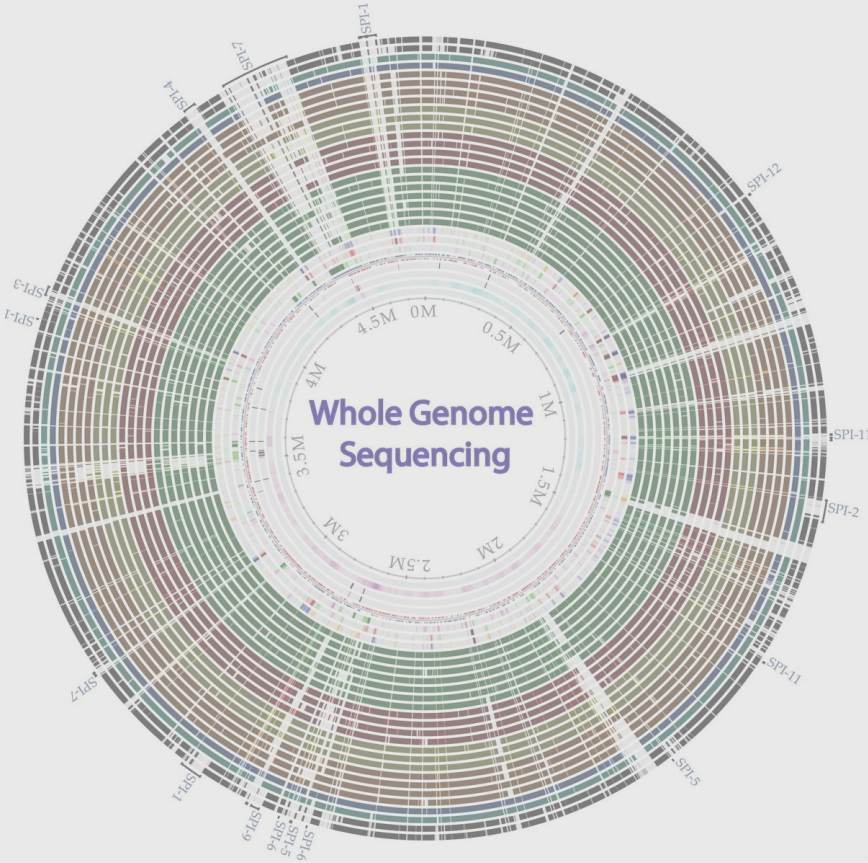
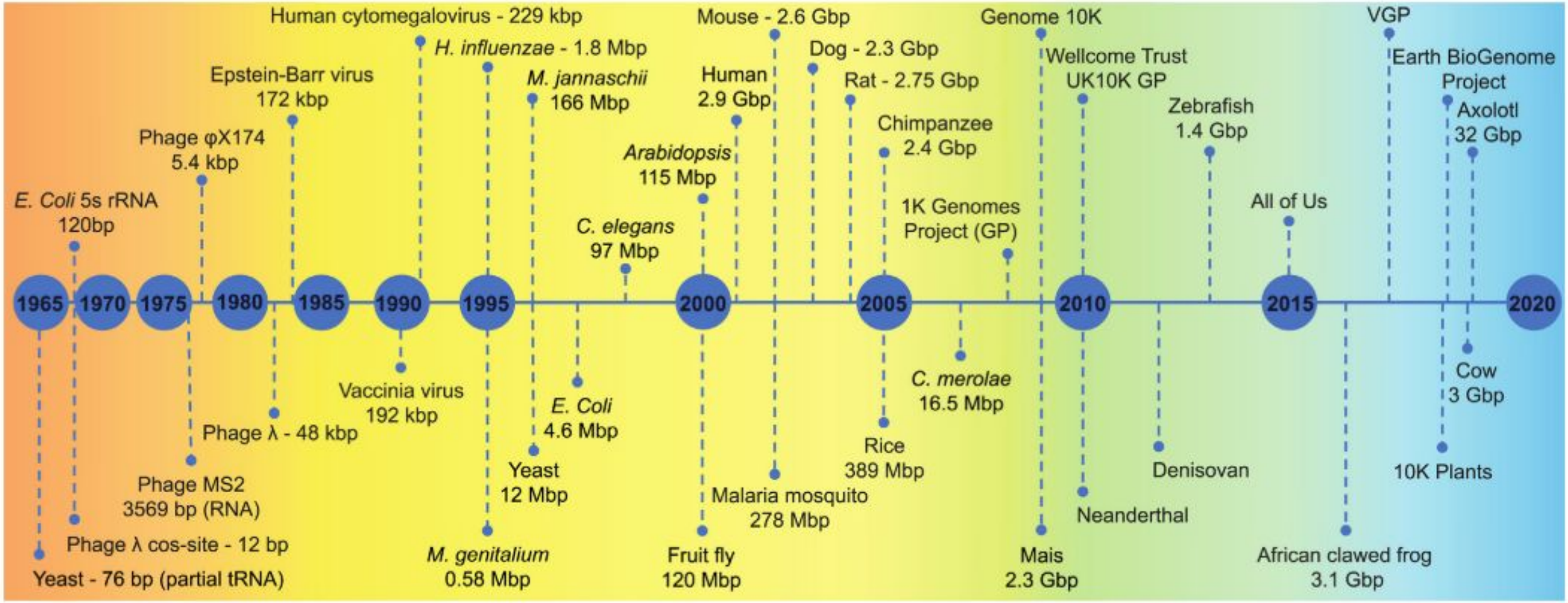
Whole-Genome Sequencing



https://www.coursera.org/learn/wgs-bacteria

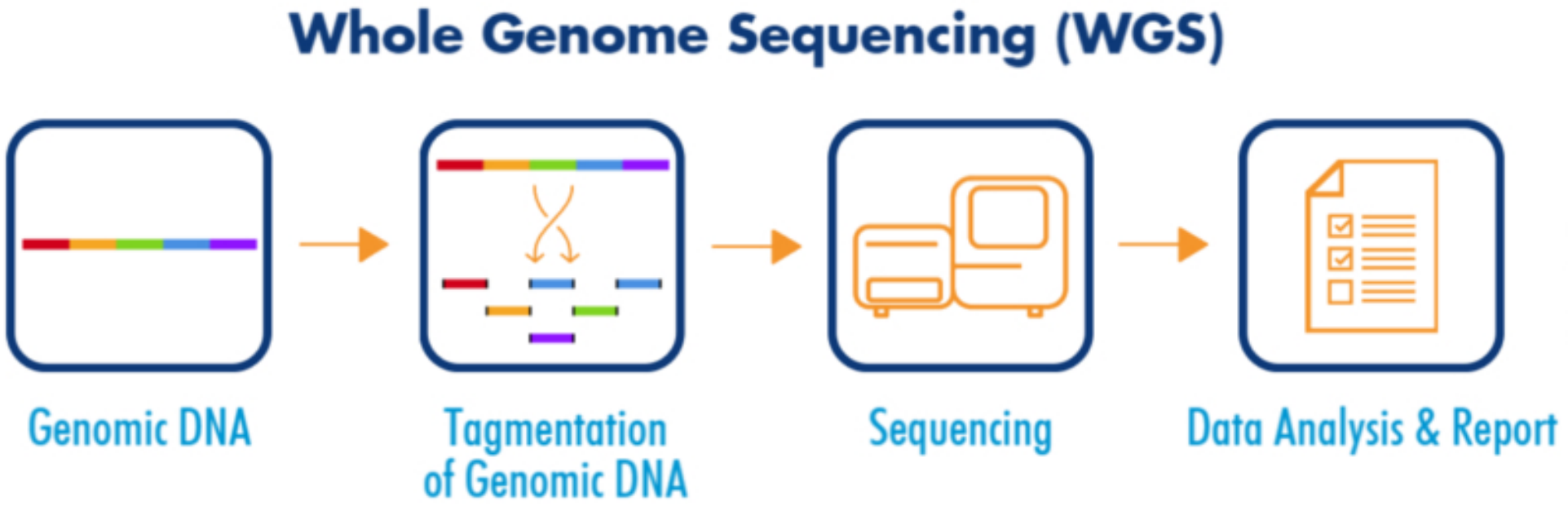
27-05-2020

Whole-Genome Sequencing

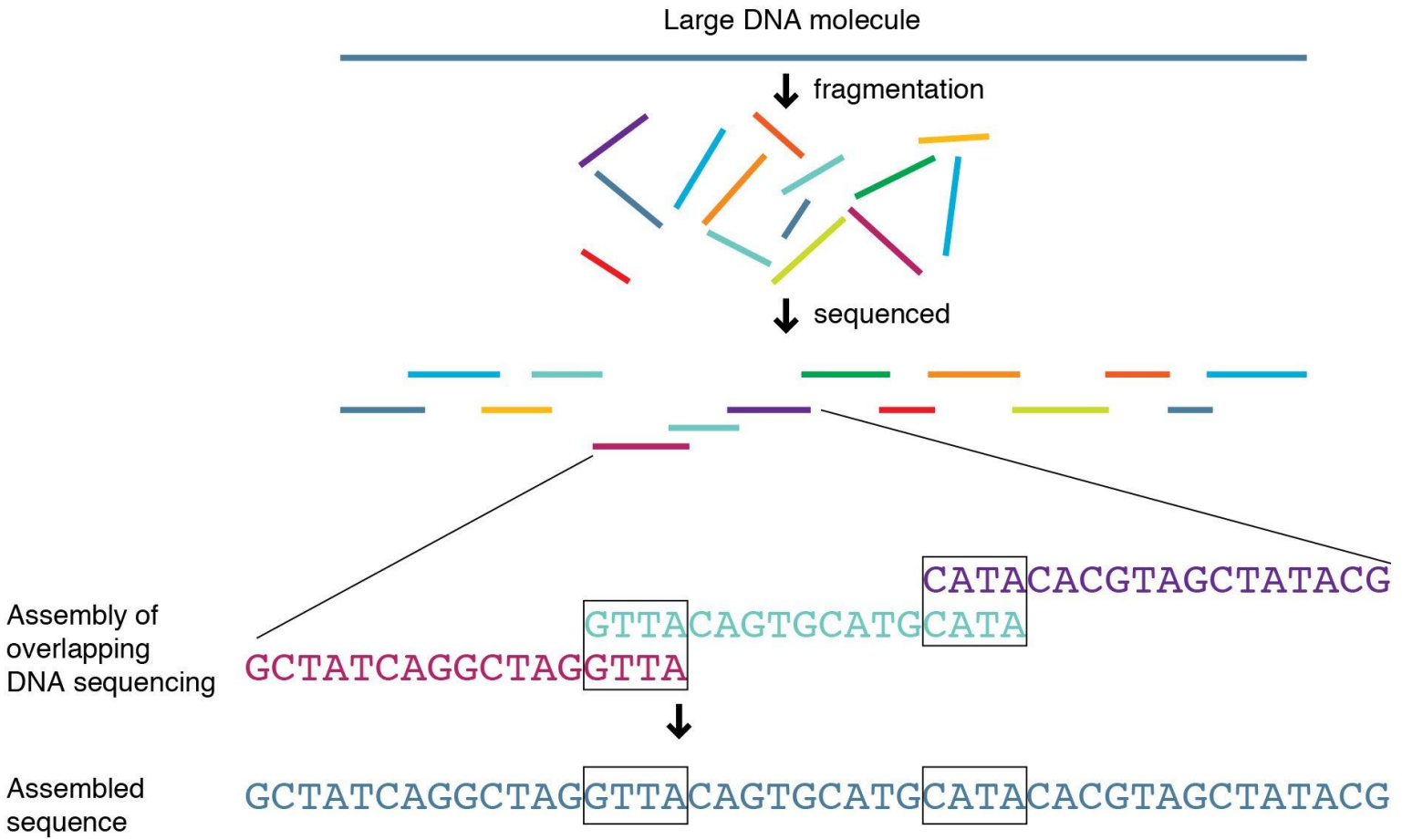


Each genome or genome project (GP) is placed under a color-coded background according to the sequencing approach adopted. Light red: early sequencing methods, Yellow: Sanger-based shotgun sequencing, Green: NGS, Light blue: TGS.

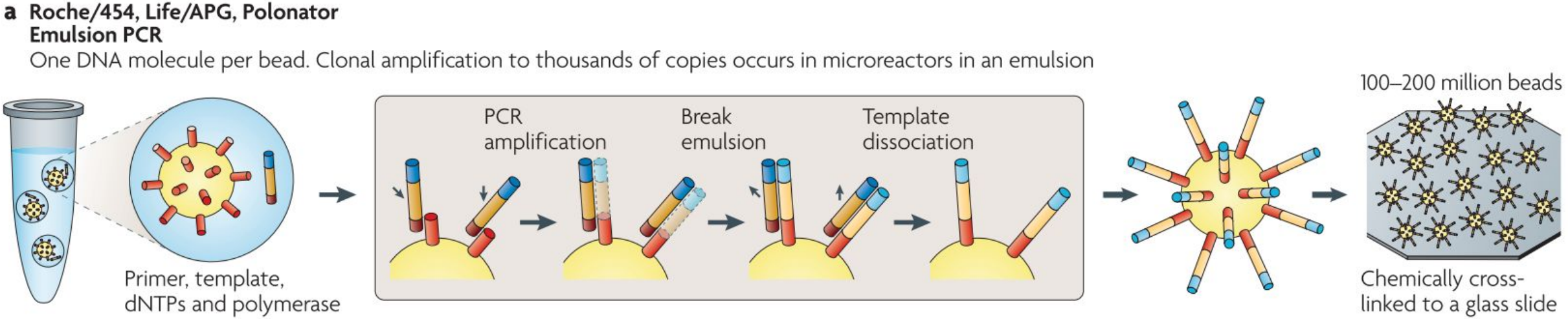
Whole-Genome Sequencing



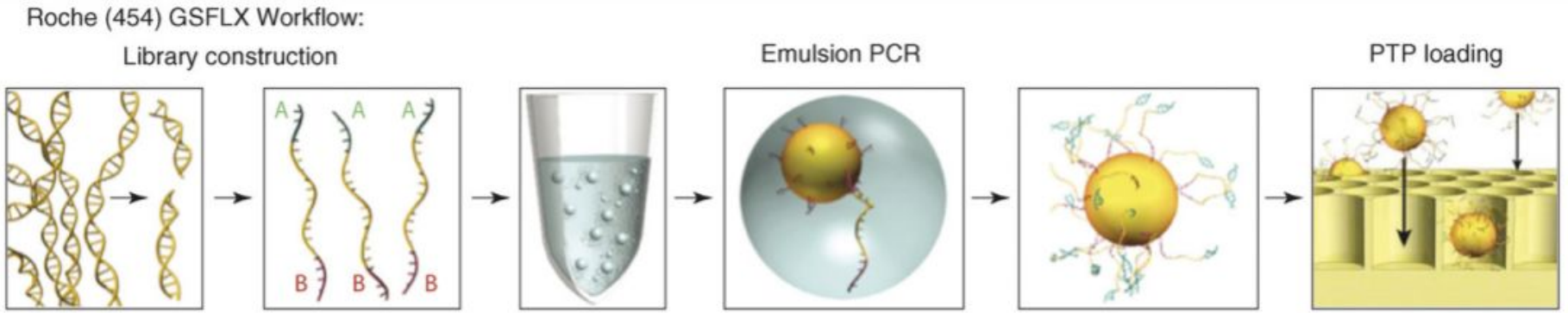
Whole-Genome Sequencing

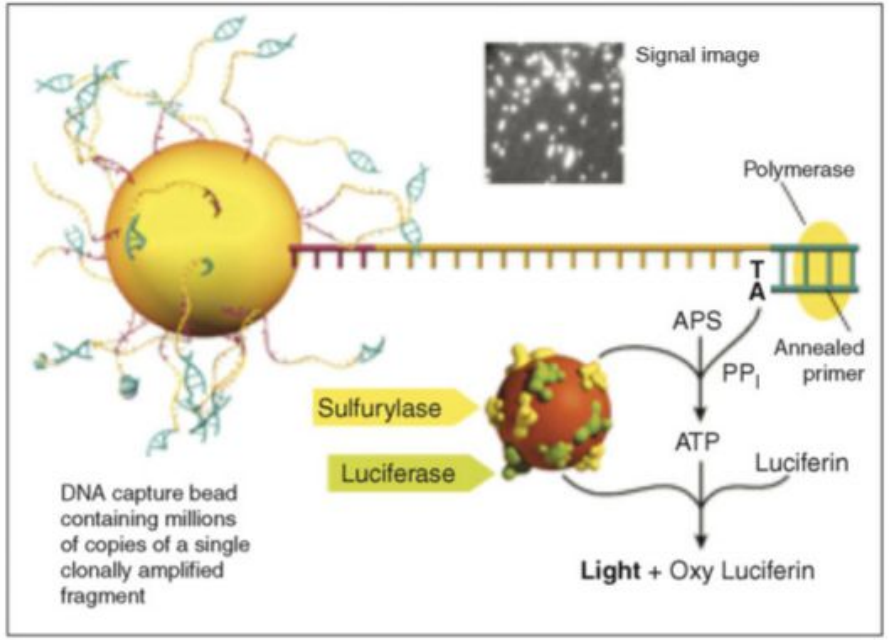


Whole-Genome Sequencing : Next Generation Sequencing



Whole-Genome Sequencing : Next Generation Sequencing

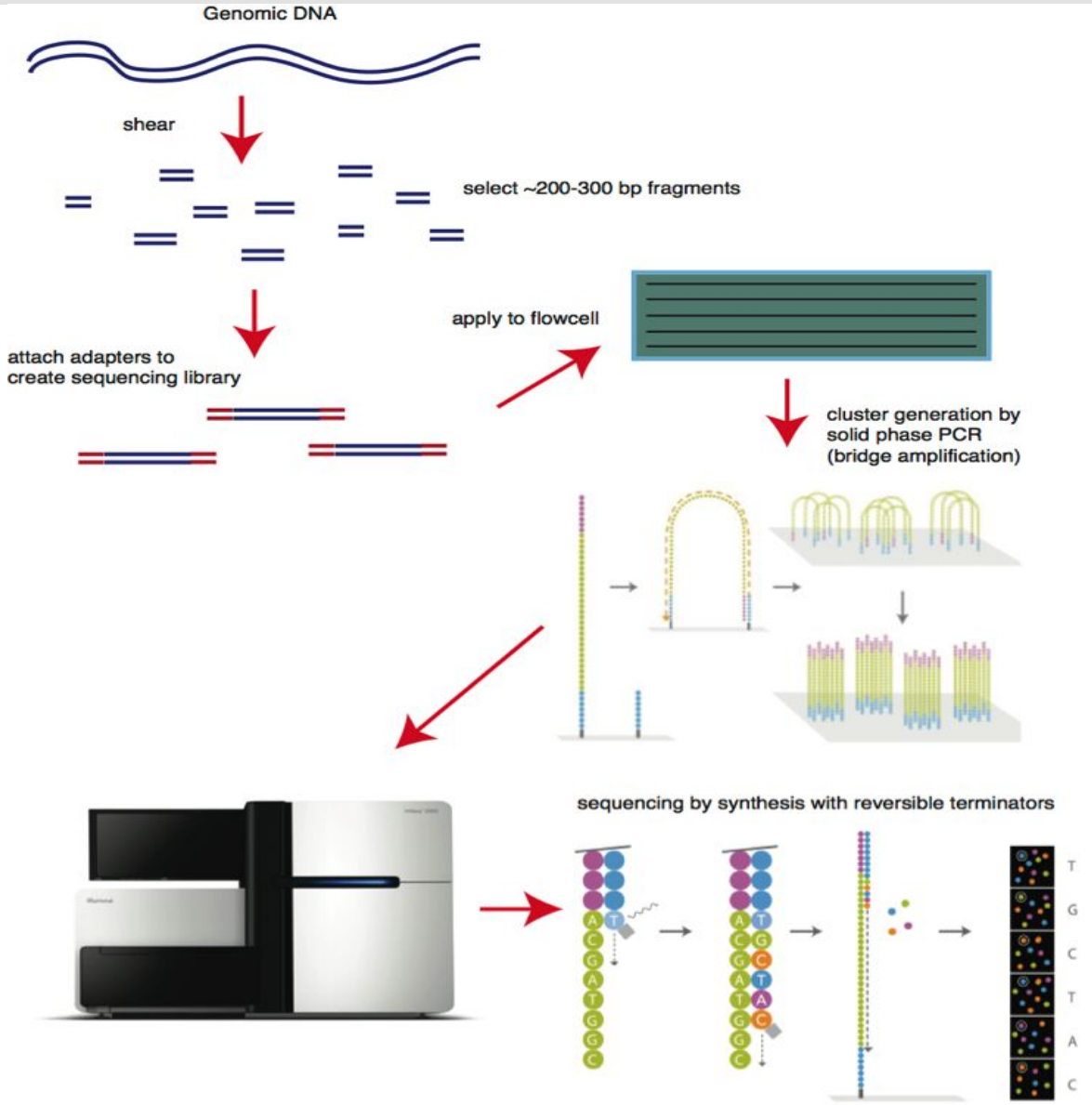




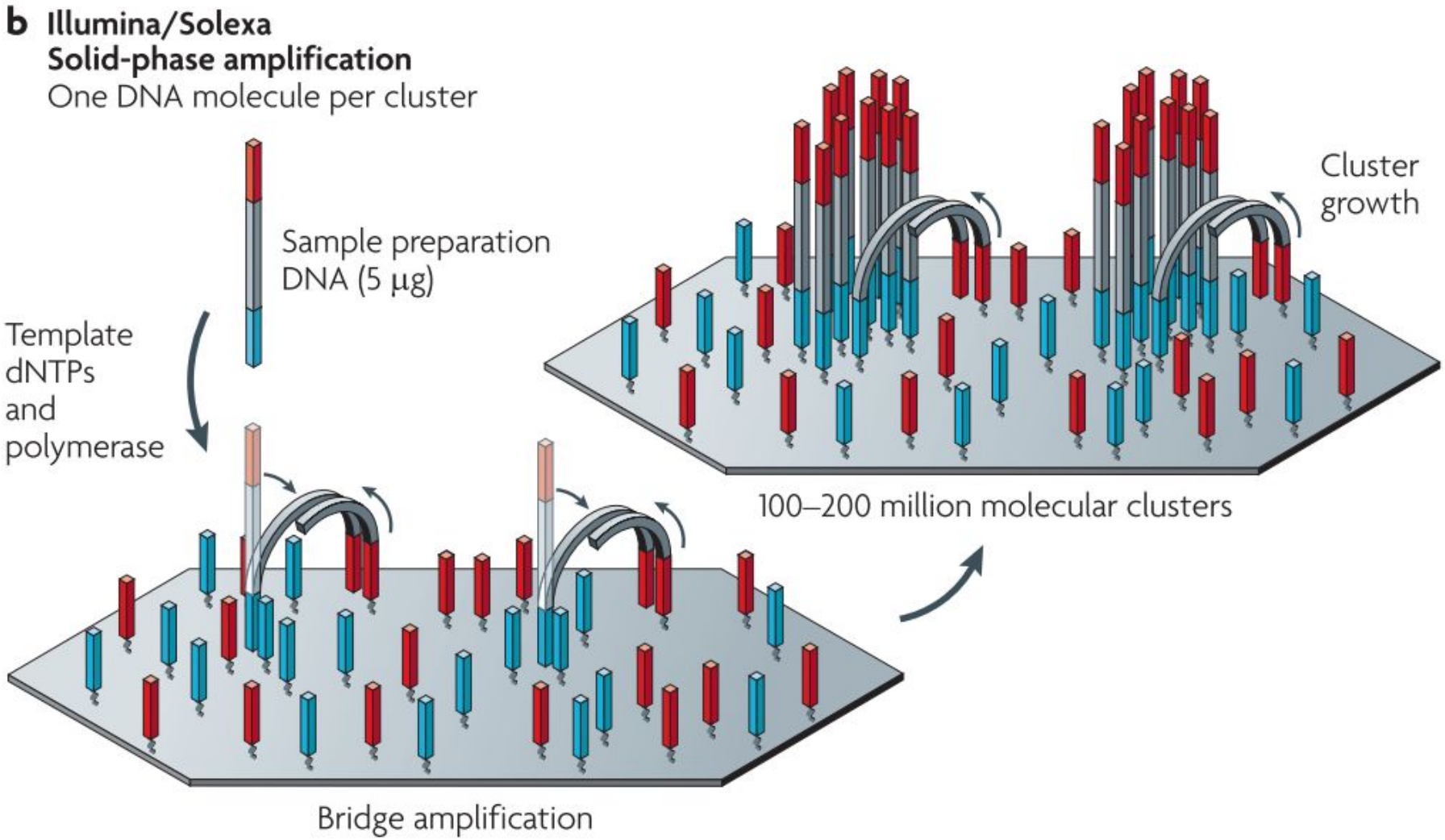
Pyrosequencing reaction

TRENDS in Genetics

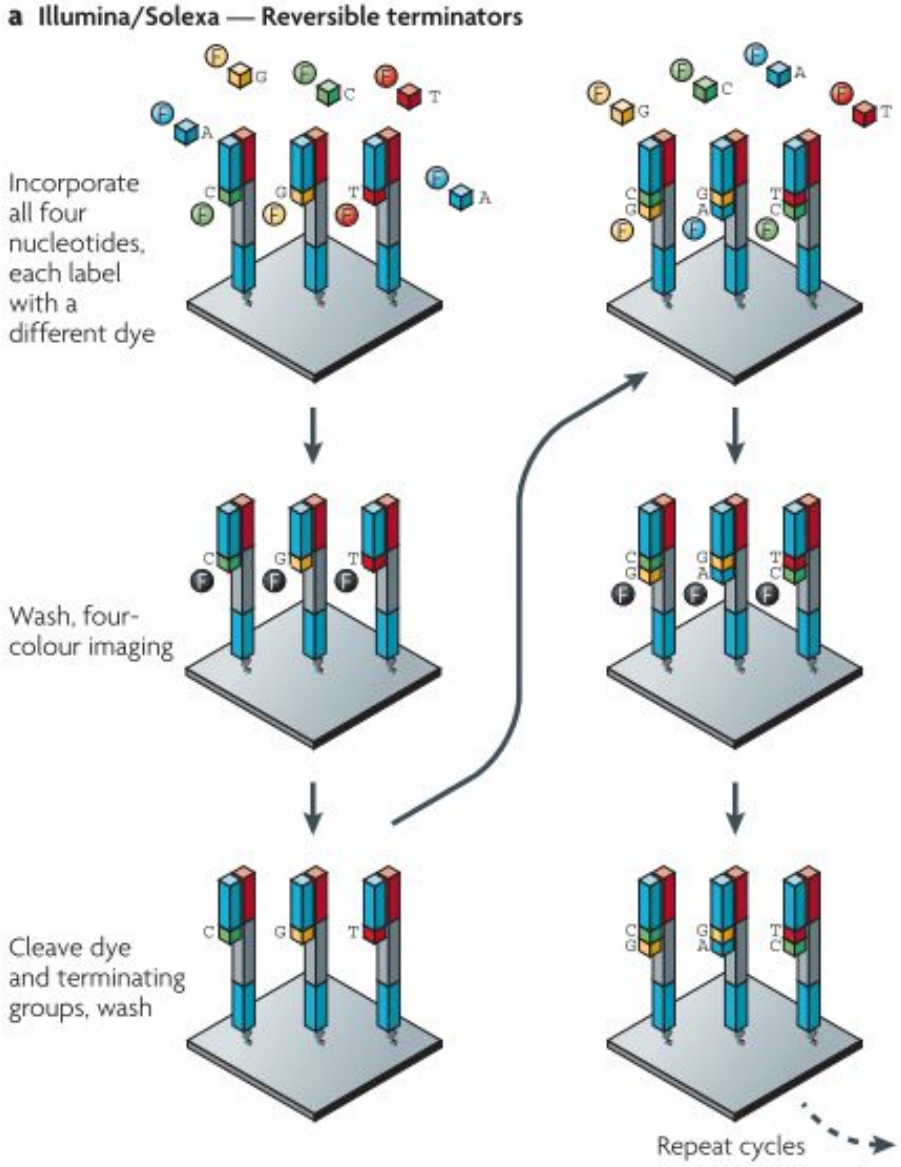
Whole-Genome Sequencing : Next Generation Sequencing

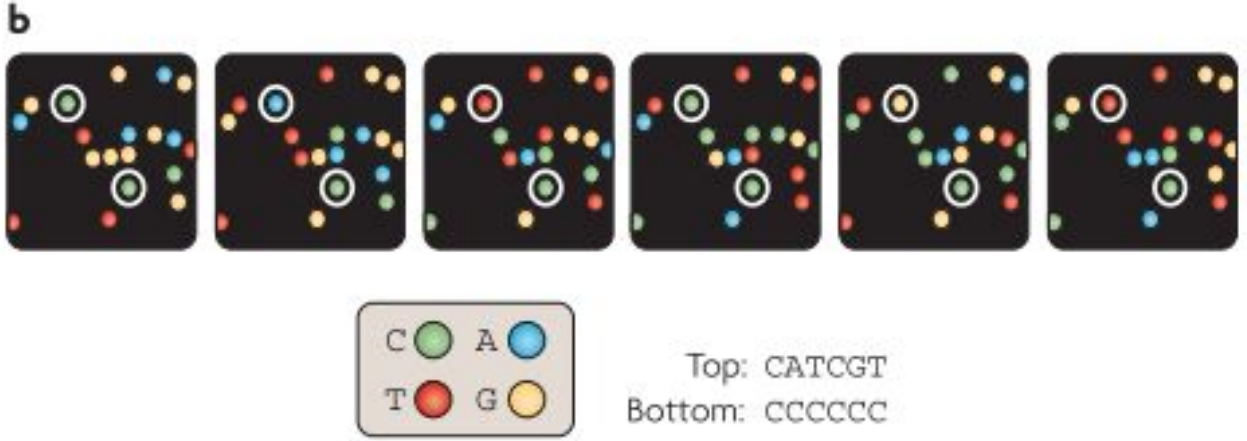


Whole-Genome Sequencing : Next Generation Sequencing



Whole-Genome Sequencing : Next Generation Sequencing





https://www.youtube.com/watch?v=fCd6B5HRaZ8

Whole-Genome Sequencing : Next Generation Sequencing

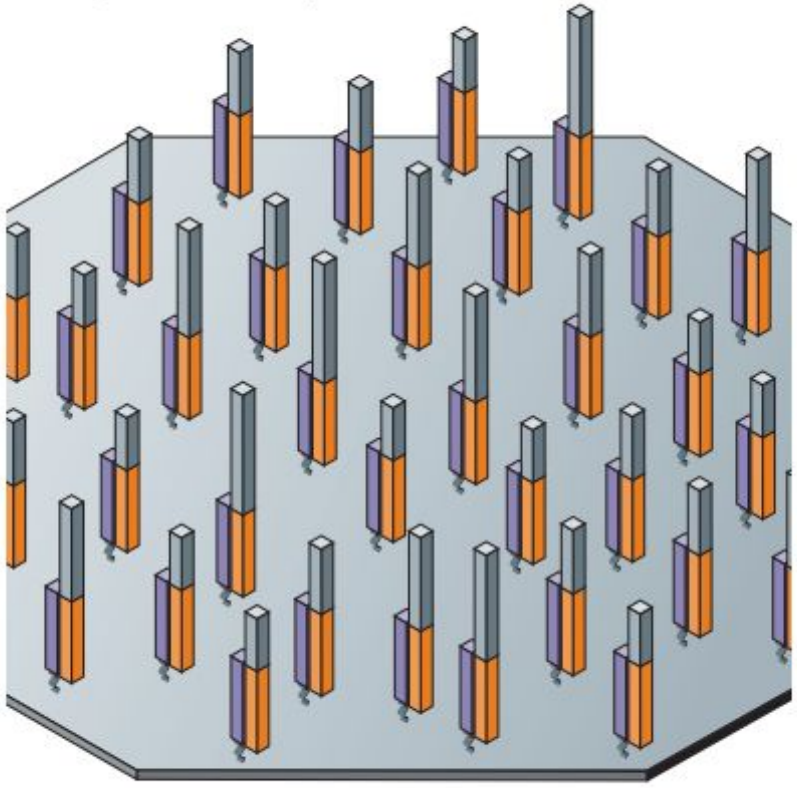
Figure 6: Sequencing Systems for Virtually Every Scale —Illumina offers innovative NGS platforms that deliver exceptional data quality and accuracy over a wide scale, from small benchtop sequencers to production-scale sequencing systems.

Whole-Genome Sequencing : Next Generation Sequencing

Feature	HiSeq2500 - Highoutput	HiSeq2500 – Rapid mode	MiSeq
Number of reads	150-180M/lane	100-150M/lane	12-15M (v2) 20-25M (v3)
Read length	2 x 100 bp	2 x 150 bp	2 x 300 bp (v3)
Yield per lane (PF data)	up to 35 Gb	up to 45Gb	up to 15 Gb
Instrument Time	~12-14 days	~2 days	~2 days
Pricing per Gb	$59 (PE100)	$53 (PE150)	$108 (PE300)

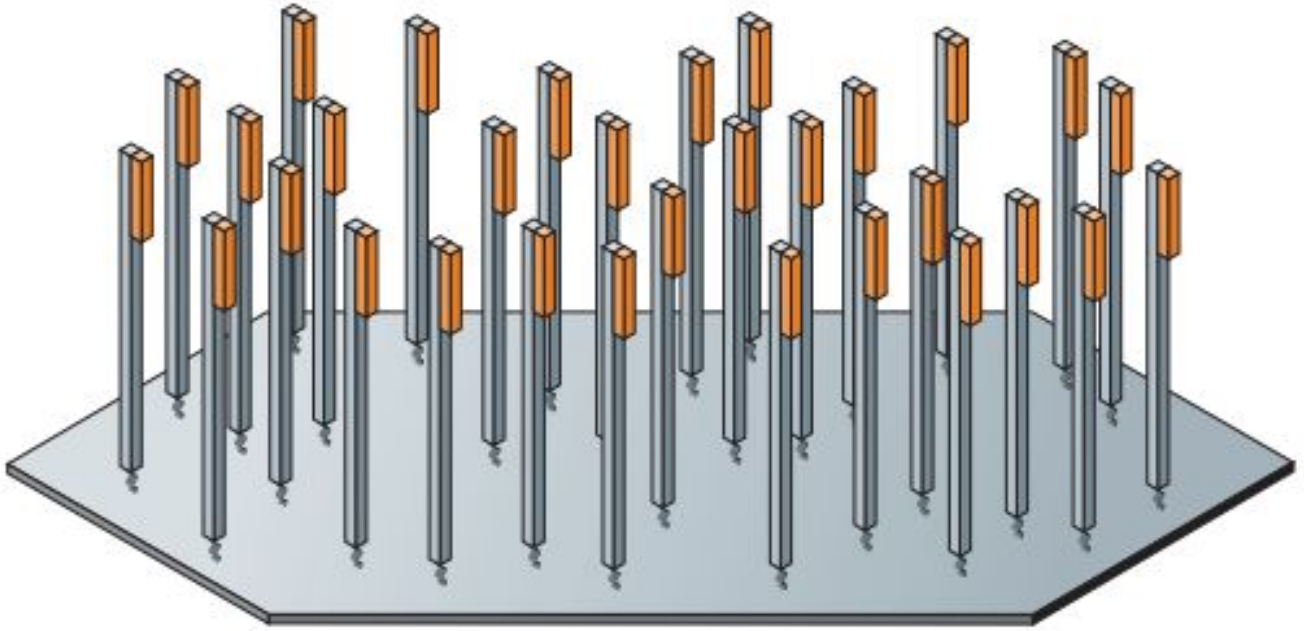
Whole-Genome Sequencing : Next Generation Sequencing

c Helicos BioSciences: one-pass sequencing
Single molecule: primer immobilized



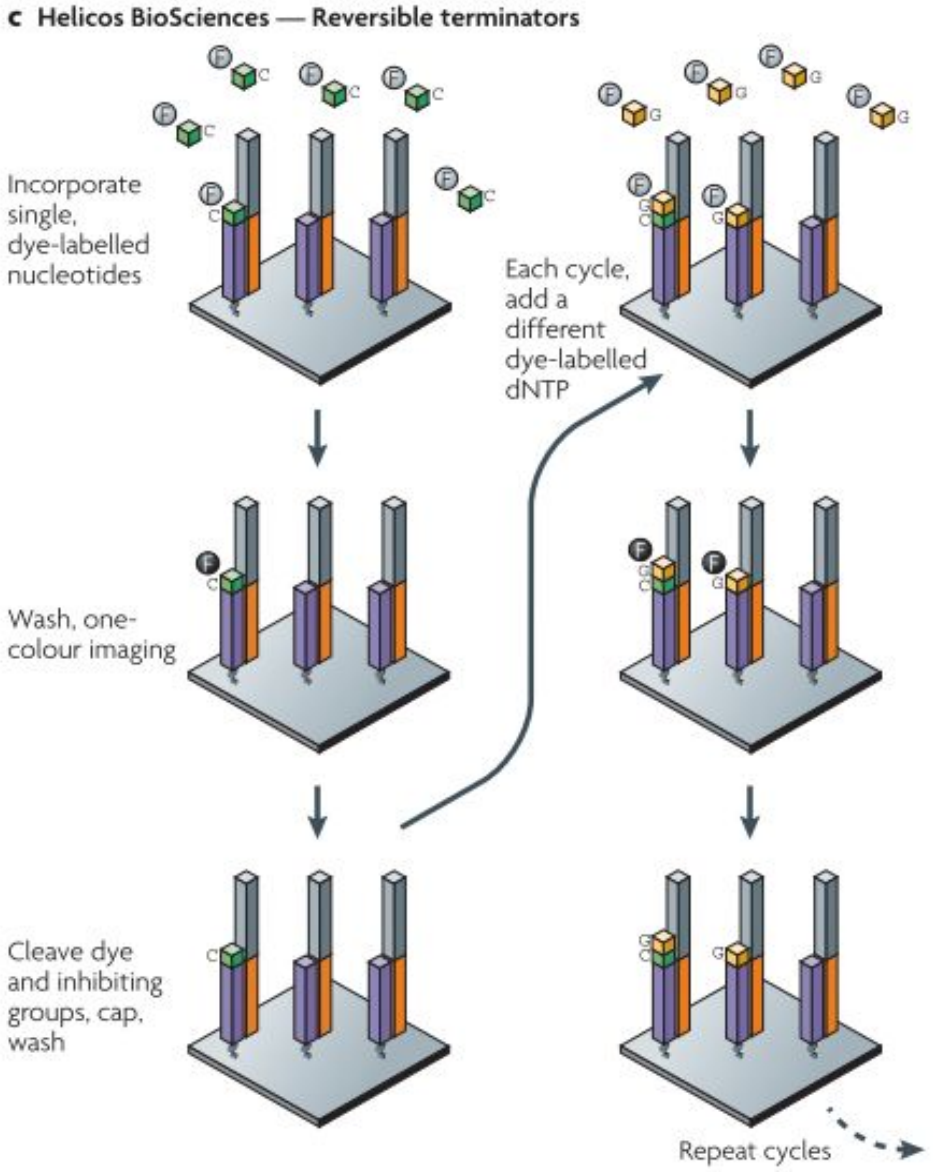
Billions of primed, single-molecule templates

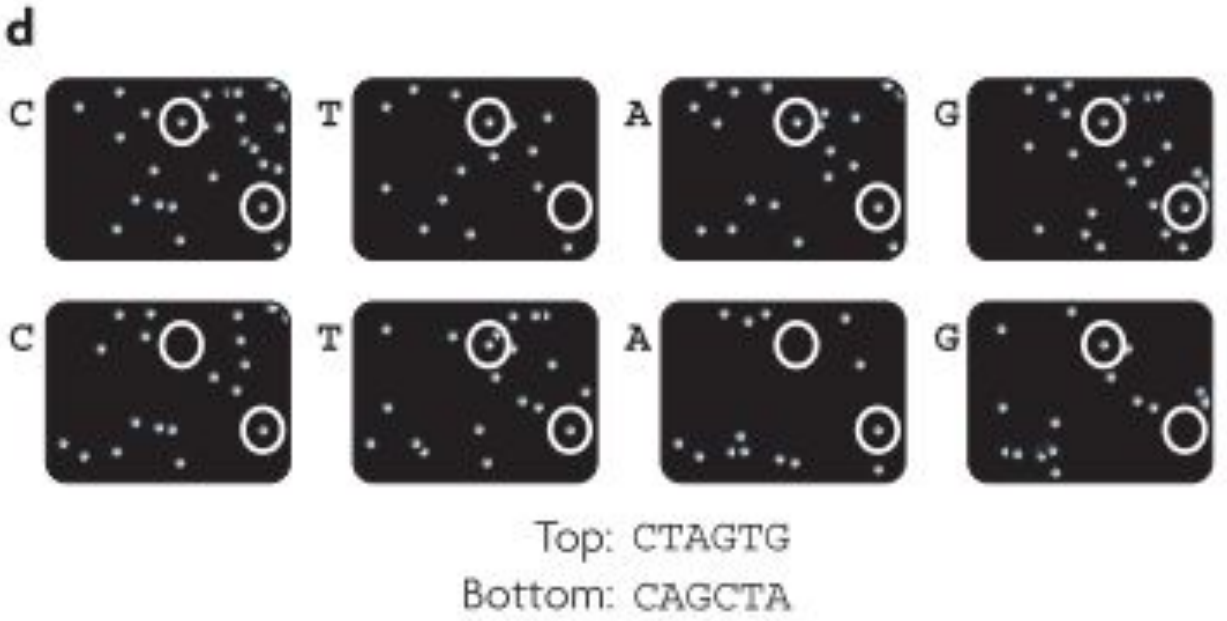
d Helicos BioSciences: two-pass sequencing
Single molecule: template immobilized



Billions of primed, single-molecule templates

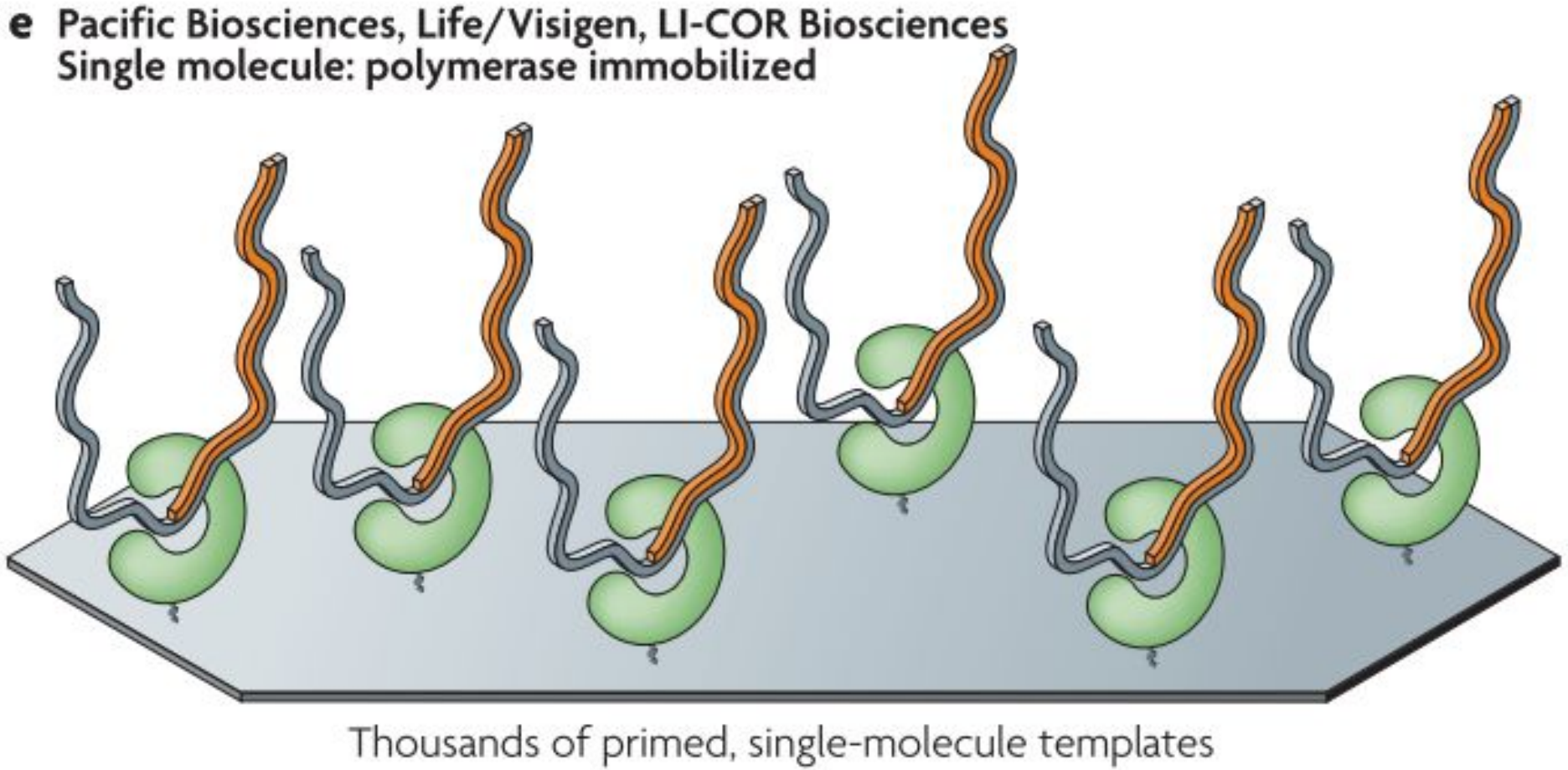
Whole-Genome Sequencing : Next Generation Sequencing





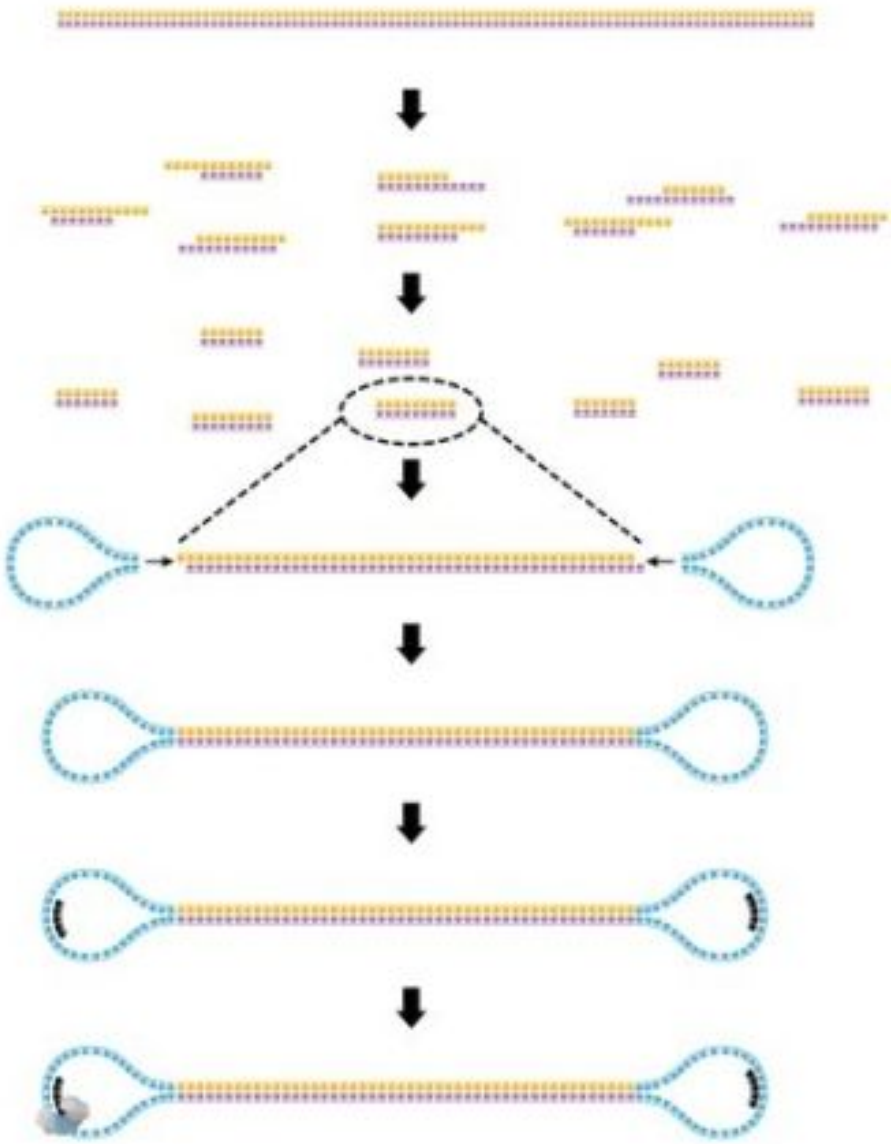
https://www.youtube.com/watch?v=73Br4kQa8j8

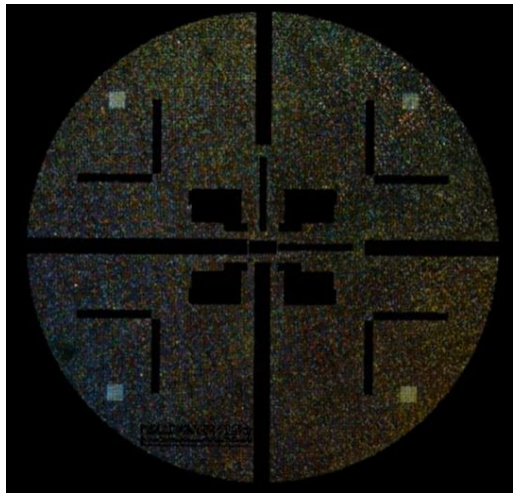
Whole-Genome Sequencing : Third Generation Sequencing



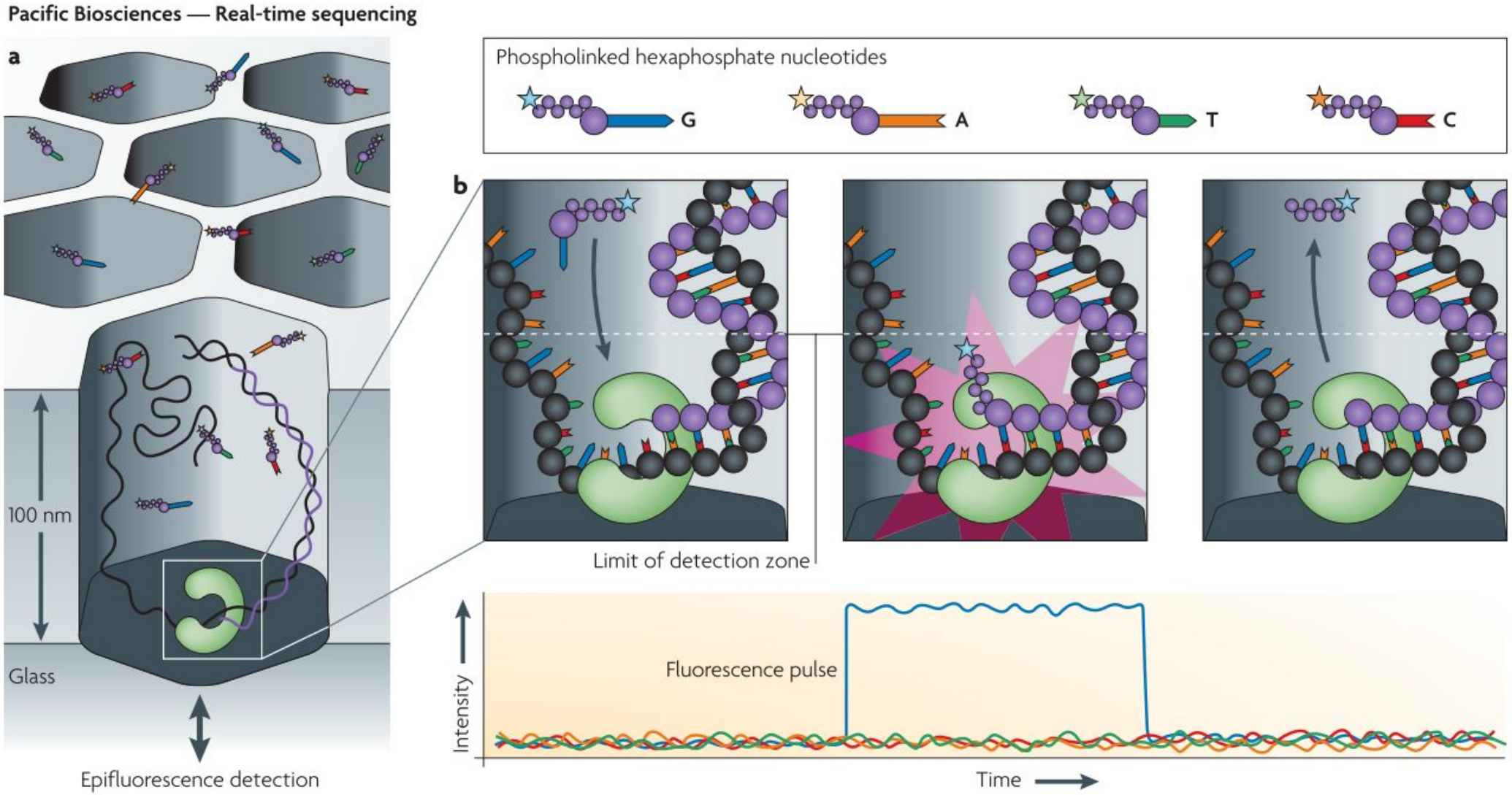
zero-mode waveguide (ZMW)

Whole-Genome Sequencing : Third Generation Sequencing

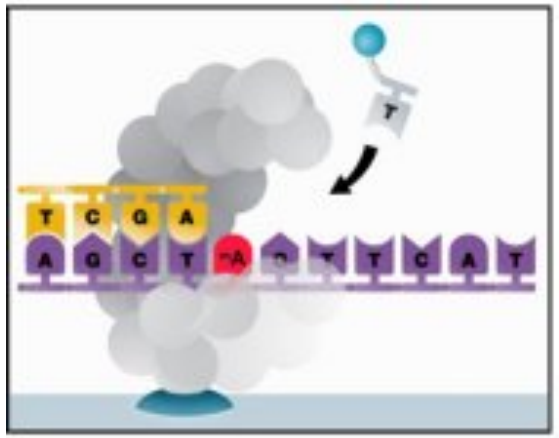


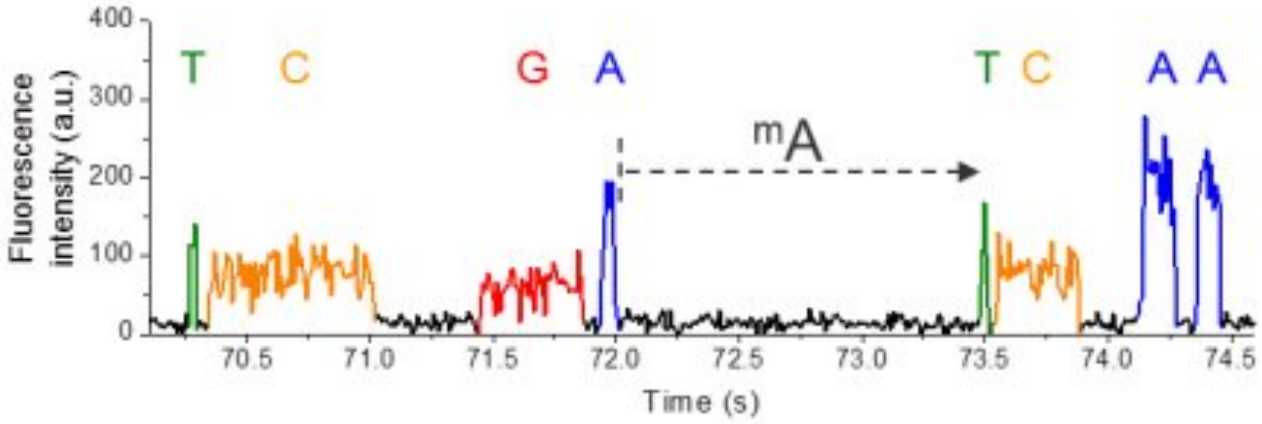


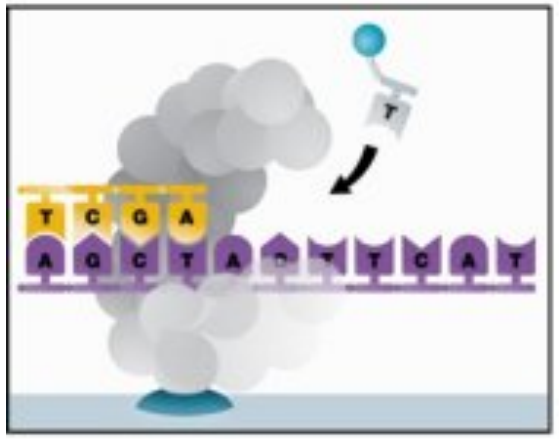
Whole-Genome Sequencing : Third Generation Sequencing

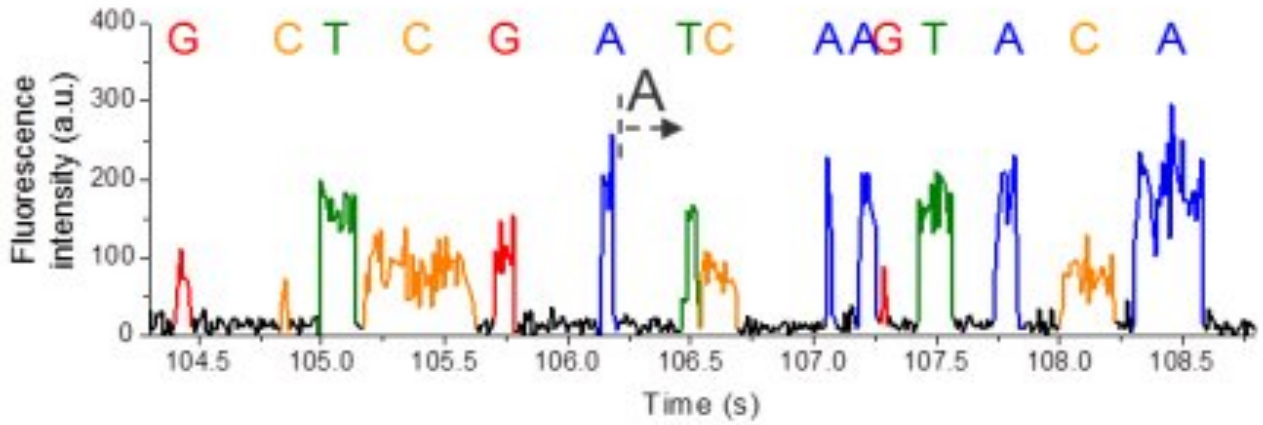


Whole-Genome Sequencing : Third Generation Sequencing



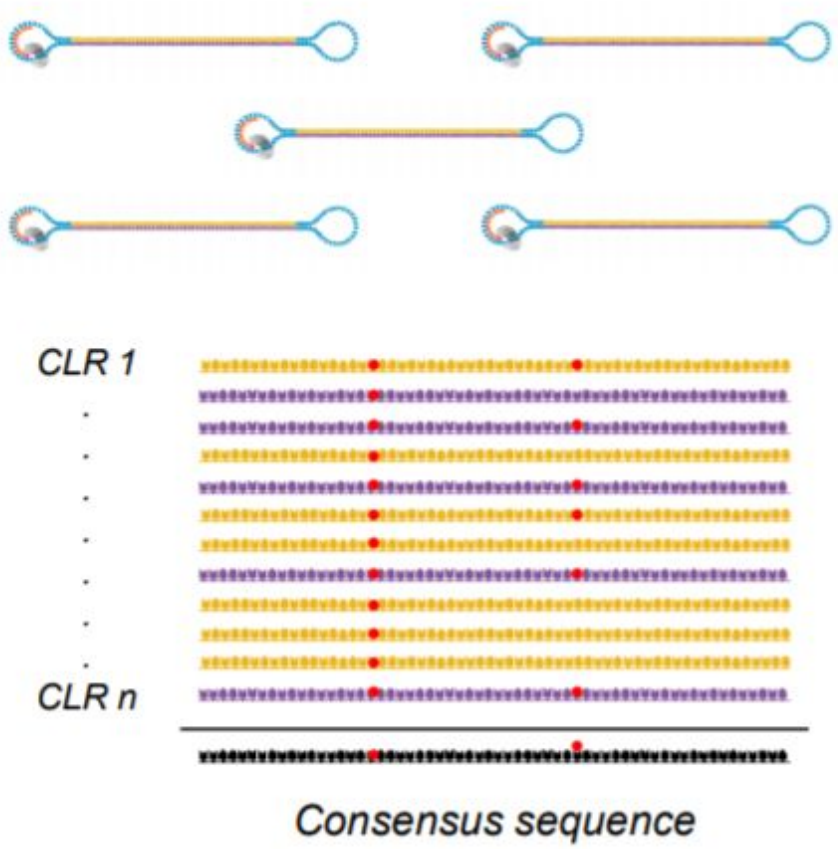






Whole-Genome Sequencing : Third Generation Sequencing

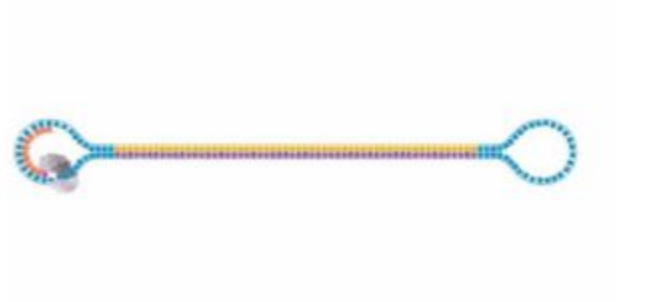
- Continuous *Long Read* Sequencing (CLR)

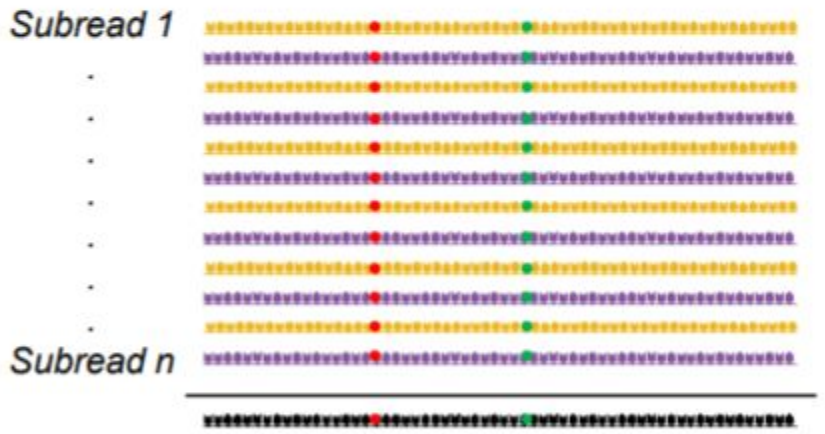


Generate reads in ten's of kilobases

Whole-Genome Sequencing : Third Generation Sequencing

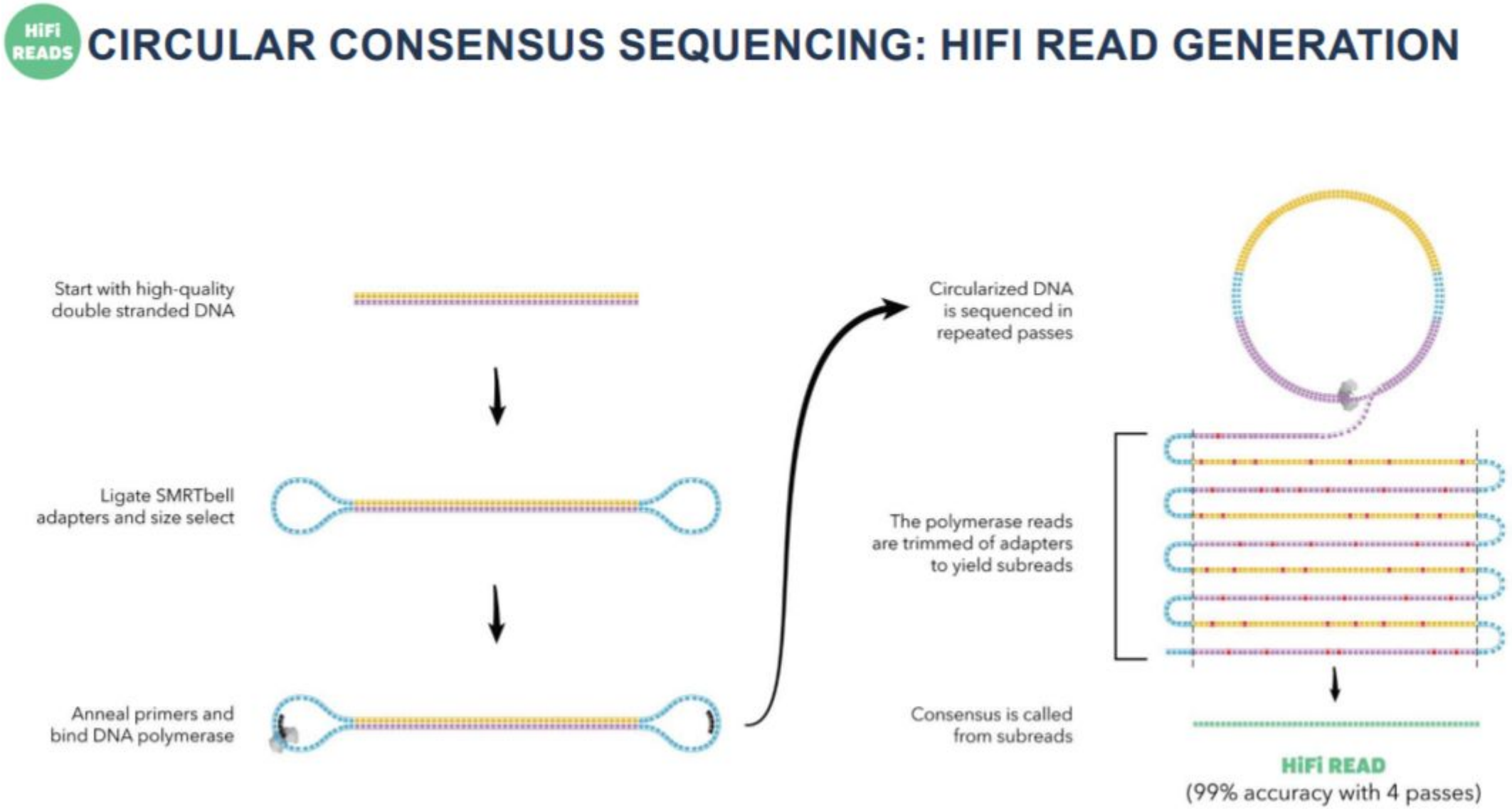
- Circular Consensus Sequencing (CCS)



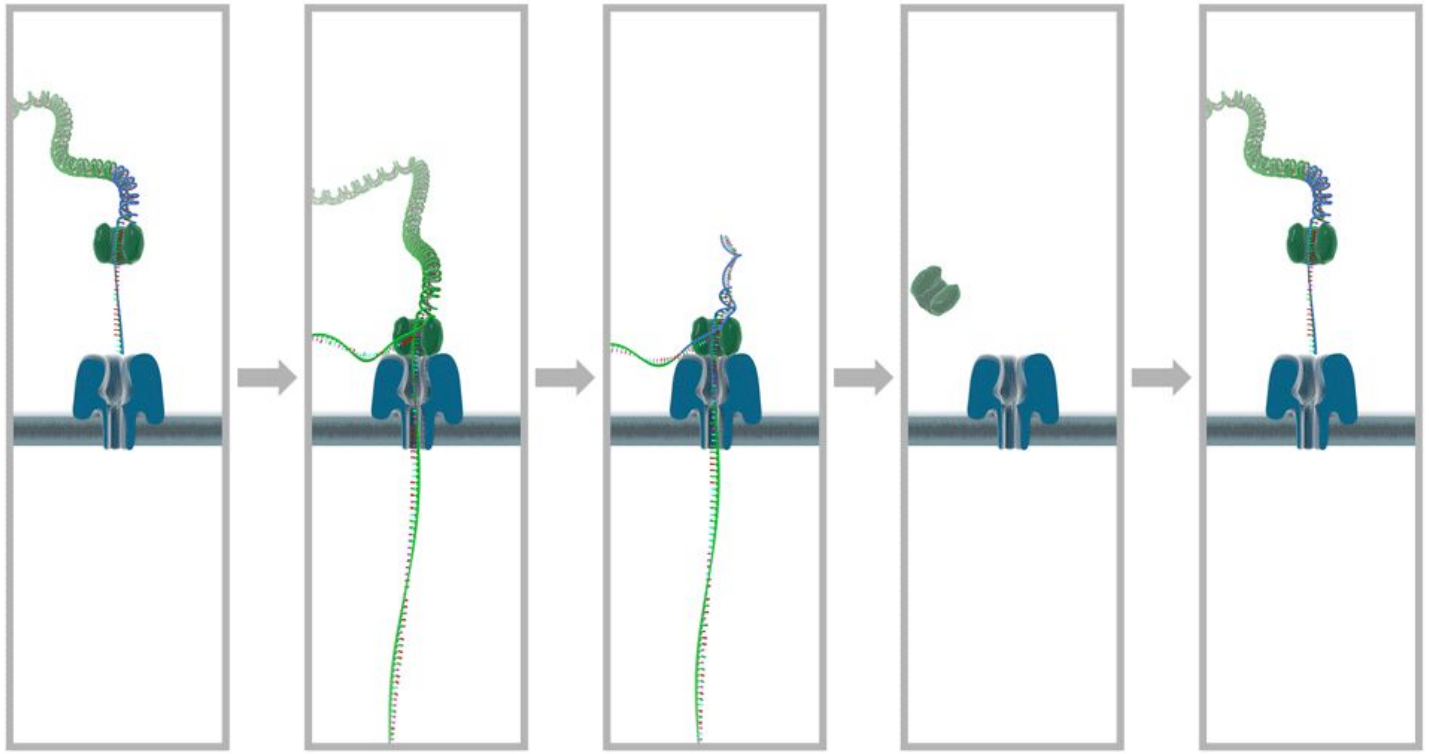


Generate high quality reads 1 – 20 kb

Whole-Genome Sequencing : Third Generation Sequencing

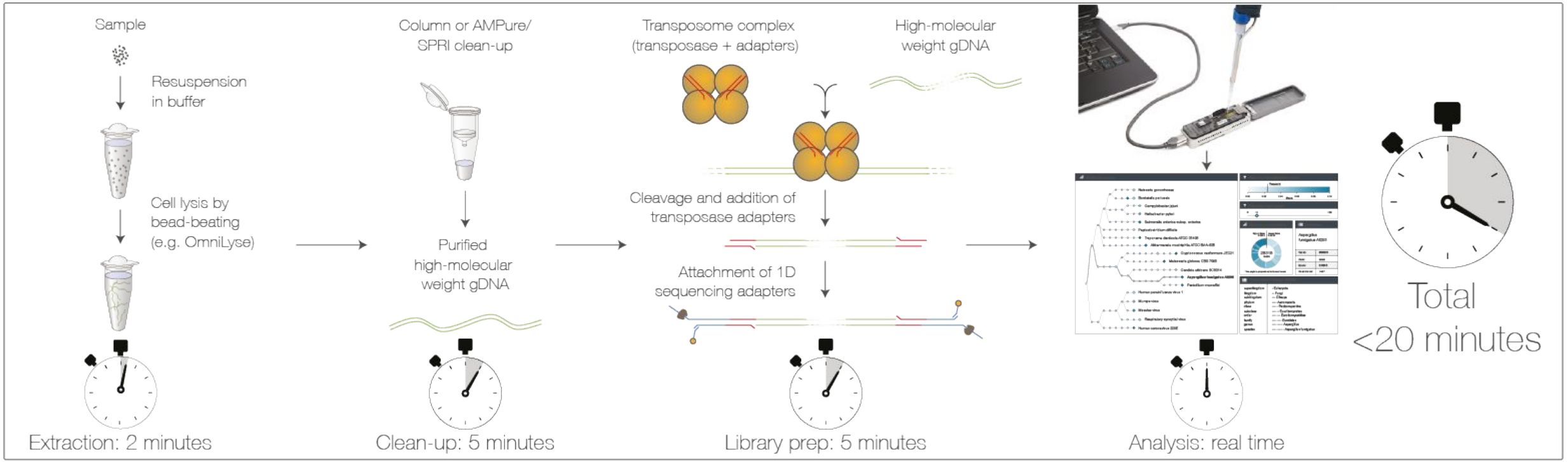


Whole-Genome Sequencing : Third Generation Sequencing



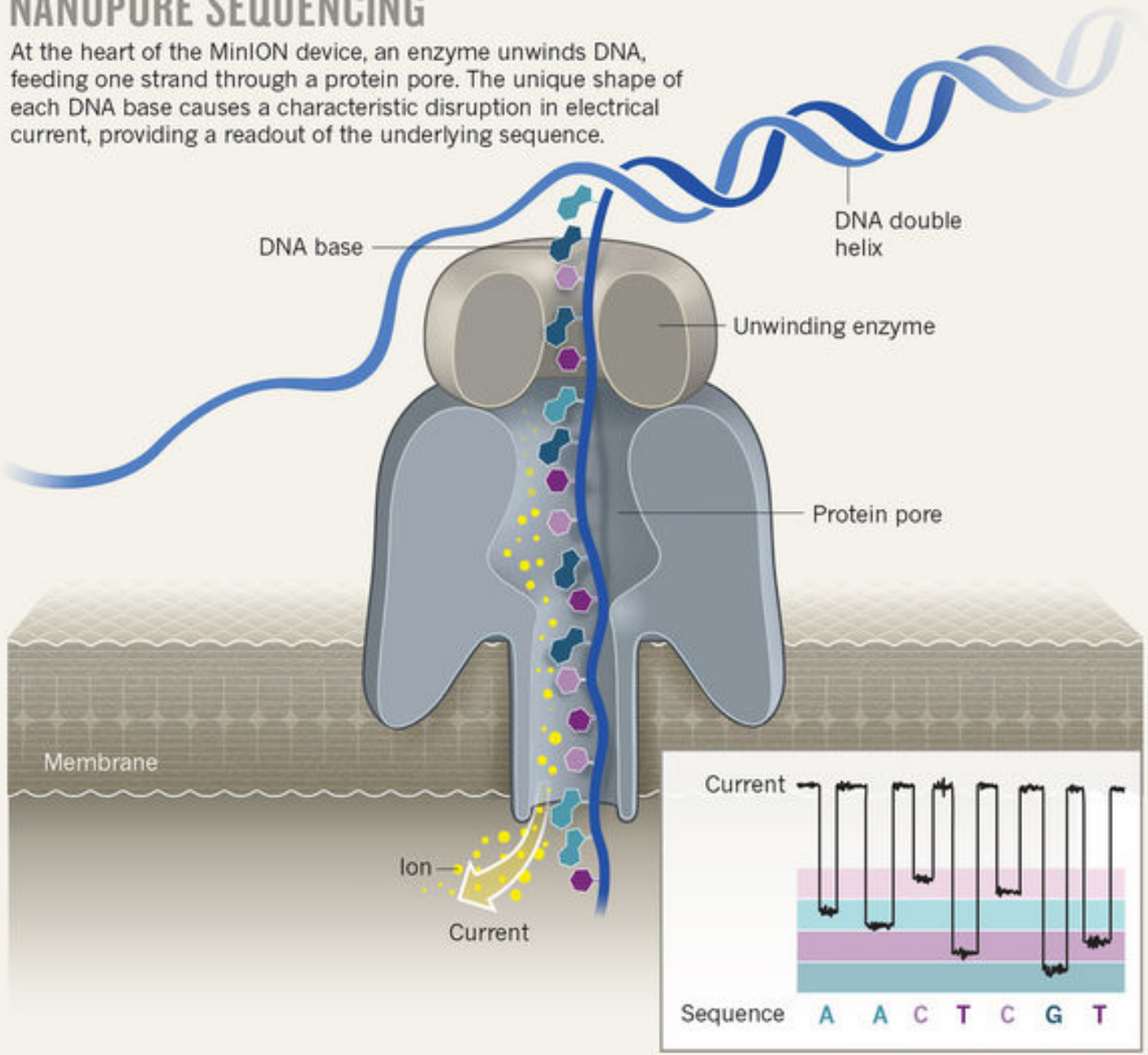
https://www.youtube.com/watch?v=E9-Rm5AoZGw

Whole-Genome Sequencing : Third Generation Sequencing

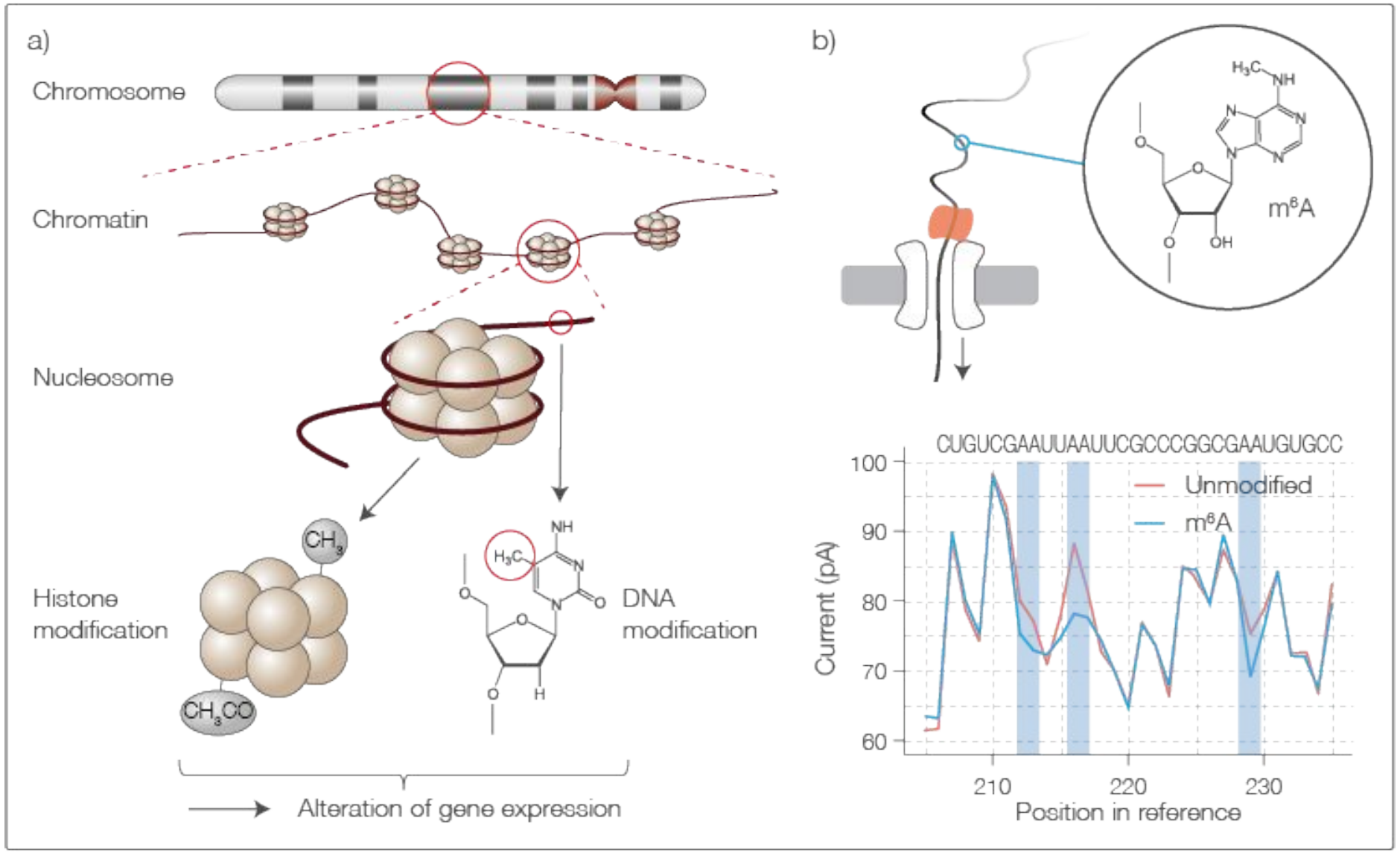


NANOPORE SEQUENCING

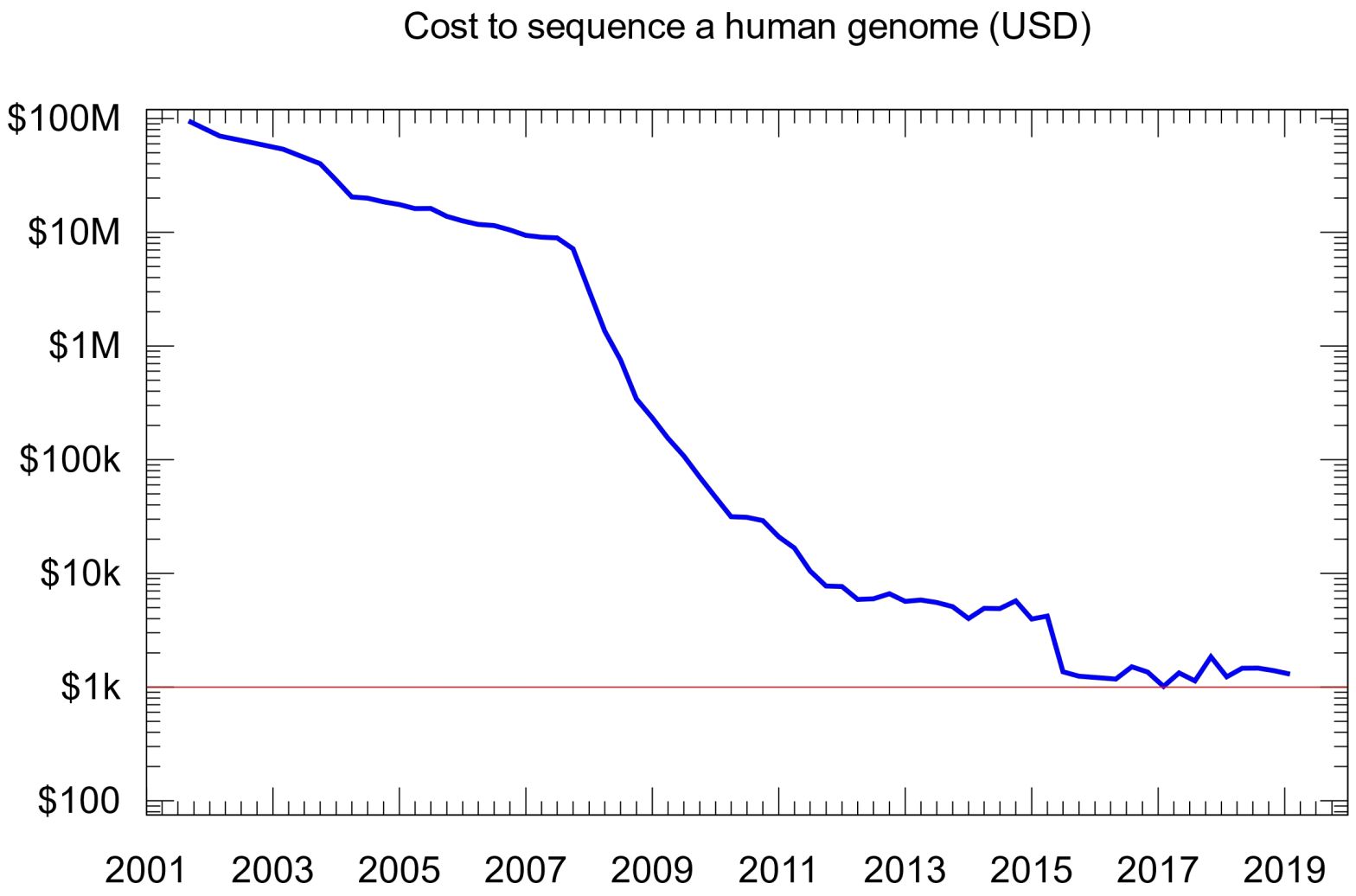
At the heart of the MinION device, an enzyme unwinds DNA, feeding one strand through a protein pore. The unique shape of each DNA base causes a characteristic disruption in electrical current, providing a readout of the underlying sequence.



Whole-Genome Sequencing : Third Generation Sequencing



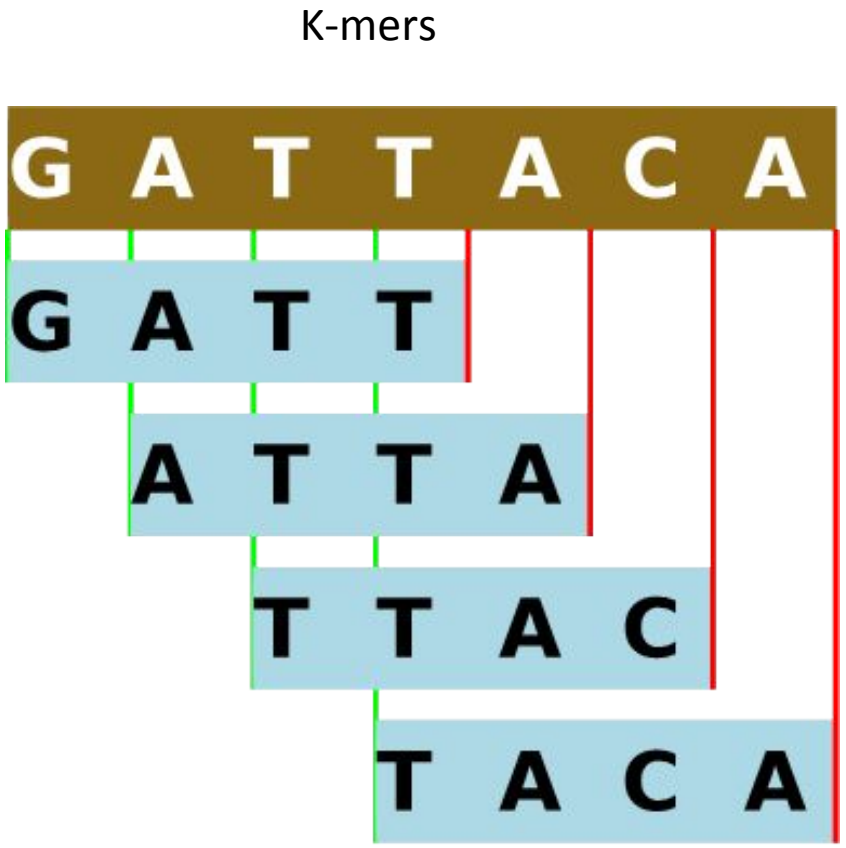
Whole-Genome Sequencing – Associated costs



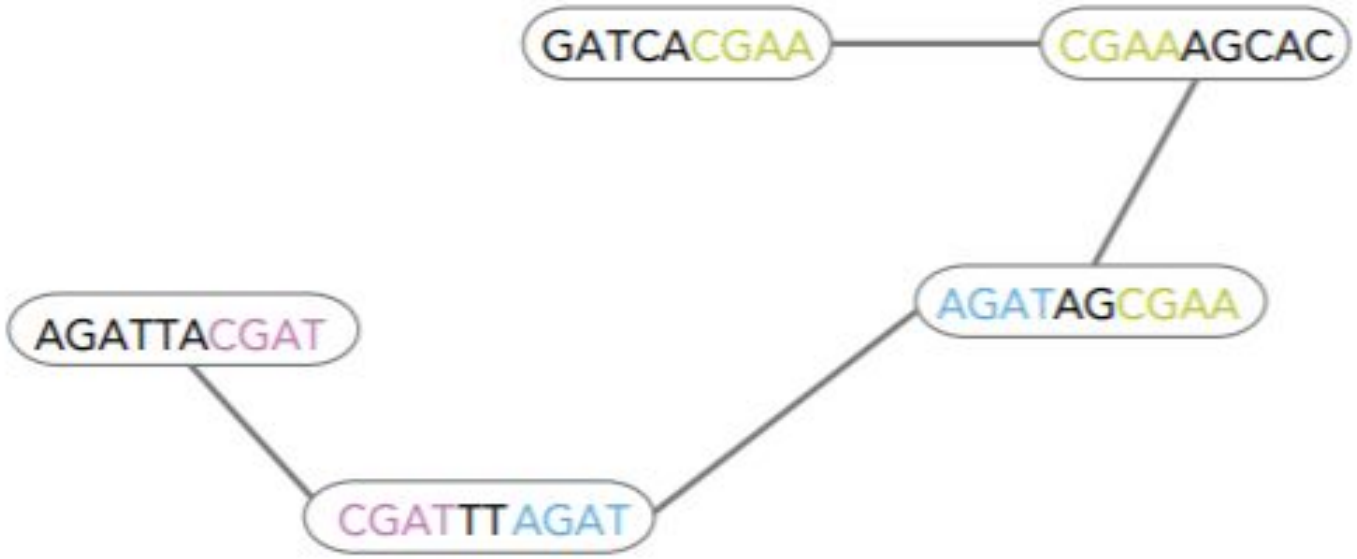
Total cost of sequencing a human genome over time as calculated by the NHGRI (NIH)

Whole-Genome Sequencing – Assembly

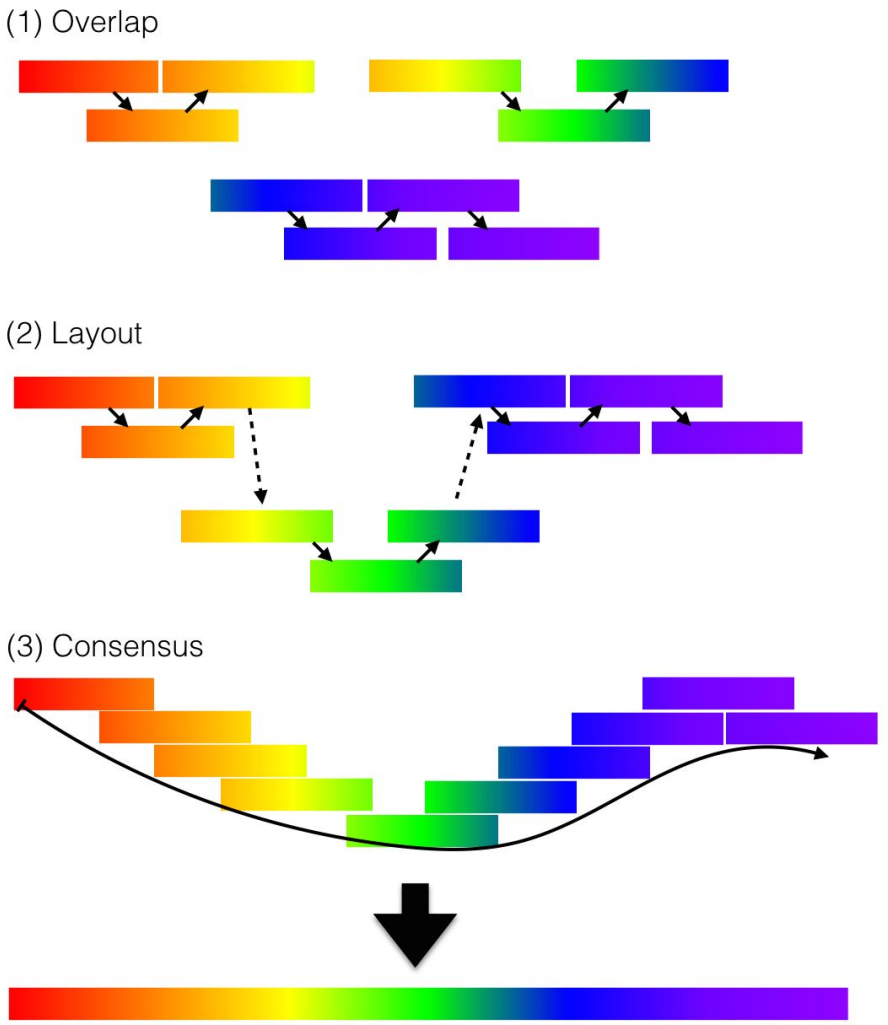
- Assembly:
  - Overlay Layout Consensus (OLC)
  - De Bruijn Graph



Whole-Genome Sequencing – OLC



Whole-Genome Sequencing – OLC

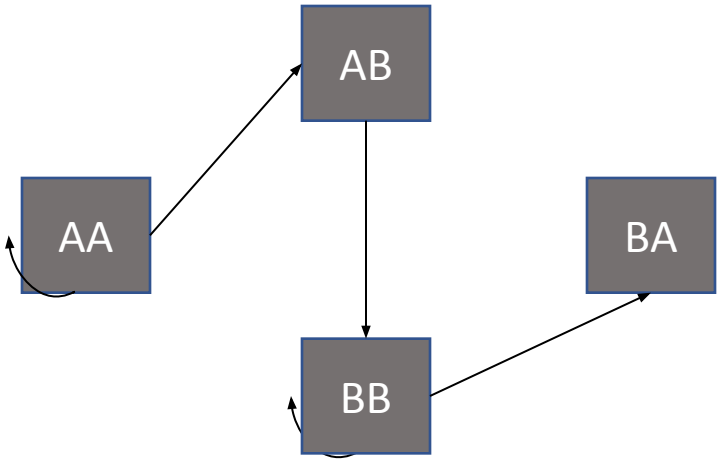


Whole-Genome Sequencing – de Bruijn graphs

Genome: **AAABBBA**

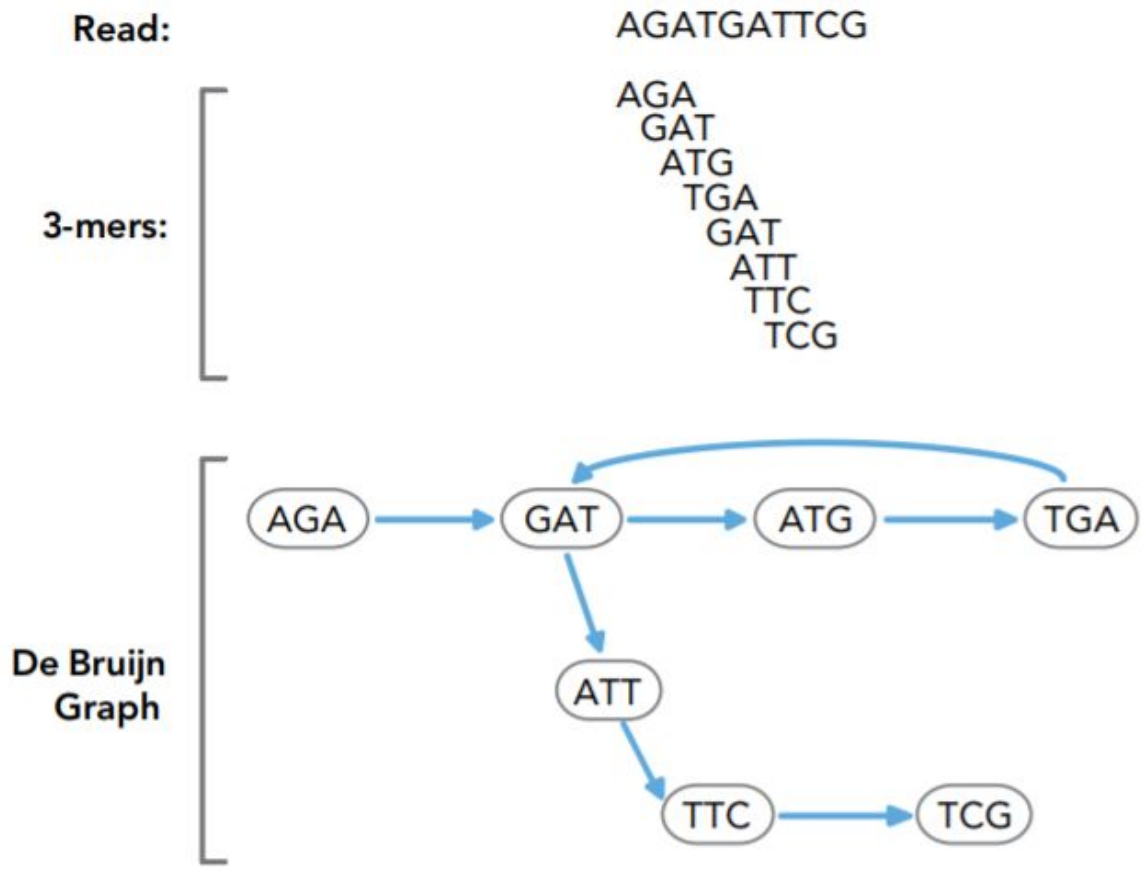
3-mers: AAA, AAB, ABB, BBB, BBA

L/R 2-mers: AA-AA, AA-AB, AB-BB, BB-BB, BB-BA



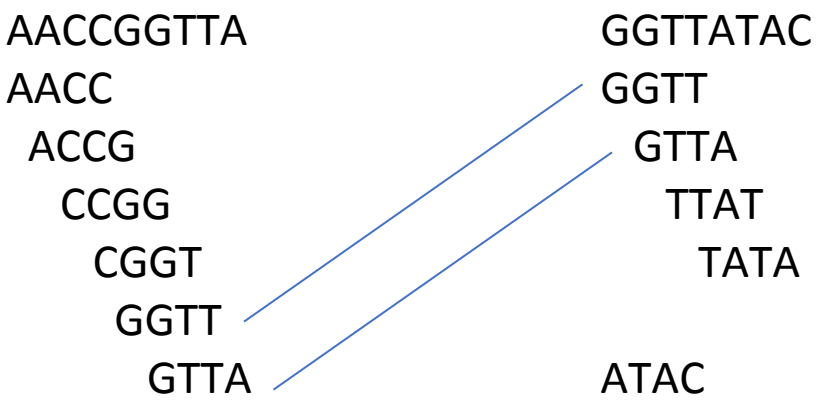
Eulerian walks

Whole-Genome Sequencing – de Bruijn graphs



The length of overlaps is k-1=2. Gray arrows indicate where all the k-mers derived from the one read are placed in the graph. Blue arrows indicate the order of the k-mers and their overlaps.

Whole-Genome Sequencing – de Bruijn graphs



AACCGGTTATA
C

Whole-Genome Sequencing – Assembly

Table 1: Overview of Tested Assemblers

Algorithm	Description	Strength	Genomes Assembled
Velvet	De Bruijn graph based Error corrections after graph is built	Fast (~30 mins) Easy to use Larger supercontig N50	Bacterial (Ref. 1; this technical note)
SOAPdenovo	De Bruijn graph based Error correction before graph is built	Easy to use Multi-threaded mode	Panda, Bacterial (Ref. 11; this technical note)
ABySS	De Bruijn graph based Can be run in parallel Distributed memory model (efficient)	Easy to use Largest contigs/scaffolds Best suited for large genomes	Human (Ref. 3; this technical note)
Forge	Overlap-layout-consensus method Modifications to accommodate Illumina reads	Largest contigs/supercontigs Good "long read" assembler	Bacterial (this technical note)

Whole-Genome Sequencing – parameters

Table 2: Effect of Coverage on Assembly Quality

Coverage	N50 contig size	Largest contig	Genome coverage
320×	95,313 bp	215,645 bp	99.47%
160×	95,368 bp	209,234 bp	99.72%
50×	97,333 bp	223,793 bp	99.72%
21×	35,828 bp	119,071 bp	99.38%

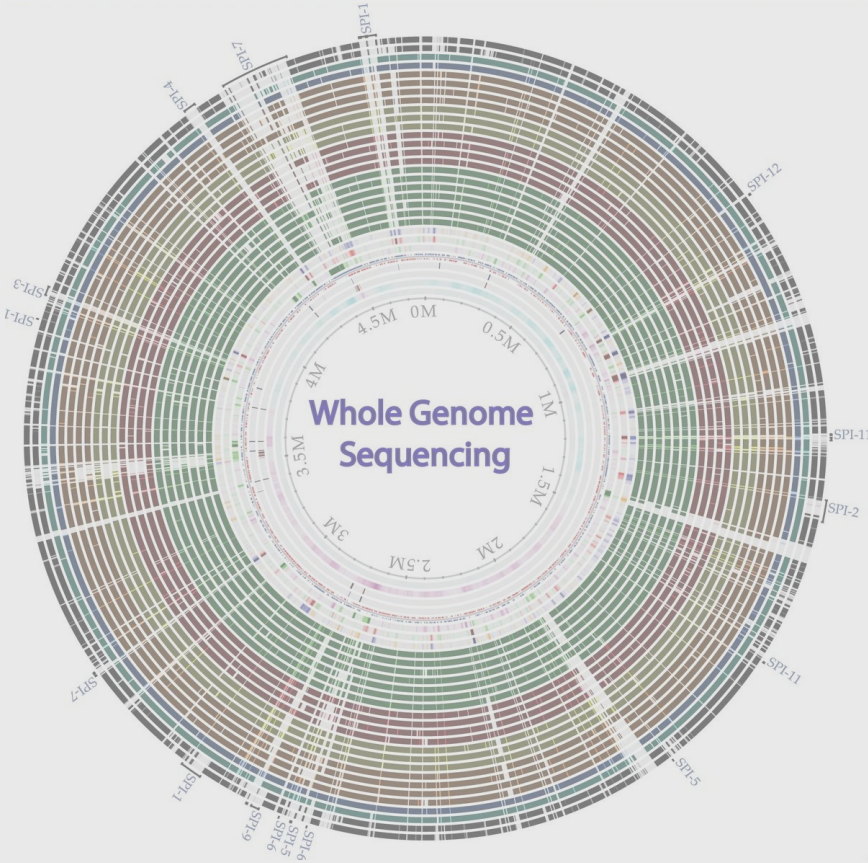
Table 3: Effect of Read Length

Sample	N50 contig size	Largest contig	Genome coverage
E. coli, 100 bp pe	132,786 bp	326,886 bp	99.87 %
E. coli, 400 bp sr	22,902 bp	127,976 bp	99.87 %
Chr. 20, 100 bp pe	70,744 bp	484,312 bp	92.69 %
Chr. 20, 400 bp sr	2,319 bp	22,823 bp	92.65 %

Table 4: Effect of Pairing Reads

Sample (100 bp reads)	N50 contig size	Largest contig	Genome coverage
E. coli, paired-end	132,786 bp	326,886 bp	99.87 %
E. coli, single read	23,326 bp	127,976 bp	99.87 %
Chr. 20, paired-end	70,744 bp	484,312 bp	92.69 %
Chr. 20, single read	2,320 bp	22,823 bp	92.43 %

Whole-Genome Sequencing



https://www.coursera.org/learn/wgs-bacteria

27-05-2020