

NGS - quality control, alignment, visualisation

Sequencing technologies

Major applications

- Transcriptome characterization
 - e.g. RNA-seq
- Epigenome characterization:
 - e.g. ATAC-seq
- DNA-protein interactions:
 - e.g. ChIP-seq
- Whole genome (assembly)
- Variant detection
- Metagenome characterization
- Any others?



Sequencing



Quality control



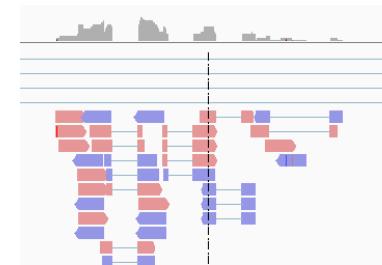
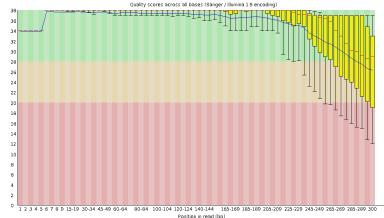
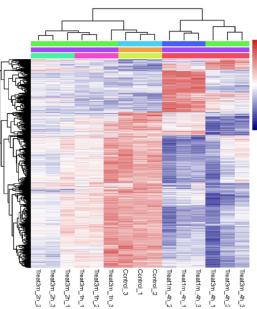
Alignment



Down-stream analysis



Visualisation



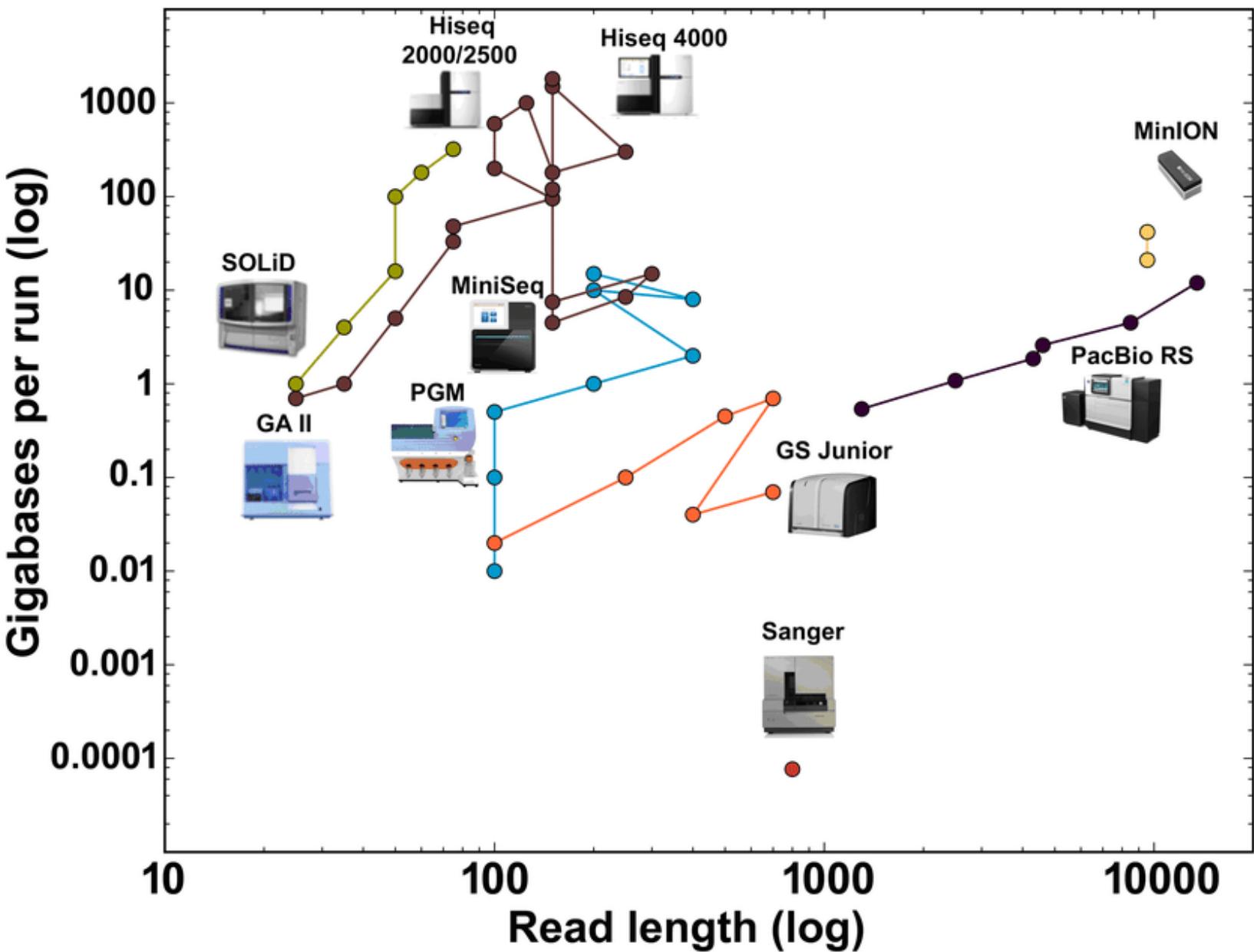


Image from: G. Silva (2016)

This course

- 2nd generation:
 - Illumina
- 3rd generation:
 - Pacific Biosciences
 - Oxford Nanopore Technology

Quiz Question 1

Illumina sequencing

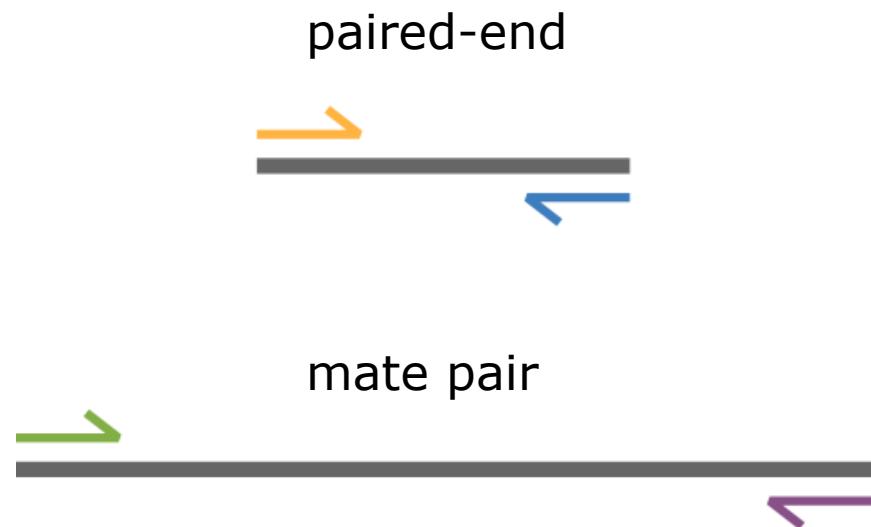
- Sequencing-by-synthesis: 2nd generation sequencing
- Massive throughput: up to 500×10^9 bases/run
- Most used platform today

illumina®

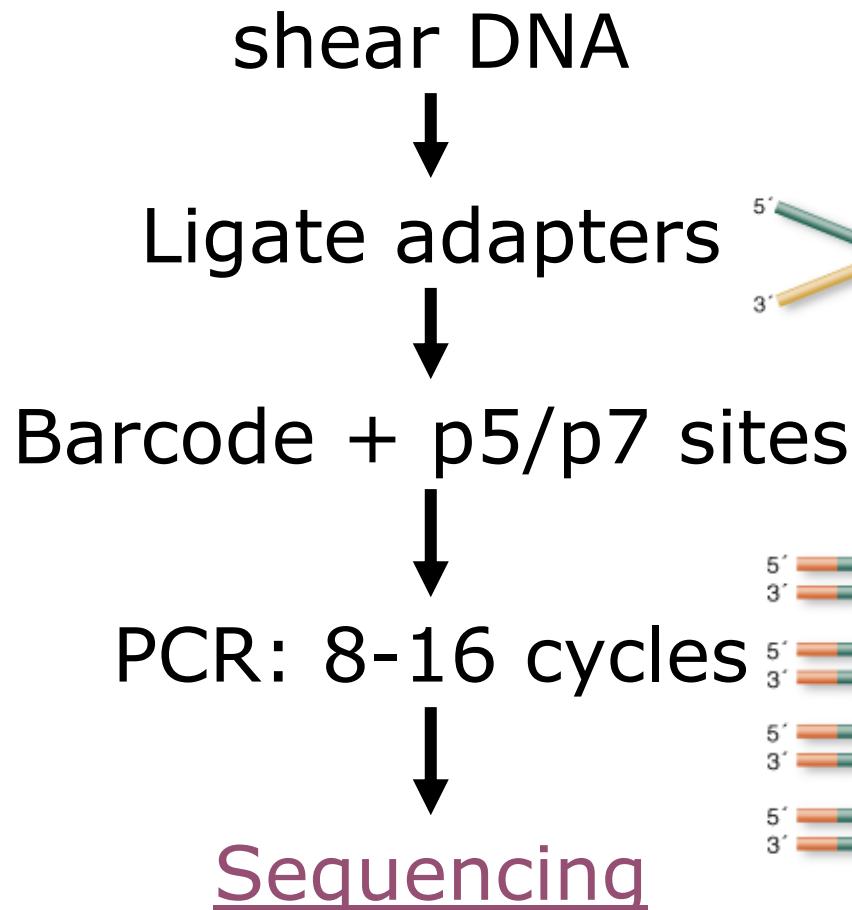


Illumina sequencing

- 50 – 300 bp
- Paired-end (or single-end)
- Multiplexing



Illumina library prep



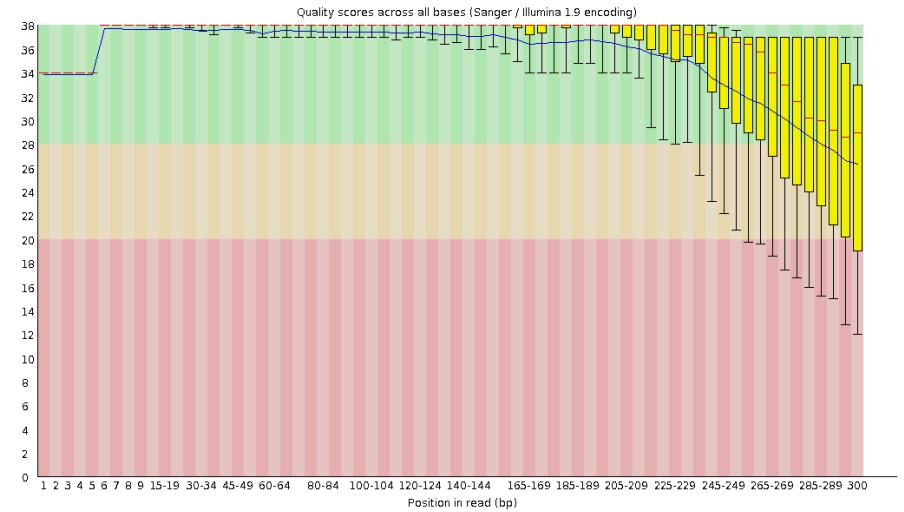
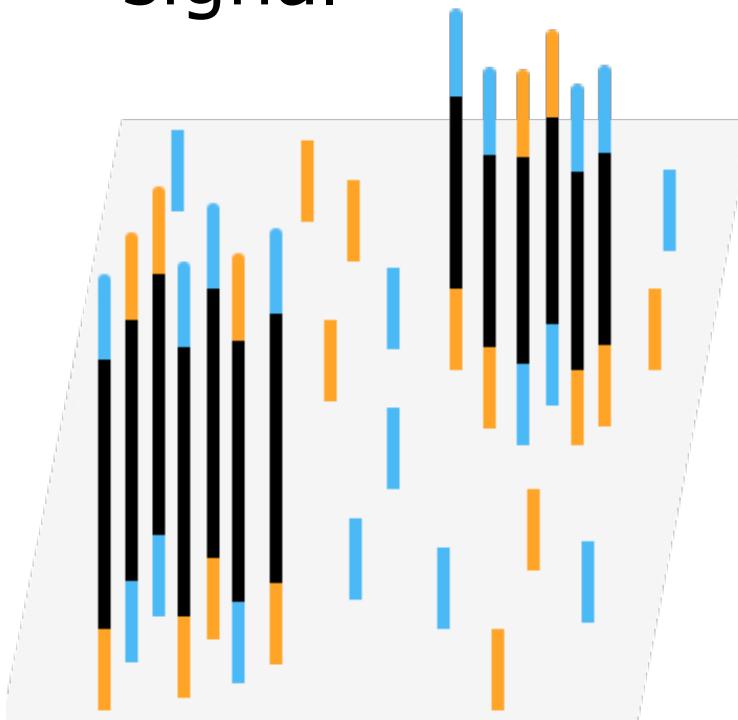
Quiz Question 2

Illumina - limitations

- Maximum read length: 300 bp
- How to reconstruct:
 - Repeats?
 - Isoforms?
 - Structural variation?
 - Haplotypes?
 - Genomes?
- Why not longer read lengths?

Illumina - limitations

- Bridge amplification
- Lengths are limited by out-of-phase of signal



Long reads (3rd generation)

- Crux: maximizing signal from a single-molecule base read-out
- Single molecule, so no out-of-phase signal
- Two frequently used platforms:
 - PacBio SMRT sequencing
 - Oxford Nanopore Technology



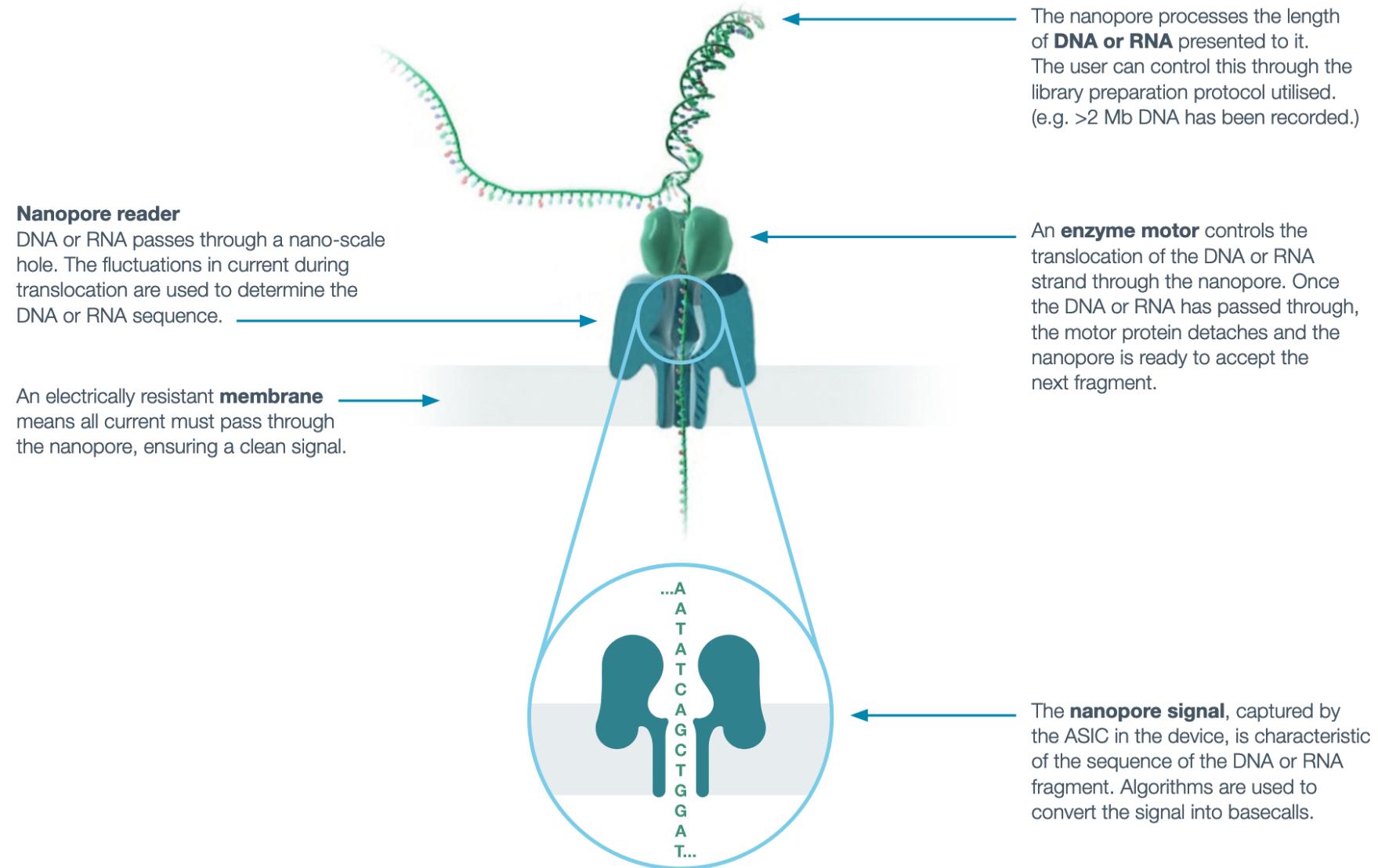
PACBIO®



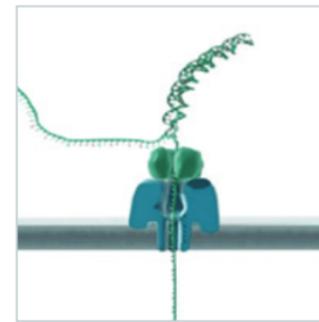
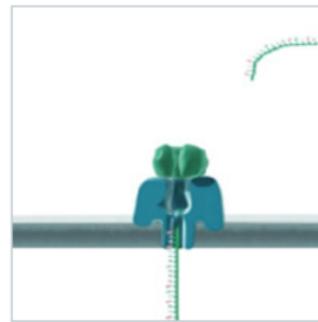
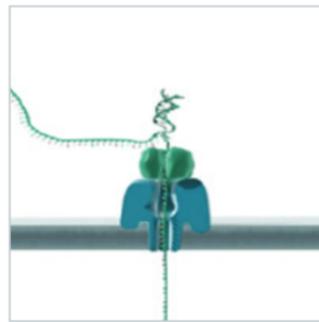
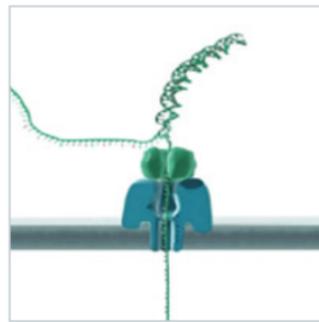
Oxford Nanopore technology

- Based on changes in electrical current
- Well-known for its scalability and portability
- ~95-97% accuracy





1D



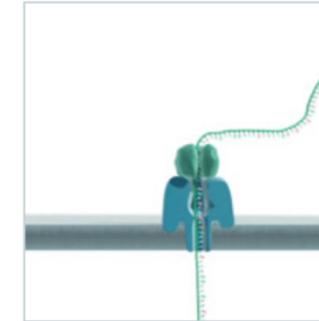
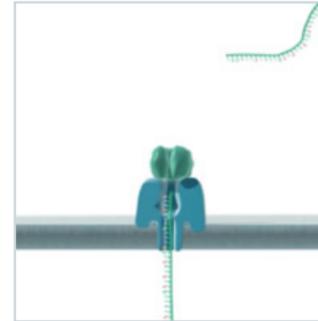
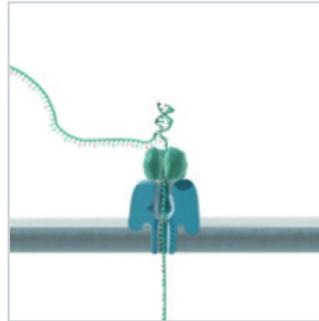
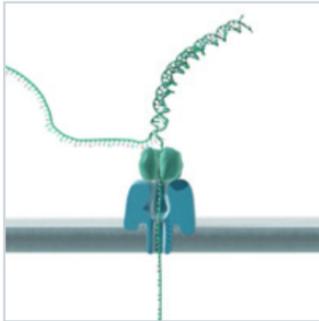
Template...

...Template...

(Exit)

Next molecule...

1D²



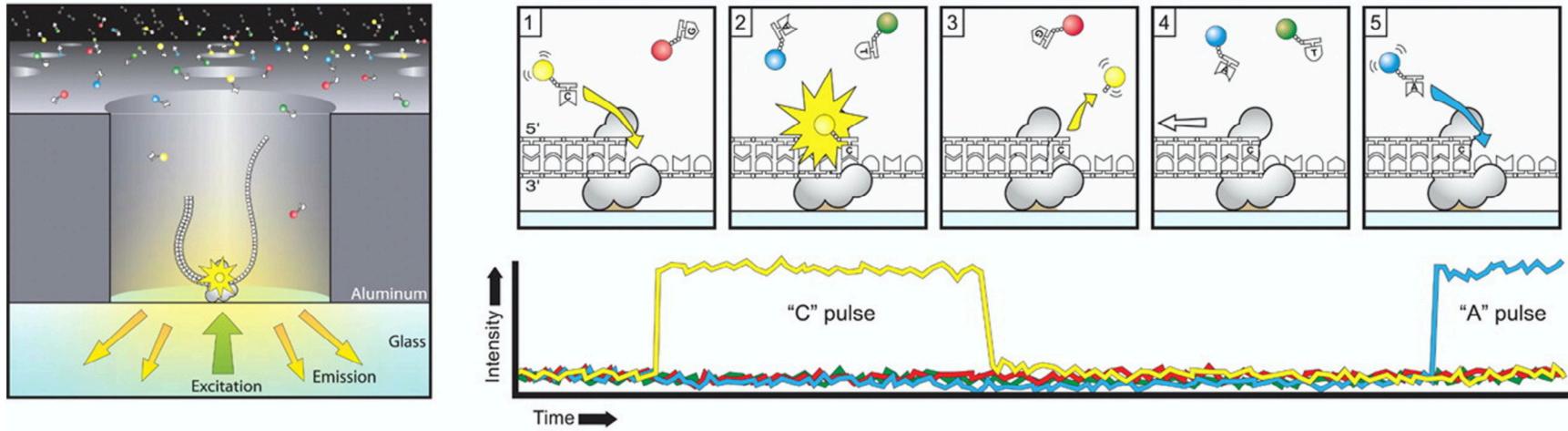
Template...

...Template...

(Exit)

...Complement

PacBio sequencing

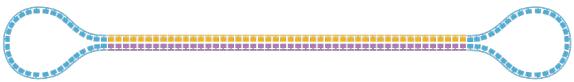


- Polymerase bound to ZMW bottom
- Circular molecules
- Single read out ~90% accuracy
- CCS (HiFi): single molecule sequenced multiple times

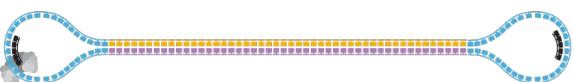
Start with high-quality double stranded DNA



Ligate SMRTbell adapters and size select



Anneal primers and bind DNA polymerase

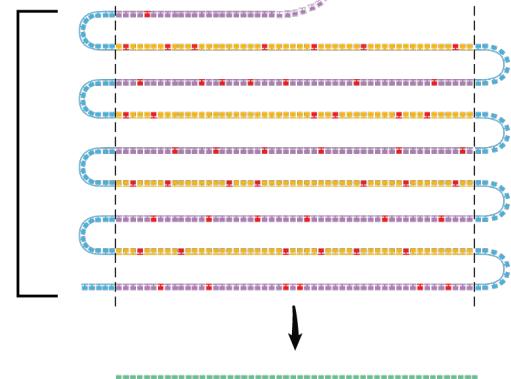


Circularized DNA is sequenced in repeated passes



The polymerase reads are trimmed of adapters to yield subreads

Consensus is called from subreads



HiFi READ
(>99% accuracy)

Quiz Question 3A and 3B