Introduction to ONT sequencing

Dr John Tembo
HerpeZ

@PandoraIDNet @AmrCovid

We will record these sessions and put them online so you can refer back to them later on

We will also put the slides up online so you can access the notes (links and image credits)

What is sequencing?

Uses and Sanger sequencing

Next generation sequencing

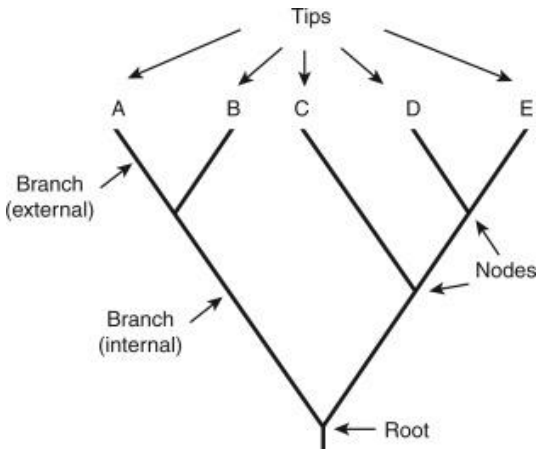
Oxford Nanopore Technologies

What is sequencing?

- Sequencing determines the order of the four bases (A, T, G and C) that make up DNA
- The first DNA sequences were obtained in the early 1970s
- Requires both laboratory and bioinformatics skills!

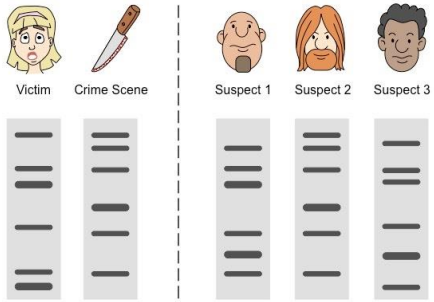
There are many uses for sequencing

Identifying species in a sample (e.g. sputum or waste water)



Seeing how closely related a group of organisms are (phylogenetic trees, evolution)

Identifying differences in genes (e.g. genetic diseases)



Forensics (e.g. DNA profiling, paternity testing etc.)

Look at whether bacteria is resistant to antibiotics

The Human Genome Project

- Began October 1990 and completed April 2003
- Researchers aimed to decipher the human genome in three major ways:
    - Determining the sequence of all the bases in the human genome
    - Mapping the locations of genes for major sections of our chromosomes
    - Producing linkage maps, so inherited traits (e.g. for genetic disease) can be tracked over generations
- The sequence is derived from the DNA of several volunteers
- Cost $2.7 billion!
- Mainly used Sanger sequencing technology

Introduction to Sanger sequencing

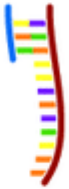
- The DNA sample is divided into four separate reactions, containing all four of the standard deoxynucleotides (dATP, dGTP, dCTP and dTTP) and a DNA polymerase (which attached the dNTPs)
- To each reaction is added only one dideoxynucleotides (ddATP, ddGTP, ddCTP, or ddTTP)
- Four separate reactions are needed in this process to test all four ddNTPs
- The ddNTP stops the DNA polymerase when it comes to a base of that type (e.g. A, T, G, C)
- The fragments are then run on a gel. The smallest move through the gel furthest and the ‘ladder’ shows the sequence of the DNA

TB

1
denature dsDNA using heat

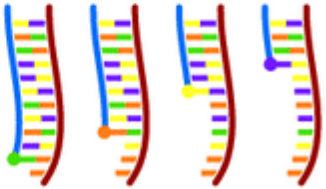
2
make multiple copies of a segment

3
attach a primer



4
add to four polymerase solutions

5
grow complementary chains until termination dye



6
denaturate the grown chains

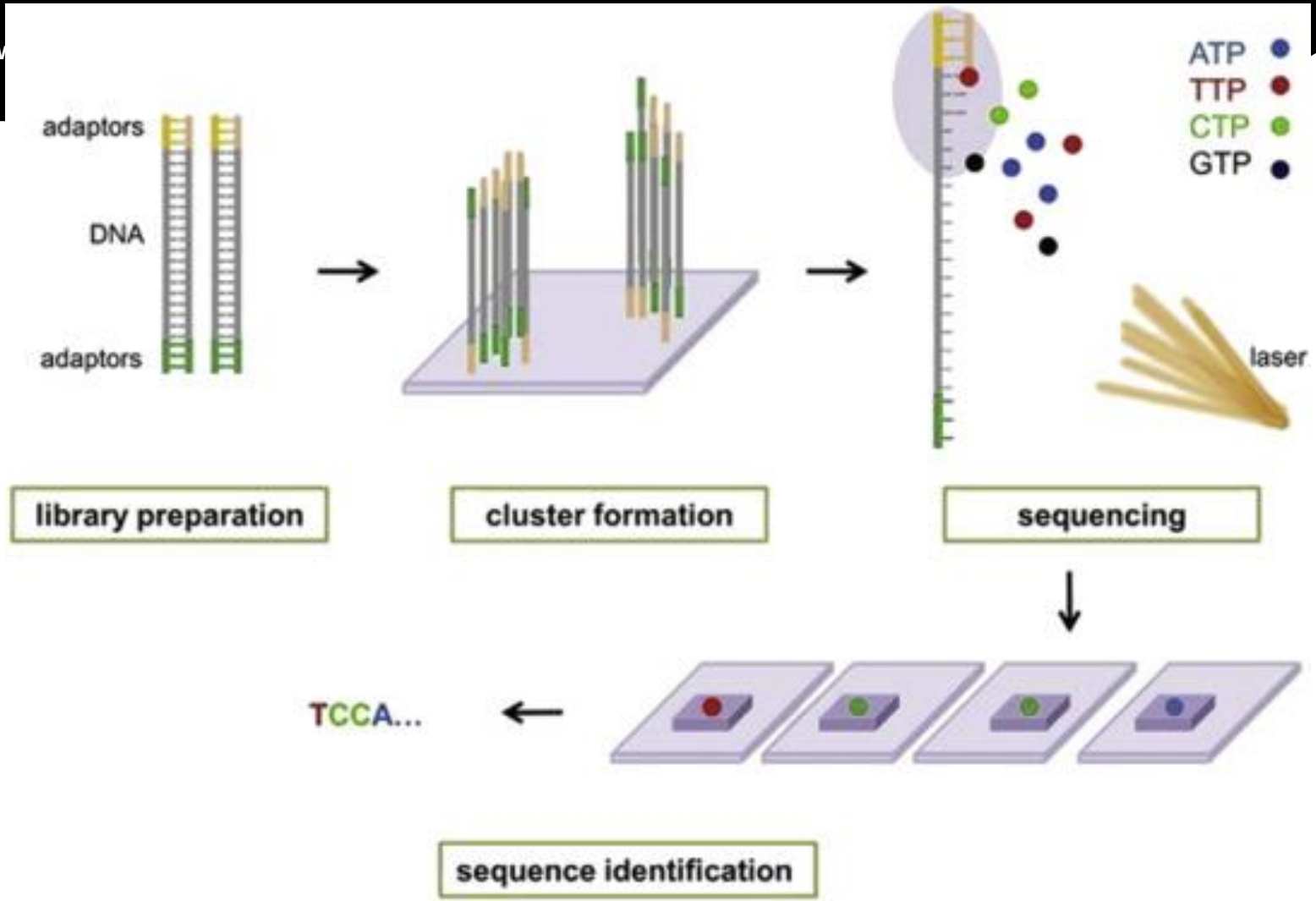
7
electrophorese the four solutions

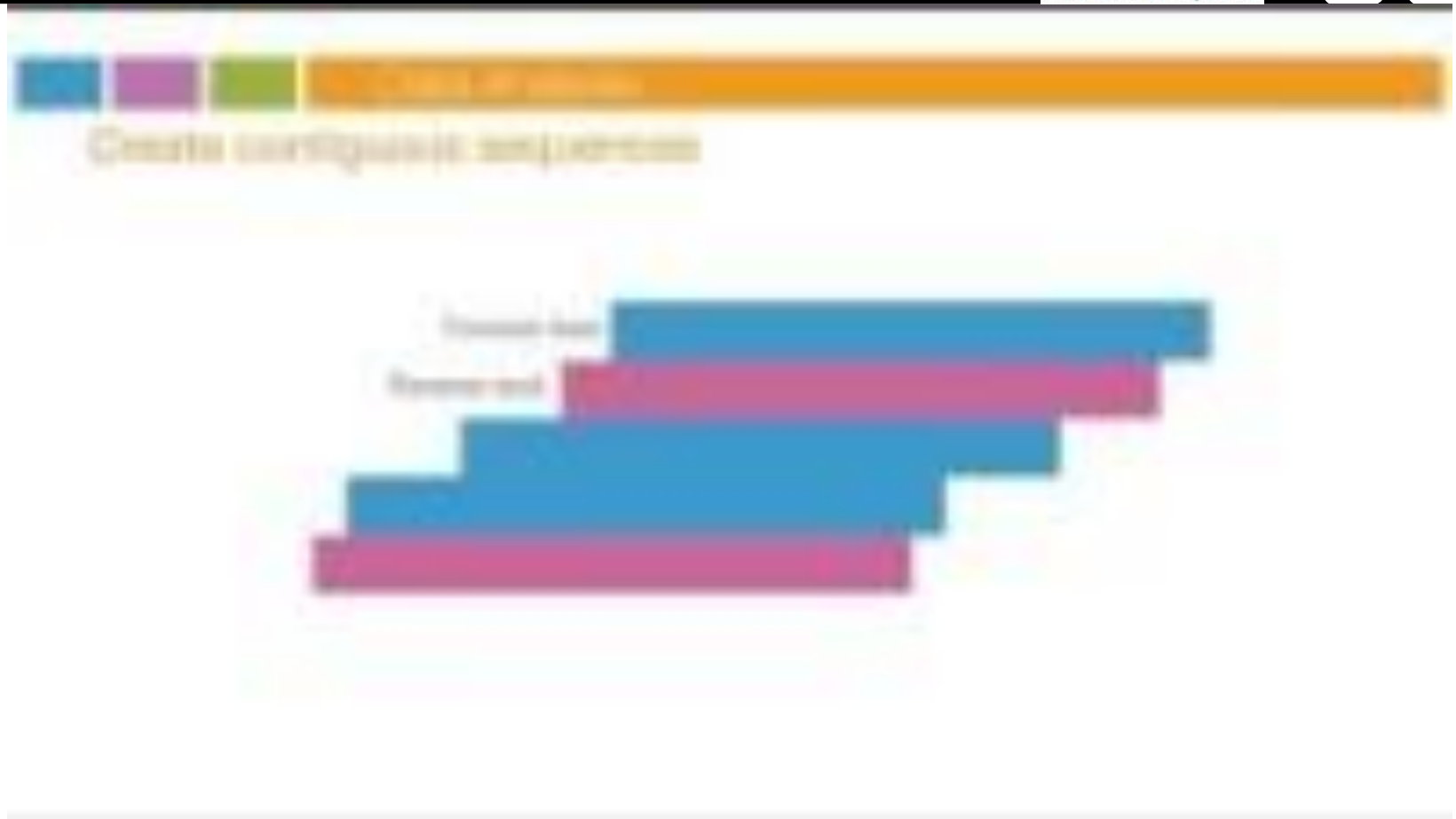
Next Generation Sequencing technology

- The Sanger method only sequences a single DNA fragment at a time
- NGS is massively parallel, sequencing millions of fragments simultaneously per run
- NGS platforms include (but are not limited to):
  - Illumina (short read) sequencing
  - Oxford Nanopore Technologies (long read) sequencing

Introduction to Illumina sequencing

- Sequencing by synthesis
- ‘Short read’ technology – DNA is cut up into 200-600 bp chunks
- The DNA is amplified, so there are lots of copies of the chunks
- They are denatured, then fluorescent complimentary bases are attached
- These fluoresce different colours, which is recorded, and the sequence is identified


adaptors
DNA
adaptors
ATP
TTP
CTP
GTP
laser
library preparation
cluster formation
sequencing
TCCA...
sequence identification

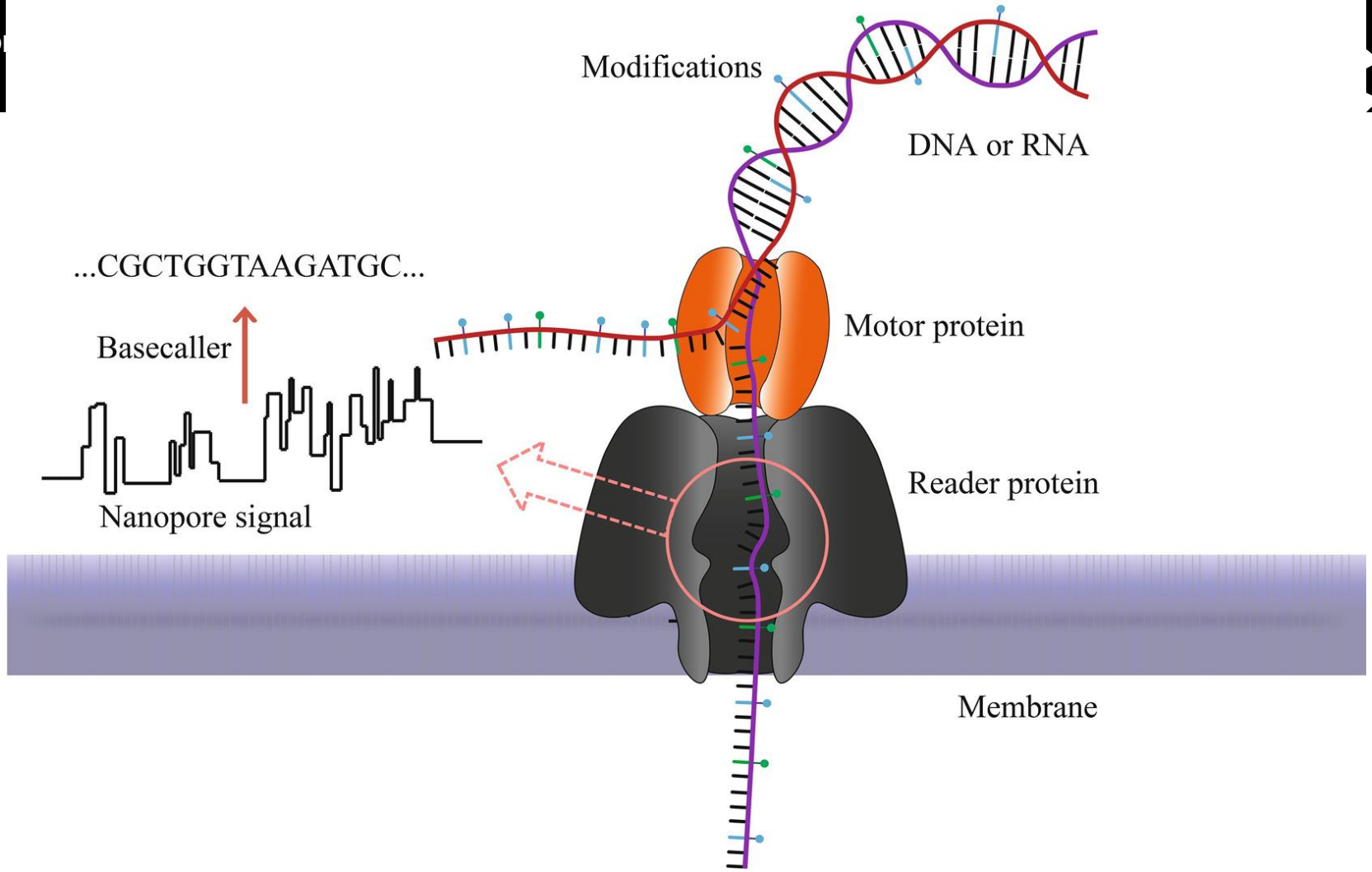


Oxford Nanopore Technologies

- To enable the analysis of any living thing, by any person, in any environment
- Shown to work in -5°C in Antarctica
- Kabobo, Democratic Republic of Congo – biodiversity (high humidity, equatorial conditions)
- Study of geothermal microbes in Iceland sequenced in a tent using only solar power (entirely ‘off-grid’)

Introduction to Oxford Nanopore Technologies

- ‘Long read’ technology
- A DNA library is prepared (proteins are added)
- Nucleic acids are passed through a protein nanopore
- As the different bases move through the nanopore, it creates a different electrical signal
- These resulting changes in the electrical signal is decoded to provide the specific DNA or RNA sequence


Modifications
DNA or RNA
...CGCTGGTAAGATGC...
Basecaller
Motor protein
Nanopore signal
Reader protein
Membrane

Oxford Nanopore Technologies

- Different ‘library preparation’ kits and sequencing platforms can be used to sequence different things

Oxford Nanopore Technologies

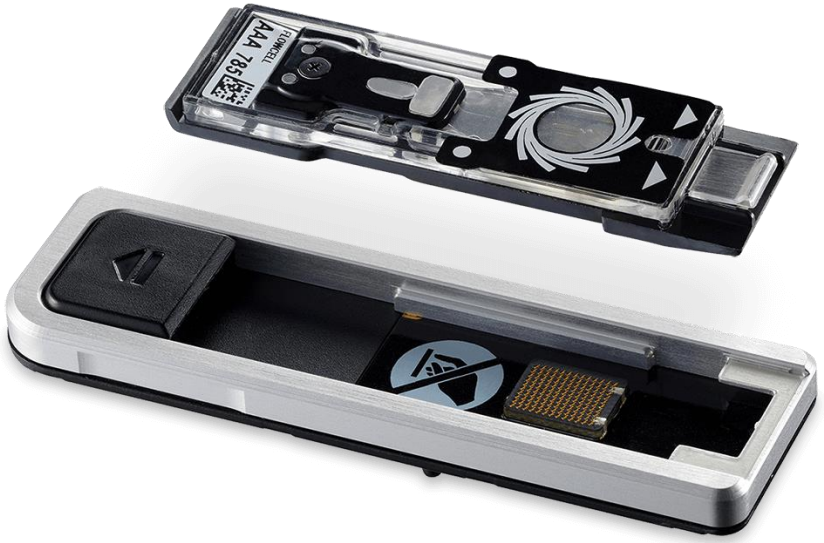
- MinION

Oxford Nanopore Technologies

- Mk1C

Oxford Nanopore Technologies

- Flongle (technically a specialised flow cell for a MinION)



Oxford Nanopore Technologies

- GridION

Oxford Nanopore Technologies

- PromethION

Oxford Nanopore Technologies

Device	No. flow cells per run	Throughput	Theoretical maximum output	Cost
Flongle	1	126 channels	2.8 Gb	From $90 (per flow cell)
MinION	1	512 channels	50 Gb	From $1,000 (starter pack)
Mk1C	1	512 channels	50 Gb	From $4,000 (starter pack)
GridION	5	5 x 512 channels	250 Gb	From $49,000 (starter pack)
PromethION	24 or 48	10,700+ channels	2,596+ Gb	From $100,455 (starter pack)

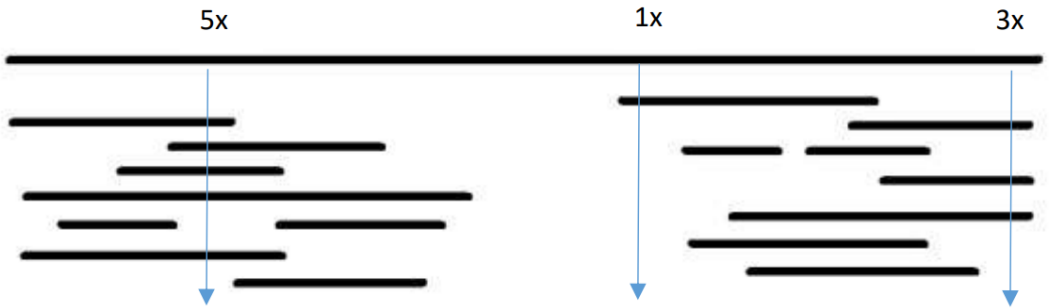
Why is getting so much information so important?

Depth vs coverage

Depth and coverage are both very important when it comes to sequencing, but they mean different things.

Depth

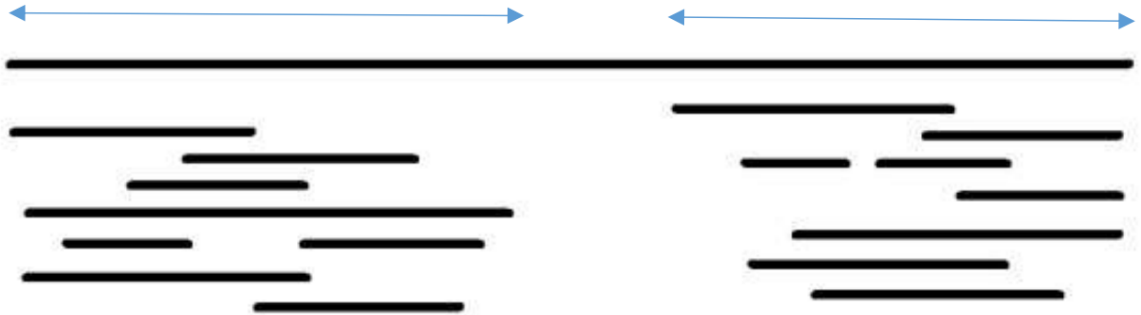
- This is the amount of times a base within a genome has been sequenced
- The greater the depth, the greater the confidence in the identity of the sequenced base
- In the image below, the complete reference genome is at the top. Below it are the sequenced contigs (a series of overlapping DNA fragments). Three bases are represented, which have varying depth of reads.



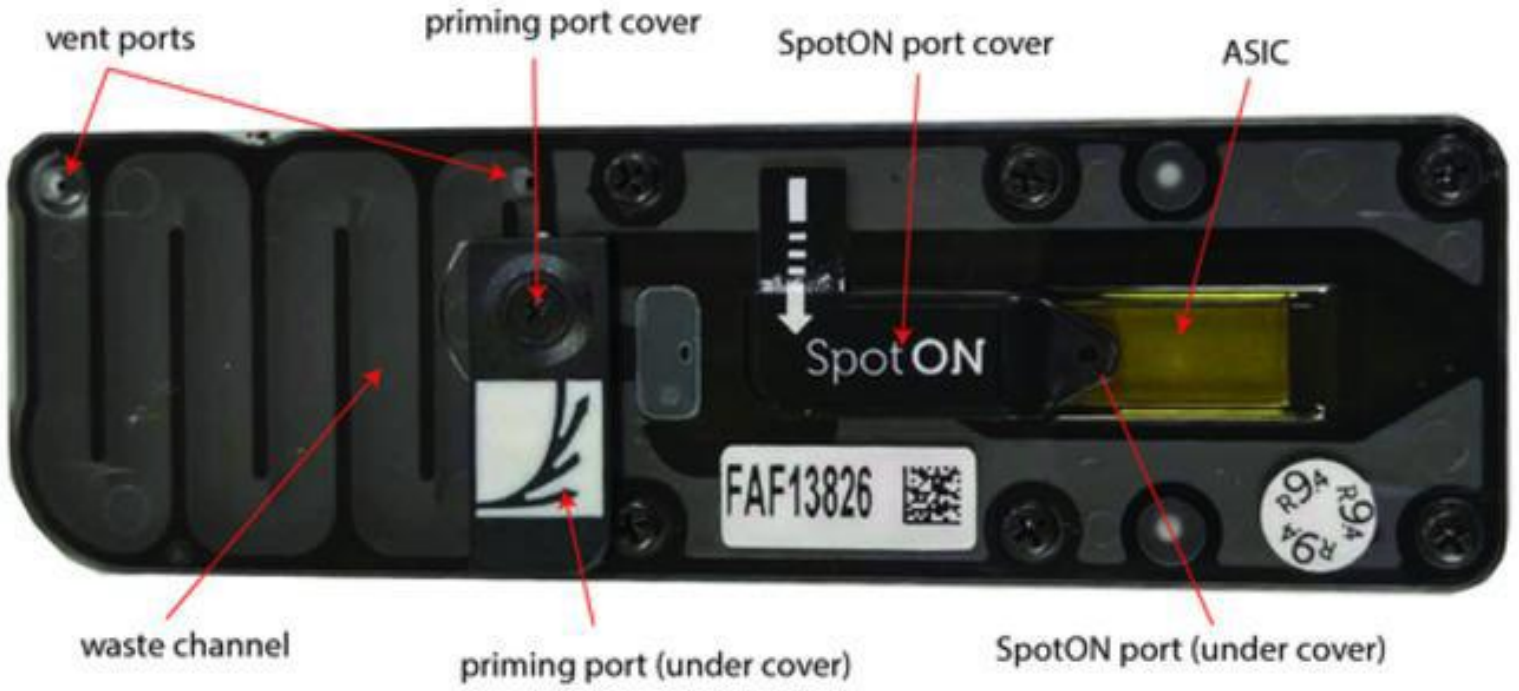
Depth vs coverage

Coverage

- The percentage of the whole genome that has been sequenced.
- For instance, in the example below, the sequenced contigs cover approximately 80% of the reference genome (at the top of the image), which means you would have sequenced 80% of the bases that make up the genome.
- You will not be able to identify genes, SNPs etc. in parts of the genome that you do not have coverage for.



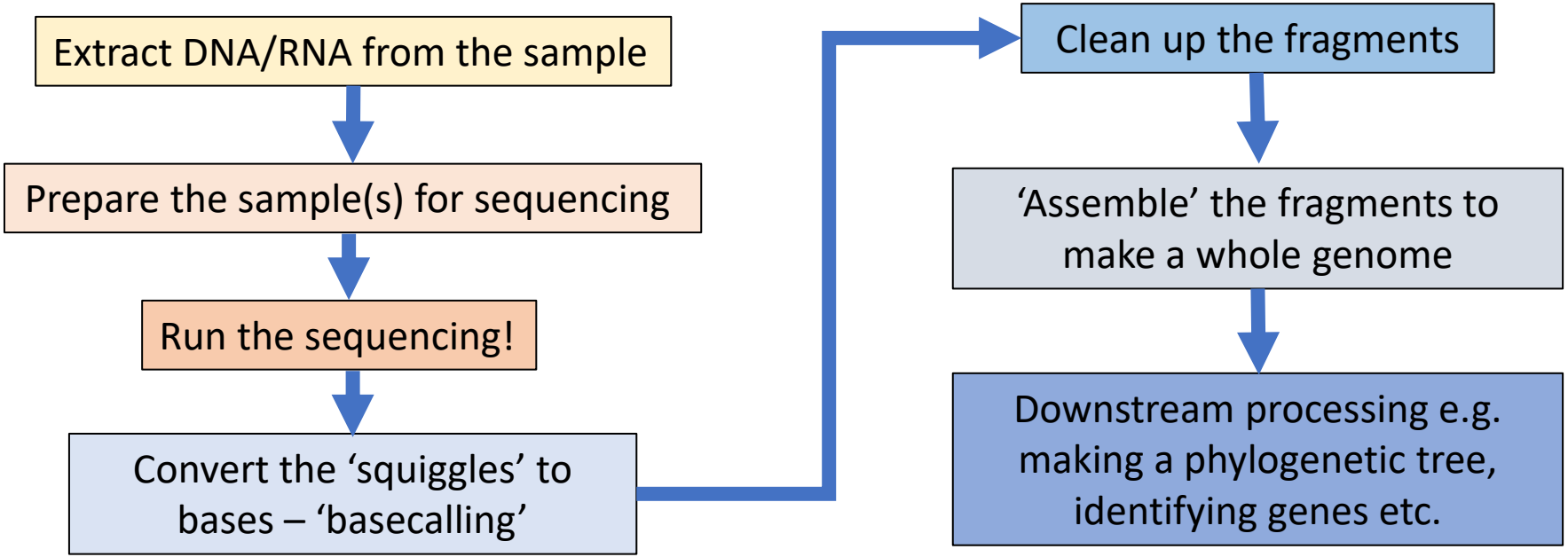
Anatomy of a Flow Cell



Different kits for ONT

- Rapid barcoding kit
- Ligation kit
- PCR based kits
- Automated kit prep

ONT basic workflow



Minimum computer requirements

Purchase the highest specification you can afford, because sequencing is constantly improving and outputting more and more data!

Component	Required specification
Operating system	Windows – 10, Linux – Ubuntu 20.04 and 18.04, OSX – Mojave, Catalina
Memory/RAM	16 GB RAM or higher
CPU	Intel i7, i9, Xeon, or better, with at least 4 cores/8 threads Ryzen 5, 7, or better, with at least 4 cores/8 threads
GPU	NVIDIA GPU RTX 2060 SUPER or better, with at least 8 GB of GPU memory
Storage	1 TB internal SSD or higher
Ports	USB3.0

Any questions?