Quadram Institute Science Health Food Innovation

Nanopore data

Alp Aydin QIB

So you've sequenced your DNA – now what?

Basic DNA formats

fasta format

>sample1	**>Label**
ACGAGCTAGCTAGCTGATCGATGCTA	**DNA sequence**

fastq format

@sample1	**@Label**
CACGTAGCTAGCTAGCTGATCGATG	**DNA sequence**
+	**+**
><>==???24114==?;;=@@342@2	**Q scores**

MinKNOW data output

- MinKNOW will output fast5 data into “C:/data/” by default (in a folder with your run name)
- If you turn on live basecalling, it will output both fast5 and fastq files
- MinKNOW temporarily creates .raw files as it is sequencing before transferring to fast5
- If a run is stopped early, let MinKNOW run to transfer .raw files to fast5
- If not, you will need to run the recover_reads command manually

MinKNOW data options

- Live basecalling (both configs)
- Demultiplexing
- Trimming barcodes
- Basecalling existing data
- Compression options (gzip) – many software actually take this format directly

Data size

- 100 gigabases of data
- fast5 = 2 terrabytes
- fast5 gzipped = 1.1 TB

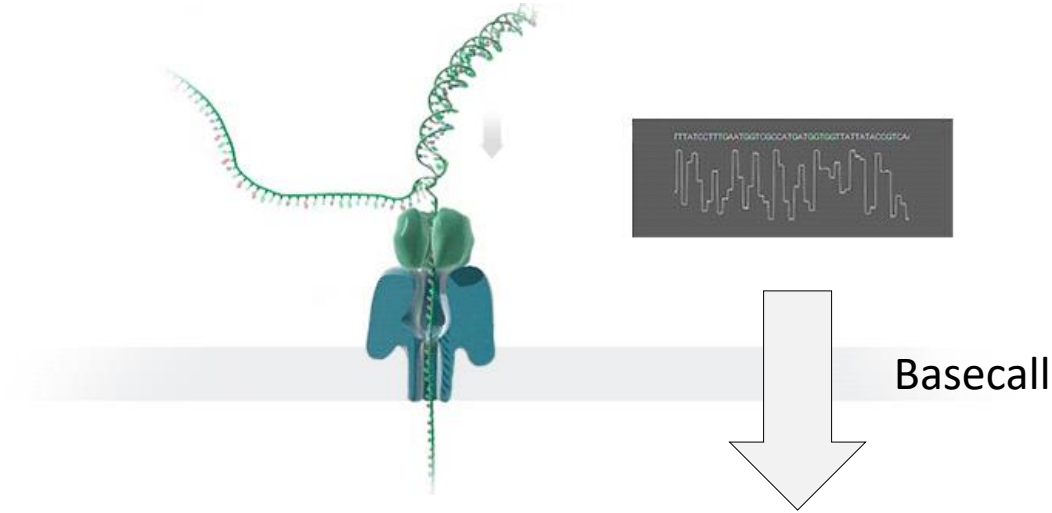
- fastq = 200 gigabytes
- fastq gzipped = 112 gigabytes

- Raw data extremely large but may be useful
- Re-basecall with new updates and features
- E.g. base modification (methylation)

- Own decision on what to keep

Basecalling

Raw data to standard readable DNA sequence format



Fast5

Basecall

AGCTGTGCTAGCTAGCTAG
+ quality of sequence

Fastq

Basecalling

- **Fast and high accuracy options**

Sample Type	Model Name	Raw Read Accuracy	Basecalling Speed on MinIT
DNA	Fast basecalling	92.1%	~166k bases/s
	High-accuracy basecalling	95.0%	~22k bases/s
RNA	High-accuracy basecalling	93.9%	~23k bases/s

De-multiplexing

CACAAAGACACCGACAACTTTCTTGCGCAGCATGCTACTGAGTCGTATTACTG…..

ACAGACGACTACAAACGGAATCGATCTTATCAGGGGTCTATCTCTCGCGAGCATGCGA…..

CACAAAGACACCGACAACTTTCTTCGAGTCGATCGATCGATCGTAGCTAGCTA…..

ACAGACGACTACAAACGGAATCGAGGTGGGTGTATATCTGCTAGCTAGCTATGCTAGCTAGCAGCT…..

CACAAAGACACCGACAACTTTCTTATCGTAGTATATATTATATATATGCTAGCTAGCAGCT…..

De-multiplexing

Barcode 1 sequences

CACAAAGACACCGACAACTTTCTTGCGCAGCATGCTACTGAGTCGTATTACTG…..

CACAAAGACACCGACAACTTTCTTATCGTAGTATATATTATATATATGCTAGCTAGCAGCT…..

CACAAAGACACCGACAACTTTCTTCGAGTCGATCGATCGATCGTAGCTAGCTA…..

Barcode 2 sequences

ACAGACGACTACAAACGGAATCGAGGTGGGTGTATATCTGCTAGCTAGCTATGCTAGCTAGCAGCT…..

ACAGACGACTACAAACGGAATCGATCTTATCAGGGGTCTATCTCTCGCGAGCATGCGA…..

Tools: Guppy or EPI2ME or qcat

Trimming

Barcode 1 sequences

CACAAAGACACCGACAACTTTCTTGCGCAGCATGCTACTGAGTCGTATTACTG…..

CACAAAGACACCGACAACTTTCTTATCGTAGTATATATTATATATATGCTAGCTAGCAGCT…..

GCGCAGCATGCTACTGAGTCGTATTACTG…..

ATCGTAGTATATATTATATATATGCTAGCTAGCAGCT…..

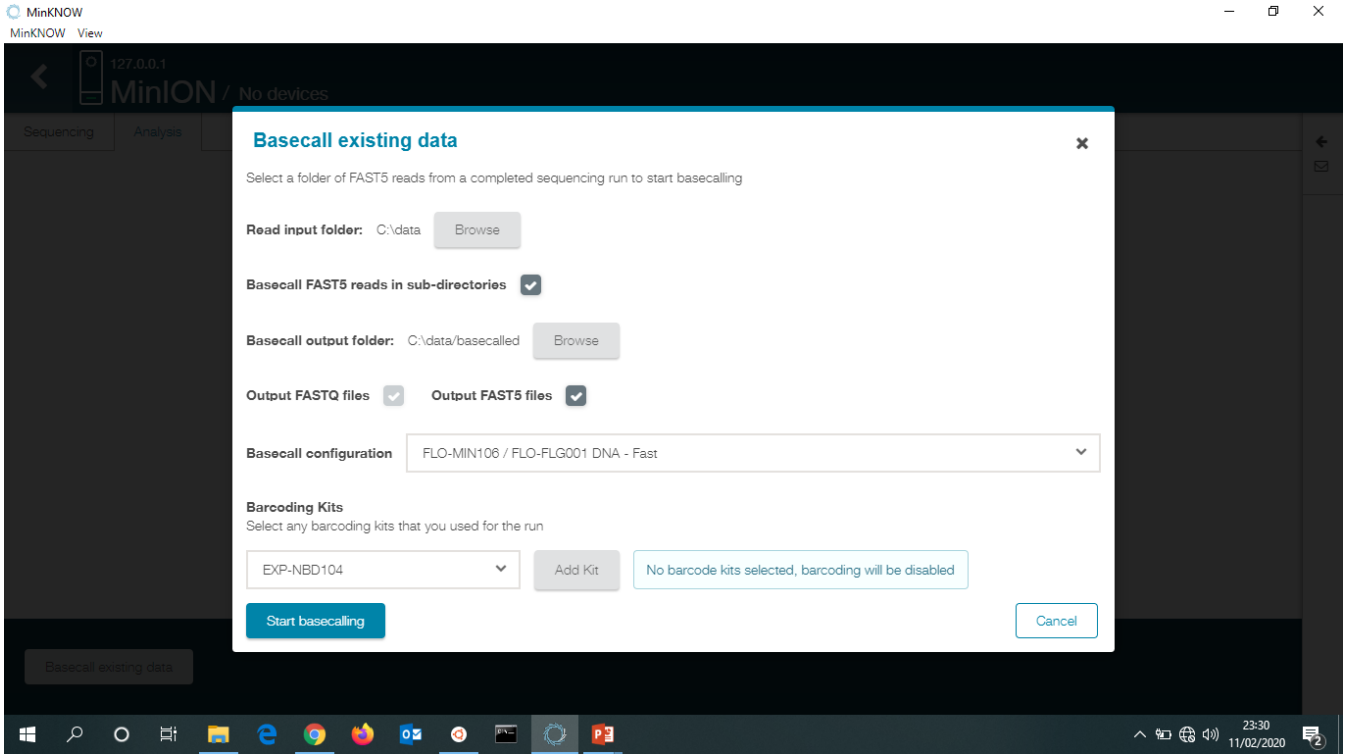
Guppy (standalone)

Example command:

```
"C:\Program Files\OxfordNanopore\ont-guppy-cpu\bin\guppy_basecaller.exe" --input_path C:\my_folder\reads --save_path C:\output_folder\basecall --config dna_r9.4.1450bps_fast.cfg
```

Q-score filtering
Demultiplex
Trimming
Compress
Resume option

Life is easier now



Computer requirements

Minimum for sequencing

Component	Configuration		
Operating Sytem	Windows	Mac	Linux
	7, 8, 10 (64 bit)	Yosemite El Capitain Sierra (64 bit)	Ubuntu 14.4*
Memory/RAM:	16GB RAM		
CPU	i7 or Xeon**		
Storage	1TB SSD		
Ports	USB3***		

** Linux products are offered under limited support and may take the team longer to respond to queries*

*** Users need to verify their i7 is a four core model or better.*

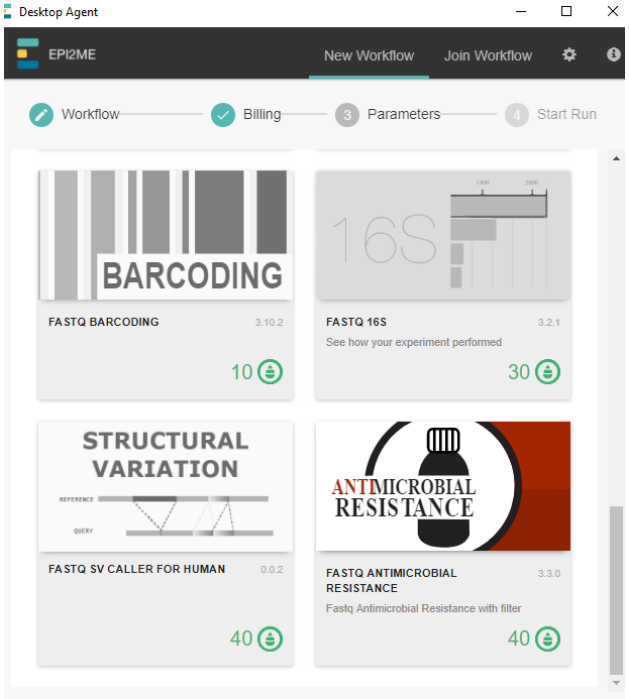
**** The MinION device is CE marked using USB3. If a user wished to use USBC they may but this invalidates the CE marking*

MinIT and Mk1c

1 TB

EPI2ME

ONT’s cloud-based analysis platform. Command-free

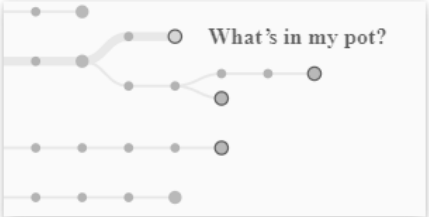


Metagenomic workflows

FASTQ 16S 3.2.2

Taxonomic assignment for 16S amplicons

30



FASTQ WIMP 3.2.2

20

FASTQ ANTIMICROBIAL RESISTANCE 3.3.2

Fastq Antimicrobial Resistance with filter

40