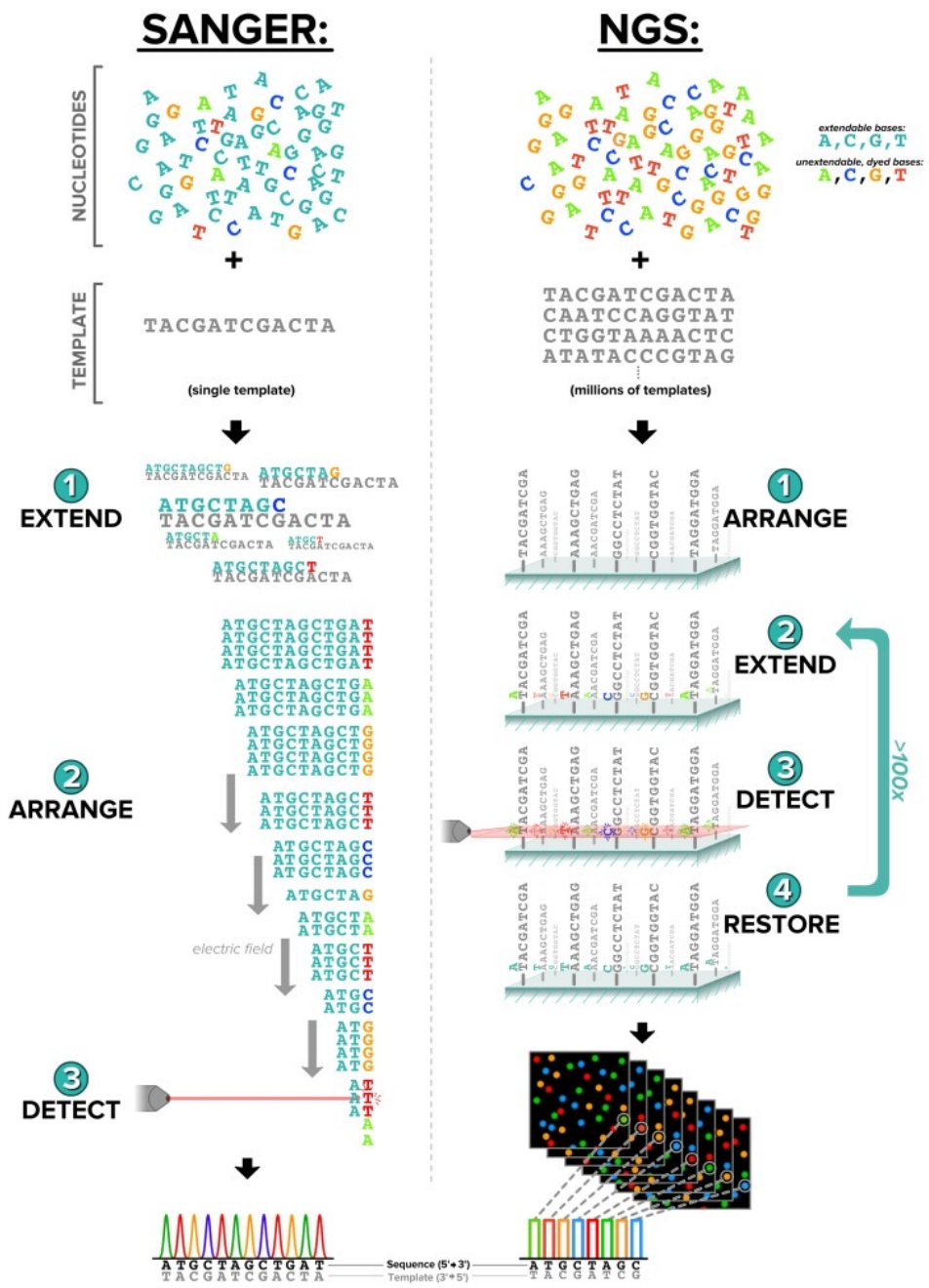
Next Generation Sequencing (NGS)

HAMIDAH SUUBI NAMAGEMBE

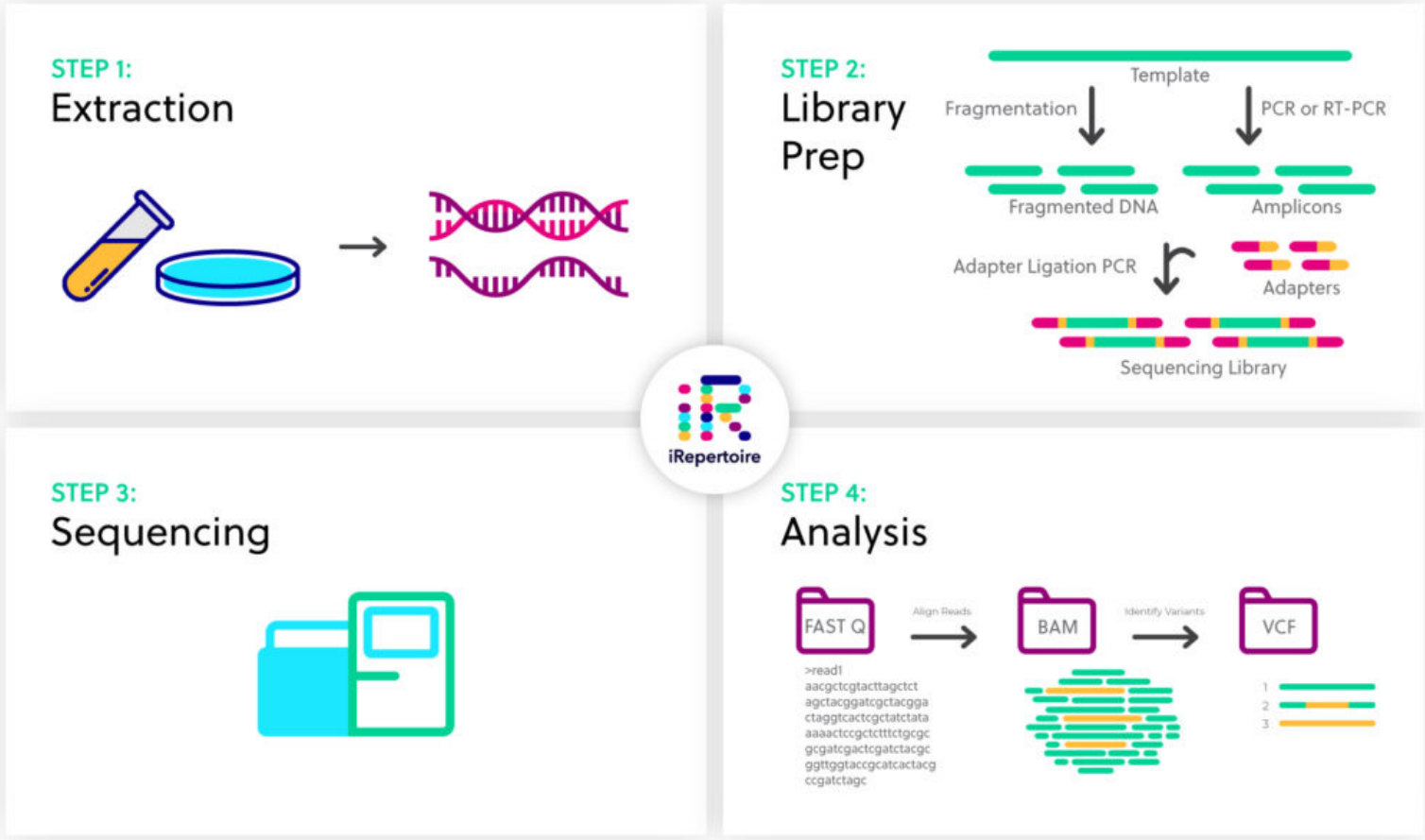
26TH JUNE 2023

Introduction

- Next-Generation Sequencing (NGS) is a high-throughput DNA sequencing method that has transformed genomic research
- NGS enables faster, more cost-effective, and more comprehensive analysis of genetic material



NGS Workflow



NGS Technologies

Short-Read NGS

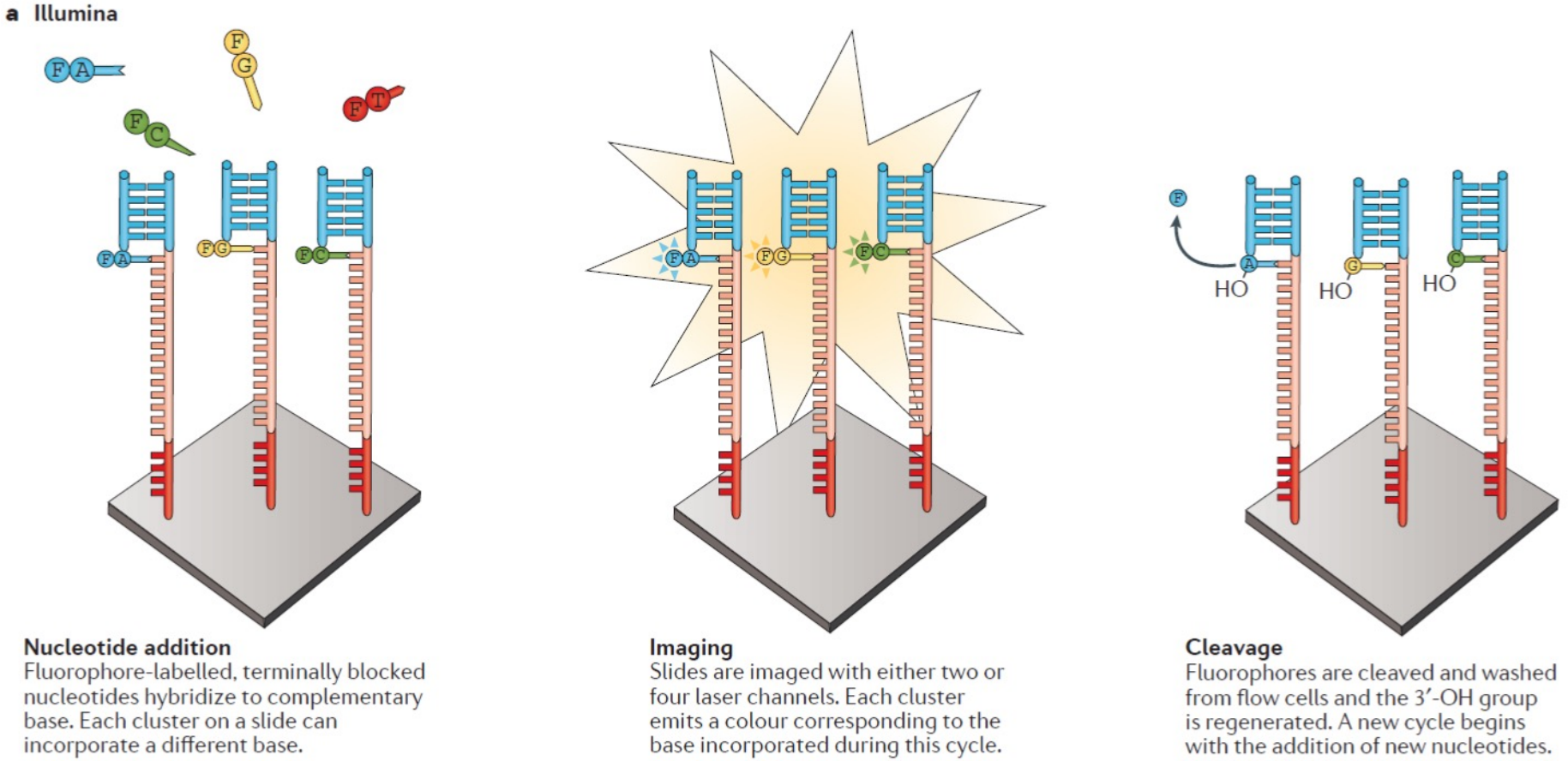
- Sequencing by synthesis (SBS) –
  i. Cyclic reversible chain termination (CRT): Illumina & Qiagen
  ii.Single nucleotide addition (SNA): Roche & Ion Torrent
- Sequencing by ligation (SBL)- Thermo Fisher & BGI

Long-Read NGS

- Realtime sequencing approaches – Pacific Biosciences & Oxford Nanopore Technologies
- Synthetic approaches- Illumina & 10X Genomics

Short-Read NGS

Sequencing by Synthesis (SBS)- *Cyclic reversible chain termination* (CRT): Illumina



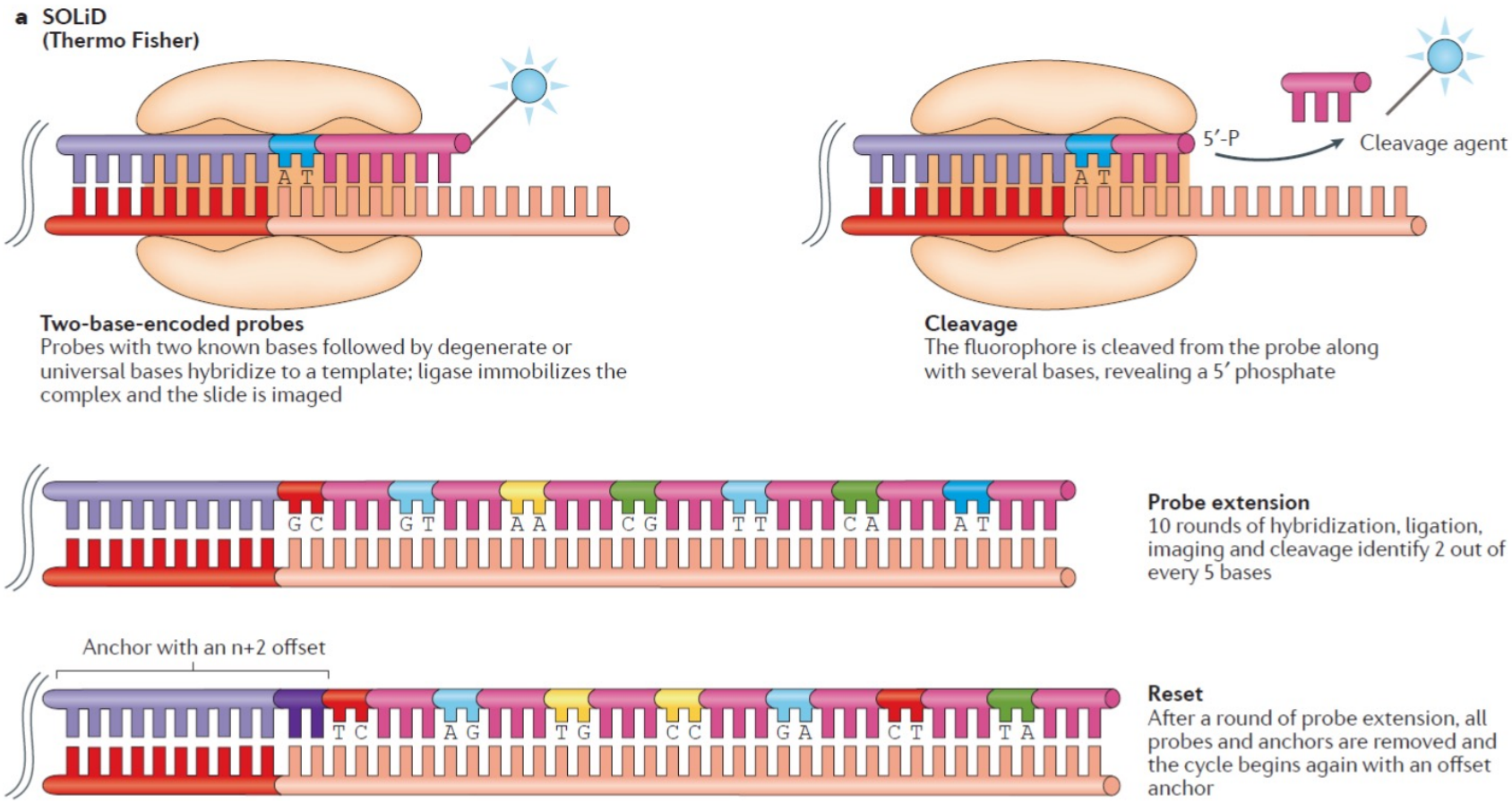
Nucleotide addition
Fluorophore-labelled, terminally blocked nucleotides hybridize to complementary base. Each cluster on a slide can incorporate a different base.

Imaging
Slides are imaged with either two or four laser channels. Each cluster emits a colour corresponding to the base incorporated during this cycle.

Cleavage
Fluorophores are cleaved and washed from flow cells and the 3′-OH group is regenerated. A new cycle begins with the addition of new nucleotides.

Short-Read NGS

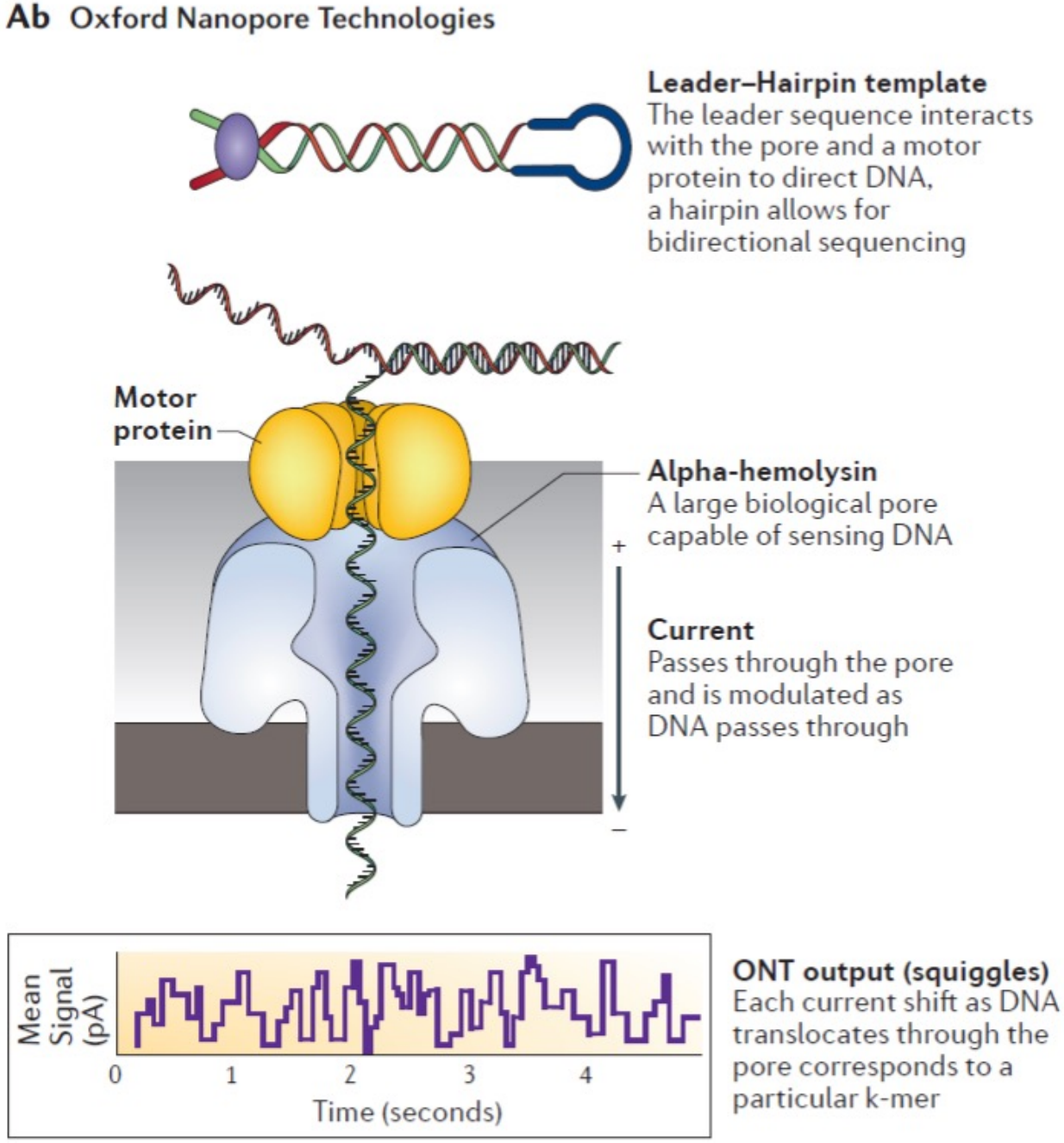
Sequencing by ligation (SBL)- Thermo Fisher



Long-Read NGS

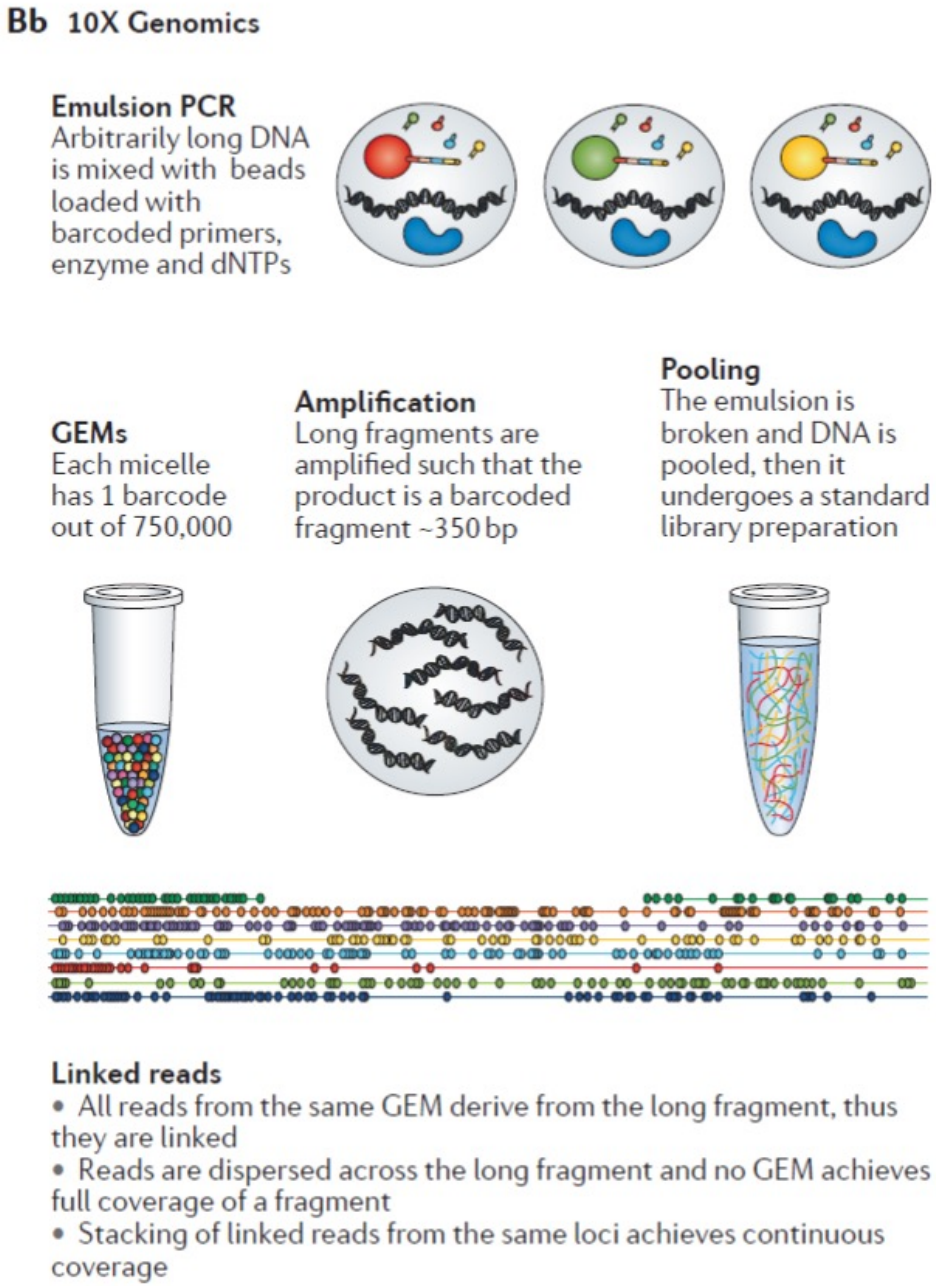
Single-molecule sequencing approaches-

ONT is based on the detection of changes in ionic current as DNA or RNA molecules pass through a nanopore



Long-Read NGS

Synthetic approaches-10X Genomics -Gel Bead-in-Emulsion (GEM) Generation: DNA molecules are encapsulated within droplets



Short-read vs Long-read

Short-read NGS involves sequencing DNA fragments of relatively small lengths, typically ranging from 50 to 300 base pairs.

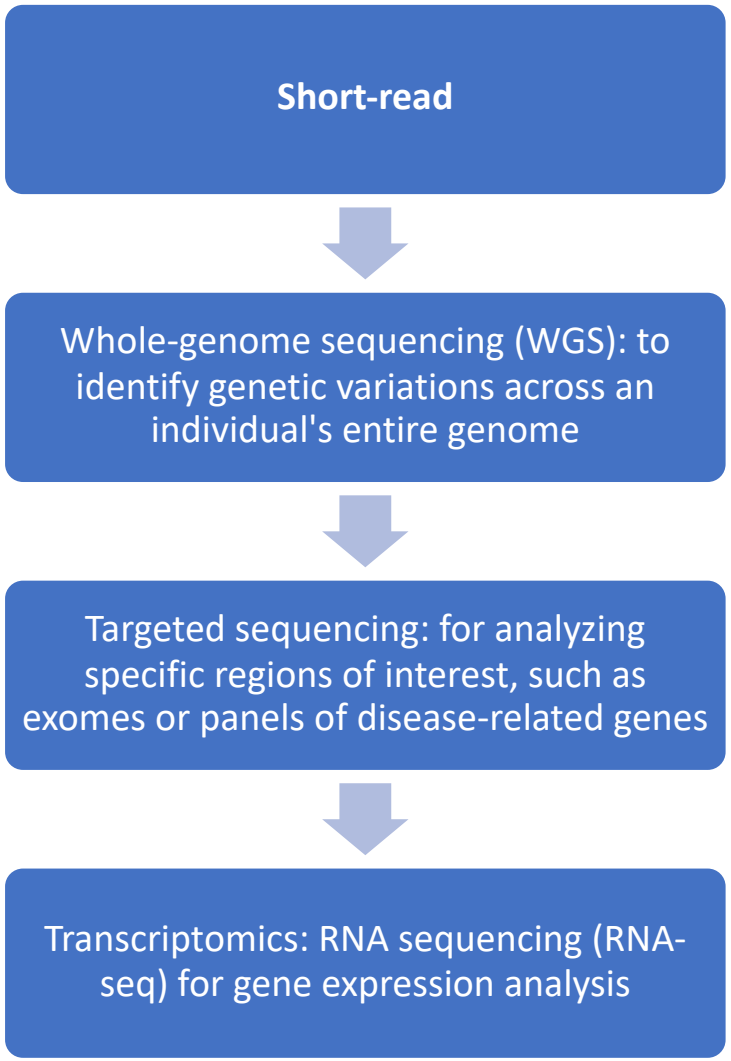
High-throughput: a large volume of sequences in a single run

Cost-effective: generally more affordable per base

High accuracy: Short reads typically have low error rates

- **Long-read NGS**-Long DNA fragments of thousands of base pairs
- Greater sequence continuity: better characterization of complex genomic regions, repetitive sequences, and structural variations
- Direct detection of modifications: such as DNA methylation

Applications of Short-read and Long-read platforms



- **Long-read**
- De novo genome assembly: for reconstructing complex genomes with high accuracy and resolving structural variations
- Characterizing repetitive regions: help resolve repetitive regions
- Studying RNA isoforms: provides insights into alternative splicing events and full-length isoforms in transcriptomics research

Data Quality Metrics

- **Depth of coverage**: how many sequence reads are present at a given position
- **Base quality**: have the correct bases been called in sequence reads)
- **Mapping quality**: have the reads been mapped to the correct position in the genome)

Table 1: Quality Scores and Base Calling Accuracy

Phred Quality Score	Probability of Incorrect Base Call	Base Call Accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1,000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%

Challenges and limitations

Data Storage: NGS generates large volumes of data that require substantial storage capacity.

Computational Requirements: Complex data analysis necessitates powerful computing

Cost: Although prices have decreased, NGS can still be relatively expensive.

Sample Quality: Poor sample quality may impact sequencing results and data interpretation.

References

- doi: 10.1177/1099800417750746
- https://irepertoire.com/ngs-overview-from-sample-to-sequencer-to-results/
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics, 17(6), 333–351. doi:10.1038/nrg.2016.49

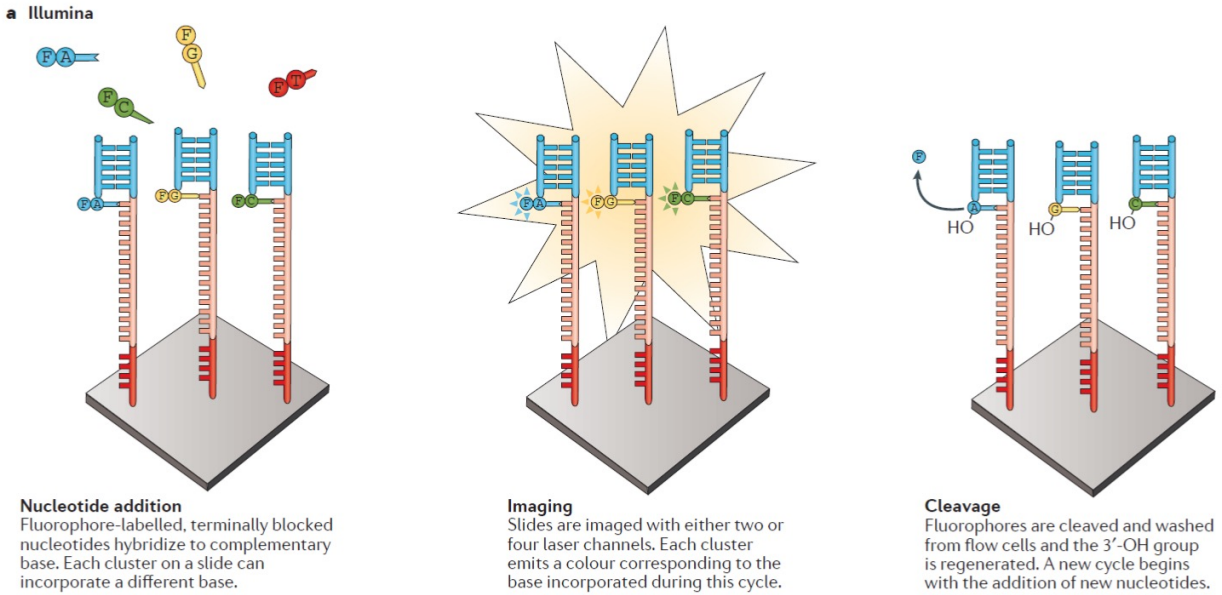
THANK YOU

NGS Technologies

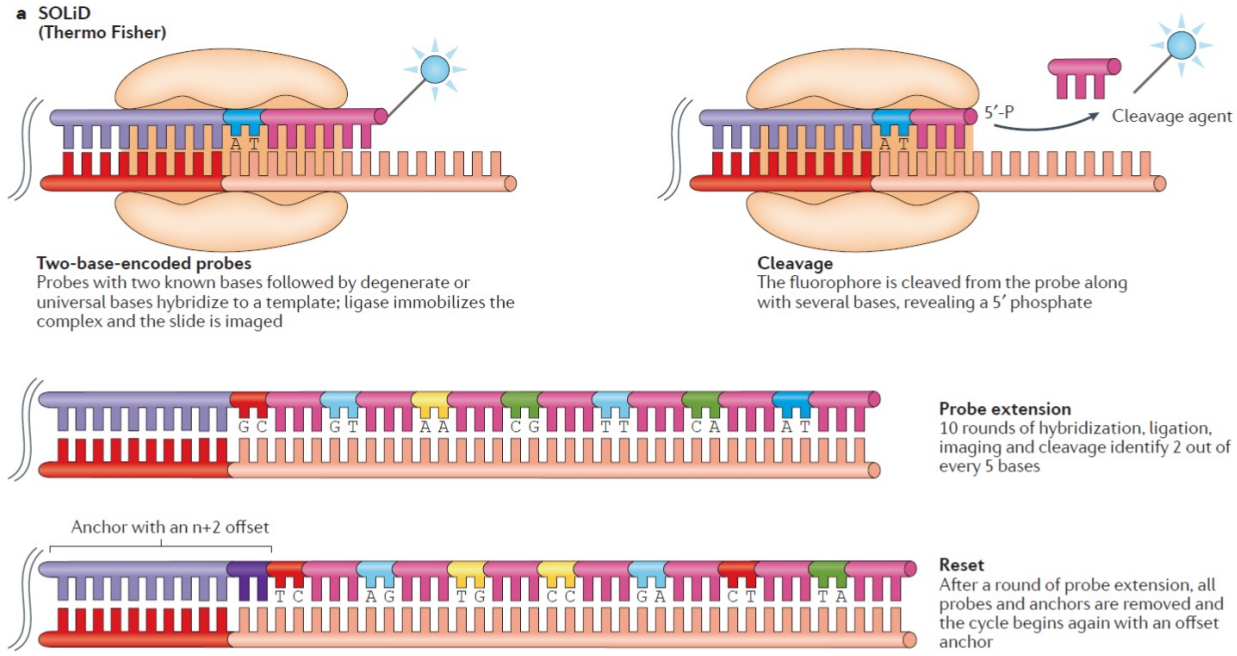
Short-Read NGS	Sequencing by synthesis (SBS) –*Cyclic reversible chain termination* (CRT): Illumina & Qiagen, and *Single nucleotide addition* (SNA): Roche & Ion Torrent
	Sequencing by ligation (SBL)- Thermo Fisher, BGI
Long-Read NGS	Realtime sequencing approaches – Pacific Biosciences, Oxford Nanopore Technologies
	Synthetic approaches- Illumina, 10X Genomics

Short-Read NGS

Sequencing by Synthesis (SBS)



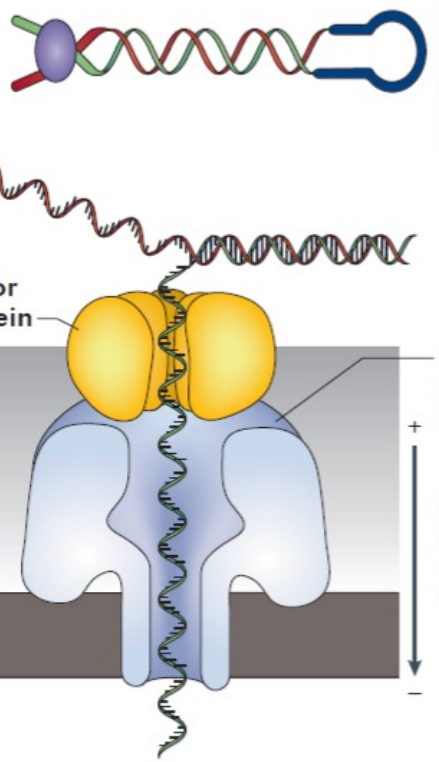
Sequencing by Ligation (SBL)



Long-Read NGS

Single-molecule sequencing approaches

Ab Oxford Nanopore Technologies

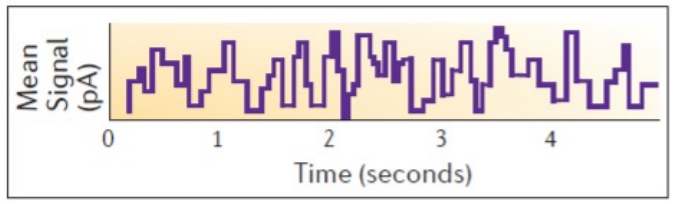


Leader–Hairpin template
The leader sequence interacts with the pore and a motor protein to direct DNA, a hairpin allows for bidirectional sequencing

Motor protein

Alpha-hemolysin
A large biological pore capable of sensing DNA

Current
Passes through the pore and is modulated as DNA passes through

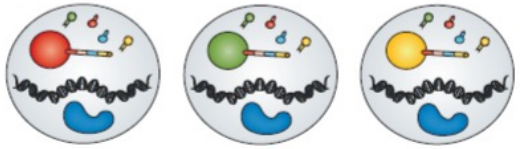


ONT output (squiggles)
Each current shift as DNA translocates through the pore corresponds to a particular k-mer

Synthetic approaches

Bb 10X Genomics

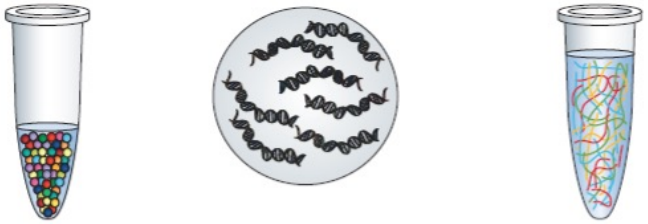
Emulsion PCR
Arbitrarily long DNA is mixed with beads loaded with barcoded primers, enzyme and dNTPs

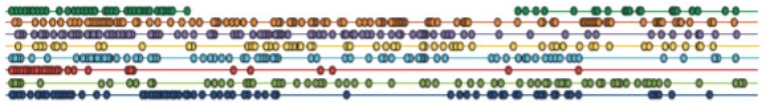


GEMs
Each micelle has 1 barcode out of 750,000

Amplification
Long fragments are amplified such that the product is a barcoded fragment ~350 bp

Pooling
The emulsion is broken and DNA is pooled, then it undergoes a standard library preparation





Linked reads

- All reads from the same GEM derive from the long fragment, thus they are linked
- Reads are dispersed across the long fragment and no GEM achieves full coverage of a fragment
- Stacking of linked reads from the same loci achieves continuous coverage

Short-read vs Long-read

Short-read NGS involves sequencing DNA fragments of relatively small lengths, typically ranging from 50 to 300 base pairs.

High-throughput: a large volume of sequences in a single run

Cost-effective: generally more affordable per base

High accuracy: Short reads typically have low error rates

- **Long-read NGS**-Long DNA fragments of thousands of base pairs.
- Greater sequence continuity: better characterization of complex genomic regions, repetitive sequences, and structural variations.
- Detection of long-range genomic information: includes intergenic regions, long-range interactions, and haplotype phasing.
- Direct detection of modifications: ONT platforms can detect DNA modifications, such as DNA methylation, directly.