

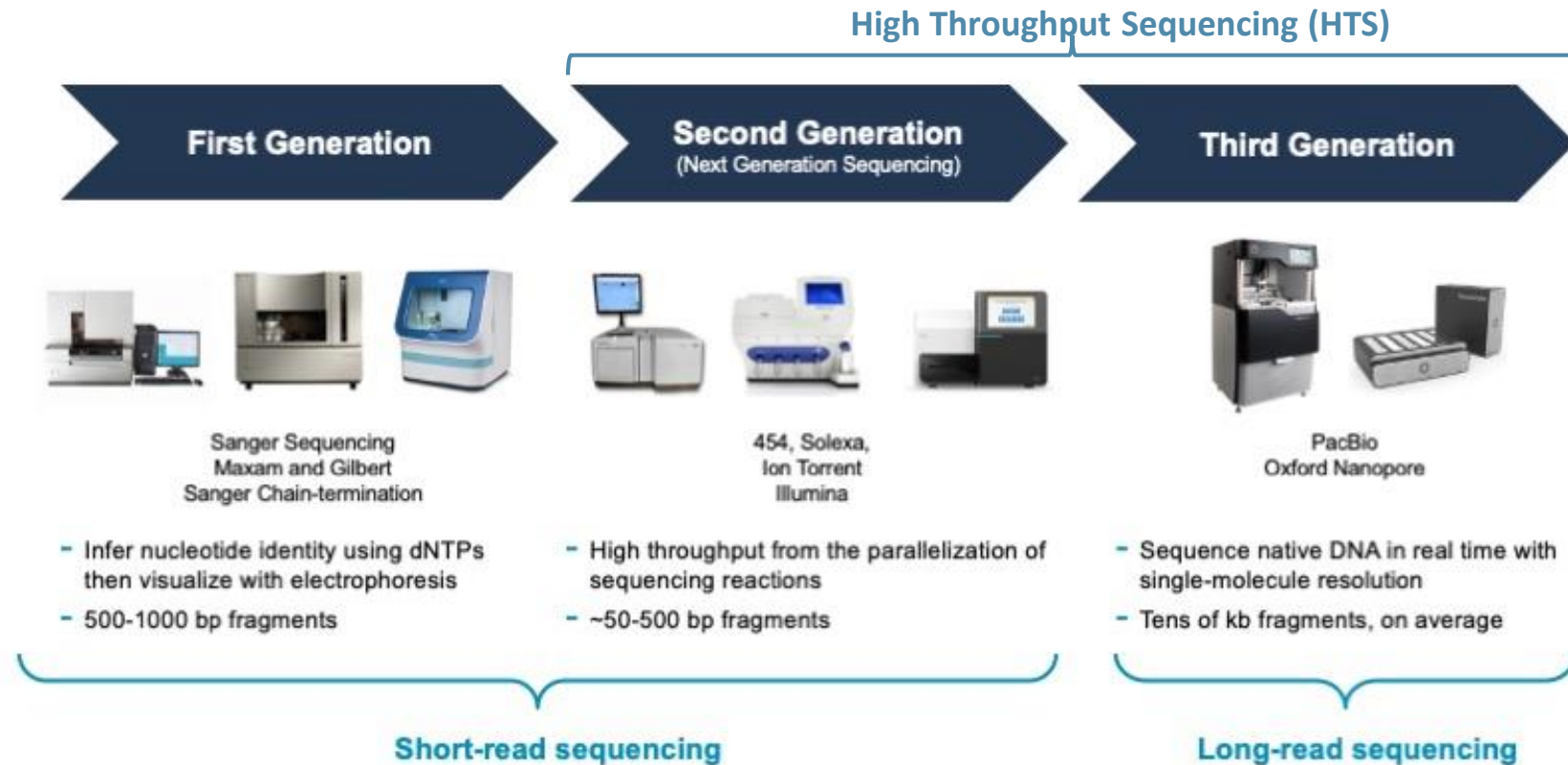
Introduction to NGS data analysis

Dr Gustavo A. Silva-Arias

Technische Universität München

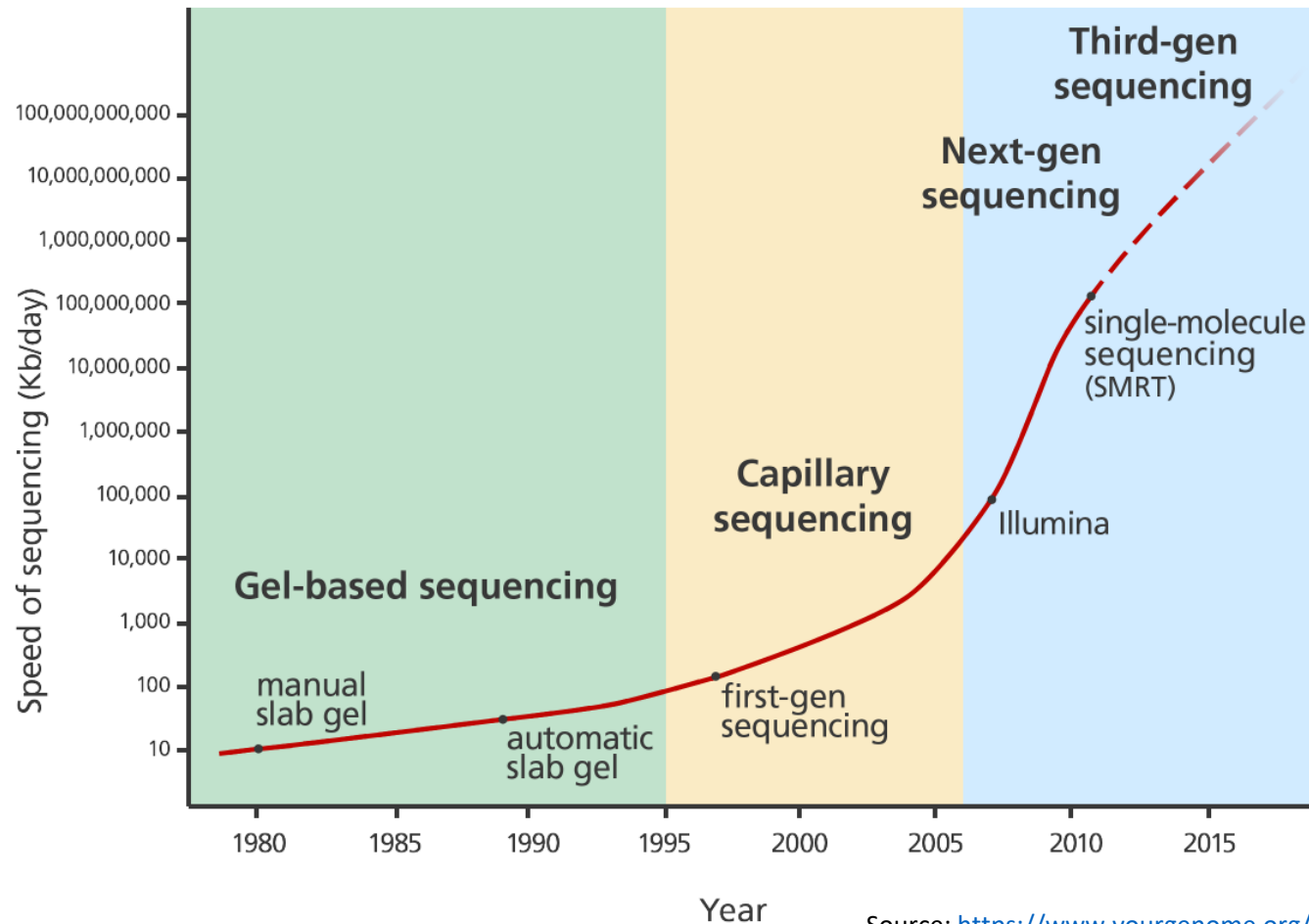
Bogotá, 23 de Agosto 2021

The evolution of sequencing technology



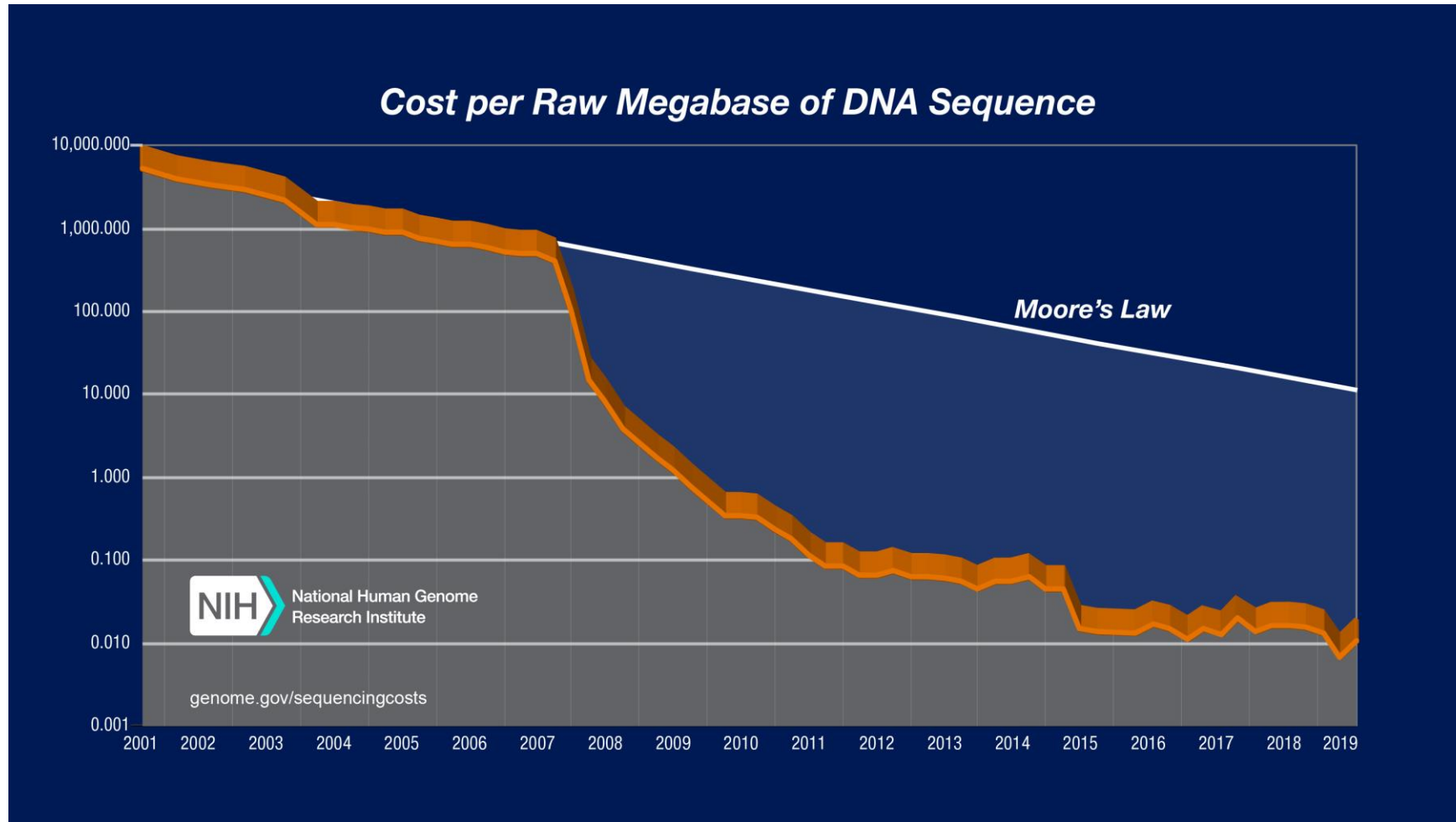
Source: <https://www.pacb.com/blog/the-evolution-of-dna-sequencing-tools/>

The evolution of sequencing technology - efficiency



Source: <https://www.yourgenome.org/stories/third-generation-sequencing>

The evolution of sequencing technology - cost



Source: <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>

First Generation: Sanger sequencing



- Frederick Sanger (1918-2013)
- British biochemist

- **Nobel Prize in Chemistry 1958:**

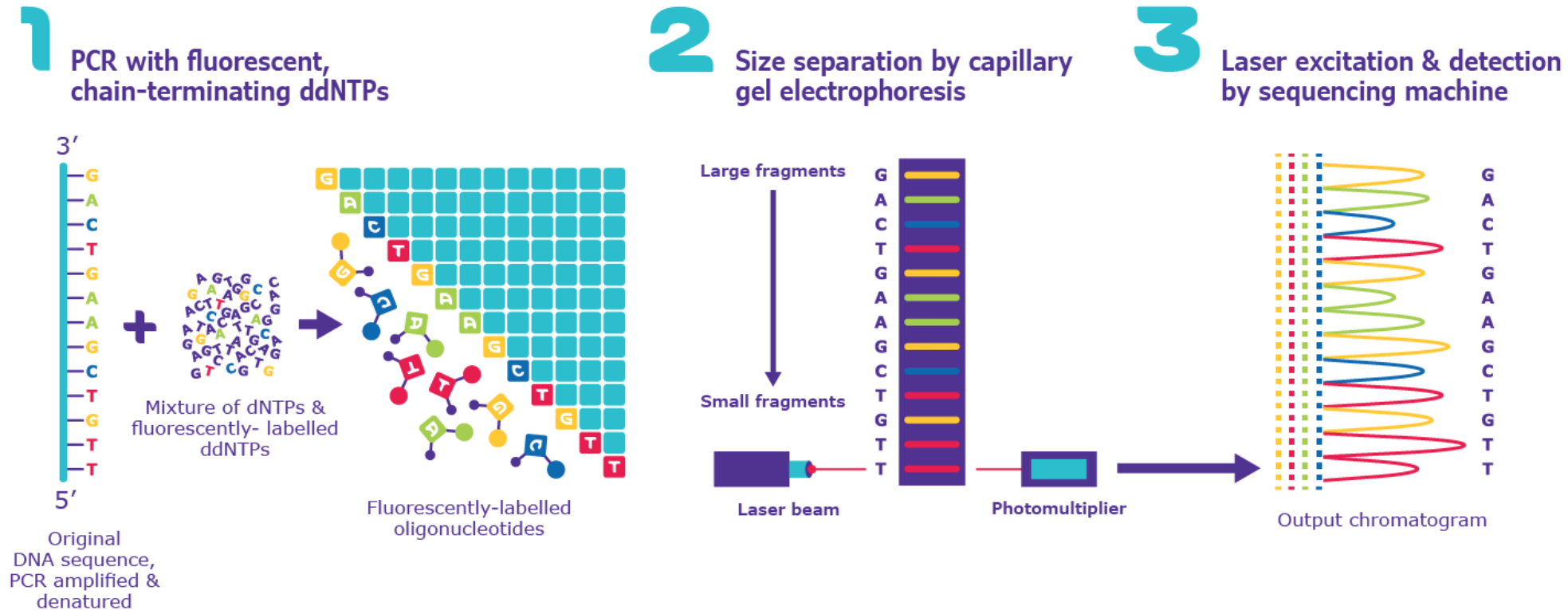
- *"for his work on the structure of proteins, especially that of insulin."*
- He determined the complete aminoacid sequence of bovine insuline using electrophoresis and chromatography

- **Nobel Prize in Chemistry 1980:**

- *"for their contributions concerning the determination of base sequences in nucleic acids."*
- Developed the "dideoxy" chain-termination method for sequencing DNA molecules, now known as the "Sanger method".

Source: <https://www.nobelprize.org/prizes/chemistry/>

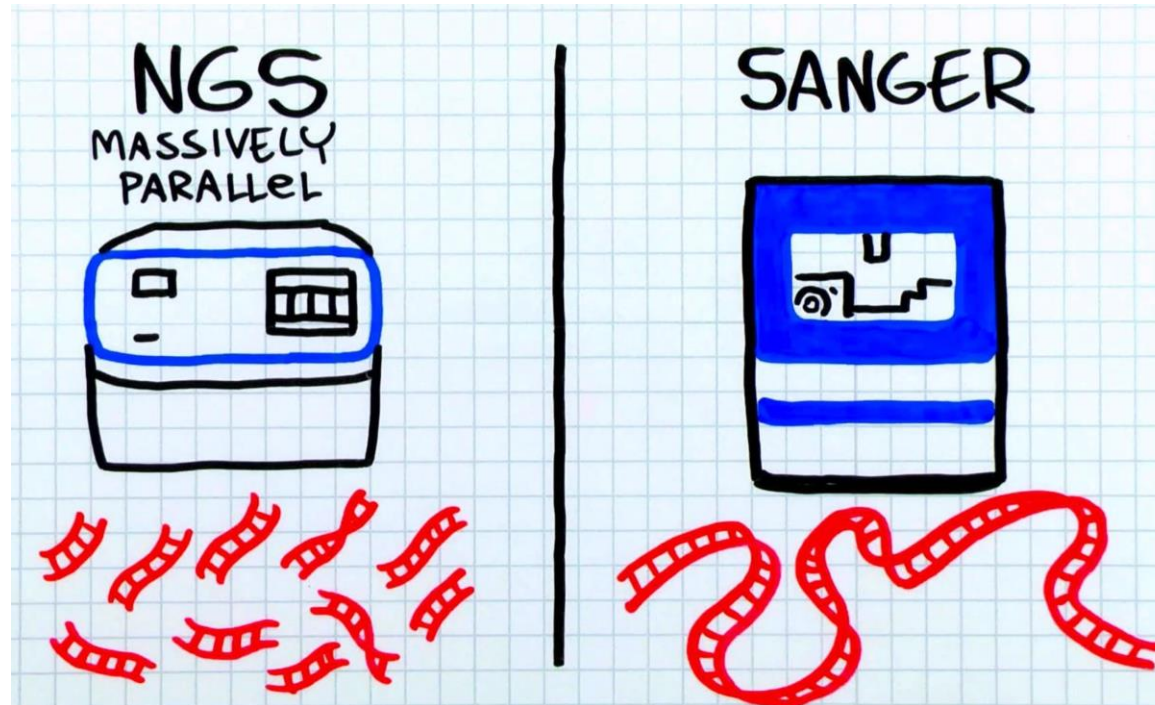
First Generation: Sanger sequencing



- Old gel „slab“ method: <https://www.youtube.com/watch?v=3M0PyxFPwkQ>
- Capillary gel method: <https://www.youtube.com/watch?v=x7PUqNA0eOA>

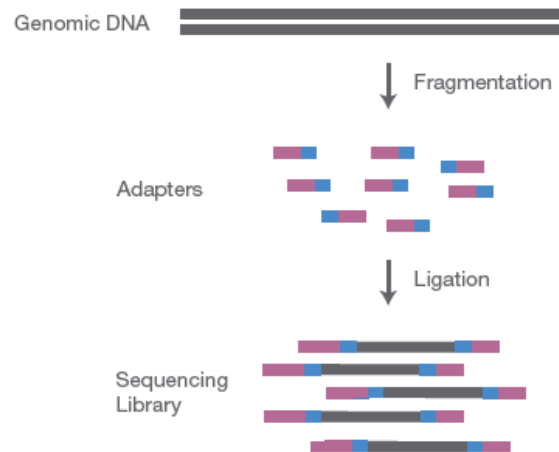
Source: <https://www.sigmaaldrich.com/technical-documents/articles/biology/sanger-sequencing.html>

Second Generation: Massive Parallel Sequencing of Short Reads

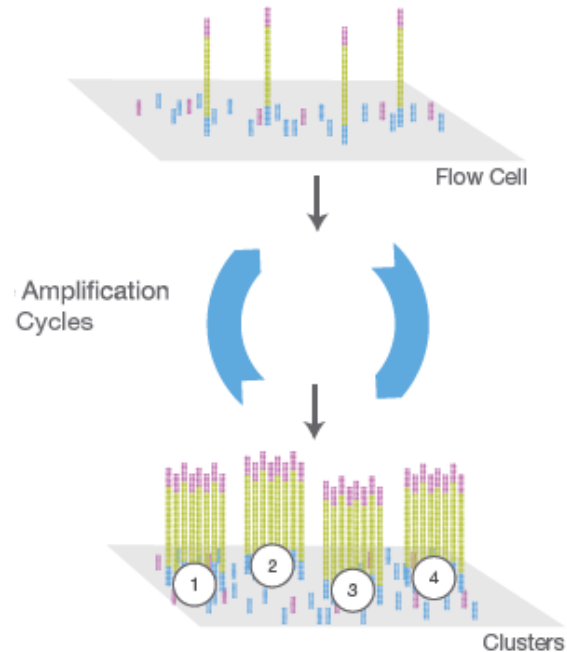


Second Generation: Massive Parallel Sequencing of Short Reads

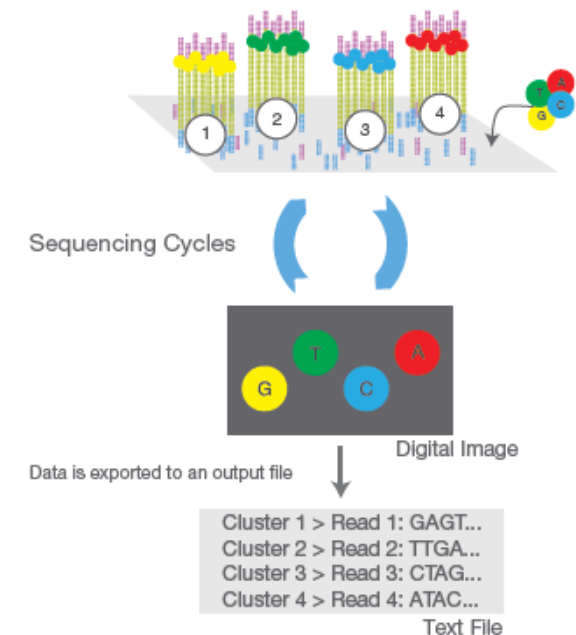
- 1. Library preparation



- 2. Template Amplification



- 3. Sequencing



Source: <https://emea.illumina.com/content/dam/illumina-marketing/documents/products/other/ivf-reproductive-genetic-health-ngs-primer-1570-2015-012.pdf>

Second Generation: Massive Parallel Sequencing of Short Reads

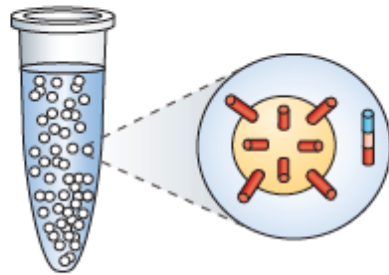
Template amplification methods

- Emulsion PCR
 - 454 (Roche)
 - SOLiD (Thermo Fisher)
 - GeneReader (Quiagen)
 - Ion Torrent (Thermo Fisher)
- Solid-phase bridge amplification
 - Illumina
- Solid-phase template walking
 - SOLiD Wilfire – Thermo Fischer
- In-solution DNA nanoball generation
 - Complete Genomics - BGI

Template amplification methods

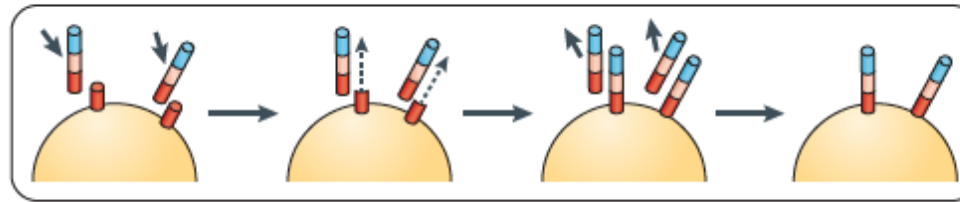
a Emulsion PCR

(454 (Roche), SOLiD (Thermo Fisher), GeneReader (Qiagen), Ion Torrent (Thermo Fisher))



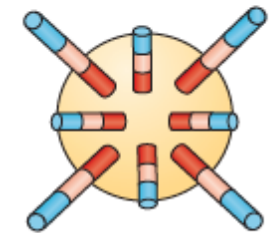
Emulsion

Micelle droplets are loaded with primer, template, dNTPs and polymerase



On-bead amplification

Templates hybridize to bead-bound primers and are amplified; after amplification, the complement strand disassociates, leaving bead-bound ssDNA templates



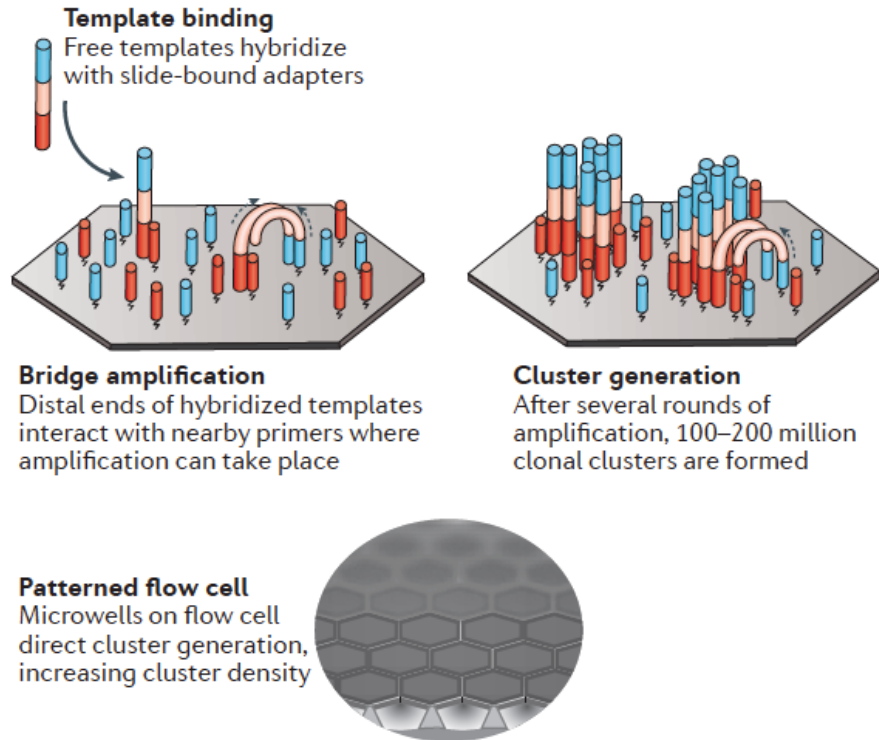
Final product

100–200 million beads with thousands of bound template

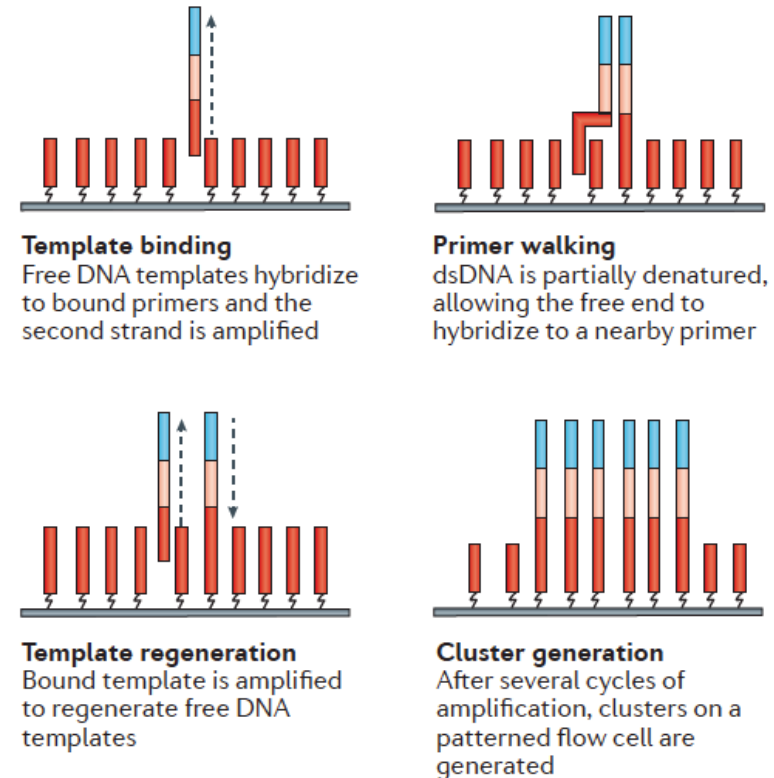
Source: Sara Goodwin *et al.* 2016 - *Nature Reviews Genetics* - <https://doi.org/10.1038/nrg.2016.49>

Template amplification methods

b Solid-phase bridge amplification (Illumina)



c Solid-phase template walking (SOLiD Wildfire (Thermo Fisher))

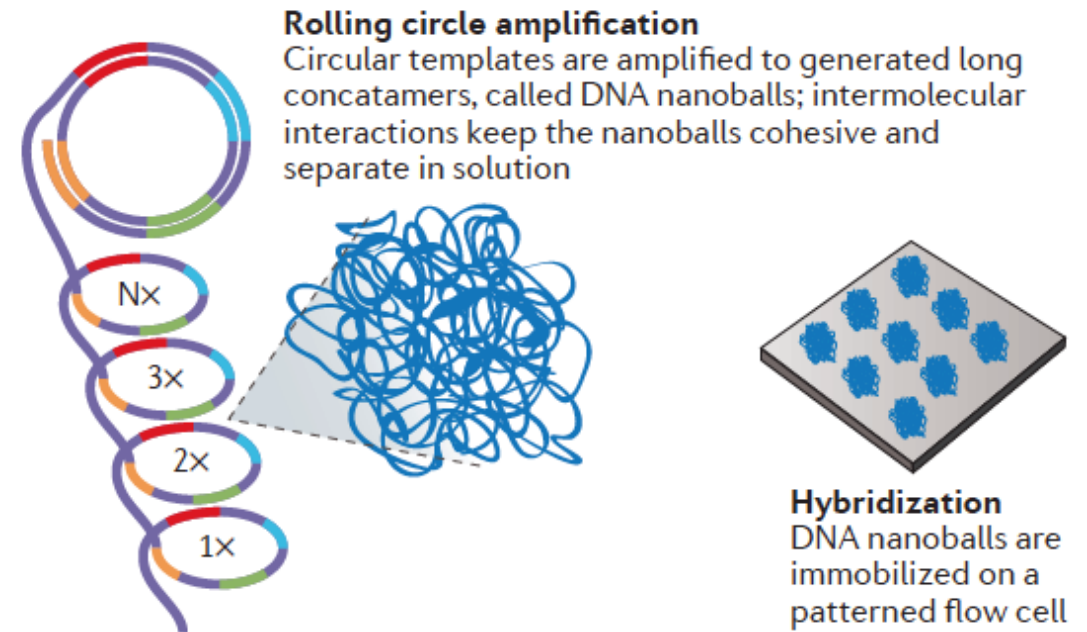
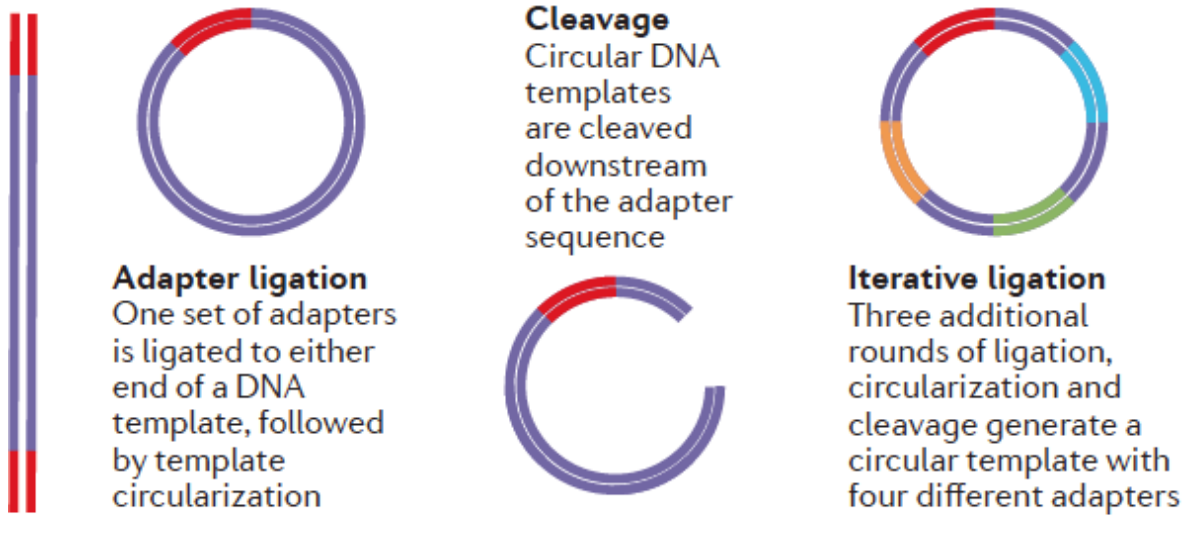


Source: Sara Goodwin *et al.* 2016 - *Nature Reviews Genetics* - <https://doi.org/10.1038/nrg.2016.49>

Template amplification methods

DNBSEQ (BGI)

d In-solution DNA nanoball generation (Complete Genomics (BGI))



Source: Sara Goodwin *et al.* 2016 - *Nature Reviews Genetics* - <https://doi.org/10.1038/nrg.2016.49>

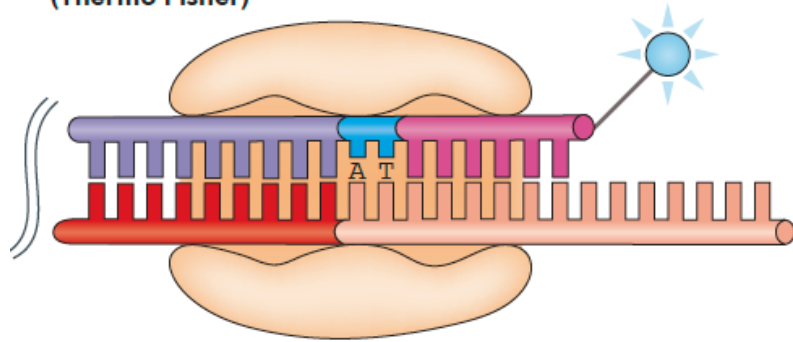
Second Generation: Massive Parallel Sequencing of Short Reads

Sequencing Methods

- By Ligation (SBL)
SOLiD (Thermo Fisher)
- By Synthesis (SBS)
 - Cyclic Reversible Termination (CRT)
Illumina
GeneReader (QiaGen)
DNBSEQ (BGI)
 - SNA: Single Nucleotide Addition
454 (Roche)
IonTorrent (ThermoFisher)

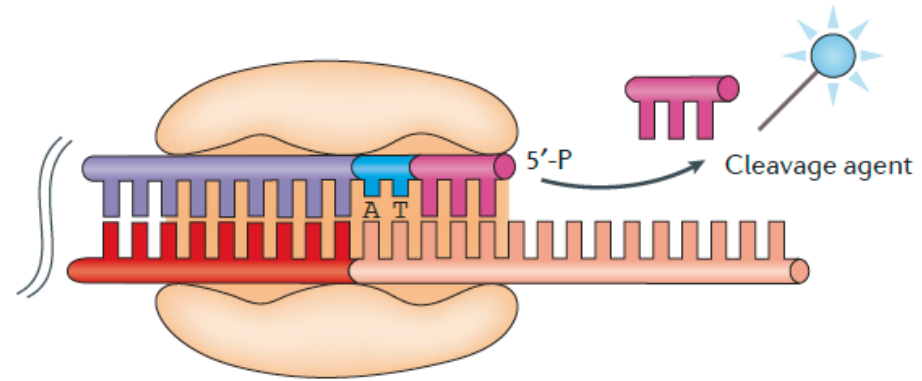
Sequencing Methods - By Ligation (SBL)

a SOLiD
(Thermo Fisher)



Two-base-encoded probes

Probes with two known bases followed by degenerate or universal bases hybridize to a template; ligase immobilizes the complex and the slide is imaged



Cleavage

The fluorophore is cleaved from the probe along with several bases, revealing a 5' phosphate

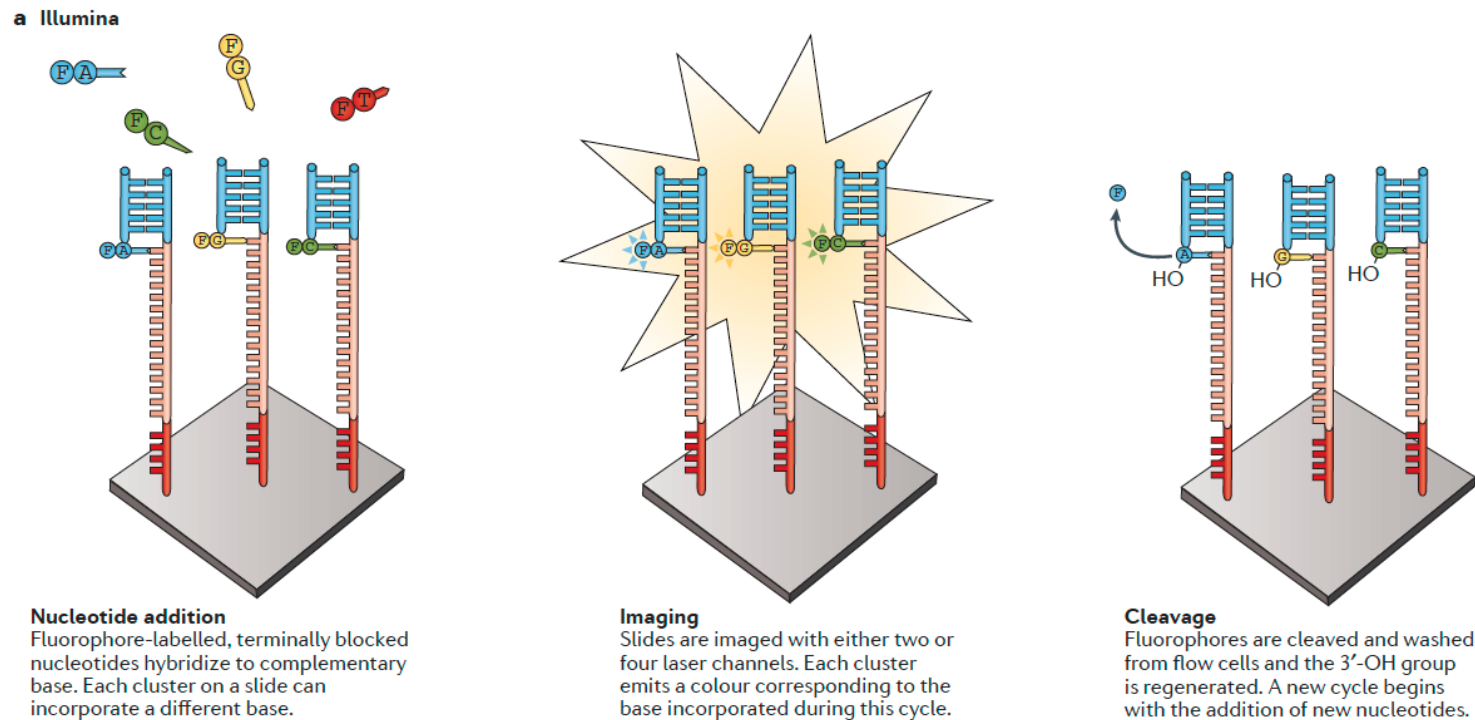
Full explanation: <https://www.youtube.com/watch?v=PPEKybWYOB4>

***Note:** DNBSEQ (BGI) also used a ligation technique but has switched to SBS to obtain longer reads

Source: Sara Goodwin *et al.* 2016 - *Nature Reviews Genetics* - <https://doi.org/10.1038/nrg.2016.49>

Sequencing Methods - By Synthesis (SBS)

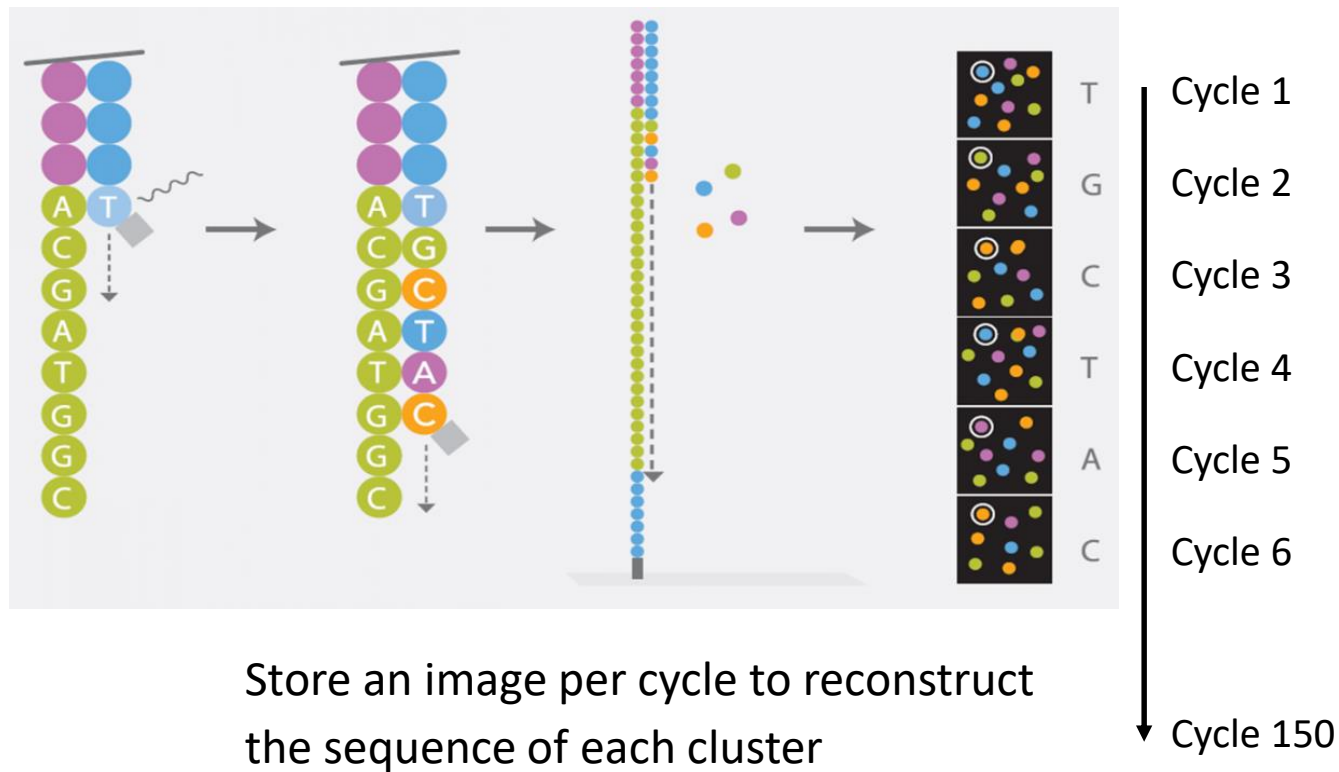
- CRT: Cyclic Reversible Termination, used by Illumina, GeneReader (QiaGen), DNBSEQ (BGI)



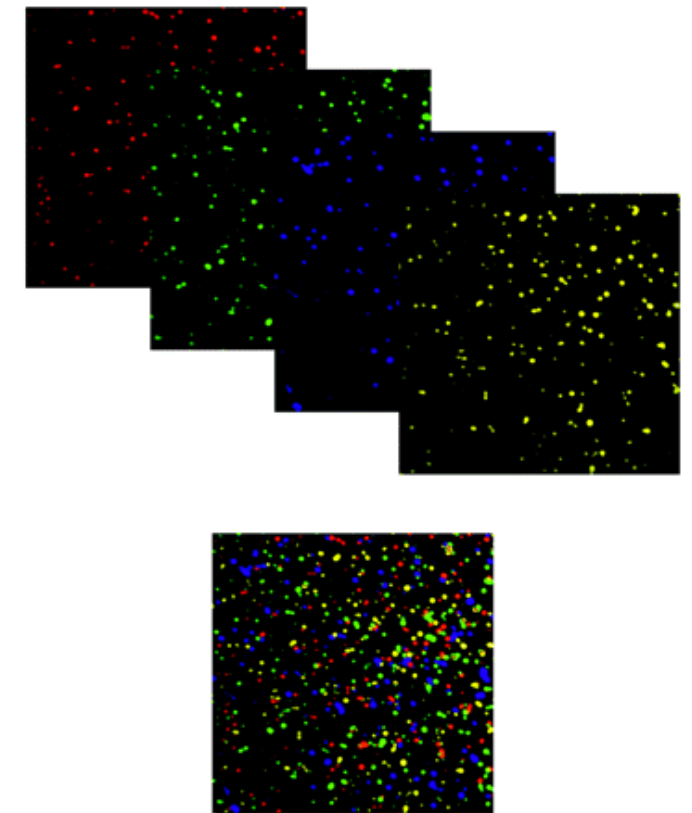
Source: Sara Goodwin *et al.* 2016 - *Nature Reviews Genetics* - <https://doi.org/10.1038/nrg.2016.49>

Sequencing Methods - By Synthesis (SBS)

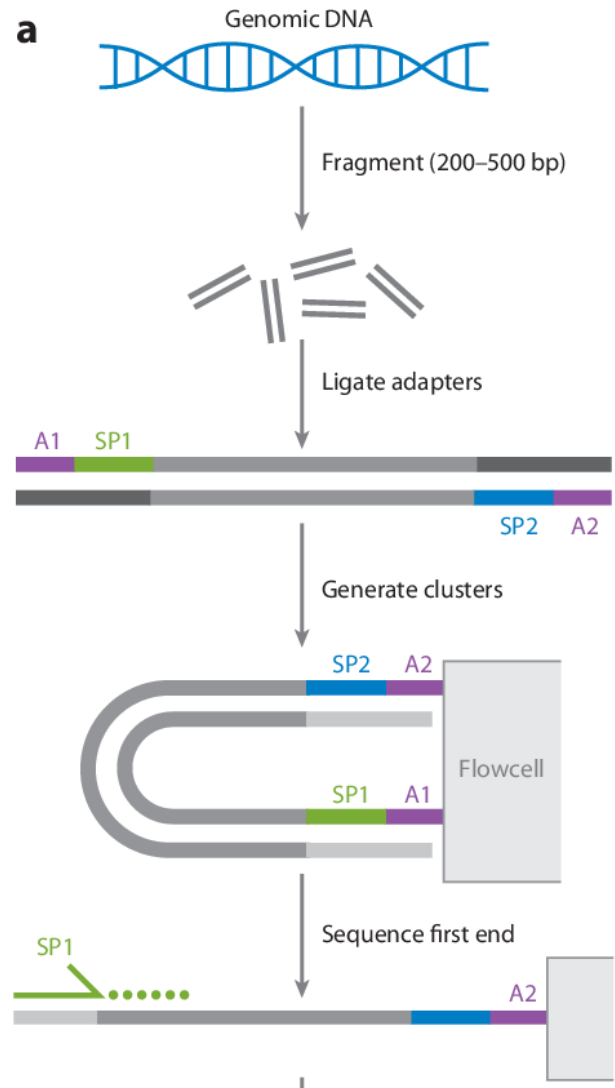
CRT: Cyclic Reversible Termination, used by Illumina, GeneReader (QiaGen), DNBSEQ (BGI)



Source: Sara Goodwin *et al.* 2016 - *Nature Reviews Genetics* - <https://doi.org/10.1038/nrg.2016.49>



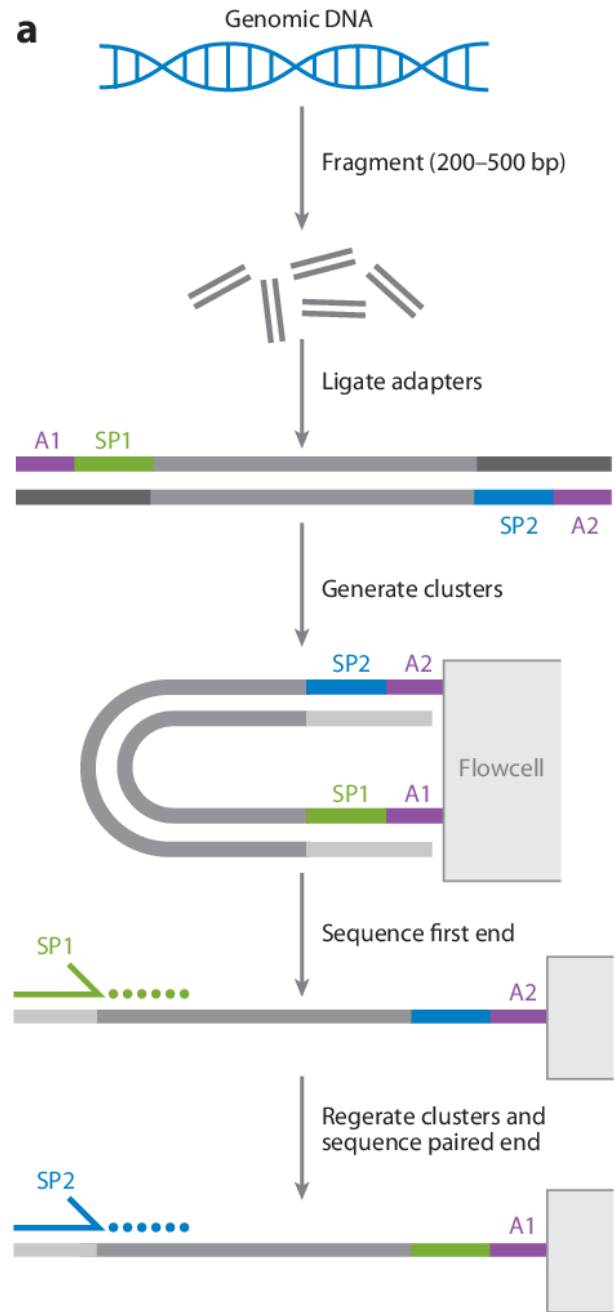
Source: <https://doi.org/10.1373/clinchem.2008.112789>



Single-End vs. Paired-End (Illumina, DNBSEQ)

Single-End

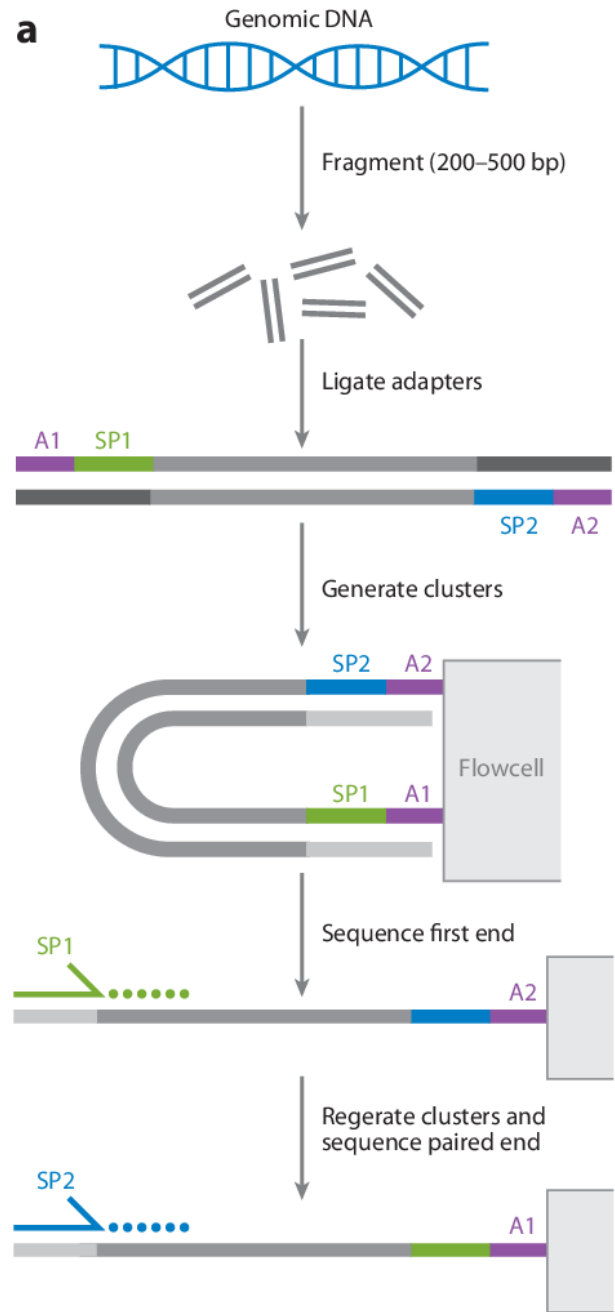
Source: Elaine R. Mardis 2013 - *Annual Review of Analytical Chemistry* - <https://doi.org/10.1146/annurev-anchem-062012-092628>



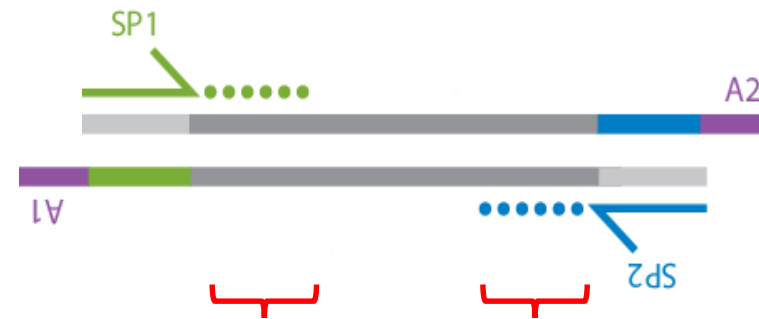
Single-End vs. Paired-End (Illumina, DNBSEQ)

Paired-End

Source: Elaine R. Mardis 2013 - *Annual Review of Analytical Chemistry* - <https://doi.org/10.1146/annurev-anchem-062012-092628>



Single-End vs. Paired-End (Illumina, DNBSEQ)



Read length 150 (x2)

Insert length

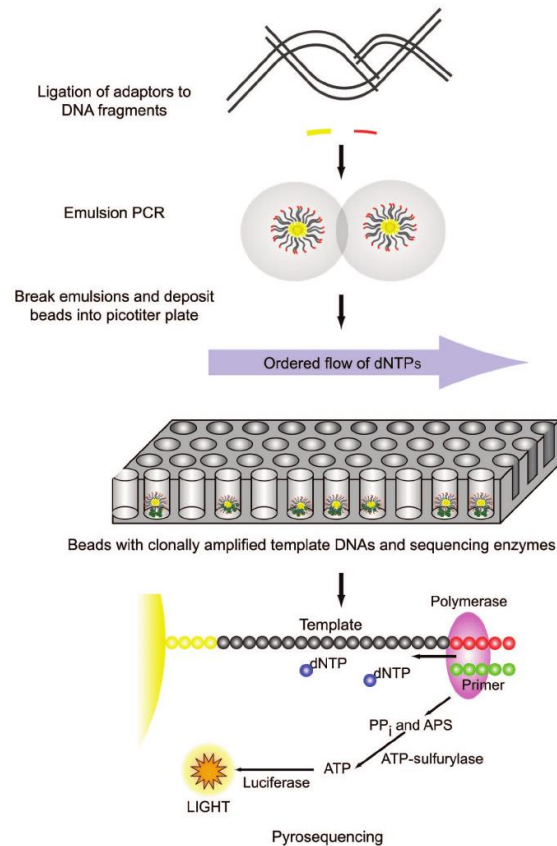
Ideally the insert length (genomic DNA fragment) should be longer than read length x 2

Paired-End

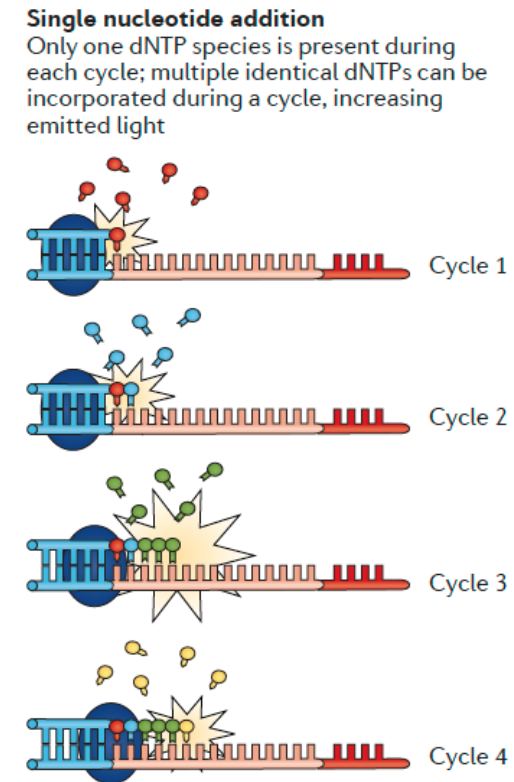
Source: Elaine R. Mardis 2013 - *Annual Review of Analytical Chemistry* - <https://doi.org/10.1146/annurev-anchem-062012-092628>

Sequencing Methods - By Synthesis (SBS)

- SNA: Single Nucleotide Addition. 454 (Roche), IonTorrent (ThermoFisher)



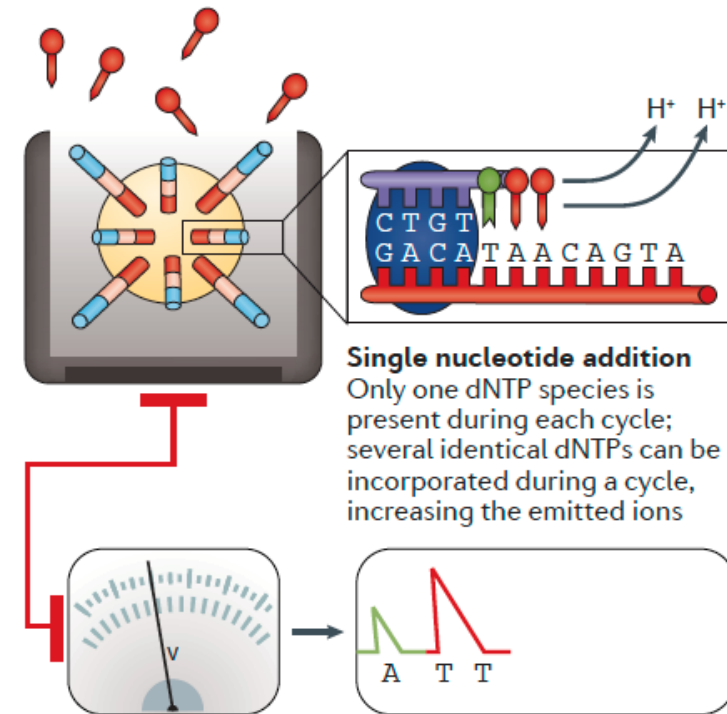
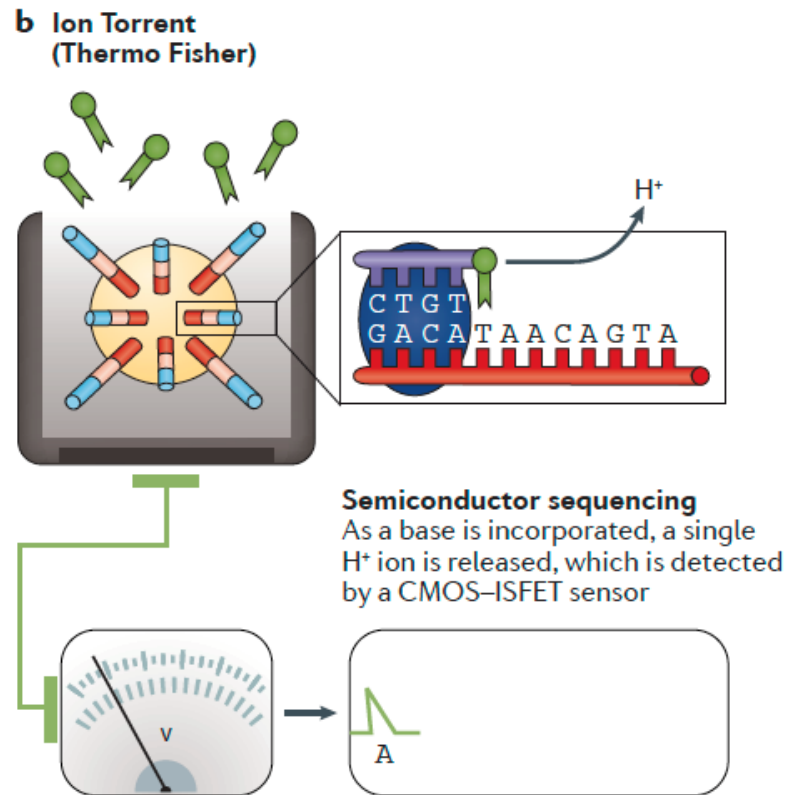
Source: <https://doi.org/10.1373/clinchem.2008.112789>



Source: Sara Goodwin *et al.* 2016 - *Nature Reviews Genetics*
<https://doi.org/10.1038/nrg.2016.49>

Sequencing Methods - By Synthesis (SBS)

- SNA: Single Nucleotide Addition, used by 454 (Roche), IonTorrent (ThermoFisher)



Source: Sara Goodwin *et al.* 2016 - *Nature Reviews Genetics* - <https://doi.org/10.1038/nrg.2016.49>

Sequencing Methods - By Synthesis (SBS)

Full explanation videos:

- Illumina [CRT]:
<https://www.youtube.com/watch?v=fCd6B5HRaZ8>
- DNBSEQ (BGI) [CRT]:
<https://www.youtube.com/watch?v=RGcpftDHpng&t>
- 454 pyrosequencing [SNA]:
<https://www.youtube.com/watch?v=bNKEhOGvcal>
- IonTorrent [SNA]:
<https://www.youtube.com/watch?v=ZL7DXFPz8rU>

Third Generation: Parallel Sequencing of Long Reads in Real Time

A Real-time long-read sequencing

Aa Pacific Biosciences

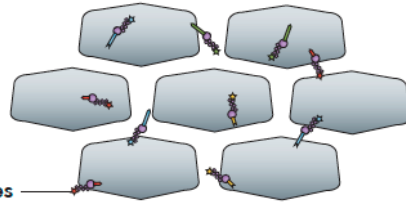
SMRTbell template

Two hairpin adapters allow continuous circular sequencing



ZMW wells

Sites where sequencing takes place

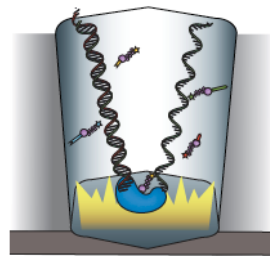


Labelled nucleotides

All four dNTPs are labelled and available for incorporation

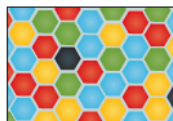
Modified polymerase

As a nucleotide is incorporated by the polymerase, a camera records the emitted light

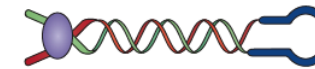


PacBio output

A camera records the changing colours from all ZMWs; each colour change corresponds to one base

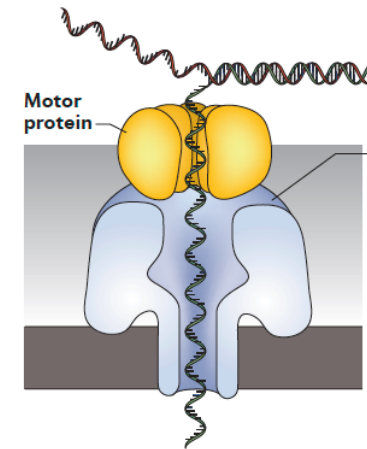


Ab Oxford Nanopore Technologies



Leader-Hairpin template

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

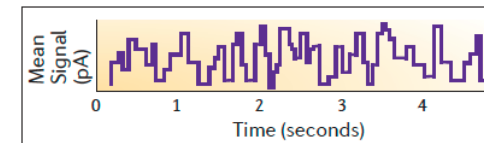


Alpha-hemolysin

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

Source: Sara Goodwin *et al.* 2016 - *Nature Reviews Genetics* - <https://doi.org/10.1038/nrg.2016.49>

Third Generation: Parallel Sequencing of Long Reads in Real Time

Full explanation videos:

- PacBio:

<https://www.youtube.com/watch?v=v8p4ph2MAvI>

<https://www.youtube.com/watch?v=NHCJ8PtYCFc>

<https://www.youtube.com/watch?v=ID8JyAbwEo>

- Nanopore:

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

<https://www.youtube.com/watch?v=CGWZvHli3i0>