Introduction to NGS technologies

Dr Gustavo A. Silva-Arias

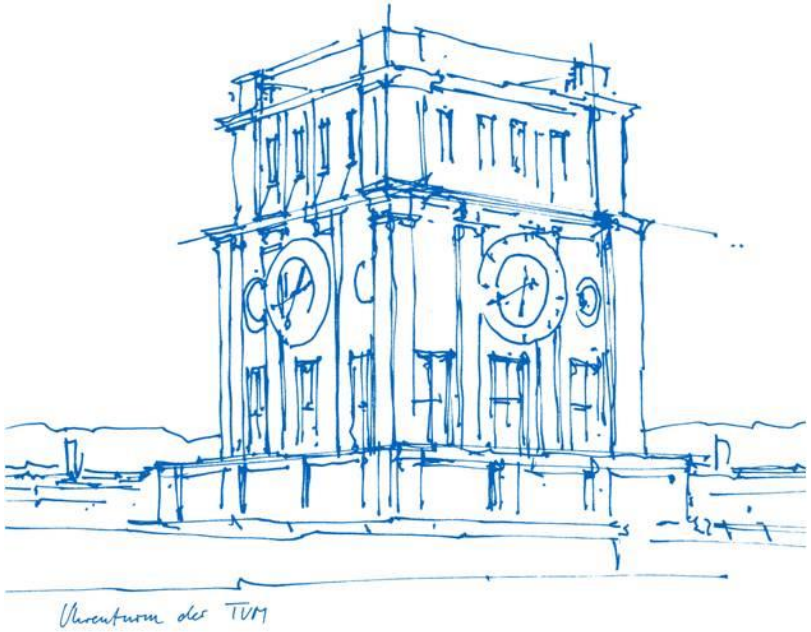
Professorship for Population Genetics

Dept of Life Science Systems

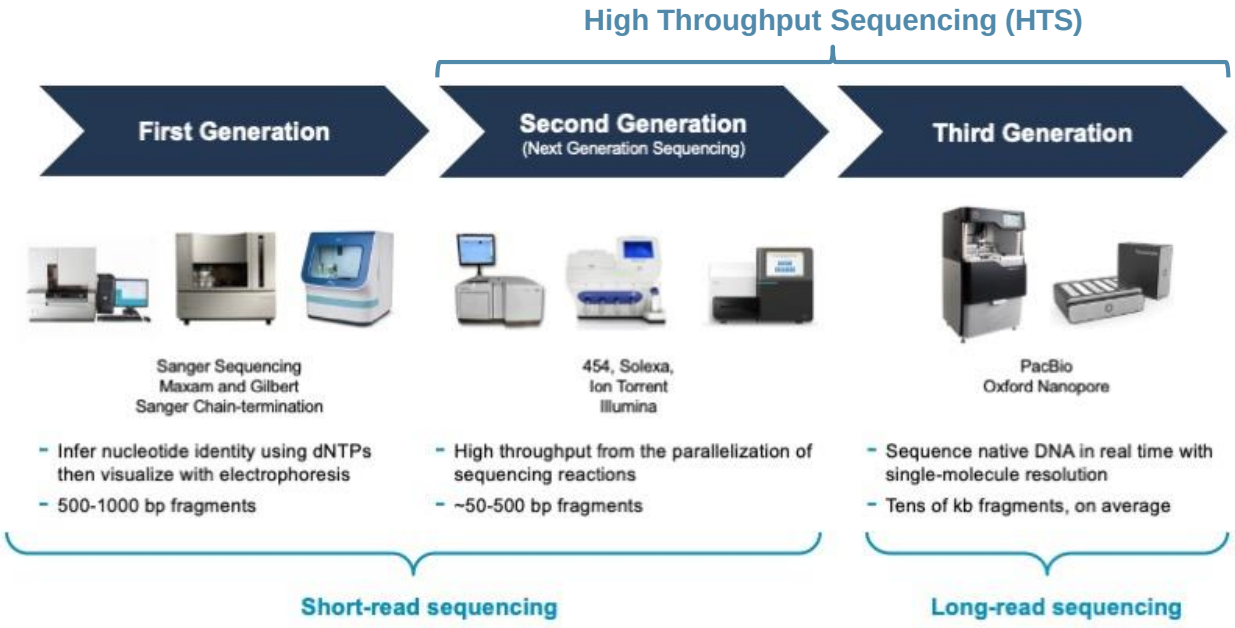
Technical University of Munich

La Paz, Cesar

1 de Agosto de 2022

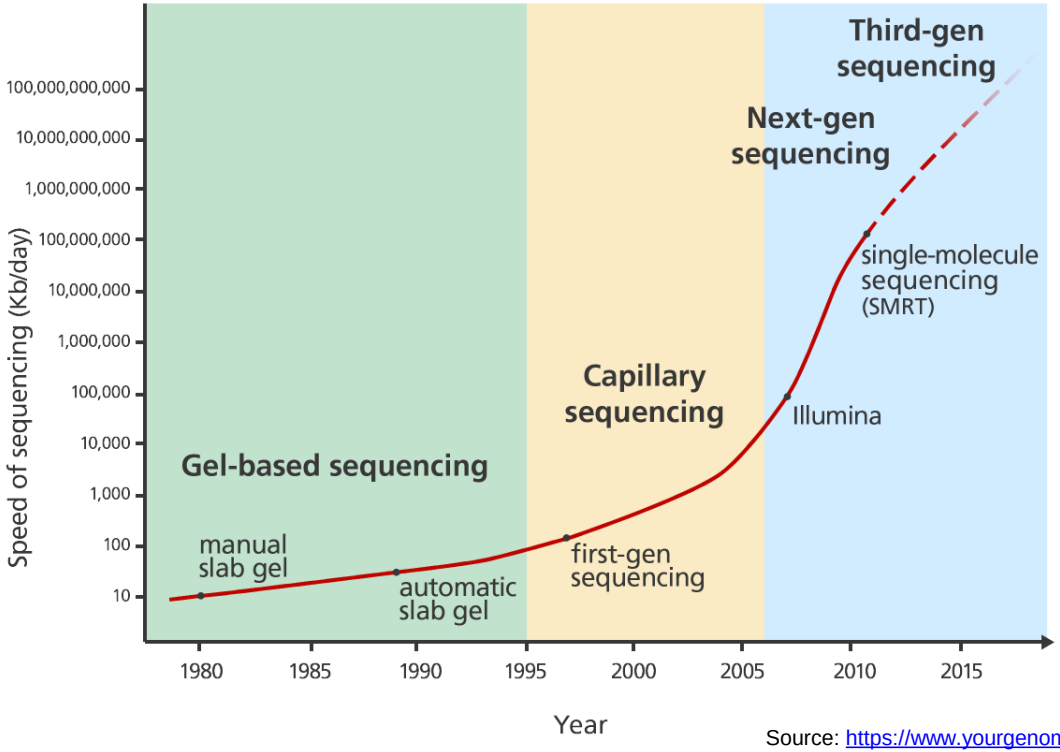


The evolution of sequencing technology



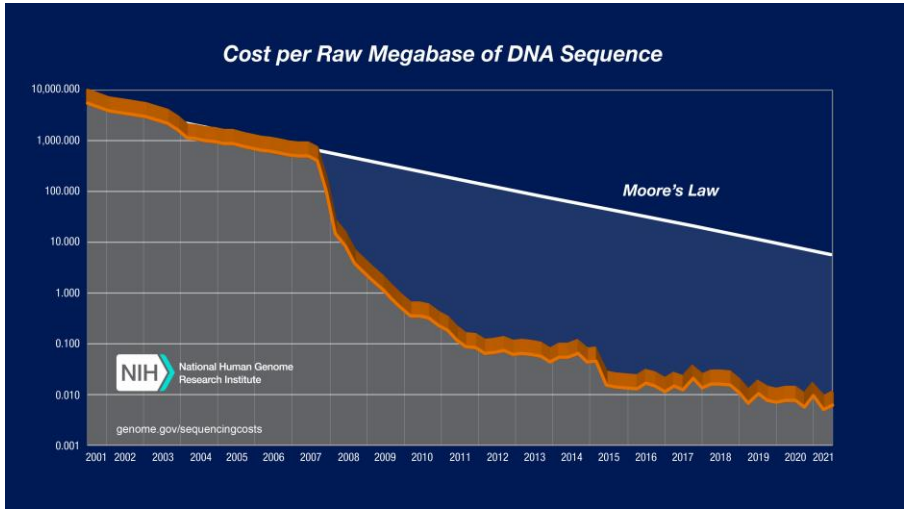
Source: https://www.pacb.com/blog/the-evolution-of-dna-sequencing-tools/

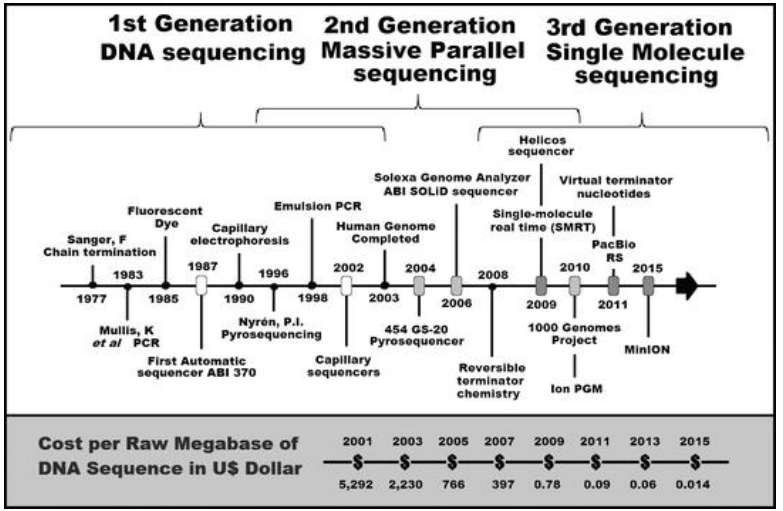
The evolution of sequencing technology - efficiency



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

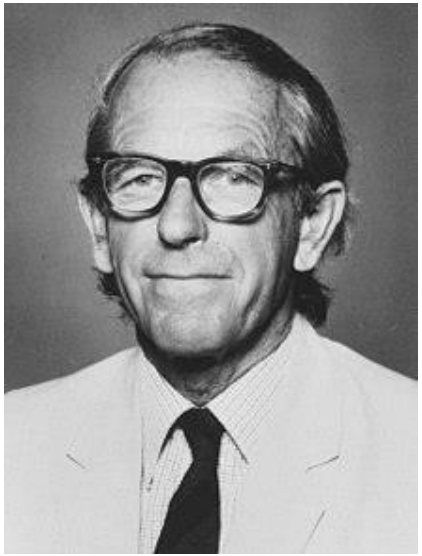
The evolution of sequencing technology - cost





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:

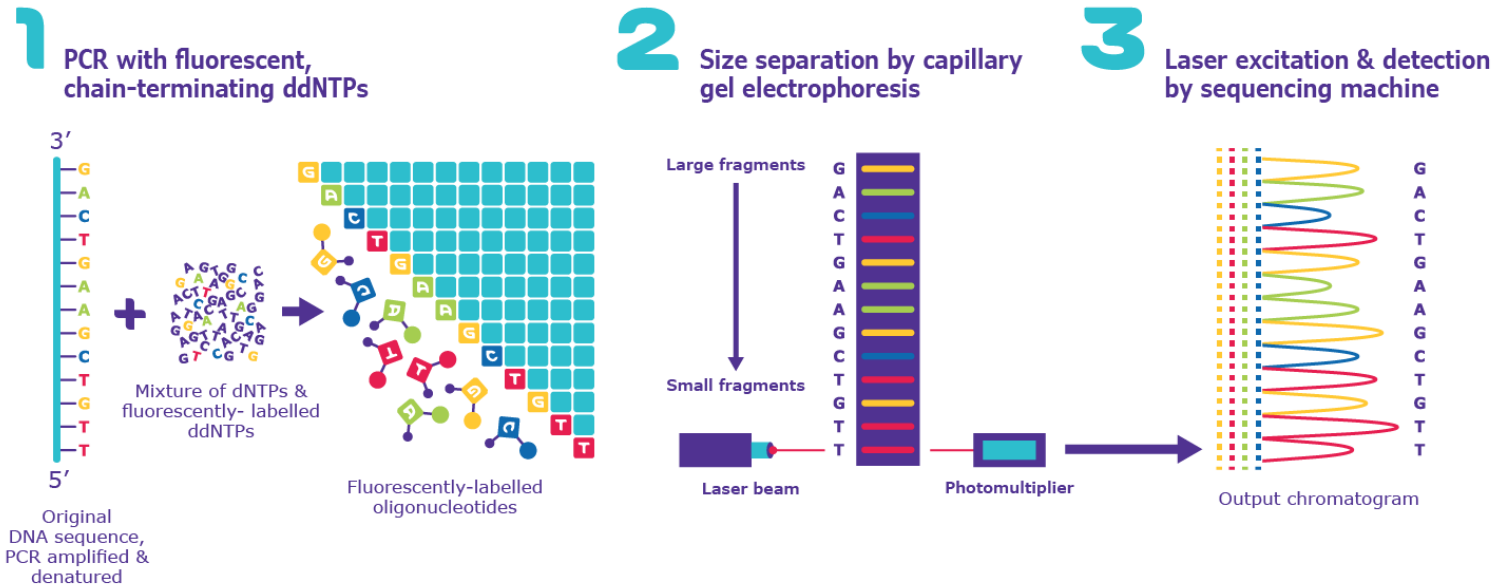
- *"for his work on the structure of proteins, especially that of insulin.“*
- He determined the complete aminoacid sequence of bovine insuline using electrophoresis and chromatography

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

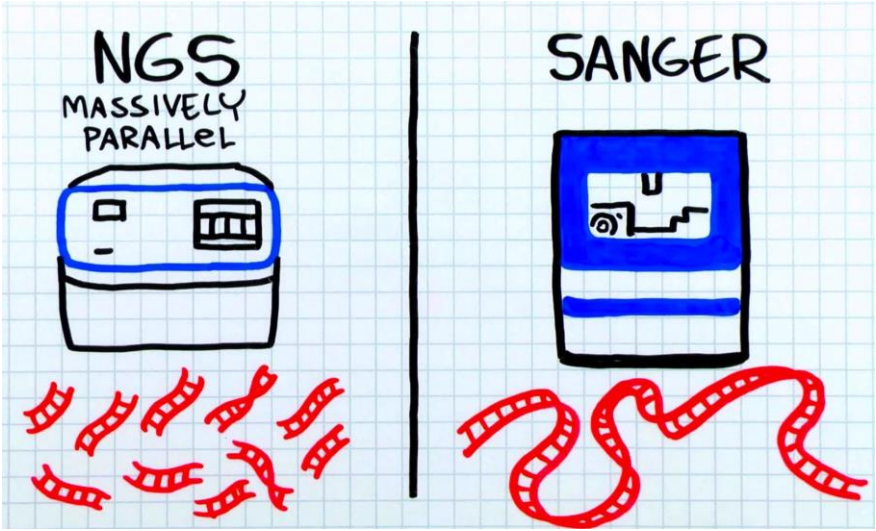


Old gel „slab“ method: https://www.youtube.com/watch?v=3M0PyxFPwkQ

Capillary gel method: https://www.youtube.com/watch?v=x7PUqNA0eOA

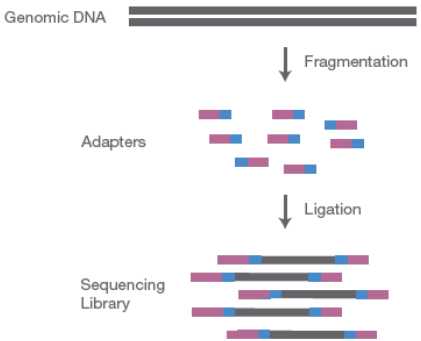
Source: https://www.sigmaaldrich.com/technical-documents/articles/biology/sanger-sequencing.html

Second Generation:
Massive Parallel Sequencing of Short Reads

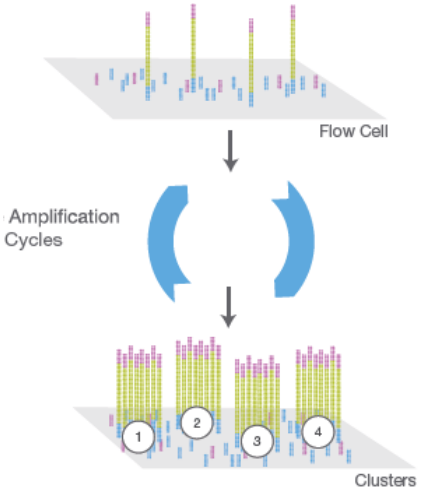


Second Generation:
Massive Parallel Sequencing of Short Reads

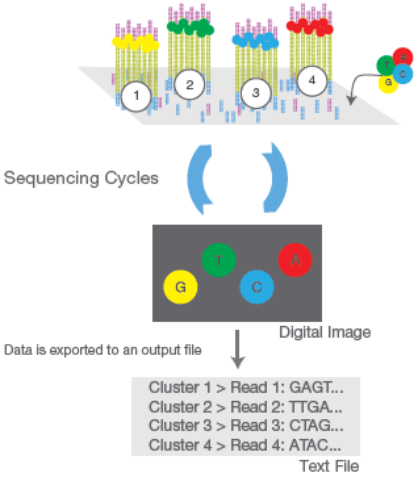
1. Library preparation



2. Template Amplification



3. Sequencing



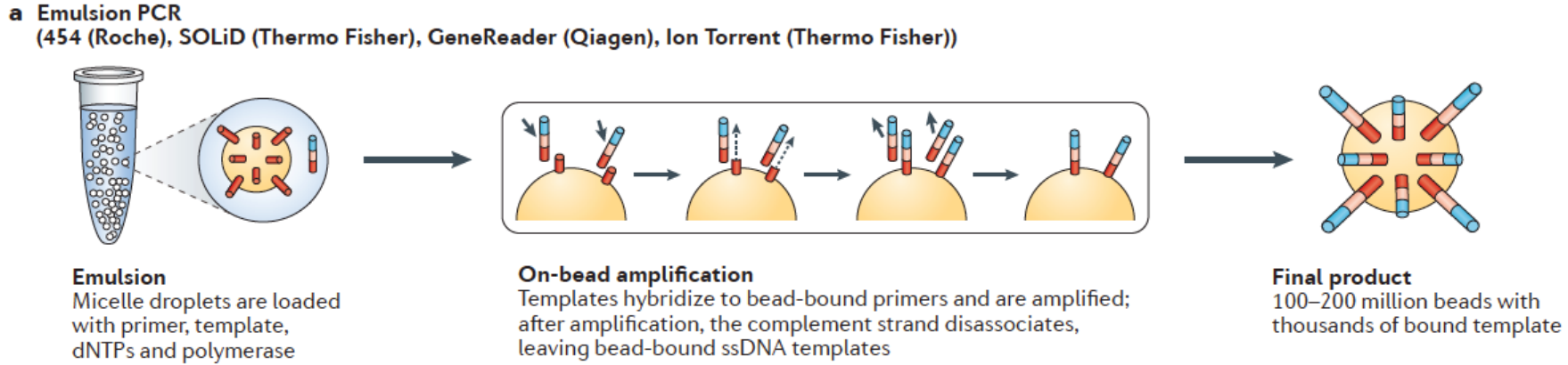
Source: https://emea.illumina.com/content/dam/illumina-marketing/documents/products/other/ivf-reproductive-genetic-health-ngs-primer-1570-2015-012.pdf

Second Generation:
Massive Parallel Sequencing of Short Reads

Template amplification methods

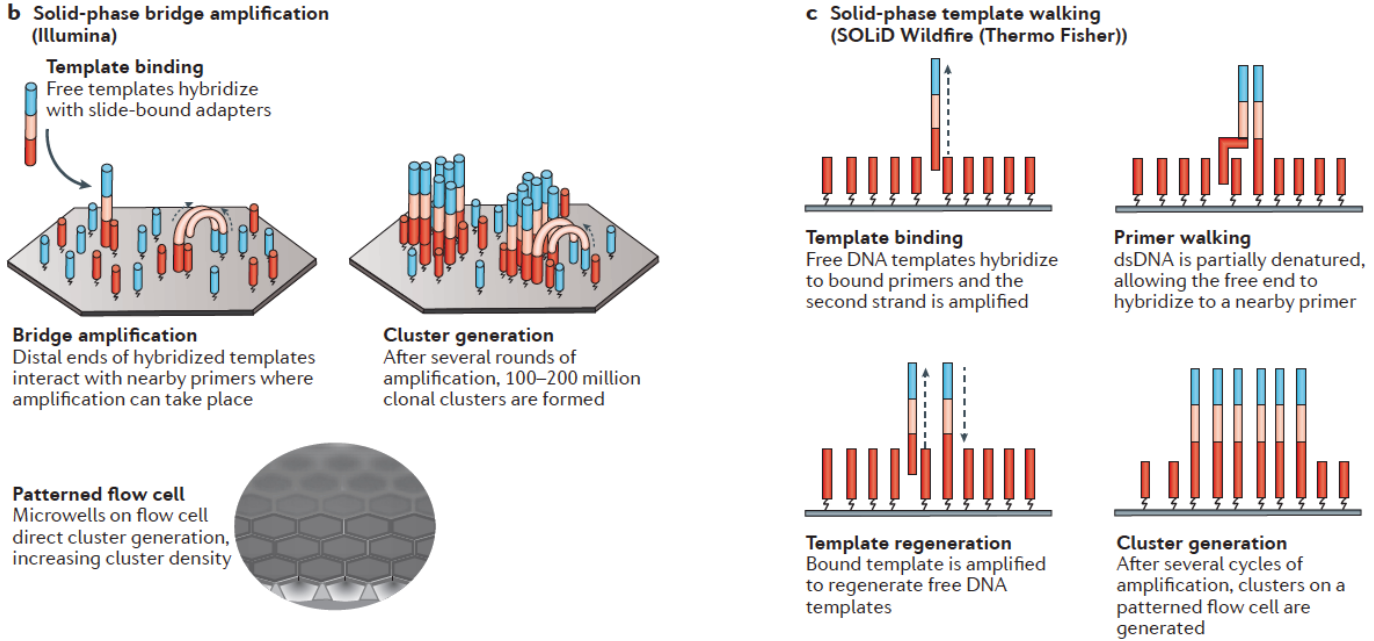
- Emulsion PCR
  - 454 (Roche)
  - SOLiD (Thermo Fisher)
  - GeneReader (Quiagen)
  - Ion Torrent (Thermo Fisher)
- Solid-phase bridge amplification
  - Illumina
- Solid-phase template walking
  - SOLiD Wilfire – Thermo Fischer
- In-solution DNA nanoball generation
  - Complete Genomics - BGI

Template amplification methods



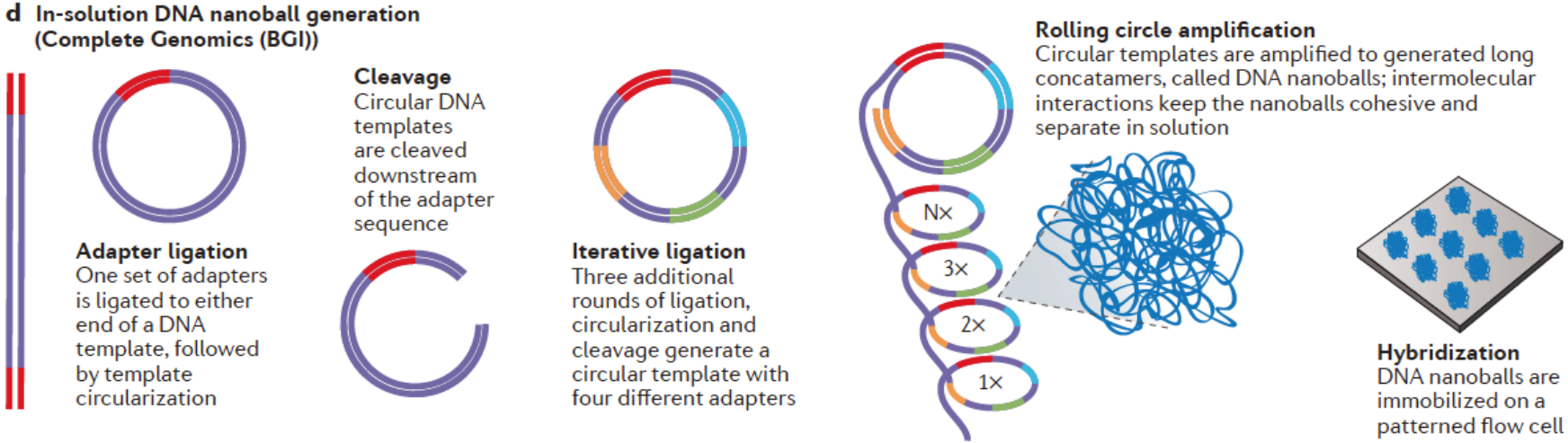
Source: Sara Goodwin *et al.* 2016 - *Nature Reviews Genetics* - https://doi.org/10.1038/nrg.2016.49

Template amplification methods



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

Template amplification methods



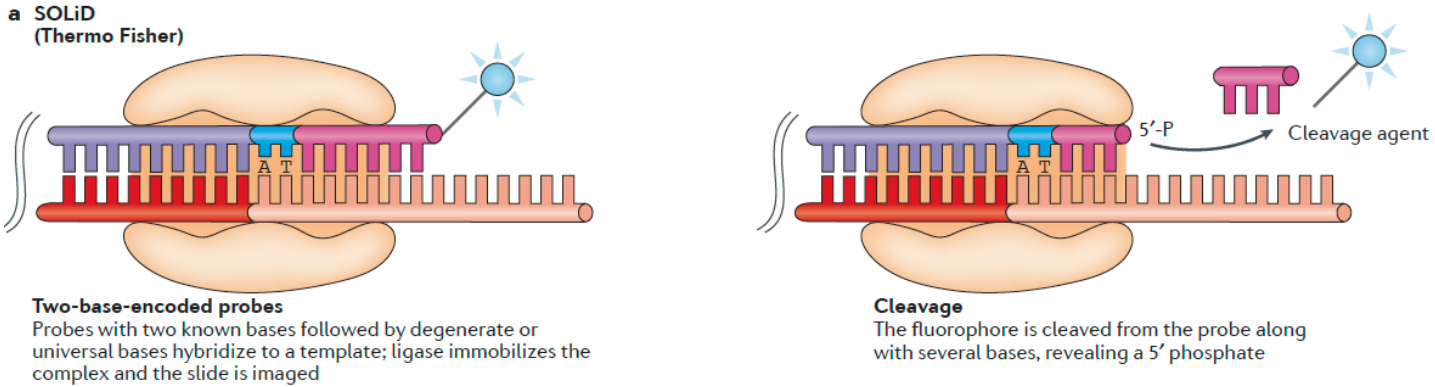
Source: Sara Goodwin *et al.* 2016 - *Nature Reviews Genetics* - https://doi.org/10.1038/nrg.2016.49

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)



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

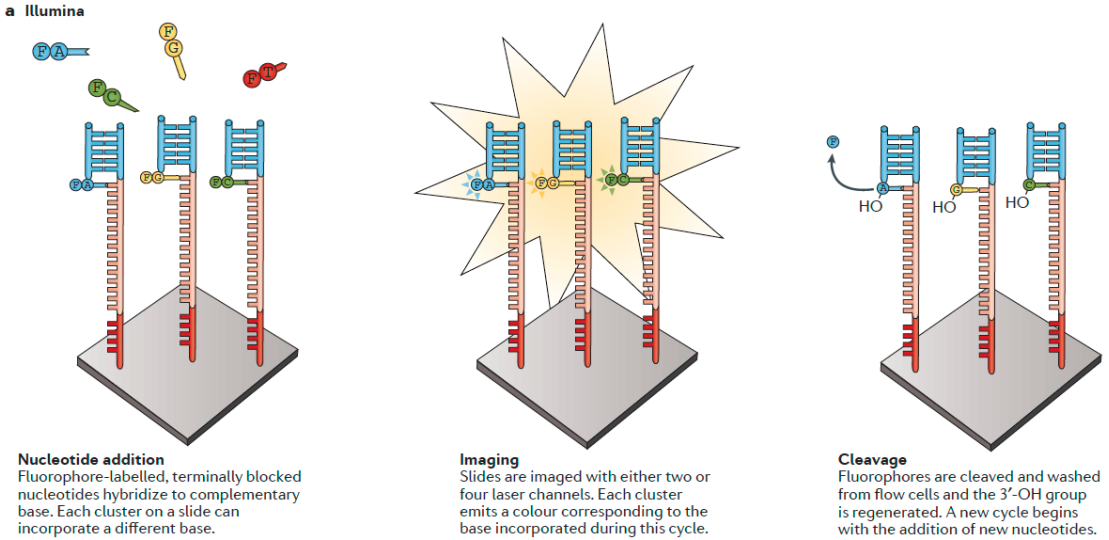
Full explanation: https://www.youtube.com/watch?v=PPEKybWYOBA

***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)

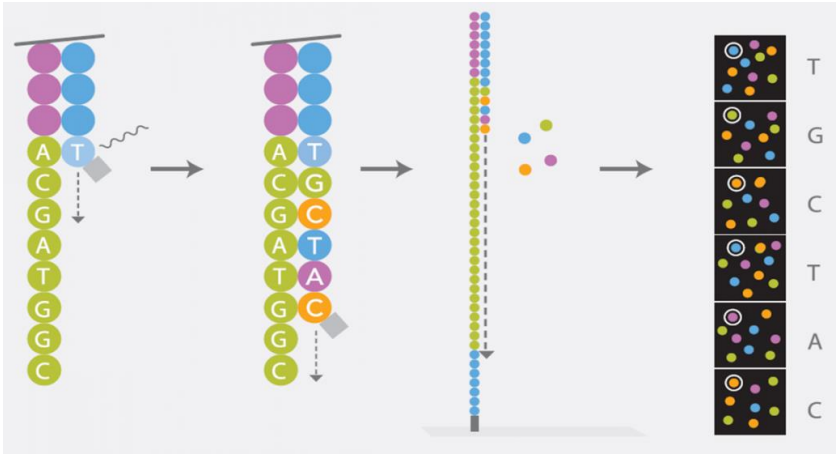
CRT: Cyclic Reversible Termination, used by Illumina, GeneReader (QiaGen), DNBSEQ (BGI)



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

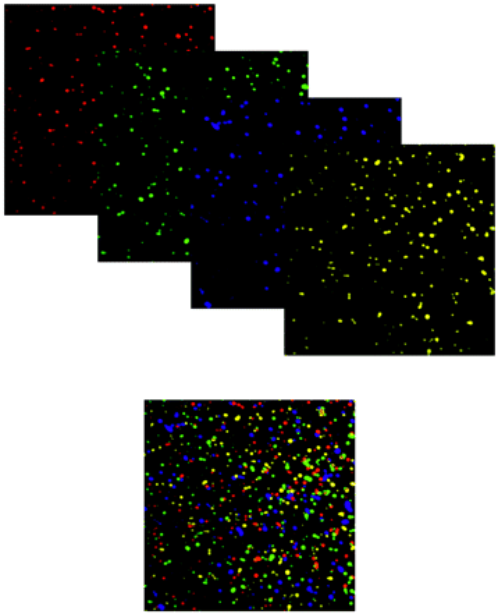
Sequencing Methods - By Synthesis (SBS)

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



Cycle 1
Cycle 2
Cycle 3
Cycle 4
Cycle 5
Cycle 6
Cycle 150

Store an image per cycle to reconstruct the sequence of each cluster



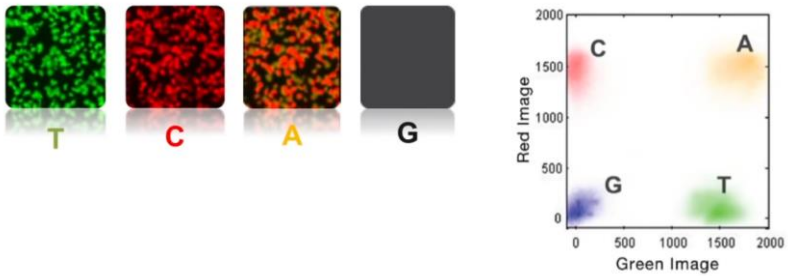
Source: Sara Goodwin *et al.* 2016 - *Nature Reviews Genetics* - https://doi.org/10.1038/nrg.2016.49

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

Sequencing Methods - By Synthesis (SBS)

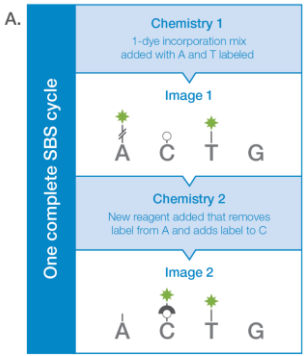
CRT: Cyclic Reversible Termination, used by Illumina, GeneReader (QiaGen), DNBSEQ (BGI)

Two-color sequencing



Two-channel SBS simplifies nucleotide detection by using two fluorescent dyes and two images to determine all four base calls. Images are taken using red and green filter bands**. Thymines are labeled with a green** fluorophore, **cytosines are labeled with a red** fluorophore, and **adenines are labeled with both red and green** fluorophores. **Guanines are permanently dark**. The MiniSeq™, NextSeq™, and NovaSeq™ Systems use two-channel chemistry.

Single color sequencing



B.

Image 1	Image 2	Result
ON	OFF	A
OFF	ON	C
ON	ON	T
OFF	OFF	G

The iSeq™ 100 System combines CMOS technology with innovative one-channel SBS chemistry to deliver high-accuracy data in a compact system

Sequencing Methods - By Synthesis (SBS)

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

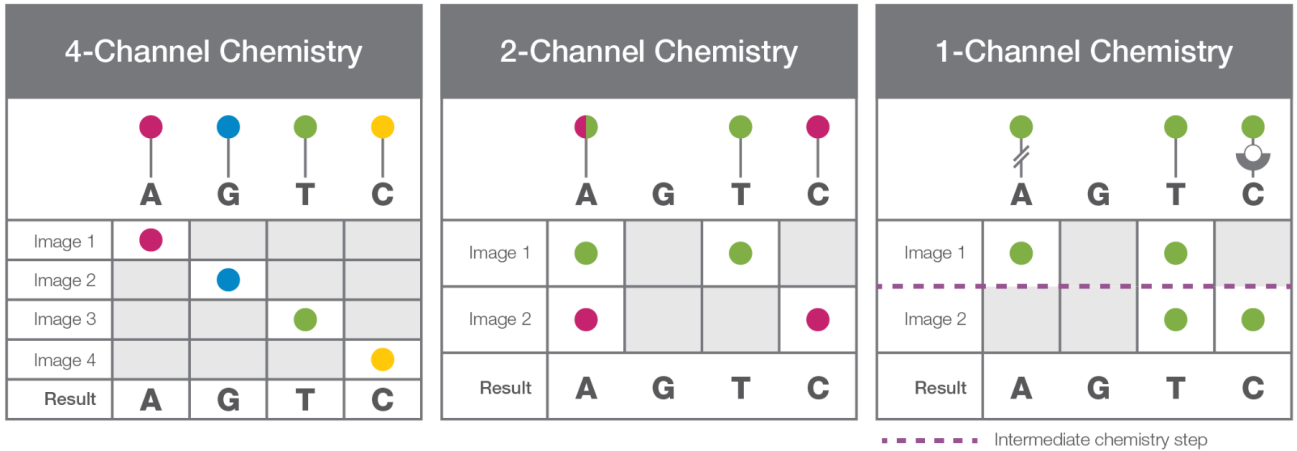
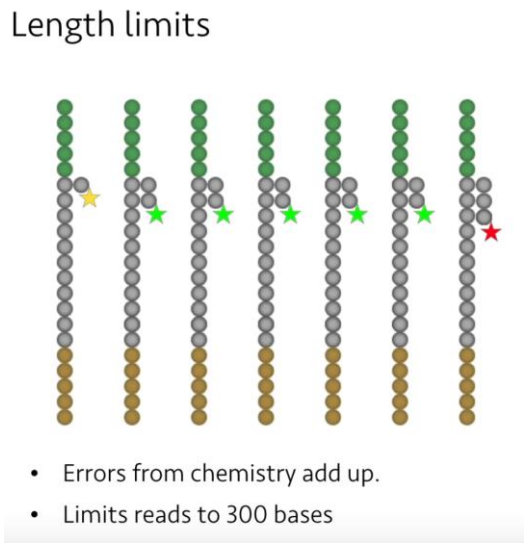


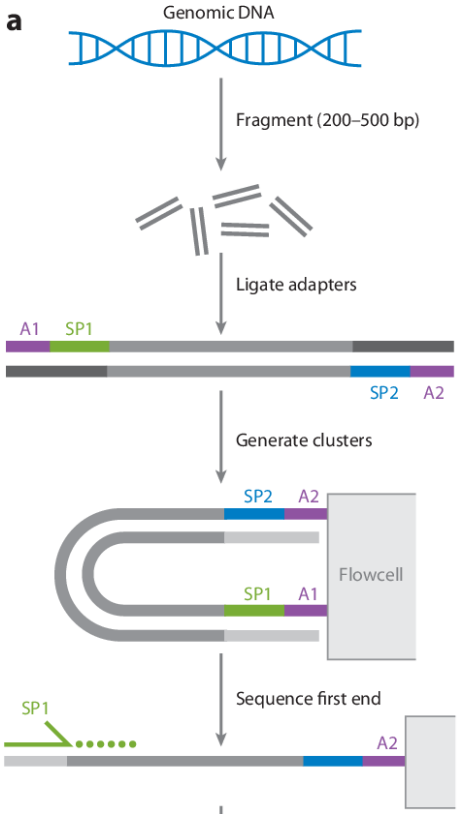
Figure 2: Four-, Two-, and One-Channel Chemistry—Four-channel chemistry uses a mixture of nucleotides labeled with four different fluorescent dyes. Two-channel chemistry uses two different fluorescent dyes, and one-channel chemistry uses only one dye. The images are processed by image analysis software to determine nucleotide identity.

Sequencing Methods - By Synthesis (SBS)

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



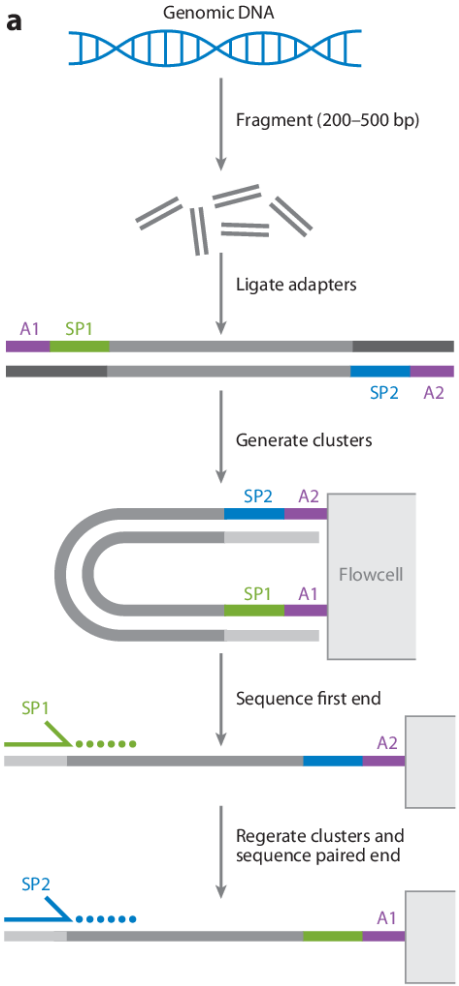
Single-End vs. Paired-End (Illumina, DNBSEQ)



Single-End

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

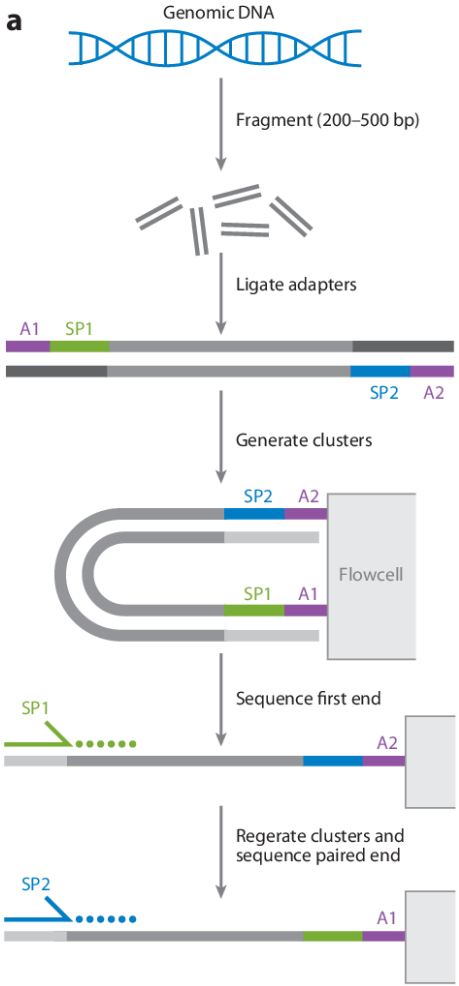
Single-End vs. Paired-End (Illumina, DNBSEQ)

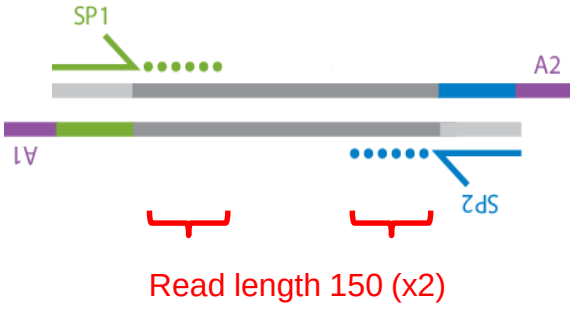


Paired-End

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)





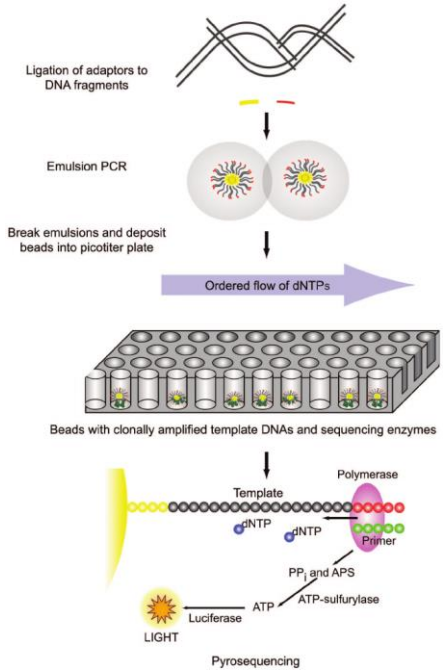
Ideally the insert length (genomic DNA fragment) should be longer than read length x 2

Paired-End

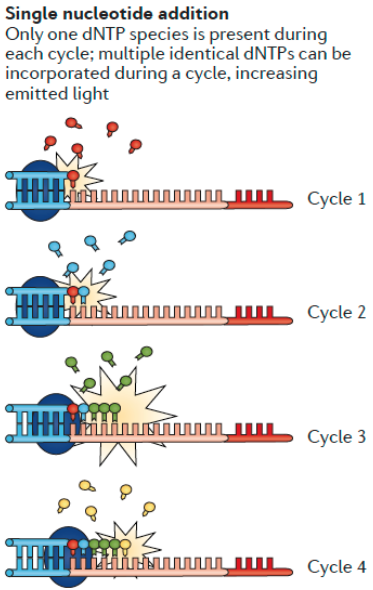
Source: Elaine R. Mardis 2013 - *Annual Review of Analytical Chemistry* - https://doi.org/10.1146/annurev-anchem-062012-092628

Sequencing Methods - By Synthesis (SBS)

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



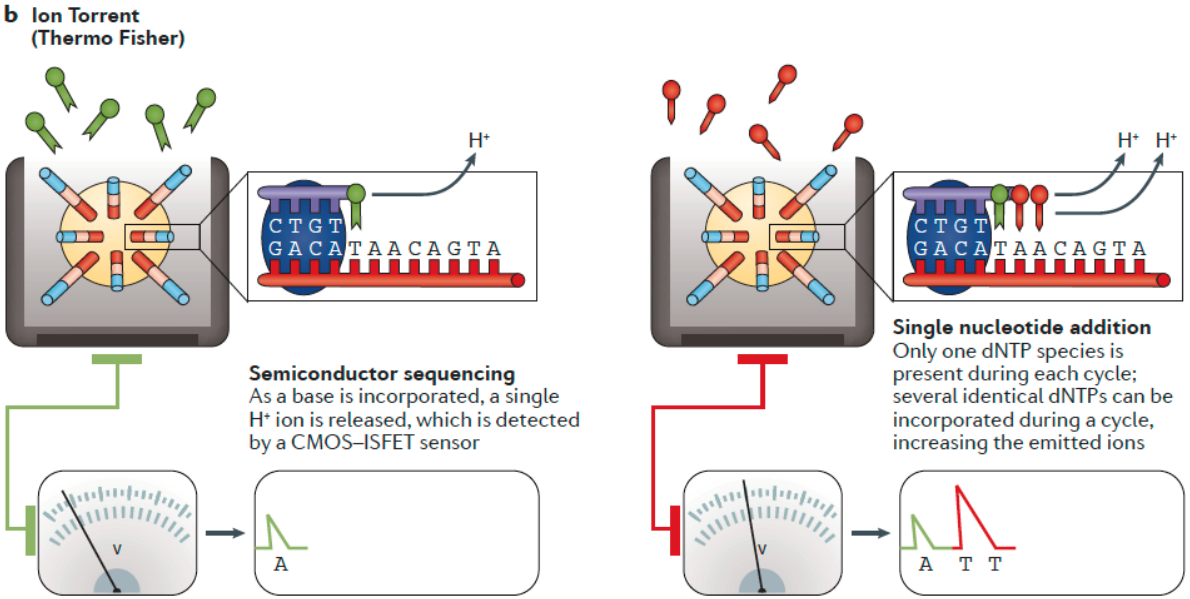
Source: https://doi.org/10.1373/clinchem.2008.112789



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:

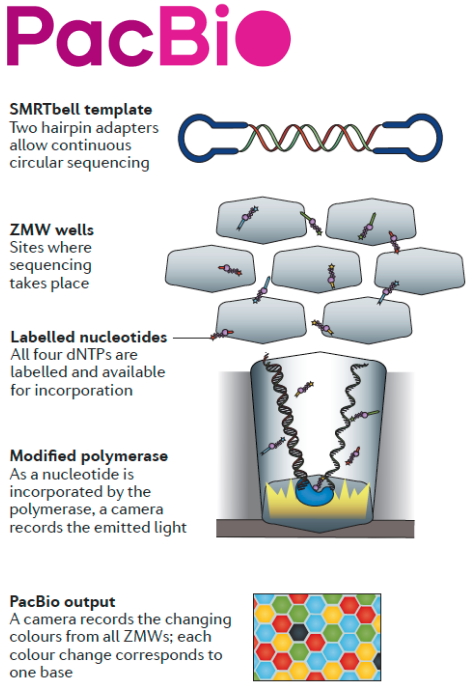
Illumina [CRT]:
https://www.youtube.com/watch?v=fCd6B5HRaZ8

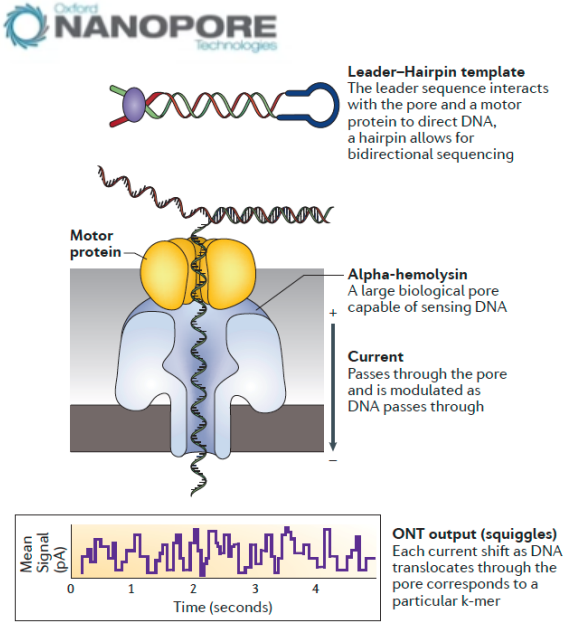
DNBSEQ (BGI) [CRT]:
https://www.youtube.com/watch?v=RGcpftDHpng&t

454 pyrosequencing [SNA]:
https://www.youtube.com/watch?v=bNKEhOGvcaI

IonTorrent [SNA]:
https://www.youtube.com/watch?v=ZL7DXFPz8rU

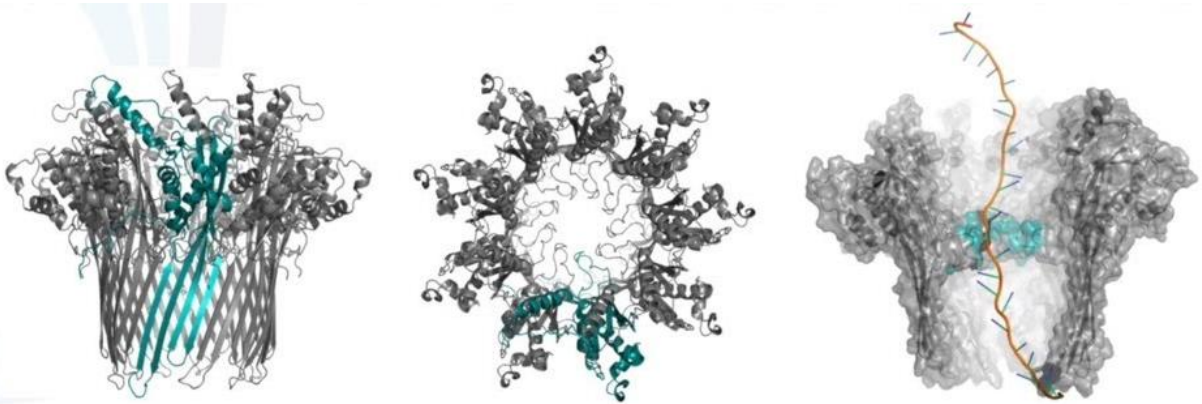
Third Generation:
Parallel Sequencing of Long Reads in Real Time





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

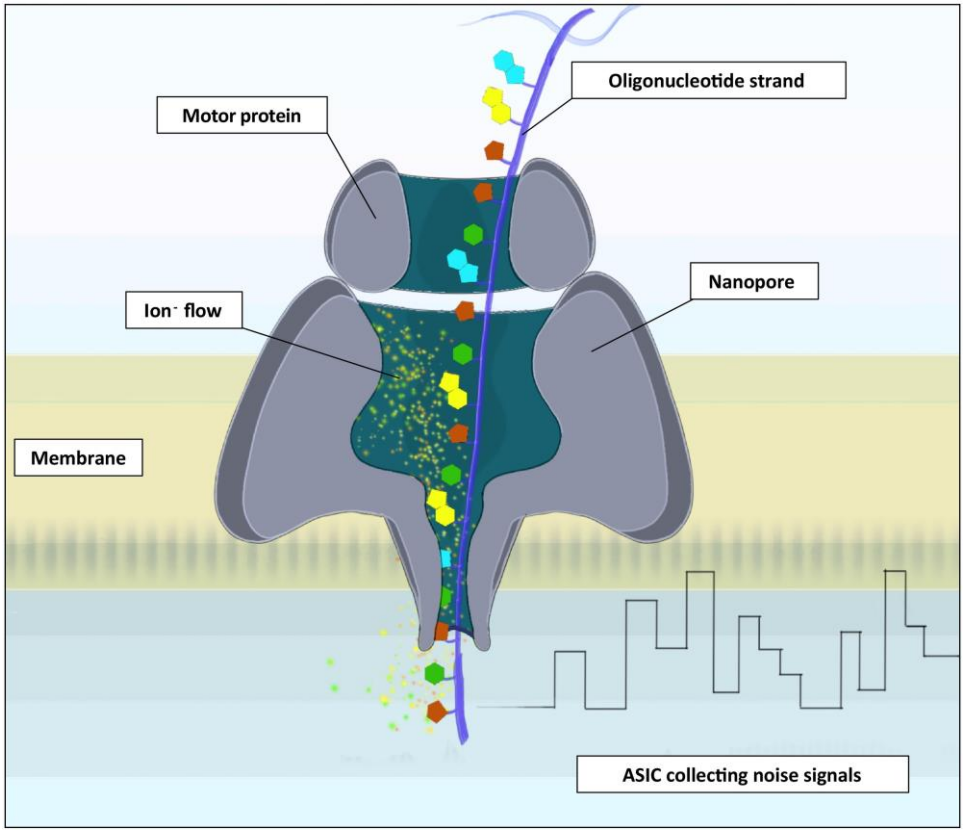
Third Generation: Oxford Nanopore



Third Generation: Oxford Nanopore

- **Really long reads**
- Can directly sequence RNA
- Sensing based on electronics
- No deterioration of the signal over time

- High error rates
  - Biased errors

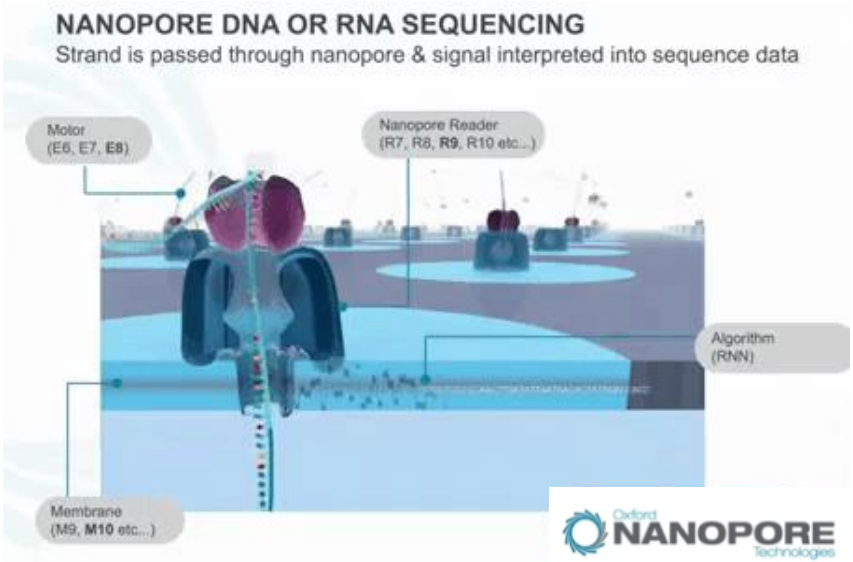


Trends in Genetics

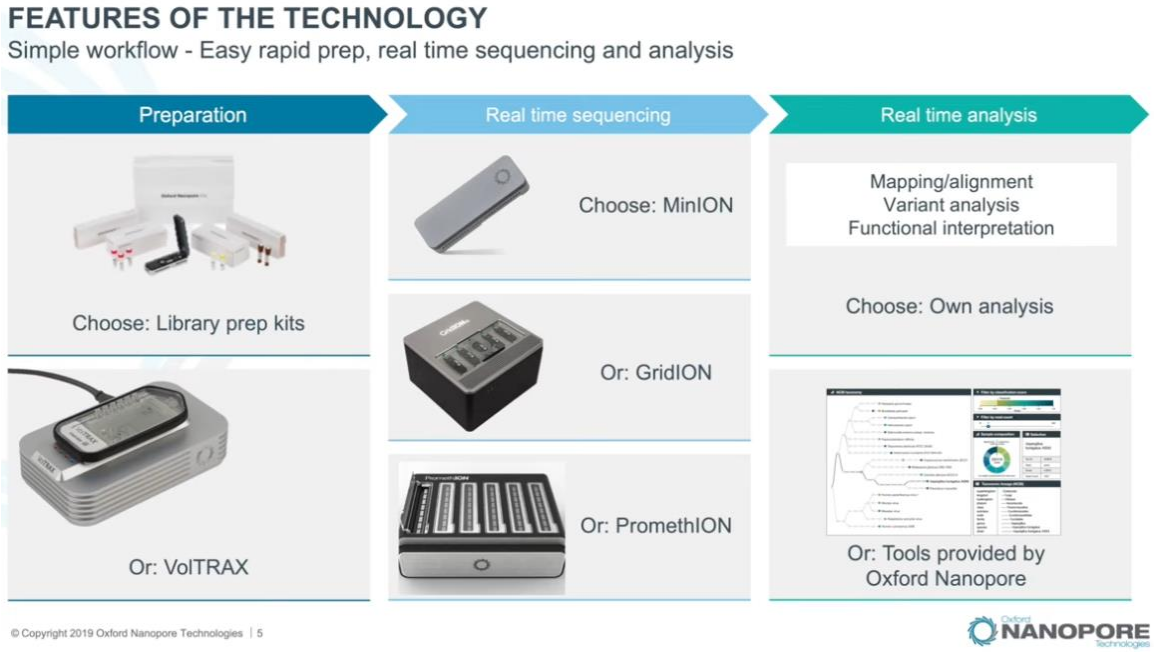
https://doi.org/10.1016/j.tig.2021.11.003

Third Generation: Oxford Nanopore

- Electronically resistant membrane
- Motor protein
  - Added to the end of dsDNA templates
  - Unzips dsDNA
  - Allows molecule to pass through the pore at certain speed DNA (450 bp/s) RNA (70 bp/s)
- Nanopore reader
  - Characteristic disruption of the electrical current created by nucleotide kmers (4) in the pore
- Base calling
  - Process raw current signal using Recurrent Neural Network algorithm



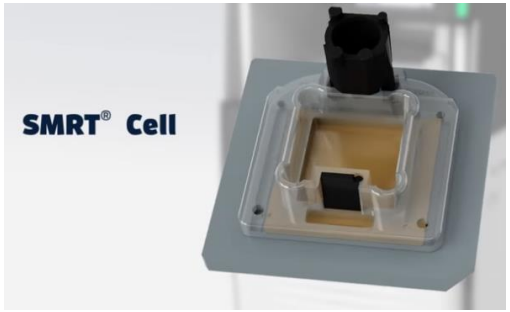
Third Generation: Oxford Nanopore



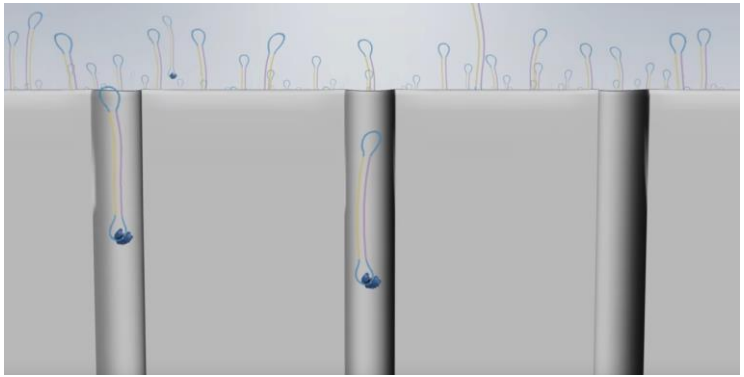
Third Generation: Pacific Biosciences

Single Molecule Real-Time (SMRT®) sequencing

Third Generation: Pacific Biosciences



Single Molecule Real-Time (SMRT®) sequencing



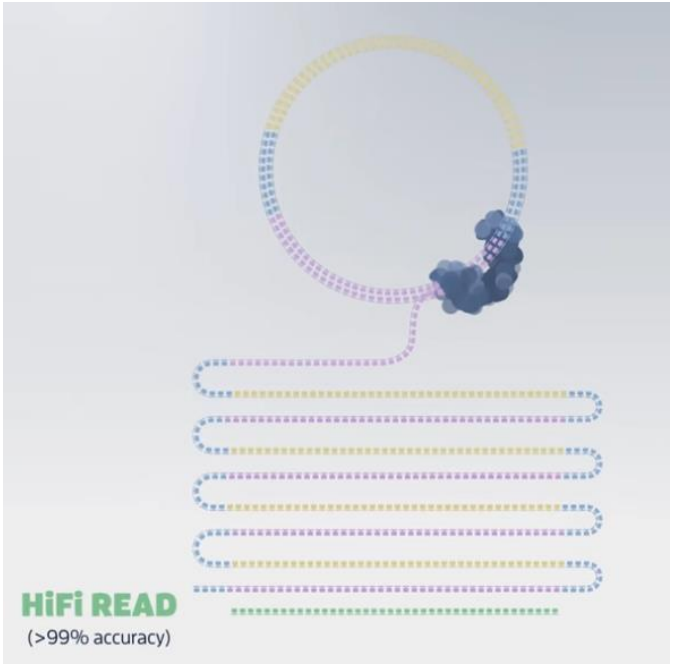
Third Generation: Pacific Biosciences

Single Molecule Real-Time (SMRT®) sequencing

Third Generation: Pacific Biosciences

Single Molecule Real-Time (SMRT®) sequencing

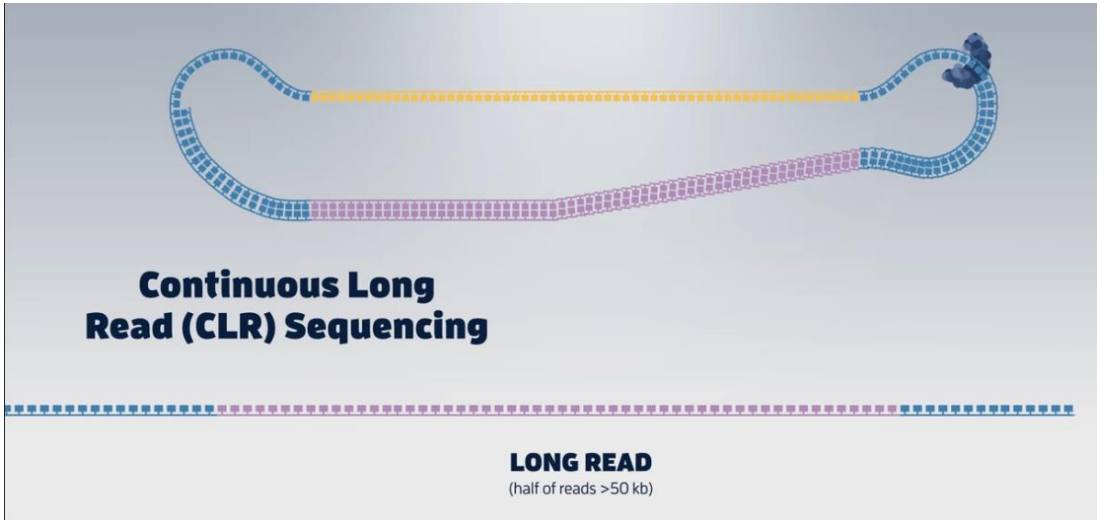
Circular Consensus Sequencing (CCS)
For highly accurate long reads



Third Generation: Pacific Biosciences

Single Molecule Real-Time (SMRT®) sequencing

Continuous Long Read (CLR)
For longest possible reads



Third Generation:
Parallel Sequencing of Long Reads in Real Time

Further explanation videos:

PacBio:

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

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

https://www.youtube.com/watch?v=_lD8JyAbwEo

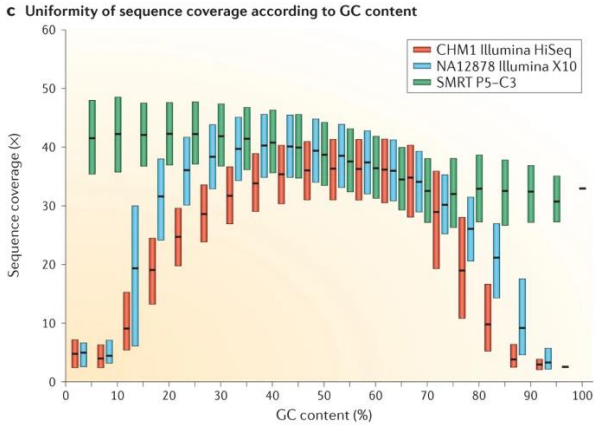
Nanopore:

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

https://www.youtube.com/watch?v=CGWZvHIi3i0

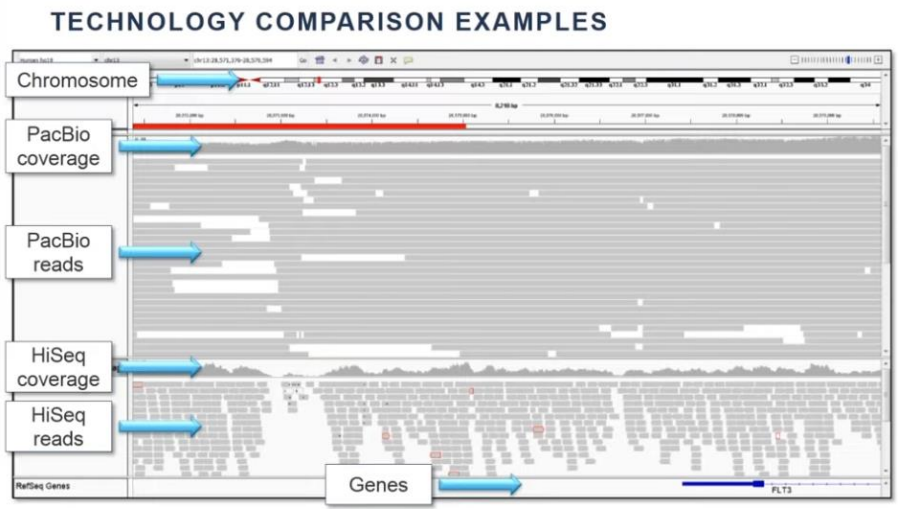
https://www.youtube.com/watch?v=sv9fFeSd3kE

2nd vs 3rd Generation:



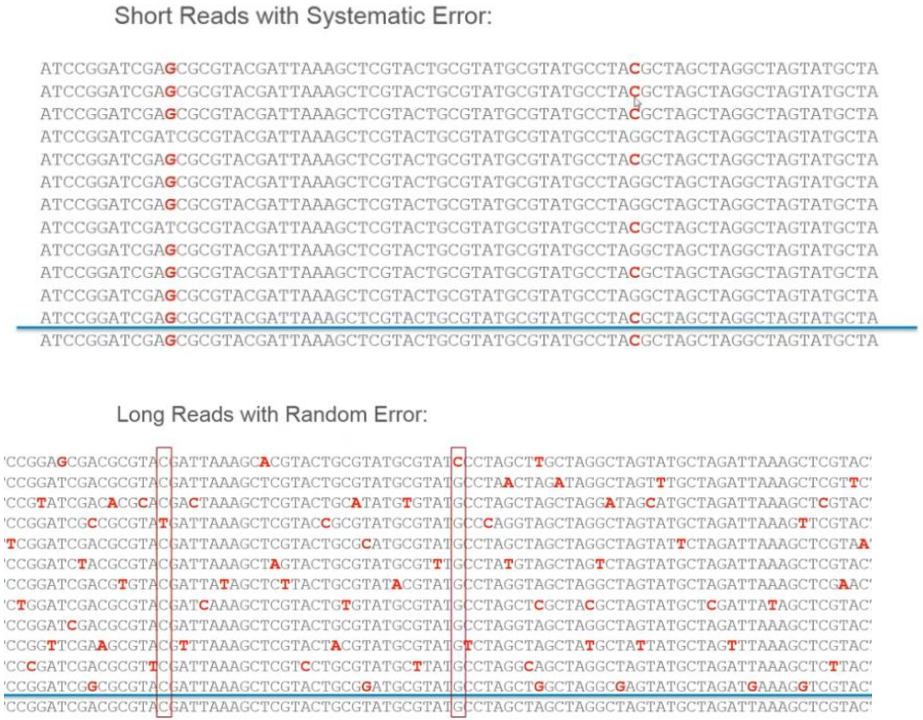
Nature Reviews | Genetics
https://doi.org/10.1038/nrg3933

Uniform coverage



2nd vs 3rd Generation:

Systematic vs random errors



Further explanation videos:
https://www.youtube.com/watch?v=mI0Fo9kaWqo

Illumina, ONT, PacBio

Introduction to NGS data analysis