The synthesized cDNA is then amplified, followed by library preparation and sequencing on the ONT MinION device (Figure 1).

Introduction

Multiple methods exist to create cDNA from total RNA (Figure 2). Here we will use the Smart-seq protocol¹. The protocol is based on the terminal deoxynucleotidyl transferase activity of the wild-type MMLV (Moloney murine leukemia virus) reverse transcriptase². The MMLV reverse transcriptase preferentially introduces cytosines³ (mainly 3-5 nucleotides) at the end of the synthesized cDNA, in a template-independent fashion, upon reaching the terminus of the RNA molecule. Different MMLV derived reverse transcriptases have been engineered that can be used for cDNA synthesis. These novel RNA dependent DNA polymerases usually lack the RNase H activity towards RNA/cDNA hybrids of the wild type enzyme. Some of them also show increased thermostability. Examples of these enzymes are the following ones: Maxima H minus (Thermo), SMARTScribe (Clontech), RevertAid (Thermo), EnzScript (enzymatics), ProtoScript II (Neb), SuperScript II (Thermo), GoScript (Promega), RevertUP II (biotechrabbit), MMLV Point Mutant (Promega), SuperScript IV (Thermo), PrimeScript II (Clontech) etc.

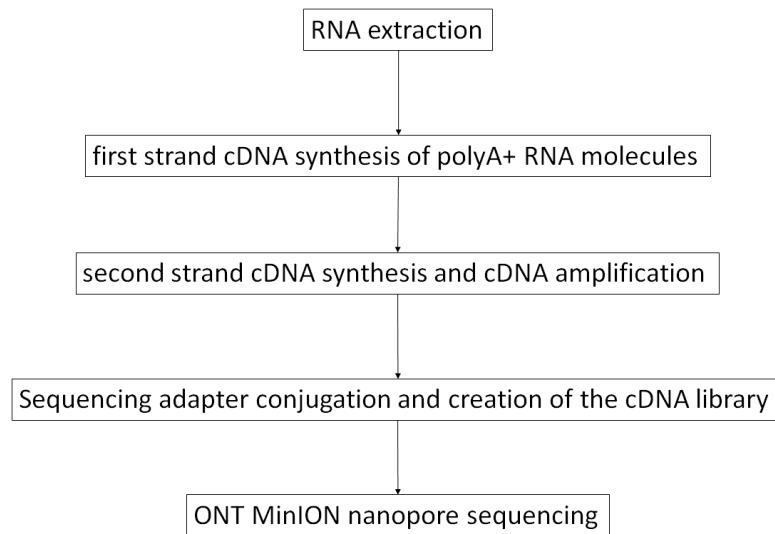


Figure 1. Schematic diagram of the procedure that is followed in this protocol.

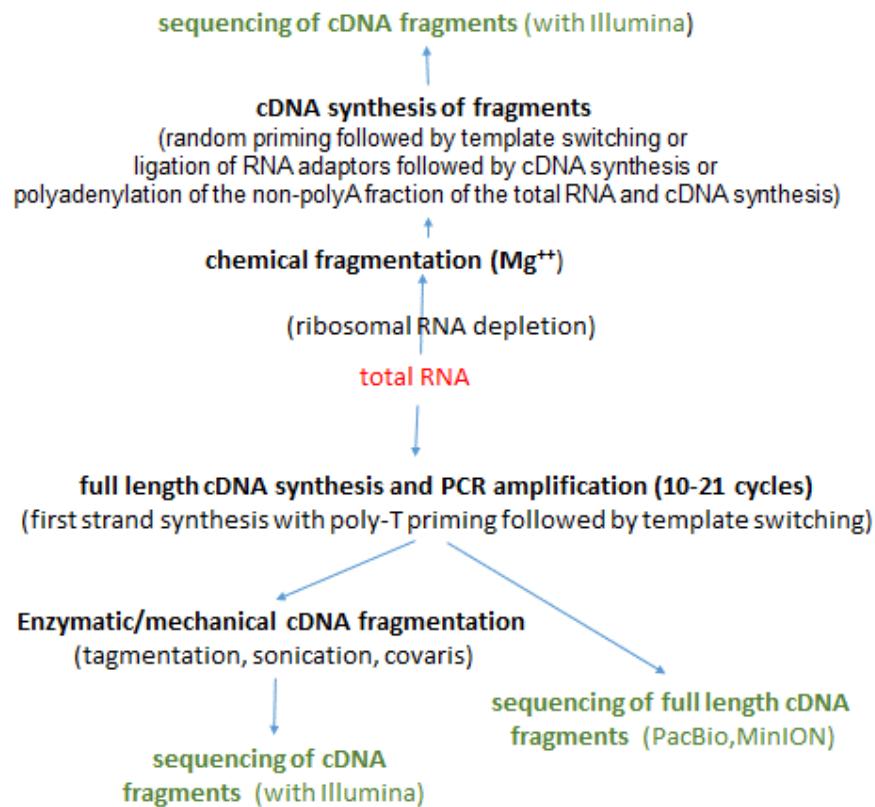


Figure 2. Different methods to synthesize cDNA molecules. The selection of the appropriate cDNA synthesis protocol depends on the sequencing platform that is going to be used.

Materials used in this protocol

➤ Reagents for bench decontamination from traces of RNases

- RNase Zap (Thermo Fischer Scientific, AM9780)

➤ General reagents

- 1M Tris-HCl pH 8.0 (Thermo Fischer Scientific, AM9855G)
- Magnetic Stand for 1.5 mL tubes (e.g. Ambion P/N AM10026) and 0.2 mL tubes
- Filtered tips (10, 20, 100, 200, 1000 µL), and respective pipettes
- Eppendorf Centrifuge 5424R or 5424 (or equivalent)
- Micro-Centrifuge for 0.2 mL PCR tubes
- Qubit Assay Tubes (Thermo Fischer Scientific; Q32856)
- Qubit Fluorometer (Life Technologies)
- VWR PCR 8-Tube Strip 0.2 mL (120 Strips) (VWR, 53509-304)
- Agilent TapeStation instrument and the corresponding assay tubes
- Eppendorf DNA LoBind Tubes 1.5 mL (022431021)
- Mixer e.g HulaMixer (Thermo Fischer Scientific), Vortex Mixer (VWR)
- BioRad Thermo Cycler T100
- Agencourt AMPure XP beads (A63880, Beckman Coulter, 5 mL)
- Ethanol 100 % (reagent grade)

➤ Reagents for RNA Extraction

- RNeasy Mini Kit (Qiagen, 74104)

➤ Reagents to assess the quality of the extracted RNA

❖ Reagents for RNA Quantification

- Qubit RNA HS Assay Kit (Thermo Fischer Scientific, Q32852)

❖ **Reagents to examine the RNA profile**

- Agilent RNA ScreenTape Ladder (Agilent; 5067-5578)
- Agilent RNA ScreenTape Sample Buffer (Agilent; 5067-5577)
- Agilent RNA ScreenTape (Agilent; 5067-5576)

➤ **Reagents for cDNA synthesis, cDNA amplification and cDNA library preparation**

❖ **Primers and sequencing adaptors for cDNA synthesis, cDNA amplification and cDNA library preparation:**

- Oxford Nanopore cDNA-PCR Sequencing kit (Oxford Nanopore Technologies; SQK-PCS109)

❖ **Reagents to remove DNA contamination in total RNA samples**

- DNA-free DNA Removal Kit (Thermo Fischer Scientific, M1906)

❖ **Enzymes and reagents for the cDNA synthesis reaction:**

- RNaseOUT 40 U/ μ L (Thermo Fischer Scientific, 10777019)
- 10 mM dNTP solution (NEB, N0447S)
- Water nuclease free PCR grade (eg. Affymetrix, 901578)
- Maxima H Minus Reverse Transcriptase (200 U/ μ L) with 5x RT buffer (Thermo Fisher Scientific, EP0752)

❖ **Reagents to spike RNA molecules of known abundance inside the sample RNA:**

- ERCC RNA Spike-In Mix (Thermo Fisher Scientific, 4456740)
- SIRV-Set 3 (Iso Mix E0 / ERCC) (Lexogen, 051.01)
- THE RNA Storage Solution (Thermo Fischer Scientific, AM7000)

❖ **Enzymes for cDNA amplification:**

- LongAMP Taq 2X Master Mix (NEB, M0287S)
- Exonuclease I (20U/ul; NEB, M0293S)

❖ **Reagents for cDNA quantification:**

- Qubit HS DNA Assay Kit (Thermo, Q32851)

❖ **Reagents to examine the cDNA profile:**

- Agilent D5000 ScreenTape (Agilent, 5067-5588)
- Agilent D5000 Reagents (Agilent, 5067-5589)
- Agilent D5000 Ladder (Agilent, 5067-5590)

➤ **Reagents for cDNA sequencing on the MinION platform**

❖ **ONT MinION Sequencing :**

- MinION SpotON FLO-MIN106 flow cells (R9.4.1)
- MinION Mk1b
- Flow Cell Wash Kit (EXP-WSH002)

Preparation of the working area

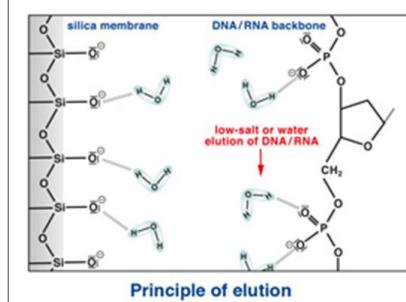
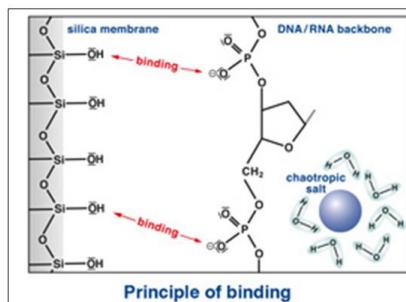
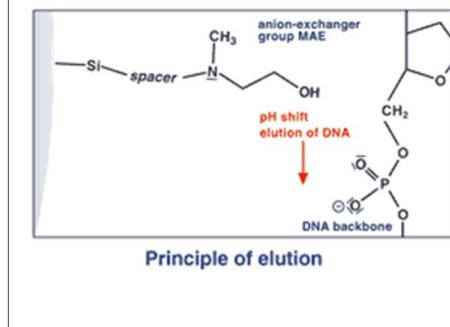
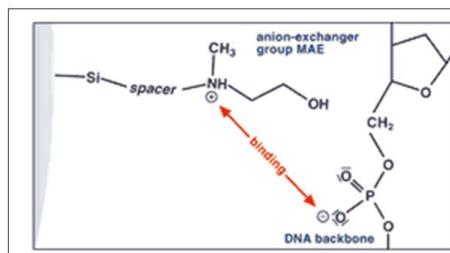
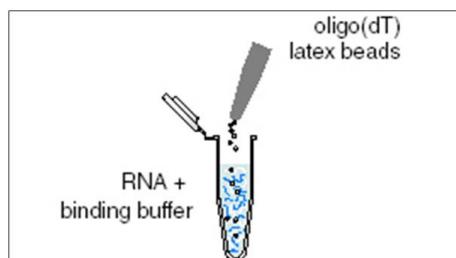
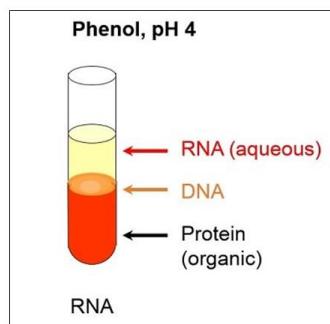
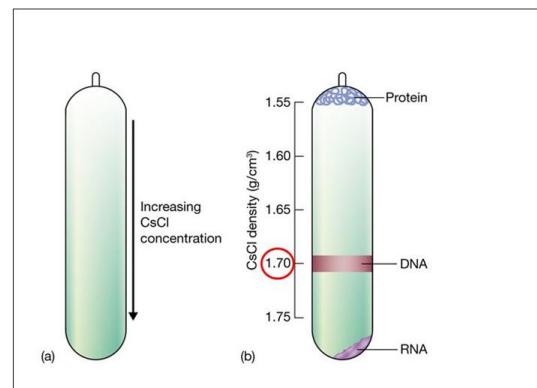
1. Clean all the surfaces and equipment with RNase Zap and maintain an RNase-free bench.
2. Thaw the RNA samples on ice.

RNA extraction

Several kits are available for isolating total or poly A+ RNA from a variety of sources.

- There are the following three chromatographic approaches that exploit the physicochemical properties of RNA to specifically enrich for RNA molecules over the DNA ones (Figure 3).
 - I. **Adsorption chromatography:** The RNA is precipitated on a silica membrane in a spin-column under a high-salt and low pH buffer. This buffer does not efficiently precipitate DNA. Kits that are using this approach are the “RNeasy Mini” (Qiagen) and the “NucleoSpin RNA II” (Clontech).
 - II. **Anion exchange chromatography:** Under low pH conditions the positively charged anion exchanger interacts with the negatively charged nucleic acids. The RNA molecules are subsequently eluted under low salt conditions whereas the DNA molecules, as they are longer, they are eluted under high salt conditions. A kit that is using this approach is the “NucleoBond RNA” (Clontech).
 - III. **Affinity chromatography:** Oligo dT covered beads can be used to enrich for polyA+ RNA molecules. A kit that is using this approach is the “NucleoTrap mRNA Mini” (Clontech).
- There are also approaches that do not rely on spin columns or bead-based protocols (Figure 3):
 - I. The RNA can be extracted with the use of a solution of acidic phenol/guanidine isothiocyanate/chloroform. Acidic phenol retains RNA in the aqueous phase (chloroform), but moves DNA into the phenol phase, because the phosphate groups on the DNA are more easily neutralized than those in RNA (DNA is less acidic/has a greater pKa than RNA). The guanidine isothiocyanate deactivates RNases. The solution with the previous components is the TRIzol Reagent (Thermo Fisher Scientific).
 - II. A classical method to extract RNA is the CsCl density gradient separation of RNA from DNA contaminants (Sambrook et al., 1989). RNA is more compacted than DNA. The

buoyant density of RNA is 1.8 g/ cm³ whereas the buoyant density of DNA is 1.7 g/ cm³ so it is easy to separate the two nucleic acids with a CsCl isopycnic centrifugation.

Adsorption chromatography:**Anion exchange chromatography:****Affinity chromatography:****Acidic phenol/guanidine isothiocyanate/chloroform extraction:****CsCl isopycnic centrifugation:****Figure 3. Different RNA extraction methods.**

RNA purity

The purity of RNA is a key factor for successful cDNA synthesis. The presence of residual organics, metal ions, EDTA, salt or nucleases in the RNA sample can have a large impact on downstream applications by inhibiting enzymatic activity or degrading the RNA.

Impurities such as salt or organic contaminants can be removed by repeated ethanol precipitation, subsequent washing with 80% ethanol and the complete removal of all remaining ethanol.

Other contaminants such as DNA, proteins and dyes may also reduce the efficiency of steps in the library preparation.

If the RNA template is from a plant or some other species with high polysaccharide /pigment levels, these glycoproteins might interfere with primer binding sites of RNA during the first-strand cDNA synthesis leading to reduced cDNA yield.

A solution of pure human RNA has the following characteristics:

- The size of the RNA, as assessed on an electrophoresis instrument (gel, Agilent Tapestation, Agilent Bioanalyzer, PerkinElmer LabChip GXII etc), is on average ~2 kb
- The quality of the RNA as assessed by Nanodrop (for samples with concentration >20 ng/ μ l), shows:
 - An absorbance 260/280 ratio of ~2.0 . A 260/280 which is lower than ~2.0 indicates the presence of DNA. A 260/280 which is lower than ~1.8 can indicate the presence of protein or phenol.
 - An absorbance 260/230 ratio of 2.0-2.2 . If the 260/230 is significantly lower than 2.0-2.2 indicates the presence of contaminants, and the RNA may need additional purification.

RNA quantification

Purpose: Quantify the amount of the isolated RNA.

Process overview: RNA is quantified on a Qubit fluorometric device using the Qubit RNA HS assay. The fluorometer is a device that measures fluorescence. For RNA, the excitation of the dye is at 630 nm and the emission is recorded at 680 nm.

General information: The assay is highly selective for RNA over double-stranded DNA (dsDNA) and is accurate for initial sample concentrations from 250 pg/ μ L to 100 ng/ μ L. Common contaminants such as salts, free nucleotides, solvents, detergents, or protein are well tolerated in the assay. The Qubit RNA HS assay is intended for total RNA, rRNA, or large mRNA molecules but not for small RNA.

To determine the purity of the sample (amount of DNA contamination) we use the Qubit RNA HS Assay Kit together with the Qubit dsDNA HS Assay Kit. These measurements give a much better indication of sample purity than the ones produced by measuring the A260/A280 ratio. We can also measure protein contamination in nucleic acid samples by running the Qubit Protein Assay.

Kit components:

Material	Amount		Concentration	Storage	Stability
	Q32852 (100 assays)	Q32855 (500 assays)			
Qubit® RNA HS Reagent (Component A)	250 µL	1.25 mL	200X concentrate in DMSO	<ul style="list-style-type: none"> • Room temperature • Desiccate • Protect from light 	When stored as directed, kits are stable for 6 months.
Qubit® RNA HS Buffer (Component B)	50 mL	250 mL	Not applicable	Room temperature	
Qubit® RNA HS Standard #1 (Component C)	1 mL	5 mL	0 ng/µL in TE buffer	<4°C	When stored as directed, kits are stable for 6 months.
Qubit® RNA HS Standard #2 (Component D)	4 × 250 µL	10 × 500 µL	10 ng/µL in TE buffer		

Figure 4. Reagents in the “Qubit RNA HS Assay Kit”. The red shadowed boxes contain information on the concentration and on the storage conditions for each one of the reagents.



Figure 5. The different reagents present in the “Qubit RNA HS Assay Kit”.

Procedure: We follow the “Qubit RNA HS Assay Kit” manual (https://assets.thermofisher.com/TFS-Artifacts/LSG/manuals/Qubit_RNA_HS_Assay_UG.pdf):

1. Prepare your sample and the standards to calibrate the Qubit Fluorometer. Set up the required number of Qubit assay tubes for standards and samples. The “Qubit RNA HS Assay” requires 2 standards.
2. Label the tube lids. Do not label the side of the tube as this could interfere with the sample read.
3. Label the lid of each standard tube as follows “ST1” and “ST2”. The tubes correspond to “standard 1” (The nucleic acid concentration for this standard is 0 ng/ul) and “standard 2” (The nucleic acid concentration for this standard is 10 ng/ul) respectively. Calibration of the Qubit Fluorometer requires the standards to be inserted into the instrument in the right order (first “standard 1” and then “standard 2”).
4. Prepare the buffer with the Qubit dye (we call it “Qubit working solution”) by diluting the “Qubit RNA HS Reagent” 1:200 in “Qubit RNA HS Buffer” (see Figure 5).
 - Prepare sufficient “Qubit working solution” to accommodate all standards and samples. For example, for 8 samples, prepare enough working solution for the 8 samples and 2 standards. Assume also that we have 1 more sample to avoid running out of “Qubit working solution” buffer.
 - The final volume of the “Qubit working solution” in each tube must be 200 μ L. Each standard tube requires 190 μ L of the “Qubit working solution”, and each sample tube requires anywhere from 180–199 μ L of the “Qubit working solution”.
 - The total volume of the “Qubit RNA HS Reagent” and “Qubit RNA HS Buffer” reagents to use is calculated as follows. In an Eppendorf tube or in a 15ml falcon tube prepare the following volumes:
 - Total volume of “Qubit RNA HS Buffer” =
(number of reactions + one extra reaction) * (200ul of “Qubit RNA HS Buffer”)

- Total volume of “Qubit RNA HS Reagent” dye =
(number of reactions + one extra reaction) * (1ul of “Qubit RNA HS Reagent”)
 - Mix the “Total volume of “Qubit RNA HS Buffer” ” with the “Total volume of “Qubit RNA HS Reagent” dye “ and vortex briefly.
5. Add 190 µL of “Qubit working solution” to each of the tubes used for standards.
 6. Add 10 µL of each Qubit standard to the appropriate tube, then mix by vortexing 2–3 seconds. Be careful not to create bubbles.
 7. Add Qubit working solution to individual assay tubes so that the final volume in each tube after adding sample is 200 µL. Your sample can be anywhere from 1–20 µL. Add a corresponding volume of Qubit working solution to each assay tube: anywhere from 180–199 µL.
 8. Add each sample to the assay tubes containing the correct volume of Qubit working solution, then mix by vortexing 2–3 seconds.
 9. Allow all tubes to incubate at room temperature for 2 minutes.
 10. Proceed to measure the RNA quantity on a “Qubit Fluorometer” device. (Notice: The instructions below can differ depending on the model type of the instrument)
 11. On the Home screen of the Qubit Fluorometer, press “RNA”, then select “RNA High Sensitivity” as the assay type.
 12. The Standards screen is displayed.
 13. On the Standards screen, press “Yes” to read the standards.

14. Insert the tube containing Standard #1 into the sample chamber, close the lid, then press “Read”. When the reading is complete (~3 seconds), remove Standard #1. The screen must show that the concentration of Standard #1 is 0 pg/ul .
15. Insert the tube containing Standard #2 into the sample chamber, close the lid, then press “Read”. When the reading is complete, remove Standard #2. The screen must show that the concentration of Standard #2 is 500 pg/ul .
16. When the calibration is complete, the instrument displays the “Sample” screen.
17. Insert a sample tube into the sample chamber, close the lid, then press “Read”. When the reading is complete (~3 seconds), remove the sample tube.
18. The instrument displays the results on the “Sample” screen. The value displayed corresponds to the concentration after your sample was diluted into the assay tube. To find the concentration of your original sample, you can record this value and perform the calculation as follows:

$$\text{Concentration of your sample} = \text{QF value} \times \frac{200}{x}$$

where:

QF value = the value given by the Qubit Fluorometer

x = the number of microliters of sample added to the assay tube

and “200” is the total volume (ul) in the measurement tube

Assess DNA contamination in the RNA extraction

Purpose: Measure the amount of DNA contamination in the extracted RNA.

Process Overview: DNA is quantified on a Qubit fluorometric device using the Qubit dsDNA HS assay. For dsDNA the excitation of the dye is at 485 nm and the emission is measured at 530 nm.

General information: The assay is highly selective for double-stranded DNA (dsDNA) over RNA and is accurate for initial sample concentrations from 10 pg/ μ L to 100 ng/ μ L.

Kit components:

Material	Amount		Concentration	Storage	Stability
	Q32851 (100 assays)	Q32854 (500 assays)			
Qubit® dsDNA HS Reagent (Component A)	250 µL	1.25 mL	200X concentrate in DMSO	<ul style="list-style-type: none"> • Room temperature • Desiccate • Protect from light 	When stored as directed, kits are stable for 6 months.
Qubit® dsDNA HS Buffer (Component B)	50 mL	250 mL	Not applicable	Room temperature	
Qubit® dsDNA HS Standard #1 (Component C)	1 mL	5 mL	0 ng/µL in TE buffer	≤4°C	
Qubit® dsDNA HS Standard #2 (Component D)	1 mL	5 mL	10 ng/µL in TE buffer		

Figure 6. Reagents in the “Qubit dsDNA HS Assay Kit”. The red shadowed boxes contain information on the concentration and on the storage conditions for each one of the reagents.

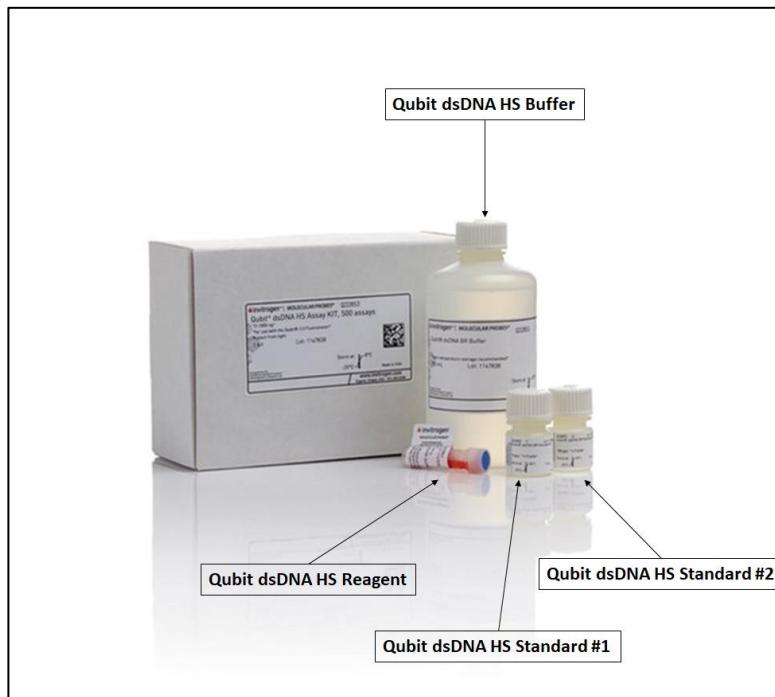


Figure 7. The different reagents present in the “Qubit dsDNA HS Assay Kit”

Procedure: The procedure is the same as the “Qubit RNA HS Assay Kit”. In the “Qubit dsDNA HS Assay Kit” we replace the “Qubit RNA HS Reagent” and the “Qubit RNA HS Buffer” with the “Qubit dsDNA HS Reagent” and the “Qubit dsDNA HS Buffer” respectively (Figure 7). In the Qubit fluorometer device we select the “DNA” tab in the Home screen. Then we select “dsDNA High Sensitivity” as the assay type.

The contaminated DNA can be as low as few ngs/ul or it can be as much as the amount of the extracted RNA. If it is far more than the total amount of the extracted RNA then either the RNA extraction is not successful or the RNA is degraded.

Assess the profile of the extracted RNA

Purpose: Asses the quality of the extracted RNA

Process overview: Total RNA is run on an Agilent Tapestation to assess the RNA profile.

General information:

➤ *Methods to assess the integrity of total RNA*

The integrity of total RNA can be visually assessed by the ratio of 28S:18S RNA on a denaturing formaldehyde agarose gel by staining either with Ethidium Bromide or with SYBR Green or with SYBR Gold (1-2 ng of RNA detection limit). The theoretical 28S:18S ratio for eukaryotic RNA is approximately 2:1. For mammalian total RNA, you should observe two bright bands at approximately 4.5 and 1.9 kb; these bands represent 28S and 18S ribosomal RNA. The ratio of intensities of these bands should be 1.5–2.5:1 .

Except from the classical denaturing/native agarose gels, there are multiple commercial instruments that let you assess the profile of the RNA library with minimal amount of material used (Agilent Bioanalyzer, Agilent Tapestation etc). In addition to assessing RNA quality, these automated systems provide a good estimation of RNA concentration. Here we use an Agilent TapeStation instrument. The quantitative range of the instrument is 25 – 500 ng/µl whereas the sensitivity of the instrument (minimal amount of RNA that can be loaded) is 5 ng/µl. The linear range of size estimation is between 100 – 6000 nt. The RIN estimation is accurate for amounts 25– 500 ng/µl.

The integrity of the total RNA can also be assessed by synthesizing cDNA. The size distribution of the cDNA is expected to reflect the size distribution of intact mRNA (see next paragraph).

➤ *Methods to assess the integrity of the mRNA*

All of the methods mentioned above can be used to assess the quality of the mRNA. However, because mRNA does not contain strong ribosomal bands, the assessment of its quality will be somewhat

subjective. Typically mRNA, after ribosomal RNA depletion, appears as a smear between 0.5 kb to 6 kb, with an area of higher intensity around 1.5 and 2 kb. This size distribution may be tissue or species-specific. If the average size of the mRNA is lower than 1.5 kb, it could be an indication of degradation.

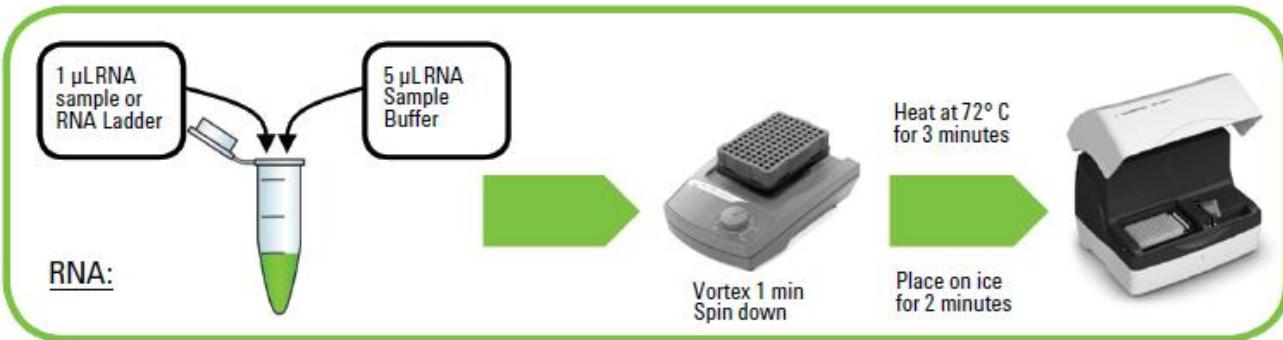


Figure 8. Overview of the quantification of RNA in the Tapestation instrument using the Agilent RNA screentape.

Procedure: We follow the Agilent RNA screentape manual (Figure 8;

https://www.agilent.com/cs/library/usermanuals/Public/ScreenTape_RNA_QG.pdf :

1. Allow reagents to equilibrate at room temperature for 30 min.
2. Vortex mix before use.
3. Thaw total RNA samples on ice.
4. If running ladder, prepare by mixing 5 µL RNA Sample Buffer (green colored tube cap) with 1 µL RNA Ladder (yellow colored tube cap). Mix up and down gently, at least 5 times, with the pipette.
5. Prepare sample by mixing 5 µL RNA Sample Buffer (green colored tube cap) with 1 µL RNA sample. Mix up and down gently, at least 5 times, with the pipette.
6. Ladder/Sample denaturation:
 - Heat ladder and samples at 72 °C for 3 min in a thermocycler. Keep the thermocycler lid temperature high to avoid evaporation of the sample.
 - Place ladder and samples on ice for 2 min

7. Transfer the solution from the PCR tube to the TapeStation strip tubes. Spin down to position the sample at the bottom of the tube.
8. Load samples into the TapeStation instrument.
9. Select the required samples on the TapeStation Controller Software.
10. Click Start and specify a filename with which to save your results.

Expected results:

Examples of the expected electropherogram profile of high-quality RNA is presented in Figure 9. The RNA integrity number (RIN) is an algorithm, developed by Agilent, for assigning integrity values to RNA measurements. It uses different features from an electrophoretic trace (area and shape of pre-, 5S-, fast-, inter-, precursor-, post-region, marker, 18S, 28S regions; Figure 10) to classify a sample as degraded or not degraded. The RIN should usually be greater than 8 (>8) or in some cases greater than 7 (>7) otherwise the molecules are assumed that are partially degraded. The degradation of RNA molecules starts at the 5' end, so we expect that the 5' coverage will be lower than the 3' coverage in a degraded sample relative to an intact one. For sequencing technologies that do not rely on full length sequencing, partially degraded samples can be sequenced with short-read sequencing technologies.

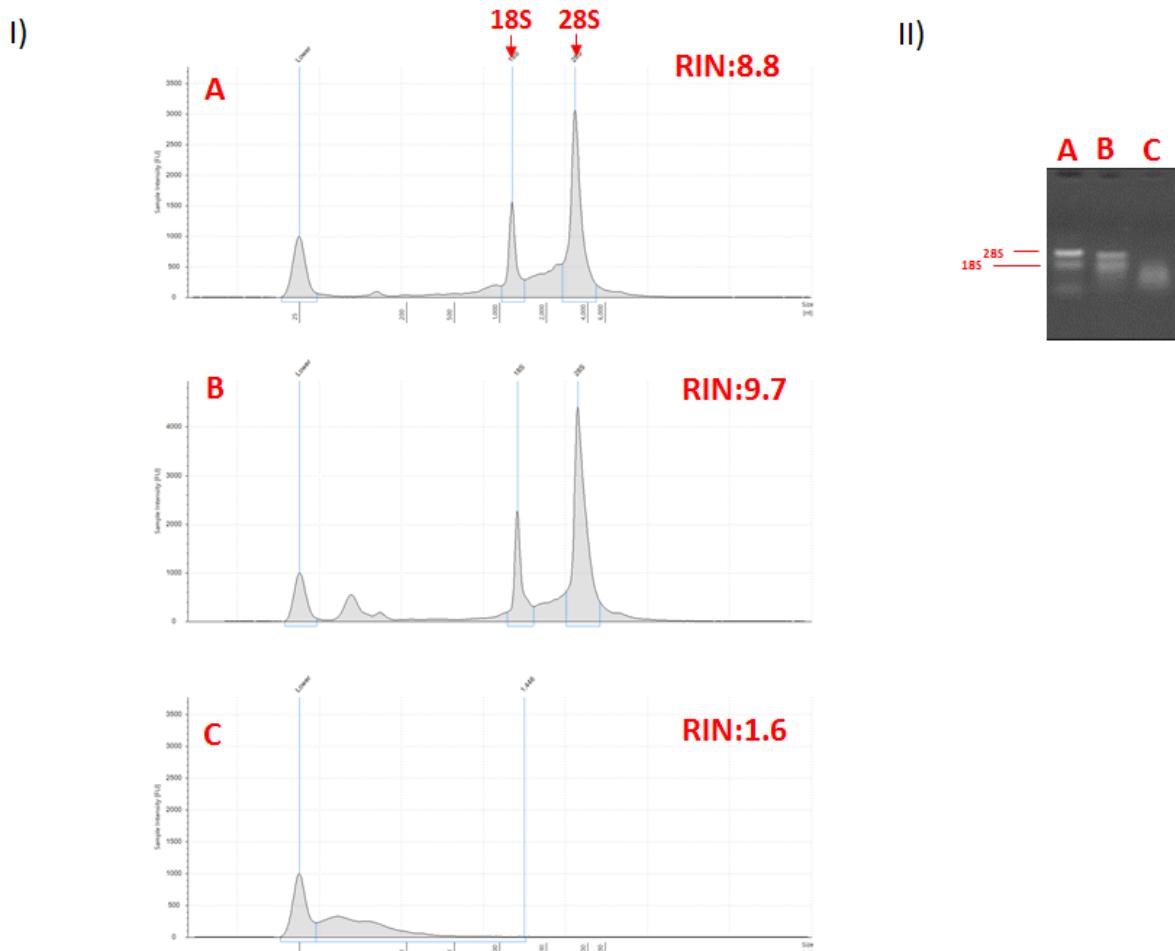


Figure 9. Electropherogram of total RNA profiles using either an Agilent Tapestation or a denaturing agarose gel. I) Total RNA profiles in the Agilent Tapestation instrument. The expected RIN value (the ratio of the 28S area to the 18S area as calculated from the provided software) for a good quality RNA (A, B) should be above 8 (or in some cases above 7). A degraded total RNA sample along with the corresponding RIN value is also presented (C). **II)** Denaturing (1.11% formaldehyde) agarose gel electrophoresis of the corresponding total RNA samples from (I).

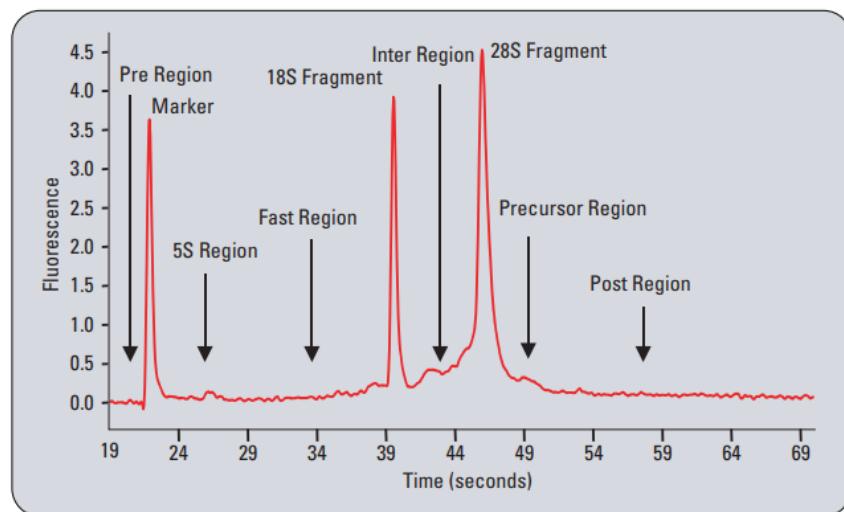


Figure 10. Electropherogram detailing the regions that are indicative of RNA quality. (From <https://www.agilent.com/cs/library/applications/5989-1165EN.pdf>)

Removal of DNA contamination from total RNA (optional)

Purpose: Remove any DNA contamination present in the total RNA sample.

Process Overview: The total RNA is treated with DNase. Then the DNase is precipitated out of solution to prevent the enzyme from causing the degradation of the synthesized cDNA.

General information: The DNase treatment is used to remove any possible DNA contamination present in the total RNA sample. The cytosine terminal transferase activity of the reverse transcriptase favors RNA/cDNA hybrids. This indicates that the DNA is not going to be amplified during the cDNA amplification step as it will lack both the poly-T sequence and the Template Switching Oligo (TSO) sequence. It can nevertheless be sequenced. For example, assuming that in 300 ngs of total RNA the contaminated DNA is 150 ngs and overall from the cDNA synthesis reaction 3 ug of cDNA were produced then we expect 5% of the sequenced reads to correspond to genomic DNA. We note here that we can always filter out genomic reads by focusing only on reads that have present at their end sequence both the adaptors used during cDNA synthesis (poly-T oligo and TSO oligo).

Kit contents:

Figure 11. The different reagents present in the in the “DNA-free DNA Removal Kit”.

Procedure: We follow the DNA-free DNA Removal Kit (https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_055739.pdf) manual:

1. Add the following reagents to a microcentrifuge tube having a total final volume of 41 μ l. We aim at the end of this procedure to have a solution with a concentration of 200 ngs of RNA per ul :

Reagent	Amount per reaction
10X DNase I Buffer	4.1 μ l
rDNase I (2 Units/ μ L)	3 μ l
RNA (10 ug; >2 μ g DNA contamination)	Variable
Water	Up to 41 μ l
Total volume	41μl

2. Incubate at 37°C for 30 minutes.
3. Add 8.2 μ l of DNase Inactivation reagent.
4. Incubate at room temperature for 2 minutes. Mix the tube occasionally 2-3 times.
5. Centrifuge at 10,000g for 1.5 minutes and transfer the supernatant, which contains the RNA, to a new tube.

Spike-In RNA

Purpose: Assess the absolute number of RNA molecules for every isoform/gene detected in the sample.

Process overview: A known amount of different types of RNA molecules are spiked inside the total RNA sample.

General information: We want to assess the absolute number of RNA molecules per isoform or gene after sequencing. For this we spike different types of RNA molecules of known abundance inside the total RNA samples. After sequencing, a standard curve is created between the known abundance of spike-in RNAs and the relative abundance as calculated from the sequencing results.

The amount of spiked RNAs depends on the fraction of mRNA molecules over the total RNA molecules. The spiked RNAs need to be introduced in the minimal amount so that they do not overtake the sequencing reaction. Additionally, the spiked RNAs amount needs to be adequate enough so that the most abundant mRNA isoform is not more abundant than the most abundant spiked RNA molecule.

Important notice: The addition of spike-ins depends on the experimental design and not all the studies need them. For example, for the majority of RNA-seq studies we only examine the relative expression levels of the different genes/isoforms between two conditions. In this case spike-ins are not usually used. Only in studies where there is biological evidence that the total amount of mRNA levels is affected, for example the early developmental stages of multiple organisms⁴⁻⁶, then the absolute expression levels for each gene/isoform can be examined across the developmental stages. Additionally, in single cell transcriptome studies, due to the small amount of starting material, the introduction of spike-ins can be used to control for technical variability between the individual single-cell cDNA synthesis reactions. In the single cell studies the spike-ins can be used to normalize the data across the individual cells.

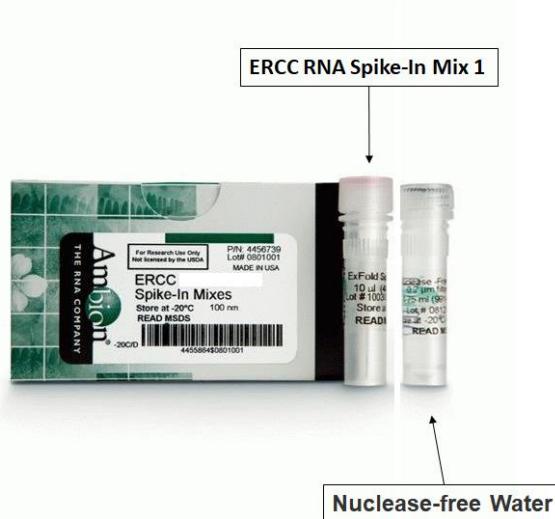
Kit contents:

Figure 12. The different reagents present in the “ERCC RNA Spike-In Mix 1 kit”.

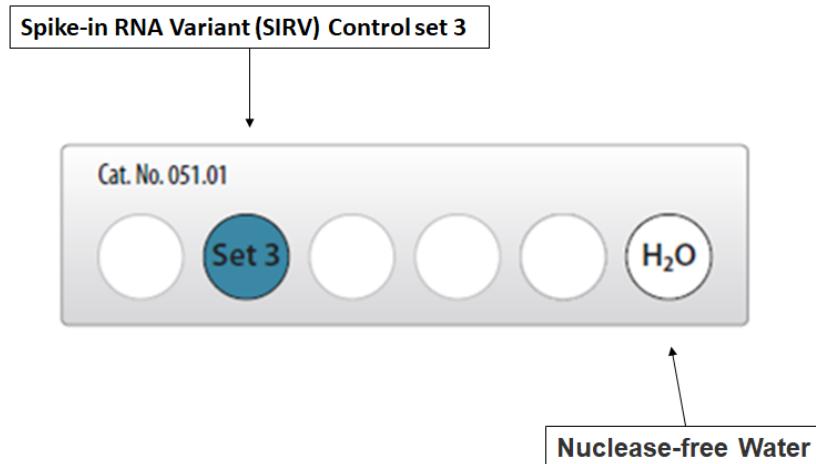


Figure 13. The different reagents present in the “Spike-in RNA Variant Control set 3 kit”.

Procedure:

- A known amount of RNA molecules is spiked in the total RNA sample. We can use two different types of spikes:
 1. ERCC RNA Spike-In Mix 1
 2. Spike-in RNA Variant (SIRVs) Control set 3 kit
- The identification and quantification of the different types of RNA molecules in the “ERCC RNA Spike-In Mix 1” can give us the following information:
 - Absolute quantification of the abundance of mRNA isoforms or genes in the sample.
 - Validates that the bioinformatics pipeline can identify 1 isoform per ERCC gene.
- The identification and quantification of the Spike-in RNA Variant (SIRVs) spiked RNAs:
 - Validates that the bioinformatics pipeline can identify the expected isoforms per SIRV gene.
- The amount of spiked RNA ($\text{mass}_{\text{spiked RNA}}$) that is going to be added in the reaction mix can be calculated as follows:

$$\text{mass}_{\text{spiked RNA}} = \text{fraction}_{\text{spiked reads}} \times \text{fraction}_{\text{target RNA}} \times \text{mass}_{\text{RNAinput}}$$

where:

mass_{spiked RNA}: mass (ngs) of spike-in RNA (SIRVs or ERCC) to be added in the sample.

fraction_{spiked reads}: desired fraction of sequenced spike-in RNA reads relative to the total amount of sequenced reads.

fraction_{target RNA}: fraction of the total RNA used in the sample, that is going to be synthesized into cDNA molecules.

mass RNA input: mass (ngs) of RNA input per sample.

- Then the volume (ul) of spike-in RNA to be used is calculated as follows:

$$\text{volume}_{\text{spike-in RNA}} = \frac{\text{mass}_{\text{spike-in RNA}}}{\text{concentration}_{\text{spike-in RNA}}}$$

where

concentration_{spike-in RNA}: concentration (ngs/ul) of the spike-in RNAs solution.

volume_{spike-in RNA}: volume (ul) from the spike-in RNAs solution to be added into the sample.

- A list with the concentrations of the different types of RNA molecules in the “ERCC stock mix 1” can be found in Figure 14 and in the webpage:

https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_095046.txt

- A list with the concentrations of the different types of RNA molecules in the “Spike-in RNA Variant (SIRVs) Control set 3 kit” can be found in Figure 15 and in the webpage:

<https://www.lexogen.com/sirvs/downloads/>

Re-sort ID	ERCC ID	GenBank*	subgroup	conc (amoles/μl)	Mw (g/mol)	conc (ng/μl)	length (nt)	GC
1	ERCC-00130	EF011072	A		30000	346,815.6	10,4044680	1073 46.0%
2	ERCC-00004	DQ516752	A		7500	169,827.2	1,2737040	531 34.7%
3	ERCC-00136	EF011063	A		1875	337,910.4	0.6335820	1047 41.7%
4	ERCC-00108	DQ668365	A		937.5	331,036.0	0.3103463	1030 49.0%
5	ERCC-00116	DQ668367	A		468.75	642,597.8	0.3012177	1999 50.1%
6	ERCC-00092	DQ459425	A		234.375	364,327.4	0.0853892	1132 50.2%
7	ERCC-00095	DQ516759	A		117.1875	170,855.0	0.0200221	535 37.8%
8	ERCC-00131	DQ855003	A		117.1875	250,887.8	0.0294009	779 47.5%
9	ERCC-00062	DQ516786	A		58.59375	331,116.2	0.0194013	1031 30.9%
10	ERCC-00019	DQ883651	A		29.296875	210,154.4	0.0061569	652 48.8%
11	ERCC-00144	DQ854995	A		29.296875	176,015.2	0.0051567	546 45.6%
12	ERCC-00170	DQ516773	A		14.6484375	333,074.2	0.0048790	1031 33.9%
13	ERCC-00154	DQ854997	A		7.32421875	175,929.0	0.0012885	545 50.1%
14	ERCC-00085	DQ883669	A		7.32421875	273,934.4	0.0020064	852 48.6%
15	ERCC-00028	DQ459419	A		3.66210938	368,832.8	0.0013507	1144 50.7%
16	ERCC-00033	DQ516796	A		1.83105469	656,082.2	0.0012013	2036 33.0%
17	ERCC-00134	DQ516739	A		1.83105469	91,205.4	0.0001670	282 31.9%
18	ERCC-00147	DQ516790	A		0.91552734	333,736.2	0.0003055	1031 36.6%
19	ERCC-00097	DQ516758	A		0.45776367	169,800.2	0.0000777	531 36.3%
20	ERCC-00156	DQ883643	A		0.45776367	161,810.4	0.0000741	502 49.0%
21	ERCC-00123	DQ516782	A		0.22888184	327,523.0	0.0000750	1030 35.7%
22	ERCC-00017	DQ459420	A		0.11444092	369,653.8	0.0000423	1144 51.4%
23	ERCC-00083	DQ516780	A		0.02861023	329,870.2	0.0000094	1036 34.8%
24	ERCC-00096	DQ459429	B		15000	359,177.0	5.3876550	1115 50.9%
25	ERCC-00171	DQ854994	B		3750	165,633.6	0.6211260	513 47.8%
26	ERCC-00009	DQ668364	B		937.5	321,807.0	0.3016941	1000 47.3%
27	ERCC-00042	DQ516783	B		468.75	328,361.2	0.1539193	1031 39.3%
28	ERCC-00060	DQ516763	B		234.375	170,806.2	0.0400327	531 31.5%
29	ERCC-00035	DQ459413	B		117.1875	366,989.6	0.0430066	1138 51.3%
30	ERCC-00025	DQ883689	B		58.59375	643,552.4	0.0377081	2002 50.1%
31	ERCC-00051	DQ516740	B		58.59375	90,967.4	0.0053301	282 34.8%
32	ERCC-00053	DQ516785	B		29.296875	330,582.2	0.0096850	1031 31.3%
33	ERCC-00148	DQ883642	B		14.6484375	162,522.4	0.0023807	502 48.6%
34	ERCC-00126	DQ459427	B		14.6484375	363,992.4	0.0053319	1132 50.6%
35	ERCC-00034	DQ855001	B		7.32421875	330,750.4	0.0024225	1027 49.2%
36	ERCC-00150	DQ883659	B		3.66210938	241,739.2	0.0008853	751 47.0%
37	ERCC-00067	DQ883653	B		3.66210938	210,062.4	0.0007693	652 47.2%
38	ERCC-00031	DQ459431	B		1.83105469	368,343.2	0.0006745	1146 47.8%
39	ERCC-00109	DQ854998	B		0.91552734	175,536.8	0.0001607	544 45.6%
40	ERCC-00073	DQ668358	B		0.91552734	196,569.2	0.0001800	611 47.1%
41	ERCC-00158	DQ516795	B		0.45776367	335,281.2	0.0001535	1041 34.6%
42	ERCC-00104	DQ516815	B		0.22888184	651,918.2	0.0001492	2036 32.9%
43	ERCC-00142	DQ883646	B		0.22888184	161,701.2	0.0000370	501 49.9%
44	ERCC-00138	DQ516777	B		0.11444092	331,846.8	0.0000380	1031 33.0%
45	ERCC-00117	DQ459412	B		0.05722046	368,368.8	0.0000211	1144 51.1%
46	ERCC-00075	DQ516778	B		0.01430512	328,053.2	0.0000047	1031 36.4%
47	ERCC-00074	DQ516754	C		15000	170,151.0	2.5522650	530 34.9%

(the table continues in the next page)

Re-sort ID	ERCC ID	GenBank*	subgroup	conc (amoles/µl)	Mw (g/mol)	conc (ng/µl)	length (nt)	GC	
48	ERCC-00113	DQ883663	C	3750	273,308.6	1.0249073	848	50.5%	
49	ERCC-00145	DQ875386	C	937.5	338,791.0	0.3176166	1050	44.4%	
50	ERCC-00111	DQ883685	C	468.75	321,970.4	0.1509236	1002	47.1%	
51	ERCC-00076	DQ883650	C	234.375	209,048.0	0.0489956	650	49.7%	
52	ERCC-00044	DQ459424	C	117.1875	376,895.0	0.0441674	1170	50.1%	
53	ERCC-00162	DQ516750	C	58.59375	169,020.2	0.0099035	531	36.3%	
54	ERCC-00071	DQ883654	C	58.59375	208,727.0	0.0122301	650	48.5%	
55	ERCC-00084	DQ883682	C	29.296875	323,056.4	0.0094645	1002	50.3%	
56	ERCC-00099	DQ875387	C	14.6484375	440,892.0	0.0064584	1370	41.6%	
57	ERCC-00054	DQ516731	C	14.6484375	91,577.4	0.0013415	282	37.6%	
58	ERCC-00157	DQ839618	C	7.32421875	331,246.4	0.0024261	1027	50.1%	
59	ERCC-00143	DQ668362	C	3.66210938	254,316.4	0.0009313	792	48.6%	
60	ERCC-00039	DQ883656	C	3.66210938	240,933.6	0.0008823	748	48.9%	
61	ERCC-00058	DQ459418	C	1.83105469	369,159.8	0.0006760	1144	50.2%	
62	ERCC-00120	DQ854992	C	0.91552734	175,216.8	0.0001604	544	48.3%	
63	ERCC-00040	DQ883661	C	0.91552734	242,349.4	0.0002219	752	52.7%	
64	ERCC-00164	DQ516779	C	0.45776367	327,370.0	0.0001499	1030	37.2%	
65	ERCC-00024	DQ854993	C	0.22888184	175,739.8	0.0000402	544	46.3%	
66	ERCC-00016	DQ883664	C	0.22888184	276,231.6	0.0000632	858	48.6%	
67	ERCC-00012	DQ883670	C	0.11444092	322,874.4	0.0000370	1002	50.9%	
68	ERCC-00098	DQ459415	C	0.05722046	371,581.2	0.0000213	1151	51.4%	
69	ERCC-00057	DQ668366	C	0.01430512	332,835.0	0.0000048	1035	49.9%	
70	ERCC-00002	DQ459430	D	15000	343,773.8	5.1566070	1069	51.4%	
71	ERCC-00046	DQ516748	D	3750	170,699.0	0.6401213	530	35.1%	
72	ERCC-00003	DQ516784	D	937.5	330,141.2	0.3095074	1031	32.8%	
73	ERCC-00043	DQ516787	D	468.75	332,733.2	0.1559687	1031	32.9%	
74	ERCC-00022	DQ855004	D	234.375	243,789.8	0.0571382	759	47.3%	
75	ERCC-00112	DQ459422	D	117.1875	367,543.8	0.0430715	1144	47.1%	
76	ERCC-00165	DQ668363	D	58.59375	282,400.0	0.0165469	880	50.0%	
77	ERCC-00079	DQ883652	D	58.59375	210,368.4	0.0123263	652	49.4%	
78	ERCC-00078	DQ883673	D	29.296875	322,705.2	0.0094543	1001	50.0%	
79	ERCC-00163	DQ668359	D	14.6484375	177,560.2	0.0026010	551	46.6%	
80	ERCC-00059	DQ668356	D	14.6484375	171,361.6	0.0025102	533	48.0%	
81	ERCC-00160	DQ883658	D	7.32421875	242,048.2	0.0017728	751	45.7%	
82	ERCC-00014	DQ875385	D	3.66210938	637,893.4	0.0023360	1977	44.4%	
83	ERCC-00077	DQ516742	D	3.66210938	90,305.2	0.0003307	281	33.5%	
84	ERCC-00069	DQ459421	D	1.83105469	369,276.0	0.0006762	1145	50.2%	
85	ERCC-00137	DQ855000	D	0.91552734	175,830.0	0.0001610	545	49.5%	
86	ERCC-00013	EF011062	D	0.91552734	264,026.2	0.0002417	816	42.9%	
87	ERCC-00168	DQ516776	D	0.45776367	329,010.4	0.0001506	1032	34.3%	
88	ERCC-00041	EF011069	D	0.22888184	369,490.6	0.0000846	1143	45.6%	
89	ERCC-00081	DQ854991	D	0.22888184	174,934.4	0.0000400	542	48.7%	
90	ERCC-00086	DQ516791	D	0.11444092	333,179.8	0.0000381	1034	32.6%	
91	ERCC-00061	DQ459426	D	0.05722046	369,065.8	0.0000211	1144	50.1%	
92	ERCC-00048	DQ883671	D	0.01430512	322,722.0	0.0000046	1000	47.7%	
				c (amoles/µl)	Mw (g/mol)	conc (ng/µl)	length (nt)	GC	
				average	1,125.163	292,410	0.3294488	909	44.2%
				median	7.324	322,798	0.0023584	1,002	47.3%
				minimum	0.014	90,305	0.0000046	281	30.9%
				maximum	30,000.000	656,082	10.4044680	2,036	52.7%
				sum	103,515.029		30.3092852		

Figure 14 (previous two pages). Abundance of the different types of RNA molecules in the “ERCC RNA Spike-In Mix 1” stock solution.

SIRV gene	SIRV ID	GenBank	conc (amoles/μl)	Mw (g/mol)	conc (ng/μl)	length (nt)	GC
SIRV1	SIRV101	KX147759.1	1000	512,587.2	0.51	1591	45.8%
	SIRV102	KX147759.1	1000	429,039.0	0.43	1330	44.7%
	SIRV103	KX147759.1	1000	449,212.6	0.45	1393	44.8%
	SIRV105	KX147759.1	1000	225,755.0	0.23	700	44.0%
	SIRV106	KX147759.1	1000	323,127.6	0.32	1003	45.1%
	SIRV107	KX147759.1	1000	247,864.8	0.25	774	44.8%
	SIRV108	KX147759.1	1000	236,291.4	0.24	732	45.6%
	SIRV109	KX147759.1	1000	160,245.8	0.16	494	45.1%
	SIRV104					1429	44.6%
	SIRV110					424	46.9%
	SIRV111					1307	44.9%
	SIRV112					1454	45.5%
SIRV2	SIRV201	KX147760.1	1000	671,531.2	0.67	2081	42.2%
	SIRV202	KX147760.1	1000	646,043.2	0.65	2001	42.2%
	SIRV203	KX147760.1	1000	230,767.2	0.23	716	41.1%
	SIRV204	KX147760.1	1000	248,226.0	0.25	770	41.7%
	SIRV205	KX147760.1	1000	176,875.6	0.18	553	42.5%
	SIRV206	KX147760.1	1000	145,954.8	0.15	454	40.1%
	SIRV207					2138	42.6%
	SIRV208					972	41.3%
	SIRV209					1052	42.5%
	SIRV301	KX147761.1	1000	805,207.4	0.81	2497	35.1%
SIRV3	SIRV302	KX147761.1	1000	592,606.4	0.59	1837	35.2%
	SIRV303	KX147761.1	1000	660,393.6	0.66	2048	35.0%
	SIRV304	KX147761.1	1000	359,013.6	0.36	1113	34.2%
	SIRV305	KX147761.1	1000	150,136.2	0.15	466	31.5%
	SIRV306	KX147761.1	1000	774,771.6	0.77	2403	36.1%
	SIRV307	KX147761.1	1000	261,206.8	0.26	809	33.6%
	SIRV308	KX147761.1	1000	162,148.8	0.16	509	41.1%
	SIRV309	KX147761.1	1000	264,910.2	0.26	826	42.7%
	SIRV310	KX147761.1	1000	198,577.8	0.20	619	39.3%
	SIRV311	KX147761.1	1000	61,473.2	0.06	191	29.8%
	SIRV312					717	42.3%
	SIRV313					1021	34.1%
	SIRV314					2609	35.9%
	SIRV315					2302	35.1%

(the table continues in the next page)

SIRV gene	SIRV ID	GenBank	conc (amoles/μl)	Mw (g/mol)	conc (ng/μl)	length (nt)	GC
SIRV4	SIRV403	KX147762.1	1000	226,783.0	0.23	700	37.9%
	SIRV404	KX147762.1	1000	201,630.4	0.20	622	38.3%
	SIRV405	KX147762.1	1000	212,884.2	0.21	656	39.6%
	SIRV406	KX147762.1	1000	208,808.4	0.21	647	42.0%
	SIRV408	KX147762.1	1000	194,260.0	0.19	600	35.8%
	SIRV409	KX147762.1	1000	513,296.4	0.51	1597	43.7%
	SIRV410	KX147762.1	1000	314,615.0	0.31	980	43.8%
	SIRV401					2283	38.7%
	SIRV402					2089	37.3%
	SIRV407					2135	38.8%
	SIRV411					1473	43.7%
SIRV5	SIRV501	KX147763.1	1000	619,743.0	0.62	1920	45.2%
	SIRV502	KX147763.1	1000	650,317.8	0.65	2014	46.4%
	SIRV503	KX147763.1	1000	180,068.2	0.18	556	43.0%
	SIRV504	KX147763.1	1000	805,965.6	0.81	2503	50.1%
	SIRV505	KX147763.1	1000	665,158.8	0.67	2059	46.6%
	SIRV506	KX147763.1	1000	187,852.4	0.19	582	51.2%
	SIRV507	KX147763.1	1000	181,593.6	0.18	563	50.1%
	SIRV508	KX147763.1	1000	683,274.0	0.68	2115	46.2%
	SIRV509	KX147763.1	1000	295,115.0	0.30	915	47.1%
	SIRV510	KX147763.1	1000	807,199.8	0.81	2504	47.6%
	SIRV511	KX147763.1	1000	185,877.2	0.19	576	50.9%
	SIRV512	KX147763.1	1000	83,553.8	0.08	259	47.5%
	SIRV513					3186	48.1%
	SIRV514					4657	48.6%
	SIRV515					2360	46.8%
	SIRV516					3020	49.2%
	SIRV517					2238	46.9%
SIRV6	SIRV601	KX147764.1	1000	472,963.0	0.47	1465	42.1%
	SIRV602	KX147764.1	1000	195,110.8	0.20	604	40.7%
	SIRV603	KX147764.1	1000	644,117.8	0.64	1999	35.4%
	SIRV604	KX147764.1	1000	505,240.4	0.51	1567	42.8%
	SIRV605	KX147764.1	1000	360,419.6	0.36	1118	43.3%
	SIRV606	KX147764.1	1000	185,678.0	0.19	575	45.2%
	SIRV607	KX147764.1	1000	194,778.8	0.19	604	46.7%
	SIRV608	KX147764.1	1000	131,524.4	0.13	407	35.1%
	SIRV609	KX147764.1	1000	166,020.0	0.17	515	48.2%
	SIRV610	KX147764.1	1000	385,213.6	0.39	1193	38.6%
	SIRV611	KX147764.1	1000	155,966.8	0.16	484	45.7%
	SIRV612	KX147764.1	1000	502,284.6	0.50	1558	42.6%
	SIRV613	KX147764.1	1000	432,465.2	0.43	1341	38.2%
	SIRV614	KX147764.1	1000	158,277.8	0.16	489	40.3%
	SIRV615	KX147764.1	1000	262,274.6	0.26	813	34.4%
	SIRV616	KX147764.1	1000	181,193.2	0.18	561	44.9%
	SIRV617	KX147764.1	1000	98,599.2	0.10	306	43.5%
	SIRV618	KX147764.1	1000	70,734.8	0.07	219	40.6%
	SIRV619					4674	40.2%
	SIRV620					2481	36.2%
	SIRV621					2749	43.8%
	SIRV622					1063	44.4%
	SIRV623					787	38.0%
	SIRV624					944	37.7%
	SIRV625					1116	39.2%

(the table continues in the next page)

SIRV gene	SIRV ID	GenBank	conc (amoles/μl)	Mw (g/mol)	conc (ng/μl)	length (nt)	GC
SIRV7	SIRV701	KX147765.1	1000	805,337.4	0.81	2492	36.4%
	SIRV702	KX147765.1	1000	736,212.4	0.74	2277	36.8%
	SIRV703	KX147765.1	1000	816,777.6	0.82	2528	36.0%
	SIRV704	KX147765.1	1000	148,304.6	0.15	458	29.5%
	SIRV705	KX147765.1	1000	805,337.4	0.81	2492	36.4%
	SIRV706	KX147765.1	1000	316,573.8	0.32	979	33.3%
	SIRV707	KX147765.1	1000	297,206.8	0.30	919	33.3%
	SIRV708					2356	36.2%
	SIRV709					2890	35.8%
	SIRV710					2570	36.0%
	SIRV711					1129	32.9%
sum							
			c (amoles/μl)	Mw (g/mol)	conc (ng/μl)	length (nt)	GC
			1,000,000	365,805	0.37	1,399	41.2%
			1,000,000	262,275	0.26	1,115	42.2%
			1,000,000	61,473	0.06	191	29.5%
			1,000,000	816,778	0.82	4,674	51.2%
			69,000,000		25.24		

Figure 15 (current and previous two pages). Abundance of the different types of SIRV molecules in the “Spike-in RNA Variant Control (SIRVs) mix set 3” stock solution.

Procedure:

1. We will use the following values in the variables presented in the above formulas:
 - The value for the “fraction spiked reads ” variable is:
 - For the ERCC RNA Spike-In Mix 1, the fraction $\text{spiked reads} = 0.03$ (We target the sequenced reads of the spiked-in RNA to be 3% of the total amount of sequenced reads)
 - For the Spike-in RNA Variant (SIRVs) Control set 3 kit, the fraction $\text{spiked reads} = 0.05$ (We target the sequenced reads of the spiked-in RNA to be 5% of the total amount of sequenced reads. From this fraction 3% is ERCC reads and 2% is SIRV reads)
 - The value for the "fraction target RNA " variable is $\text{fraction target RNA} = 0.01$ (We assume that the polyA+ fraction is 1% of the total RNA. We note here that the polyA+ fraction can be between 1%-5% of total RNA)
 - The value for the “mass RNA input ” is $\text{mass RNA input} = 300 \text{ ngs}$ (We will use in the cDNA synthesis reactions 300 ngs of total RNA)
 - Based on the above values we will need to use:
 - For the ERCC RNA Spike-In Mix 1, the $\text{mass spiked RNA} = 0.09 \text{ ngs}$
 - For the Spike-in RNA Variant (SIRVs) Control set 3 kit, the $\text{mass spiked RNA} = 0.15 \text{ ngs}$
2. The concentration of the stock solutions are:
 - The “ERCC RNA Spike-In Mix 1” tube contains 10 μl of ERCC RNAs at a concentration of 103.515 fmoles/ μl or 30.3 ng/ μl .

- Each tube of the “Spike-in RNA Variant (SIRVs) Control set 3 kit” contains 10 µl of SIRV RNAs and it has 3.78 fmoles/ul of SIRVs RNA and 5.59 fmoles/ul of ERCC RNAs with a total nucleic acid concentration of 3.03 ng/ul .
3. Prepare the appropriate dilution of each Spike-In Mix needed. In the new diluted solution the “mass _{spiked RNA}“ for either the “ERCC RNA Spike-In Mix 1” or the “Spike-in RNA Variant (SIRVs) Control set 3 kit” should correspond to 1 ul of the final diluted volume. So we need to have the following dilutions:
- For the “ERCC RNA Spike-In Mix 1” we are going to dilute 340.6 times the stock solution. So in 339.6 ul of “THE RNA solution” add 1 ul from the “ERCC RNA Spike-In Mix 1” stock solution (new concentration= 0.09 ng/ul). Afterwards we will have to take Volume _{spike-in RNA} = ((0.09 ngs)/(0.09 ng/ul))=1 ul of the diluted solution.
 - For the “Spike-in RNA Variant (SIRVs) Control set 3 kit” we are going to dilute 20.2 times the stock solution. So in 19.2 ul of “THE RNA solution” add 1 ul from the “Spike-in RNA Variant (SIRVs) Control set 3 kit” stock solution (new concentration= 0.15 ng/ul). Afterwards we will have to take Volume _{spike-in RNA} = ((0.15 ngs)/(0.15 ng/ul))=1 ul of the diluted solution.

ONT MinION sequencing kit

The protocol that follows is based on the “cDNA-PCR Sequencing (SQK-PCS109) PCS_9085_v109_revC_04Feb2019” protocol for cDNA synthesis and adapter conjugation. The reagents necessary for the cDNA synthesis and cDNA amplification are inside the “SQK-PCS109 MinION library kit” box and are shown in Figure 16. The DNA sequences of the different primers used in the “SQK-PCS109 MinION library kit” are presented in Figure 17.

The kit is recommended for users who:

- would like to identify and quantify full-length transcripts
- want to explore isoforms, splice variants and fusion transcripts using full-length cDNAs
- have a low starting amount of RNA (*That is the reason that we perform cDNA amplification*)
- would like to generate high amounts of cDNA data
- wish to start from either total RNA or polyA+ RNA

cDNA-PCR Sequencing Kit contents									
VNP : VN primer					RAP : Rapid adapter				
SSP : Strand-switching primer					SQB : Sequencing buffer				
cPRM : cDNA primer					SQT : Sequencing tether				
EB : Elution buffer					LB : Loading beads				
Contents	Colour			No. of tubes					
VN Primers (VNP)	White cap, blue label			1					
Strand-switching primers (SSP)	White cap, pink label			1					
cDNA Primers (cPRM)	White cap, grey label			1					
Elution Buffer (EB)	Black			2					
Rapid Adapter (RAP)	Green			1					
Sequencing Buffer (SQB)	Red			2					
Sequencing Tether (SQT)	Purple			1					
Loading Beads (LB)	Pink			1					

Figure 16. The different reagents present in the SQK-PCS109 ONT MinION library preparation kit.

SQK-PCS109 primer sequences

Poly-T primer (VN Primer) in the VNP tube (concentration is at 2 μ M):

5' - **ACTTGCCTGTCGCTCTATCTTC** TTTTTTTTTTTTTTTTTVN - 3'

Strand-Switching Primer (Template Switching Oligo) in the SSP tube (concentration is at 10 μ M):

5' - **TTTCTGTTGGTGCTGATATTGC**T mGmGmG - 3'

(The mG modification corresponds to 2' O-Methyl modified RNA nucleotides)

cDNA amplification primers (Rapid attachment primers) in the cPRM tube (concentration is at 10 μ M):

➤ *Forward (PR2):*

5' - **TTTCTGTTGGTGCTGATATTGC** - 3'

➤ *Reverse (3580F):*

5' - **ACTTGCCTGTCGCTCTATCTTC** - 3'

Sequencing adaptors (“Adapter Y”) in the RAP tube:

➤ *Sequence of the “Adapter Y” top strand:*

5'-
GGCGTCTGCTTGGGTGTTAACCTTTTTTTAATGT**ACTTCGTTCAGTTACGTA**
TTGCT -3'

➤ *Sequence of the “Adapter Y” bottom strand:*

5'- **GCAATACGTAACGAACTGAA** -3'

Figure 17. Sequence information of the primers or sequencing adaptors used in the SQK-PCS109 ONT MinION library preparation kit. The sequence of the Forward and Reverse cDNA amplification primers is indicated on the Template Switching Oligo and on the poly-T primer with red and green color respectively. For the sequencing adaptor, the orange color marks the

bottom strand and the reverse complement of the bottom strand on the top strand (For the complete construct see Figure 29). The concentration of the primers is also presented in the picture. The 2' O-Methyl oligo modification is best characterized as an RNA analog that offers stability against general base hydrolysis and nucleases, as well as increased binding affinity of the RNA:DNA duplex (Tm of duplexes increases by 1 - 4°C per addition).

Procedure overview

In the following parts we will start with the cDNA synthesis and cDNA amplification reactions and we will end with the sequencing reaction as per the ONT cDNA-PCR Sequencing (SQK-PCS109) protocol. The different steps of the protocol are presented in Figure 18.

In brief:

- We will use the strand-switching protocol to prepare full-length cDNAs from total RNA
- We will amplify the cDNAs produced by PCR, adding rapid attachment primers during the PCR step
- We will attach sequencing adapters to the PCR products
- We will prime the flow cell, and load the cDNA library into the flow cell
- We will start a sequencing run using the MinKNOW software, which will collect raw data from the device and convert it into basecalled reads

The difference of this protocol with the previous versions of the protocol is that the attachment of the sequencing adaptors is not an enzymatic based ligation reaction but rather a copper catalyzed 1,3 dipolar cycloaddition reaction (click chemistry) between an alkyne (modification present on the synthesized cDNA) and an azide (modification present on the sequencing adaptor). The ability to attach adaptors via click chemistry offers a faster sample preparation time, higher yields of conjugated molecules and considerably improved sensitivity which enables the detection of the enriched target from much lower input amounts of starting material. The limitation of this protocol is that only the ONT provided primers can be used for the cDNA synthesis/amplification reactions whereas custom made protocols should rely on a kit that offers a ligation-based attachment of the sequencing adaptors (the ligation efficiency of the sequencing adaptors is ~50-60%).

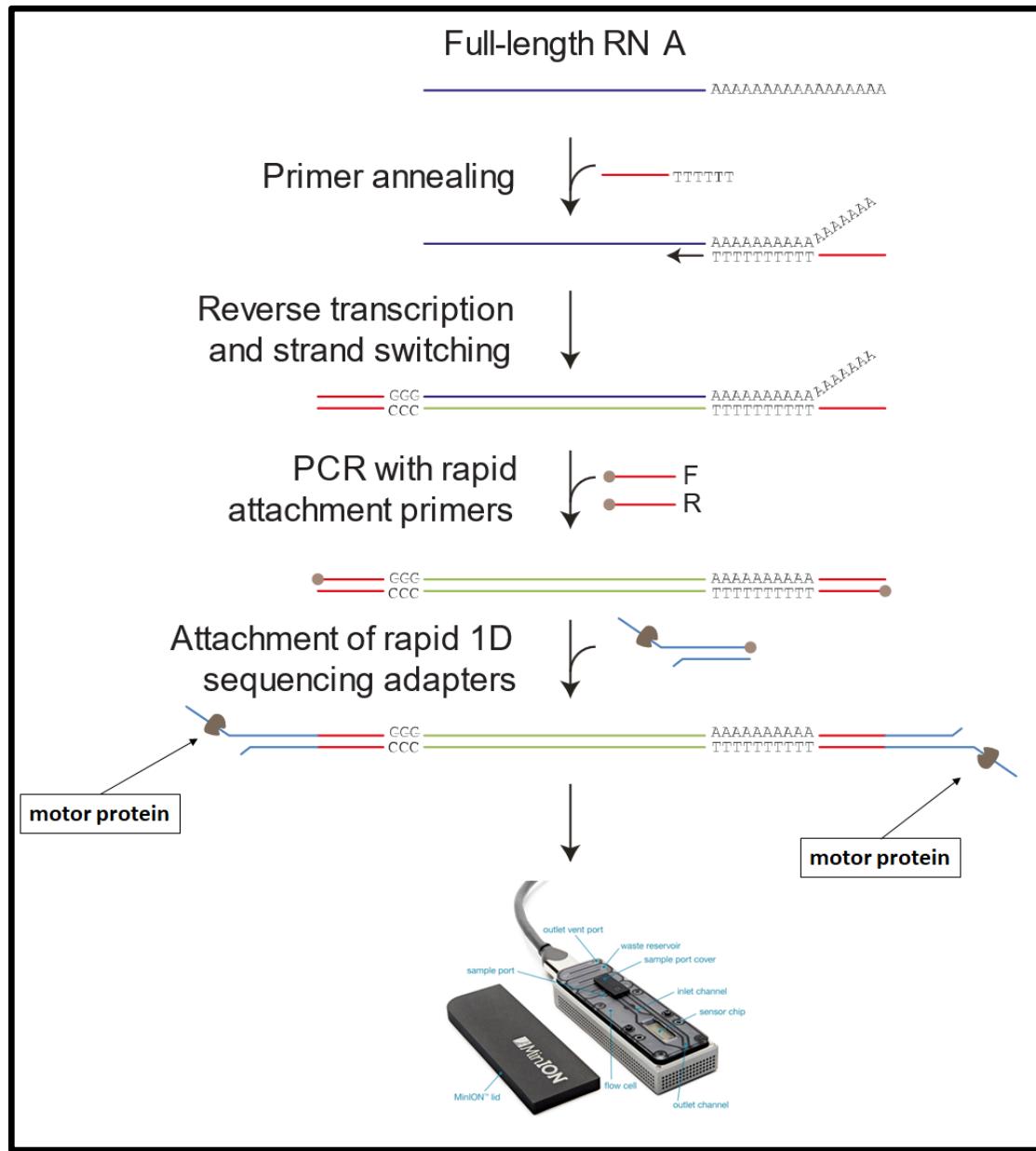


Figure 18. Overview of the protocol for cDNA sequencing on the ONT MinION platform. The cDNA synthesis, cDNA amplification, attachment of the sequencing adapters are presented in the picture. The whole procedure ends with sequencing on the ONT MinION device.

cDNA synthesis

Purpose: Reverse transcribe the polyA+ mRNA molecules into cDNA

General Information: In this protocol we use the Maxima H Minus Reverse Transcriptase to reverse transcribe the mRNA into cDNA. Afterwards we will use the LongAmp Taq DNA Polymerase for cDNA amplification. Other reverse transcriptase/cDNA amplification enzyme combinations can be used, for example the SuperScript IV enzyme (SSIV) as the reverse transcriptase and the Advantage 2 High fidelity polymerase for cDNA amplification⁷ or even the KAPA HiFi polymerase for cDNA amplification. The SSIV enzyme can be a good alternative option as the reverse transcription takes place at a higher temperature (~55°C) which helps resolve difficult templates. The intermediate products of the cDNA synthesis/cDNA amplification reactions are presented in Figure 19.

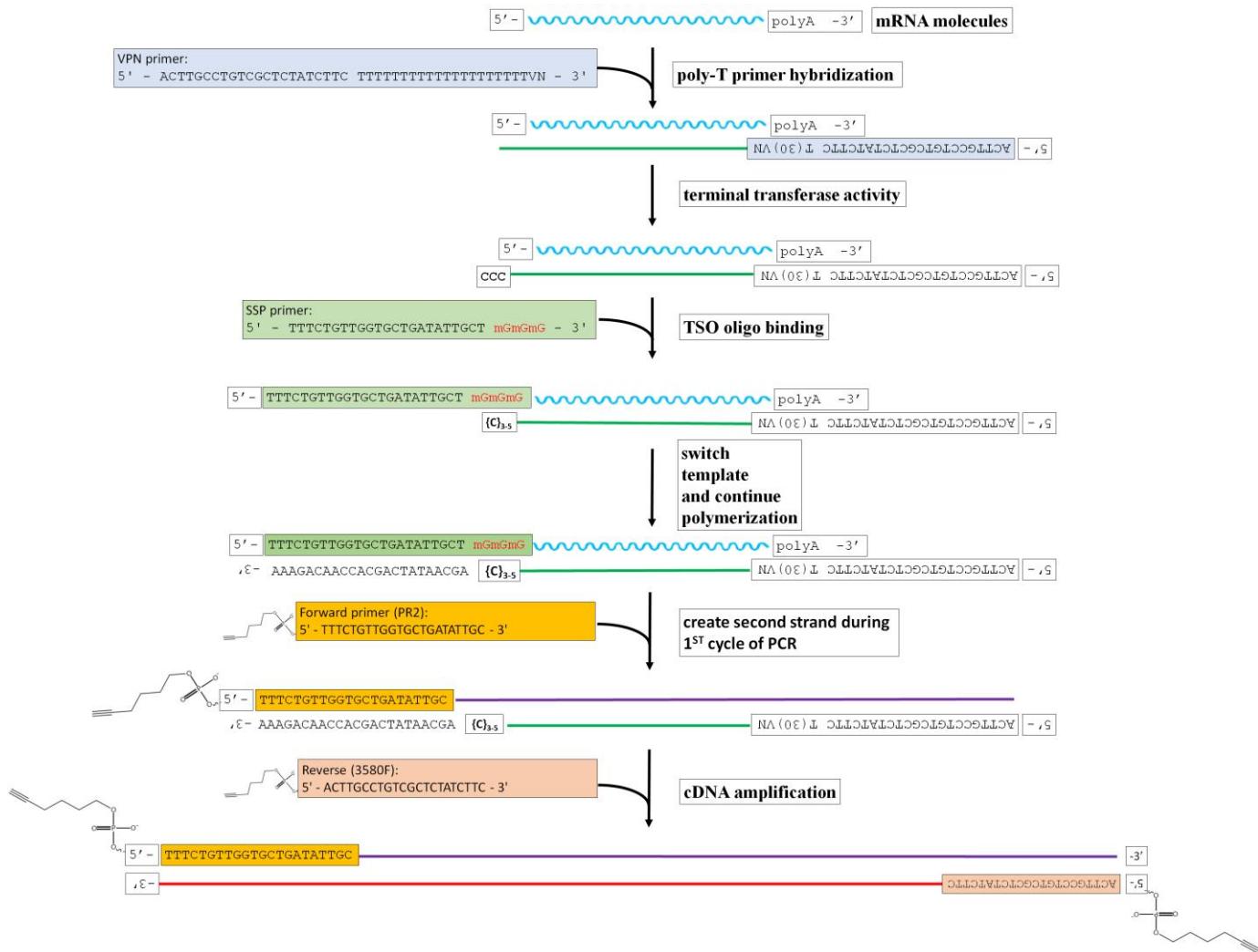


Figure 19. cDNA synthesis with the template switching method. Structure of the transcripts created after cDNA amplification. The incorporation of the poly-T primer (VN primer), the Template Switching Oligo (TSO; SSP primer), and the cDNA amplification primers (Forward and Reverse primers) in the synthesized cDNA molecules is presented for the different steps of the template switching protocol. The alkyne incorporated on the 5' end of the Forward and Reverse cDNA amplification primers is also indicated in the figure.

Procedure:➤ **polyT primer binding on the polyA+ RNA molecules**

For consistency, ONT recommends starting the library preparation from PolyA-enriched RNA; however, total RNA can also be used. As the proportion of PolyA+ RNA present in a total RNA sample may vary between organisms and tissues, some optimization of input amount and/or PCR cycles is usually required.

Steps:

1. Label the following 0.2 ml PCR tubes: “**pre-RT**” (1 PCR tube), “**RT**” (1 PCR tube), “**PCR**” (1 to 4 PCR tubes).
2. The “**pre-RT**” mix contains the polyT primer. Prepare the following reaction in the tube labelled “**pre-RT**” (Either the diluted ERCC RNA Spike-In Mix 1 or the diluted SIRVs Control set 3 kit is used):

	pre-RT mix	µl / reaction
1	diluted ERCC RNA Spike-In Mix 1 or diluted SIRVs Control set 3 kit	1
2	300 ng of total RNA	x
3	VPN	1
4	dNTP Mix (stock: 10 mM each)	1
5	Nuclease free water	(8 - x)
	Total =	11

Important: We note here that the ONT protocol suggests using 1 ng of polyA+ RNA or the equivalent 50 ngs of total RNA. Here we suggest using 300 ngs of total RNA. Given that the concentration of the polyT primer is 2 µM there should be enough primer to accommodate the mRNA content (usually 1-5% of total RNA) present in the 300 ngs of total RNA. The logic behind this, is that if there is enough total RNA available using a higher amount of RNA in the reaction, without saturating the reverse transcription

reaction, will avoid an unnecessary extra number of PCR cycles that might introduce PCR amplification artifacts. Practically we will take the same amount of synthesized cDNA with a lower number of PCR cycles.

3. Mix gently by flicking the tube, and spin down.
4. Incubate at 65° C for 5 minutes and then snap cool on a pre-chilled freezer block.

➤ **Reverse transcription reaction**

Steps:

1. In the tube called “**RT**”, mix together the following:

	RT mix	µl / reaction
1	5x RT buffer	4
2	RNaseOUT	1
3	Nuclease-free water	1
4	Strand-Switching Primer (SSP)	2
	Total =	8

2. Mix gently by flicking the tube, and spin down.
3. Add the “**RT**” mix to the “**pre-RT**” mix, mix by flicking the tube and spin down.
4. Incubate at 42° C for 2 minutes.
5. Add 1 µl of Maxima H Minus Reverse Transcriptase. The total volume is now 20 µl.
6. Mix gently by flicking the tube, and spin down.
7. Incubate on a thermocycler using the following protocol:

Cycle step	Temperature	Time	No. of cycles
Reverse transcription and strand-switching	42° C	90 mins	1
Heat inactivation	85° C	5 mins	1
Hold	4° C		

➤ **cDNA amplification reaction**

Overview: The PCR primers anneal to the complementary section at the end of the poly-T primer, and the strand-switching primer (Figure 17). Therefore, only full-length transcripts with both a poly-T primer and a strand-switching primer that was incorporated during the reverse transcription step will be amplified.

Important: *Each PCR reaction uses 5 µl of reverse-transcribed RNA (out of a 20 µl reaction). Therefore, sufficient material is available to perform four PCR reactions per reverse transcription reaction. Do NOT use all 20 µl of the reverse transcription reaction in a single PCR reaction.*

Steps:

1. In 1 tube named “PCR mix”, prepare the following reaction at room temperature. If all the “RT reaction” from the previous step is going to be used, then in 4 tubes prepare the following reaction in each one of them separately (200 µl of reaction in total):

	PCR mix	µl/reaction
1	2x LongAmp Taq Master Mix	25
2	cDNA Primer (cPRM)	1.5
3	Nuclease-free water	18.5
4	Reverse-transcribed RNA sample	5
	Total =	50

2. Amplify using the following cycling conditions. The number of cDNA amplification PCR cycles should lead to the production of 400-1000 ngs of cDNA per 50 µl of cDNA amplification reaction (Total 1000-2000 ngs of amplified cDNA if all the reverse transcription reaction of the 300 ngs of total RNA is used – 200 µl of cDNA amplification reaction). Otherwise adjust the cDNA amplification PCR cycles. You can do between 10-18 PCR cycles instead of the recommended 14 cycles.

Cycle step	Temperature	Time	No. of cycles
Initial denaturation	95°C	30 secs	1
Denaturation	95°C	15 secs	14
Annealing	62°C	15 secs	14
Extension	65°C	50 secs per kb	14
Final extension	65°C	6 mins	1
Hold	4°C		

3. Add 1 µl of NEB Exonuclease 1 (20 units) directly to each PCR tube. (*The role of the Exonuclease 1 is to degrade the unincorporated primers as it catalyzes the removal of nucleotides from single-stranded DNA in the 3' to 5' direction*)
4. Incubate the reaction at 37° C for 15 min, followed by 80° C for 15 min. (*At 80° you heat inactivate the Exonuclease I*)
5. If only one amplification reaction is used transfer the 50 ul in a clean 1.5 ml Eppendorf DNA LoBind tube. Otherwise if multiple amplification reactions are used, pool the reactions together in a clean 1.5 ml Eppendorf DNA LoBind tube.

➤ **Agencourt AMPure XP cleanup of cDNA amplification products**

1. Let AMPure XP beads equilibrate at RT for 30 minutes before use (vortex first and let them in RT).
2. Vortex AMPure XP beads until evenly mixed, then add 0.8X sample volume of Agencourt AMPure XP beads to the cDNA amplification reaction.
3. Pipet the entire volume up and down to mix thoroughly. Place the sample tubes on a Hula mixer (rotator mixer) for 5 minutes to let the DNA bind to the beads. Briefly spin the samples to collect the liquid from the side of the tube.
4. Prepare 500 µl of fresh 70% ethanol in nuclease-free water.
5. Place the sample tubes on the magnetic separation device for ~2 minutes until the liquid appears completely clear, and there are no beads left in the supernatant.
6. While the samples are on the magnetic separation device, pipette out the supernatants. Keep the samples on the magnetic separation device. Add 200 µl of freshly made 70% ethanol to each sample without disturbing the beads. Wait for 30 seconds and carefully pipette out the supernatant containing contaminants. DNA will remain bound to the beads during the washing process.
7. Add again 200 µl of freshly made 70% ethanol to each sample without disturbing the beads. Wait for 30 seconds and carefully pipette out the supernatant.
8. Briefly spin the samples to collect the liquid from the side of the wall. Place the samples on the magnetic device for 30 seconds, then remove all the remaining ethanol with a pipette.

9. Allow the beads to dry for ~30 seconds, but do not dry the pellet to the point of cracking.
10. Remove the tube from the magnetic rack and resuspend pellet in 12 µl of Elution Buffer (EB).
11. Incubate at room temperature for 10 mins.
12. Pellet beads on magnet until the eluate is clear and colorless.
13. Remove and retain 12 µl of eluate into a clean 1.5 ml Eppendorf DNA LoBind tube.
 - Remove and retain the eluate which contains the cDNA library in a clean 1.5 ml Eppendorf DNA LoBind tube
 - Dispose of the pelleted beads
14. Analyze 1 µl of the amplified cDNA for size, quantity and quality:
 - a. The cDNA can be quantified with a Qubit HS DNA Assay Kit. Use 0.5 ul of the amplified cDNA to quantify it. An estimate of the final cDNA yield is indicated in Figure 20 for different cDNA amplification cycles.
 - b. The profile of the cDNA will be analysed on an Agilent Tapestation as is presented in the next section. Use 0.5 ul of the amplified cDNA to check its profile. The expected cDNA electropherogram profile is presented in Figure 20. This profile picture was generated on the “PerkinElmer LabChip” instrument. The signal intensity distribution of the cDNA library on the High Sensitivity DNA LabChip spans the region from 400bp and above. We note here that the electropherogram profile can be slightly different in an Agilent D5000 screentape.
15. If the quantity of amplified cDNA is above 200 fmol (200 ngs for fragments with average size 1.5 kb), the remaining cDNA can be frozen and stored for another sequencing experiment (in this case, library preparation would start from the Adapter Addition step). We recommend avoiding multiple freeze-thaw cycles to prevent DNA degradation.

Expected results:

The number of cDNA amplification cycles depends:

- I. on the fraction of RNA molecules with a polyA tail to the total RNA molecules.
- II. the first strand cDNA synthesis efficiency of the reverse transcriptase.

Increased number of PCR cycles can create artifacts due to:

- **strand invasion or template switching:** The more denaturing and annealing processes happen the higher the chances that hybrid molecules are going to be created (for example 30% of the molecules after 30 cycles of PCR amplification⁸). This occurs when molecules with partial complementarity anneal with each other. For example, molecules with partial complementarity are the different isoforms of the same gene.
- **daisy chains:** During PCR the primers will eventually be used up and limit further reactions. The depletion of primers causes daisy chaining. Since the denaturing and annealing processes continue without primers, the single-stranded DNA molecules reanneal to themselves. However, this reannealing does not always occur with another complementary strand and due to the fact that all of the synthesized molecules have the same ends, end to end hybridization between different molecules can prime the PCR extension. It is this imperfect match up that causes 'tangles'. These tangles look like a daisy chain.
- **preferential amplification of short transcripts at the expense of bigger ones:** The amplification efficiency of the short transcripts, sometimes, is higher than the longer ones⁹.

To reduce the PCR artifacts, it is advisable to perform the minimum number of cDNA amplification cycles that will give a reasonable amount of working material. For example, in Figure 20 we see that 14 cycles of cDNA amplification can give approximately ~10.5 ug of amplified cDNA per 300 ng of starting amount of total RNA. Nevertheless, comparison of the cDNA electropherogram of 14 cycles with the other cycles (9-13 cycles) shows the existence of artifacts at the size range of cDNA molecules more than 3000 bp (red shaded area in Figure 20). In this case we note that the density of the size distribution is higher in the area of very large molecules, for the 14 cycles, indicating the presence of

fusion molecules (molecules produced from strand invasion or daisy chains). Note also the complete absence of free primers in cycle 14. The ideal cDNA amplification cycles are the ones that produce 1-2 ug of amplified cDNA material from 300 ng of starting amount of total RNA. In the case of Figure 20 this corresponds to 11 or 12 cycles of PCR cDNA amplification. We note here that MCF7 is a highly proliferative cell line with a high mRNA content (~5% of total RNA) whereas in case of quiescent cells the mRNA content can be lower than the MCF7 one (~1% or less of total RNA). In that case, the number of PCR cycles needs to be increased proportionally to achieve the same yield, and this is the reason that 14 cycles are recommended in the cDNA amplification reaction (5 time less mRNA corresponds theoretically to 2-3 extra cycles of PCR)

As per the ONT recommendation, libraries with artificial high-molecular weight product instead of the expected smear, are typically associated with poor sequencing performance. ONT suggests that repeating the cDNA amplification PCR with fewer cycles can remedy this problem.

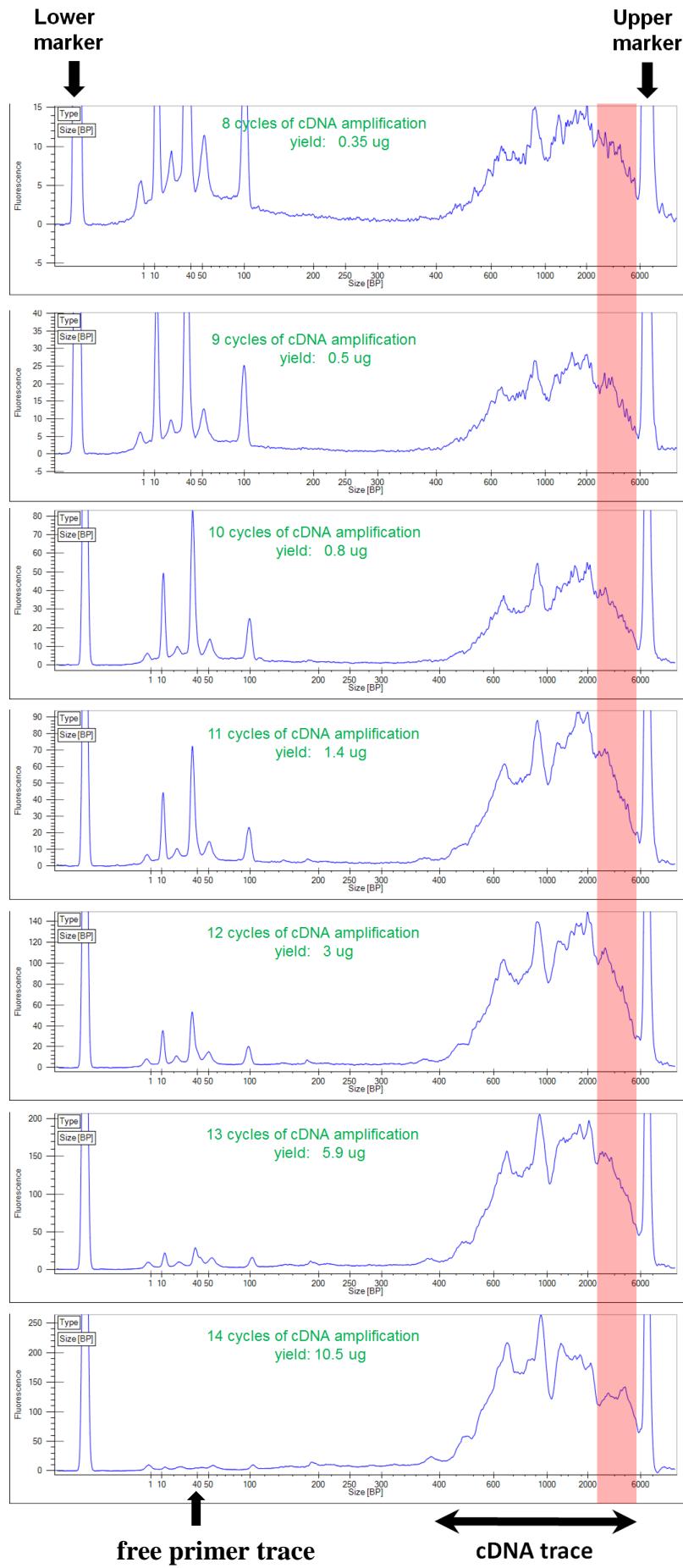


Figure 20 (previous page). Typical cDNA profile from a PerkinElmer LabChip GX High Sensitivity DNA Chip for the cDNA amplification cycles 8-14. MCF7 total RNA was processed with a similar but not exactly the same protocol as the one used here⁷. The yield of the cDNA produced per PCR cycle is also indicated. This yield corresponds to a starting amount of 300 ngs of total RNA. Despite the different cDNA synthesis/amplification protocols we expect the yield of the different PCR cycles to be close to the one presented in the image above. If less than 300 ngs of total RNA are used, the yield of the cDNA synthesis/cDNA amplification reaction should be proportionally downgraded. In this case to achieve a specific yield the number of cycles should be increased. The yield was calculated after measuring the cDNA present in the PCR reaction with a dsDNA specific dye (Qubit HS DNA Assay Kit). No Ampure XP beads cleanup was performed before measurement. Upon Ampure XP beads cleanup someone should expect to have ~70%-80% of the indicated yield from each cycle. The lower and upper marker traces of the LabChip reagents are also indicated. For cycles 11-14 in total 70 ng of cDNA were run on the LabChip lane whereas for the cycles 8,9,10 we run 34, 47 and 65 ng of cDNA respectively. We note here that MCF7 is a highly proliferative cell line with a high mRNA content whereas in case of quiescent cells the mRNA content will be lower than the MCF7 one. In that case, the number of PCR cycles needs to increase to achieve the same yield as the MCF7 cell line.

Assess the profile of the extracted cDNA

Purpose: Asses the quality of the extracted cDNA

Process Overview: cDNA is run on an Agilent Tapestation or equivalent platform to assess the cDNA profile.

General information: The profile of the produced cDNA can be assessed using an Agilent TapeStation instrument. The quantitative range of the instrument is 0.1 – 50 ng/ μ l whereas the sensitivity of the instrument (minimal amount of cDNA that can be loaded) is 0.1 ng/ μ l. The linear range of size estimation is between 100 – 5000 nt.

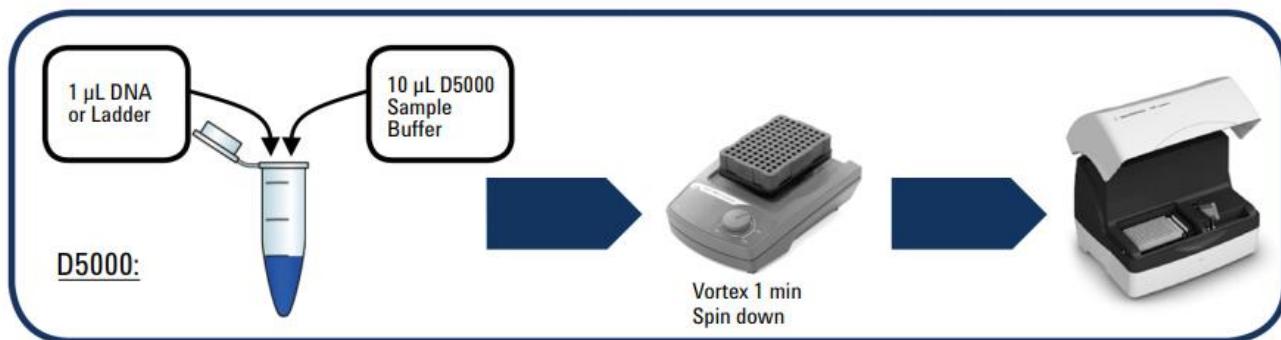


Figure 21: Overview of the quantification of cDNA in the Tapestation instrument.

Procedure: We follow the Agilent D5000 screentape manual

(https://www.agilent.com/cs/library/usermanuals/public/ScreenTape_D5000_QG.pdf):

1. Allow reagents to equilibrate at room temperature for 30 min.
2. Vortex mix before use.
3. Use the corresponding strip tubes to add the sample and the ladder.
4. Prepare ladder by mixing 10 μ L D5000 Sample Buffer with 1 μ L D5000 Ladder in a tube strip.
5. Prepare sample by mixing 10 μ L D5000 Sample Buffer with 1 μ L DNA sample in the same tube strip.

6. Spin down, then vortex using IKA vortexer and adaptor at 2000 rpm for 1 min.
7. Spin down to position the sample at the bottom of the tube.

Introduction to the ONT MinION nanopore sequencing platform

Nanopore devices are a class of single molecule bio-sensors. These include both protein pores embedded in a lipid bilayer and “solid-state” nanopores, where the membrane material is made of a highly insulating, solid material, such as Si3N4, SiO2, polymers, or graphene. The operating principle of a nanopore device is relatively simple¹⁰. It consists of a liquid cell, which is separated into two compartments by the thin membrane (Figure 22). The latter encompasses a small aperture, the nanopore, which is the only connection for ions and fluid to be transferred between the two compartments. Moreover, each compartment contains one electrode, which is in rapid redox equilibrium with the surrounding electrolyte (Figure 22 A, C). For small pores (less than 100nm in diameter), the nanopore resistance ‘R’ is usually much larger than the solution resistance R_s . Hence, upon application of a constant voltage ‘E’, an ion current is induced that depends on the cross-sectional area and length of the pore channel, the conductivity of the electrolyte, and for very small pores also to a significant extent on the surface charge density inside the pore (Figure 22 B). The majority of the potential (V_E) thus drops at the nanopore, creating a relatively strong electric field (E) ($E = -\nabla V_E$). It is the electric field that is the major driving force for the translocation of charged biomolecules, such as DNA or RNA. The translocation of ssDNA through the nanopore causes a drop in the current that is characteristic of the bases in contact with the pore at that time (Figure 22 B).

In the ONT MinION platform a sensor measures the current in the nanopore several thousand times per second (3,000 Hz). The raw current measurements are compressed into a sequence of ‘events’, each being a mean current value with an associated variance and duration. The raw current measurements or the corresponding events, plotted over time, are referred to as a ‘squiggle plot’ (Figure 23). The base-caller then models the characteristics of 4096 possible DNA 6-mers.

Characteristics of the ONT MinION flowcell

The flow cell is composed of: a fluidic chamber, the sensor array, a common reservoir, a common electrode, ion selective membrane, a printed circuit board (PCB) and an application specific integrated circuit (ASIC) (Figure 24). Standard manufacturing processes are used to create the nanopore devices (Figure 25).

In more details:

- **Flow cell compartments:**

- The flow cell has three compartments: the common electrode reservoir which is separated from the bulk by a diffusion barrier, the bulk which is separated from the individual wells by a membrane, and the individual wells (Figure 22). The flow cells are shipped with the storage buffer (yellow) in all three compartments to maintain osmotic balance. The storage buffer contains salt and the standard redox couple that enables the current to run through the nanopore. The storage buffer also contains a QC DNA molecule that gives a distinct signal which helps to identify functional pores during Platform QC.

- **Sensor array:**

- The sensor array is where the sequencing chemistry happens. The array is made up of many sensing wells, each designed to record through a single nanopore (Figure 26).
- The sensor area of the flow cell is comprised of many individual sensor wells, each designed to hold a single nanopore. These sensor wells have a micro patterned structure, created using photolithographic techniques, sitting on a silicon substrate (Figure 25). The well provides the geometry for forming a membrane so that a single nanopore can be inserted. The well is approximately 90 µm deep with a platinum electrode at the base of the well (Figure 26). This electrode is then connected to the PCB through vias in the silicon layer. On the top of the wells are pillars (Figure 26) that are patterned to control the movement of a hydrophobic pretreatment (Figure 25). The wells are also patterned with additional features to control the pretreatment distribution (Figure 26).
- The flow cells have a SpotON port added so that small volumes of sample can be added directly on to the array. The volume of the main chamber area over the sensor array is 100 µl. Using the SpotON inlet, sample volumes of 75 µl can be used either by a dropwise application, or by pipetting straight above the sensor array (Figure 24A).

- **Membranes:**

- For each of the wells, a membrane is formed over the array so that the ionic fluid underneath the membrane is trapped. The membrane provides an insulating layer, so that once a nanopore is inserted, communication between the electrodes is dominated by the signal through that pore. A triblock amphiphilic polymer and a synthetic pretreatment oil is used to form the membrane (Figure 25). The use of this triblock amphiphilic polymer gives the flow cell membranes increased robustness to physical disruption, chemicals, and biological materials.

- **ASIC:**

- The ASIC is a high-density array of low-noise amplifier circuits that is used to measure and provide a digital readout of the current flow in the electrochemical circuit between the individual well and shared common electrode. In addition, the ASIC receives commands from the host software system to provide control functionality for the sensor array, including: acquisition frequency, signal filtering, sensor current range, multiplex input selection, electrode bias potential generation, and deselection of sensor inputs with broken membranes to avoid saturation of the measurement circuits.
- The flow cell ASIC is designed to operate across a wide range of conditions depending on the application. Typical operating cases are acquisition frequency of 2 to 20 kHz while measuring 10s to 1000s of pico-amperes. The applied electrode bias potential can be controlled using scripts in MinKNOW with a range of ± 1 V.
- The ASIC also has the ability to use the applied potential to unblock any of the channels that are not sequencing. A reversal of the potential can “flick” any stray DNA or contaminants out of the pore on a per channel basis and reset the channel to an “open pore” state to allow the next strand to be sequenced.
- The MinION flow cell has 2048 active well electrodes organized hexagonally on the surface of the array. The active wells are arranged in two blocks of 32x32 with a set

of four inactive wells separating them. There are also inactive wells present at the edges of the array where the gasket sits (Figure 27).

- The MinION flow cell has 512 signal reading channels, each connected to 4 wells which may contain at least one nanometer-scale biological pore (nanopore) embedded in an electrically-resistant membrane bilayer (Figure 26). Each channel provides data from one of the four wells at a time, allowing up to 512 independent DNA molecules to be sequenced simultaneously. The order of use of the wells is defined by the allocation of wells to 4 well-groups during an initial examination of the well conductivity at the beginning of each sequencing run. The choice of a single well for each of the recording channels is referred to as multiplexing. Multiplexing the nanopore array is used to improve the yield of channels containing a single pore and the sequencing output of the consumable. Most sequencing scripts perform a ‘MUX selection’ where the best 512 wells are chosen as the first group. As a nanopore channel overtime can stop functioning, every 90 minutes another ‘MUX scan’ is performed in order to select the best 512 wells from the ones that are left. Additionally, if during the sequencing run a nanopore channel stops functioning then another microwell with a working nanopore, from the remaining 3 microwells of the same signal reading channel, is selected. In the past, in the microwells that were not in use at any given time, there was a 20% leakage of current. This resulted in a slow depletion of the electrochemistry and eventual degradation of the microwell sequencing performance even though MinION was not using them to sequence molecules. In this version of the platform the current leakage through non sequencing microwells has been significantly reduced.

- **Electrodes:**

- Electrodes are used to create a circuit with the membrane and the buffer, present in the sensor array, so that electrophysiological measurements can be made as the ion flow changes through the pores.

- **Waste chamber:**

- The waste chamber has a total capacity of 1.9 ml. The outlet port allows excess sample to be removed with a pipette.

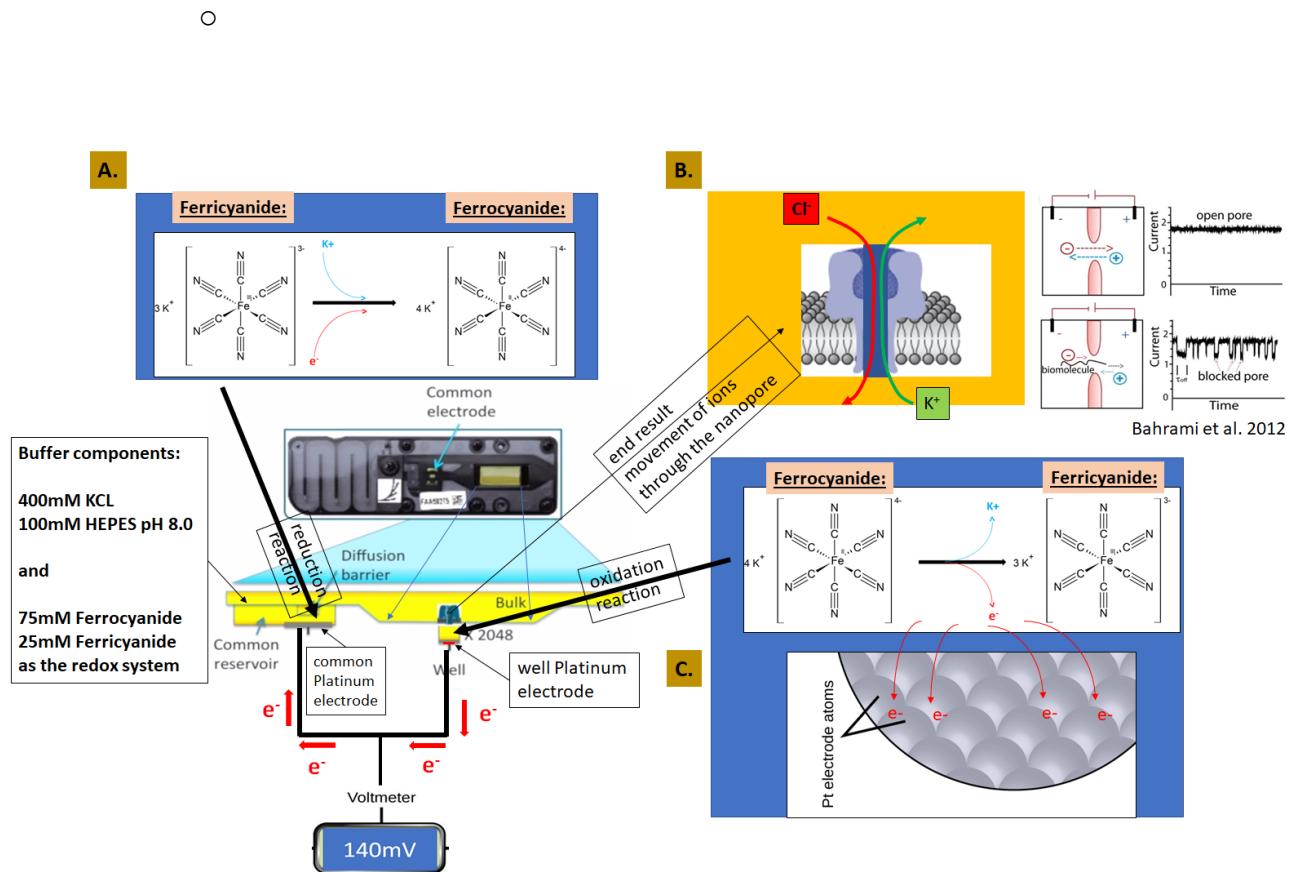


Figure 22. Characteristics of the ONT MinION nanopore sequencing platform. The redox reactions taking place in the vicinity of each electrode are presented in the figure (A, C). When a voltage is applied across the membrane, an ion current flows through the nanopore. The translocation of ssDNA through the nanopore causes a drop in the current that is characteristic of the bases in contact with the pore at that time (B).

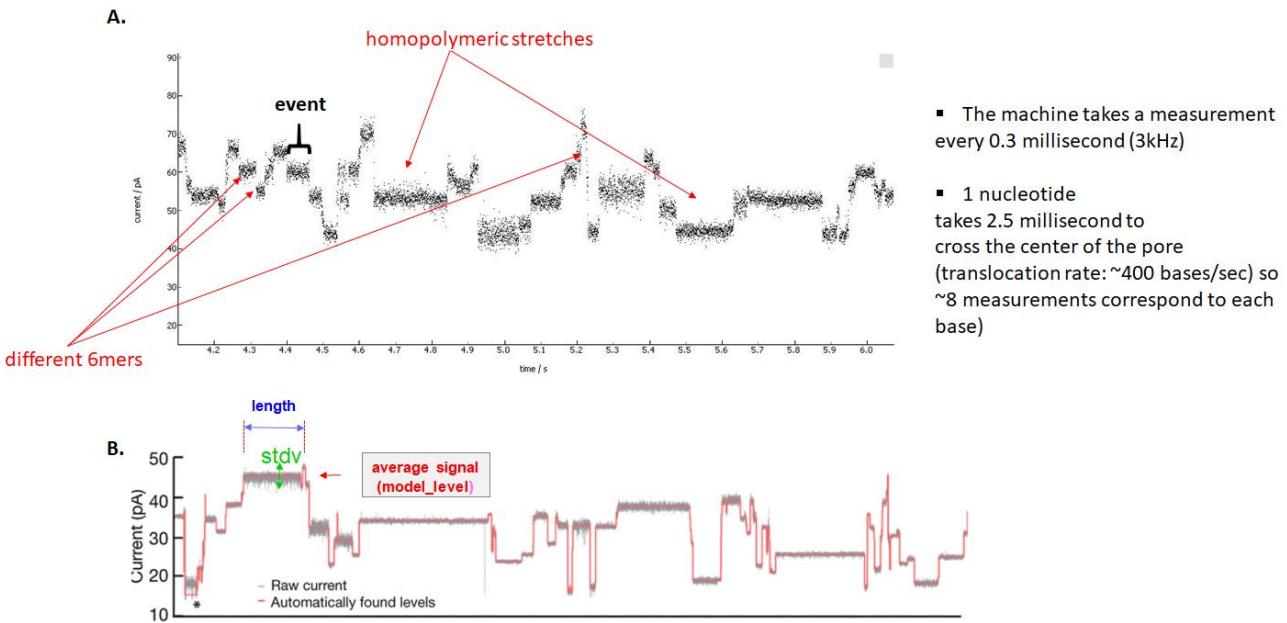


Figure 23. Signal generation from each nanopore in the ONT MinION sequencing platform. A)

The raw signal creates the “squiggle plots”. For $R_{\text{pore}} = 3 \text{ GigaOhms}$ and $V = 180 \text{ mV}$ then the measured current is $I = V/R \rightarrow I = 60 \text{ pA}$. B) The squiggle plots are then segmented into events. Each event corresponds to either one distinct DNA 6mer or to a homopolymeric stretch of DNA 6mers. The duration of each event and the accuracy of the measurements corresponding to each event (mean value and standard deviation) are also recorded.

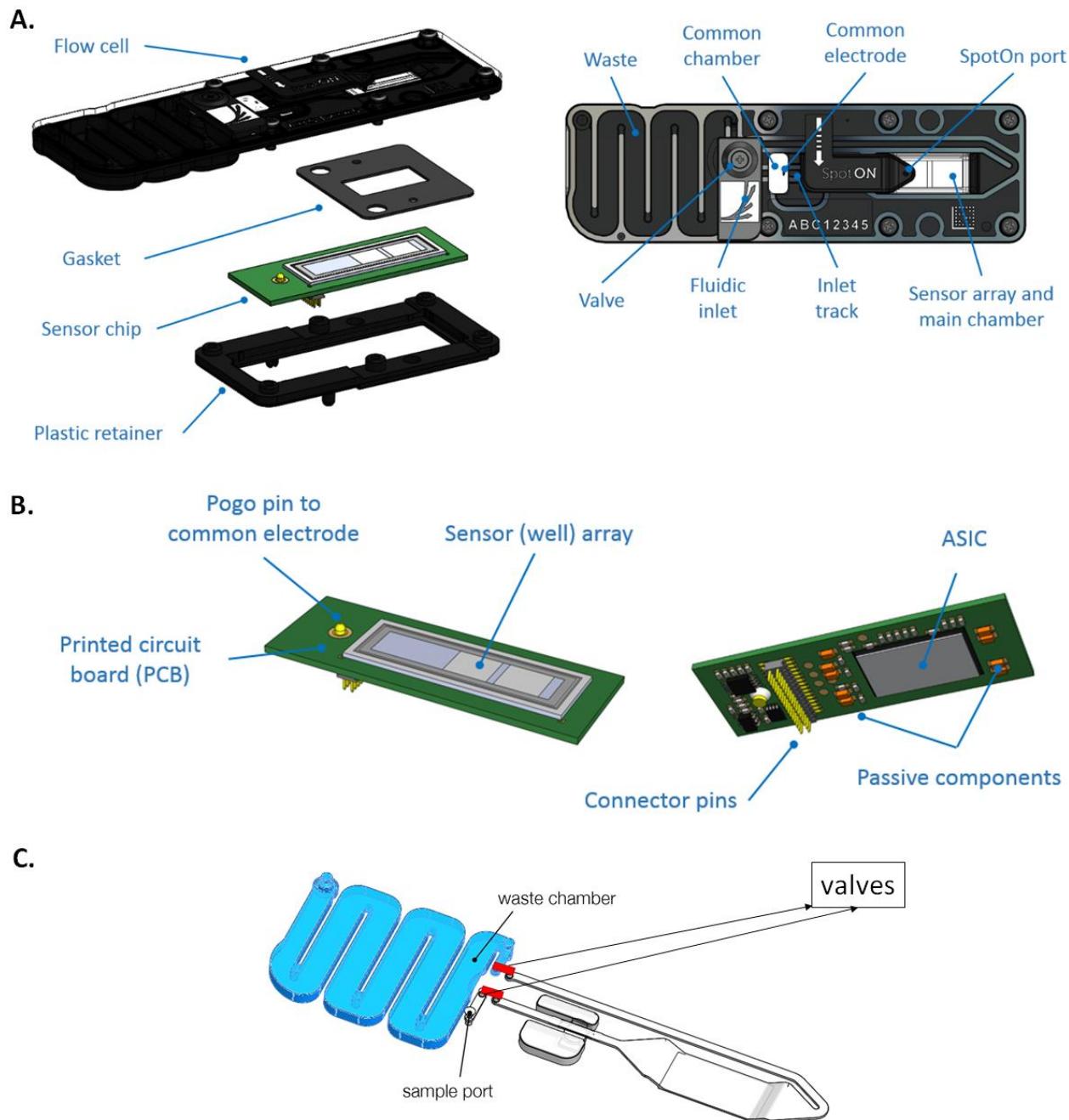


Figure 24. Characteristics of the ONT MinION flow cells. A) The flow cells contain the proprietary sensor array, Application-Specific Integrated Circuit (ASIC) and nanopores that are needed to perform a complete single-molecule sensing experiment. B) Electronics present on the flow cell C) The path of the liquid through the flow cell is regulated by two valves: one immediately downstream of the priming port, and the other between the end of the array and the waste chamber (valves shown in the diagram in red).

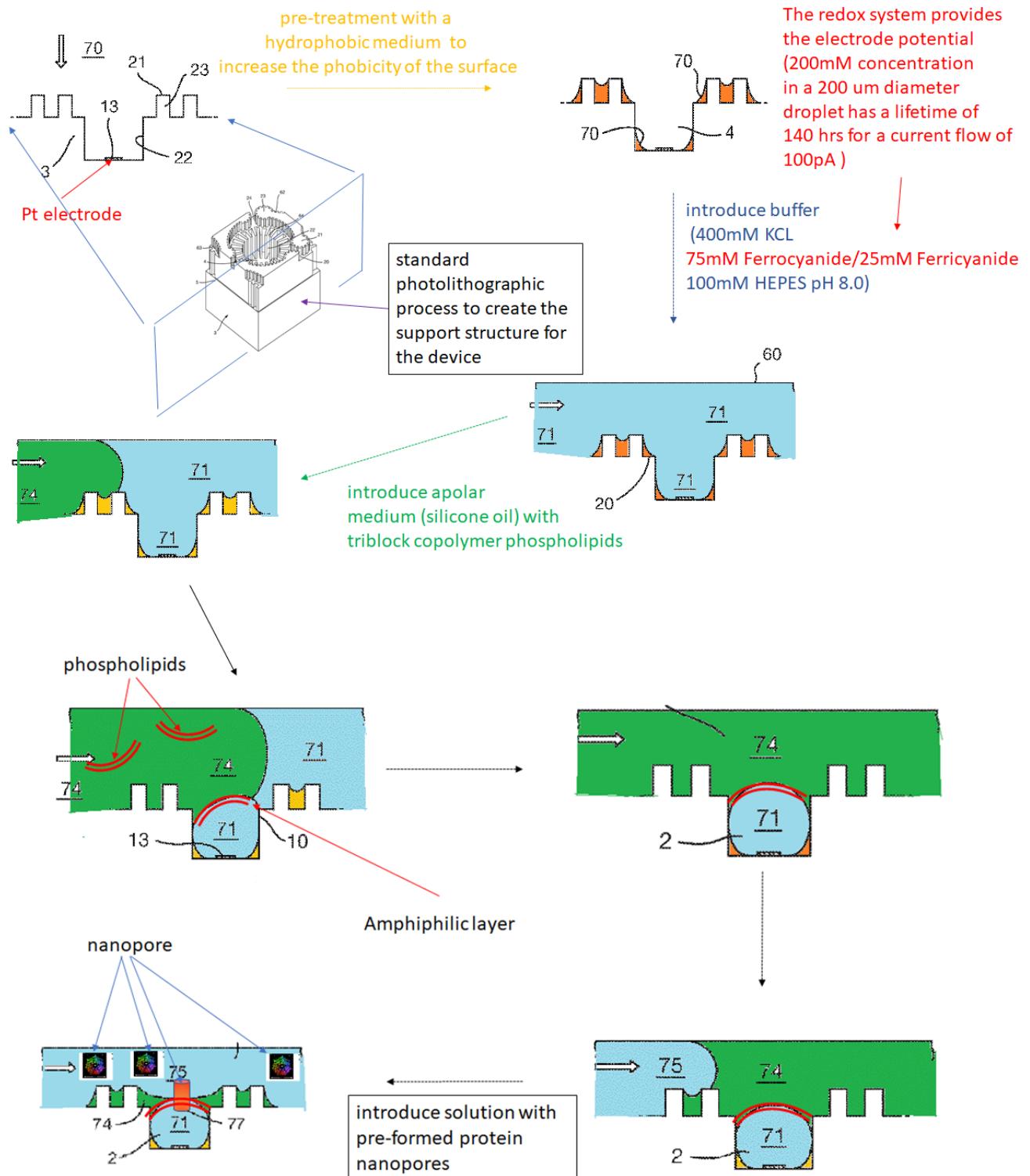


Figure 25. Manufacturing process of the ONT MinION sensor array based on the ONT patents.

We note here that the current manufacturing process of the MinION platform can be significantly different from the one presented on the corresponding patents. The figures present the sequence of

events that happen during the introduction of the electrolyte solution, the engineered amphiphilic triblock copolymers that will create the lipid bilayer, and the introduction into the device, at the last step, of the solution with the pre-formed nanopores.

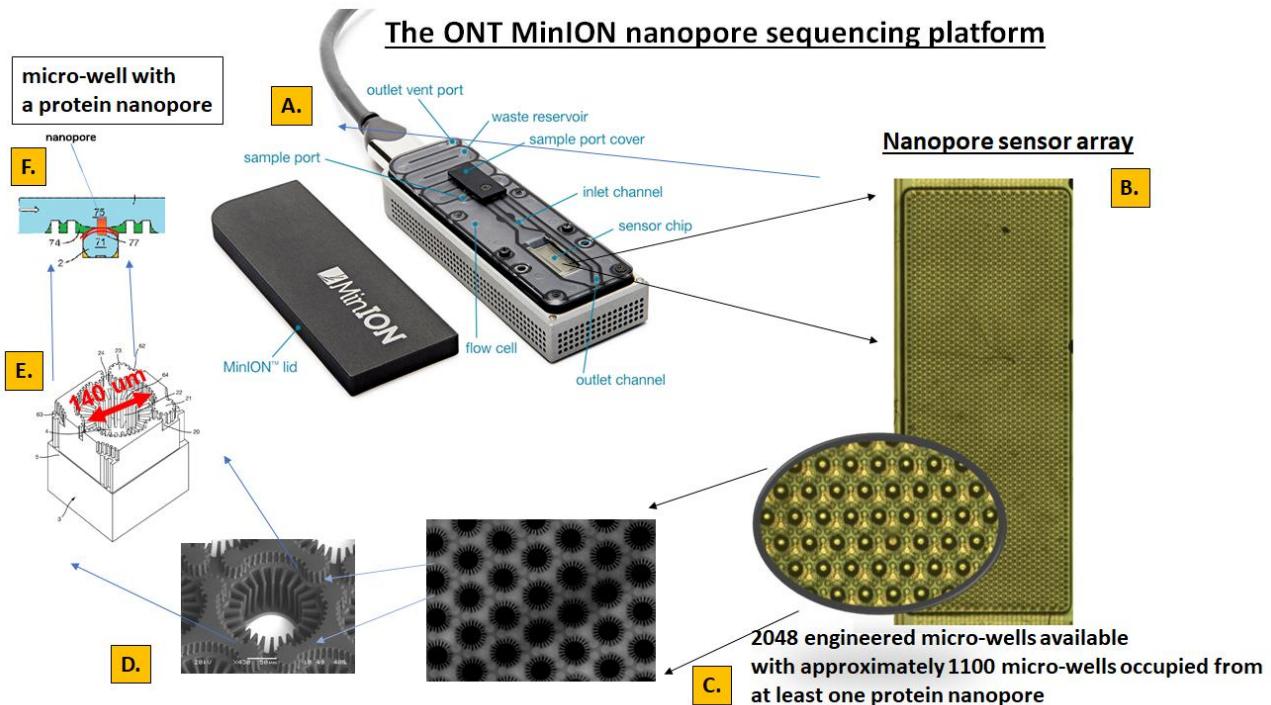


Figure 26. Details of the ONT MinION sensor array. A) A fluidic chip is used to pass a solution of DNA or RNA molecules on top of a nanopore sensor array. B) The nanopore sensor array consists of 2048 engineered micro-wells where a lipid bilayer can be used to create a sealed compartment between the micro-well volume and the rest of the device. At least one nanopore protein can be present in at least half of the micro-wells. The occupancy of the nanopore protein incorporation into the lipid bilayer follows a poisson distribution with just a few micro-wells having more than one nanopore. This implies that a large fraction of the micro-wells (for example at least 40%) do not have a nanopore to avoid overpopulating each micro-well with more than one protein nanopore. C) Magnification of the array of micro-wells. D) Scanning Electron Microscopy image of a single sensor well. E) Dimensions of each micro-well in the sensor array. F) A micro-well occupied from a protein nanopore. We note here that a change in the manufacturing process has increased the number of available nanopores from the ~1100 nanopores indicated in the picture to ~1500-1600 nanopores per flowcell. Practically during the introduction of the solution with the nanopores (last picture in Figure 25), in order to help them insert into the membrane, a voltage is applied that transiently disrupts the continuity of the lipid bilayer and this helps the nanopores insert more efficiently in the lipid bilayer.

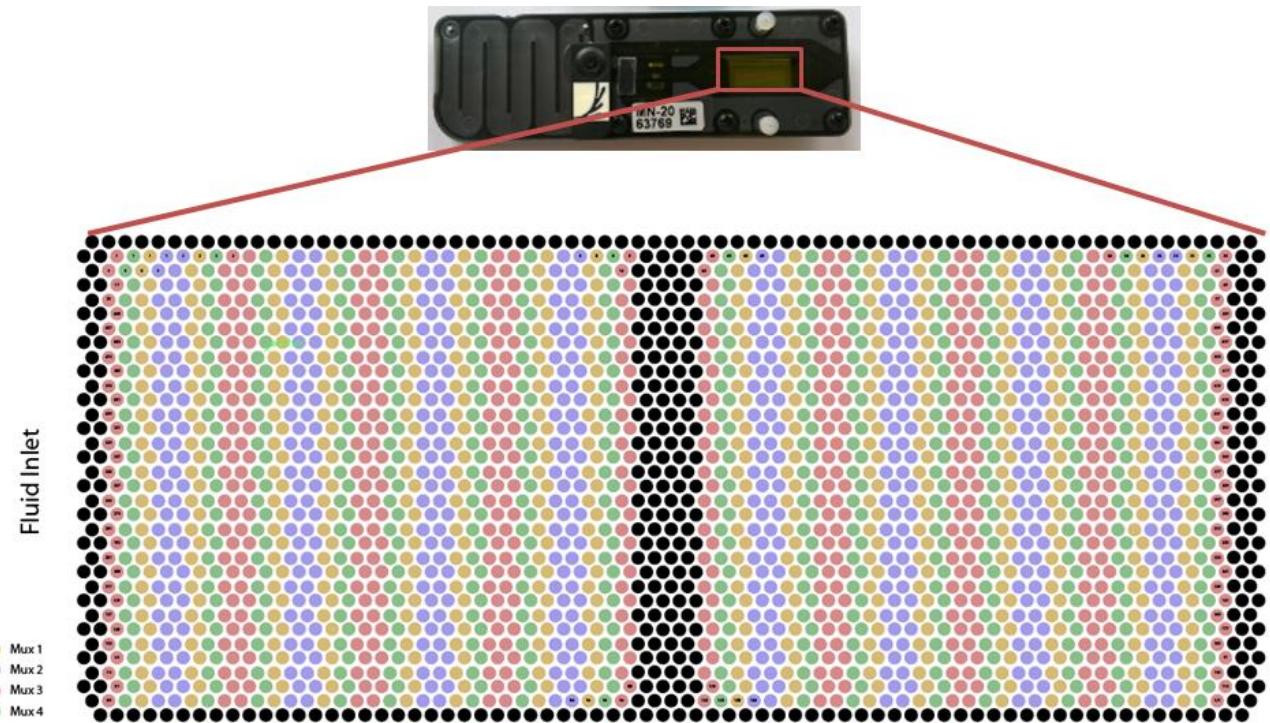


Figure 27. Multiplexing groups (Mux) in the ONT MinION sensor array. Each hexagon corresponds to a micro-well. The colored hexagons represent micro-wells connected to the measurement channels. The black micro-wells are not connected to any measurement channel. The different colors correspond to the 4 different groups (Mux 1,2,3 and 4) of microwells that are examined for the existence of a nanopore during the “platform QC” step and before the sequencing run step.

cDNA sequencing on MinION

Purpose: Sequencing of the cDNA library of poly-A transcripts on a nanopore MinION sequencing platform.

Process Overview: cDNA is conjugated with the Nanopore ONT MinION adaptors through Click chemistry and then it is sequenced on the ONT MinION device.

Procedure:

ONT MinION Library preparation

Attachment (chemical conjugation) of the sequencing adapters

General information: The cDNA molecules will be conjugated to the ONT MinION sequencing adaptors. The sequencing adaptors have a region where the motor protein is pre-bound (Figure 29). That is the reason for storing them in -20°C to reduce the dissociation of the motor protein from the adaptor molecules. A cholesterol tether is also bound on the adaptors. The role of the cholesterol tether is to restrict the diffusion of the cDNA molecules on the lipid bilayer (Figure 30). The click chemistry reaction between the cDNA and the sequencing adaptors is presented in Figure 31. A schematic representation of the conjugation of the sequencing adaptors on the cDNA biomolecules and their subsequent binding on the CsgG nanopore is presented in Figure 32.

Procedure:

1. Make up 100-200 fmol (200ngs) of amplified cDNA, and make the volume up to 12 µl in Elution Buffer (EB). *As can be seen in Figure 28 a solution of 200 ngs of cDNA, with an average fragment size of 1.5 kb, corresponds to ~200 fmoles. We note here that only 0.1% of the molecules that are introduced into the MinION are sequenced.*

Mass	Molarity if fragment length = 0.5 kb	Molarity if fragment length = 1.5 kb	Molarity if fragment length = 3 kb
50 ng	154 fmol	51 fmol	26 fmol
100 ng	308 fmol	103 fmol	51 fmol
200 ng	616 fmol	205 fmol	103 fmol
300 ng	925 fmol	308 fmol	154 fmol

Figure 28. Molarity of DNA fragments of different lengths. The average cDNA fragment length is 1.5 kb

2. Add 1 μ l of Rapid Adapter (RAP) to the amplified cDNA library. *The amount of the azide carrying adaptors in the conjugation reaction is usually 2-3 times more than the amount of the alkyne carrying biomolecule.*
3. Mix by pipetting and spin down.
4. Incubate the reaction for 5 minutes at room temperature.
5. Spin down briefly.
6. The prepared library is used for loading into the flow cell. Store the library on ice until ready to load.

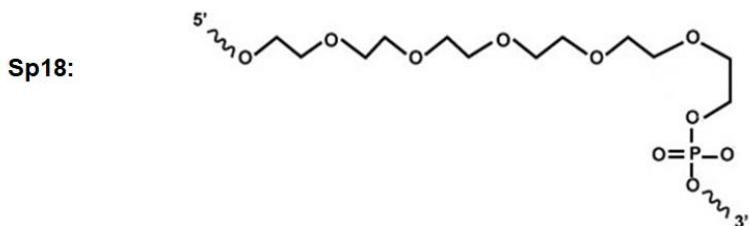
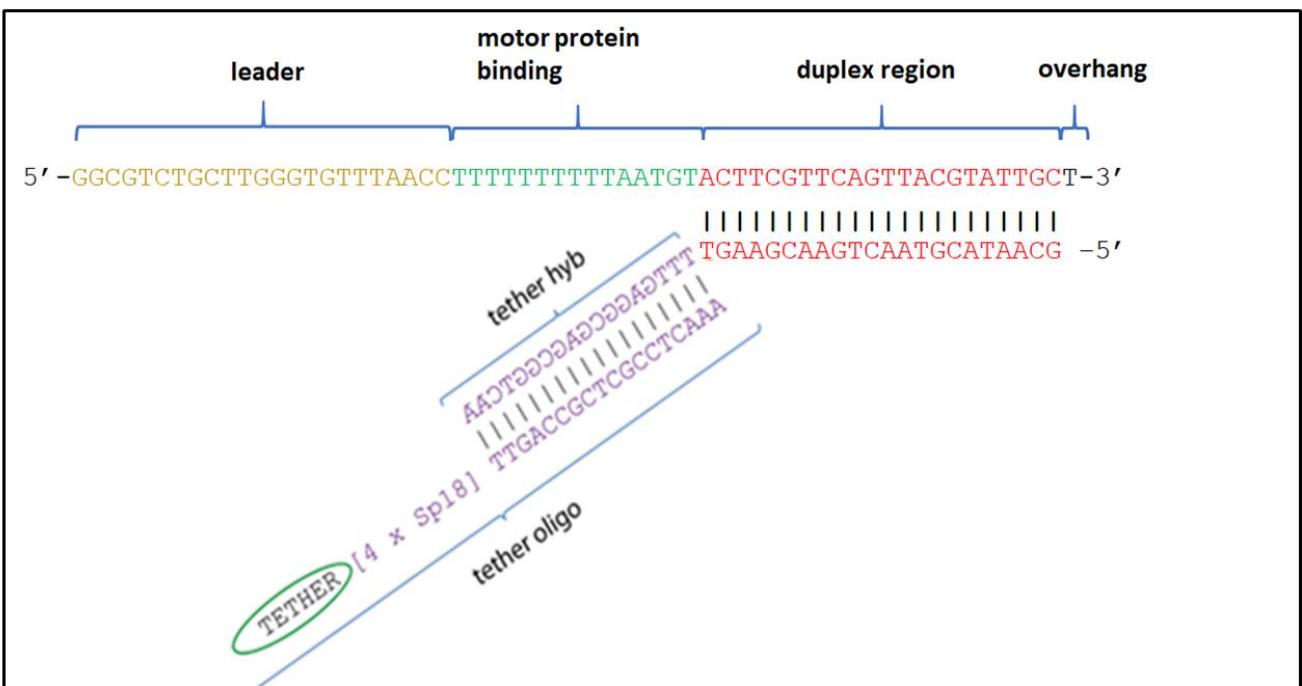
Adapter Y:

Figure 29. ONT sequencing adaptor (Y adaptor) structure. The tethered cholesterol moiety is also indicated. We note here that the cholesterol tether is not pre-bound on the sequencing adaptors during their chemical conjugation with the cDNA. The cholesterol tethers are introduced first inside the MinION flow cell and once the cDNA is added they hybridize with the sequencing adaptors.

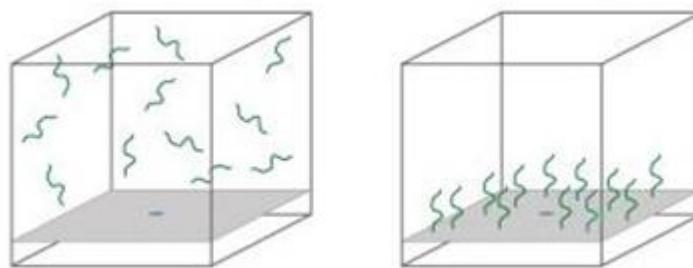


Figure 30. The cholesterol tethers restrict diffusion of the cDNA molecules on the lipid bilayers instead of having to diffuse on the volume of the whole buffer.

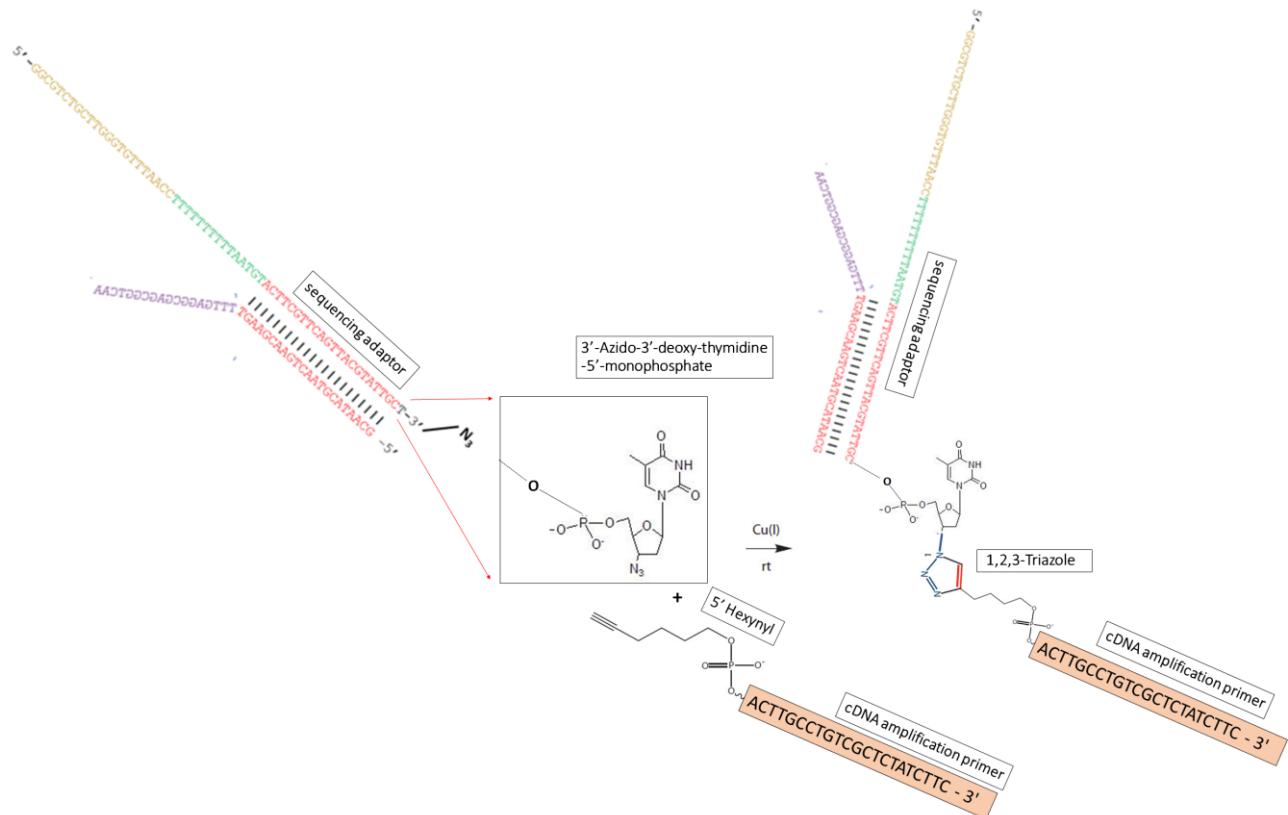


Figure 31. Chemical conjugation of the sequencing adaptor on the cDNA amplification primer through click chemistry. The cDNA amplification primer carries the alkyne and the sequencing primer

the azide residue. The Copper(I) catalyzed 1,3-dipolar cycloaddition of the azide with the alkyne is indicated. The produced 1,2,3-Triazole is also presented.

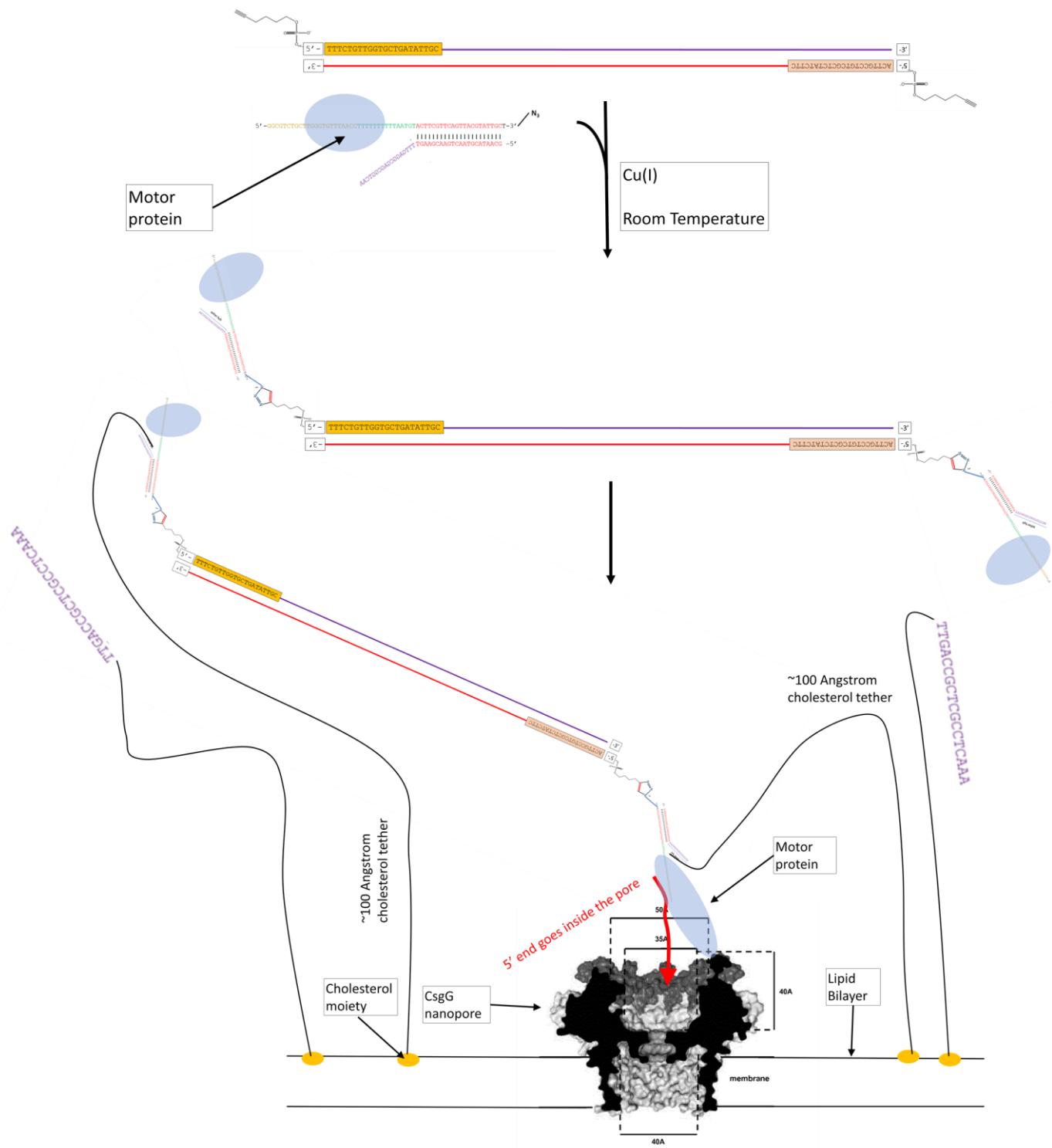


Figure 32. Schematic representation of the conjugation of the sequencing adaptors on the DNA biomolecules and their subsequent binding on the CsgG nanopore. The cholesterol tethers are first introduced in the flowcell where they insert into the lipid bilayer. Afterwards the sequencing adaptor conjugated cDNA molecules are introduced in the flow cell and these molecules hybridize with the free

DNA end of the cholesterol tethers which eventually restricts their movement on the lipid bilayer and brings them next to the nanopore. In the last picture the dimensions of the molecules are not to scale.

Quality Control of the flow cell (“Platform QC”)

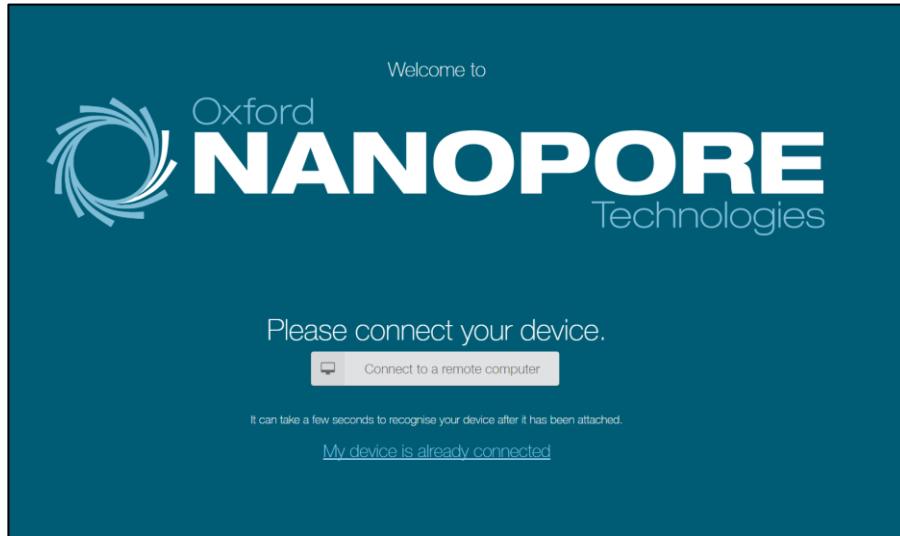
General information: Flow cells are shipped with a “Quality Control” DNA molecule (QC DNA) present in the buffer. This molecule produces a distinctive nanopore signal. The software uses this signal to validate the integrity of the nanopore array before use. It scans its micro-well to see if a pore is present, with the aim of determining the total number of functional pores present in a chip. (Figure 27). The Platform QC script first applies a triangle wave to assess the membrane. Then 180mV are applied for 30s, and different metrics contained in the signal, such as current levels, noise measurements and events that can help to classify the signal, are recorded. The pores are classified as:

- single pore: one nanopore in the micro-well
- multiple pores: multiple nanopores in the micro-well
- saturated: leaky membrane
- zero current: no nanopore in the micro-well

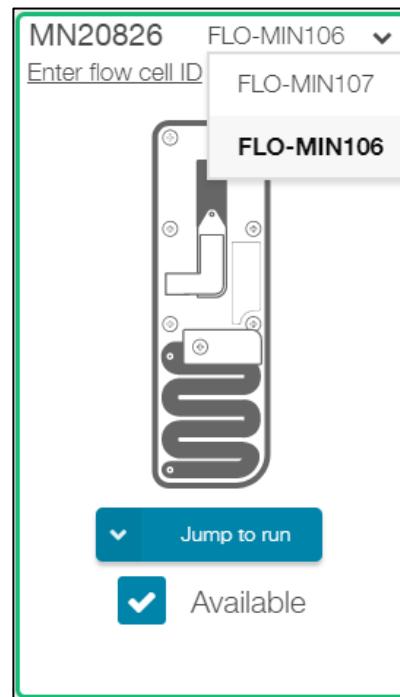
To class a single pore, the software searches for an open pore current level within a given range, combined with the distinct signal of the QC DNA strand that is present in the flow cell storage buffer. Active pores are reported in four groups, each of which may be used in turn when running long experiments e.g. a 48 hour sequencing run. Shorter experiments will use fewer than four groups.

Procedure:

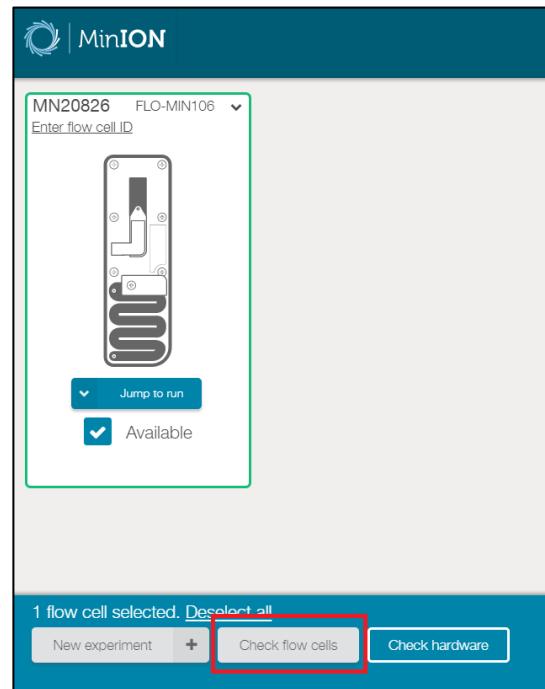
1. Take the flow cell to be tested out of the fridge and equilibrate it at room temperature for 30 minutes
2. Insert the flow cell into the MinION and close the lid.
3. Double-click the MinKNOW icon located on the desktop to open the MinKNOW GUI.
4. If your MinION was disconnected from the computer, plug it back in.



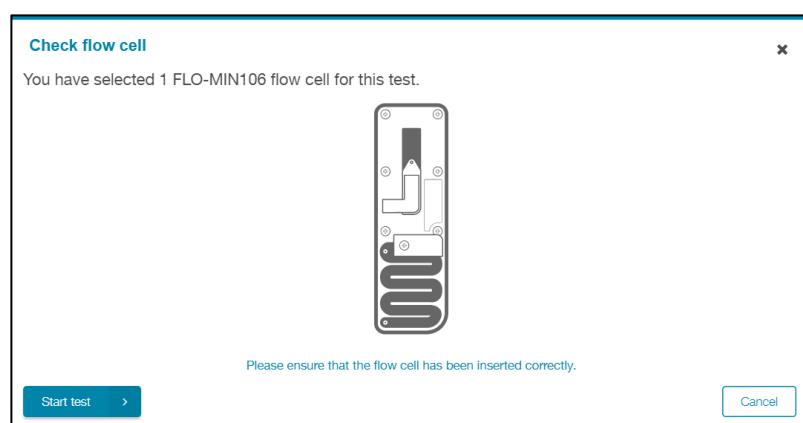
5. Choose the flow cell type from the selector box (you can find the flow cell type on the flow cell box).
6. Enter flow cell ID.
7. Then check the "Available" box.



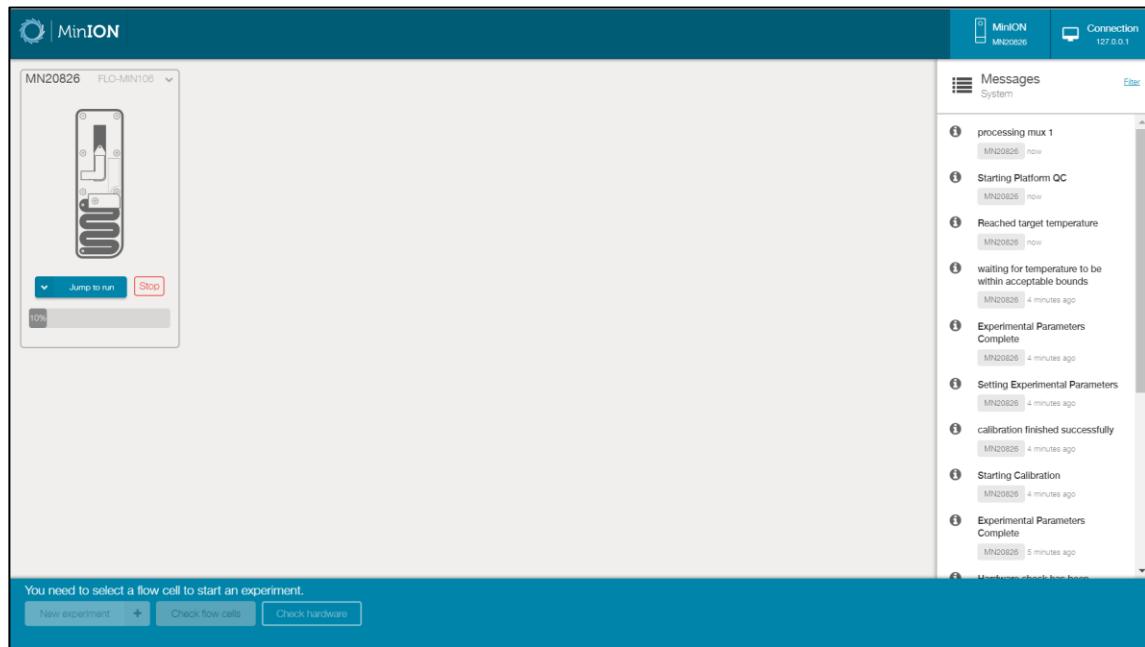
8. Click "Check flow cells" at the bottom of the screen.



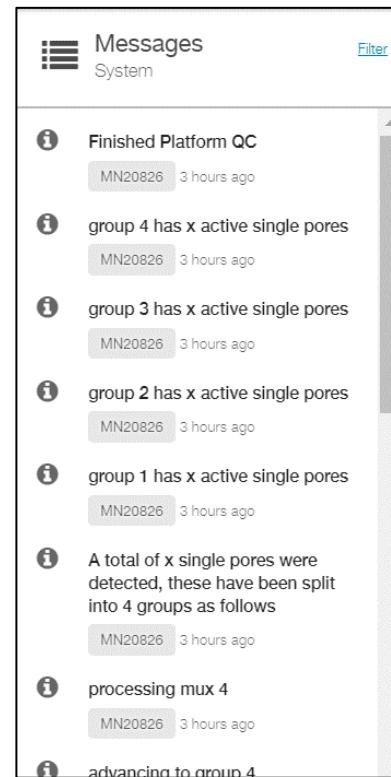
9. A screen will load displaying the number of flow cells selected for your test. It is important to choose the correct flow cell type in order to obtain an accurate result at Platform QC
10. Click "Start test".



11. The flow cell will become greyed out and will show the run progress.



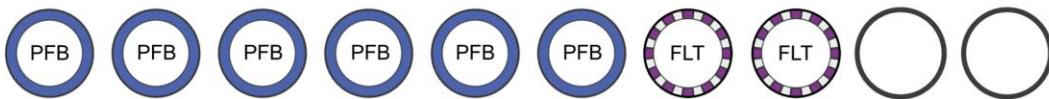
12. Check the number of active pores available for the experiment, reported in the System History panel when the check is complete.



13. The total number of pores available will be reported in the notification panel and should be at least **800 pores**.

Prepare the reagents for MinION sequencing

The extra reagents needed at this step are inside the “PromethION Flow Cell Priming Kit” box and are the following:



PFB : PromethION Flush Buffer

FLT : Flush Tether

Contents	Description	No. of tubes
PFB (blue cap)	PromethION Flush Buffer	6
FLT (purple stripe cap)	Flush Tether	2

1. Thaw at room temperature the following reagents:

- From the “SQK-PCS109 MinION library kit” box:
 - Sequencing Buffer (SQB)
 - Loading Beads (LB)
- From the “PromethION Flow Cell Priming Kit” box:
 - one tube of Flush Tether (FLT)
 - one tube of PromethION Flush Buffer (PFB)

As soon as thawing is complete place the tubes on ice.

2. Mix the Sequencing Buffer (SQB) and PromethION Flush Buffer (PFB) tubes by vortexing, spin down and return to ice.
3. Spin down the Flush Tether (FLT) tube, mix by pipetting, and return to ice.

4. Open the lid of the nanopore sequencing device and slide the flow cell's priming port cover clockwise so that the priming port is visible.

Priming the flow cell with the running buffer

The priming mix used below will flush away from the nanopores the already present “Quality Control” DNA molecules that are used to assess the number of pores (previous section). The priming mix will also introduce the substrate co-factor (ATP) of the motor enzyme that is used for the DNA translocation process in the pore. For example, the ATP is used as the substrate for the ATPase domain of the helicase which acts as the translocation motor. **The DNA translocation speed** is determined by:

- the intrinsic properties of the motor protein that translocates the DNA through the nanopore (for example the direct RNA sequencing uses a different and a slower motor protein than the one used for DNA sequencing)
- experimental conditions such as temperature and **co-factor concentration** (Figure 33).

The ATP substrate co-factor is present in both the PromethION Flush Buffer (PFB) and the sequencing buffer (SQB) used in the next session. The cholesterol tethers are found in the Flush Tether (FLT) tube. Because the cholesterol tethers tend to aggregate, they are not pre-bound on the sequencing adaptors as they will cause the aggregation of the molecules. In the contrary they are first introduced in the flowcell where they insert into the lipid bilayer. Afterwards the sequencing adaptor conjugated cDNA molecules are introduced in the flow cell and these molecules hybridize with the free DNA end of the cholesterol tethers which eventually restricts their movement on the lipid bilayer and brings them next to the nanopore.

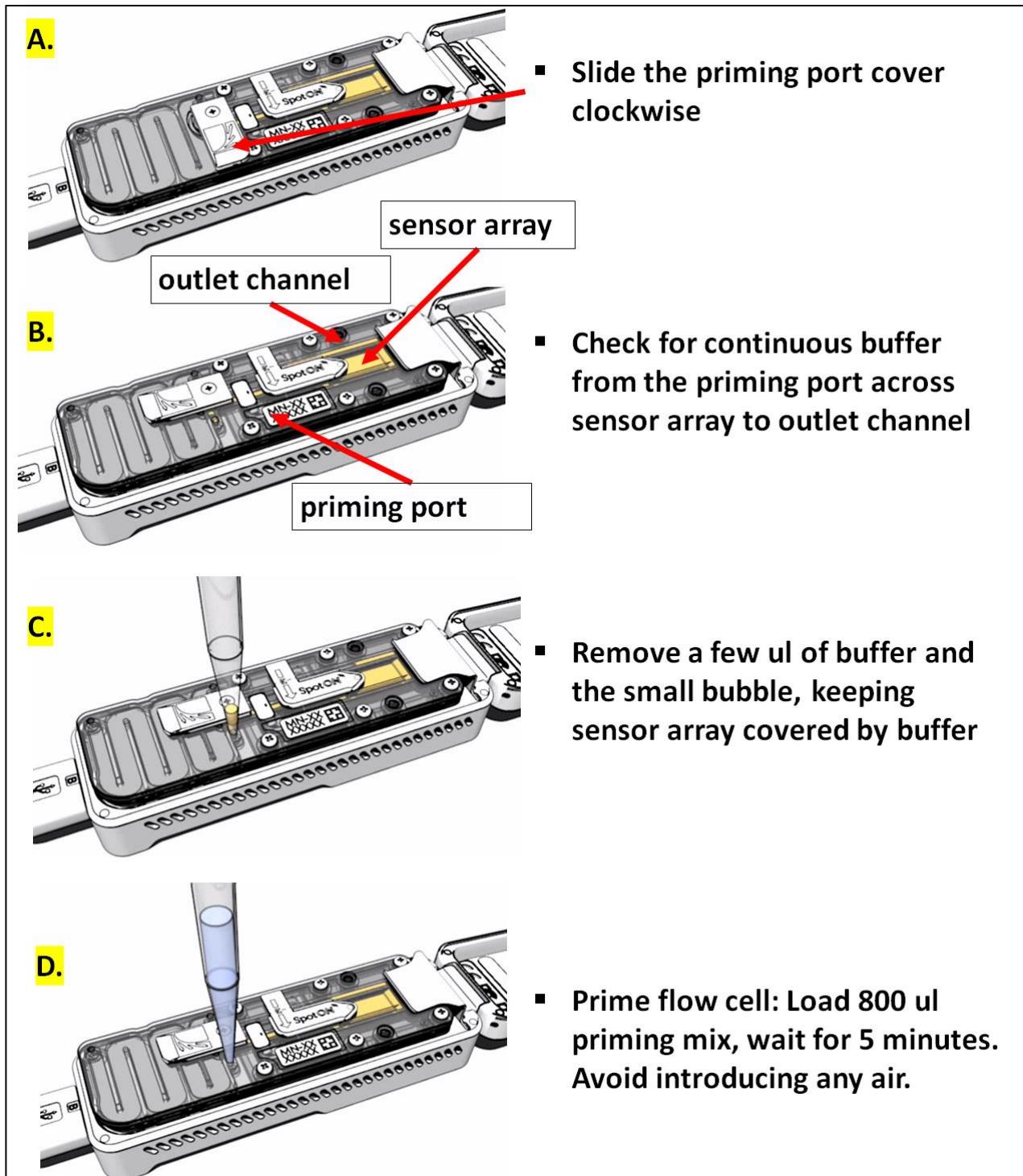
IMPORTANT: Take care to avoid introducing any air during pipetting. Care must be taken when drawing back buffer from the flow cell. The array of pores must be covered by buffer at all times. Removing more than 20-30 µl risks damaging the pores in the array.

Steps:

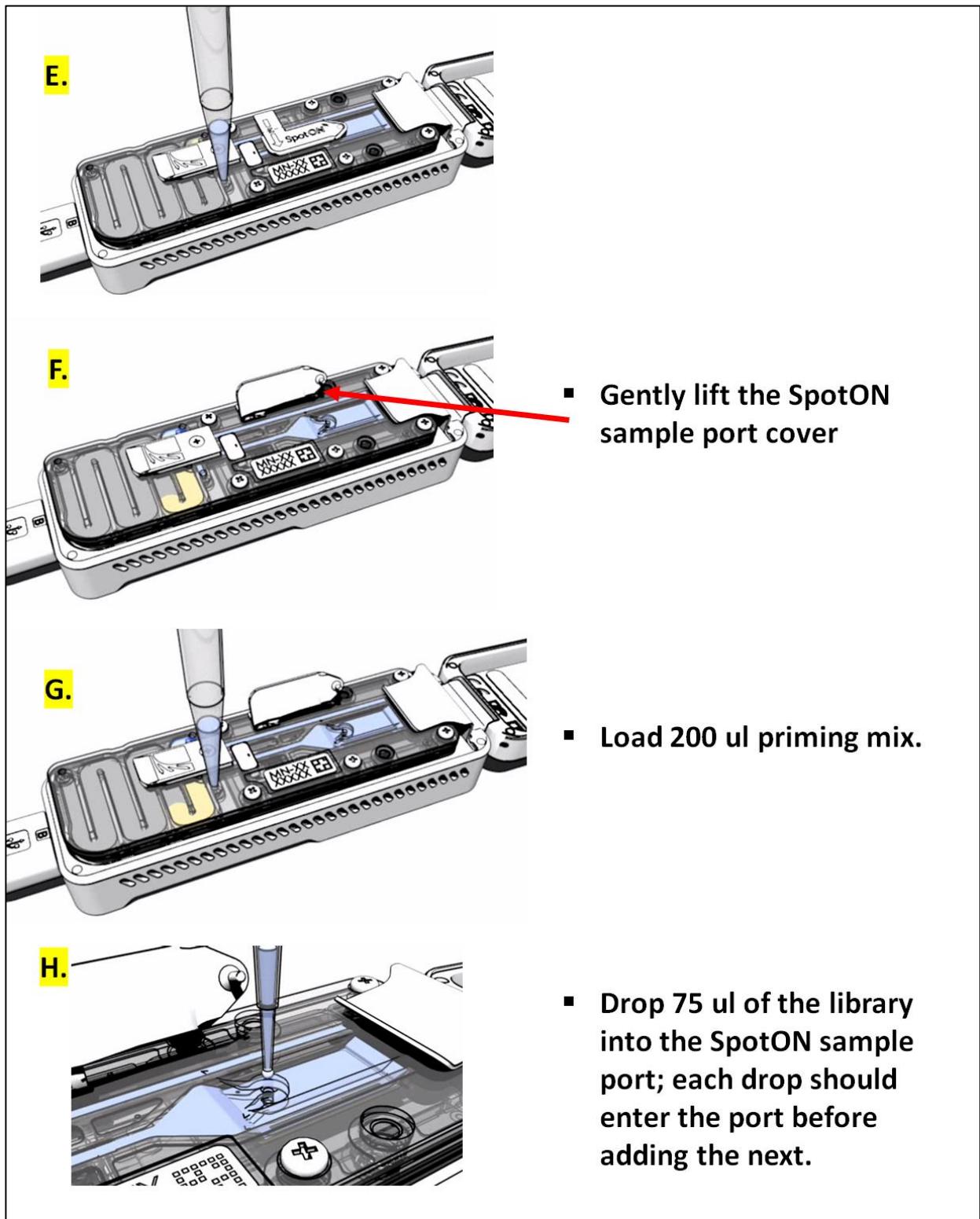
1. Flip back the MinION lid and slide the priming port cover clockwise so that the priming port is visible (Figure 34A).
2. After opening the priming port, check for small bubble under the cover. Draw back a small volume to remove any bubble (a few μ l):
 - I. Set a P1000 pipette to 200 μ l .
 - II. Insert the tip into the priming port.
 - III. Turn the wheel until the dial shows 220-230 μ l, or until you can see a small volume of buffer entering the pipette tip (Figure 34C).
 - IV. Visually check that there is continuous buffer from the priming port across the sensor array (Figure 34B).
3. Prepare the flow cell priming mix: add 46 μ l of thawed and mixed Flush Tether (FLT) directly to the tube of thawed and mixed PromethION Flush Buffer (PFB), and mix by pipetting up and down. **The new solution is called priming mix.**
4. Load 800 μ l of the priming mix into the flow cell via the priming port, avoiding the introduction of air bubbles (Figure 34D). Wait for 5 minutes. *The step practically replaces the storage buffer with the sequencing running buffer* (Figure 35).

Motor protein	Fuel concentration	Speed, b/s
E8	30 mM	450
E7	11 mM	250
E6	10 mM	70

Figure 33. Type of motor protein and fuel (co-factor) concentration (ATP) as determinants of sequencing speed.



(the picture continues in the next page)



(the picture continues in the next page)

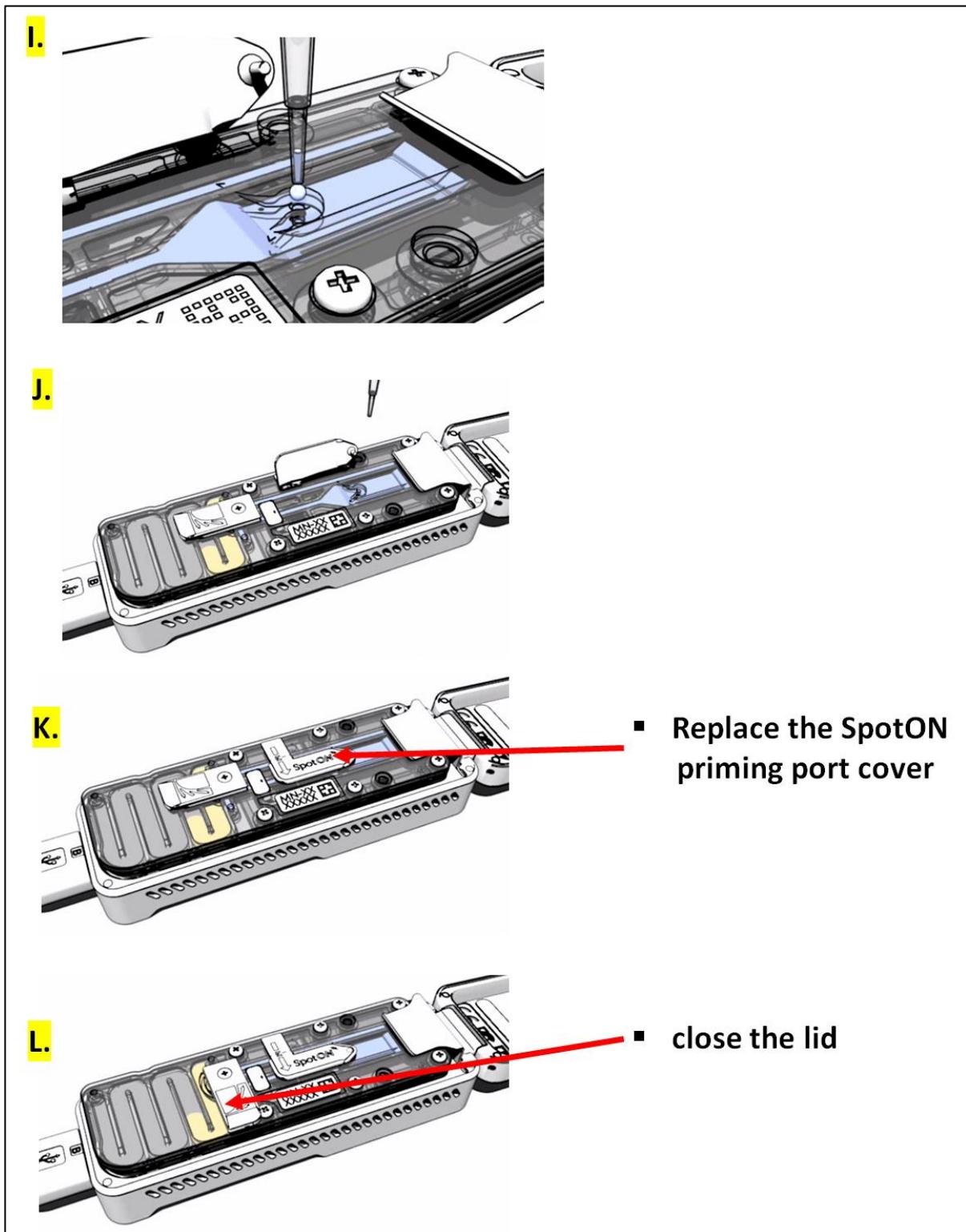


Figure 34 (current and previous two pages). Procedure for priming the ONT MinION flow cell and subsequently loading the cDNA library.

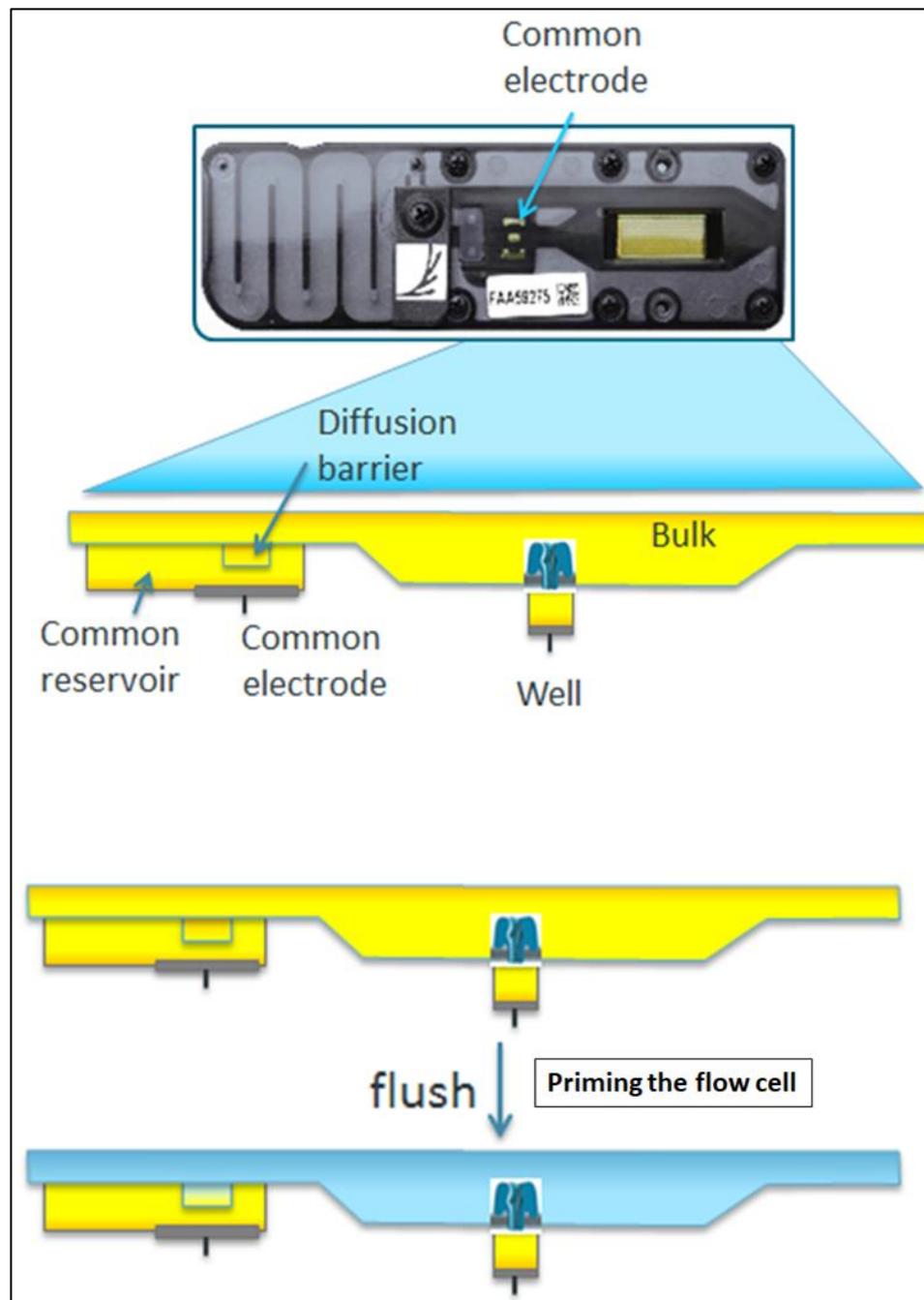


Figure 35. Effect of the priming mix buffer on the different compartments of the ONT MiniON flowcell. Before library loading, the flow cell is flushed with “priming mix” (shown as blue), which displaces the storage buffer from the bulk compartment. The storage and running buffers contain different concentrations of charged species, and will affect the pores and DNA translocation differently.

Library loading on the flow cell

IMPORTANT: The Loading Beads (LB) tube contains a suspension of beads. These beads settle very quickly. It is vital that they are mixed immediately before use. *The “Library Loading Beads” are practically Sepharose beads that are used to trap the library of cDNA molecules. As the Sepharose beads precipitate, due to gravity, on the surface of the lipid bilayer they bring the cDNA molecules close to the lipid bilayer. The free DNA end of the cholesterol tethers, that are already inserted into the lipid bilayer, will hybridize with the sequencing adaptor conjugated cDNA molecules and will restrict the movement of the cDNA molecules on the lipid membrane. The electric field will then drive the DNA molecules into the nanopore.*

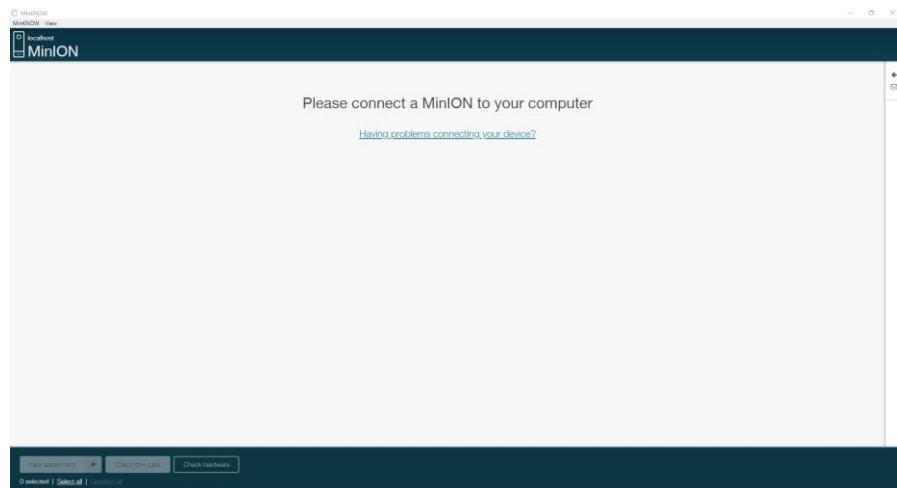
1. In a new tube, prepare the library for loading as follows:

Reagent	Volume
Sequencing Buffer (SQB)	37.5 µl
Loading Beads (LB), mixed immediately before use	25.5 µl
DNA library	12 µl
Total	75 µl

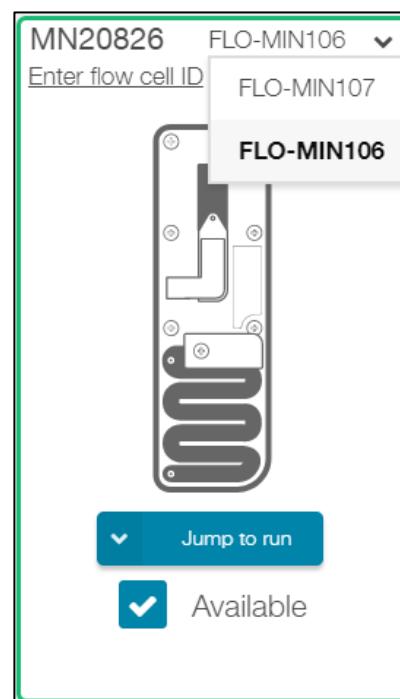
2. Gently lift the SpotON sample port cover to make the SpotON sample port accessible (Figure 34F).
3. Load 200 µl of the “priming mix” (see previous step) into the flow cell via the priming port (not the SpotON sample port), avoiding the introduction of air bubbles (Figure 34G).
4. Mix the prepared library gently by pipetting up and down just prior to loading.
5. Add 75 µl of sample to the flow cell via the SpotON sample port in a dropwise fashion (Figure 34H). Ensure each drop flows into the port before adding the next (Figure 34I).
6. Gently replace the SpotON sample port cover, making sure the bung enters the SpotON port, close the priming port and replace the MinION lid (Figure 34K).

Starting a sequencing run

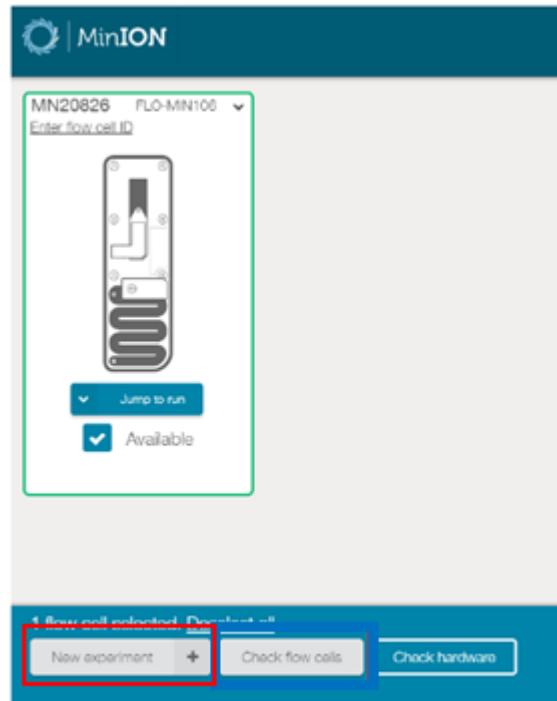
1. Double-click the MinNOW icon located on the desktop to open the MinNOW GUI.
2. If your MinION was disconnected from the computer, plug it back in.



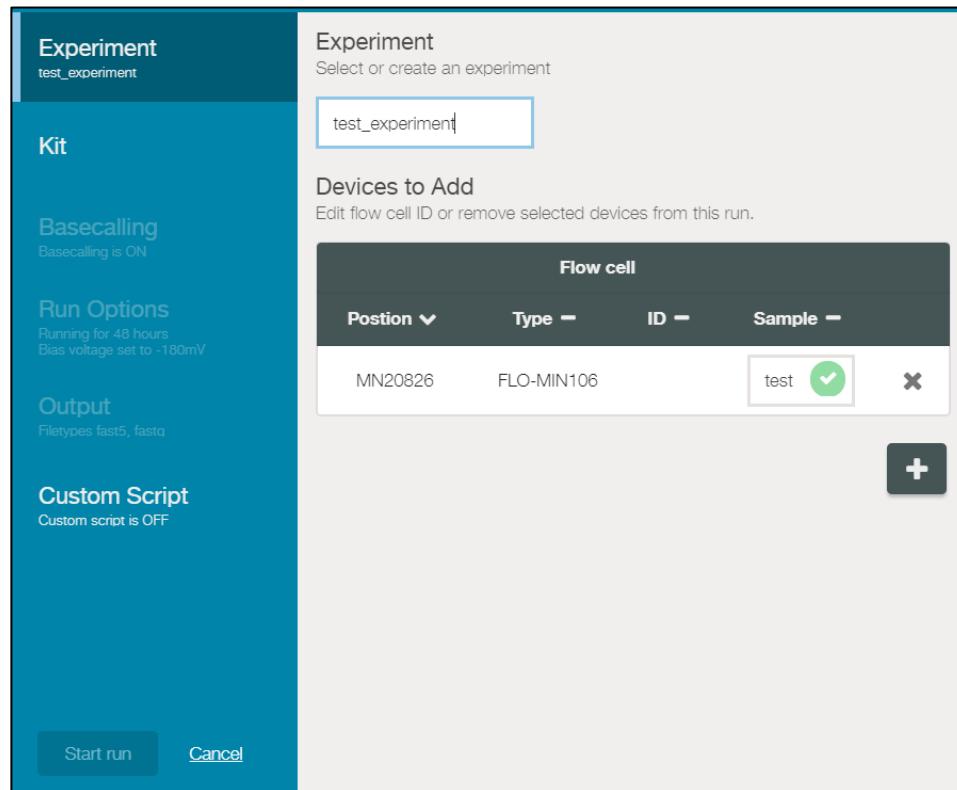
3. Choose the flow cell type from the selector box (you can find the flow cell type on the flow cell box). Then check the "**Available**" box.



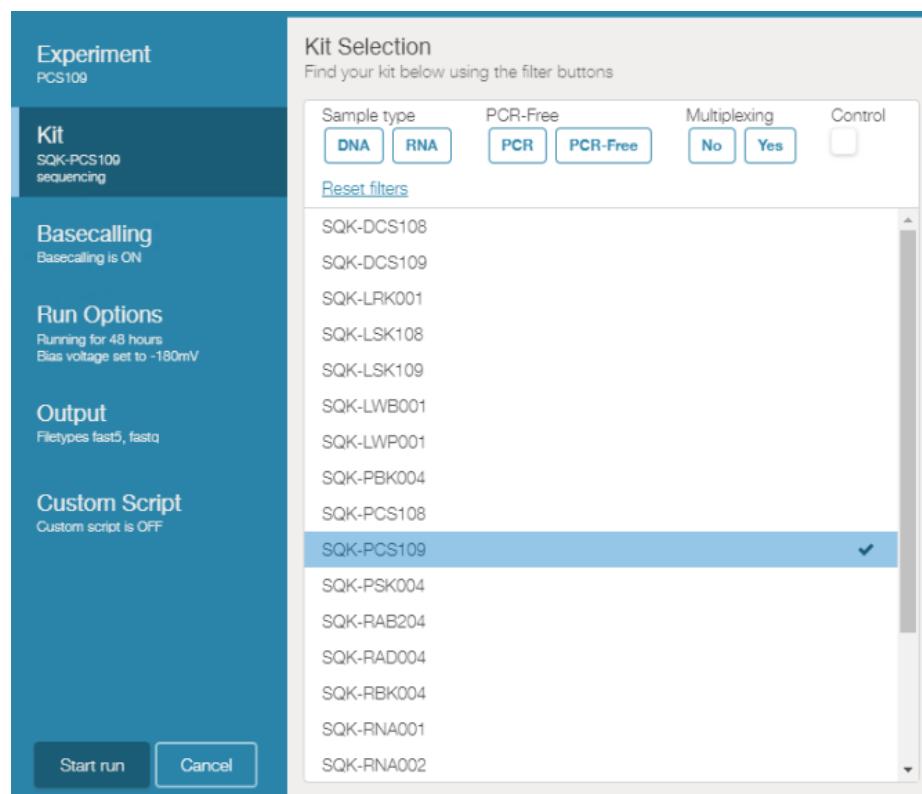
4. Click the "New Experiment" button at the bottom left of the GUI.



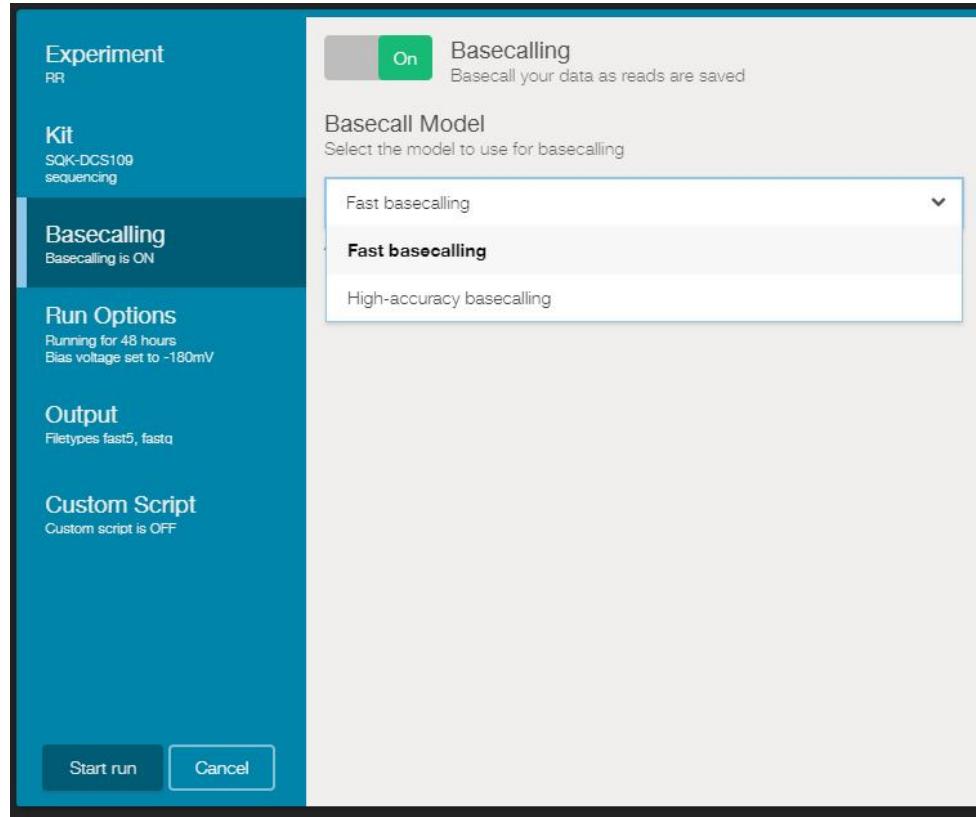
5. On the New experiment popup screen, select the running parameters for your experiment from the individual tabs.
6. The experiment name tab will show the chosen flow cell. An experiment name can then be assigned.
7. The other tabs will not become available until an experiment name has been provided.



8. Click the “Kit” tab. This selection will provide a dropdown of available kits. Select **SQK-PCS109**.



9. Click the “**Basecalling**” tab. Select live basecalling (“On” option) on the instrument.



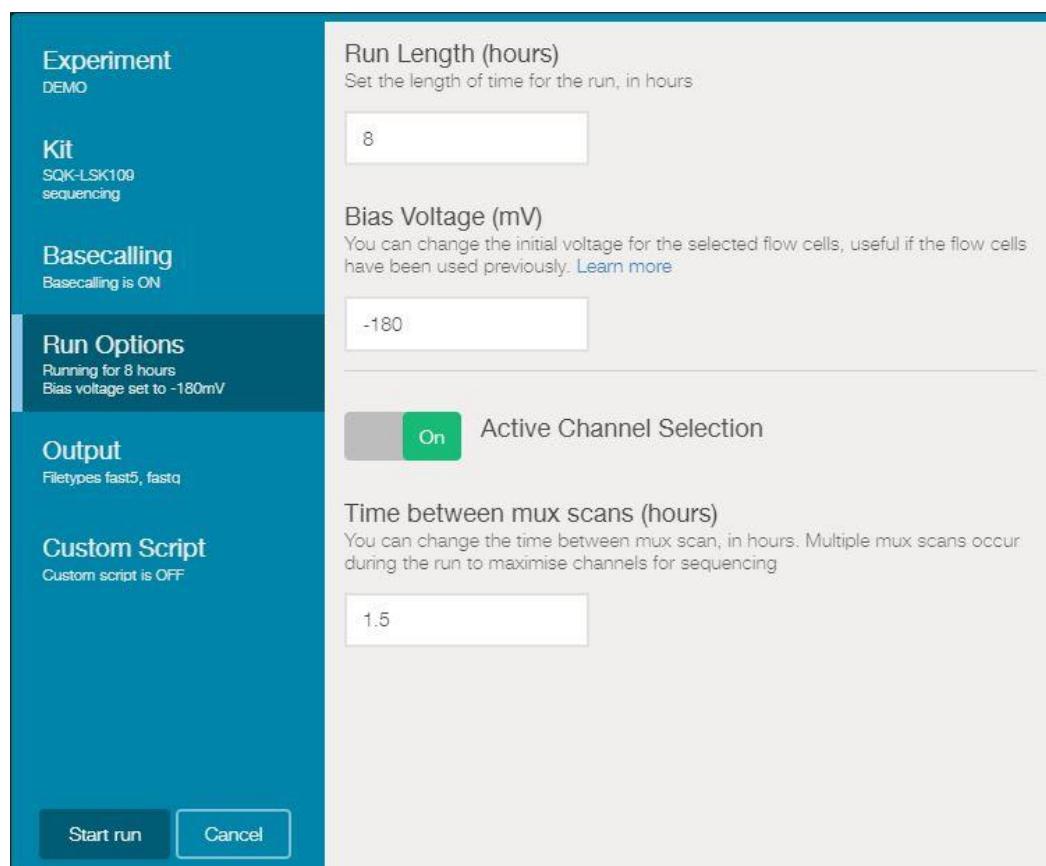
10. Select the “**Fast basecalling**” option.

- Model selection The MinKNOW's basecaller - Guppy - provides two basecalling models for nanopore data. These are:
 - Fast - This model is able to keep up with a full experiment on a GridION or MinIT. It implements a lite version of the Flip-flop algorithm. This delivers the same level of accuracy as the previous default transducer models, which have been deployed prior to the introduction of Flip-flop.
 - HAC - The High accuracy (HAC) Flip-flop model provides a higher consensus/raw read accuracy than the 'Fast' model. It contains a more computationally intense Flip-flop architecture that can deliver higher accuracy

using the same data produced by nanopore sequencing. It is currently 5-8 times slower than the fast model so users should ensure their data transfer, diskspace and device utilization is scaled appropriately for this.

- The run options tab provides variables for run time and starting voltage. By default these will be 48 hours and -180 mV. **If the flow cell was already used in a previous run the last bias voltage value should be used on the console.**

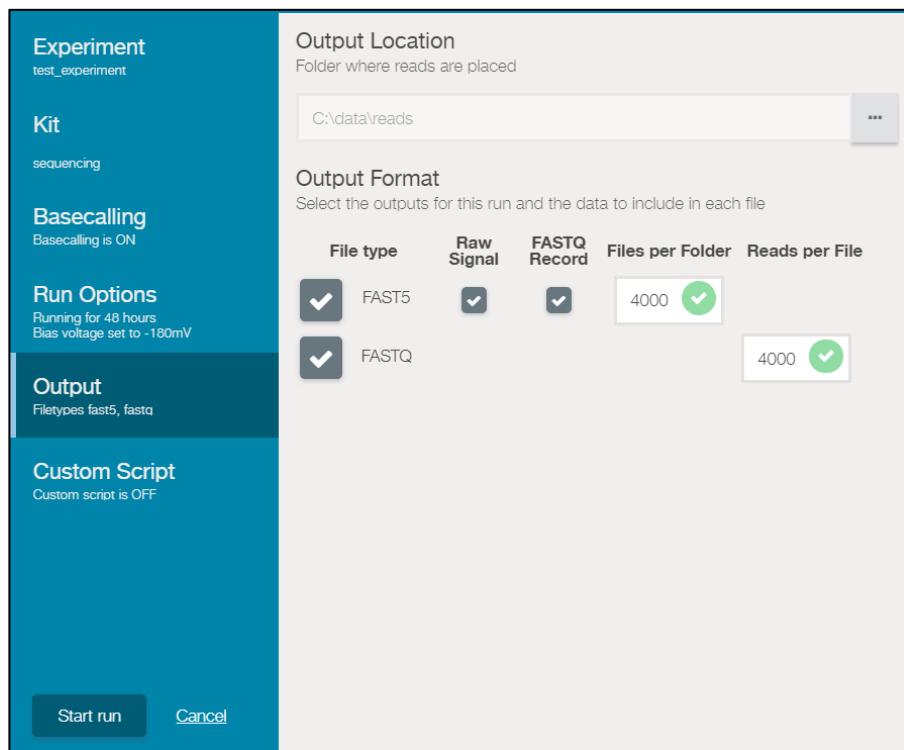
- Active Channel Selection refers to a feature introduced in MinKNOW v2.2. If a channel is in the "Saturated" or "Multiple" state, the software instantly switches to a new channel in the group. If a channel is "Recovering", MinKNOW will attempt to revert the channel back to "Pore" or "Sequencing" for ~5 minutes, after which it will select a new channel in the group. This maximizes the number of channels sequencing at the start of the experiment.



- Specify to produce:
 - Both .fast5 and .fastq files

13. The options available for each file type are:

- Output settings - FASTQ: The number of basecalled reads that MinKNOW will write in a single file. By default this is set to 4000. **Change the value to 10000000 to output everything into one file.**
- Output settings - FAST5: The number of files that MinKNOW will write to a single folder. By default this is set to 4000
- There are additional options for the information contained in .fast5 files. By default these are all stored within the .fast5 file:
 - - raw data
 - - event data
 - - .fastq basecalls



14. Click "Start run".

15. The pop up box will disappear, and the flow cell will become greyed out.



16. Allow the script to run to completion.
17. The MinKNOW Experiment page will indicate the progression of the script; this can be accessed through the "Experiment" tab that will appear at the top right of the screen.
18. Monitor messages in the Message panel in the MinKNOW GUI
19. **Important:** To ensure that the experiment runs to completion, all sleep modes (including screensavers and log-offs) should be disabled on the host computer.

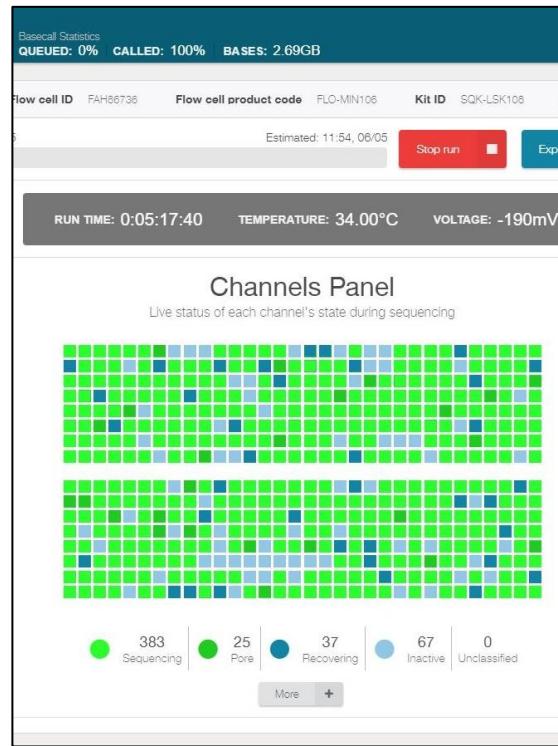
The running experiment screen

As the sequencing protocol starts, a scan of the sensor array begins before the sequencing. This multiplex scan “**MUX scan**” allows MinKNOW to prioritize the order in which the nanopores are used, maximizing the data output in the initial stages of the run. In a similar fashion to the Platform QC, a 180 mV voltage is applied to each MUX in turn. If there are library DNA strands in any of the pores, these are now used to classify the micro-wells. Micro-wells are prioritized for sequencing based on 'strand' first, then 'pore', then 'unavailable' pores:

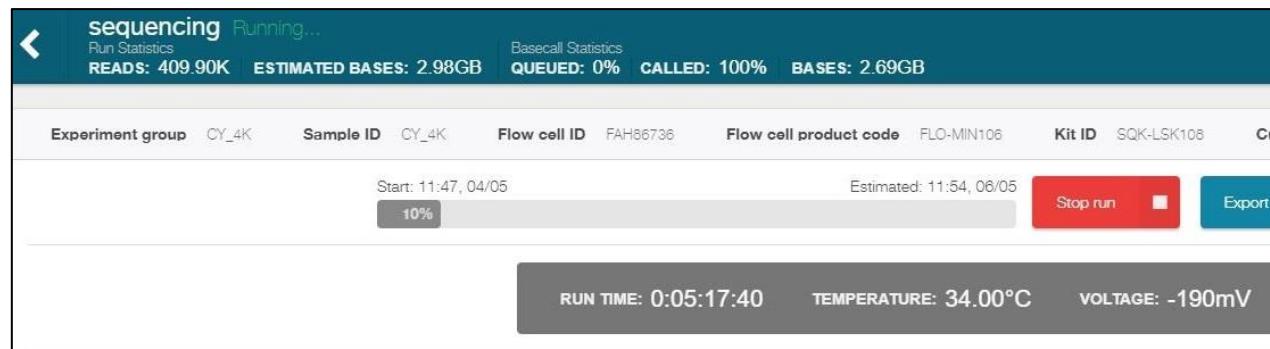
- “strand”: a DNA strand translocating through a single pore
- “pore”: an open pore without DNA
- “unavailable”: signals that the software is not confident about

The software then groups the best 512 micro-wells for sequencing plus 2nd, 3rd and 4th choices that can be utilized during a sequencing run. These are the groups 1, 2, 3 and 4 reported on the GUI. Sequencing then begins from group 1 but in this version of the software the groups 2, 3 and 4 are not used as is. On the contrary every 90 minutes another scanning of the sensor array takes place in order to select the best 512 wells from the ones that are left. This ensures that maximal sequencing data are collected in the shortest possible time. Additionally, if during the sequencing run a nanopore channel stops functioning then another microwell with a working nanopore, from the remaining 3 microwells of the same signal reading channel, is selected. At the end of the sequencing run, MinION can be restarted to keep sequencing the same sample with the rest of the remaining functional pores until there is no functional nanopores left.

As the experiment progresses, you can obtain real-time feedback for your flow cell from the running experiment screen. For additional information regarding the plots on this screen, see the following:

Experiment summary information panel:

In more details we see the following panels:

Run and basecall statistics panel:

- The status bar at the top of the screen will display information relating to the experiment as a whole:
 - Run statistics: The total number of reads and bases produced across the experiment
 - Basecall statistics: There are two values for basecalled reads:

- Basecalled reads as a percentage of the total reads produced across the experiment.
This gives an indication as to the size of the queue for reads to be basecalled.
 - Total number of reads basecalled across the experiment
 - Run time: The duration of the experiment
 - Temperature: The heatsink temperature of the selected position
 - Voltage: The applied potential of the position at that point in time
- Check the temperature is approximately 34° C (Figure 36).
- The MinION is able to maintain a temperature of 34° on a typical lab bench when the local ambient conditions are between 19.5° C and 24.5° C. However, there are a number of external factors which can disrupt the local conditions and which need to be taken into account, for example warm air expelled from laptops, or cool air from a fan or air conditioning system increasing airflow around the MinION.
 - The MinION takes approximately 10 minutes to get to temperature.

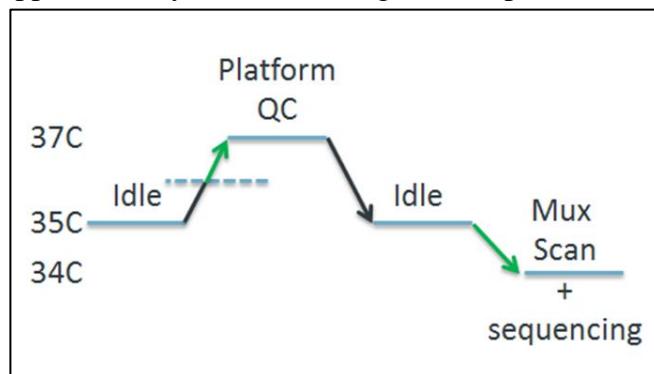
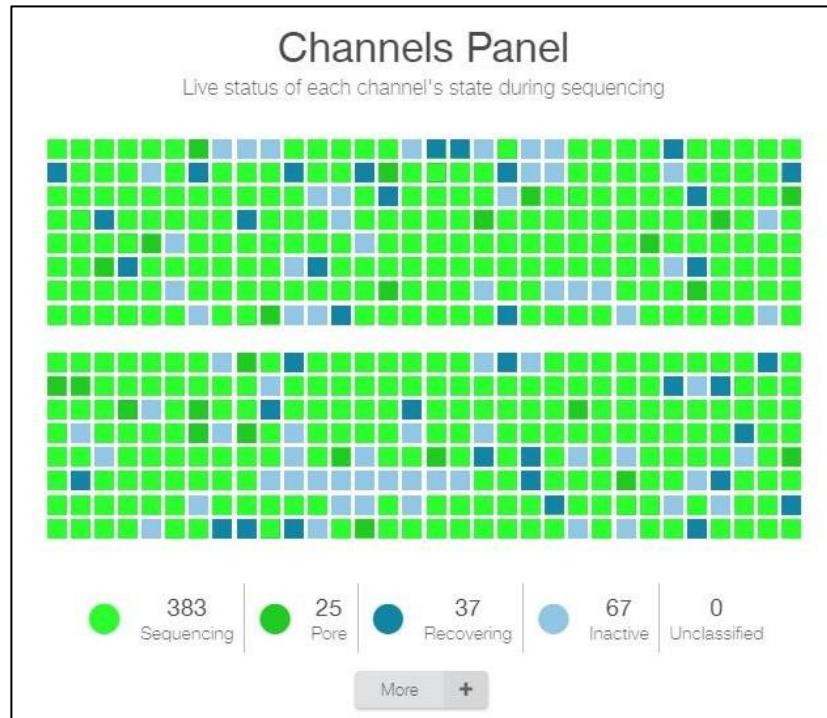


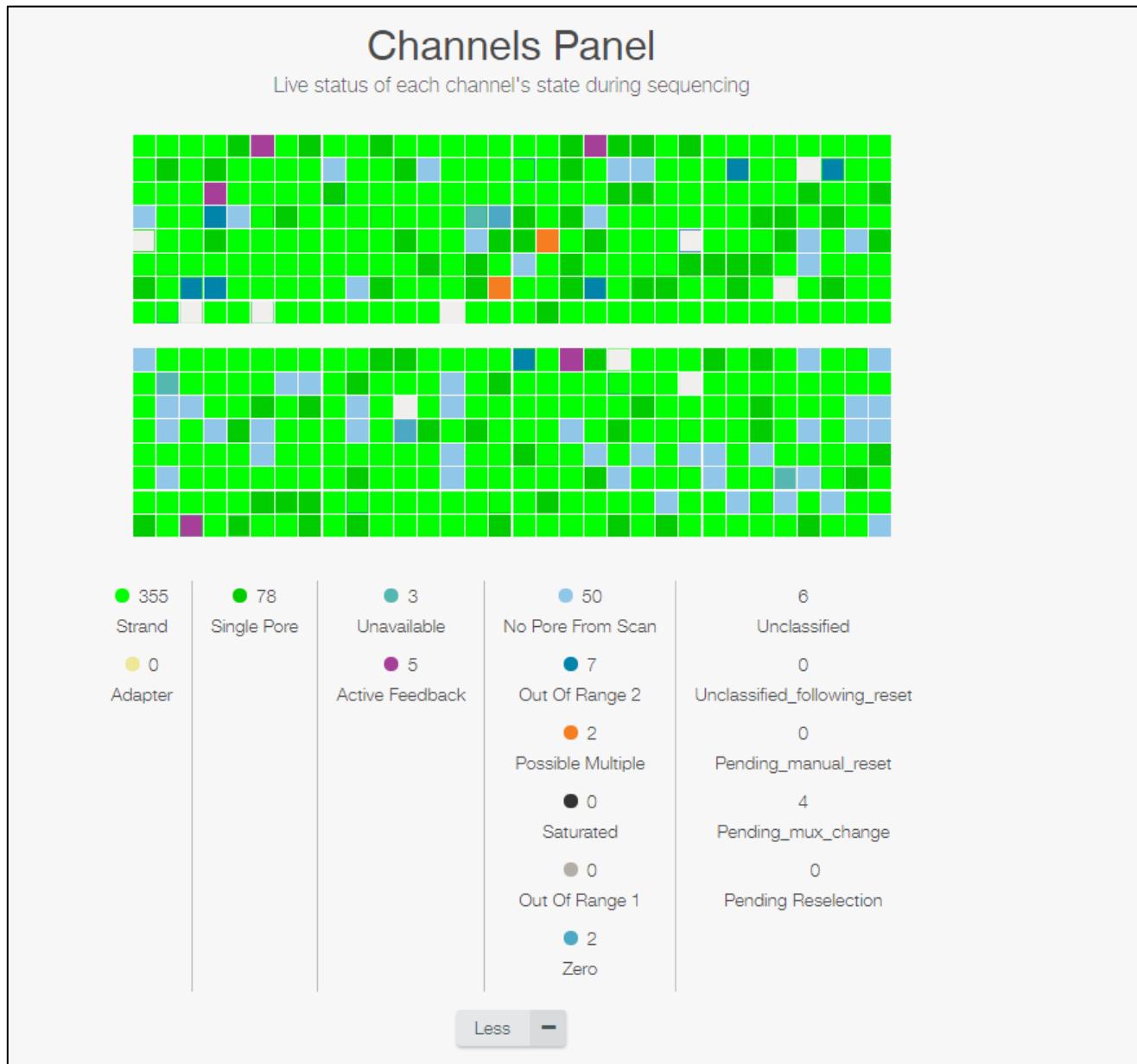
Figure 36. A range of set temperatures is used for various steps of the sequencing run. Higher temperatures increase the mobility of charged species and thereby increase the current.

Channel status panel:

- Check the number of active pores reported in the MUX scan are similar (within 10-15%) to those reported at the end of the Platform QC.
 - If there is a significant reduction in the numbers, restart MinNOW.
 - If the numbers are still significantly different, close down the host computer and reboot.
 - When the numbers are similar to those reported at the end of the Platform QC, restart the experiment on the Connection page. There is no need to load any additional library after restart.
- Stopping the experiment is achieved by clicking "Stop experiment" button at the top of the screen.
- Check pore occupancy in the channel panel at the top of the experimental view.
 - I. A good library will be indicated by a higher proportion of light green channels in “Sequencing” mode, than in “Pore” mode. The combination of “Sequencing”

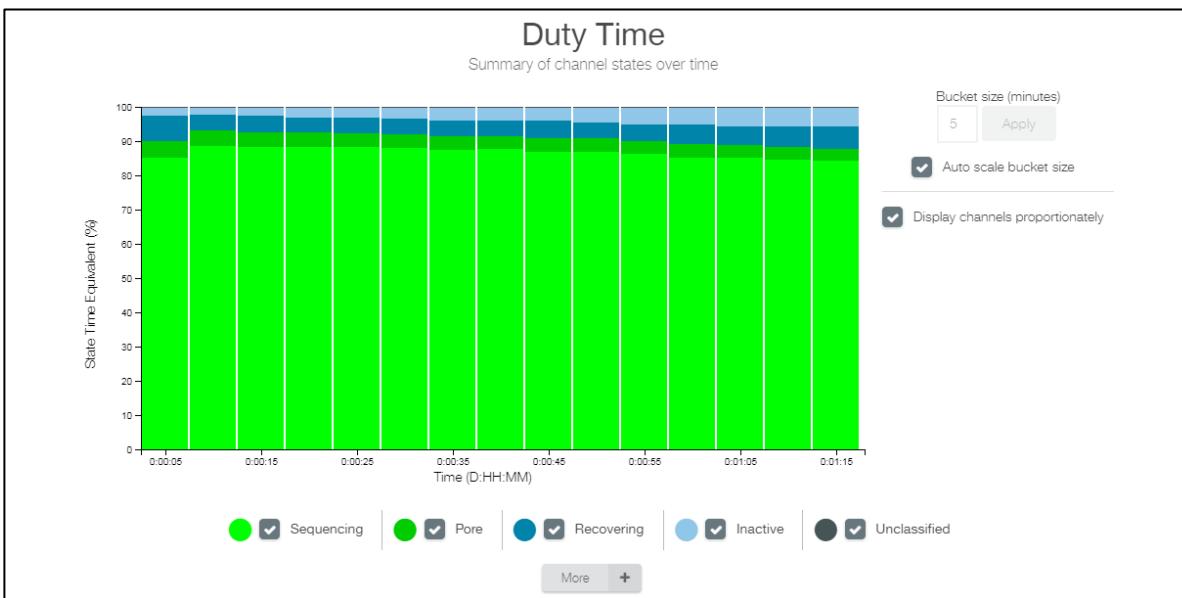
and “Pore” modes indicates the number of active pores at any point in time. A low proportion of Sequencing channels will reduce the throughput of the run.

- II. **Recovering** indicates channels that may become available for sequencing again. A high proportion of this may indicate additional clean up steps are required during your library preparation.
- III. **Inactive** indicates channels that are no longer available for sequencing. A high proportion of these as soon as the run begins may indicate an osmotic imbalance.
- IV. **Unclassified** are channels that have not yet been assigned one of the above classifications
 - Clicking on the "More" button reveals a more detailed array of channel states:
 - a) **Strand:** the channel is in strand
 - b) **Adapter:** the channel has just captured a new strand
 - c) **Single pore:** the channel appears to show a single pore. Available for sequencing
 - d) **Unavailable:** the channel appears to show a pore that is currently unavailable for sequencing
 - e) **Active feedback:** the channel is reversing the current flow to eject the analyte
 - f) **No pore from scan:** the Mux scan has not detected a pore in the well
 - g) **Out of range 2:** current level is between 10 and 9999 pA. Currently unavailable for sequencing
 - h) **Possible multiple:** the channel appears to show more than one pore. Unavailable for sequencing
 - i) **Saturated:** the channel has switched off due to current levels exceeding hardware limitations
 - j) **Out of range 1:** current level is between -5 and -9999 pA. Currently unavailable for sequencing
 - k) **Zero:** Current level is between -5 and 10 pA. Currently unavailable for sequencing.

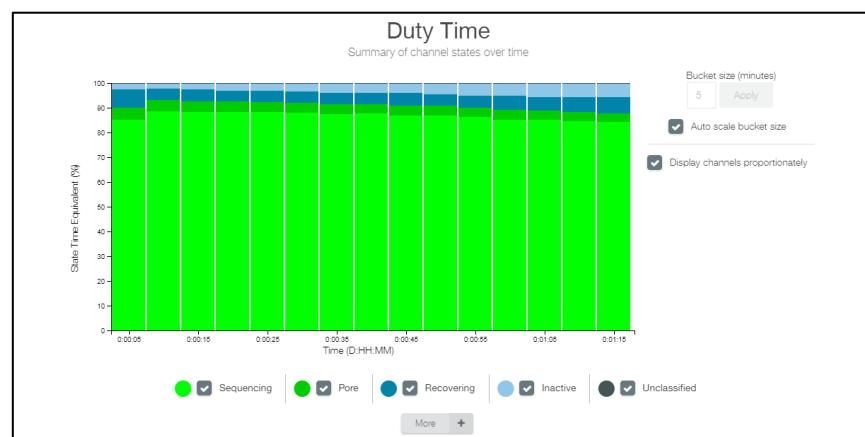
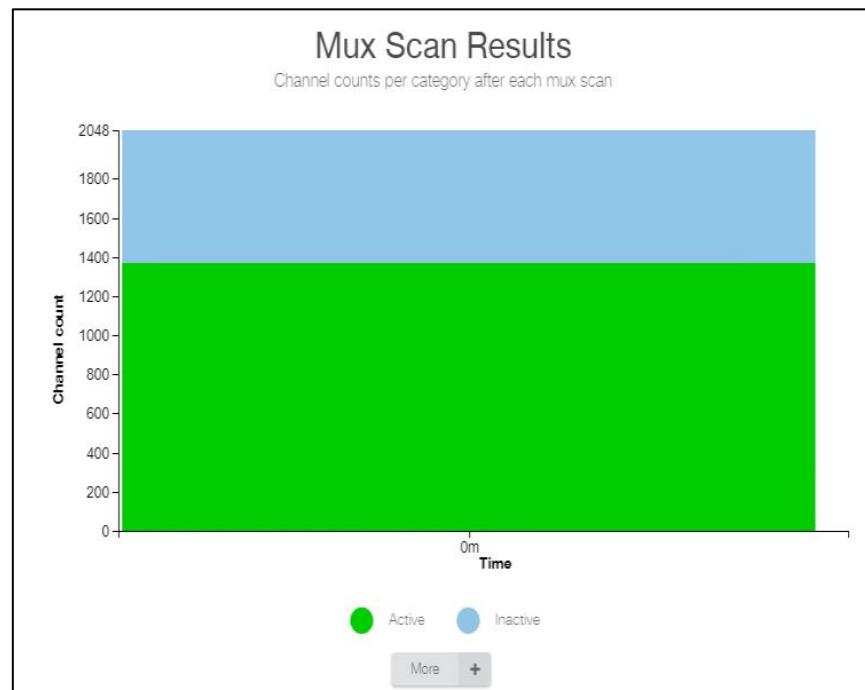


- Monitor the pore occupancy
 - It is recommended that you monitor the pore occupancy for the first 30 minutes of your sequencing experiment. If fewer than ~50% (for cDNA sequencing) of all active pores are in strand then it is probably good to stop the run, wash the flowcell and store it for use in another run. **A less than optimal pore occupancy will lead to a higher than expected number of pores in the open state, which results in an increased rate of electrolyte utilization and a rapid decline in the number of**

“good” pores following the platform QC step. This will impact the output of the run and the longevity of the flow cell.

*Duty time plots panel:*Duty time plot of the first 13 minutesDuty time plot of the first 75 minutes

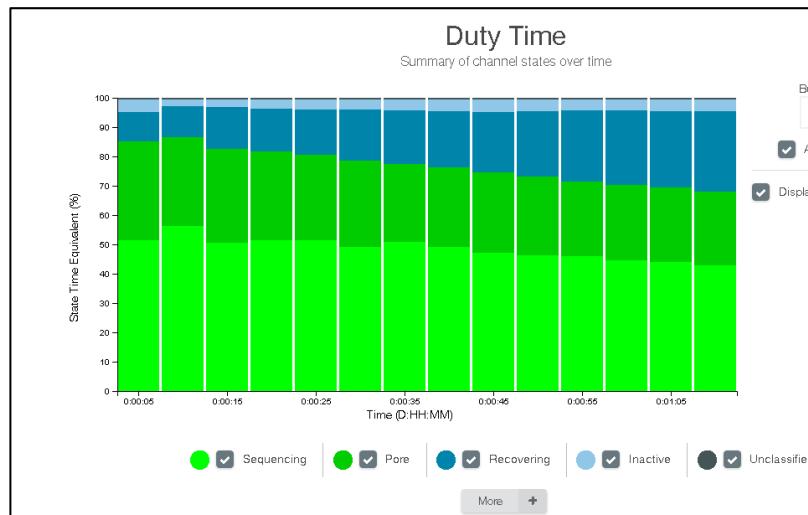
- The duty time plot summarizes the channel states over time.
- Each bar shows the sum of all channel activity in a particular amount of time. This time bucket defaults to 1 minute, and scales to 5 minutes automatically after reaching 48 buckets. However, bucket size can be adjusted in the "Bucket size" box to the right of the graph.
 - The graph populates over time and can be used as a way to assess the quality of the sequencing experiment and make an early decision whether to continue with the experiment or to stop the run.
 - The Duty Time feature in the MinNOW software can be used to judge the quality of your experiment.
 - It is recommended to observe the duty time plot populating over the first 30 min-1 hr of the sequencing run. By this time, the channel state distribution will give an indication whether the DNA/RNA library is of a good quality, and whether the flow cell is performing well.
 - **If Active Channel Selection is enabled during the run, the software instantly switches to a new channel in the group if a channel is in the “Saturated” or “Multiple” state, or after ~5 minutes if a channel is “Recovering”. This feature maximizes the number of channels sequencing at the start of the experiment, however this may also result in an artificially high number of "Sequencing" or "Pore" channels in the duty time plot. For this reason, the Mux Scan Results plot, which shows the true distribution of channel states at the point of the most recent mux scan, is better to be used.**
 - Below are some examples of good and bad sequencing runs:
 - ✓ **Good quality library:** A good quality library will result in most of the pores being in the "Sequencing" state, and very few in "Pore", "Recovering" or "Inactive". A library that looks like this is likely to give a good sequencing throughput.

Duty time plot:**Mux scan results:**

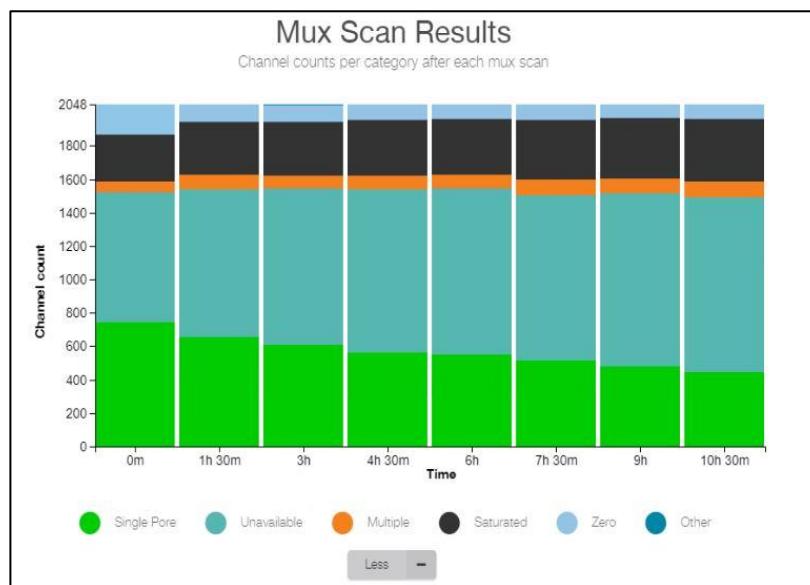
- ✓ **Channel blocking:** Under certain conditions (usually the presence of contaminants in the library or due to secondary structure present on the

molecules), pores may become blocked and therefore unable to sequence. This manifests itself as a build-up of "Recovering" pores over time.

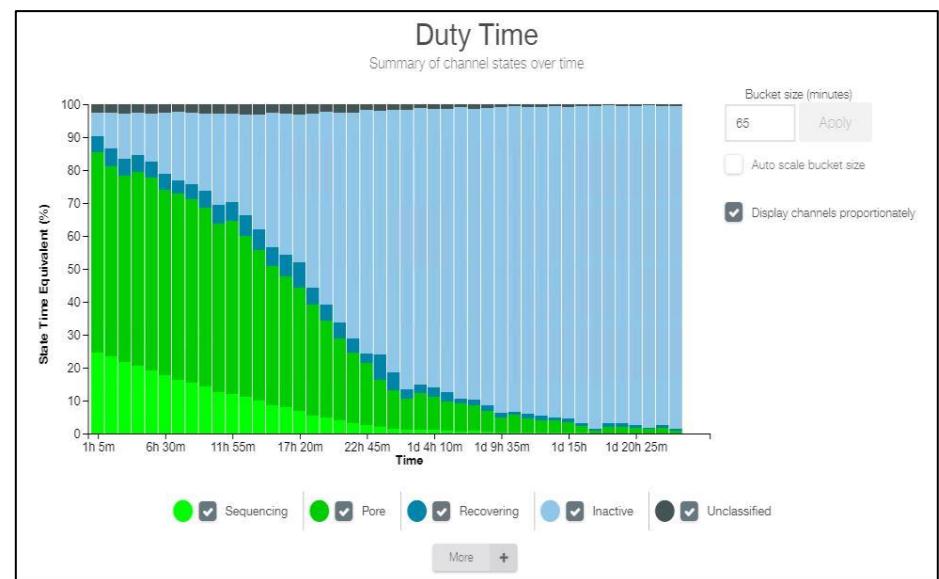
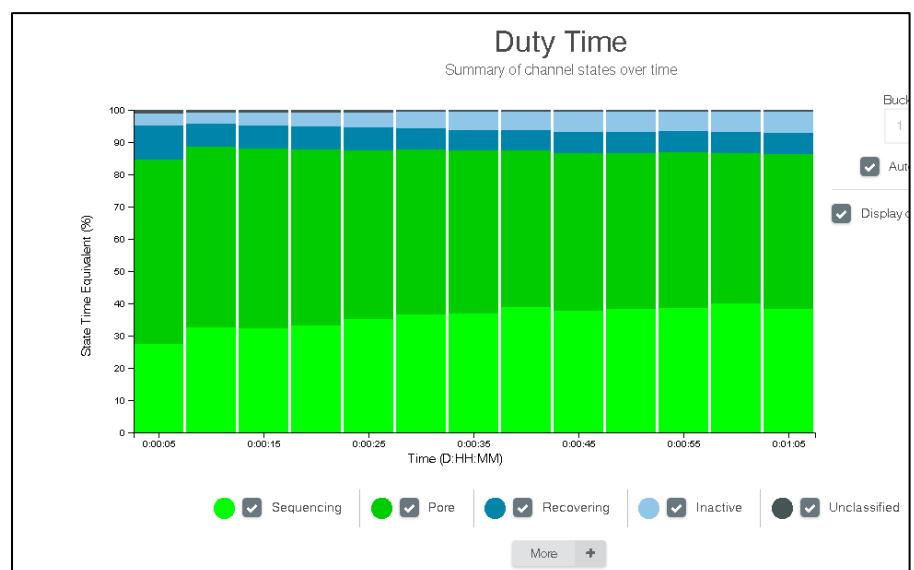
Duty time plot:



Mux scan results:



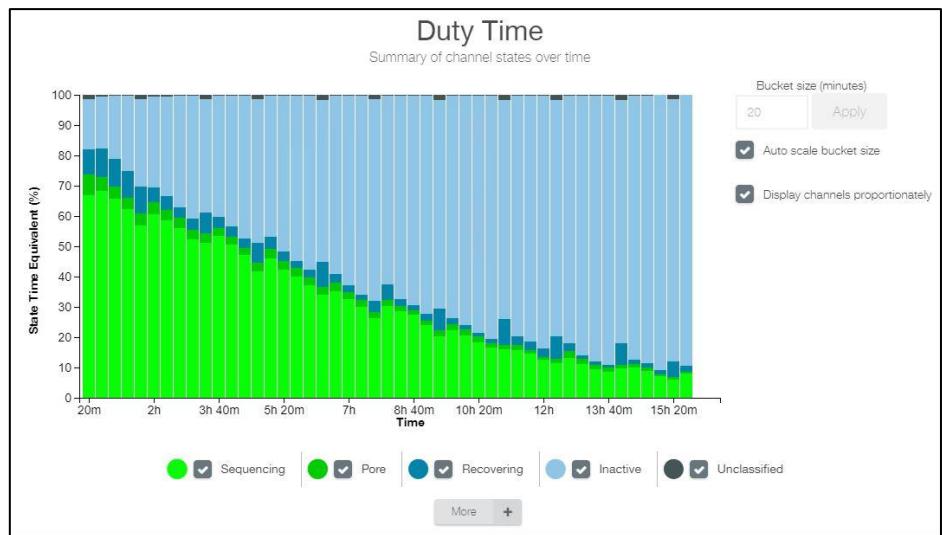
- ✓ **Low pore occupancy:** If there was insufficient starting material, or some sample has been lost during library prep, or the sequencing adapters did not ligate well to the strand ends, the duty time plot will show a high ratio of "Pore" to "Sequencing" states, meaning that only a limited number of pores are sequencing at any one time.

Duty time plot with Active Channel Selection switched on:Duty time plot with Active Channel Selection switched off:

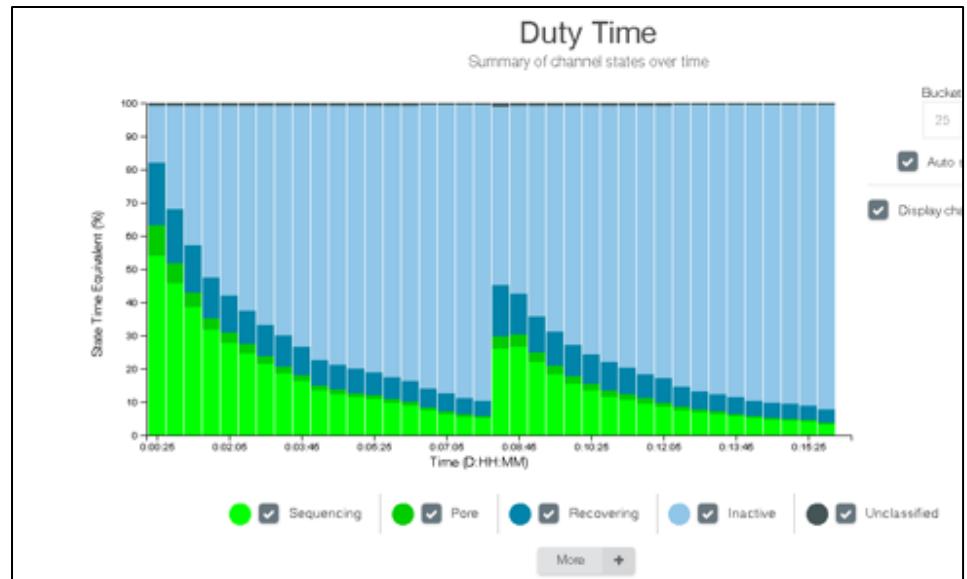
- ✓ **Flow cell failure due to osmotics or surfactants:** If the duty time plot shows a high number of 'Inactive' channels building up over time, this could indicate that the channels or membranes have been damaged by e.g. air

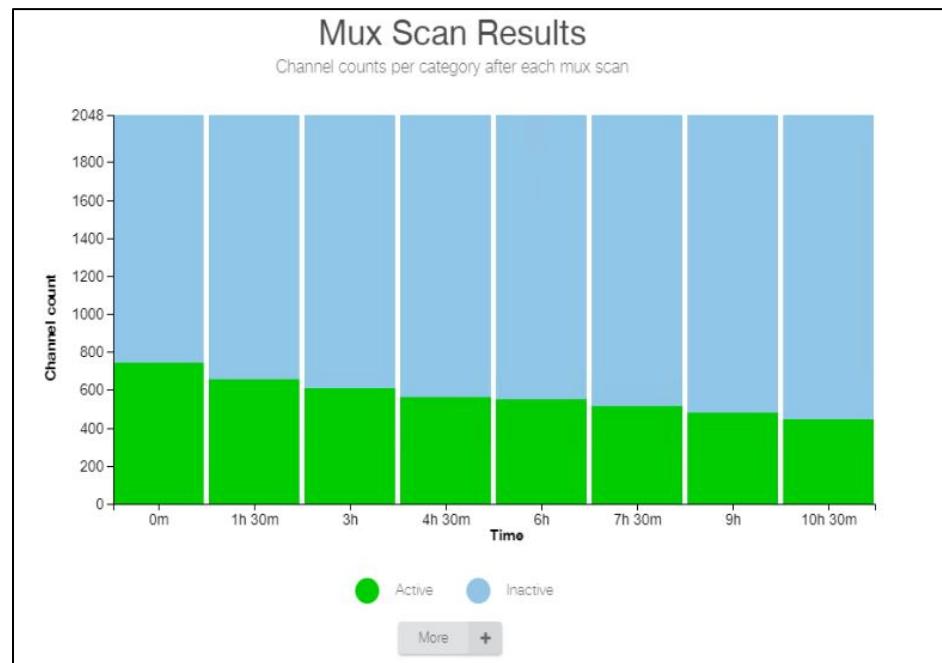
bubbles, osmotic imbalance, or the presence of detergents or surfactants in the library.

Duty time plot with Active Channel Selection switched on:



Duty time plot with Active Channel Selection switched off:

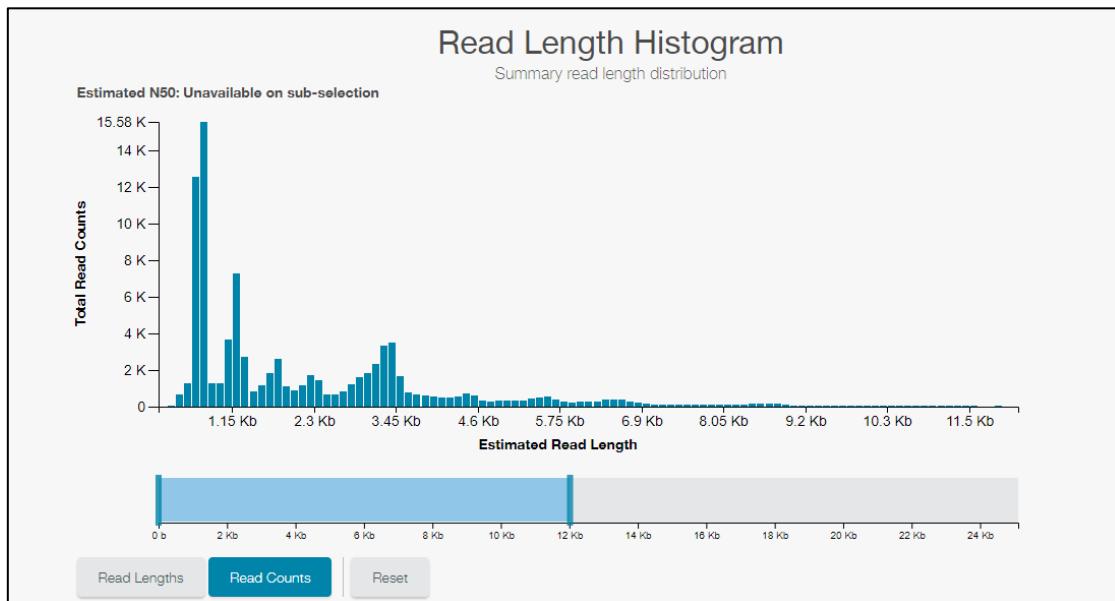


Mux scan results plot:

➤ Read length histogram panels:

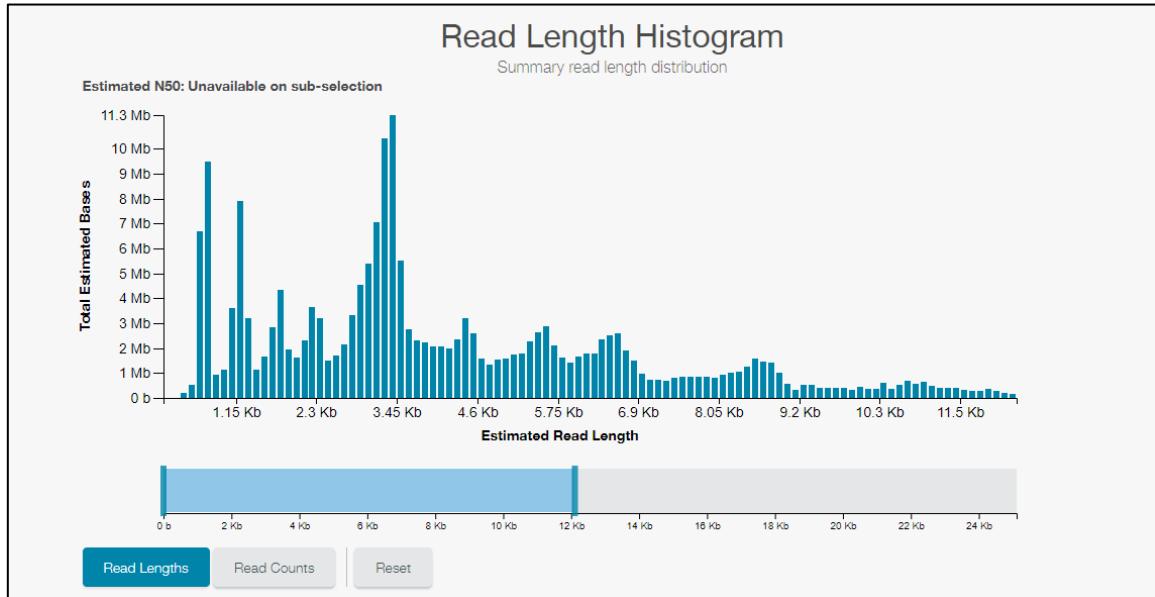
There are two cumulative histograms that can be toggled between in MinKNOW. The histograms reflect expected read lengths or cumulative number of bases for the experimental design being used. These are:

- **Read count plot histogram:** this shows the number of reads per bin and the distribution of these bins' aggregate values. This enables the user to understand how the read lengths vary in number and size; this helps the user visualise the N50 value. **(N50 value is defined as the cDNA molecule length where the sum of the lengths of all the cDNA molecules of size N50 or longer contain at least 50 percent of the total number of bases sequenced)**



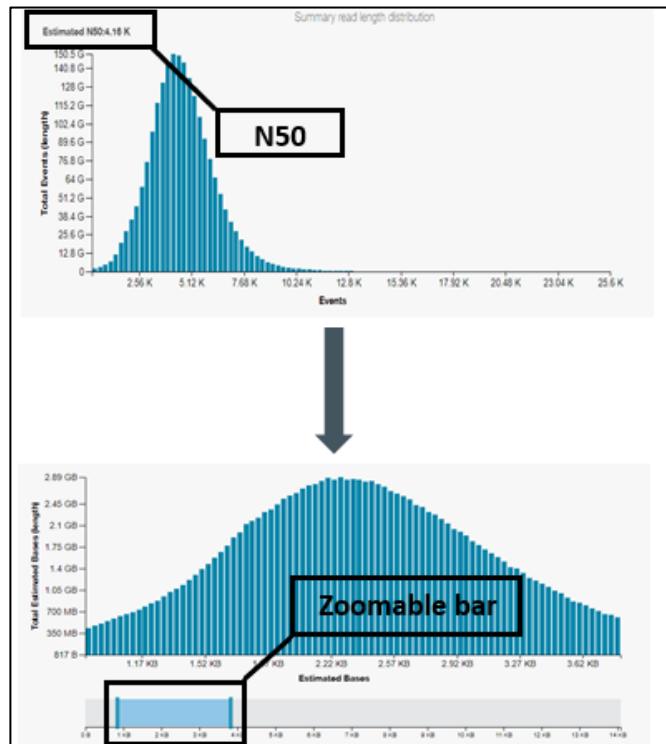
Read count plot histogram. The x-axis corresponds to the read length. The y-axis corresponds to the total number of estimated reads in that read length bin.

- **Read length histogram:** this shows the cumulative number of bases vs the read length. This enables the user to compare the skew from more numerous reads or longer reads towards the N50 i.e. if more bases are implicated in a particular bin (through numerosity or length) then this will pull the averages in this direction.



Read length plot histogram. The x-axis corresponds to the read length. The y-axis corresponds to the total number of estimated bases in that read length bin.

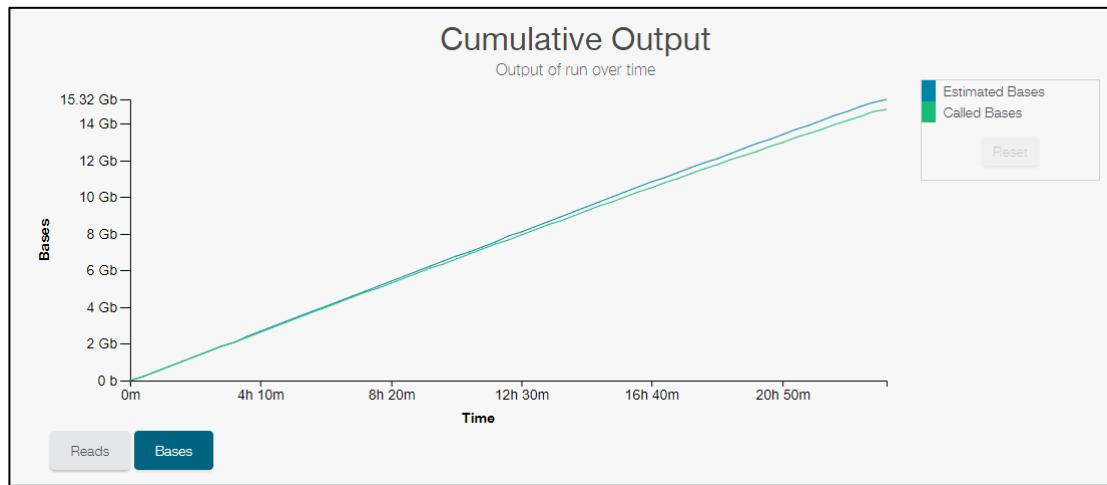
- The N50 value is presented (only for the whole set of passed reads) in the top left corner of the histogram. Each histograms X-axis (read length) can be zoomed into using the scaled bar under the histogram (shown with the blue box as labelled below). This facilitates the user to visualize distribution of 1. read counts or 2. number of bases with read length, more accurately over a size of interest. The reset button refocuses the zoom bar and histogram for the entire 'passed read' data set.



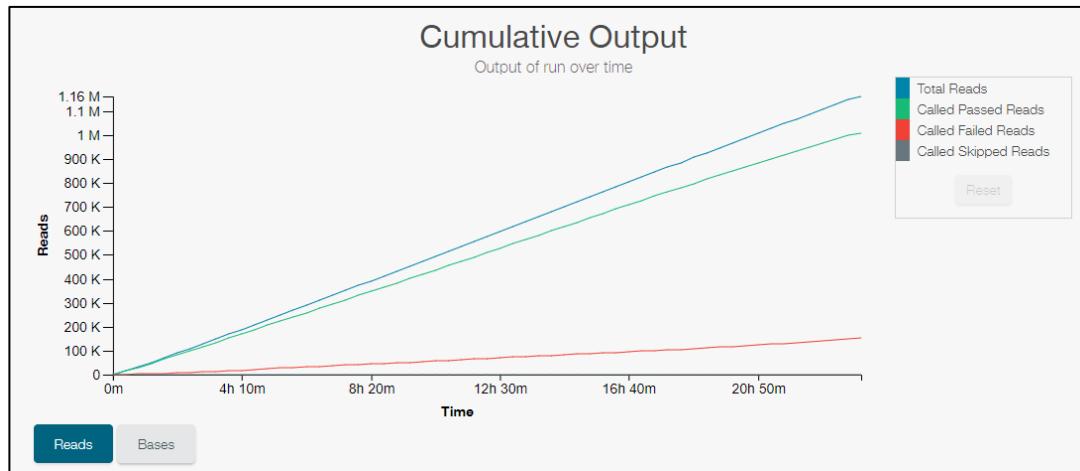
➤ Cumulative output panels:

The cumulative output graphs show:

- the number of bases that have been sequenced and basecalled



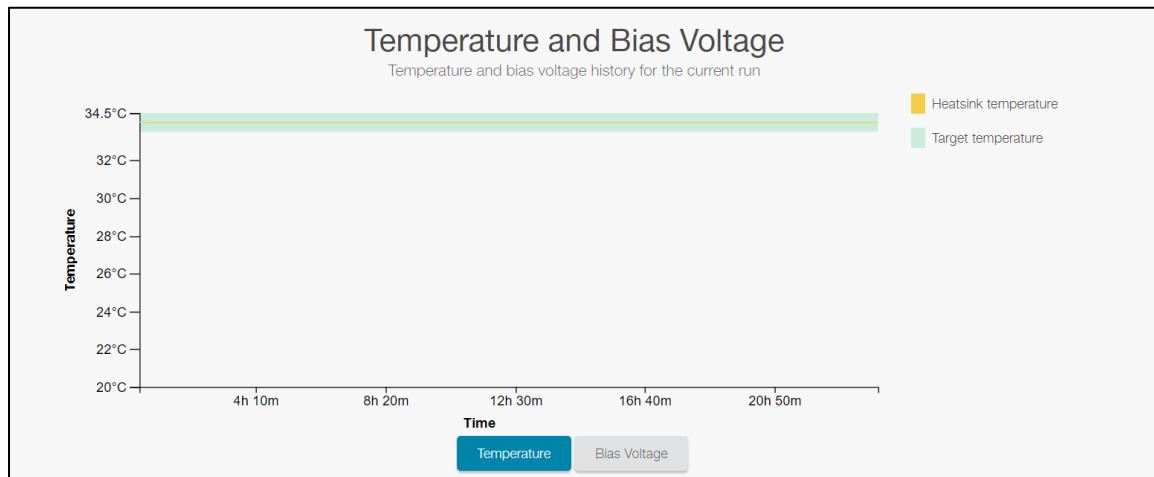
- the number of reads that have been sequenced and basecalled; and whether the reads have passed or failed the quality filters



➤ Additional panels:

The new graphs are displayed with 10-minute bin intervals over the course of the experiment and include temperature, bias voltage, translocation speed, qscore, cumulative and individual output, temperature and bias voltage graph.

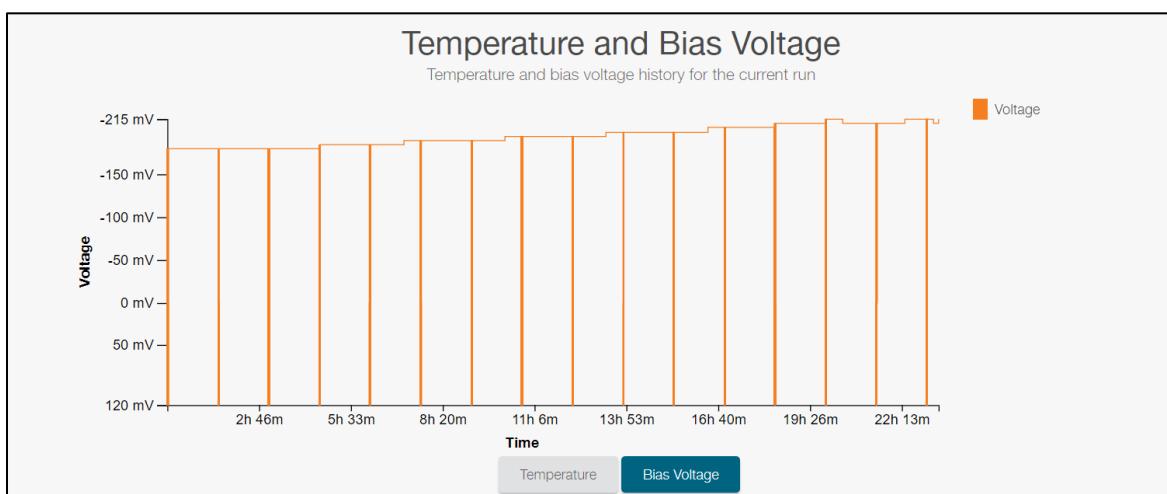
- Temperature vs time graph:** The temperature graph gives a real-time representation of the temperature below the flow cell. If the temperature reading drifts out of the target zone, please consult Technical Services, otherwise the quality of your data may be compromised.



- **Bias voltage v time graph:** The bias voltage graph provides the running voltage in real-time. MinKNOW will automatically adjust the applied voltage based on the strand range and will naturally drift to higher voltages as the electrochemistry in the well is depleted. This graph is useful for running a flow cell multiple times.

If you set the voltage for a subsequent run as the final running voltage of the previous run, then MinKNOW will find it easier to identify the appropriate running voltage.

You will notice drops in the voltage at regular intervals and these will correspond to the MUX scans that are defaulted to occur every one and a half hours. Here, each MUX - the options available to each channel - will be scanned to look for its availability for sequencing. The common voltage is reversed before and after each MUX assessment for clearer results.



- **Translocation speed and Qscore graphs** (Note: These graphs are only present if Basecalling is turned on.)
 - **Translocation speed vs time:** The translocation speed graph gives a new insight into the workings of the nanopore sequencing.

Important: The rate at which the DNA/RNA is passed through a nanopore by the motor protein is proportional to the concentration of the ATP in the flow cell and ATP turnover will be proportional to the concentration of motor protein present in solution and runtime. As a result, we therefore recommend not to overload a flow cell with library, otherwise more motor proteins will turn over ATP in the bulk solution, limiting the ATP available for the motors above the pore. This practically means that less ATP will be available in the solution which will result in a slower translocation speed and the sequencing quality of the reads will eventually drop.

The translocation graph gives a real time representation of the speed at which DNA/RNA strands are passed through the pore. If the translocation speed drops below this window, then quality may be compromised, as well as potential output, as strands take longer to move through the pore.

If the translocation speed drops below 300 - 350 bases per second and a lowering of the quality score is observed, then the ATP can be replenished through the addition of new priming mix buffer in the flow cell (Figure 37). There is no risk of losing cDNA molecules during the addition of new buffer as the cDNA molecules are immobilized on the lipid bilayer through the cholesterol tethers.



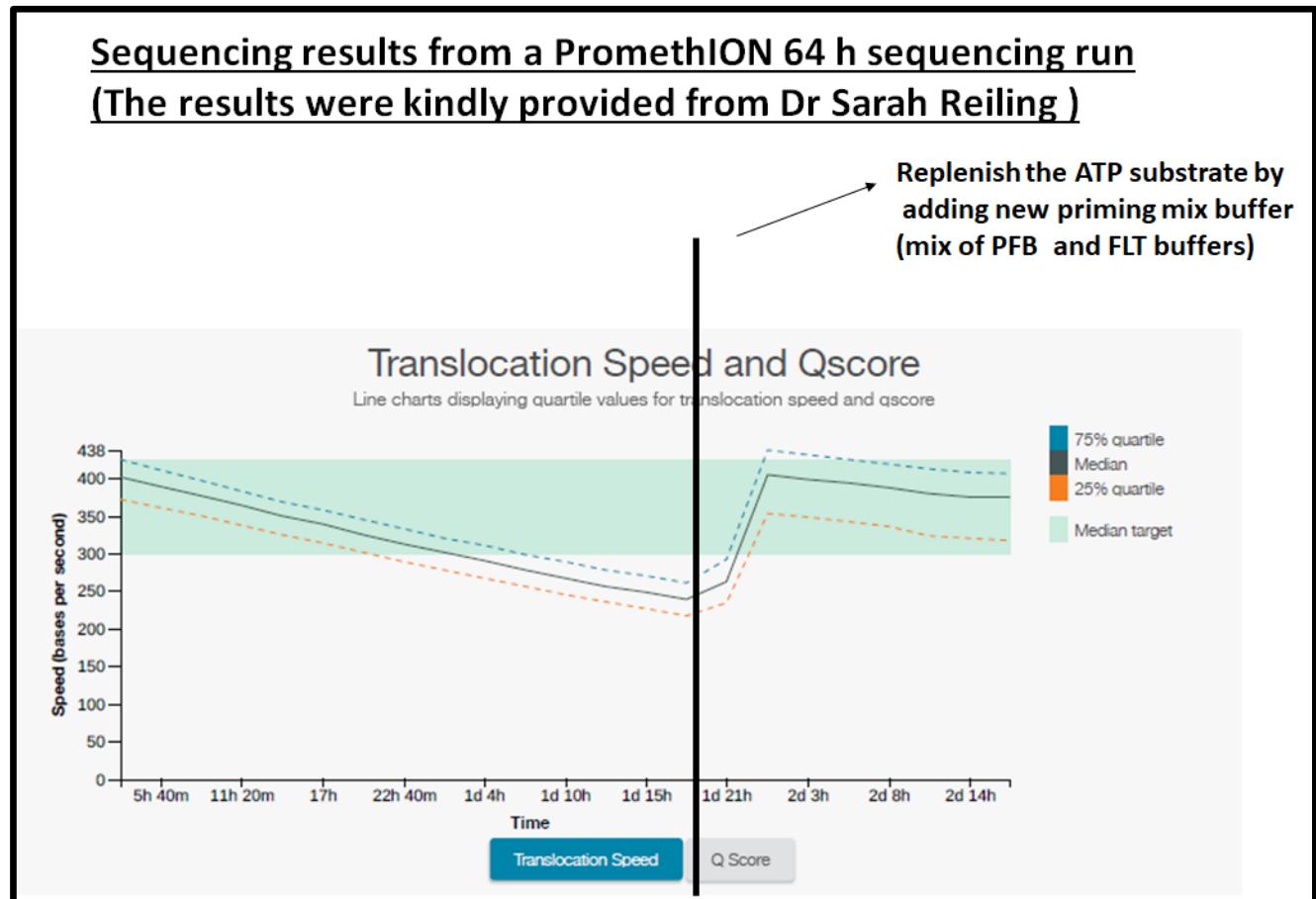
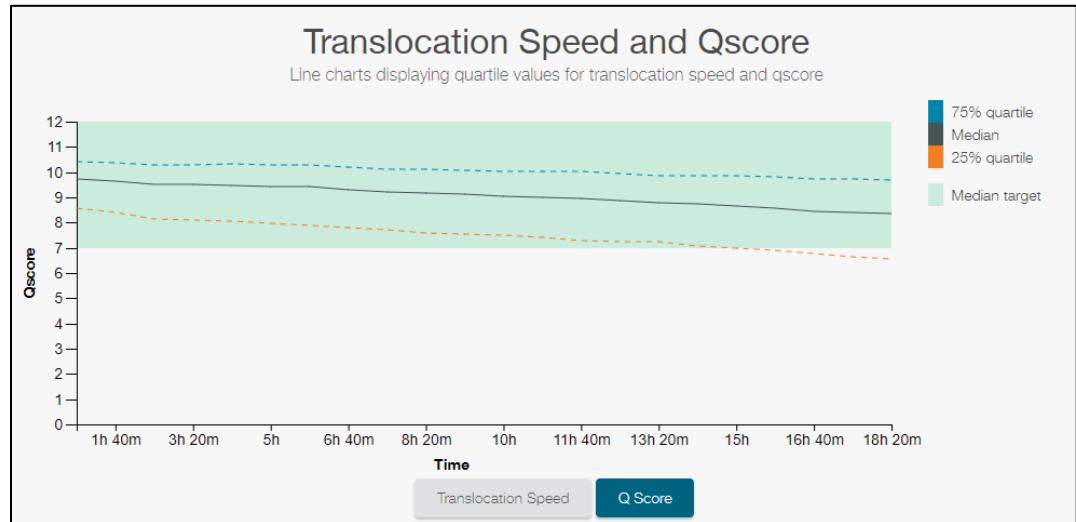
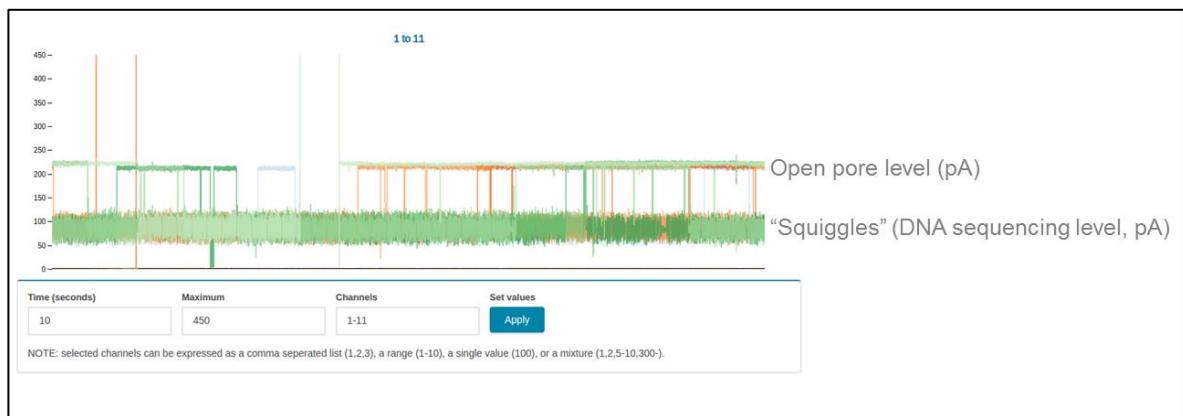


Figure 37. Effect of replenishing the ATP on the translocation speed during a 64 hours PromethION flow cell run. The ATP was replenished 1 day and 21 hours after the start of the sequencing run on the PromethION instrument. The PromethION flow cells have ~6,500 nanopores available for sequencing (active single pores) compared to ~1,500 nanopores available for sequencing on the MinION flow cell.

- **Qscore vs time:** The Qscore graph gives a live representation of the median strand Qscore over time, plus a window that gives the pass/fail folder divide.

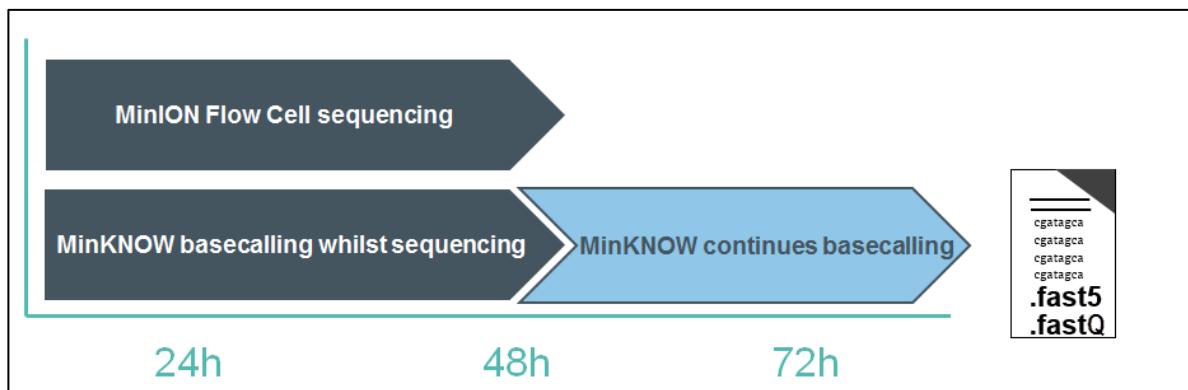


Trace viewer panel:



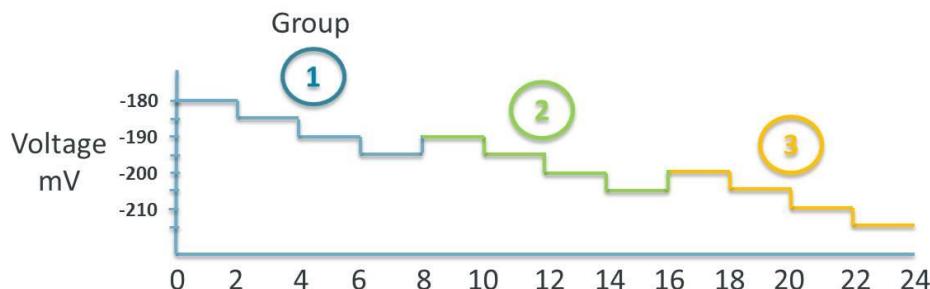
- The trace viewer displays the current levels from individual channels. By default it is set to show 10 channels, this number can be changed through the selection boxes beneath the viewer.
Additional parameters that can be altered:
 - Time: The length of time plotted on one screen
 - Maximum: The highest current level to be shown on the y axis

- The length of the sequencing scripts are indicated during selection, e.g. 48 hours. However, if live basecalling is selected, a minimum-spec laptop may not keep up with the speed of data acquisition, and not all reads will be basecalled by the end of the experiment. The reads that had not finished being basecalled during the experiment will continue to be basecalled in Catch-Up mode.



- **Voltage drift during a sequencing run**

Once the MinKNOW script proceeds to sequencing, the system goes through a cycle of voltage changes. The current scripts start with an applied voltage of -180 mV, which is optimal for basecalling with R9.4 chemistry. However in the process of a run, the voltage drifts due to the depletion of the redox chemistry in the bulk solution. In order to keep the voltage constant (*and thus the current at constant levels*), the applied voltage is decreased over the sequencing time to account for the drift from the individual well electrodes. The voltage can also be raised during a MUX group change. Although not all the individual wells of a new MUX group have depleted their redox chemistry, some chemistry would have been depleted from the common electrode, which is why the voltage does not fully return to -180 mV.



References

- 1 Ramskold, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* **30**, 777-782, doi:10.1038/nbt.2282 (2012).
- 2 Zajac, P., Islam, S., Hochgerner, H., Lonnerberg, P. & Linnarsson, S. Base preferences in non-templated nucleotide incorporation by MMLV-derived reverse transcriptases. *PLoS One* **8**, e85270, doi:10.1371/journal.pone.0085270 (2013).
- 3 Schmidt, W. M. & Mueller, M. W. CapSelect: a highly sensitive method for 5' CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs. *Nucleic Acids Res* **27**, e31 (1999).
- 4 Bayega, A. *et al.* Transcriptome landscape of the developing olive fruit fly embryo delineated by Oxford Nanopore long-read RNA-Seq. *bioRxiv*, 478172, doi:10.1101/478172 (2018).
- 5 Owens, N. D. L. *et al.* Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. *Cell Rep* **14**, 632-647, doi:10.1016/j.celrep.2015.12.050 (2016).
- 6 Sandler, J. E. & Stathopoulos, A. Quantitative Single-Embryo Profile of Drosophila Genome Activation and the Dorsal-Ventral Patterning Network. *Genetics* **202**, 1575-1584, doi:10.1534/genetics.116.186783 (2016).
- 7 Bayega, A. *et al.* Transcript Profiling Using Long-Read Sequencing Technologies. *Methods Mol Biol* **1783**, 121-147, doi:10.1007/978-1-4939-7834-2_6 (2018).
- 8 Bolisetty, M. T., Rajadinakaran, G. & Graveley, B. R. Determining exon connectivity in complex mRNAs by nanopore sequencing. *Genome Biol* **16**, 204, doi:10.1186/s13059-015-0777-z (2015).
- 9 Oikonomopoulos, S., Wang, Y. C., Djambazian, H., Badescu, D. & Ragoussis, J. Benchmarking of the Oxford Nanopore MinION sequencing for quantitative and qualitative assessment of cDNA populations. *Sci Rep* **6**, 31602, doi:10.1038/srep31602 (2016).
- 10 Albrecht, T. How to understand and interpret current flow in nanopore/electrode devices. *ACS Nano* **5**, 6714-6725, doi:10.1021/nn202253z (2011).