

Long-read sequence analysis

Sequencing technologies

What is a long read?

- Short read: 50-300 bp, often paired-end
- Long read: > 1kb, up to 20 Mb

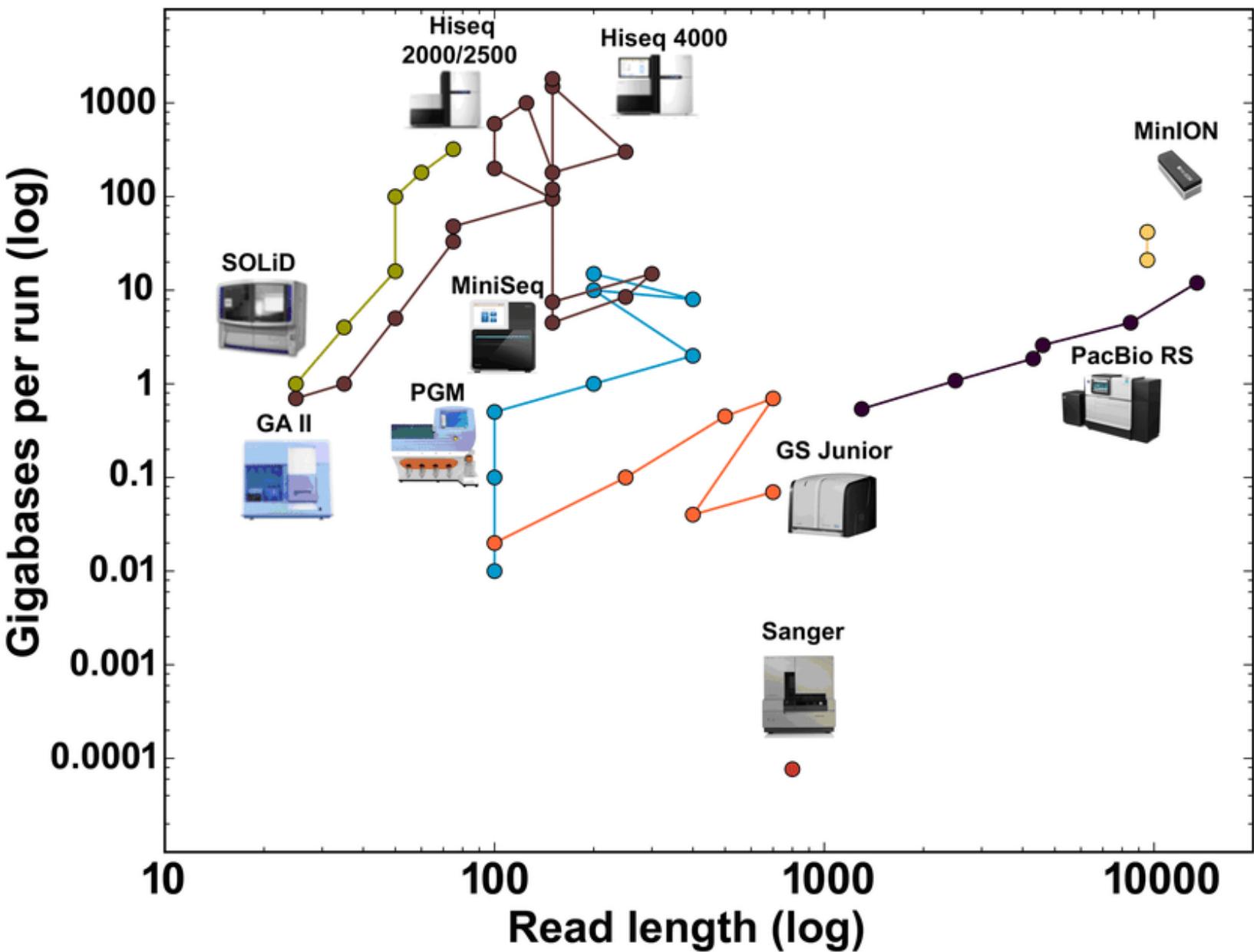


Image from: G. Silva (2016)

Illumina sequencing

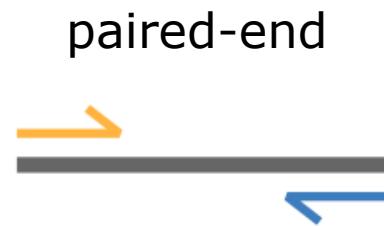
- Sequencing-by-synthesis: 2nd generation sequencing
- Massive throughput: up to 500×10^9 bases/run
- Most used platform today

illumina®



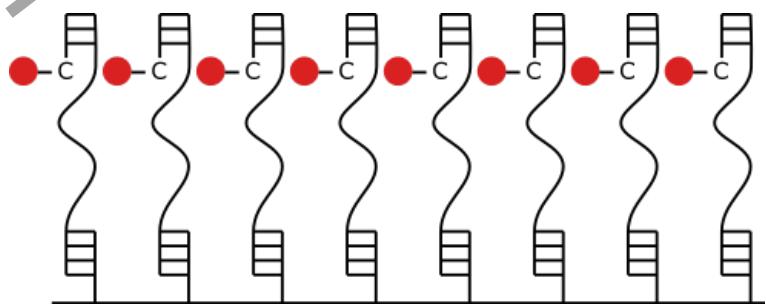
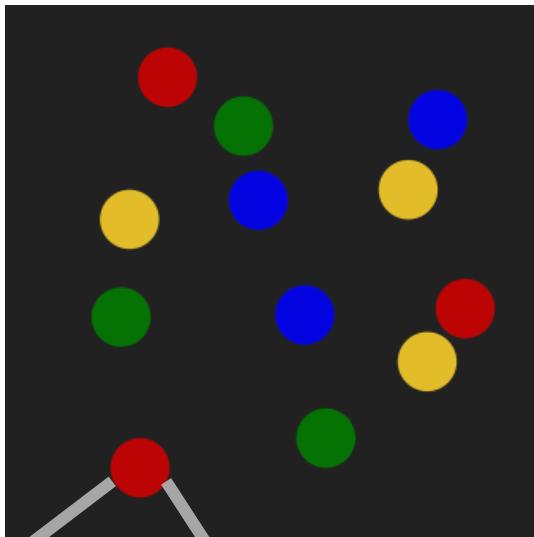
Illumina sequencing

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

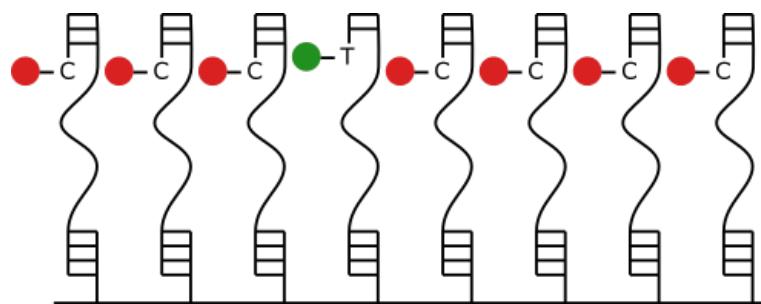


Illumina - limitations

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



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



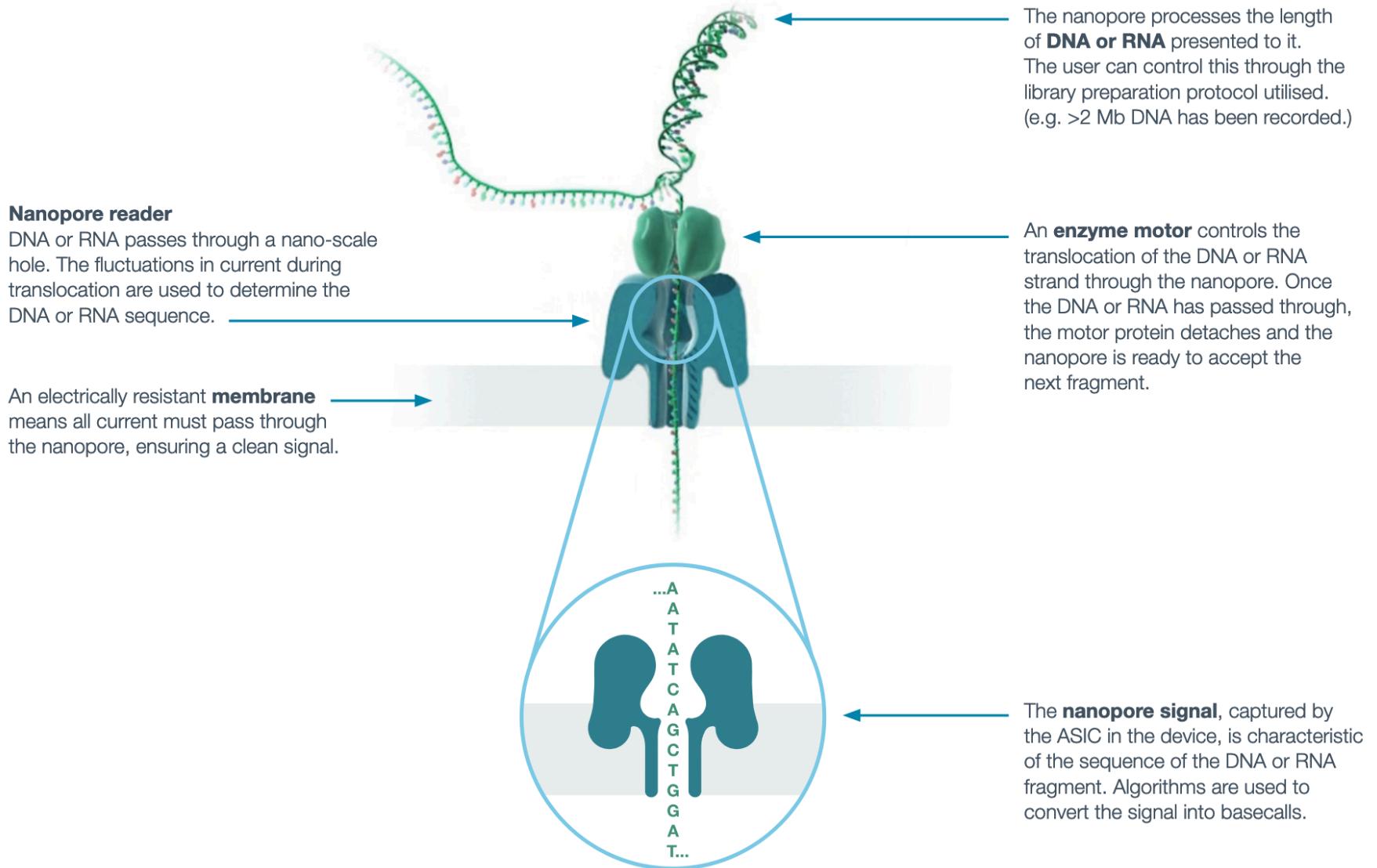
PACBIO®



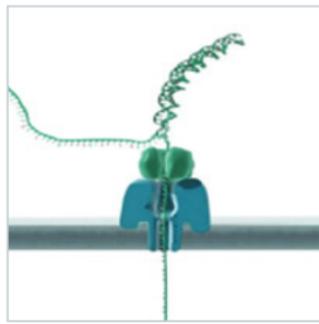
Oxford Nanopore technology

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

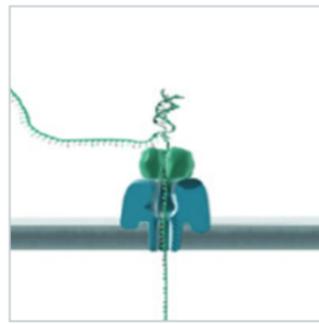




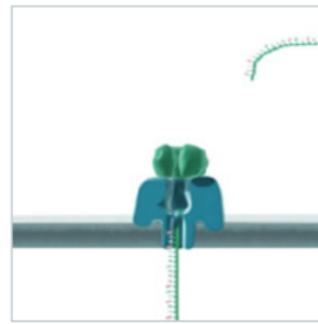
1D



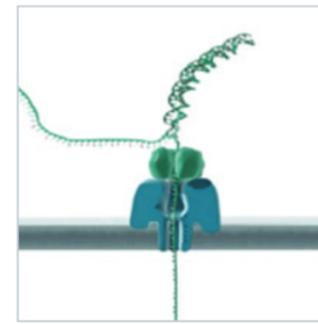
Template...



...Template...

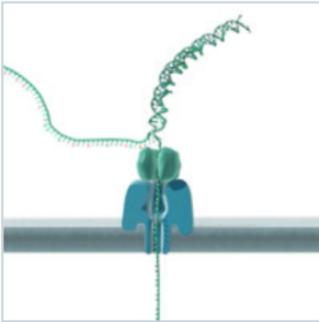


(Exit)

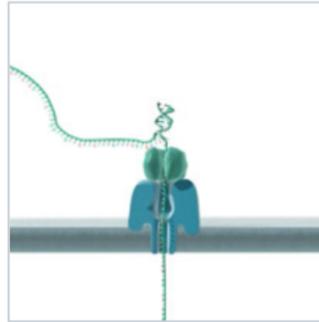


Next molecule...

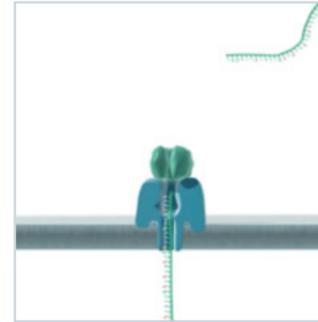
1D²



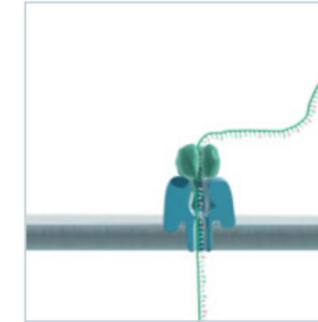
Template...



...Template...

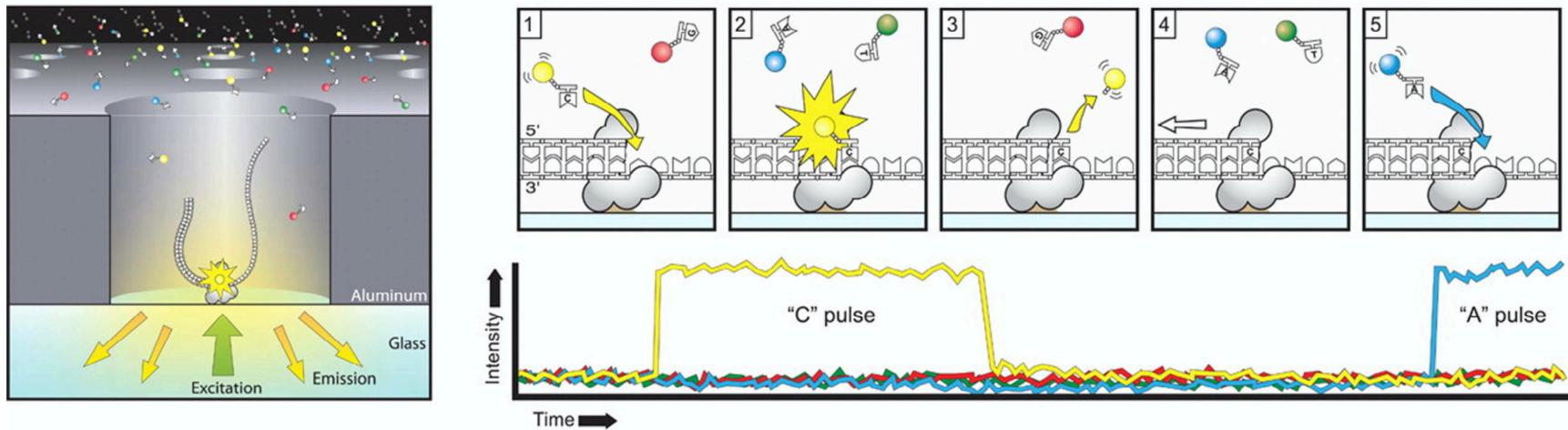


(Exit)

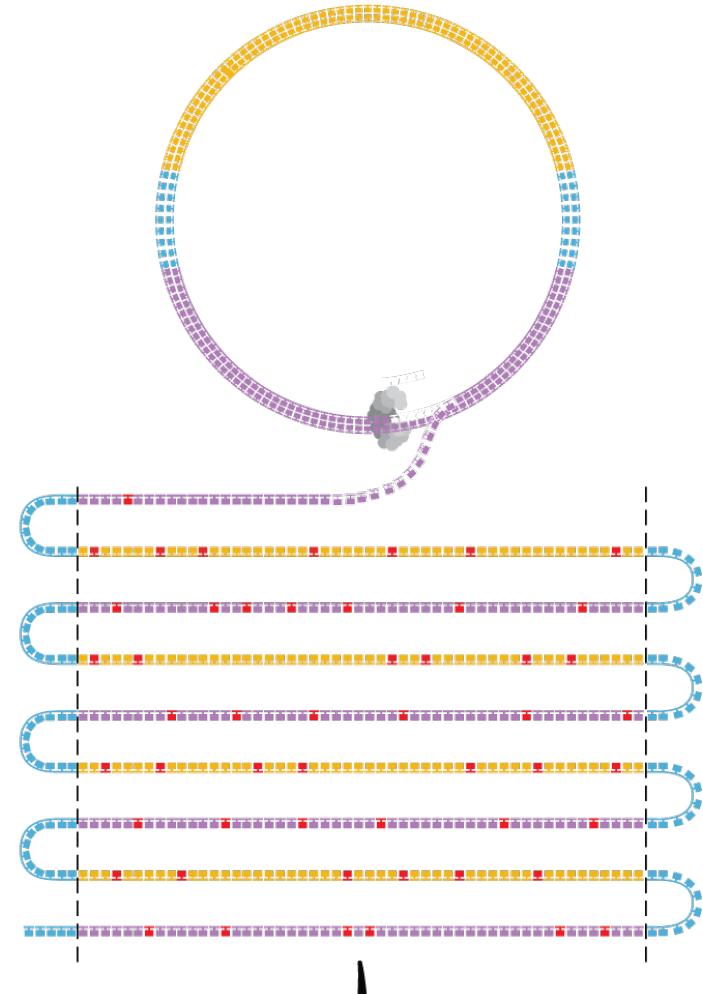
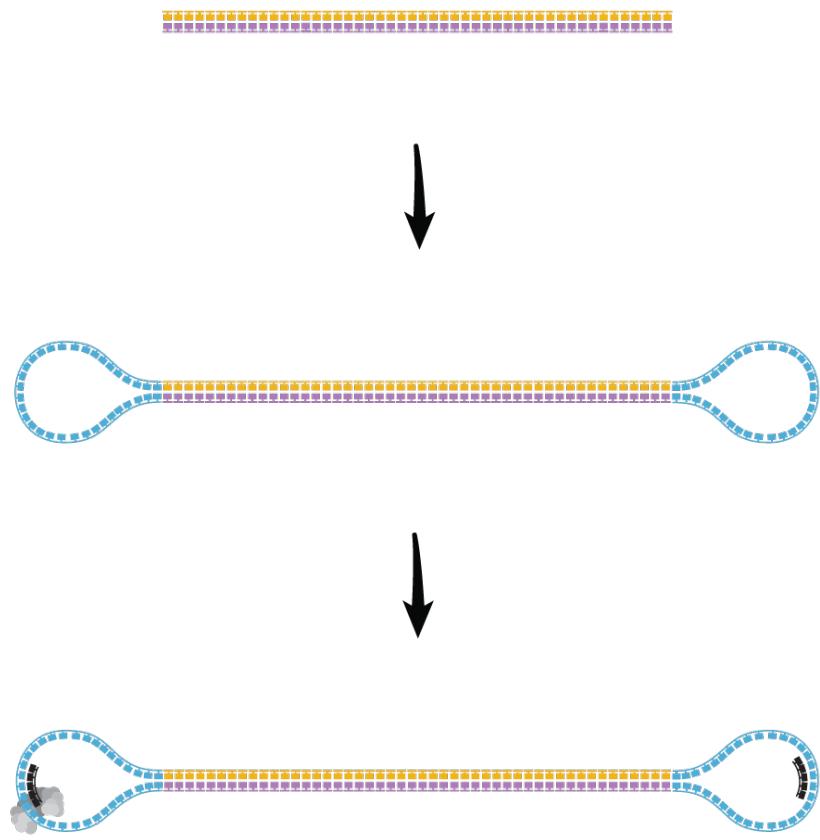


...Complement

PacBio sequencing



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



HiFi READ
(>99% accuracy)

	ONT	PacBio
Read accuracy	~95-97%	~90% (>99% CCS)
Read length	up to 20 Mb	up to 30-40 kb
RNA base modifications	Yes (m6A) ¹	No
DNA base modifications	Yes (m5C, m6A) ²	Yes (m5C, m6A, hm5C) ³
Throughput (BIF)	~250M reads/run	~4M CCS reads/run

1. Liu, H., et al (2019). Accurate detection of m6A RNA modifications in native RNA sequences. *Nature Communications*, 10(1), 1–9
2. Liu, Q., et al (2019). Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. *Nature Communications*, 10(1).
3. Flusberg, B. A., et al (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods*, 7(6), 461–465