

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



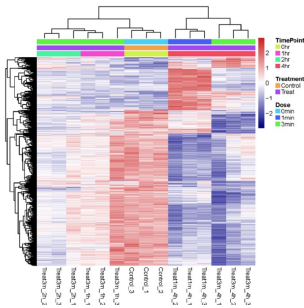
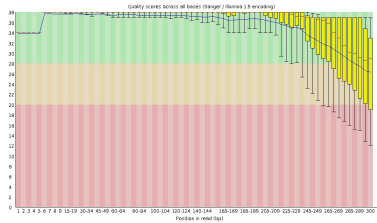
Alignment



Down-stream analysis



Visualisation



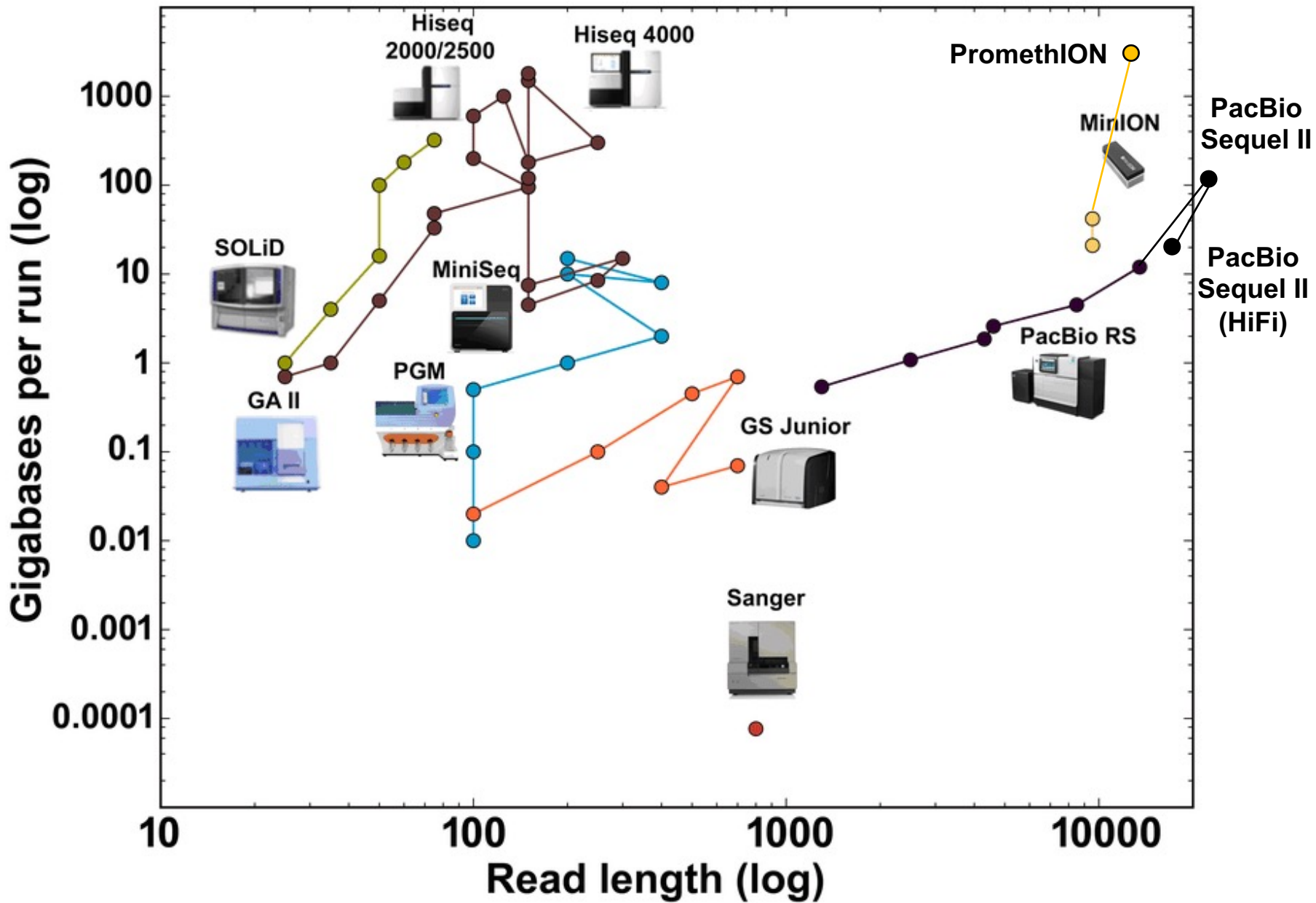


Image from: G. Silva (2016)

This course

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

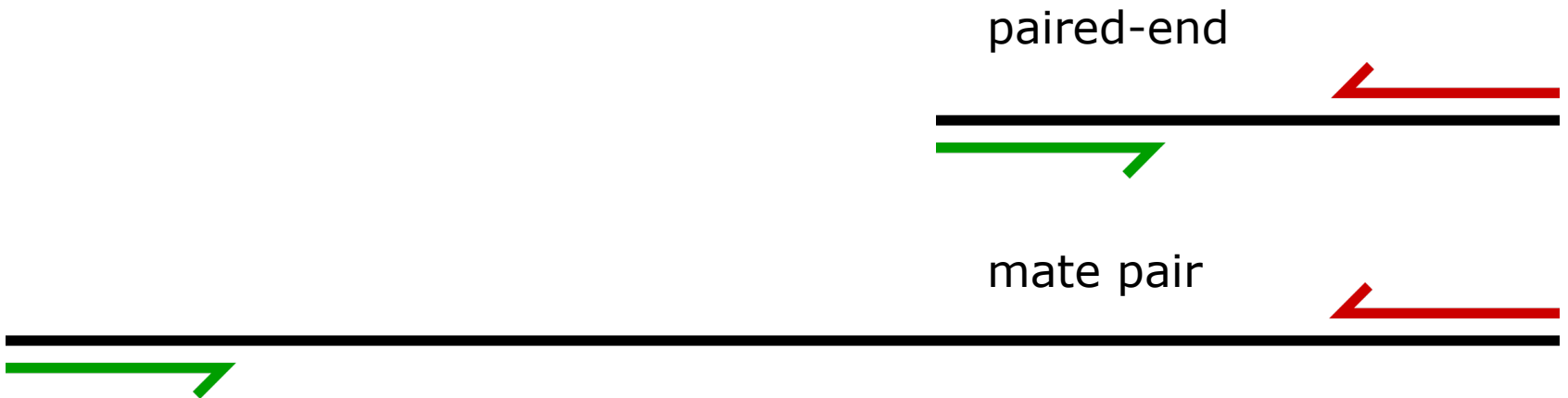
Quiz Question 2

Illumina sequencing

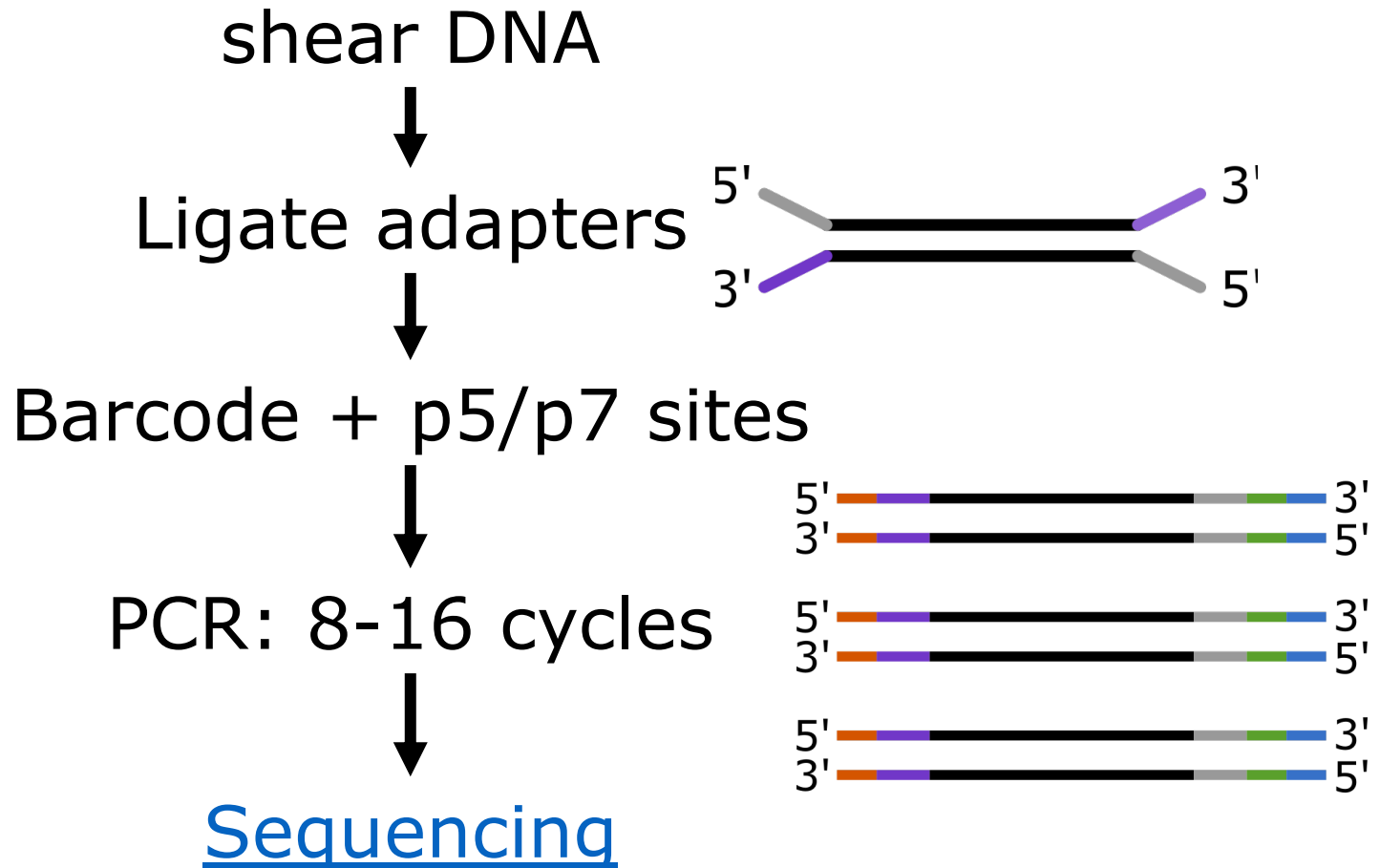
- Sequencing-by-synthesis: 2nd generation sequencing
- Massive throughput: up to 500×10^9 bases/run
- Most used platform today

Illumina sequencing

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



Illumina library prep



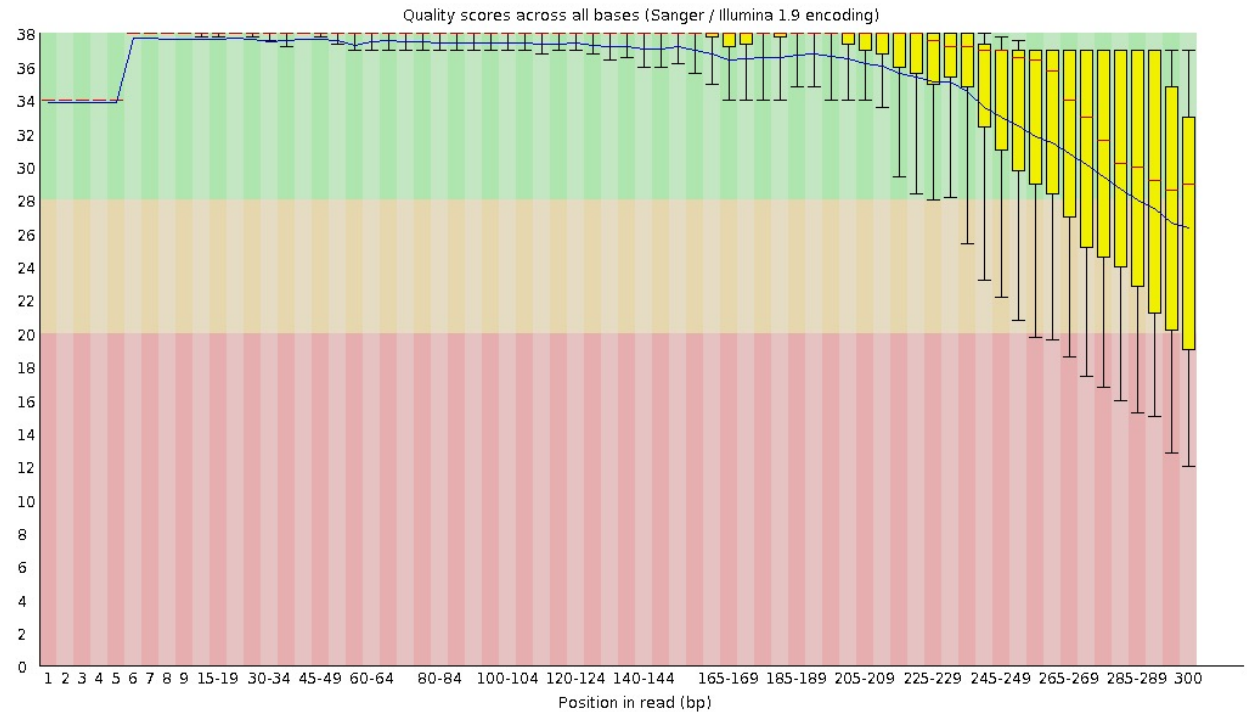
Quiz Question 3

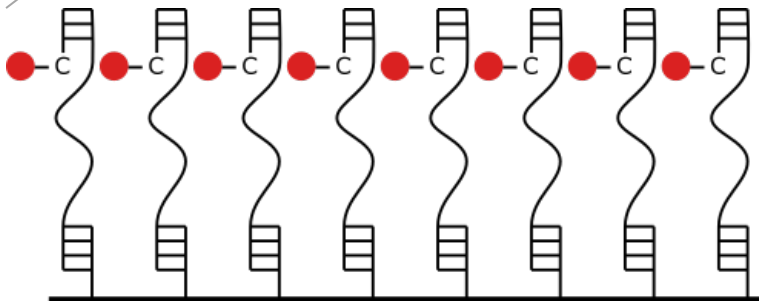
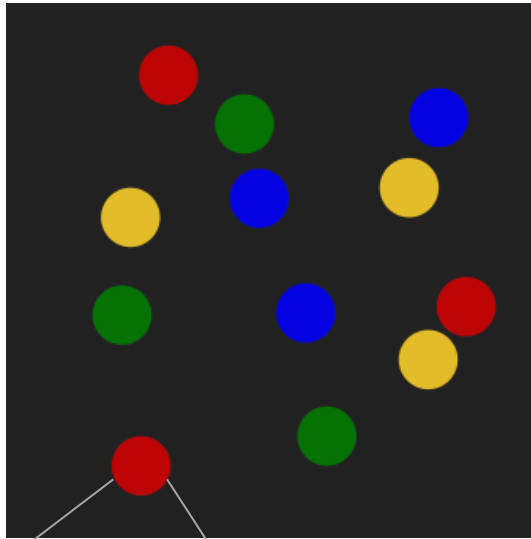
Illumina - limitations

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

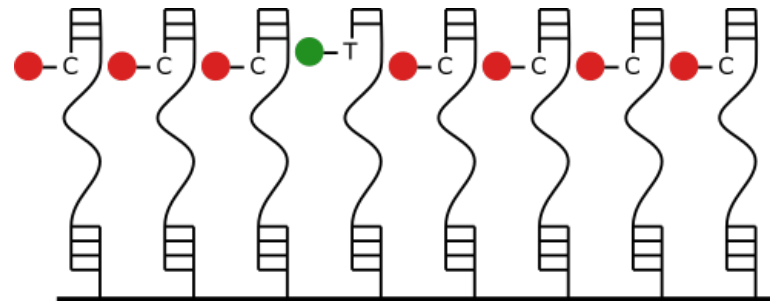
Illumina - limitations

Sequence quality declines towards the end





in phase



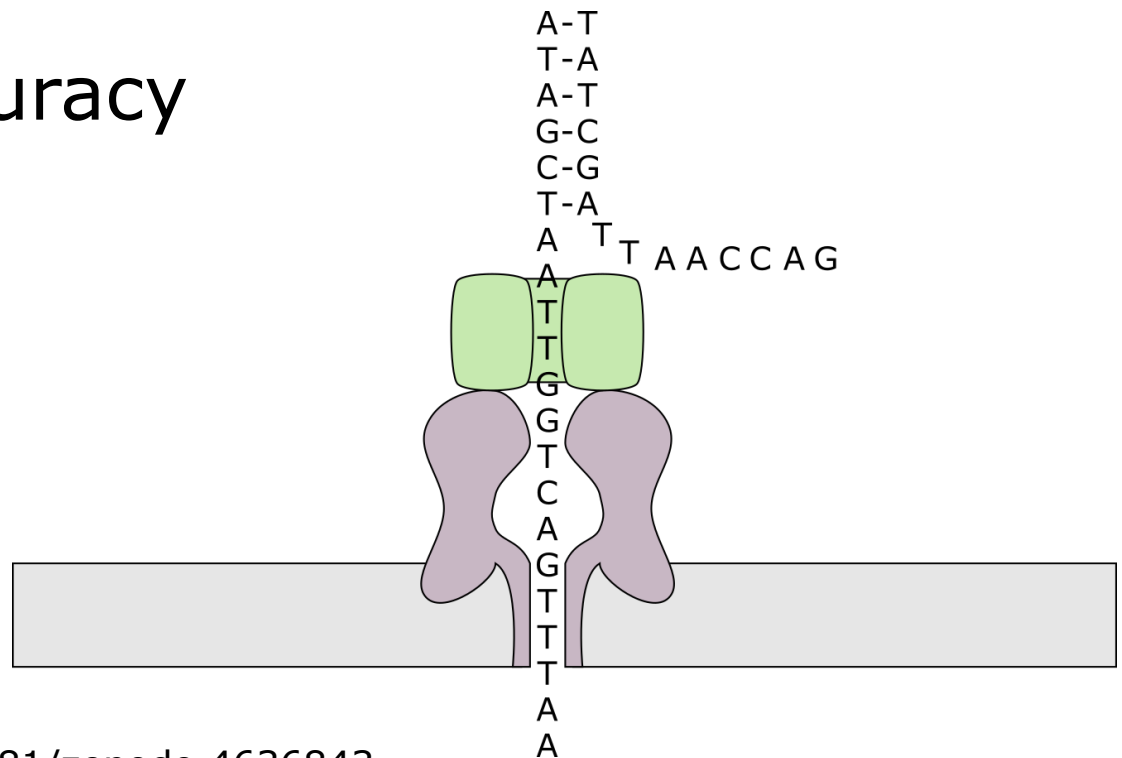
out of phase

Long reads (3rd generation)

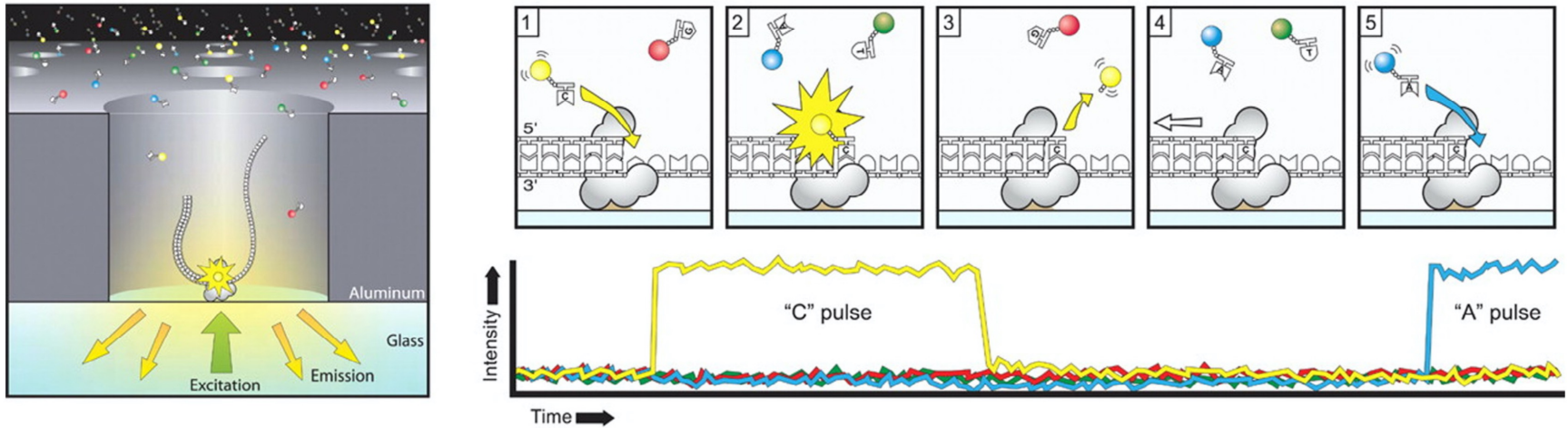
- Crux: maximizing signal from a single-molecule base read-out
- Single molecule, so no out-of-phase signal
- Two frequently used platforms:
 - PacBio SMRT sequencing
 - Oxford Nanopore Technology

Oxford Nanopore technology

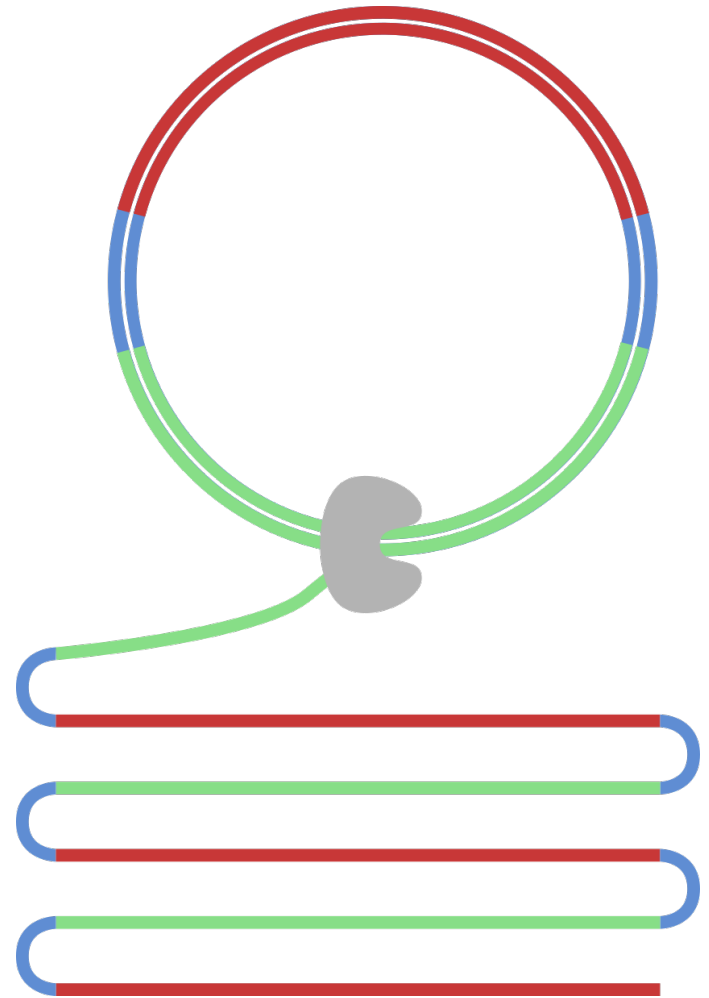
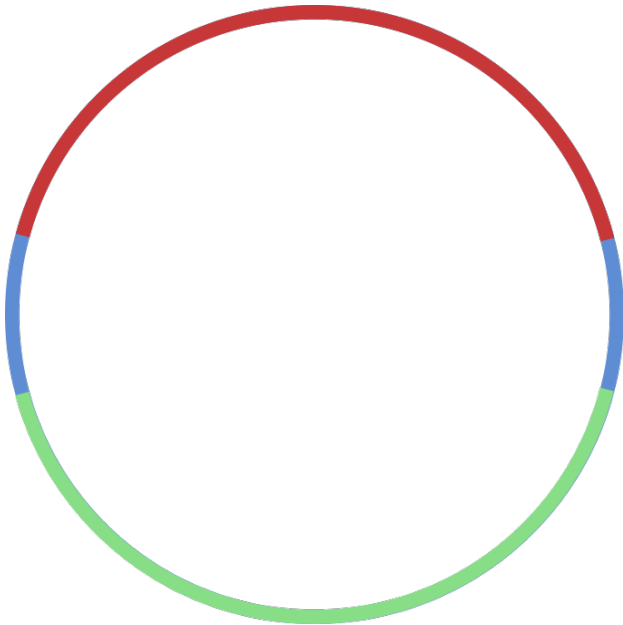
- Based on changes in electrical current
- Well-known for its scalability and portability
- ~95-97% accuracy



PacBio sequencing



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



Hi-Fi read

Quiz Question 4A and 4B