Polio Sequencing Consortium

Overview of Direct Molecular Detection and Nanopore Sequencing (DDNS) of Poliovirus

Lecture and discussion

Poliovirus Direct Detection and Nanopore Sequencing (DDNS)

Polio Sequencing Consortium

sample collection
(stool)

processing
(chloroform)

RNA extraction
(manual, automated)

nested PCR
(barcoded primers)

nanopore sequencing
(pooled VP1 products)

analysis & reporting
(PIRANHA)

PoSeCo | Poliovirus Sequencing Consortium

| Poliovirus Investigation Resource Automating Nanopore Haplotype Analysis

sample08 report 2022-09-12

Table 1 | Summary of sample content

Search:

Sample	Barcode	Reference Group
sample08	barcode33	Sabin1-related
sample08	barcode33	Sabin3-related

VP1 sequences

```
>sample08|barcode33|Sabin1-related|Poliovirus1-Sabin_AY184219|1|268:AT
GGGTTAGGTCAGATGCTTGAAAGCATGATTGACAACACAGTCCGTGAAACGGTGGGGGCGGCAACGTCTAGAGACGCTCTCCCAAACACTGAAGCCAGTGGACCAGCACACTCCAAGGAAATTCCGGCACTCACCGCAGTGGAAACTGGGGCCAC
AAATCCACTAGTCCCTTCTGATACAGTGCAAACCAGACATGTTGTACAACATAGGTCAAGGTCAGAGTCTAGCATAGAGTCTTTCTTCGCGCGGGGGTGCATGCGTGGCCATTTTAACCGTGGATAACTCAGCTTCCACCAAGAATAAGGATAAGC
TATTTACAGTGTGGAAGATCACTTATAAAGATACTGTCCAGTTACGGAGGAAATTGGAGTTCTTCACCTATTCTAGATTTGATATGGAATTTACCTTTGTGGTTACTGCAAATTTCACTGAGACTAACAATGGGCATGCCTTAAATCAAGTGTAC
CAAATTATGTACGTACCACCAGGCGCTCCAGTGCCCGAGAAATGGGACGACTACACATGGCAAACCTCATCAAATCCATCAATCTTTTACACCTACGGAACAGCTCCAGCCCGGATCTCGGTACCGTATGTTGGTATTTCGAACGCCTATTCACA
CTTTTACGACGGTTTTTCCAAAGTACCACTGAAGGACCAGTCGGCAGCACTAGGTGACTCCCTCTATGGTGCAGCATCTCTAAATGACTTCGGTATTTTGGCTGTTAGAGTAGTCAATGATCACAACCCGACCAAGGTCACCTCCAAAATCAGAG
```

Download

www.polionanopore.org

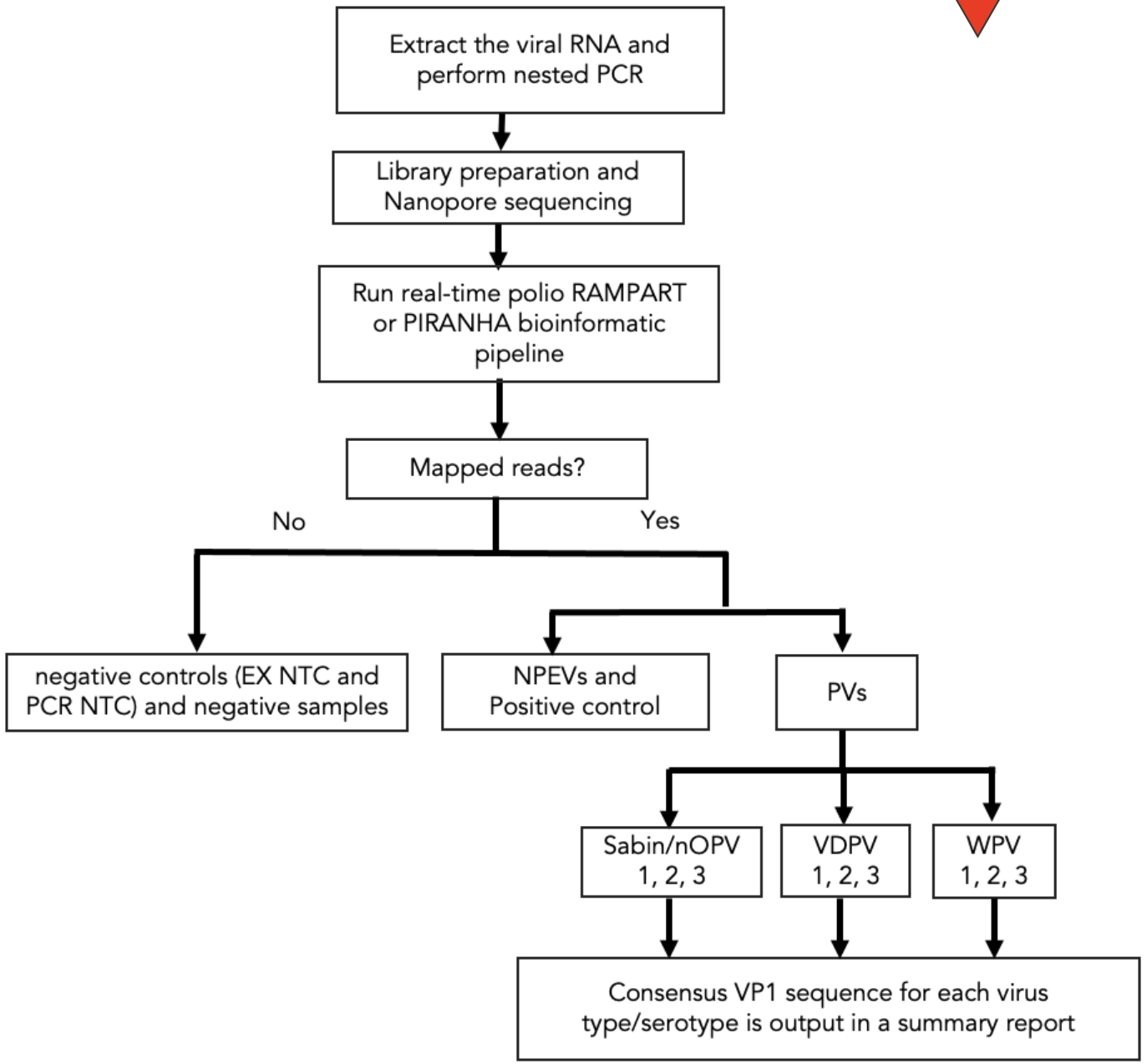
Polio Sequencing Consortium

Advantages of DDNS over cell-culture-ITD-Sanger sequencing

- Protocol detects and generates a VP1 sequence for poliovirus in stool samples 2-3 days after arrival in the laboratory
- Non-inferior sensitivity and specificity compared with cell-culture
- Accurate consensus sequences for every poliovirus detected
- Faster detection of outbreaks
- Cost-effective at ~$15 per sample assuming >90 samples/week
- Ongoing training in GPLN labs to implement method and generate further data on performance and costs -> accepted/recommended by GPLN SWG

Polio Sequencing Consortium

Diagnostic algorithm



Polio Sequencing Consortium

Example Piranha report

Piranha report 2022-12-08

a.

Sample	Barcode	Sample Classification	Reference group	Number of mutations
ENV001	barcode01	Sabin-like	Sabin3-related	1
ENV001	barcode01	Sabin-ilke	Sabin2-related	2
ENV002	barcode02	VDPV	Sabin2-related	12
ENV003	barcode03	WPV1	WPV1	NA
ENV004	barcode04	VDPV	Sabin2-related	10

b.

Sample	Barcode	Sabin1-related	Sabin2-related	Sabin3-related	WPV1	WPV2	WPV3	NonPolioEV	unmapped
ENV001	barcode01	0	488	252	0	0	0	0	25
ENV002	barcode02	0	1100	0	0	1	0	35	12
ENV003	barcode03	0	0	0	339	0	0	0	0
ENV004	barcode04	0	0	0	138	0	0	0	0
negative	barcode05	0	0	0	0	0	0	1	10

Showing 5 of 96 entries

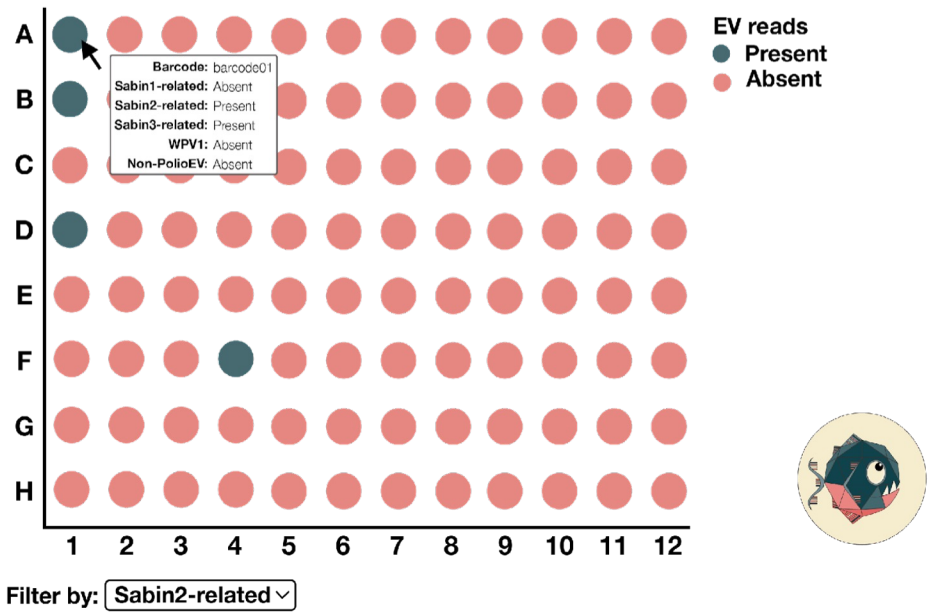
c.

Identical Sequences	Sequence IDs
1	ENV001|barcode01|Sabin2-related|2|161:CT;427:GA ENV030|barcode30|Sabin2-related|2|161:CT;427:GA
2	ENV022|barcode22|Sabin3-related|1|17:CT ENV024|barcode24|Sabin3-related|1|17:CT

d.

Pass	Sample	Barcode	Sabin1-related	Sabin2-related	Sabin3-related	WPV1	WPV2	WPV3	NonPolioEV	unmapped
✓	negative	barcode05	0	0	0	0	0	0	1	10
	positive	barcode05	0	0	0	0	0	0	45	0

e.



Polio Sequencing Consortium

http://polionanopore.org

- Protocols maintained on protocols.io – can be reached through

www.polionanopore.org

- Forums on the protocols.io site to raise any queries you have about the protocol
- New version of the protocols will be published here.

Polio Sequencing Consortium

QA/QC

- Careful QA/QC essential (clean PCR, data management)
- QC guidelines available and QA programme under development
- SOPs include recommended best practices for preventing and detecting cross-contamination, minimising sample handling
- Positive control (lyophilized CV-A20 standard prepared by NIBSC) and negative (water) control included in each sample batch from RNA extraction step
- Sequencing accuracy can be confirmed by testing reference panel and repeats/comparison with alternative sequencing platforms in a subset
- Long term monitoring of lab performance (including NPEV detection) and proficiency testing
- Accreditation of GPLN laboratories for DDNS will follow a similar process as for virus isolation, ITD, sequencing and environmental surveillance based on minimum requirements checklist

Polio Sequencing Consortium

How this training is organised

- Scheduled presentations and guided lab work
- An overview of the method
- Guided use of the protocol by trainee lab teams
- Independent performance of the protocol by the lab teams with support
- Practice the use of Oxford nanopore MinKNOW sequencing software
- Practice the use of the graphical user interface for Piranha (‘piranhaGUI’)

Polio Sequencing Consortium

Expected outcomes

- Be able to perform the DDNS protocol including RNA extraction, nested PCR, barcode file preparation, nanopore sequencing, report generation by PIRANHA
- Understand the QC/QA process
- Be ready to implement in your own laboratories with GPLN and PSC support
- Appreciate the other potential applications of poliovirus nanopore sequencing including generation of whole genomes

Polio Sequencing Consortium

Post training expectations

- Trainees should identify any additional equipment their lab would require to perform DDNS
- Trainees should identify the areas in their lab where the steps of the DDNS protocol could be performed
- Trainers should explain the DDNS method to their other lab members
- Trainees should liaise with the WHO for procurement of required equipment and reagents for DDNS implementation in their lab
- Trainees should provide a point of contact for the PSC to provide additional reagents (e.g. the positive control virus)
- Further training should be arranged to support implementation of DDNS in their own laboratory

Polio Sequencing Consortium

Acknowledgments

Study team

Imperial College London: Alex Shaw, Catherine Troman, Isobel Blake, Joyce Akello, Laura Cooper, Shannon Fitz, Nicholas Grassly

NIBSC: Manasi Majumdar, Erika Bujaki, Javier Martin

NIH Islamabad: Adnan Khurshid, Yasir Arshad, Masroor Alam

University of Edinburgh: Áine O'Toole, Rachel Colquhoun, Corey Ansley, Zoe Vallance, Andrew Rambaut

INRB DRC: Tresor Kabeya, Emmanuel Lokilo, Yogolelo Riziki, Amuri Aziza, Eddy Lusamaki, Jean-Claude Makangara, Marceline Alonga, Yvonne Lay, Bibiche Nsunda, Elisabeth Pukuta, Steve Ahuka, Jean-Jacques Muyembe, Placide Mbala

Biosurv International: Catherine Pratt

Collaboration and funding

BMGF: Ananda Bandyopadhyay, Kathleen Rankin and colleagues

WHO HQ: Ousmane Diop, GPLN SWG, Surveillance Group

WHO regional offices: Salmaan Sharif (EMRO), Anfumbom Kitu Womeyi Kfutwah (Jude) (AFRO), Julius Chia (AFRO), Eugene Saxentoff (EURO)

Polio Sequencing Consortium

2. Sequencing Run preparation, RNA extraction and RT-PCR, nested PCR, Library preparation and flow cell loading

Lecture and discussion

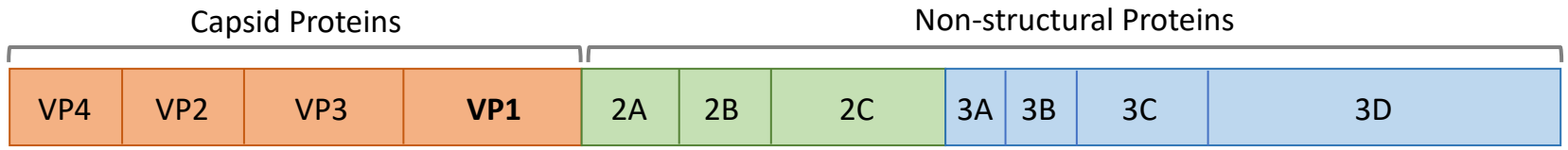
Polio Sequencing Consortium

Planning a run

- The protocol requires 2 days once RNA is extracted
- Can process up to 96 samples (including negatives) by barcoding samples
- Negative and positive run controls should be included in every run
- Samples are pooled to allow simultaneous sequencing
- Samples currently only tested once, but contamination likely to be identified through identical sequences for non-vaccine strains.

Polio Sequencing Consortium

Our PCR strategy



Pan-Enterovirus RT-PCR (primers for A+C enteroviruses, inc. nOPV2 : 5'NTR + Cre/nOPV-MM-R

- SuperScript III One-Step RT-PCR System

VP1 PCR Nest

- DreamTaq PCR Master Mix
- Barcoded primers

Barcode with flanking sequences

Library preparation overview

Polio Sequencing Consortium

Part 1: RNA extraction, RT-PCR and PCR

Part 2: Pooling, cleaning and end-prep

Part 3: Preparation for sequencing

Extracted RNA

panEV rt-PCR

VP1 PCR with barcoded primers

Pool samples

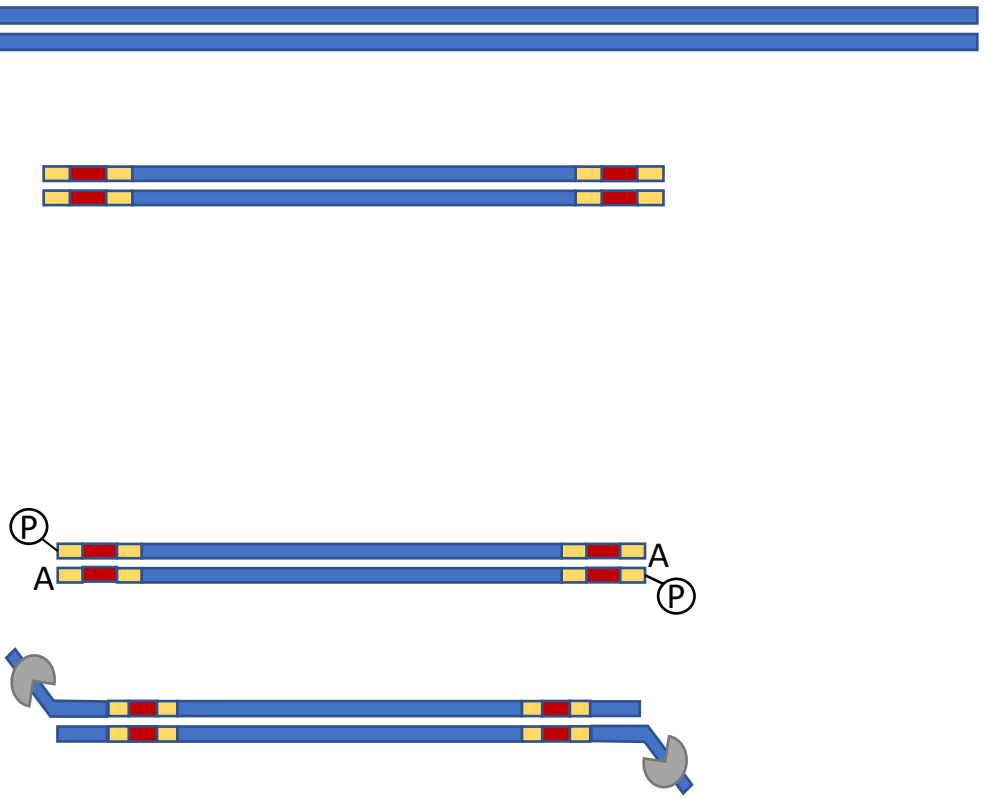
Ampure cleanup/concentration

End-prep & dA-tailing

Ligate sequencing adapter

Ampure cleanup & quantify

Load flow cell



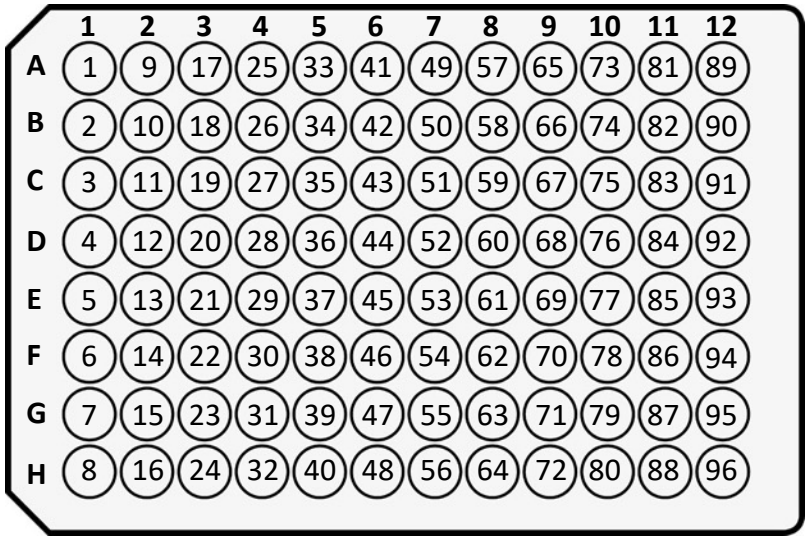
Polio Sequencing Consortium

Run preparation

- All sample metadata can be entered into a csv file and each sample assigned a barcode.
- The analysis software (PIRANHA) will append the sequencing results to each sample.

	A	B	C	D	E
1	barcode	sample	EPID	institute	...
2	barcode01	sample01	ARA-HIG-TOR-22-01		
3	barcode02	sample02	ARA-HIG-TOR-22-01		
4	barcode03	sample03	ARA-KAN-ERF-22-01		
5	barcode04	sample04	ARA-KAN-ERF-22-01		
6	barcode05	sample05	ARA-HIG-TOR-22-02		

The barcoded primers are stored in the format shown below, each number being a unique barcoded sequence attached to both Q8 and Y7



Polio Sequencing Consortium

Run controls

- Both negative and positive controls go through the whole process from RNA extraction to sequencing
- Negative run control: Nuclease free water used for RNA extraction
- Positive run control: CVA20 formulated by NIBSC, distributed freeze-dried and known to amplify in both PCR steps
- Instructions for use of the positive control are provided with the vials
- The controls demonstrate if the required sensitivity and specificity of the method is achieved
- They confirm negative or positive signals and can indicate if any issues arise during processing steps

Polio Sequencing Consortium

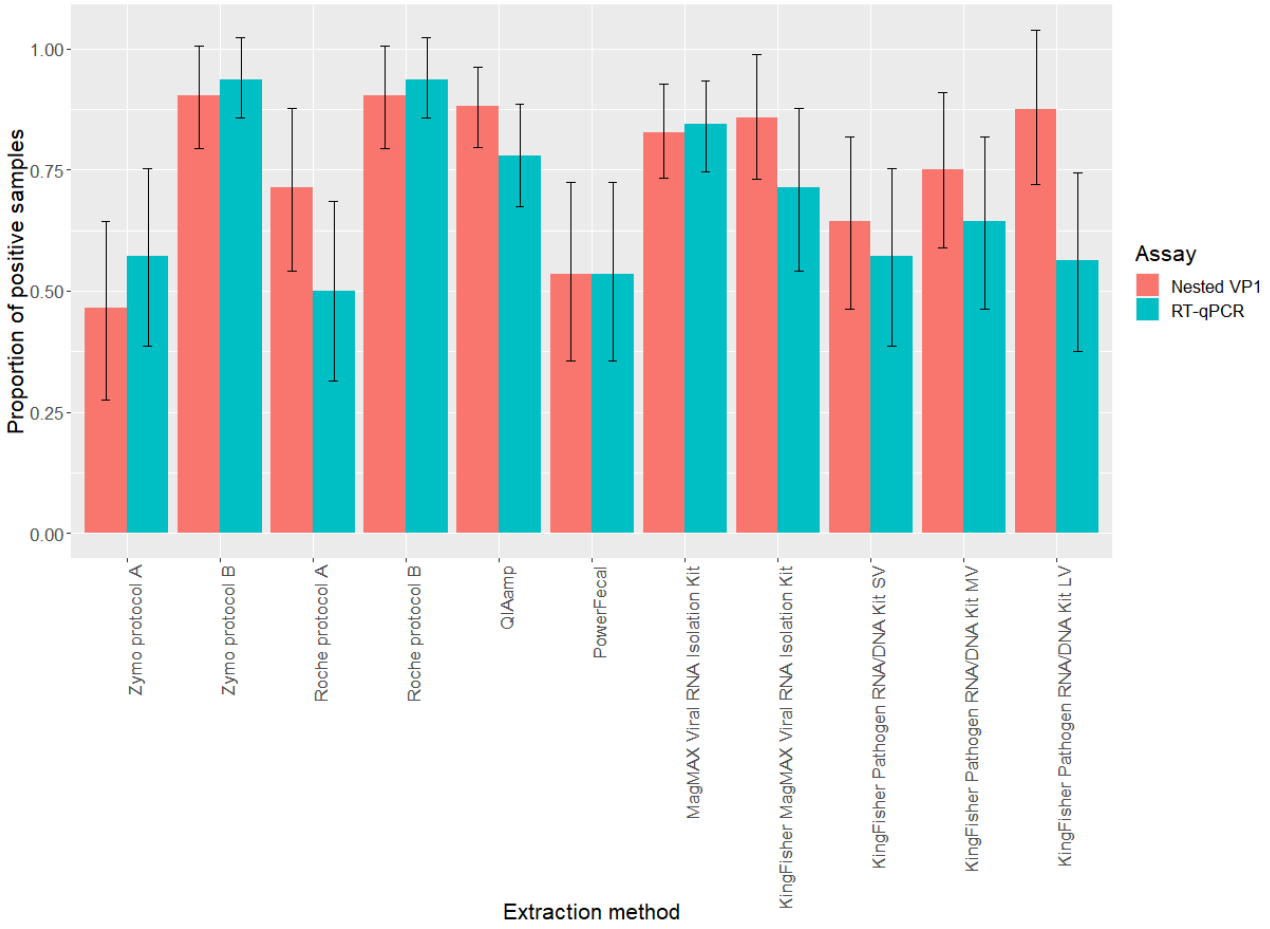
Performance of different RNA extraction kits

Approach

- 11 different RNA extraction methods have been compared
- Performance for nested VP1 PCR, pan-poliovirus qPCR (part of ITD) and near full-genome PCR assays examined
- We also considered processing time, flexibility, ease of use and cost

Conclusions

- There is significant variability between RNA extraction methods in their assay performance
- Optimal kits were:
  - manual: Roche HighPure, MagMAX Viral RNA
  - automated: Kingfisher MagMAX Viral RNA



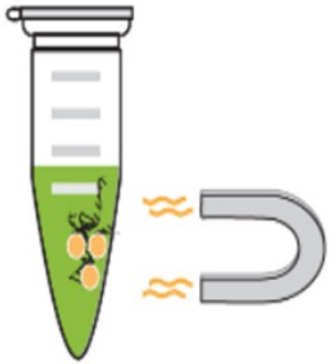
full details: Akello, Bujaki et al. 2023 *Microbiol.Spectr.* e0425222

Polio Sequencing Consortium

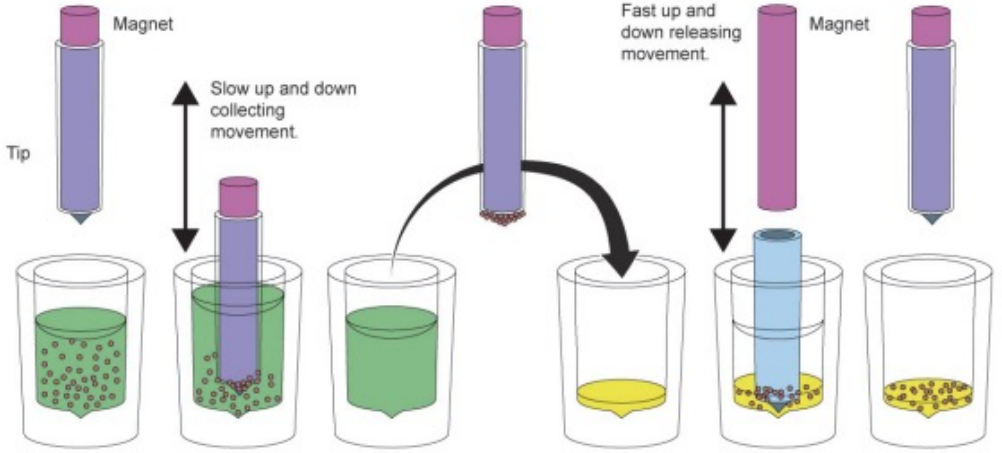
RNA extraction

- Recommended kit: MagMAX Viral RNA Isolation kit Cat no: AM1939

  Individual kit components are also available in large format.
- Magnetic bead-based method
- Can be performed manually in microcentrifuge tubes or in deep-well plates using KingFisher equipment

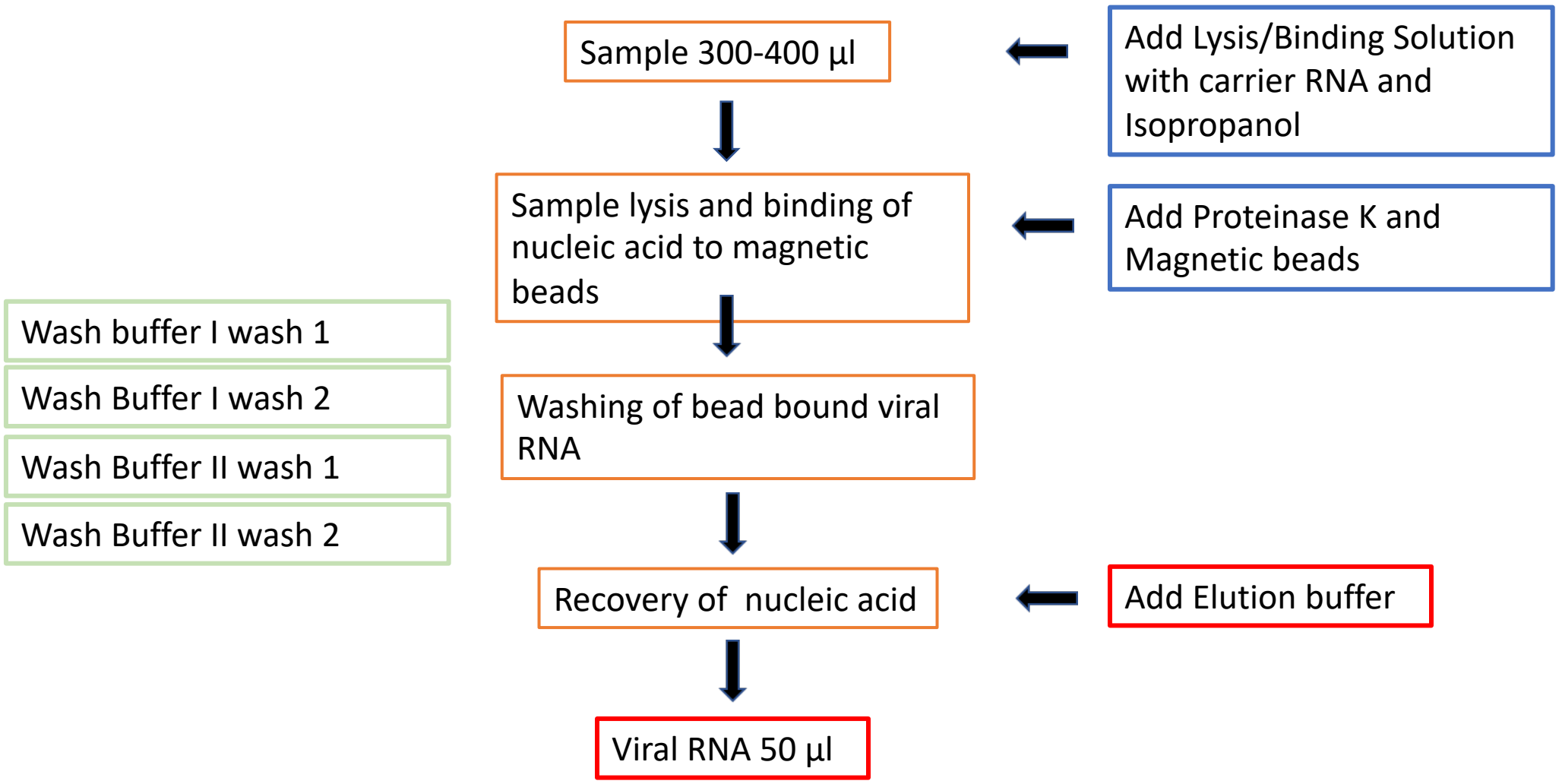


or



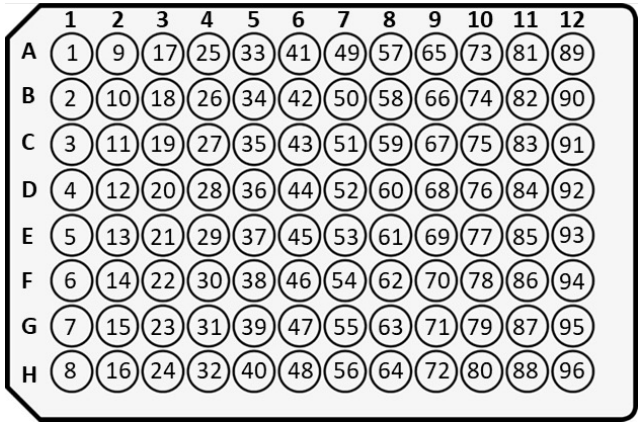
Polio Sequencing Consortium

RNA extraction - MagMAX Viral RNA Isolation Kit



Setting up the panEV RT-PCR

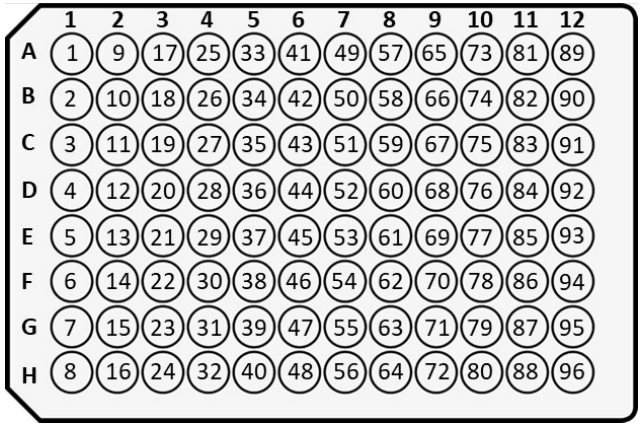
Mastermix plate
(19 µl reaction mix per well)



Add 5 µl

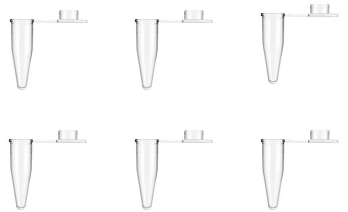
Extracted RNA

Automated



or

Manual

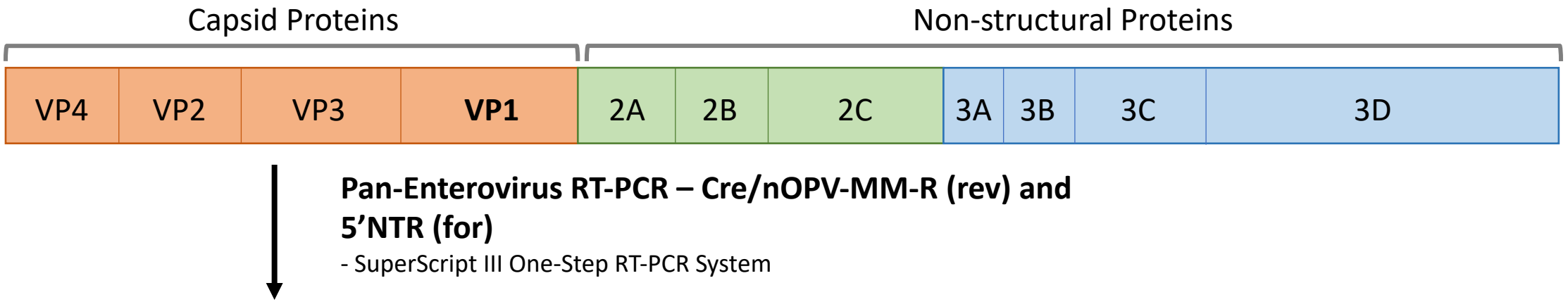


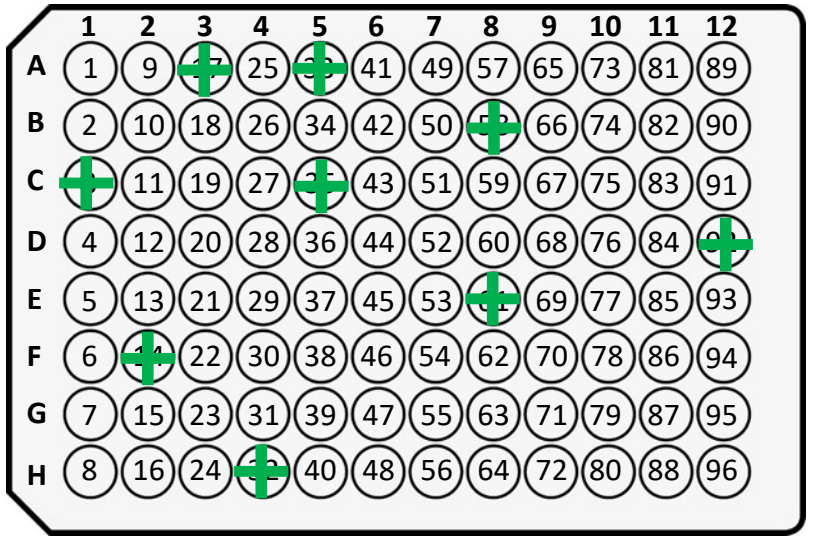
	1 Reaction (µL)
2x Master Mix	12.5
SSIII Platinum Taq mix	1
Reverse Primer (10 mM, Cre/nOPV2-MM-R mix)	1
Nuclease free Water	4.5
Total	19

- 30 minutes of incubation at 50 °C (reverse transcription)
- Add 1 µL of 10µM forward primer (5'NTR) for the PCR

Polio Sequencing Consortium

panEV PCR



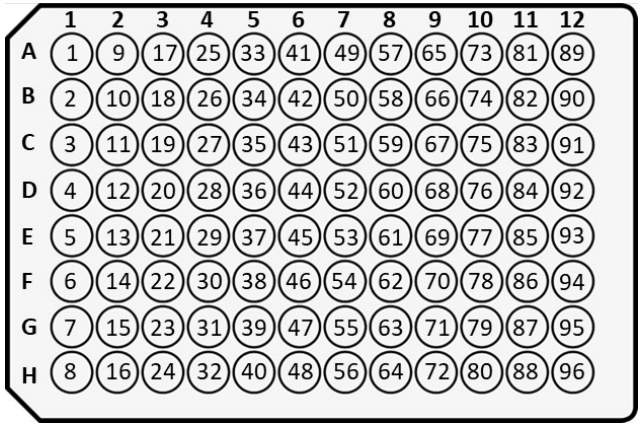


- Can run a gel to confirm, expect a band around 4kb.
- Only indicates the presence of an enterovirus.
- We do not run this gel routinely
- We do not exclude samples at this stage

Polio Sequencing Consortium

Nested VP1 PCR

- Performed with Dreamtaq
- Uses 2 µl of each of the panEV reaction (no cleaning required)
- Primers Q8 and Y7 for poliovirus specific amplification of VP1
- Primers have a barcode to allow multiplexing of samples; each sample is assigned a barcode, which is the same in each of the primers.
- Barcoded Q8 and Y7 primers can be organised in a 96 well plate for ease of addition.
- Duration: 2 hrs for PCR



Polio Sequencing Consortium

VP1 PCR primers

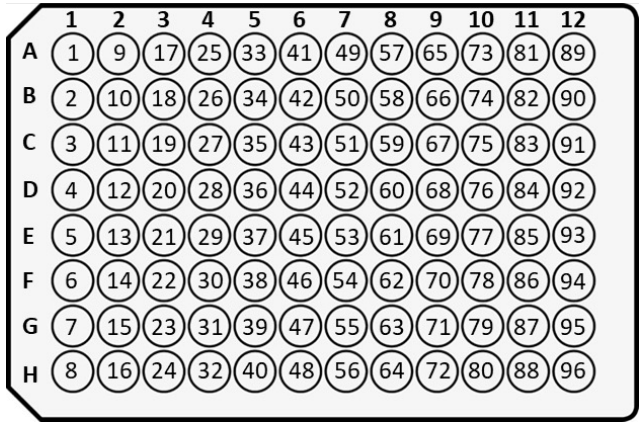
Flanking sequence

Y7 GGTGCTGACCGAGATCCTACGAATGGAGTGTTTAACCTGGGTTTGTGTCAGCCTGTAATGA

Barcode sequence **Primer sequence**

TACACCTTRTCTCTGGAGAATCCAATTACCGAGATCCTACGAATGGAGTGTGTCGTGG Q8

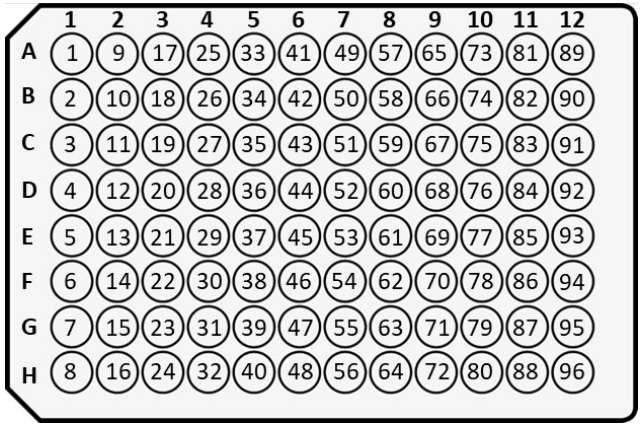
Primer sequence **Barcode sequence**



https://community.nanoporetech.com/technical_documents/chemistry-technical-document/v/chtd_500_v1_revt_07jul2016/barcoding-kits

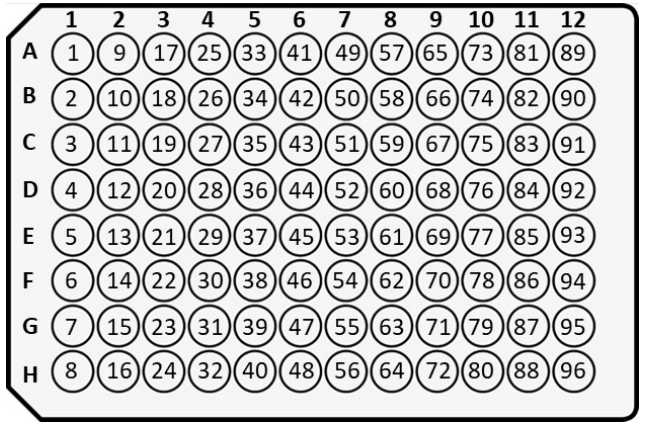
Setting up the nested VP1 PCR

Mastermix plate
(21 µl reaction mix per well)



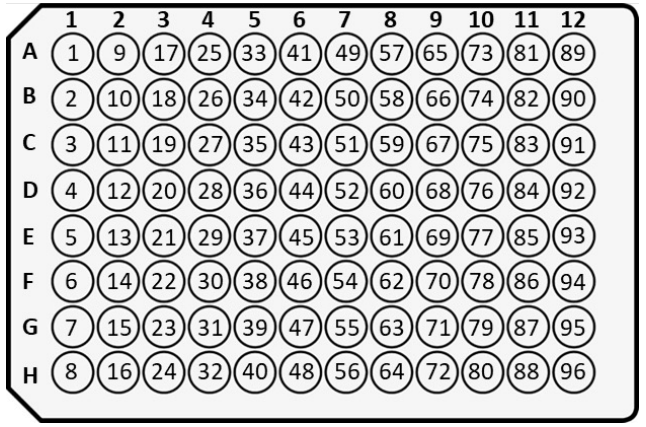
Add 2 µl

Q8/Y7 barcoded primer plate (10µM)



Add 2 µl

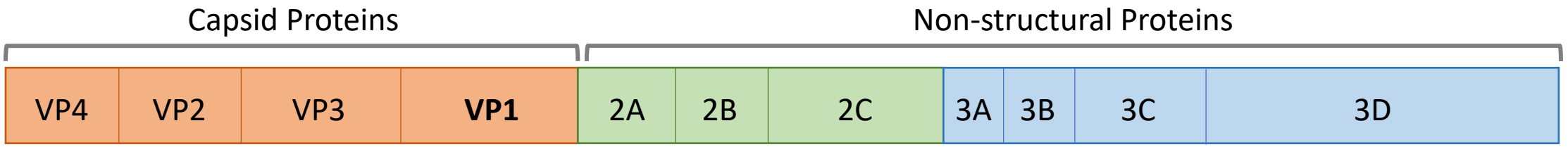
panEV reaction plate

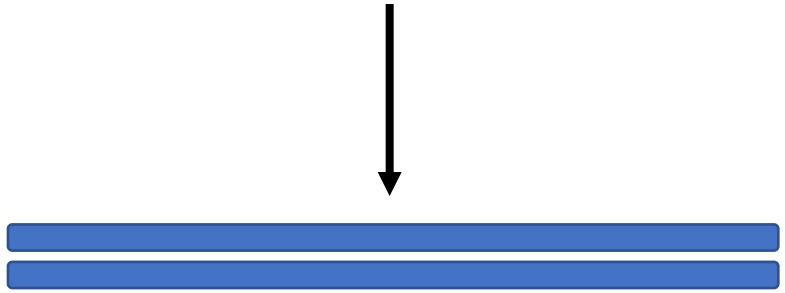


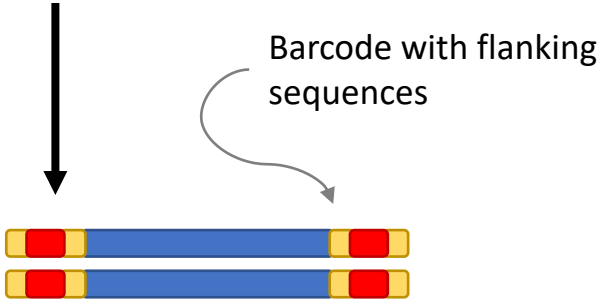
	1 Reaction (µL)
Water	8.5
DreamTaq 2x master mix	12.5
Total volume	**21**

Polio Sequencing Consortium

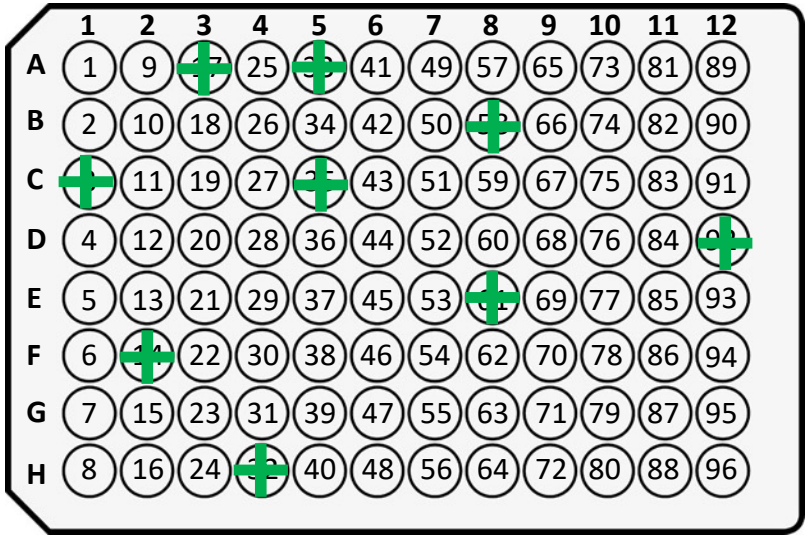
Post PCR







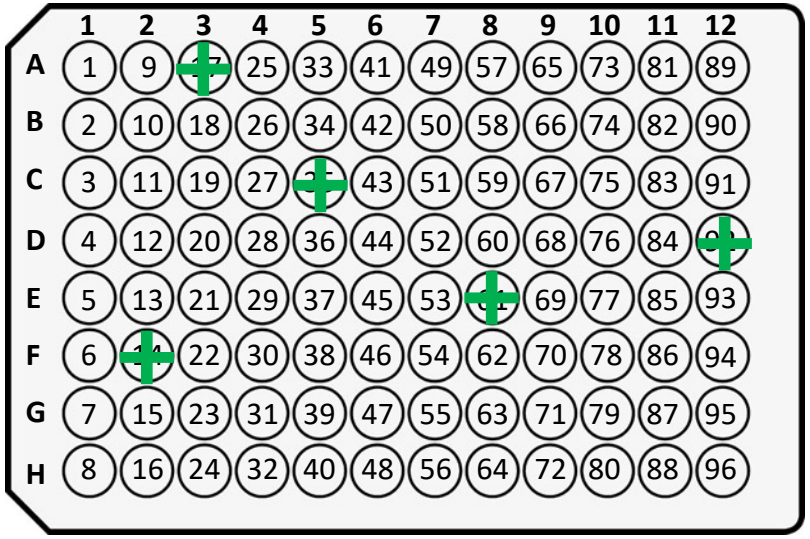
- Can run a gel to confirm, expect a band around 1.2kb.
- Only indicates the presence of poliovirus or enterovirus with similar primer binding sites
- We typically run a selection of samples to confirm PCR success

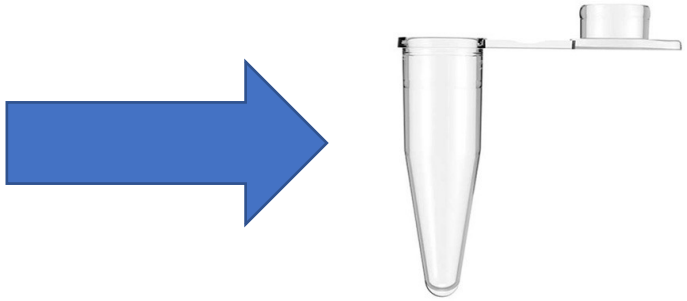


Polio Sequencing Consortium

Library Pooling

- Barcoded samples equally pooled by volume (2µl each)
- Previously pooled equal ng of product, we have found for routine stool testing that it is easier to pool by volume without losing sensitivity





Polio Sequencing Consortium

Why pool everything?

- Equal volume pooling removes quantification and dilution steps
  - Reduces cost
  - Reduces scope for error
  - Quicker
- Samples tend to have a fairly narrow range of concentrations after the nested PCR- generally 10-100 ng/µl
- Do not need great sequencing depth with stool samples- likely low complexity
- Negative samples do not detract from the success of sequencing

Polio Sequencing Consortium

Library cleaning and concentration

- Clean and concentrate the pool using 1:1 ratio of AmpureXP beads.

- In detail:
  - Add ampure beads to the pool which will bind the DNA and allow it to be separated by a magnet
  - Wash the pool with 80 % ethanol twice
  - Allow to briefly dry (bead pellet will become matt brown)
  - Resuspend pellet in 51 µl of water
  - Use magnet to pull the beads out of solution
  - Retain 50 µl of the eluate (the concentrated pool).

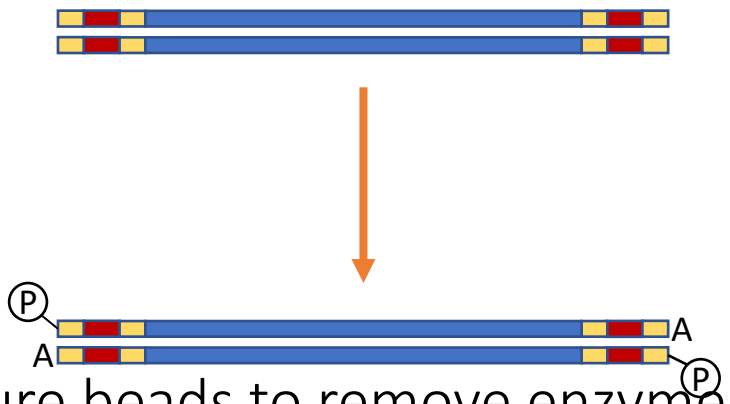
Ampure beads, Beckman Coulter

Pause point – store library at 4°C

Polio Sequencing Consortium

End preparation

- Treatment of DNA with Ultra II End-prep
- Ensures DNA has 3’ dA tailed and 5’ phosphorylated ends making them sticky for ligation of the sequencing adapter

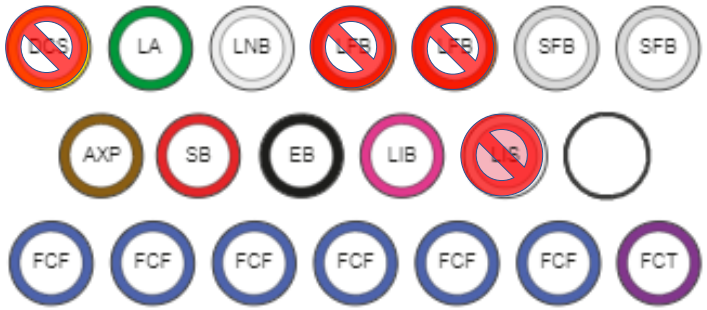


- Clean sample with ampure beads to remove enzyme

Polio Sequencing Consortium

Library preparation

- Performed using the Ligation Sequencing Kit from Oxford Nanopore
- Current version is LSK-114
- Contents:



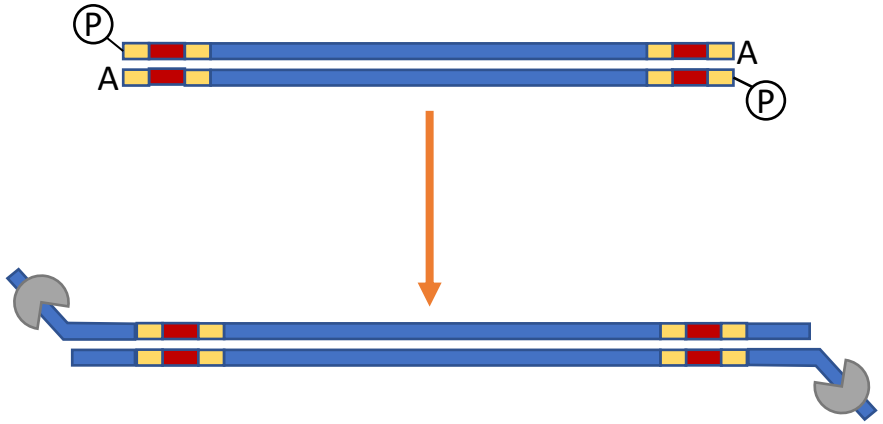
DCS : DNA Control Strand
LA : Ligation Adapter
LNB : Ligation Buffer
LFB : Long Fragment Buffer
SFB : Short Fragment Buffer
AXP : AMPure XP Beads
SB : Sequencing Buffer
EB : Elution Buffer
LIB : Library Beads
LIS : Library Solution
FCF : Flow Cell Flush
FCT : Flow Cell Tether

There vials in the kit that we will not use in this protocol (these are crossed out to the right)

Polio Sequencing Consortium

Adaptation for Sequencing

- Ligate the Ligation Adapter (LA) which includes the motor protein



- The motor protein controls the speed at which the DNA passes through the pore
- Ligation performed with Quick T4 Ligase (supplied by NEB) and the LNB ligation buffer

Polio Sequencing Consortium

Cleaning of Library

- Clean using Ampure beads

BUT

washing is performed with Short Fragment Buffer (SFB) and elution in elution buffer (EB)

- SFB used for washing where no size selection is required.
- Long fragment buffer (LFB) can be used to select for DNA >3kb

Polio Sequencing Consortium

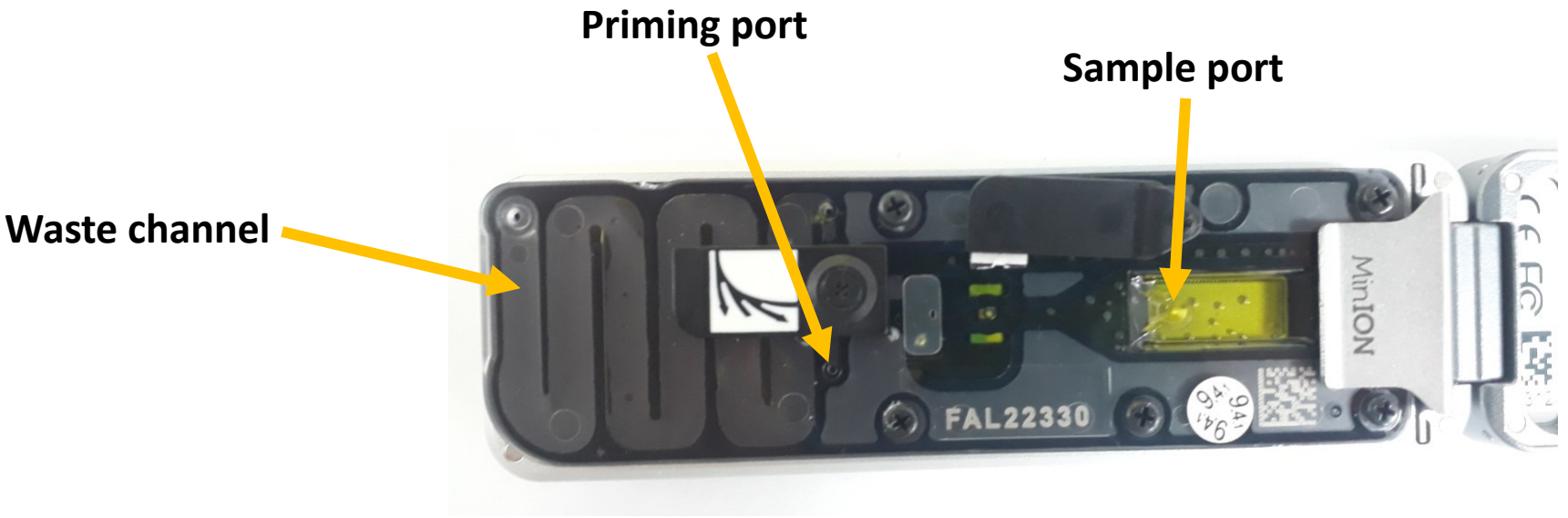
Priming of the flow cell

- Flowcell priming – Flow Cell Flush (FCF), Flow Cell Tether (FCT), and BSA

- The function of the tethers is to help pull DNA down towards the pores for sequencing
- The BSA (not included in the kit) is said to improve stability and throughput
- Added via the priming port in two steps
  - First time with sample port closed
  - Second time with sample port open

Polio Sequencing Consortium

Loading the MinION



In this picture both the priming and sample loading ports are open.

Polio Sequencing Consortium

Final Library

- The mix loaded onto the MinION contains:
  - Your adapted DNA library
  - Sequencing buffer (SB)
  - Library beads (LB)

- Loaded via the spot-on port, one drop at a time.

Polio Sequencing Consortium

Loading the MinION

- Before flowcell priming/loading any small air bubbles are removed from beneath the priming port ( )
- Waste reagents are removed from the waste port ( ) with both priming port and sample port covers **closed**
- ***Video guide: See "Loading the MinION"***

World Health Organization
REGIONAL OFFICE FOR THE Eastern Mediterranean

Polio Sequencing Consortium

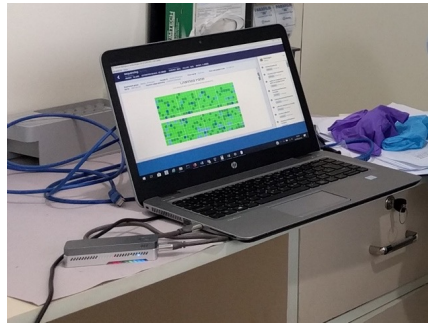
3. Nanopore sequencing

Lecture and discussion

Polio Sequencing Consortium

Library Preparation and Nanopore Sequencing

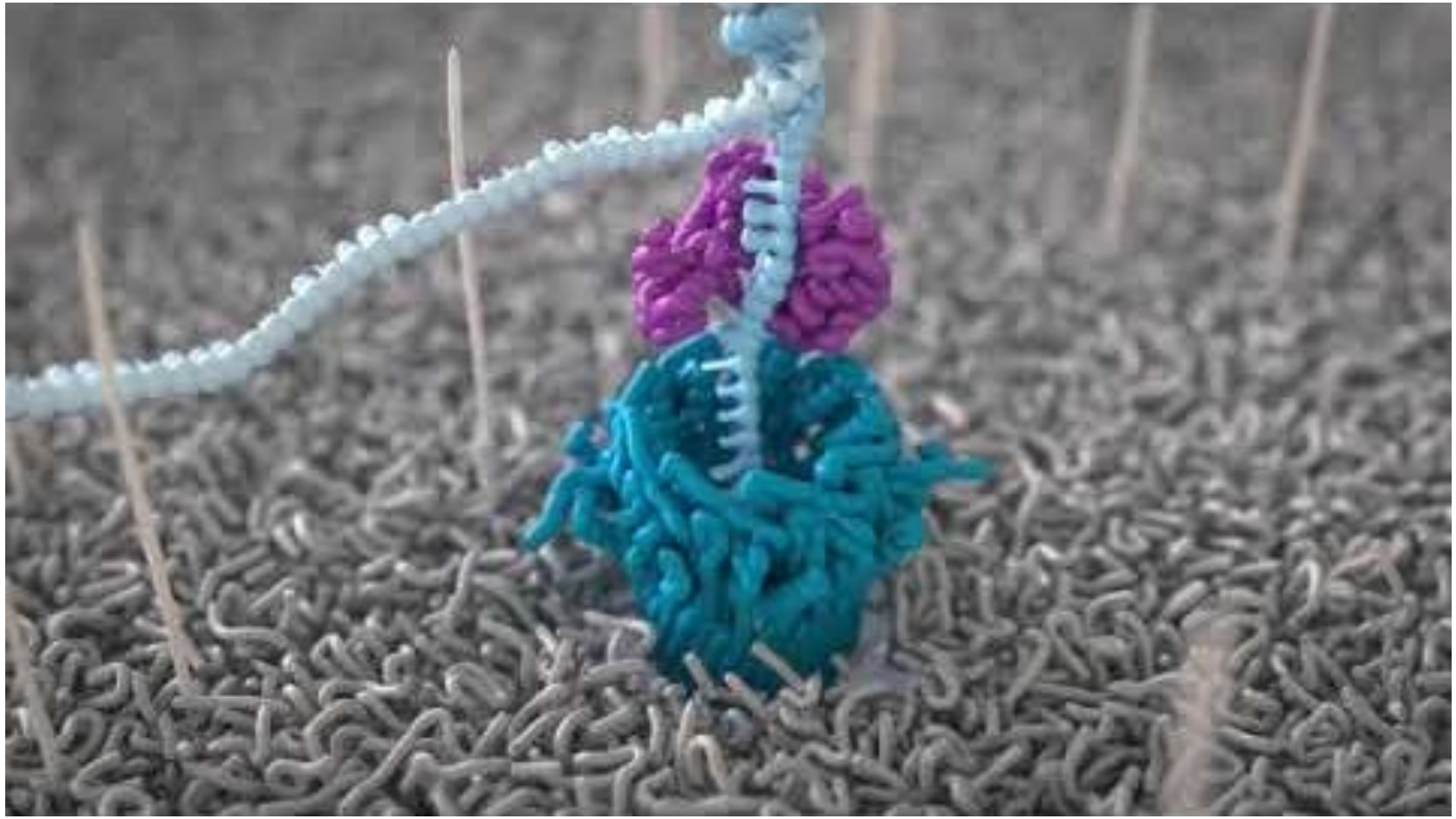
Extracted RNA → RT-PCR with panEV primers → Nested PCR with barcoded VP1 primers → Sequencing library preparation



PIRANHA (command line) or piranhaGUI (point and click)

Polio Sequencing Consortium

Nanopore Sequencing



https://youtu.be/RcP85JHLmnI

Polio Sequencing Consortium

MinION Sequencing

- Long Reads- potentially 100,000s of bp
- Fast- 1,000,000 reads per hour and can process in real time
- “Cheap” (depending on multiplexing)

but

- Error rate per base (~0.5% with current highest accuracy settings) - can still give accurate consensus sequences.

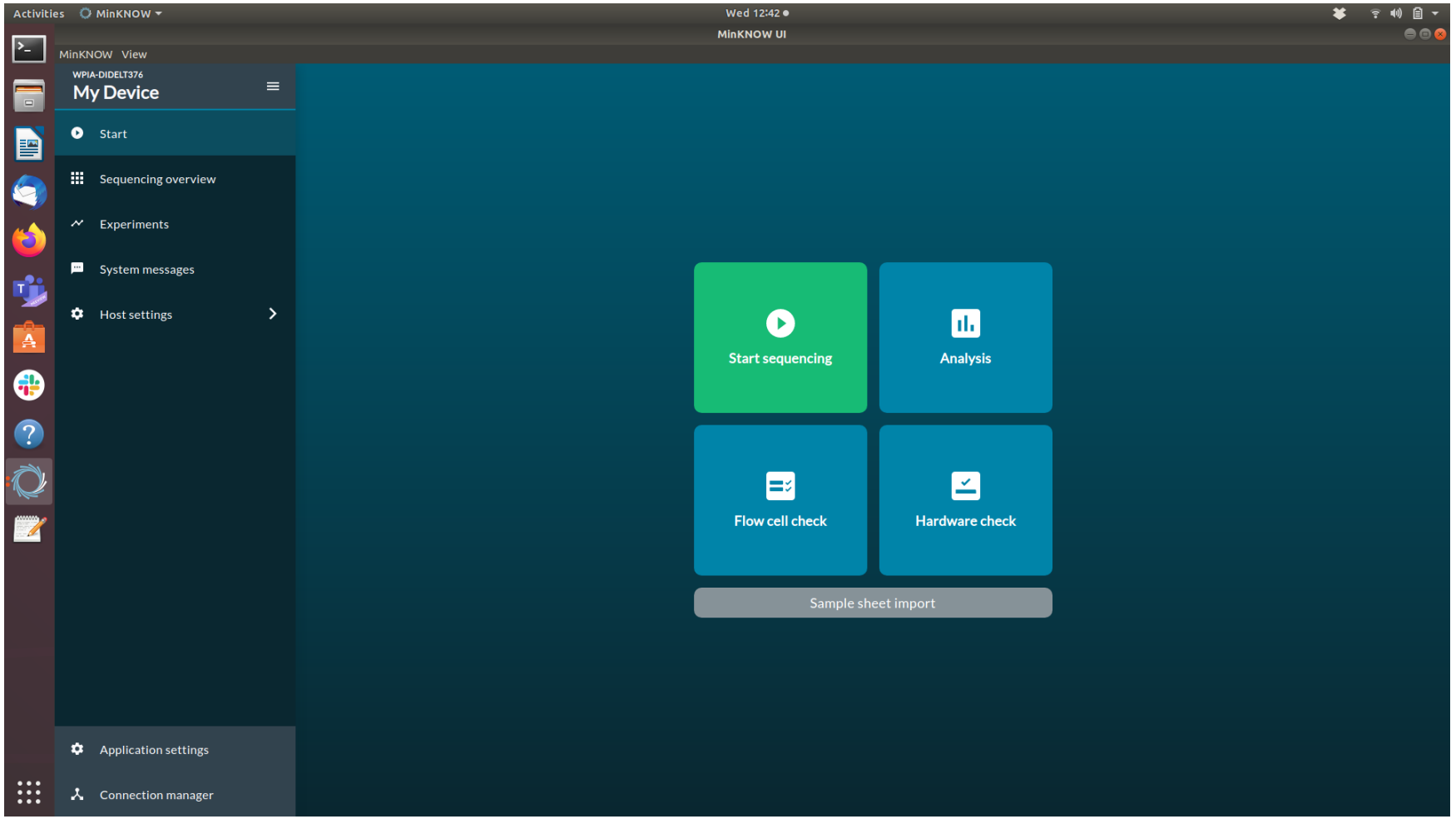
Polio Sequencing Consortium

MinKNOW

- Easy to use for setting up and running your sequencing run
- Install following the instructions on the Nanopore website
- Basecalling and demultiplexing performed within MinKNOW

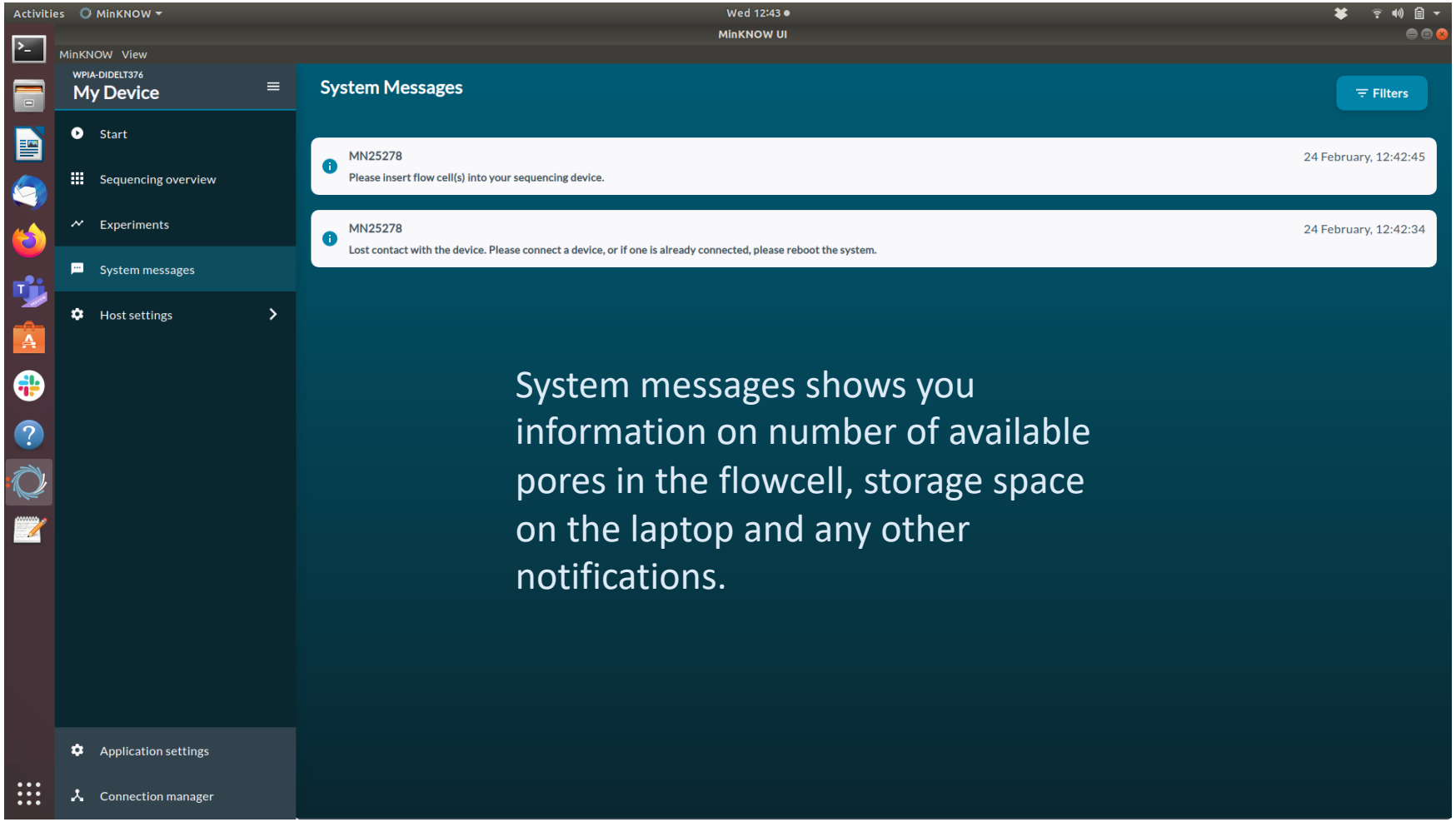
Polio Sequencing Consortium

MinKNOW GUI



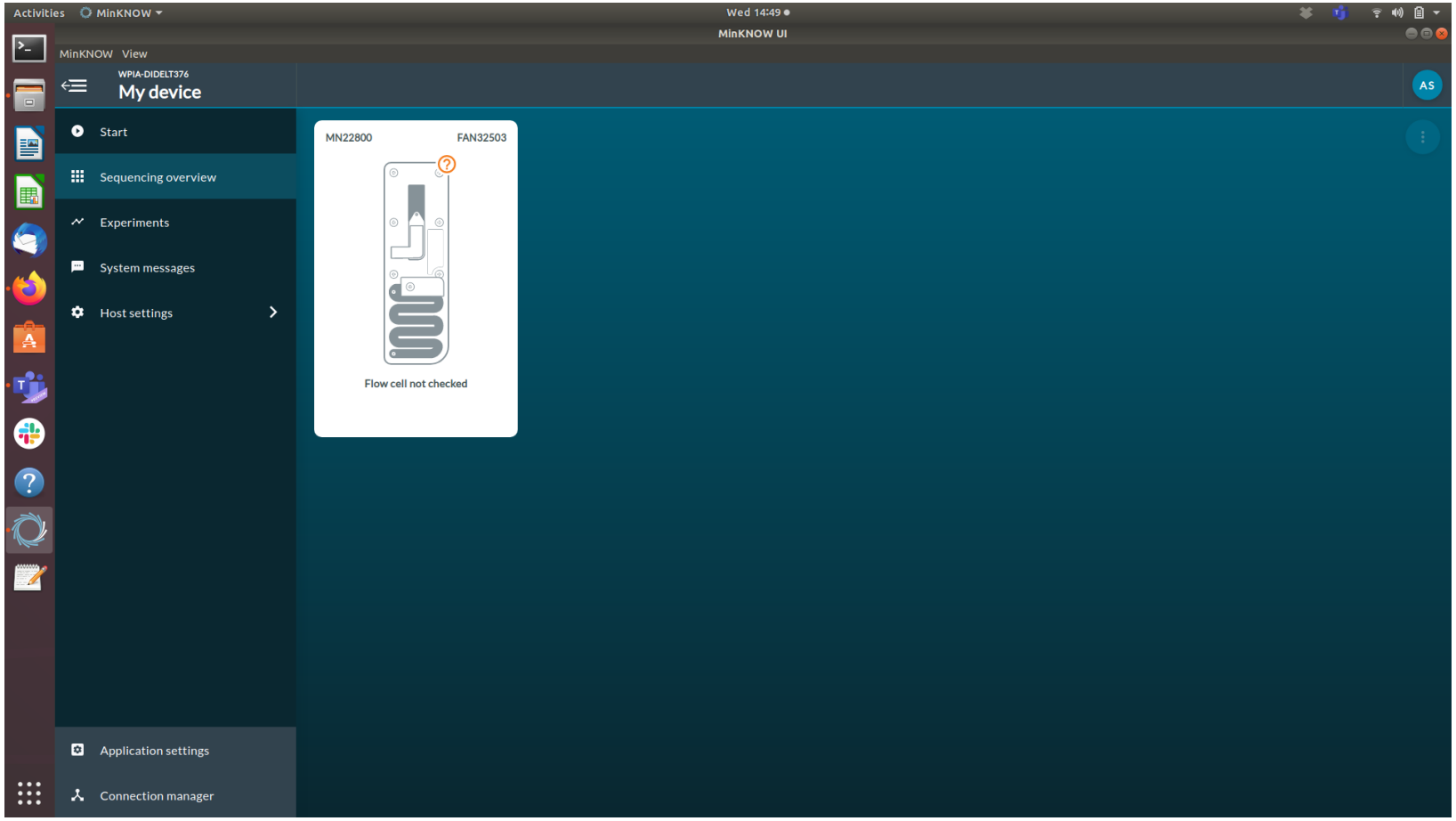
Polio Sequencing Consortium

System messages



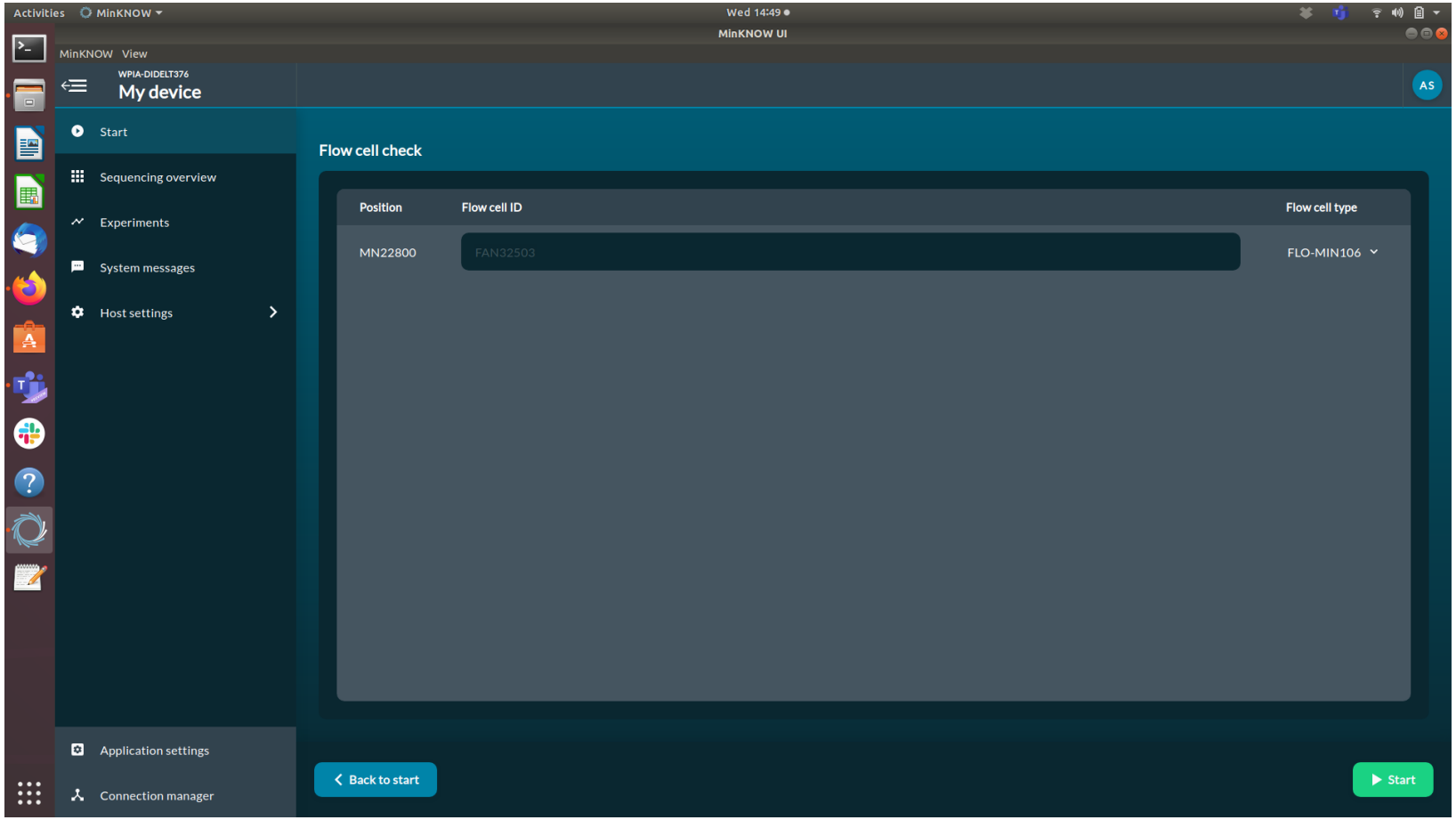
Polio Sequencing Consortium

Sequencing overview



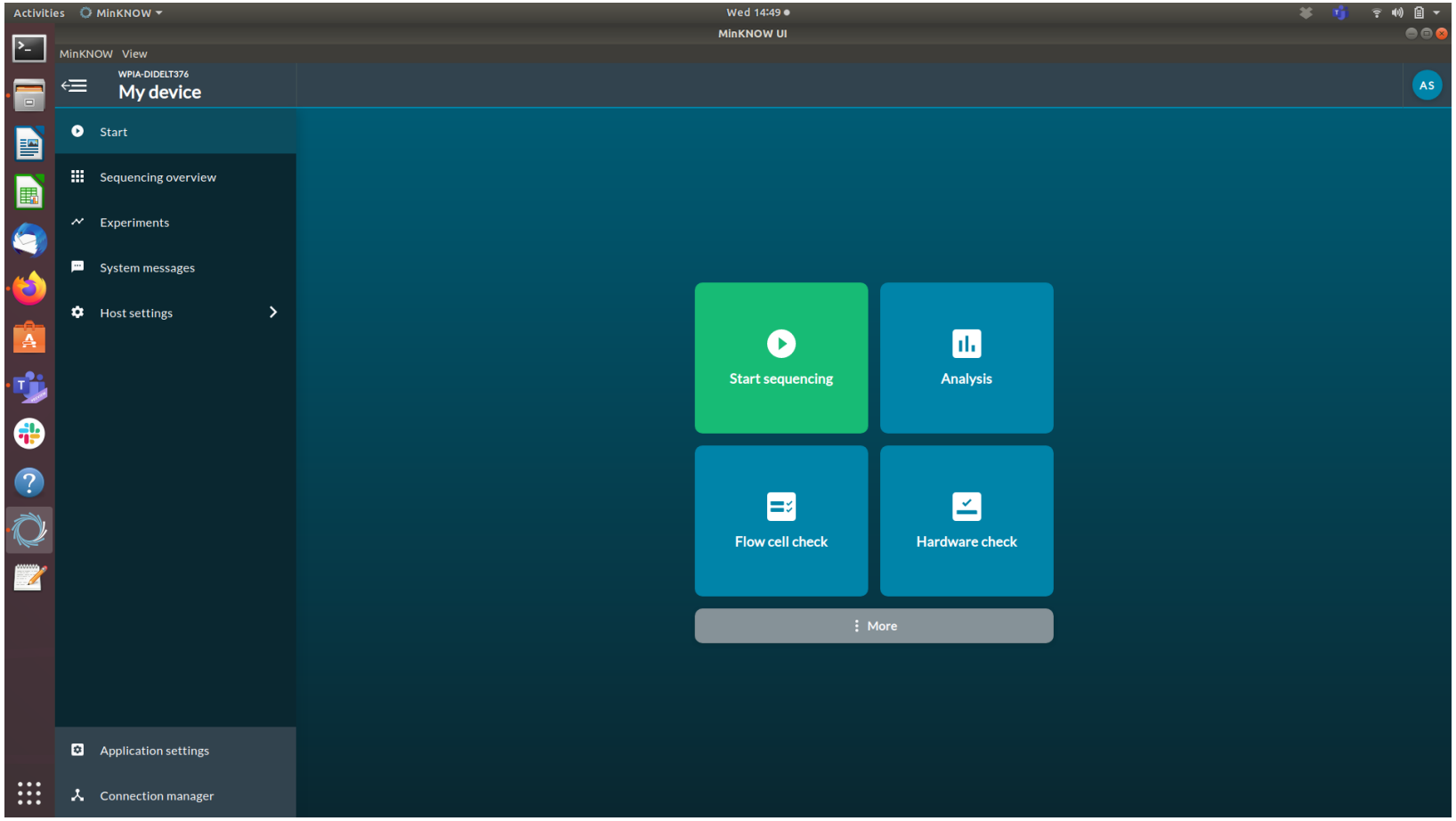
Polio Sequencing Consortium

Running a flow cell check



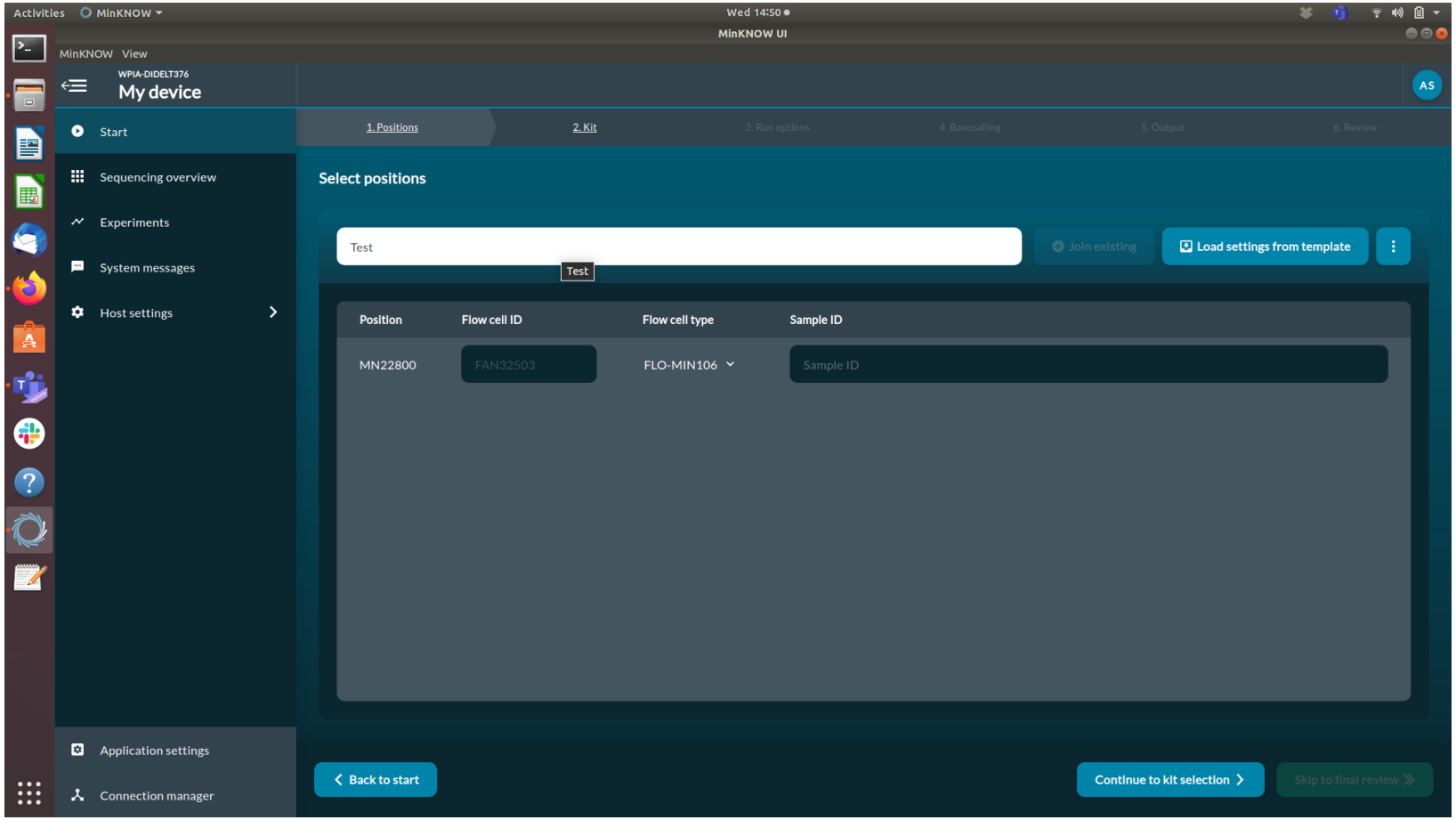
Polio Sequencing Consortium

Starting a sequencing run



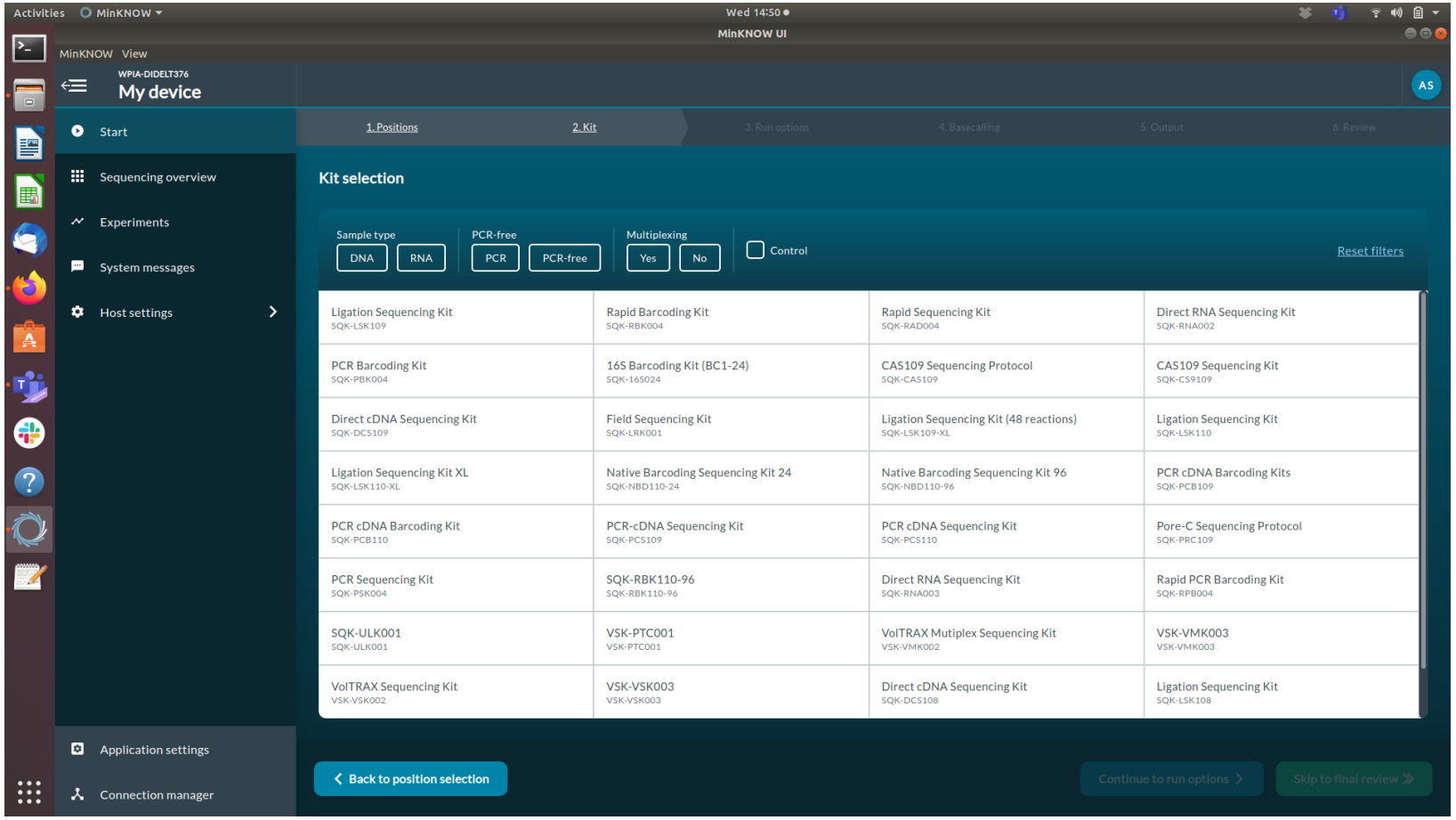
Polio Sequencing Consortium

Starting a sequencing run



Polio Sequencing Consortium

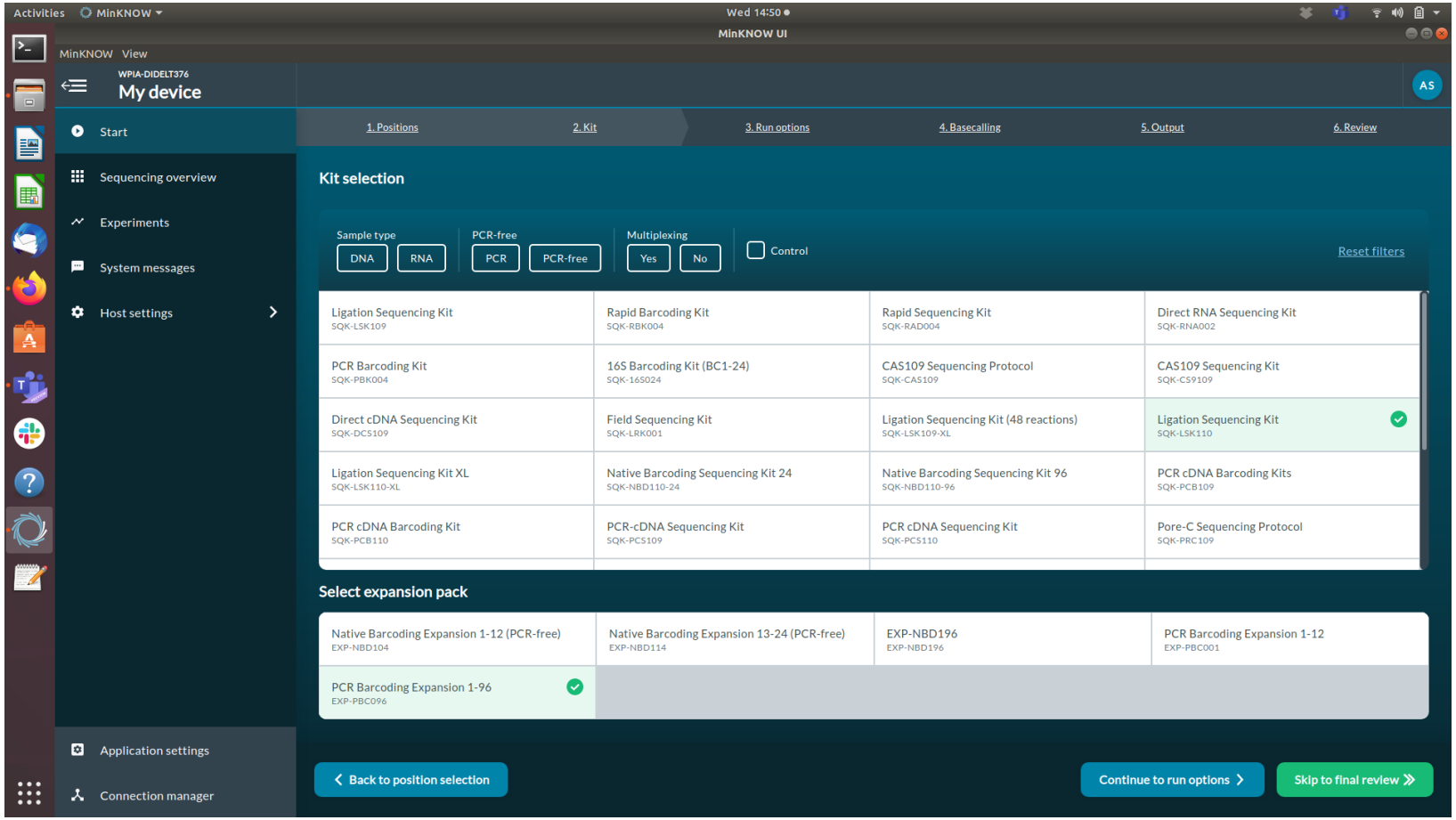
Selecting the sequencing kit



For v14 chemistry and sequencing with the barcoded VP1 primers select SQK-LSK114

Polio Sequencing Consortium

Selecting barcodes

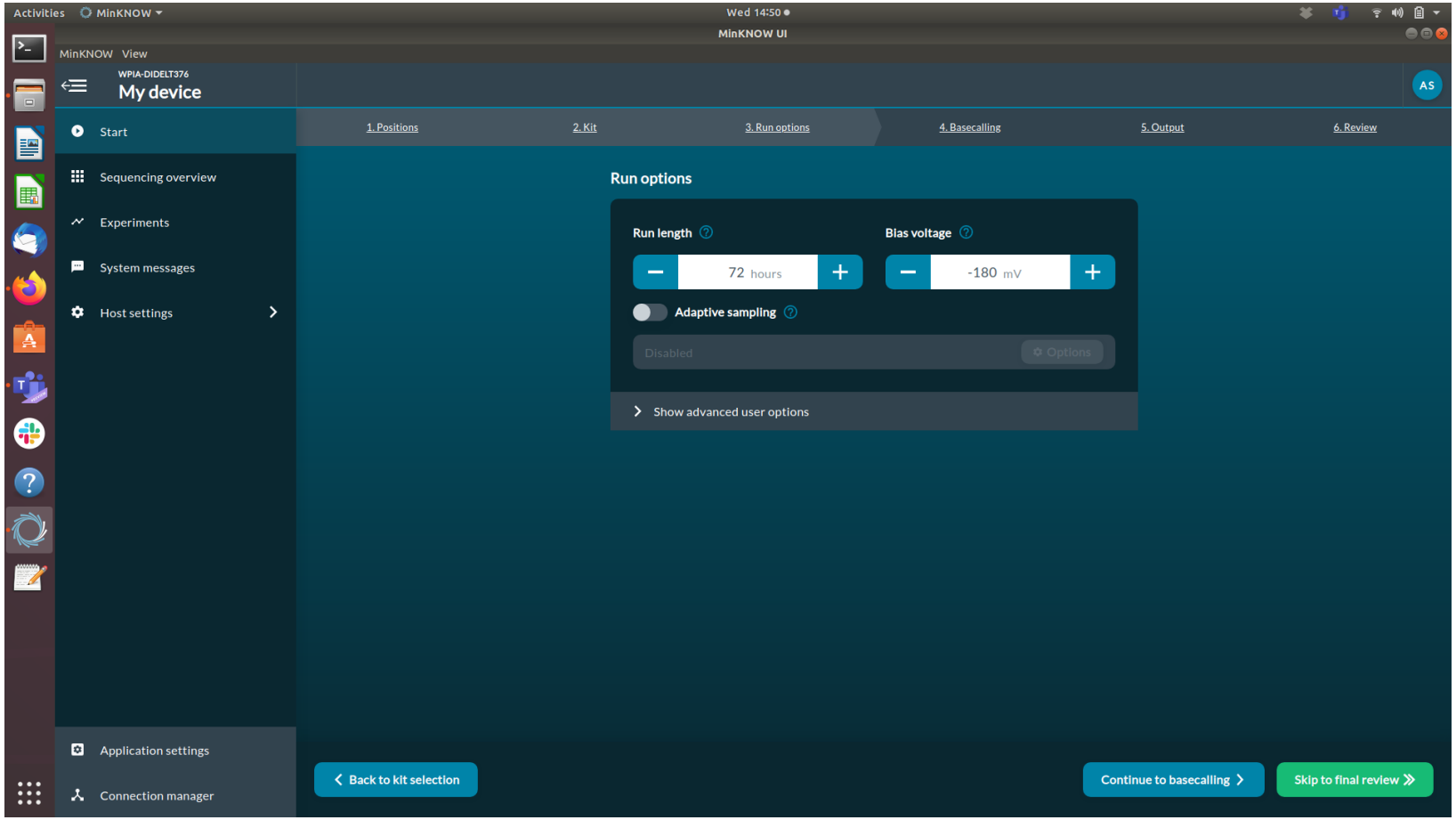


For the barcoded VP1 primers select EXP-PBC096

(note- this step may not be available in the most recent software version, but can be performed in the analysis tab after the run)

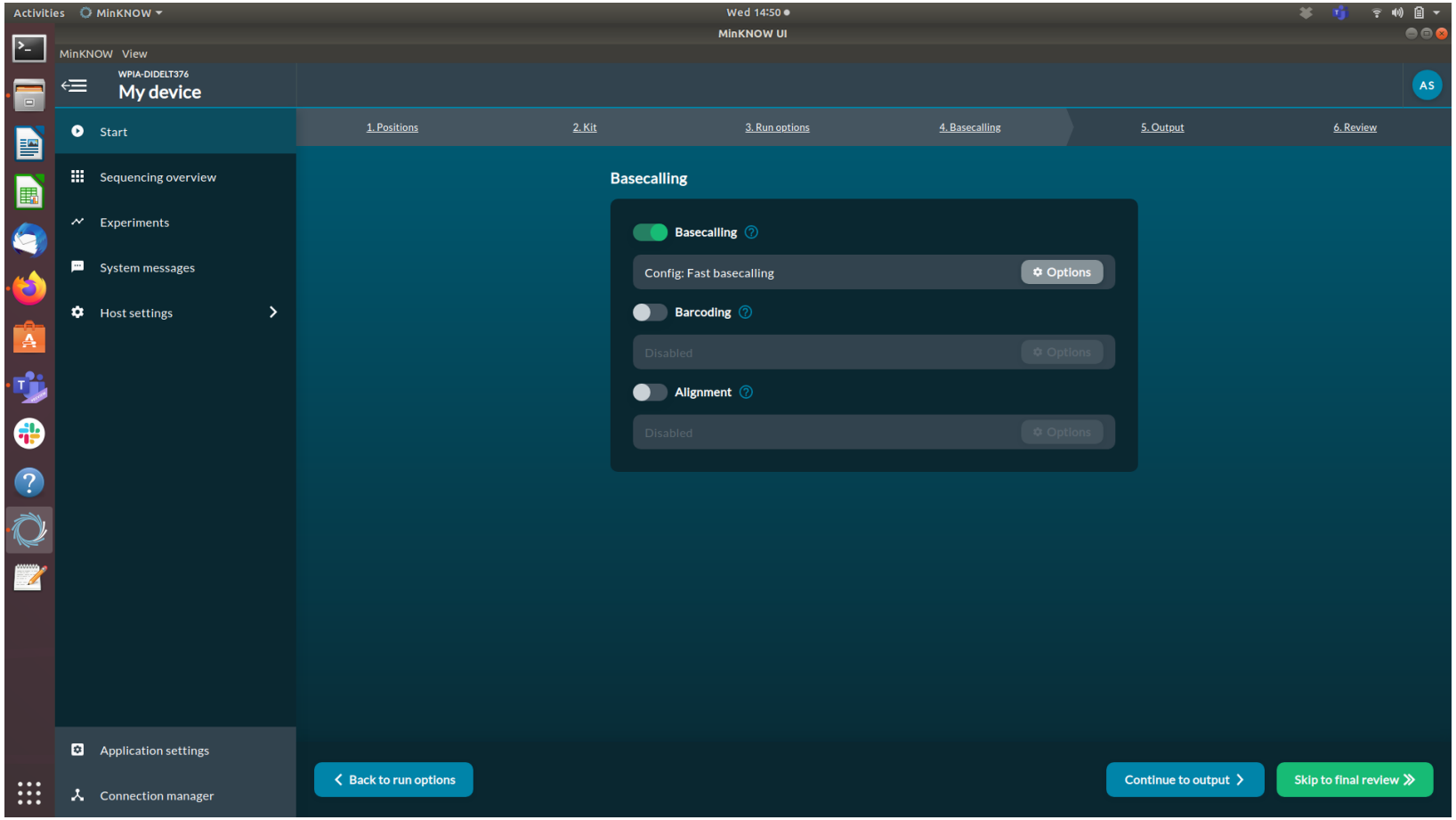
Polio Sequencing Consortium

Setting run length and voltage



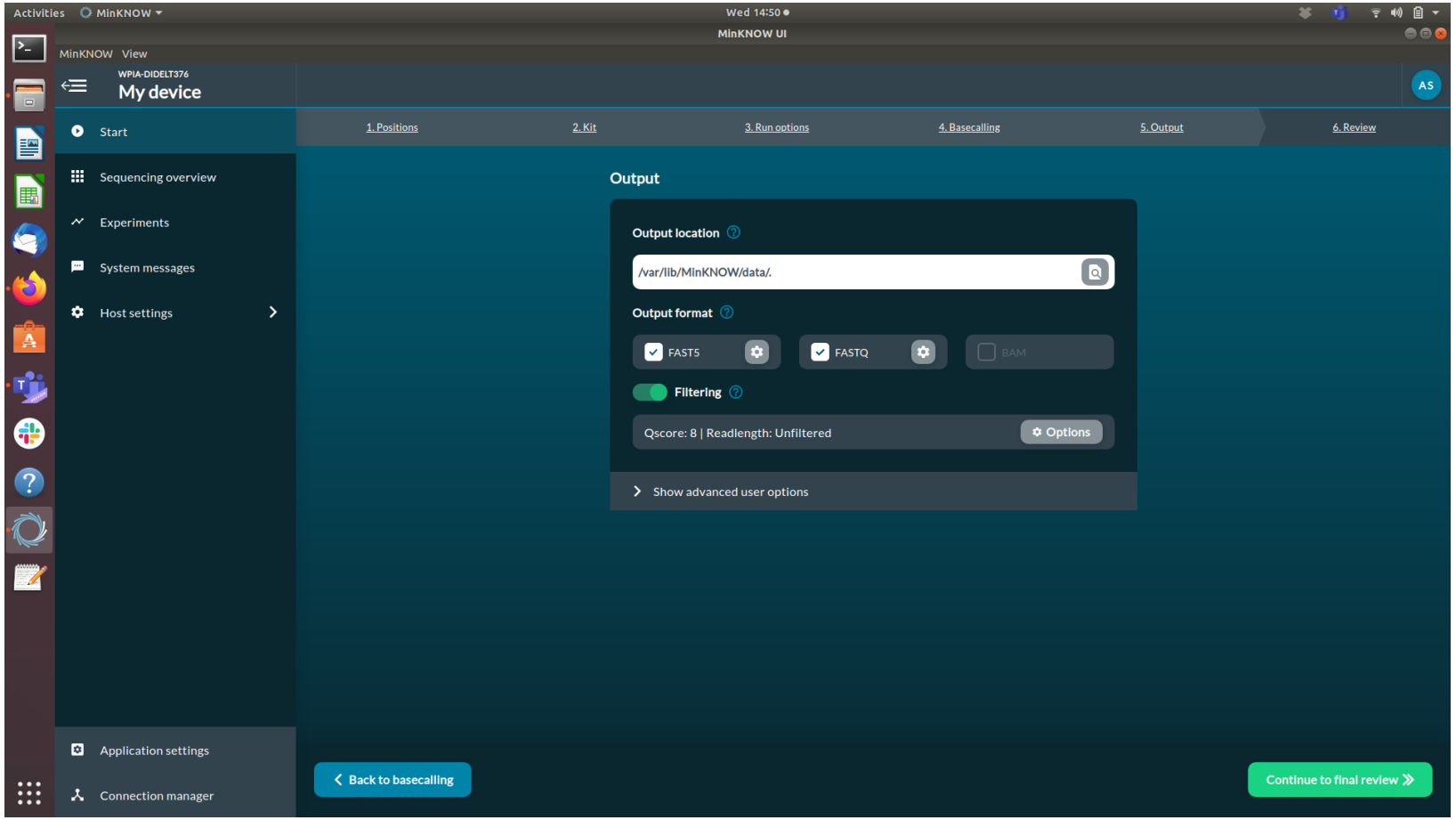
Polio Sequencing Consortium

Basecalling options



Polio Sequencing Consortium

Output options

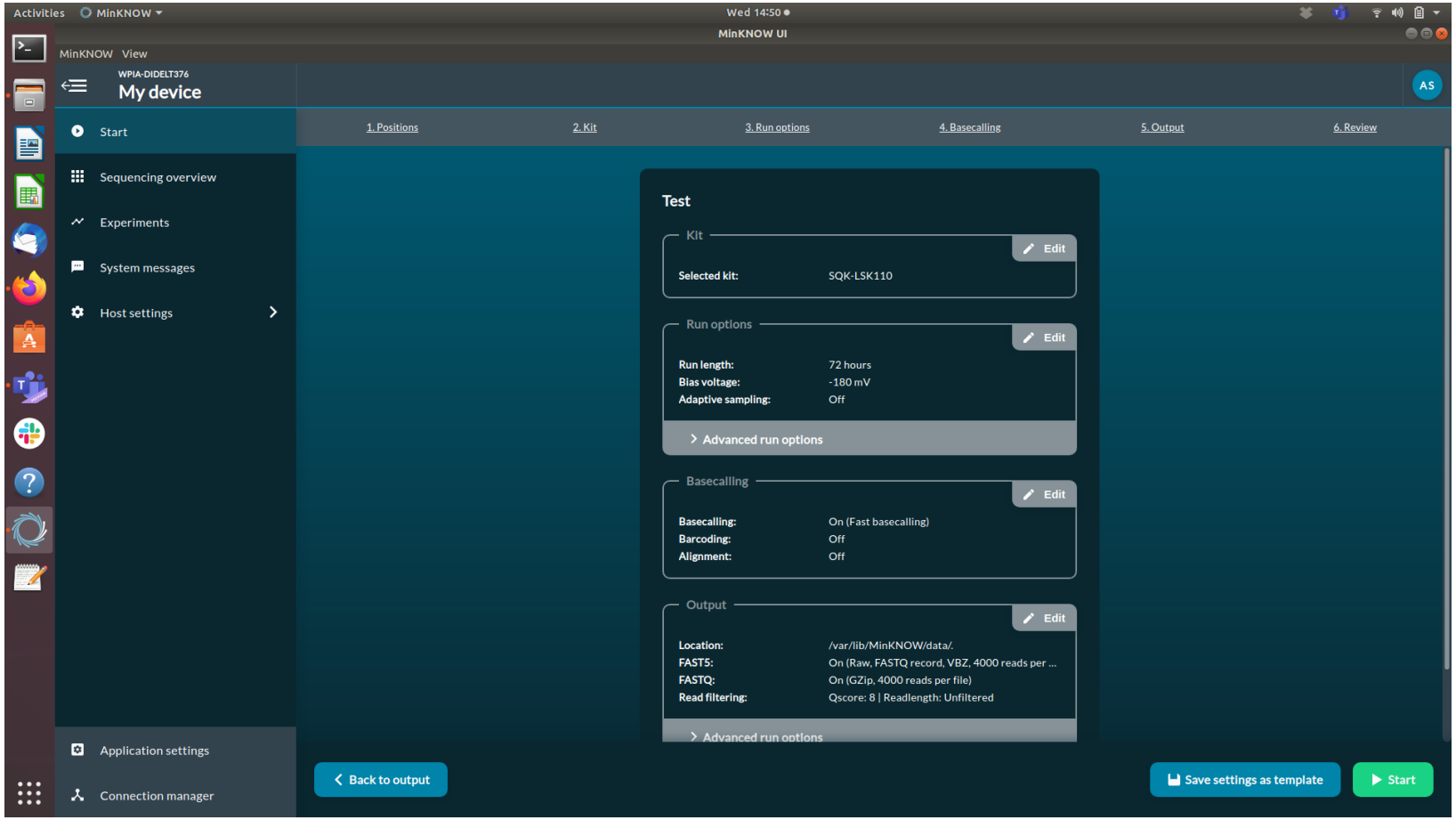


Select an analysis folder on your desktop for easy access

On a Linux computer the default is var/lib/minknow/data

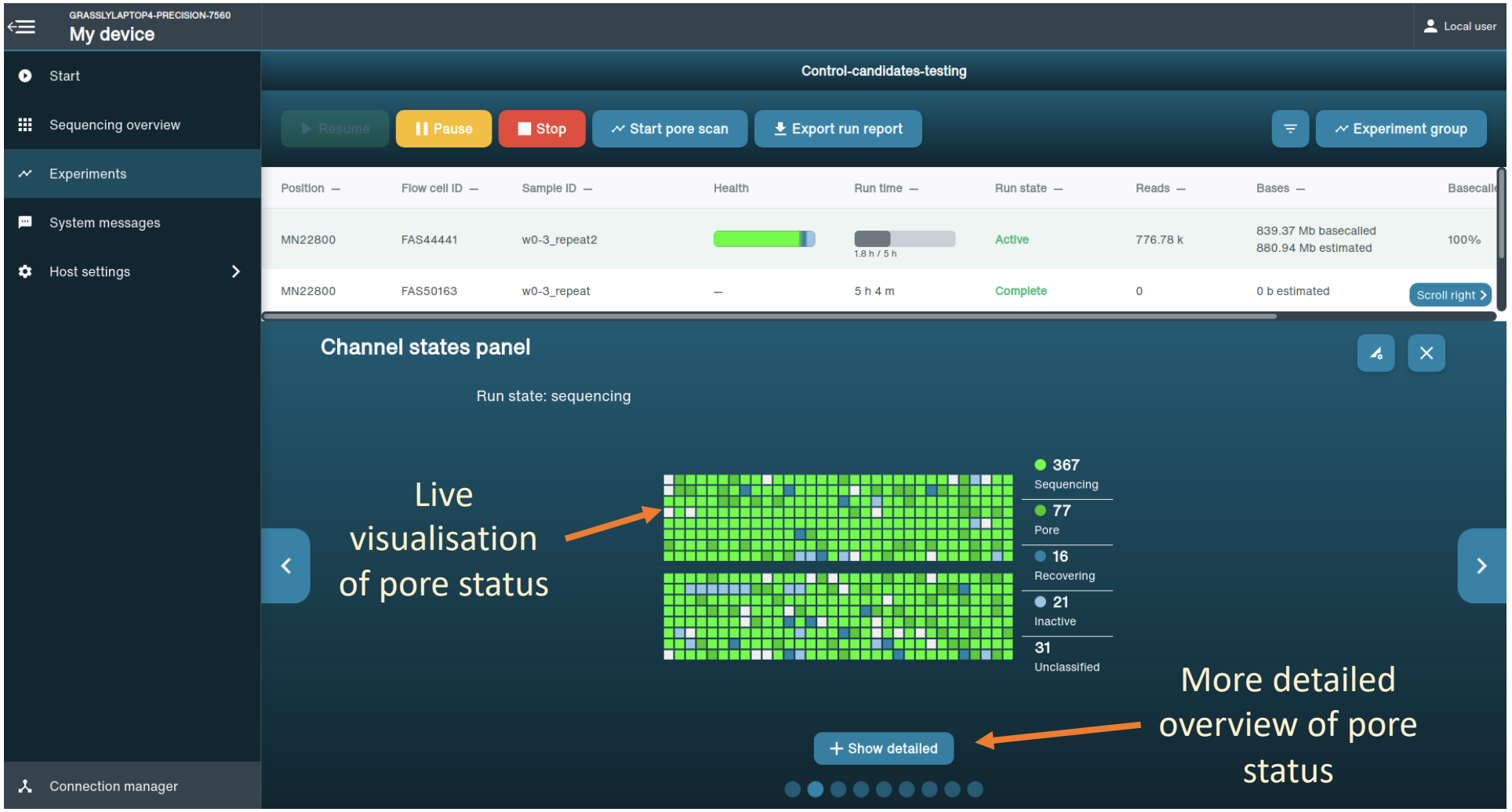
Polio Sequencing Consortium

Review the settings



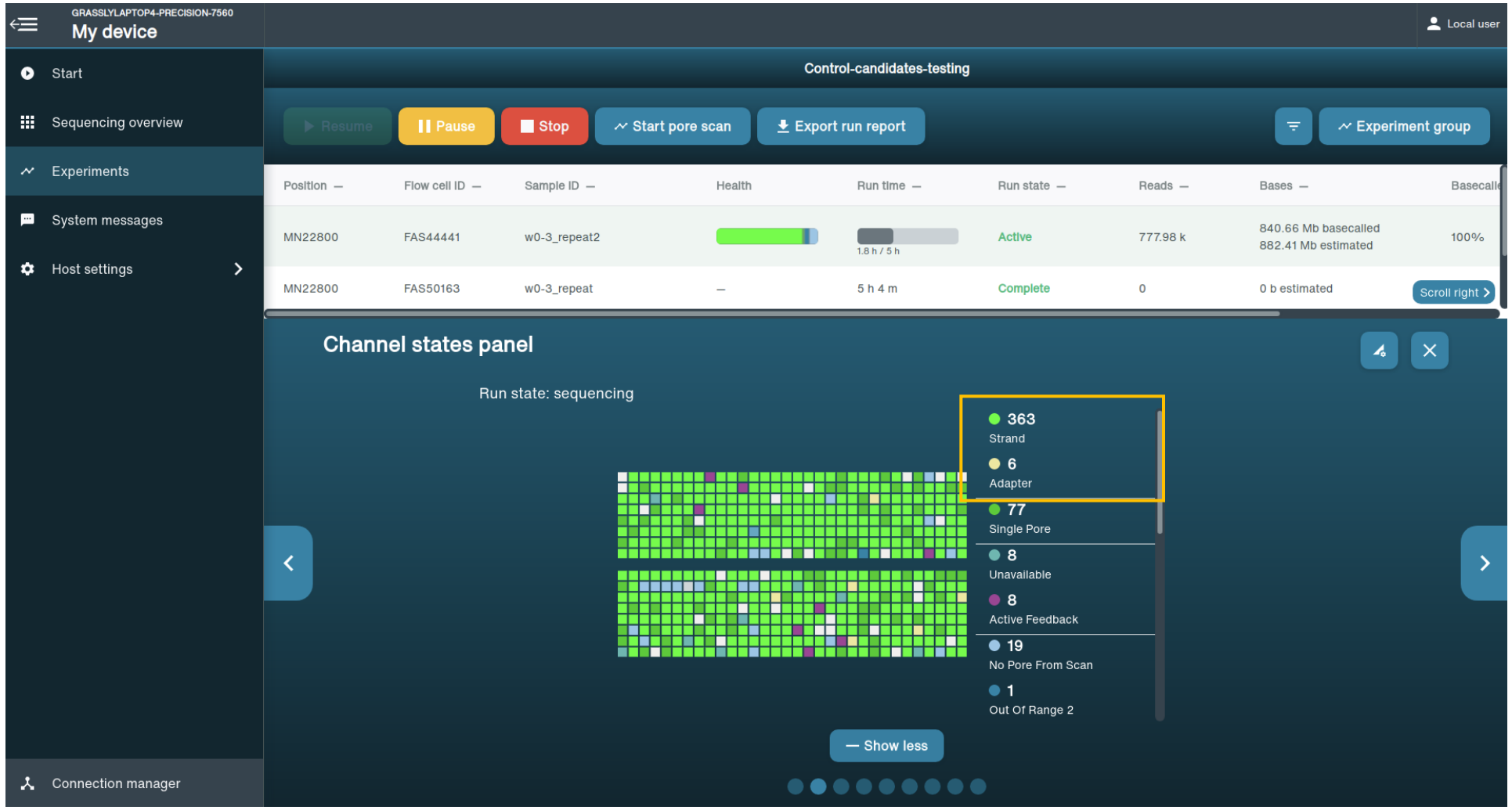
Polio Sequencing Consortium

Sequencing Run



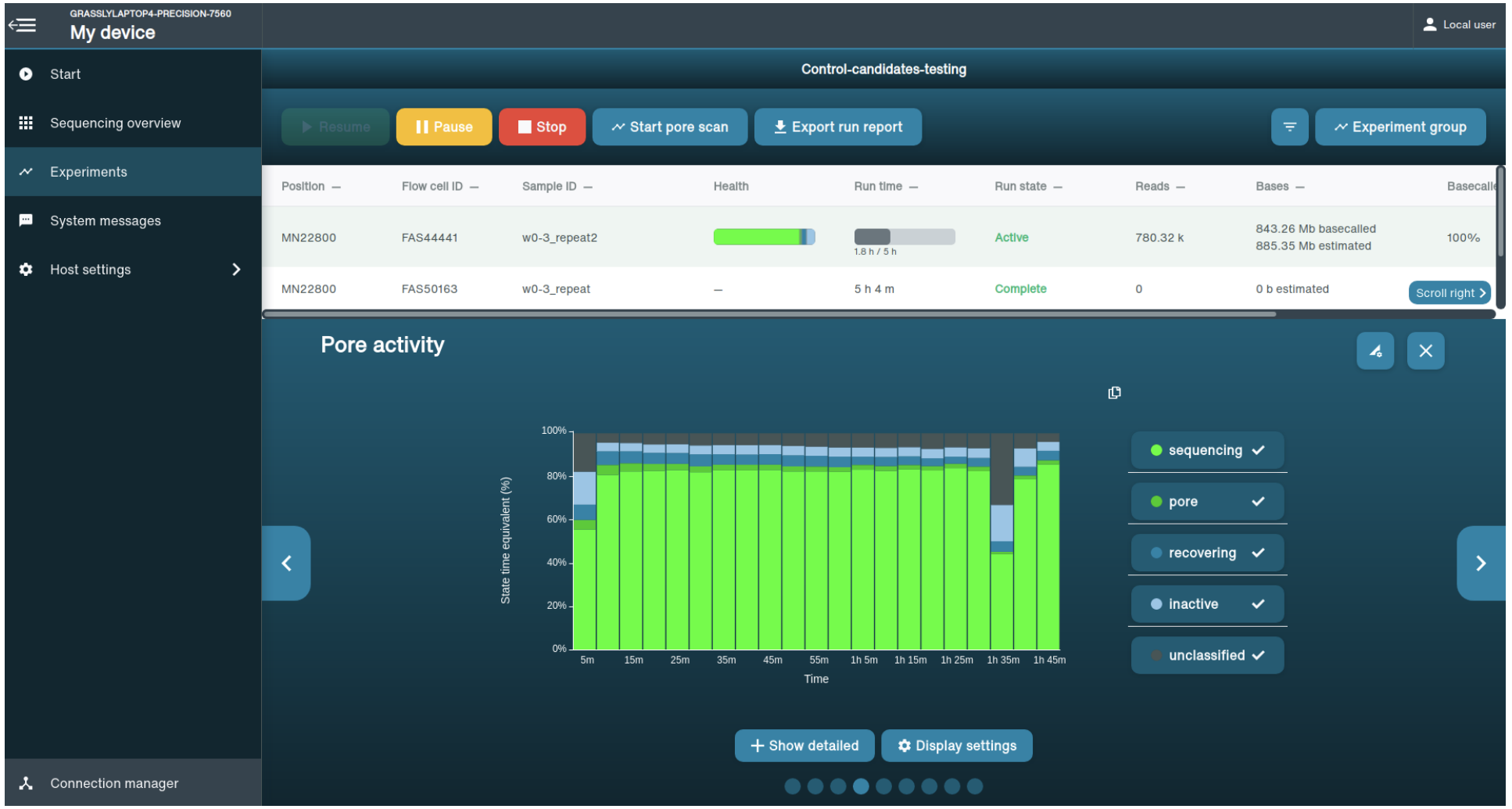
Polio Sequencing Consortium

Sequencing Run



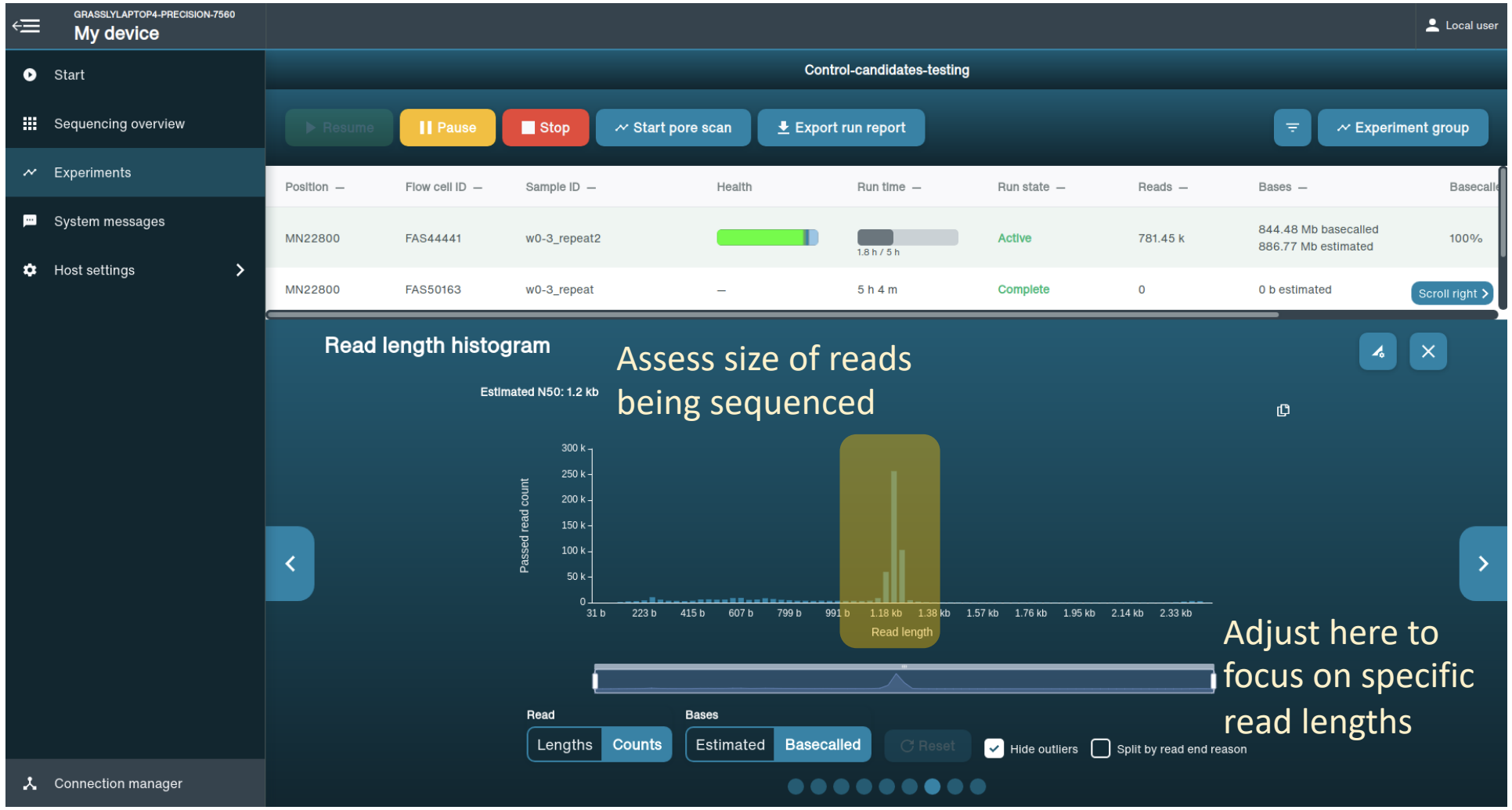
Polio Sequencing Consortium

Sequencing Run



Polio Sequencing Consortium

Sequencing Run

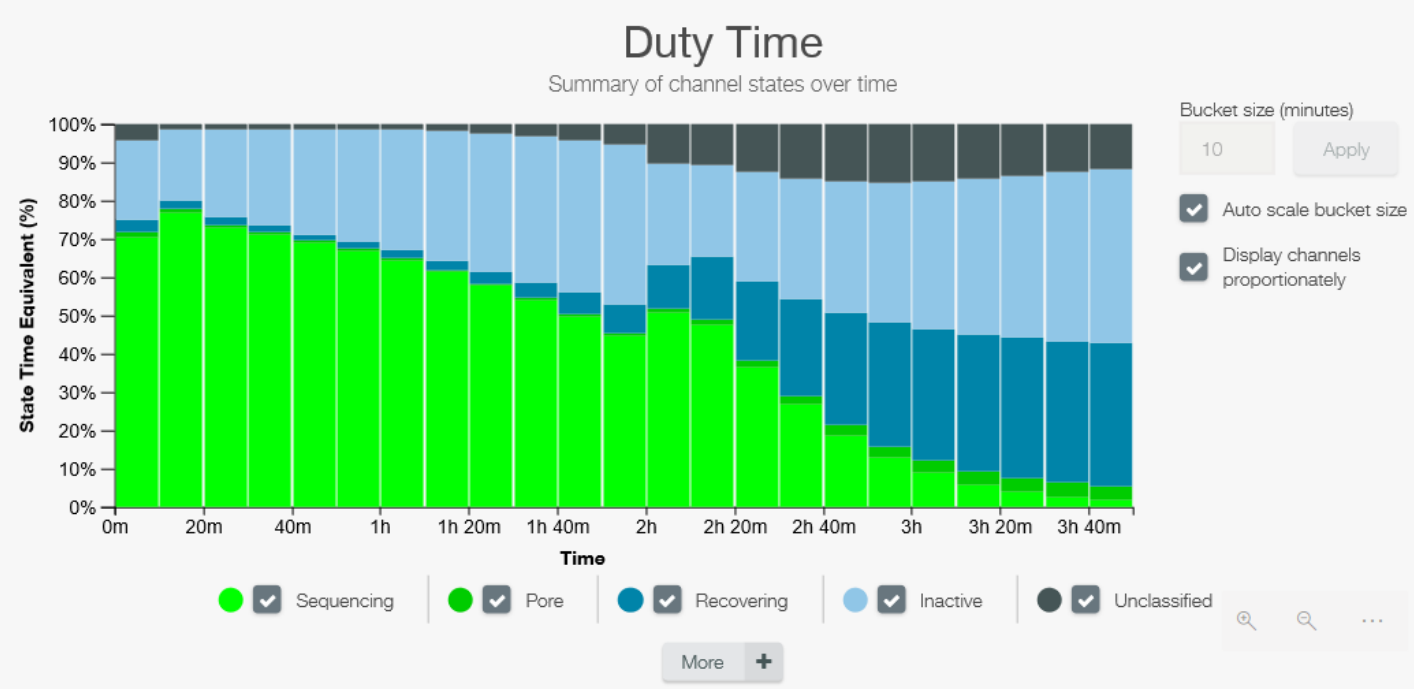


Polio Sequencing Consortium

Sequencing Run

Polio Sequencing Consortium

Examples of a bad start



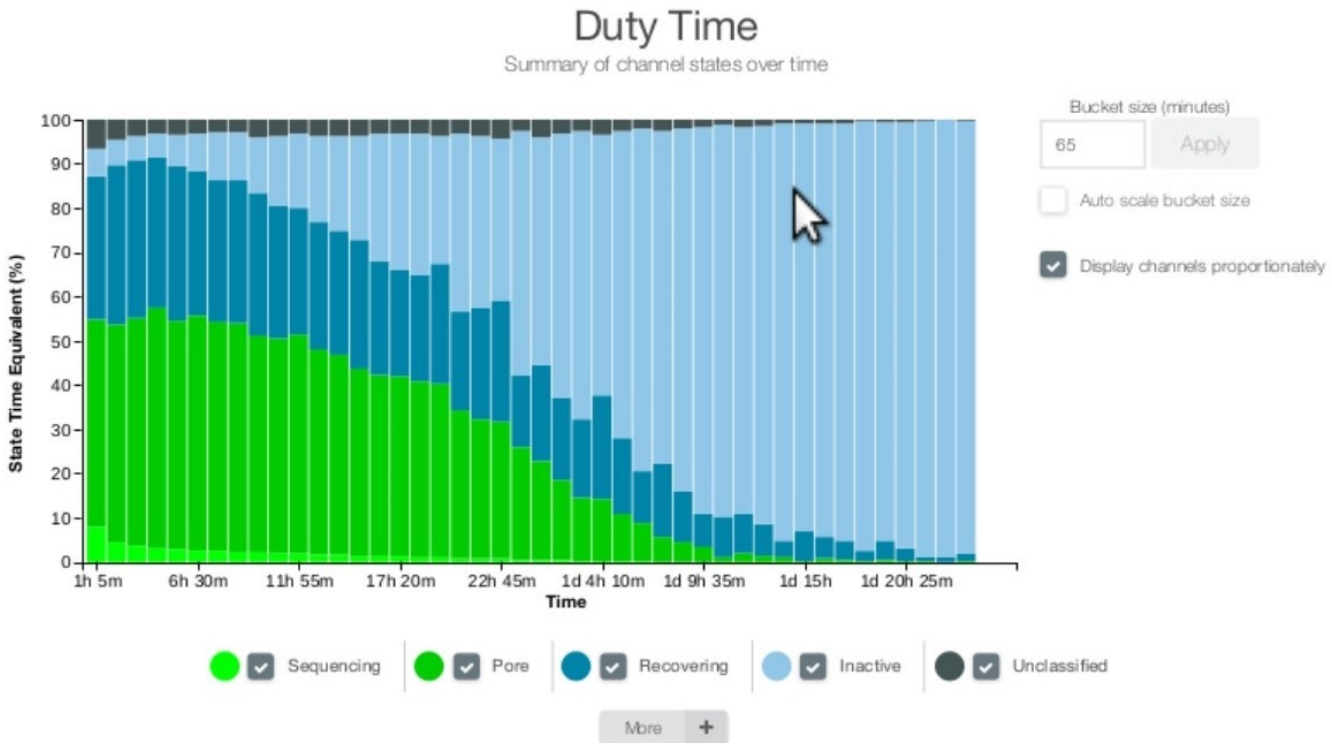
Rapid pore death may be caused by:

- Contamination
- Using an old flowcell
- Software fault

https://community.nanoporetech.com

Polio Sequencing Consortium

Examples of a bad start



Low pore occupancy may be caused by:

- Not loading enough sample onto the flowcell
- Failed adaptation for sequencing

https://community.nanoporetech.com

Polio Sequencing Consortium

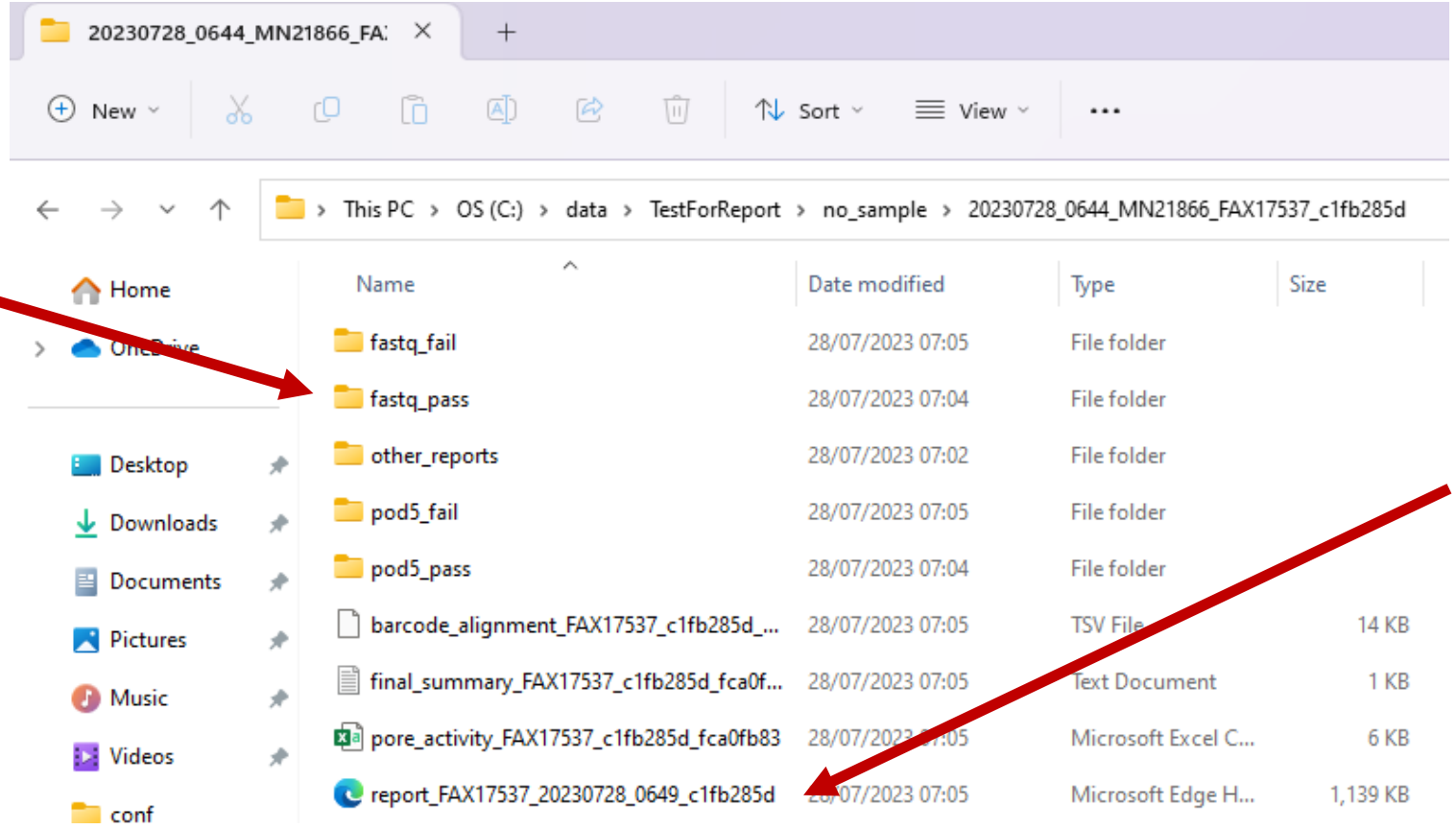
4. Post-sequencing run checks

Lecture and discussion

Polio Sequencing Consortium

1. Confirm that your run created data

Confirm that there is data in the fastq_pass folder

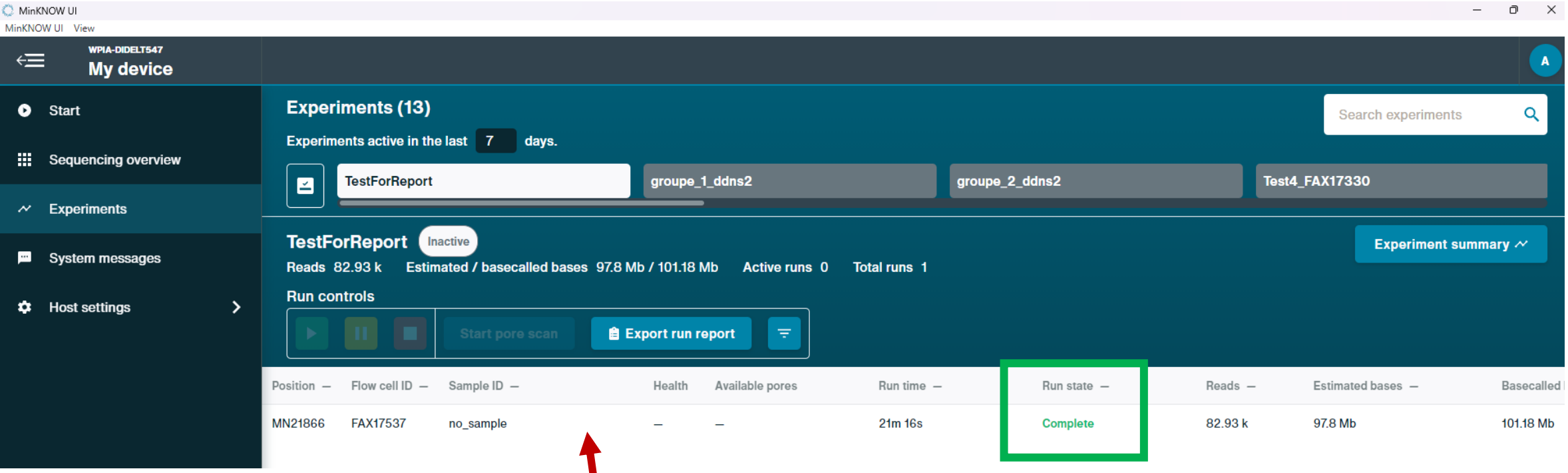


Confirm that your run report has been created (this contains a summary of the run performance)

If there is no data, you can restart your run in MinKNOW.
If there is not enough data, you can click “join existing” when you start a run to ADD the data to your prior run

Polio Sequencing Consortium

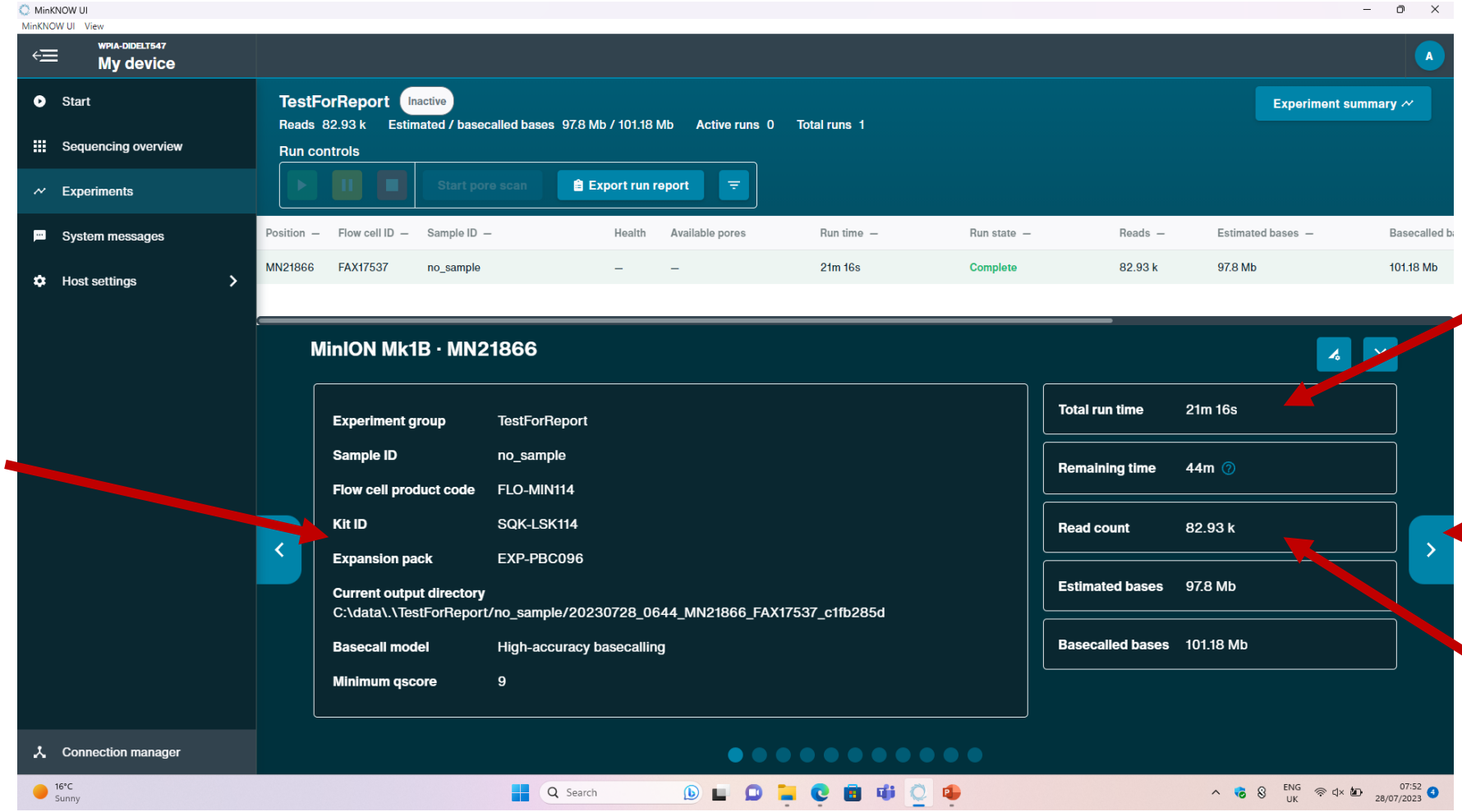
2. Check your run is complete in MinKNOW



Click on your run to see more details (will take you to the screen on the next slide)

Polio Sequencing Consortium

3. Check your run performance



Your run settings

Confirm your run length (4 hours for routine stool testing by DDNS)

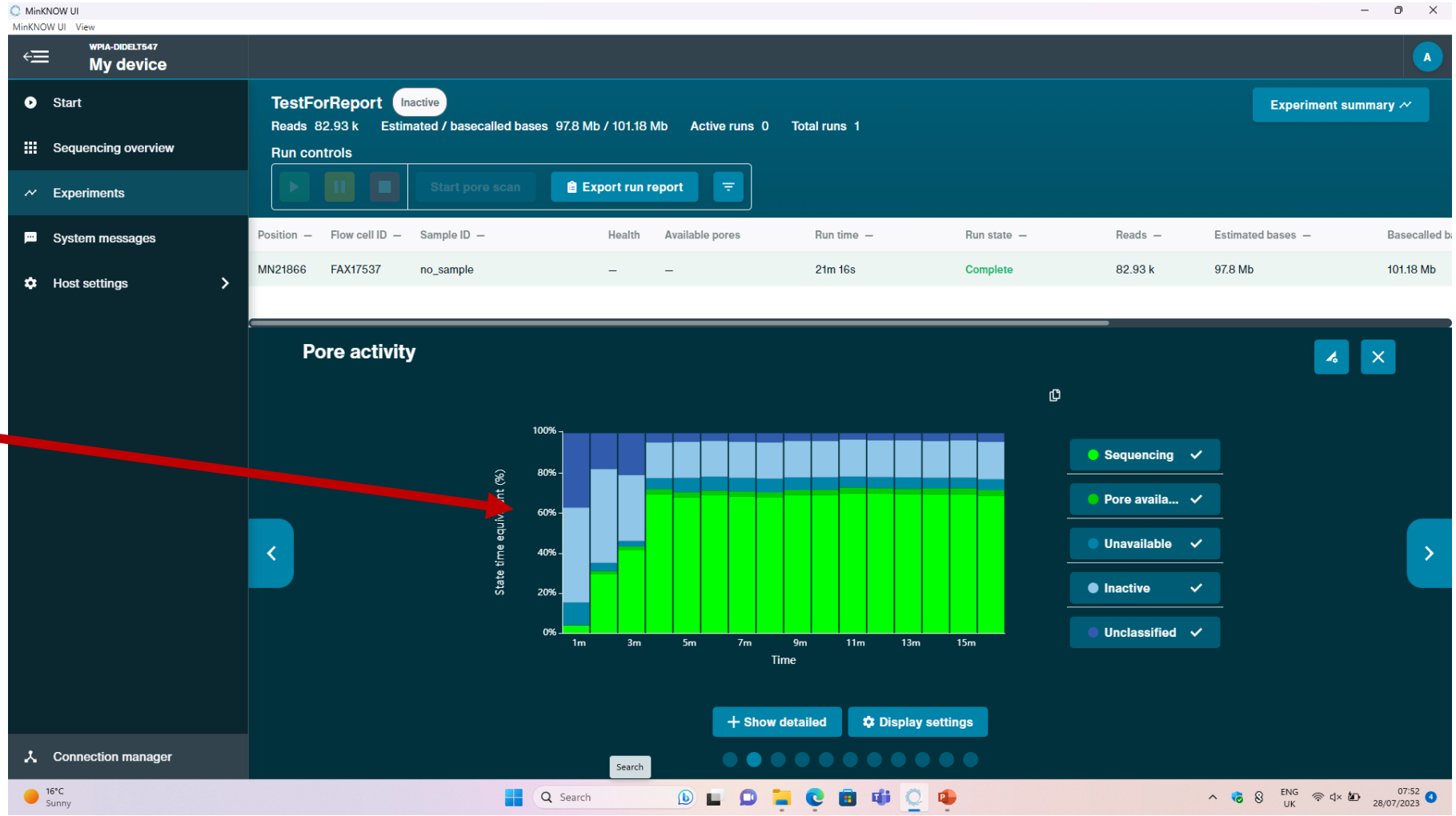
Click for more figures

The total number of sequencing reads generated

Polio Sequencing Consortium

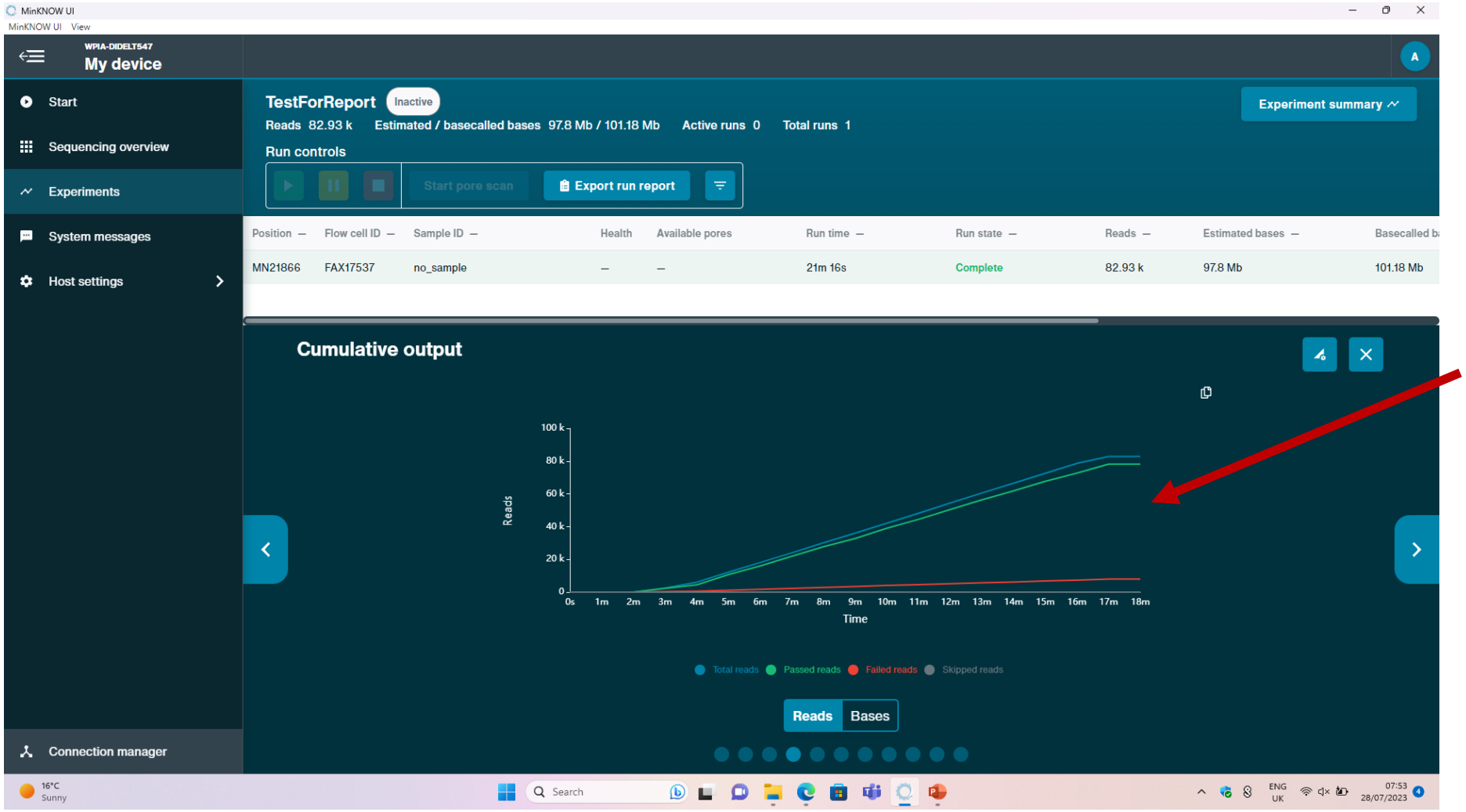
4. Check the number of pores remaining

Check how many pores are remaining at the end of your run (and again after washing). Green pores are available to sequence.



Polio Sequencing Consortium

5. Check the quality of your data

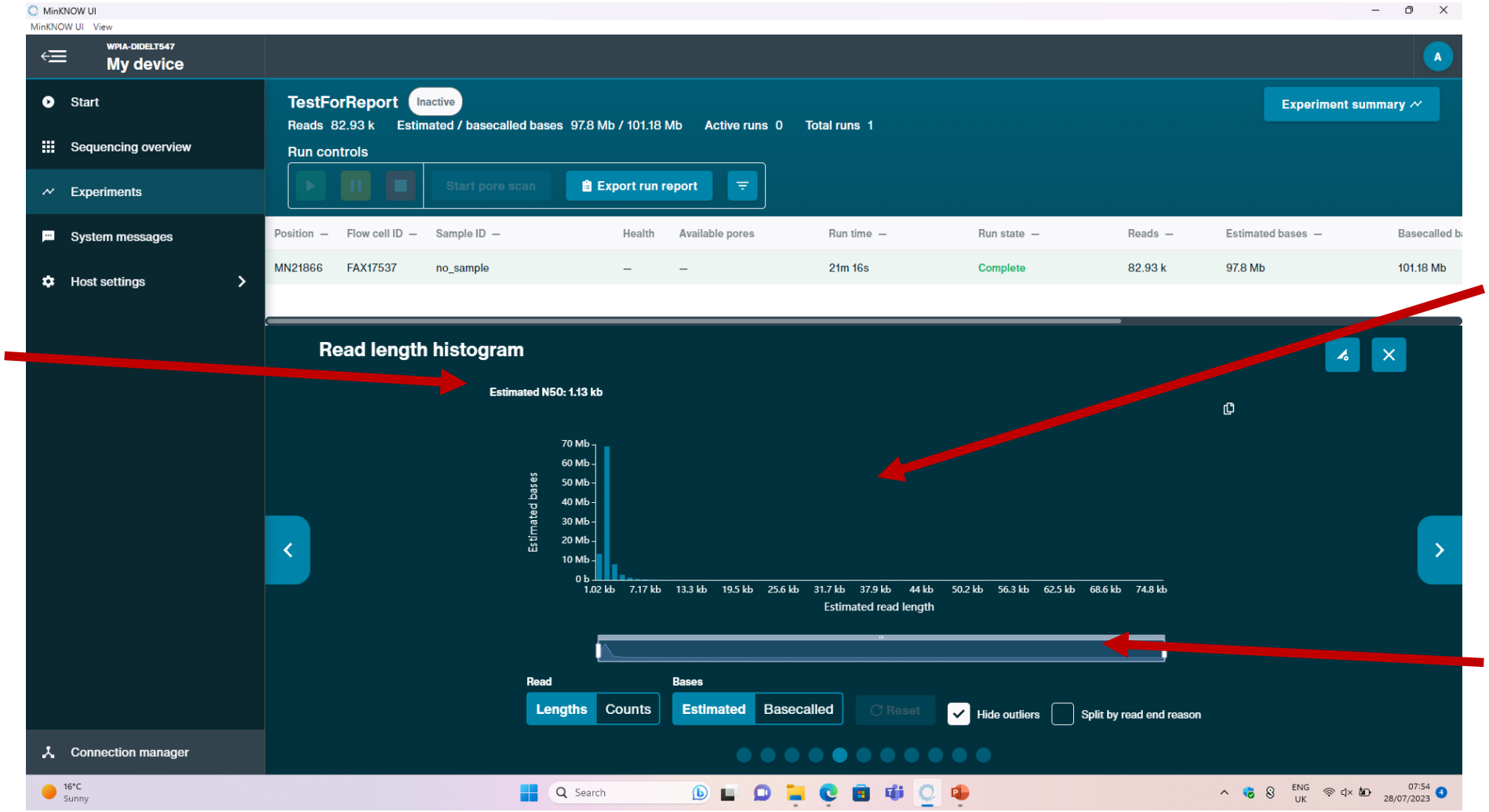


Check the proportion of reads that are passing quality checks (green line). The vast majority of reads should pass.

Polio Sequencing Consortium

6. Check the length of your sequences

Check the average read length (N50)- for VP1 by DDNS this should be ~ 1.1 kB

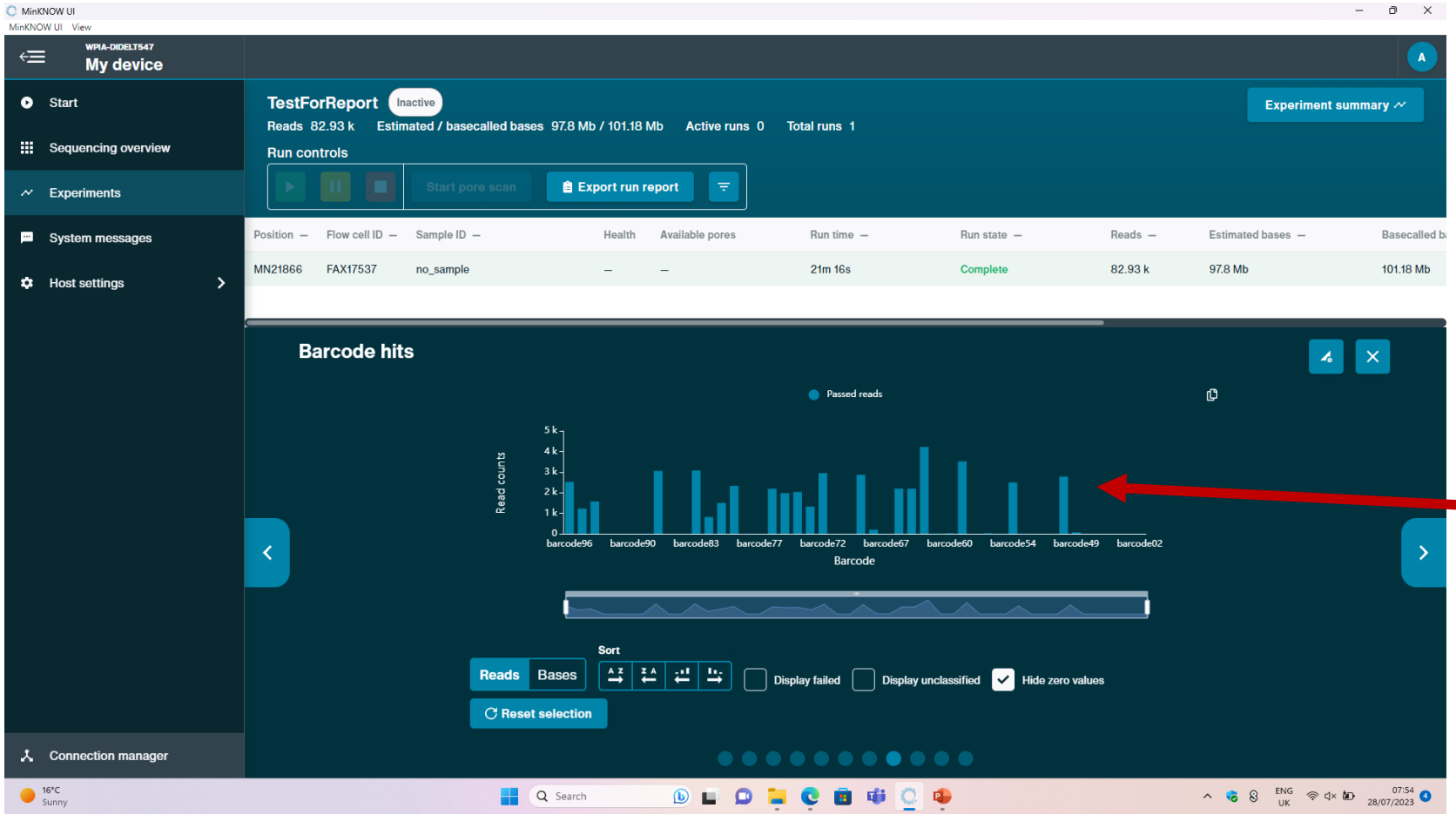


This chart shows the distribution of read lengths

Use this slider to change the width of the chart

Polio Sequencing Consortium

7. Check your data has been demultiplexed



If barcoding was turned on, this chart will show the data assigned to each barcode

If your data has not been demultiplexed you can go to “Start”, “Analysis” and “Barcoding” in MinKNOW to perform the demultiplexing.

Polio Sequencing Consortium

5. Flow cell washing

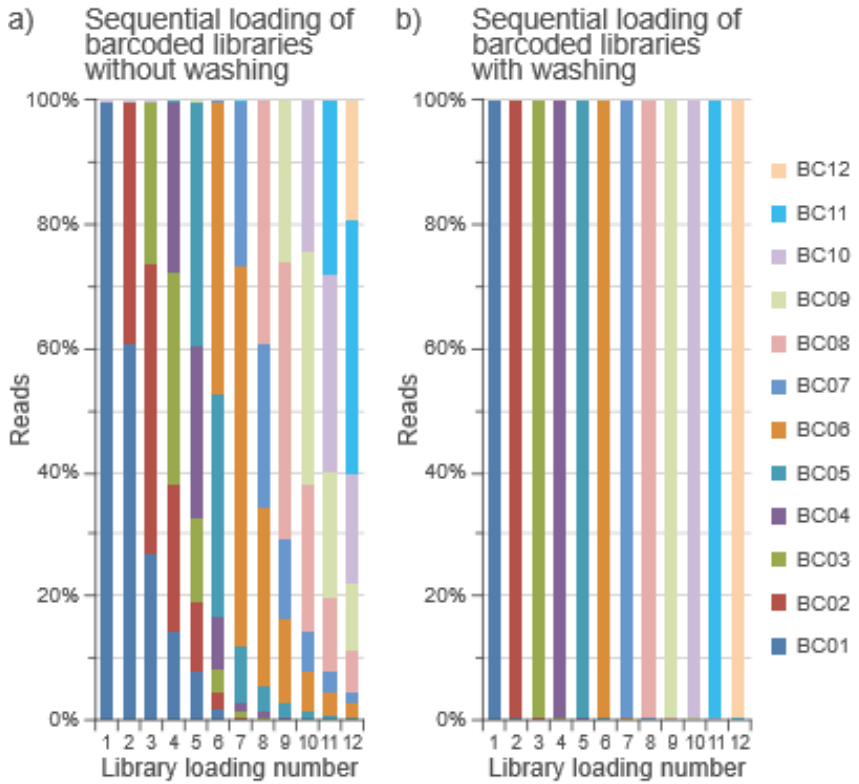
Lecture and discussion

Polio Sequencing Consortium

Washing a flow cell after a sequencing run

- Wash kit (WSH004) contains
  - Wash mix (DNAse I, WMX)
  - Wash diluent (enzyme buffer, DIL)
  - Storage buffer (S)

- DNAse I removes DNA that remains in the pores after the run ends
  - Avoids contaminating the next run



https://store.nanoporetech.com/catalog/product/view/id/122/

Polio Sequencing Consortium

Protocol

- Combine 2ul of the DNAse enzyme (WMX) with 398ul of the wash buffer (DIL)
- Remove any air bubbles under the priming port
- Put the 400ul wash solution onto the flow cell through the priming port with the SpotOn port closed
- Incubate at room temperature for 60 minutes

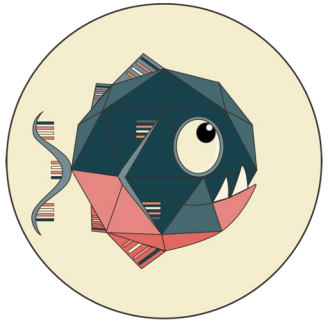
Polio Sequencing Consortium

Protocol

- For immediate use:
  - Flush the flow cell with Flush Buffer as described for flow cell priming

- For storage:
  - Add 500ul of the storage buffer (S) through the priming port, put the flow cell back into its packaging
  - Store in the fridge until the next use

Polio Sequencing Consortium

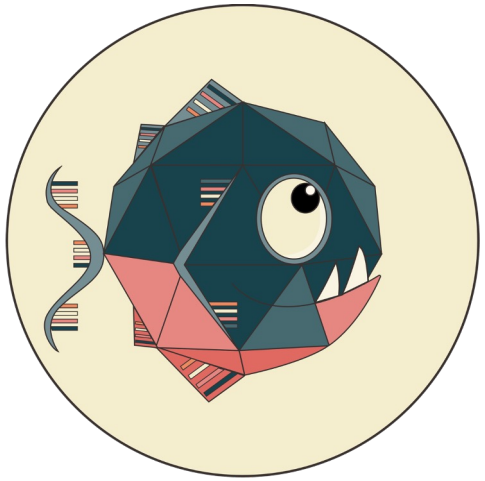


6. Installation and use of piranha

Poliovirus investigation resource

automating nanopore haplotype analysis

Lecture and discussion



Polio Sequencing Consortium

piranha

Poliovirus investigation resource automating nanopore haplotype analysis

Áine O'Toole[1,*] Rachel Colquhoun[1], Corey Ansley[1], Catherine Troman[2], Daniel Maloney[1], Zoe Vance[1], Joyce Akello[2], Erika Bujaki[3], Manasi Majumdar[3], Adnan Khurshid[4], Yasir Arshad[4], Muhammad Masroor Alam[4], Javier Martin[3], Alexander G Shaw[2], Nicholas C Grassly[2,$], Andrew Rambaut[1,$]

Affiliations

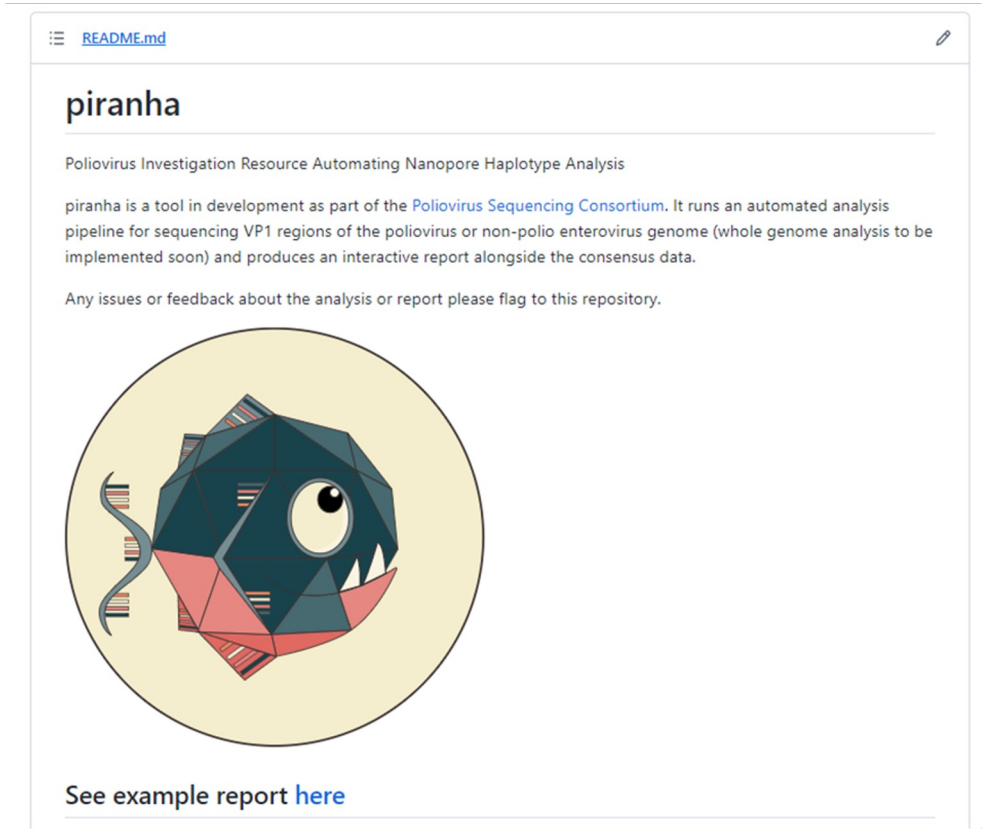
1. Institute of Ecology & Evolution, University of Edinburgh, United Kingdom
2. Department of Infectious Disease Epidemiology, Imperial College London, United Kingdom
3. Division of Virology, National Institute for Biological Standards and Control (NIBSC), Hertz, United Kingdom
4. Department of Virology, National Institute for Health, Islamabad, Pakistan

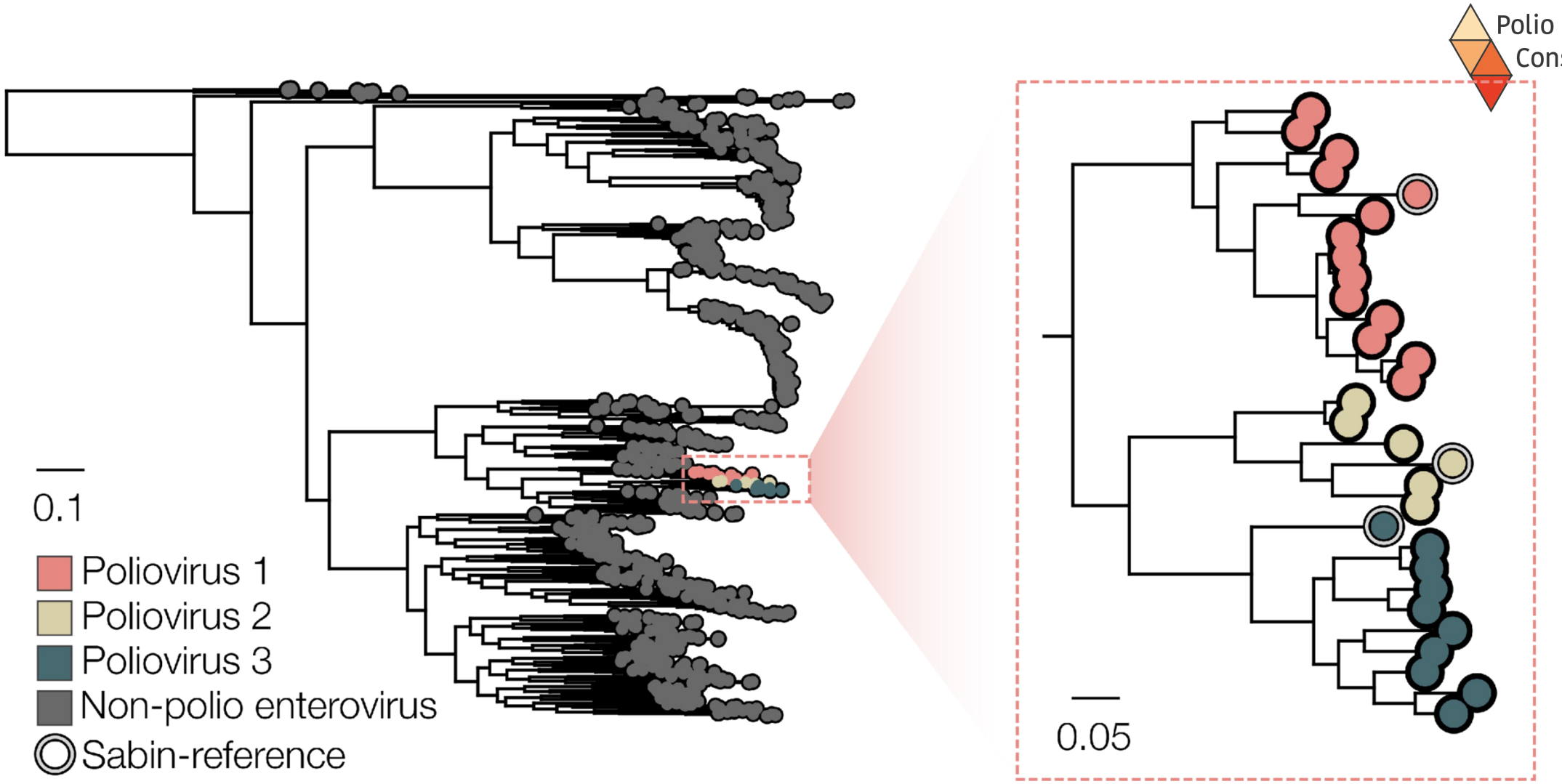
Polio Sequencing Consortium

PIRANHA - overview

- Takes demultiplexed fastq files, maps to reference sequences, filters by length, and produces consensus sequences and reports as output
- Requires command line
- Installation instructions and basic usage available on github:

https://github.com/polio-nanopore/piranha





Background database: 959 VP1 sequences


Mixed Sample
Reference groups
Polio Sequencing Consortium
Consensus/ haplotype

Polio Sequencing Consortium

PiranhaGUI

- Uses *Docker* to allow running of PIRANHA on different operating system (e.g. Windows, Linux) and to avoid using the command line!

- Available on github here:

https://github.com/polio-nanopore/piranha

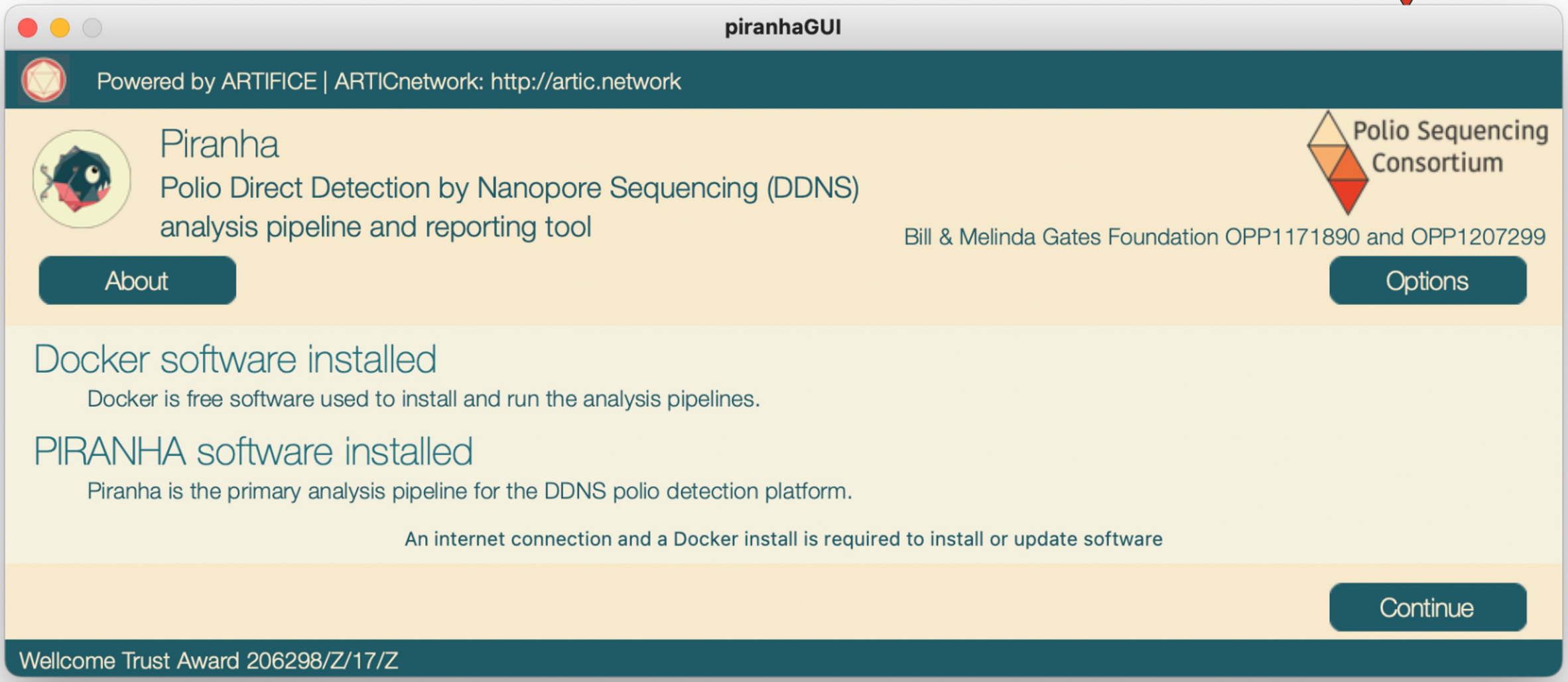
- Requires same input as piranha and produces the same output

Polio Sequencing Consortium

PiranhaGUI Installation

- Download from https://github.com/polio-nanopore/piranha/releases
- To install for windows, download: PiranhaGUIv1.5.0_installer_windows.exe
- Double click the downloaded file to install
- Open the PiranhaGUI when installed
- Click on the “Install Docker” button to go to the Docker website to download Docker
- Click on the “Install piranha” button to install piranha inside piranhaGUI.

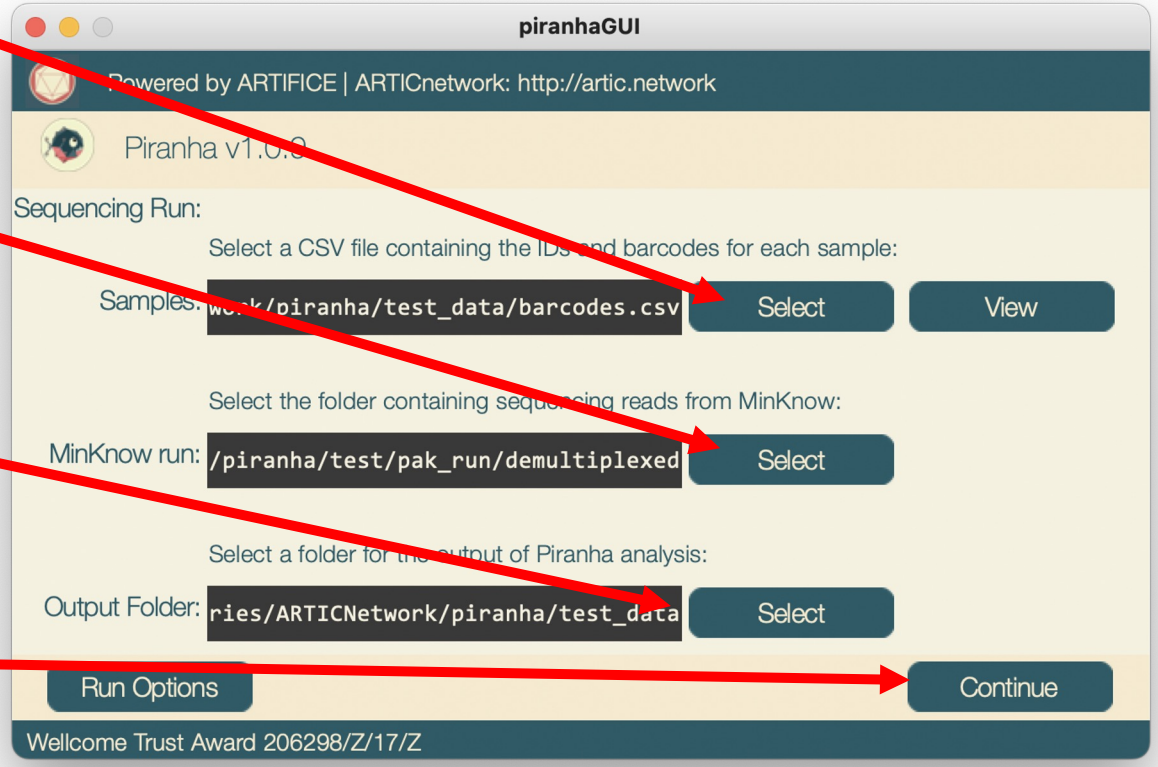
Polio Sequencing Consortium



Polio Sequencing Consortium

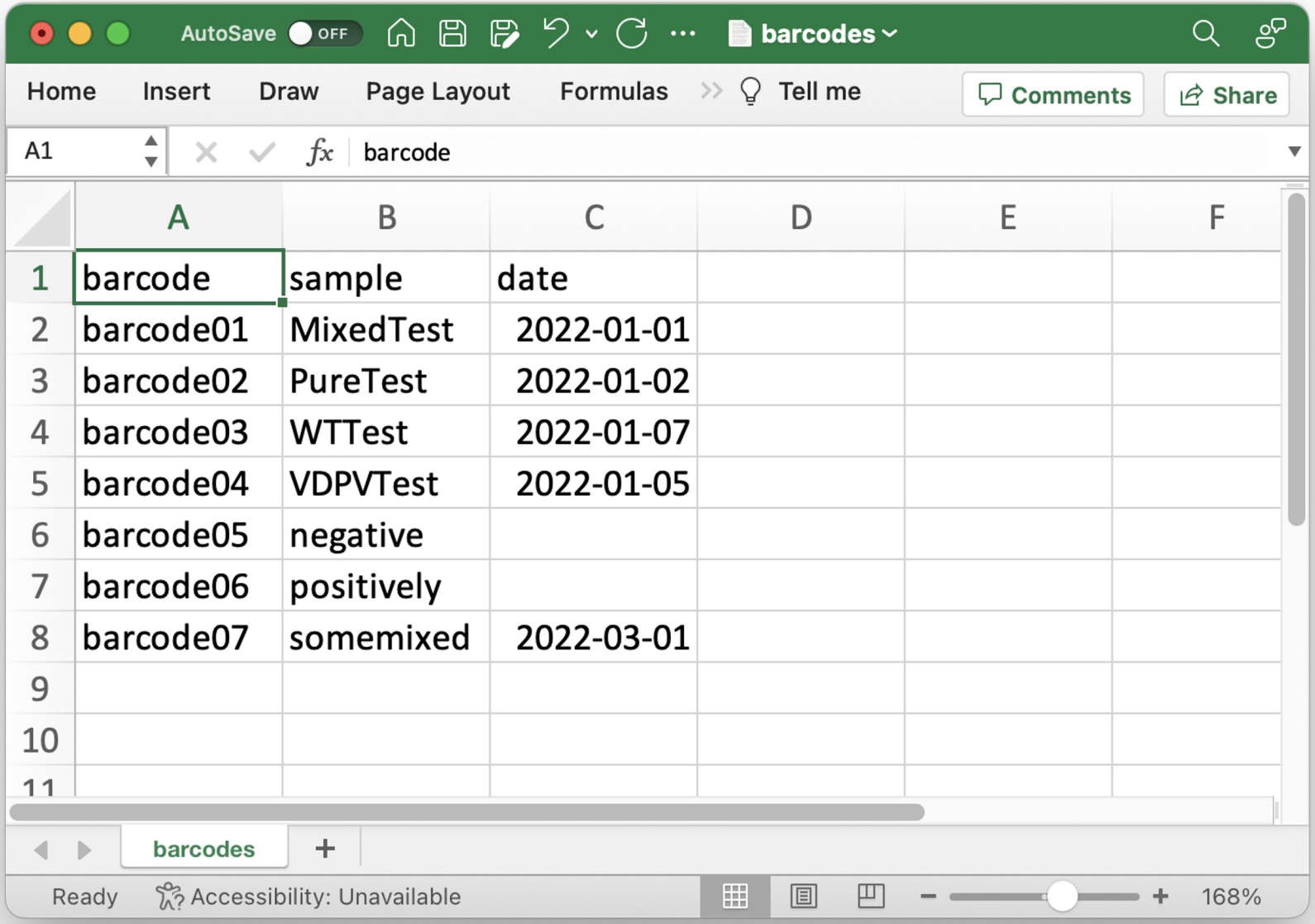
Running PiranhaGUI

- Select your barcode.csv file
- and your demultiplexed sequencing data
- Select an output folder for your results
  - Click “Continue”



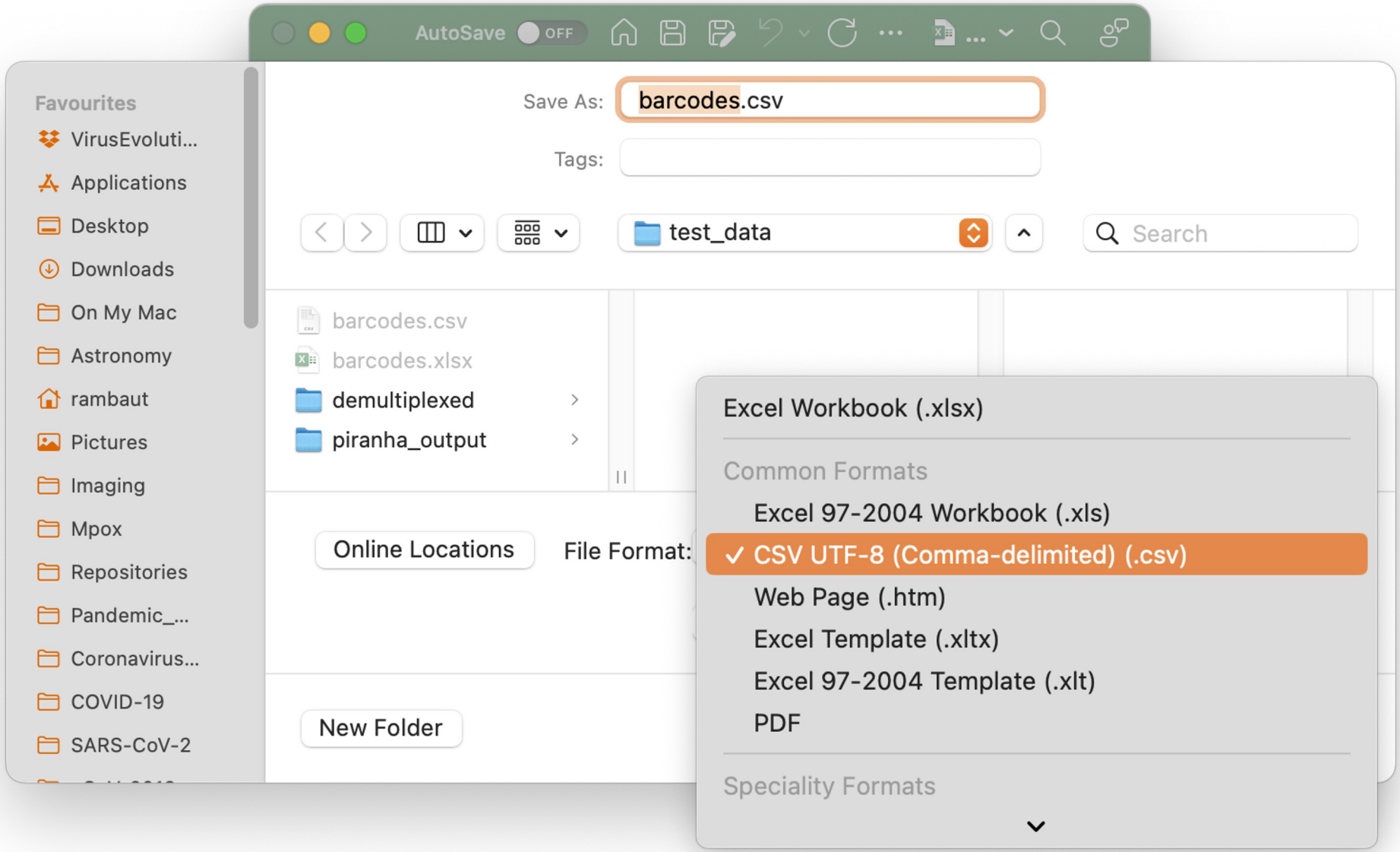
Barcodes and Samples

Polio Sequencing Consortium



	A	B	C	D	E	F
1	barcode	sample	date			
2	barcode01	MixedTest	2022-01-01			
3	barcode02	PureTest	2022-01-02			
4	barcode03	WTTest	2022-01-07			
5	barcode04	VDPVTest	2022-01-05			
6	barcode05	negative				
7	barcode06	positively				
8	barcode07	somemixed	2022-03-01			
9						
10						
11						

Polio Sequencing Consortium



Polio Sequencing Consortium

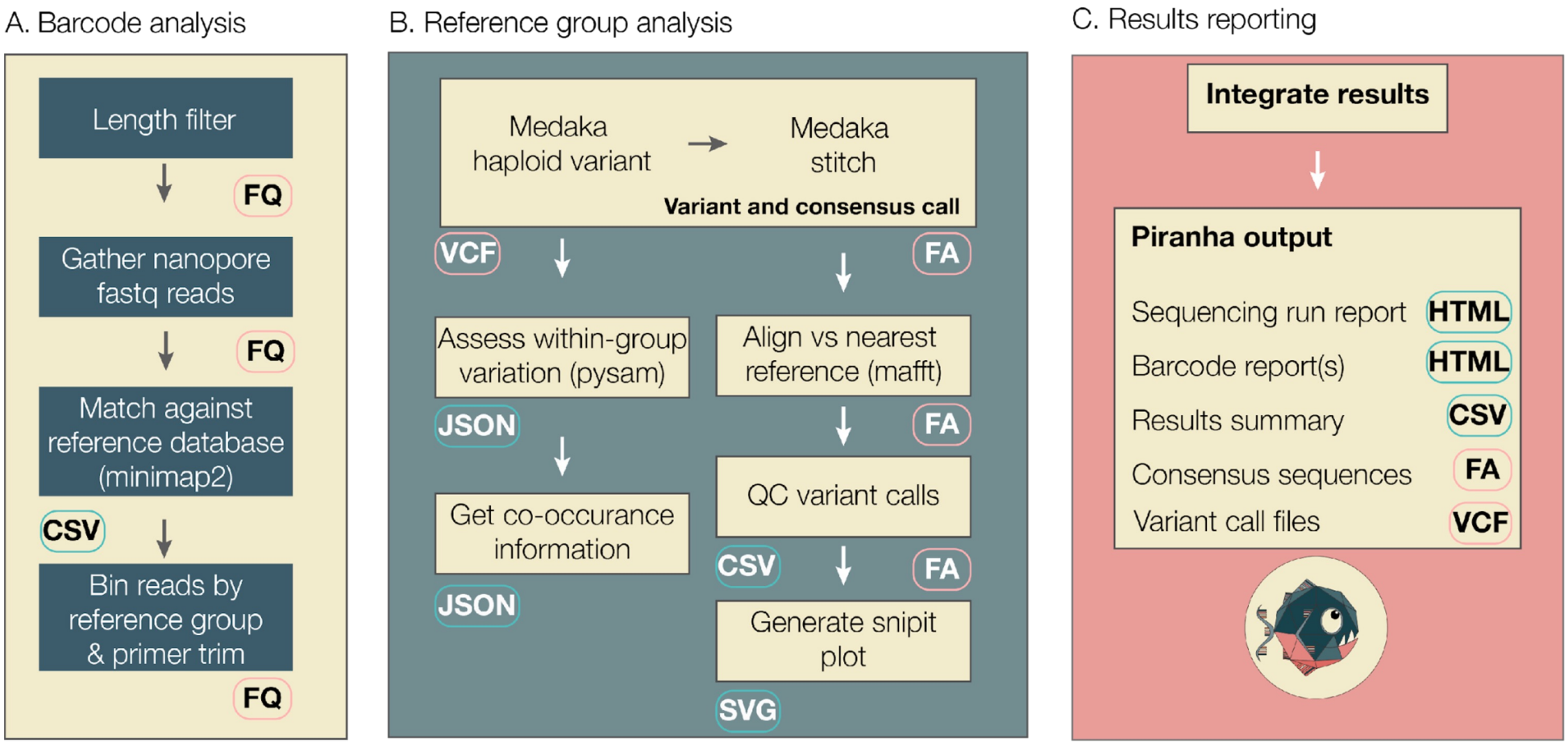
barcodes.csv — Edited

```
barcode,sample,date,well
barcode01,MixedTest,2022-01-01,A01
barcode02,PureTest,2022-01-02,B01
barcode03,WTTest,2022-01-07,C01
barcode04,VDPVTest,2022-01-05,D05
barcode05,negative,,D06
barcode06,positively,,D07
barcode07,somemixed,2022-03-01,H12
```

Polio Sequencing
Consortium

Polio Sequencing Consortium

Workflow schema



Example report

Piranha report 2022-12-08

a.

Sample	Barcode	Sample Classification	Reference group	Number of mutations
ENV001	barcode01	Sabin-like	Sabin3-related	1
ENV001	barcode01	Sabin-like	Sabin2-related	2
ENV002	barcode02	VDPV	Sabin2-related	12
ENV003	barcode03	WPV1	WPV1	NA
ENV004	barcode04	VDPV	Sabin2-related	10

b.

Sample	Barcode	Sabin1-related	Sabin2-related	Sabin3-related	WPV1	WPV2	WPV3	NonPolioEV	unmapped
ENV001	barcode01	0	488	252	0	0	0	0	25
ENV002	barcode02	0	1100	0	0	1	0	35	12
ENV003	barcode03	0	0	0	339	0	0	0	0
ENV004	barcode04	0	0	0	138	0	0	0	0
negative	barcode05	0	0	0	0	0	0	1	10

Showing 5 of 96 entries

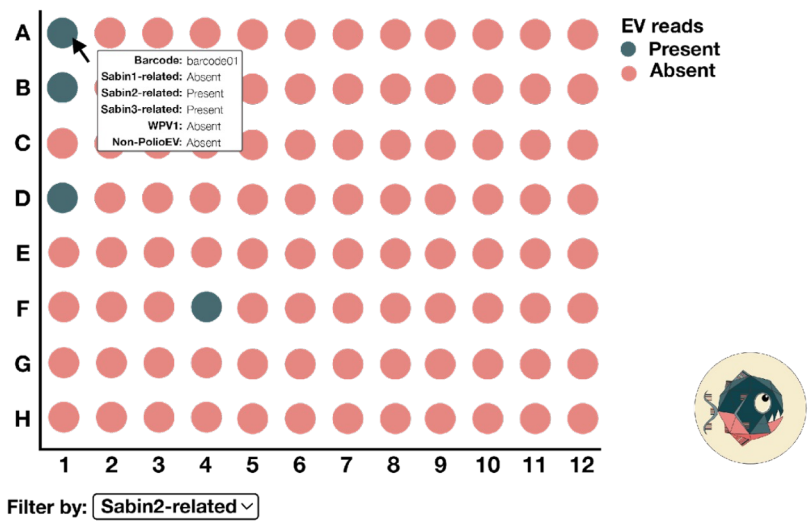
c.

Identical Sequences	Sequence IDs
1	ENV001|barcode01|Sabin2-related|2|161:CT;427:GA ENV030|barcode30|Sabin2-related|2|161:CT;427:GA
2	ENV022|barcode22|Sabin3-related|1|17:CT ENV024|barcode24|Sabin3-related|1|17:CT

d.

Pass	Sample	Barcode	Sabin1-related	Sabin2-related	Sabin3-related	WPV1	WPV2	WPV3	NonPolioEV	unmapped
✓	negative	barcode05	0	0	0	0	0	0	1	10
	positive	barcode05	0	0	0	0	0	0	45	0

e.



Polio Sequencing Consortium

Polio Sequencing Consortium

ENV001 report 2022-12-08

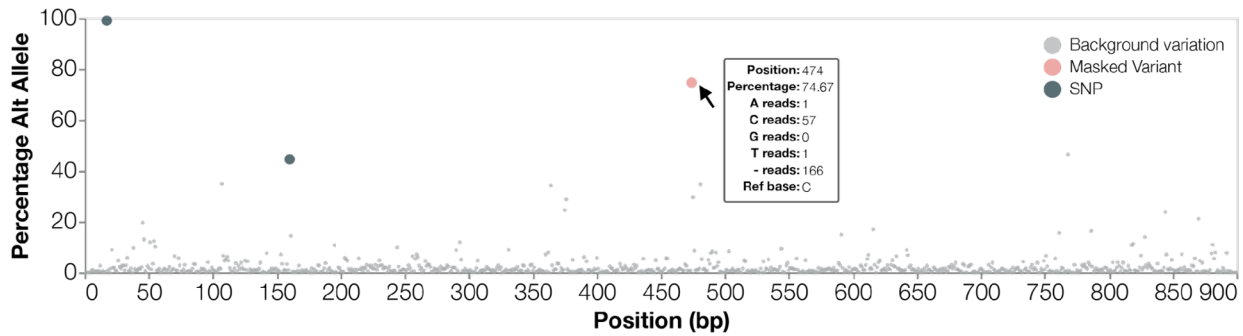
a.

Sample	Barcode	Reference group
ENV001	barcode01	Sabin3-related
ENV001	barcode01	Sabin2-related

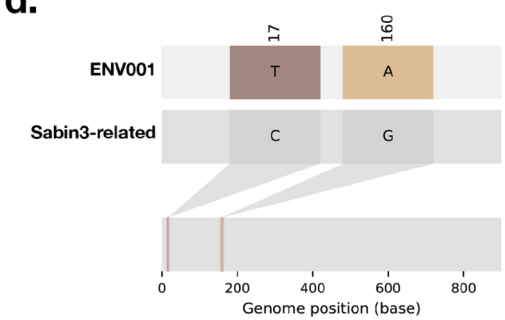
b. VP1 sequences

>ENV001|barcode01|Sabin3-related|Poliovirus3-Sabin_AY184221|2|17:CT;160:GA|date=2022-03-01
GGTATTGAAGATTTGATTTCTGAAGTTGCACAGGGCGCCCTAACTTTGTCACTCCCGAAGCAACAGGATAGCTTACCTGATACTAAGGCCAG
TGGCCCGGCGCATTCCAAGGAGGTACCTGCACTCACTGCAGTCGAGACTGGAGCCACCAATCCTCTGACACCATCCGACACAGTTCAAA
CGCGCCACGTAGTCCAACGACGCAGCAGGTCAGAGTCCACAATAGAATCATTCTTCGCACGCGGGGCGTGCGTCGCTATTATTGAGGTGG
ACAATGAACAACCAACCACCCGGGCACAGAAACTATTTGCCATGTGGCGCATTACATACAAAGATACAGTGCAGTTGCGCCGTAAGTTGGA

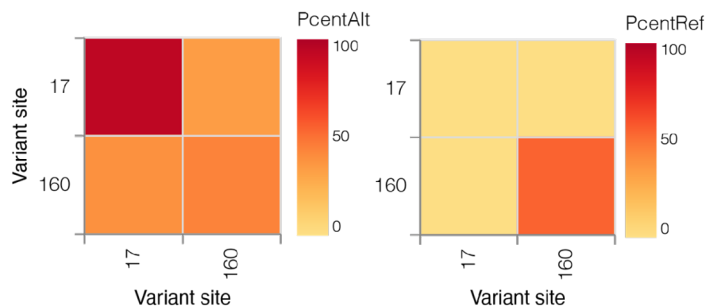
c.



d.



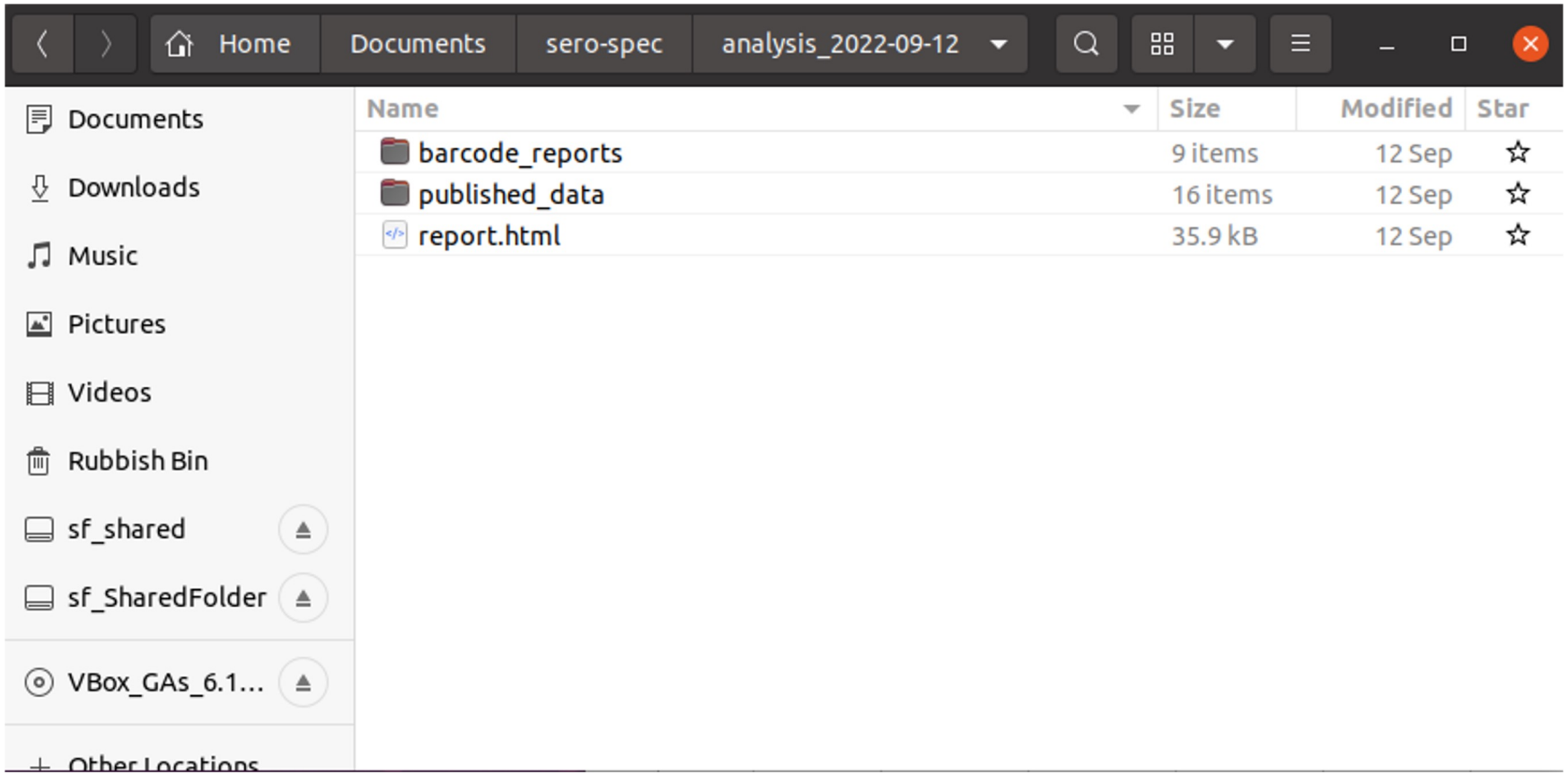
e.



Polio Sequencing Consortium

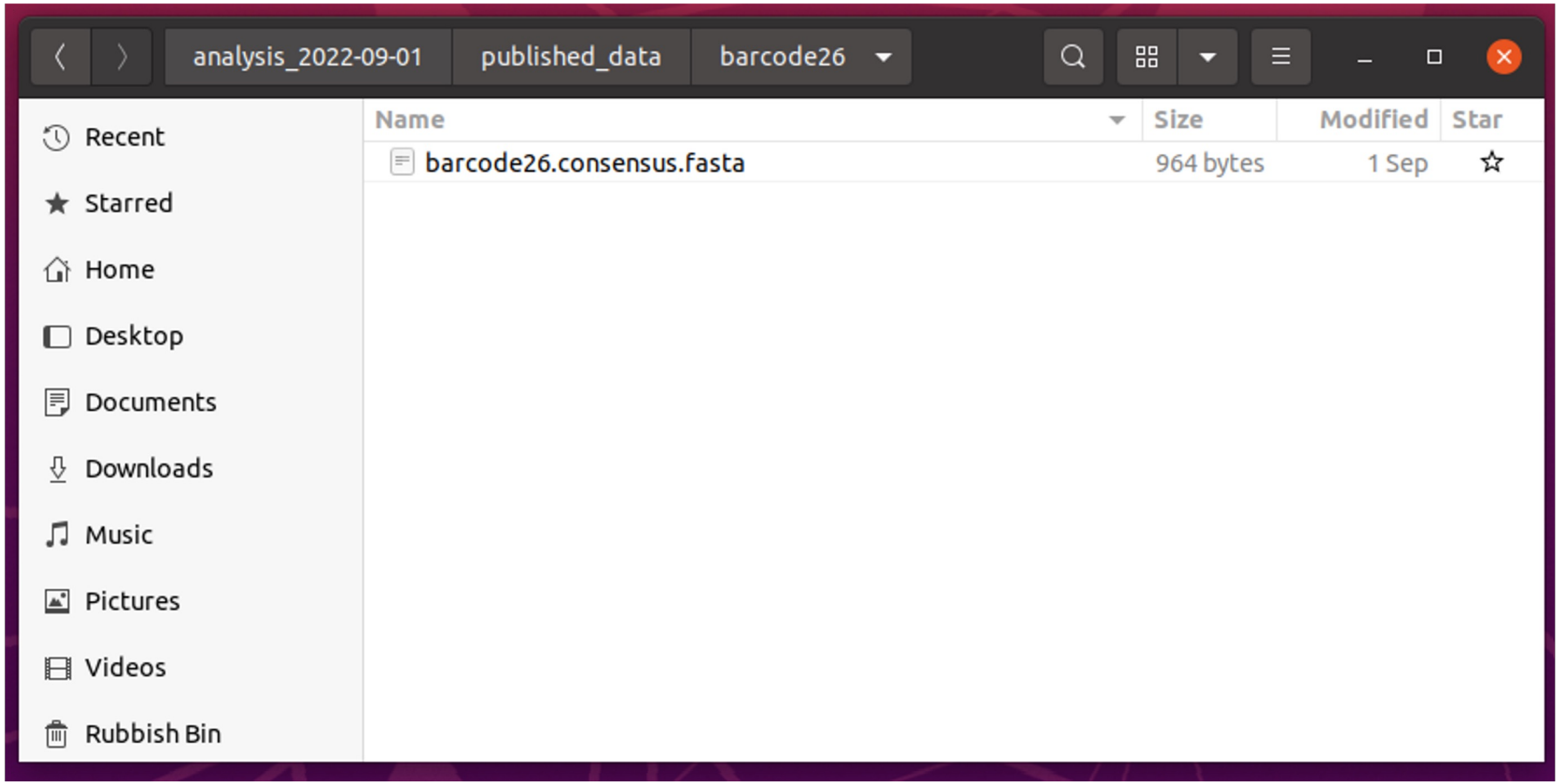
PIRANHA - output

- Outputs a summary report html, individual barcode reports, and consensus sequences as a fasta (both individual and in one file)



Polio Sequencing Consortium

PIRANHA fasta output



Polio Sequencing Consortium

Things to look out for

- Spaces

  Don’t put spaces in folder names

- Special characters

  Try to stick to using only ‘–’ and ‘_’ when naming samples, files, and directories: e.g., `run_2023-07-26_barcodes.csv`

- Capital letters

  Some bioinformatics software is case sensitive so be careful when writing commands and file paths

- Dates: always use the ISO 8601 date format: `2023-07-26`

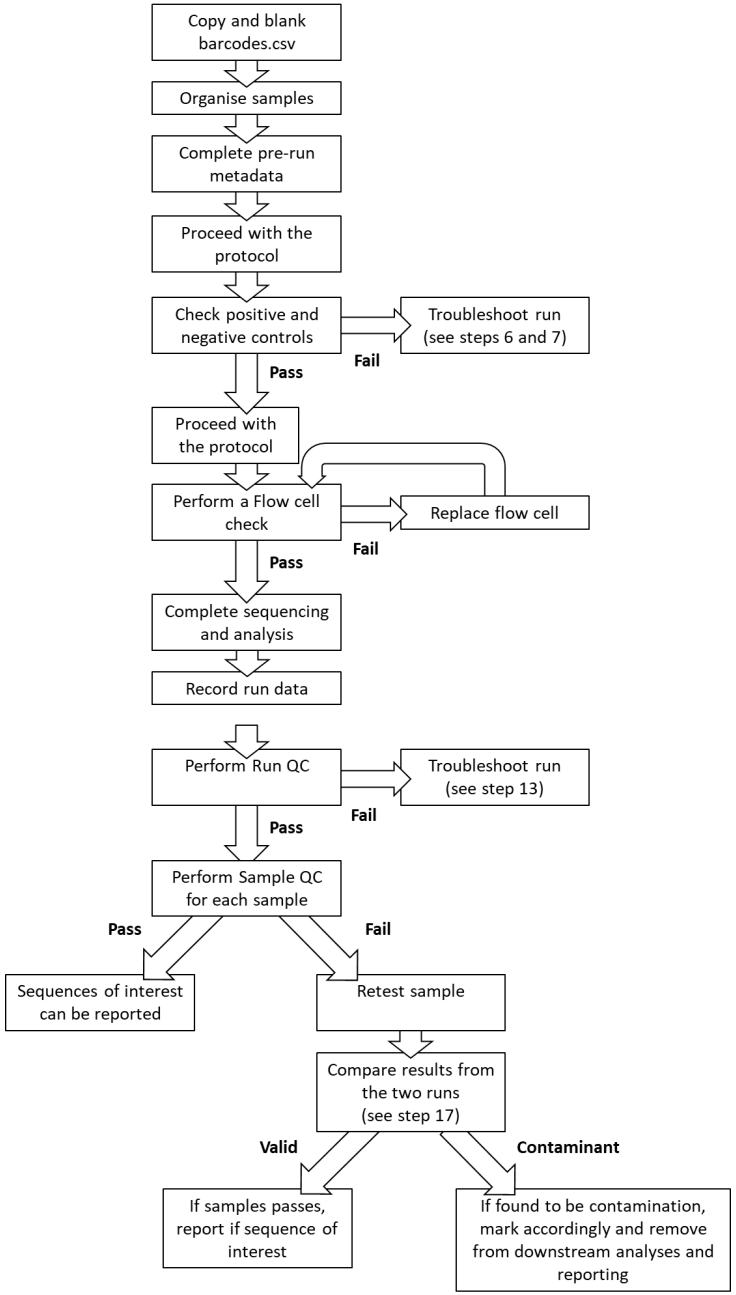
Polio Sequencing Consortium

7. Quality Control and Reporting

Lecture and discussion

Polio Sequencing Consortium

QC for NS of PV



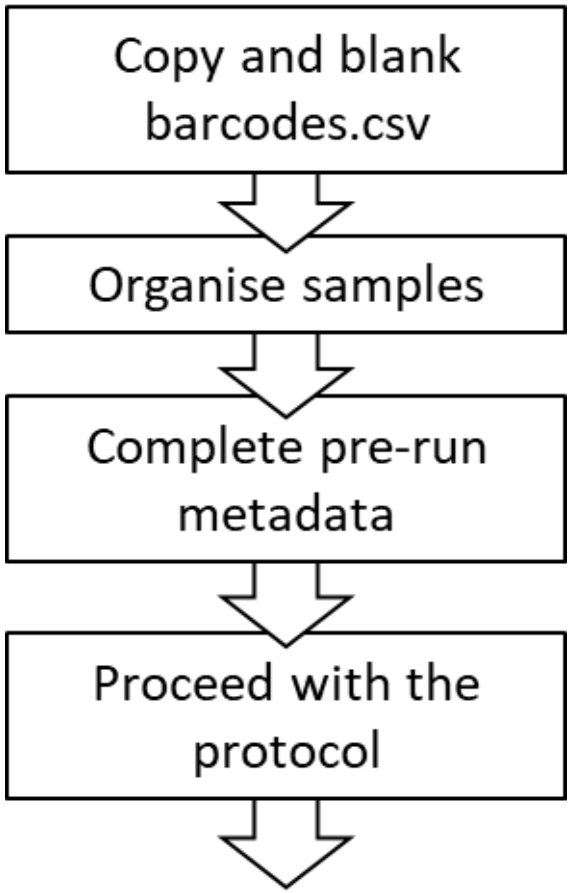
- QC documents in progress:

SOP_Metadata and QC v4.0.docx

- Feedback very welcome!
- Spreadsheet contains an example output file with QC data. This is built by PIRANHA using your barcodes.csv file.
- Word document explains the procedure in detail
- This QC routine is designed for routine direct testing of poliovirus from stool samples where the majority of samples will be negative. Other applications may require other QC thresholds (e.g. minimum pore numbers, run duration)

Polio Sequencing Consortium

Run setup



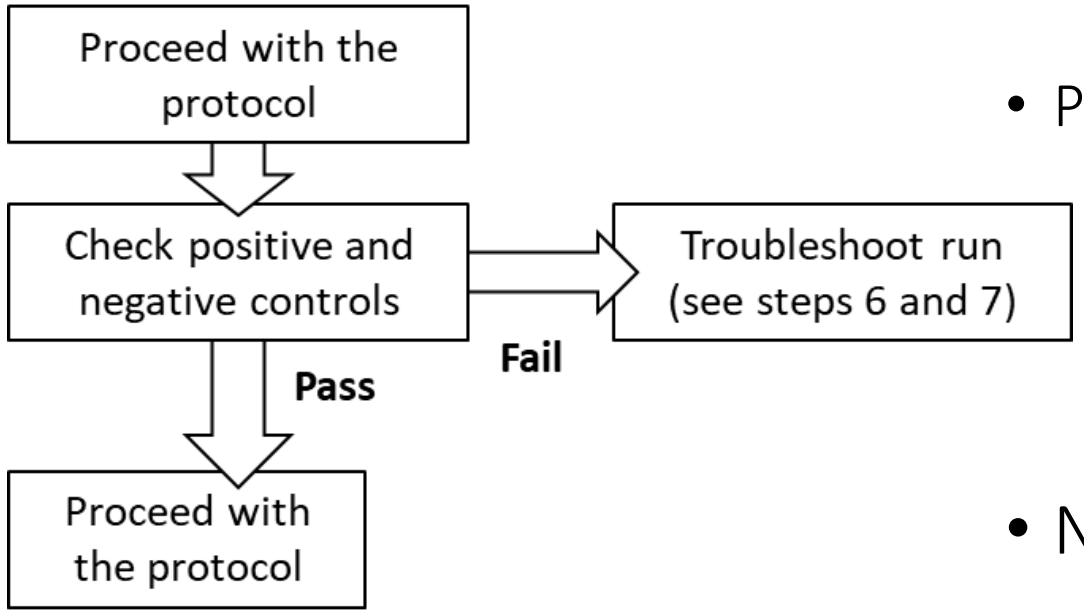
	A	B	C	D	E	F	G	H	I
1	sample	barcode	EPID	institute	...	Sabin1-related|closest_refere	Sabin1-related|num_reads	Sabin1-re	Sabin1-rel
2	sample01	barcode01	ARA-HIG-TOR-22-01				0		
3	sample02	barcode02	ARA-HIG-TOR-22-01			Poliovirus1-Sabin_AY184219	3556	1	99.89
4	sample03	barcode03	ARA-KAN-ERF-22-01			Poliovirus1-Sabin_AY184219	2243	1	99.89
5	sample04	barcode04	ARA-KAN-ERF-22-01			Poliovirus1-Sabin_AY184219	1453	0	100
6	sample05	barcode05	ARA-HIG-TOR-22-02			Poliovirus1-Sabin_AY184219	1278	0	100

(optional pre-run metadata in columns D,E, and can inset more)

- First columns are the barcode.csv file- more metadata can now be added to this file
- Aim to separate cases from the same region if possible- these viruses may be similar, making QC difficult.

Polio Sequencing Consortium

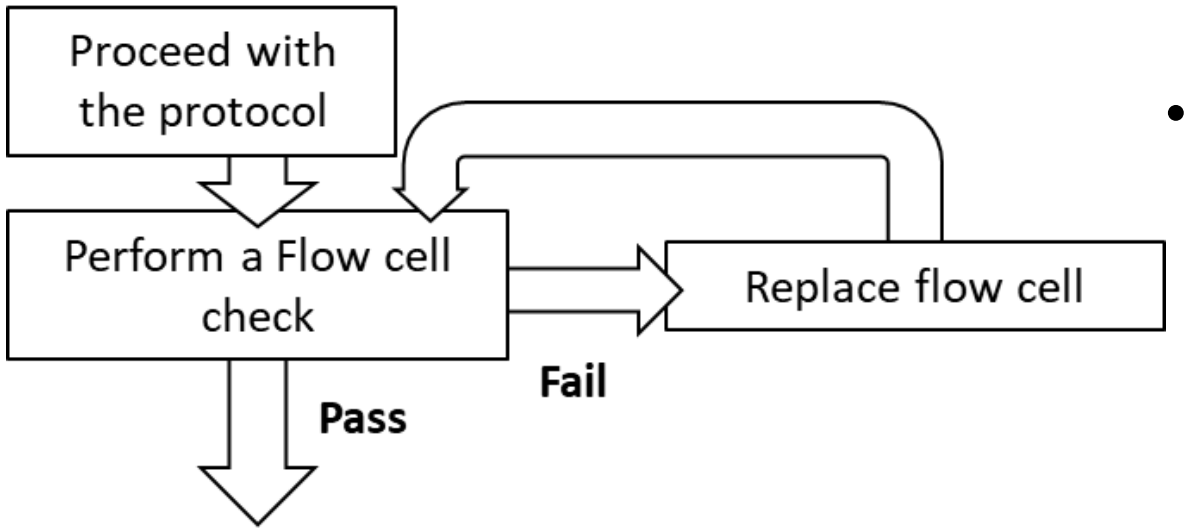
RNA extraction and nested RT-PCR QC



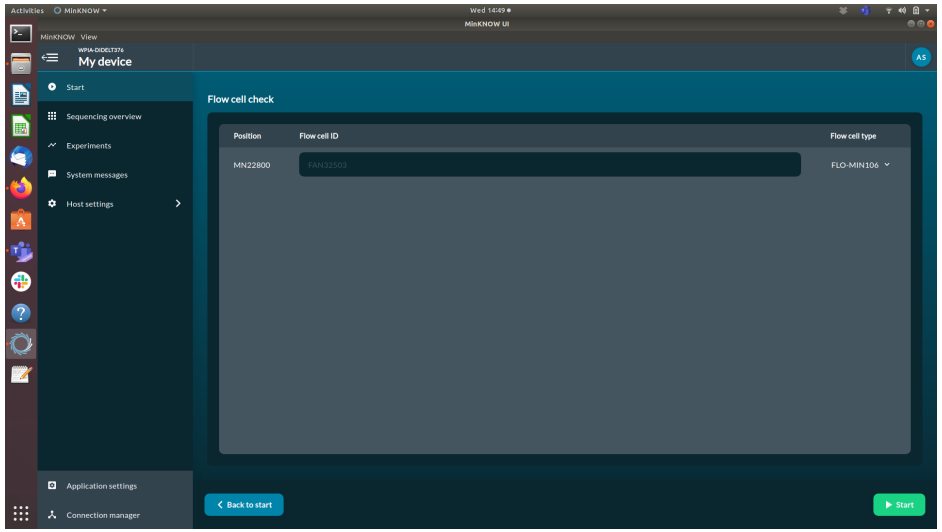
- Failed RT-PCR or failed PCR reactions
  - Note these in the barcodes.csv (they need to be repeated on later runs)
- Positive control
  - Two per day of RNA extractions performed (first and last extractions)
  - Should yield a ~1,400 bp amplicon on a gel or tapestation
- Negative control
  - Two per day of RNA extractions performed (first and last extractions)
  - Should have no amplicon.

Polio Sequencing Consortium

Flow cell check

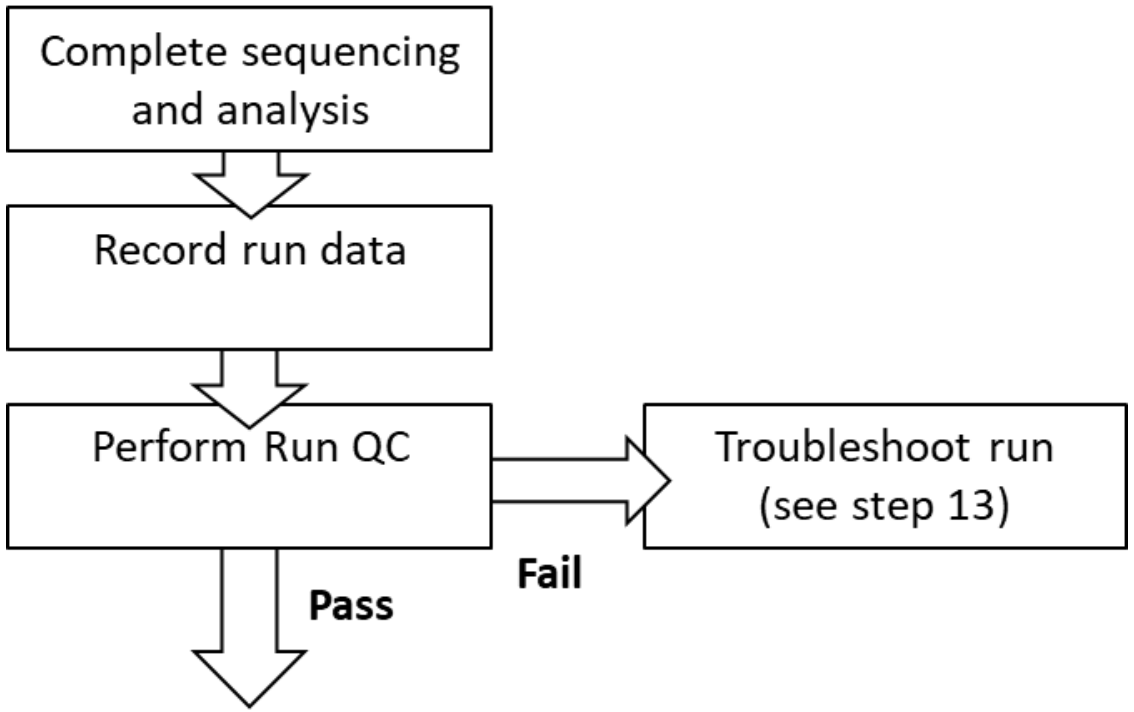


- Should be > 400 pores available for routine direct testing from stool.
- Record flow cell ID- can be useful to identify contamination between runs.



Polio Sequencing Consortium

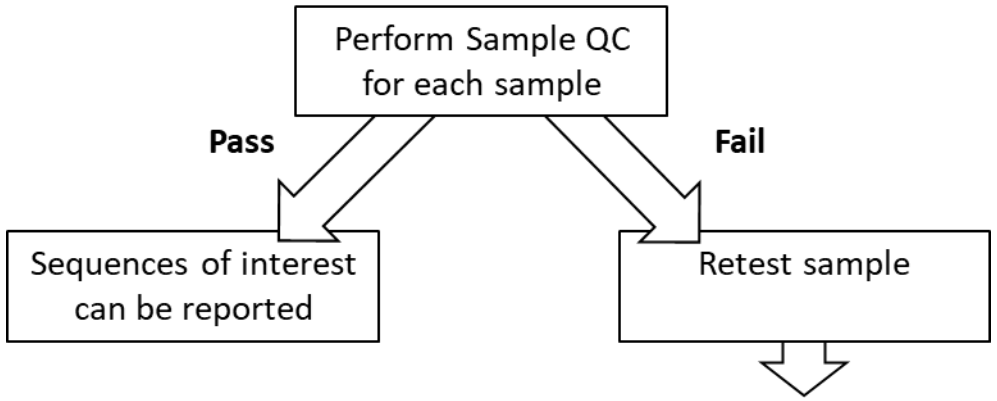
Sequencing Run QC



- PIRANHA will add sequencing and QC data to the barcodes.csv and save the file as a report.
- Check run QC
  - Did Positive control have > 500 reads mapping to A20?
  - Does the Negative control have <50 reads mapping to PV or NPEVs?
  - Did the run continue to completion and with at least 400 pores available (excluding shifts between pore groups)

Polio Sequencing Consortium

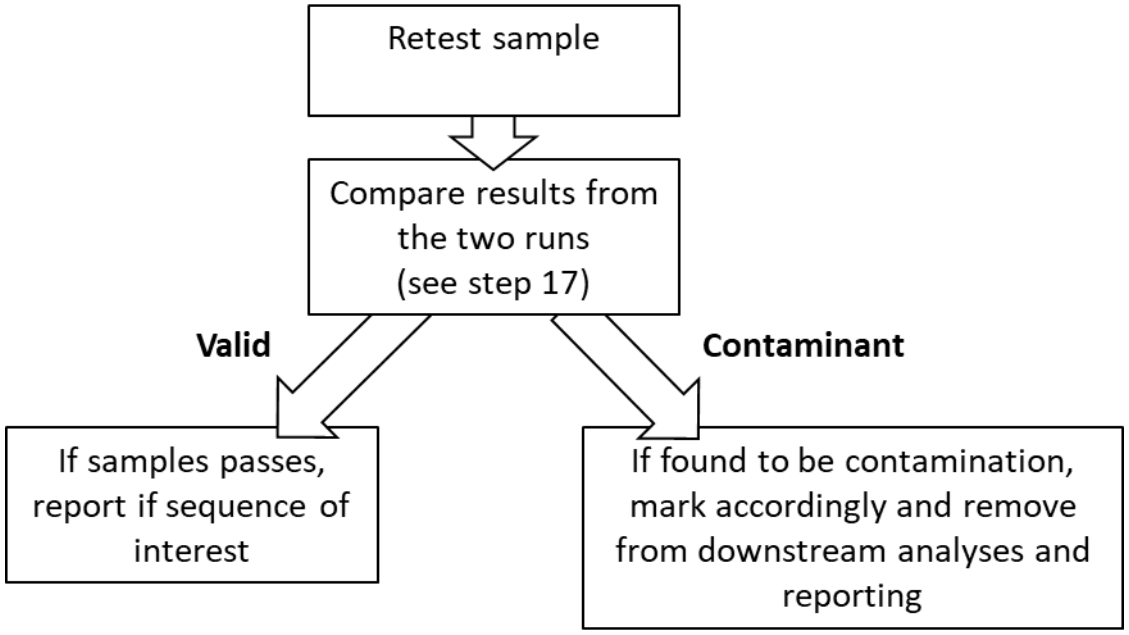
Sample QC



- If run passes, check samples where PV was detected. For VDPVs:
  - Samples with the same EPID (i.e. from the same case) should be no more than 2 nucleotides different from each other over VP1.
  - Samples adjacent to each other on the plate (on either axis) with different EPIDs should differ by >1 nucleotides over VP1.
  - Samples with <1000 reads should be >1 nucleotides different over VP1 from sequencing reads from previous runs that were generated using the same barcode and the same flow cell ID.

Polio Sequencing Consortium

Sample retesting



- Try to avoid using the same layout if checking a group of samples.
- Flag as a QC-check on the new sequencing run.
- Annotate the original run depending on the result- can report the sample if it passes the repeat.
- Talk to the appropriate person to arrange the retesting

Polio Sequencing Consortium

End