

# **NGS - quality control, alignment, visualisation**

Geert van Geest

Interfaculty Bioinformatics Unit, UniBe

Training group, SIB

# 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



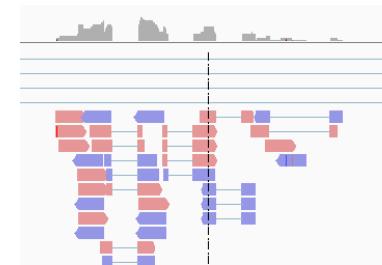
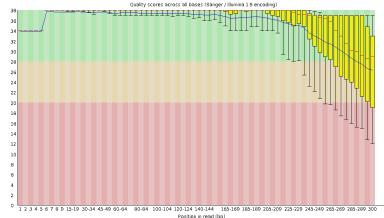
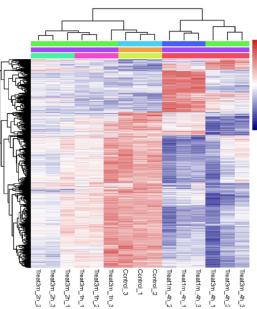
Alignment

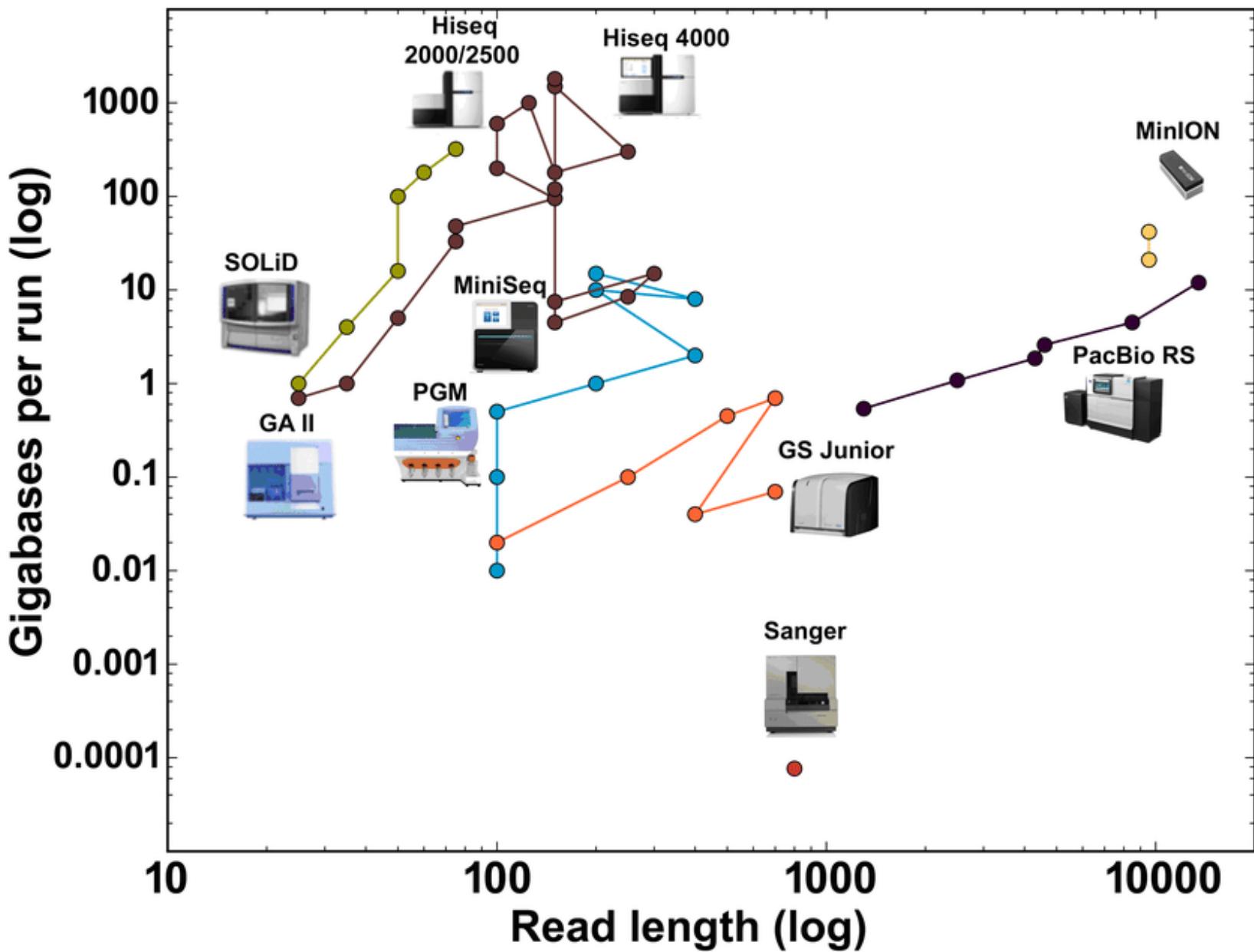


Down-stream analysis



Visualisation





# Quiz Question 1

# Illumina sequencing

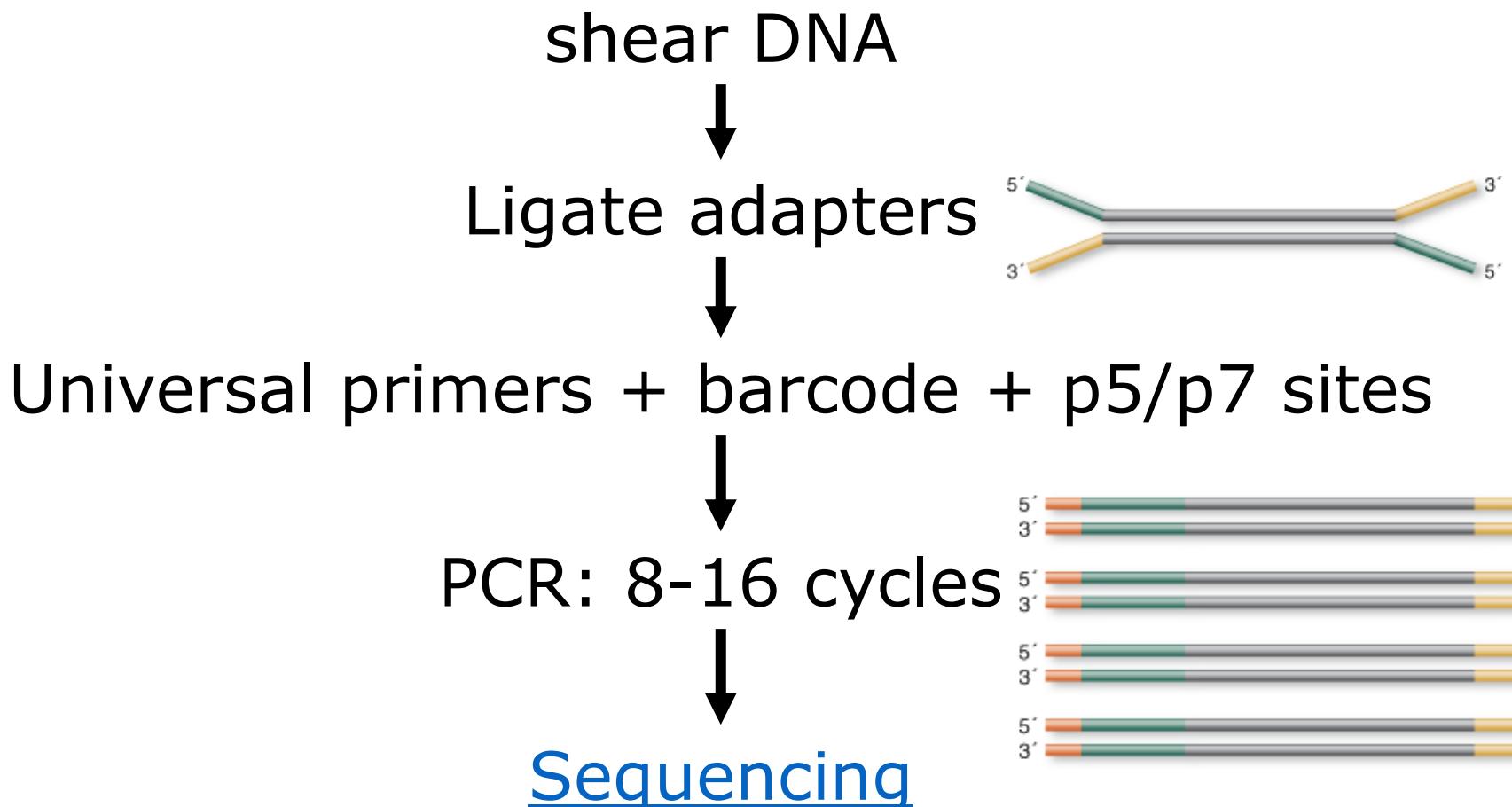
- Sequencing-by-synthesis: 2nd generation sequencing
- Massive throughput
- Most used platform today

illumina®

# Illumina sequencing

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

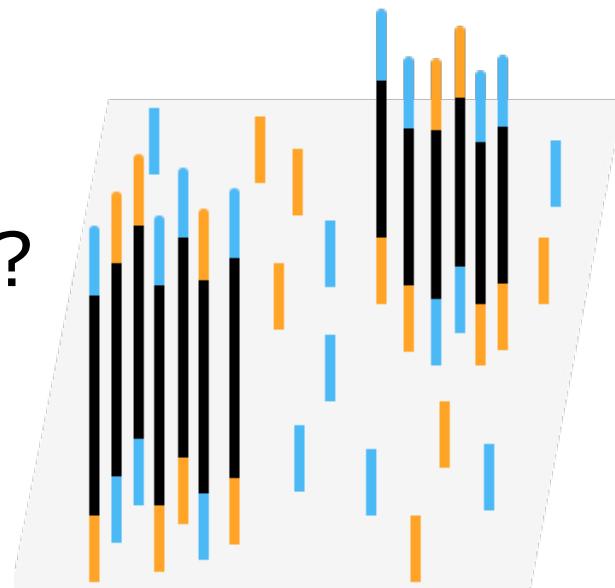
# Illumina library prep



# Quiz Question 2

# Illumina - limitations

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



# 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



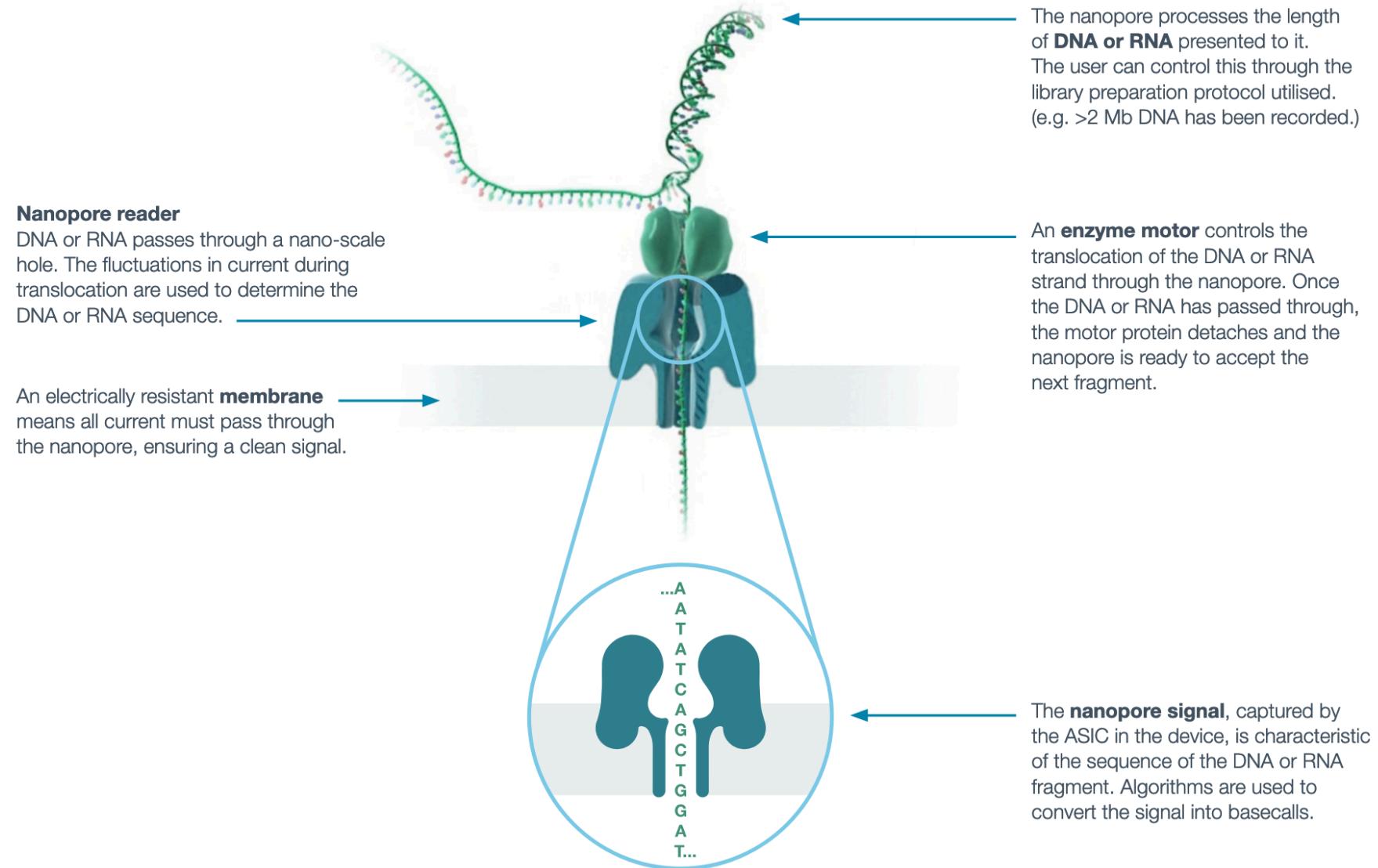
PACBIO®



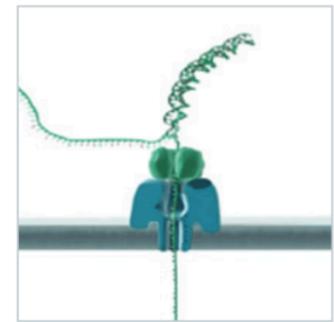
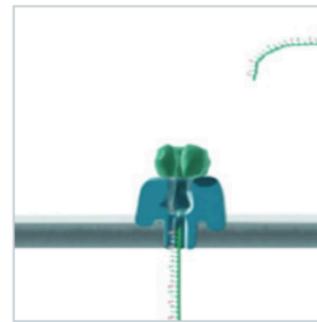
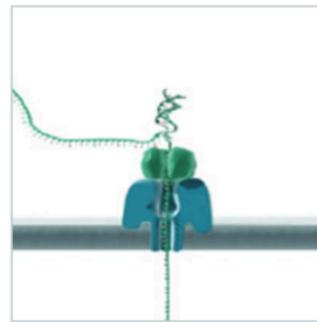
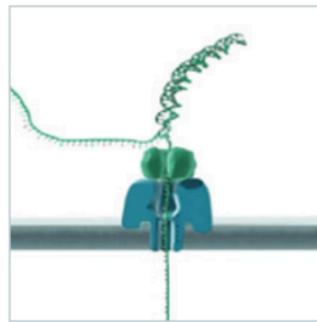
# Oxford Nanopore technology

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





**1D**



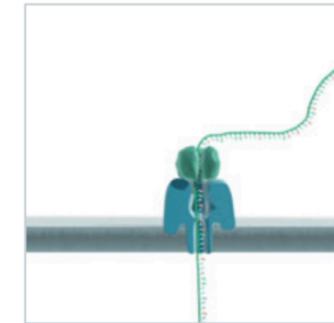
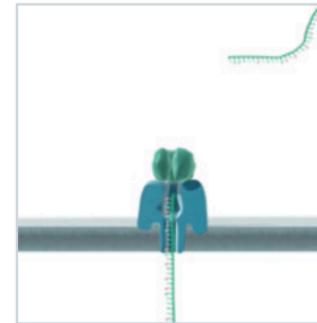
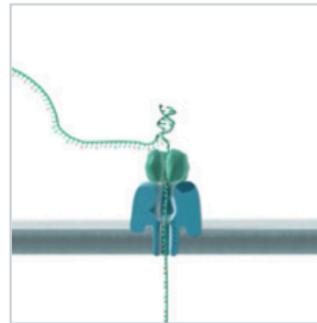
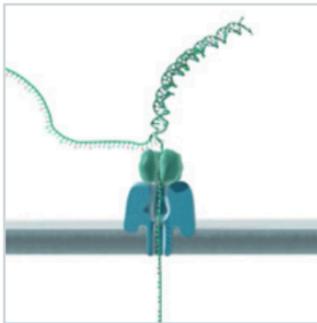
Template...

...Template...

(Exit)

Next molecule...

**1D<sup>2</sup>**



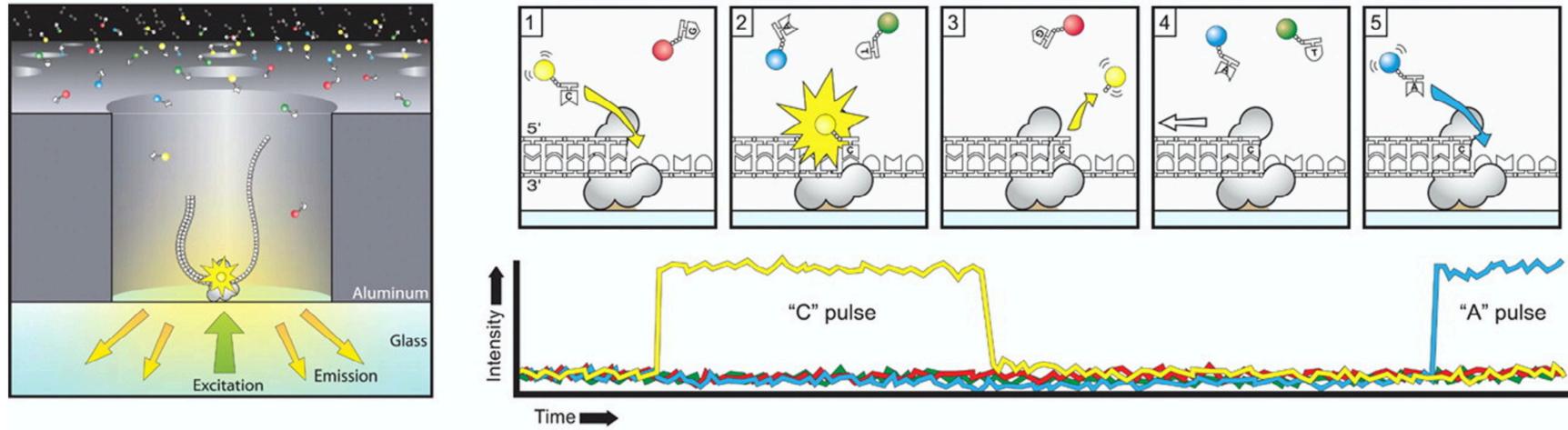
Template...

...Template...

(Exit)

...Complement

# PacBio sequencing



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

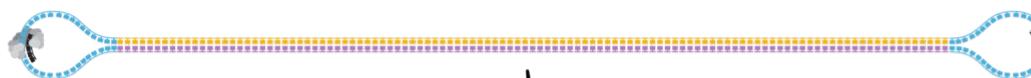
Start with high-quality  
double stranded DNA



Ligate SMRTbell  
adapters and size select



Anneal primers and  
bind DNA polymerase



Circularized DNA  
is sequenced in a  
single pass



The polymerase reads  
are trimmed of adapters  
to yield subread



During assembly,  
consensus is called from  
multiple molecules

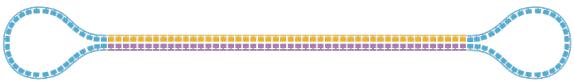
### LONG READ

(Half of Reads >50 kb)

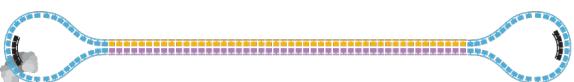
Start with high-quality double stranded DNA



Ligate SMRTbell adapters and size select



Anneal primers and bind DNA polymerase

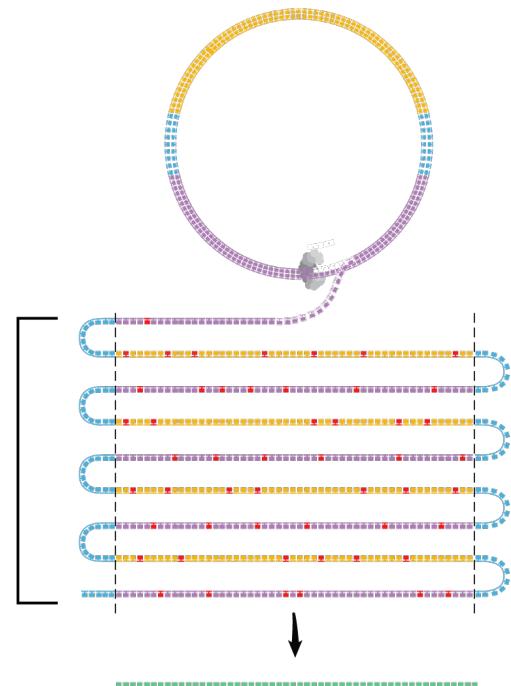


Circularized DNA is sequenced in repeated passes



The polymerase reads are trimmed of adapters to yield subreads

Consensus is called from subreads



**HiFi READ**  
(>99% accuracy)

# Quiz Question 3A and 3B