Illumina sequencing technology

Céline Keime
keime@igbmc.fr

Illumina sequencing technology

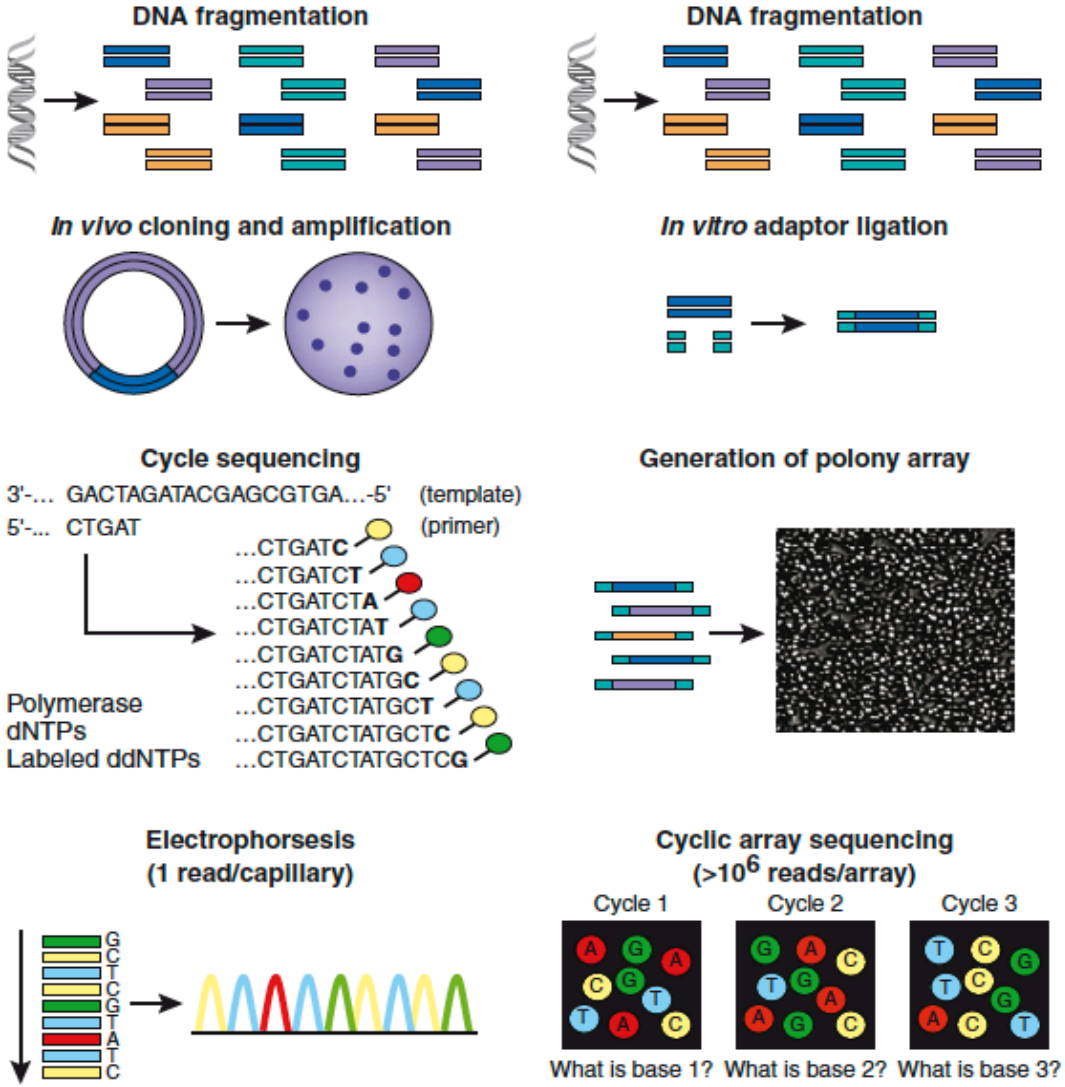
- Introduction to second generation sequencing
- Library preparation
- Amplification
- Sequencing
- Illumina sequencers and throughput
- Comparison between different generations of sequencing technologies

Illumina sequencing technology

- Introduction to second generation sequencing
- Library preparation
- Amplification
- Sequencing
- Illumina sequencers and throughput
- Comparison between different generations of sequencing technologies

Conventional vs second generation sequencing

Conventional sequencing 2nd generation sequencing



Three main technologies

- SOLiD DNA Sequencer Applied Biosystems
  - Shendure et al., 2005
- Genome Sequencer FLX+ Roche (formerly 454)
  - Margulies et al., 2005
- Illumina (formerly Solexa)
  - Fedurco et al., 2006

➔ Important decrease of

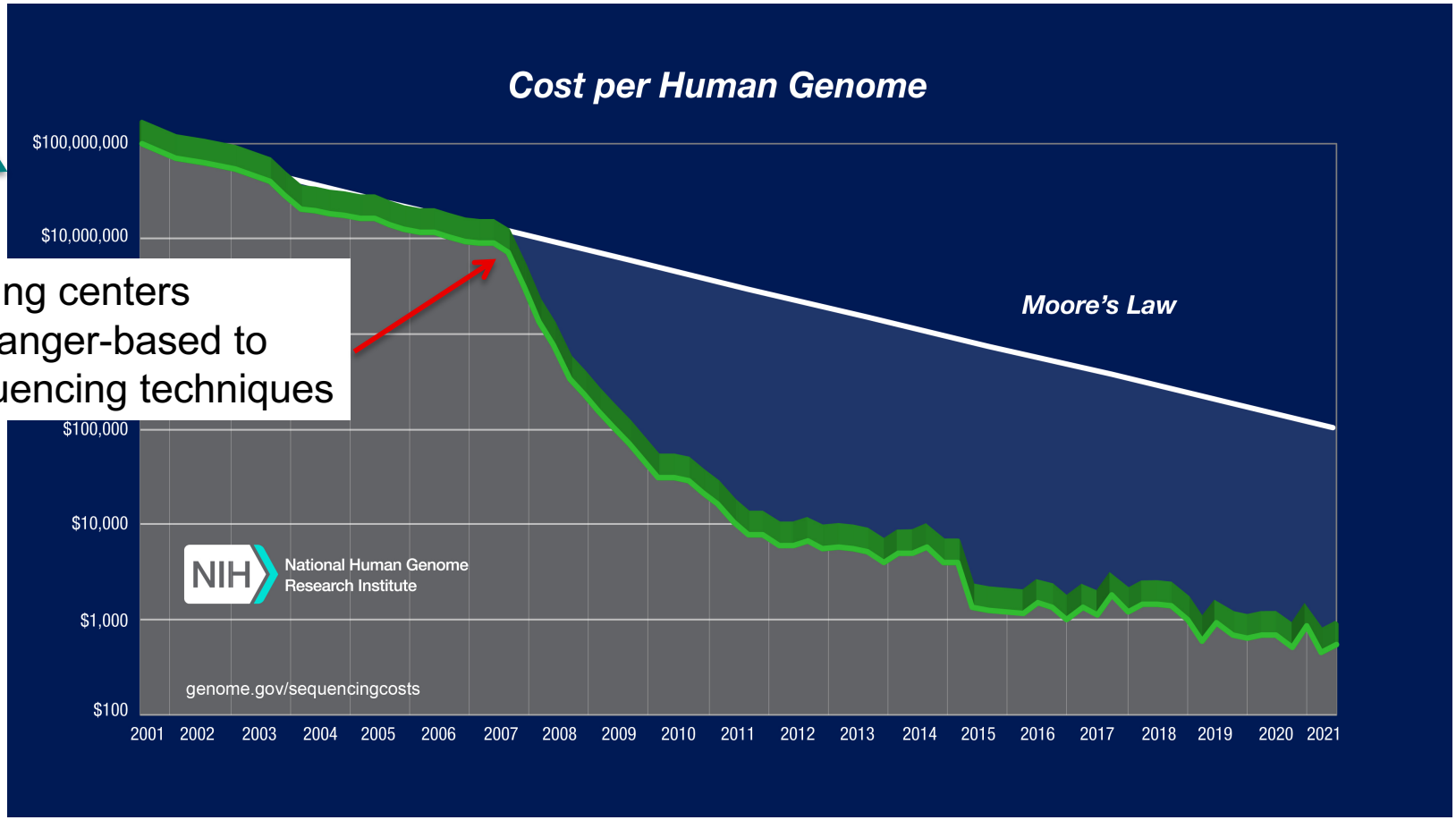
- Cost per base
- Time needed to obtain sequences

Shendure et al., Nature Biotechnology 2008

Decrease of sequencing costs

Log scale !

When the sequencing centers transitioned from Sanger-based to 2nd generation sequencing techniques

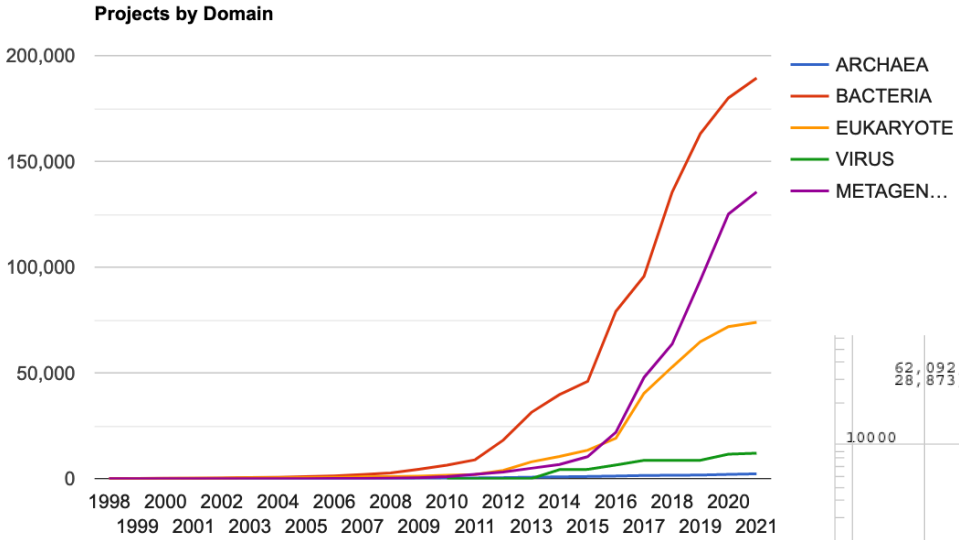


http://www.genome.gov/sequencingcosts/

Human genome with 30X coverage for Illumina sequencing
More information on how these costs are calculated :
https://www.genome.gov/sequencingcostsdata/

Increase of data volume

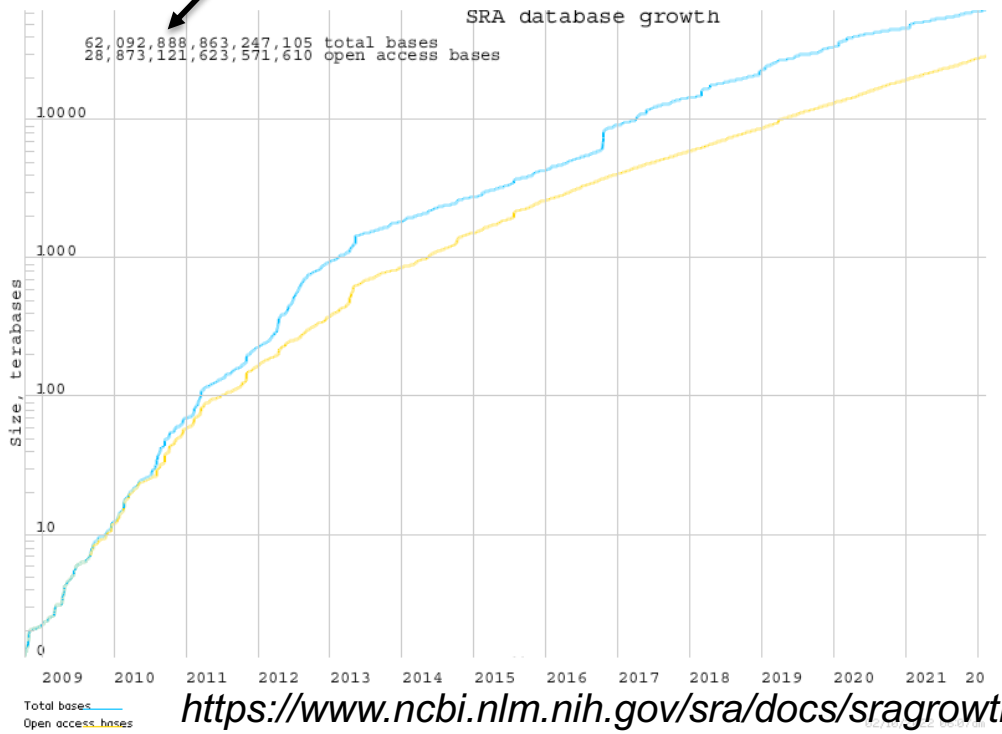
Projects in Genome Online Database



https://gold.jgi.doe.gov/statistics

Bases in Sequence Read Archive

~ 62,000 Tera bases

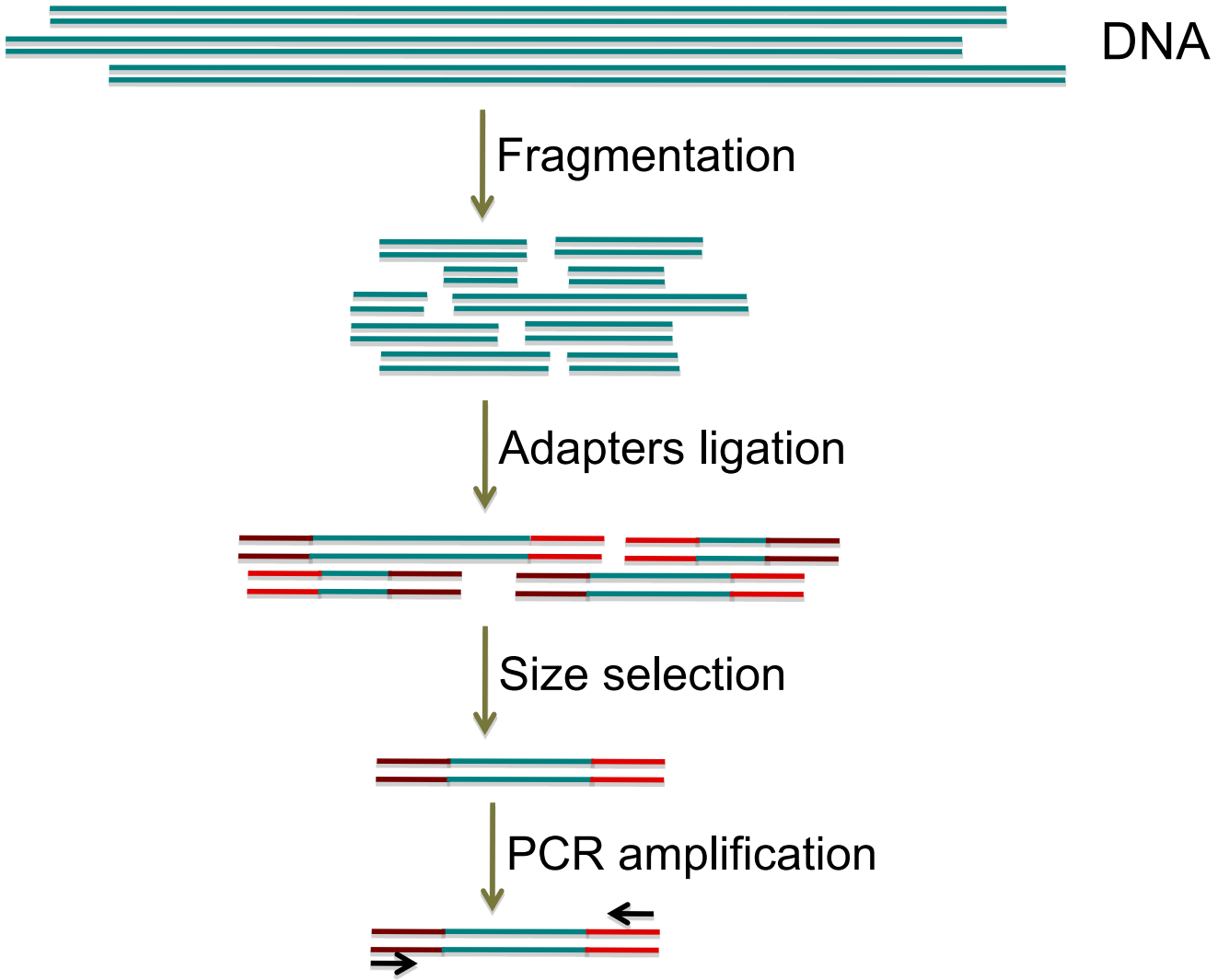


https://www.ncbi.nlm.nih.gov/sra/docs/sragrowth/

Illumina sequencing technology

- Introduction to second generation sequencing
- Library preparation
- Amplification
- Sequencing
- Illumina sequencers and throughput
- Comparison between different generations of sequencing technologies

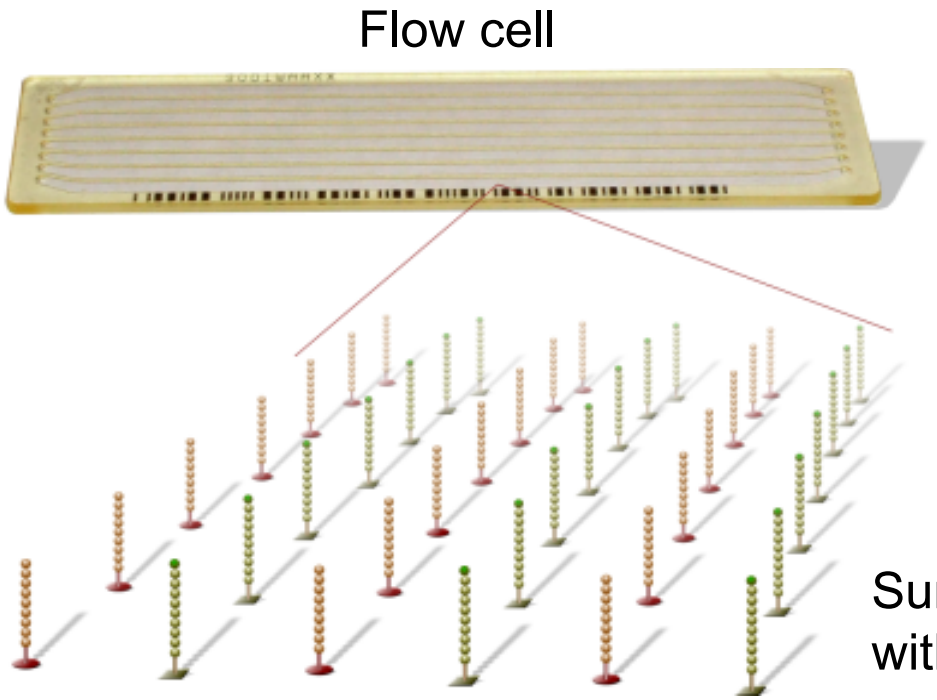
Library preparation



Illumina sequencing technology

- Introduction to second generation sequencing
- Library preparation
- Amplification
- Sequencing
- Illumina sequencers and throughput
- Comparison between different generations of sequencing technologies

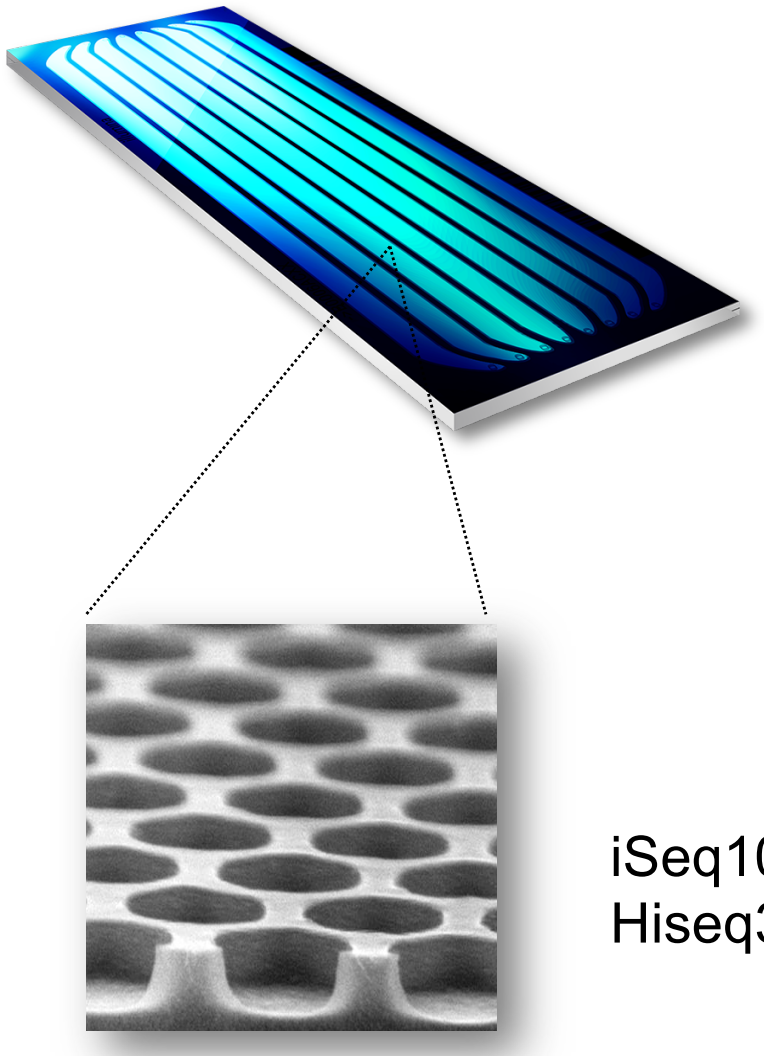
Fixation of DNA fragments on a solid support



https://www.illumina.com

Fixation of DNA fragments on a solid support

Patterned Flow Cell : billions of ordered wells

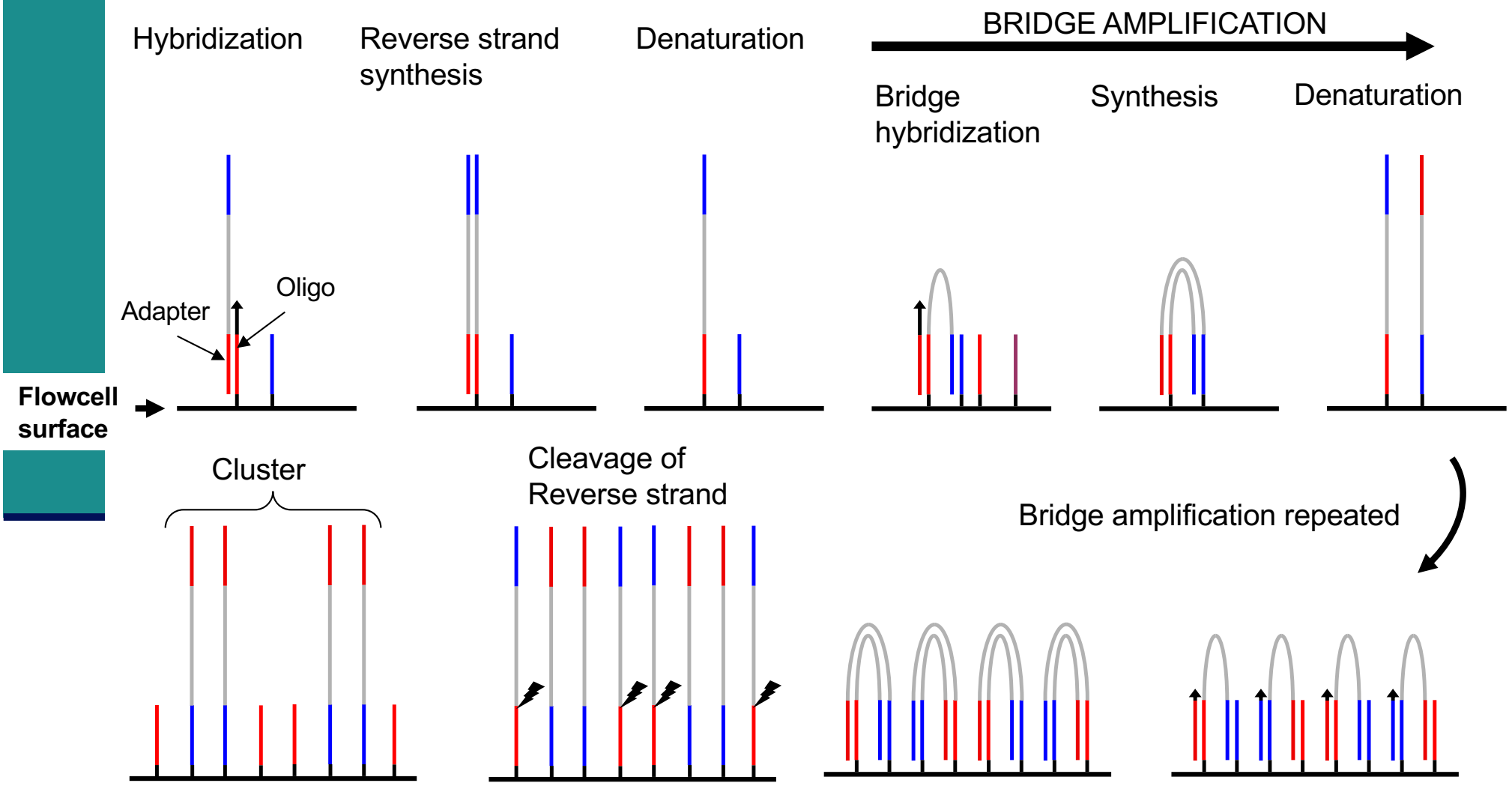


→ More reads
→ Faster run time

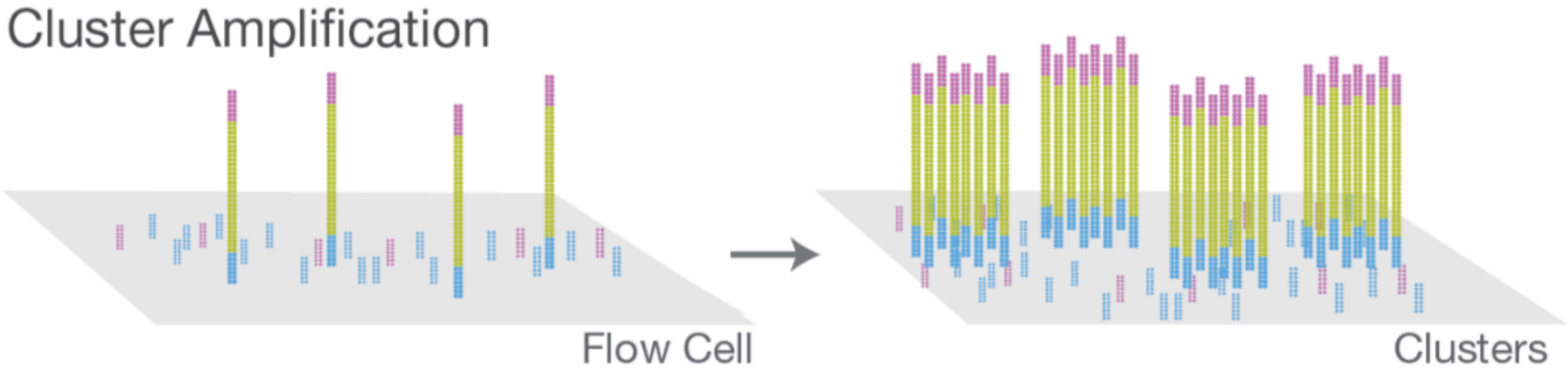
iSeq100, NextSeq1000/2000,
Hiseq3000/4000/X, NovaSeq6000 sequencers

https://www.illumina.com

Amplification : method



Amplification : result



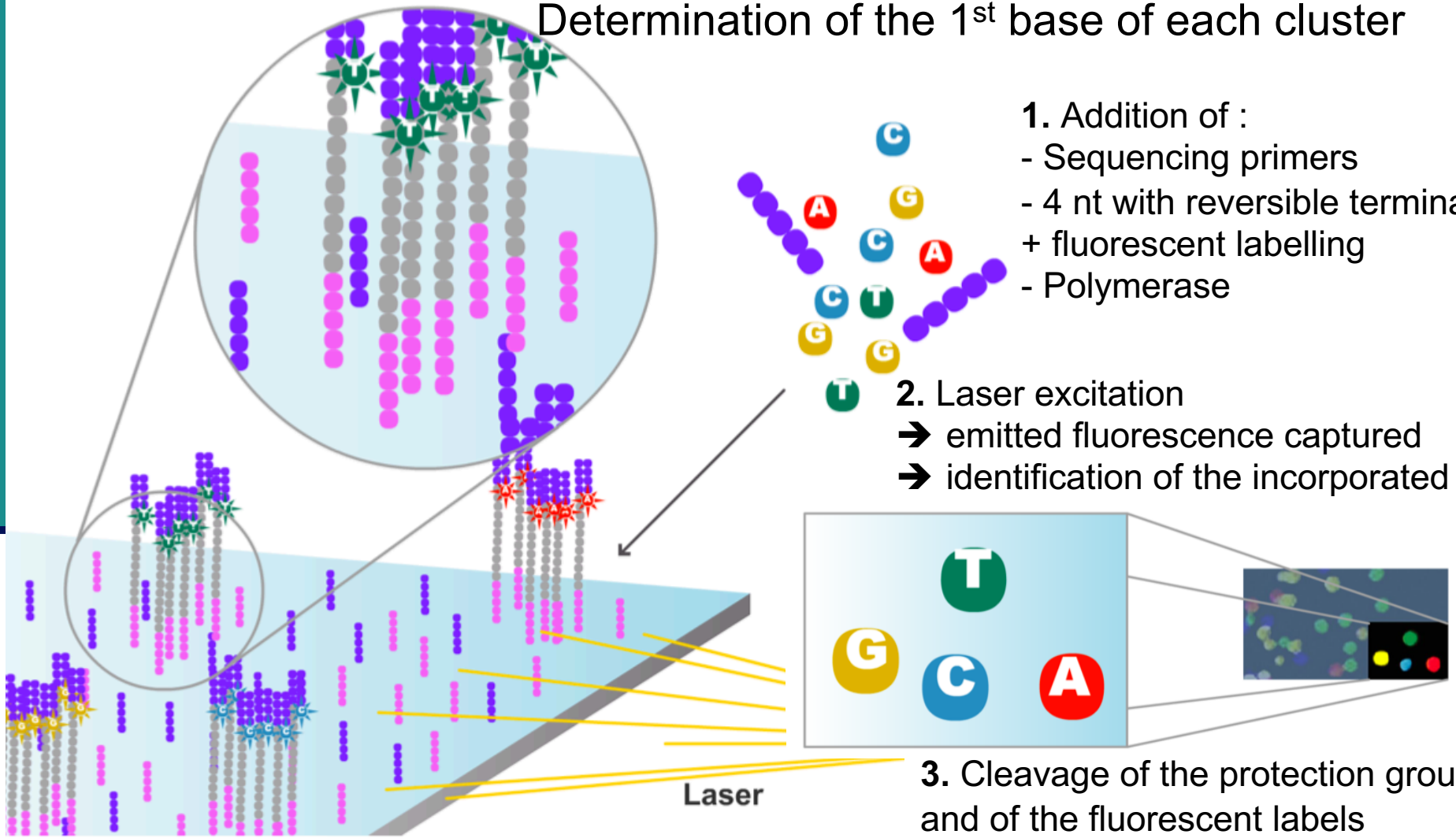
https://www.illumina.com

Illumina sequencing technology

- Introduction to second generation sequencing
- Library preparation
- Amplification
- Sequencing
- Illumina sequencers and throughput
- Comparison between different generations of sequencing technologies

Sequencing by synthesis

First sequencing cycle:
Determination of the 1st base of each cluster



1. Addition of :
- Sequencing primers
- 4 nt with reversible terminators
+ fluorescent labelling
- Polymerase

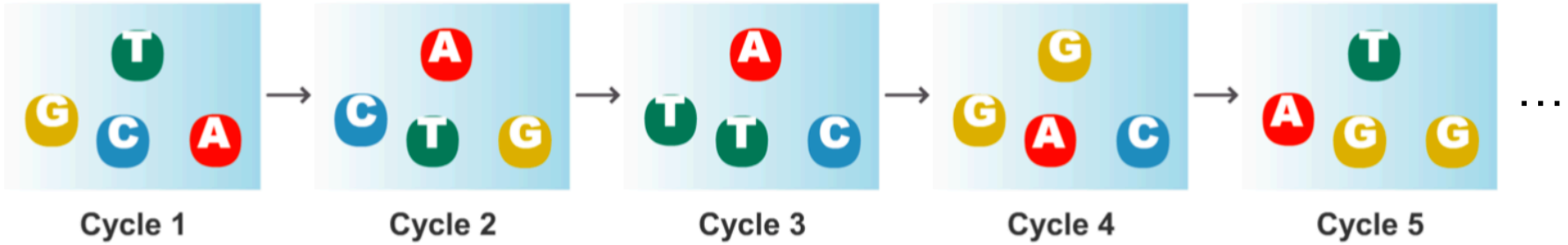
2. Laser excitation
➔ emitted fluorescence captured
➔ identification of the incorporated base

3. Cleavage of the protection groups and of the fluorescent labels

https://www.illumina.com

Sequencing by synthesis

As many sequencing cycles as the number of bases needed in the resulting read



eg cluster 1 = TAAGT…
➔ 1 read

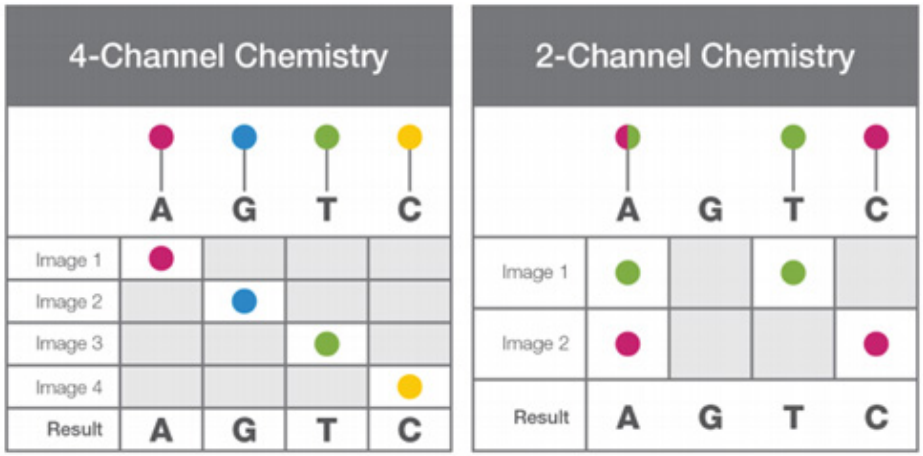
Each lane (top and bottom) is divided into swaths composed of tiles
→ 1 image per tile per color, for each cycle

https://www.illumina.com

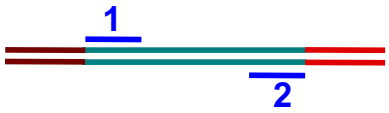
2-channel sequencing by synthesis

- Evolution of the original 4-channel SBS technology
  - 2 fluorescent dyes
  - Only 2 images per cycle instead of 4
  - Accelerates sequencing and data processing times

- Clusters seen in red images → C
- Clusters seen in green images → T
- Clusters observed in both red and green images → A
- Unlabelled clusters → G

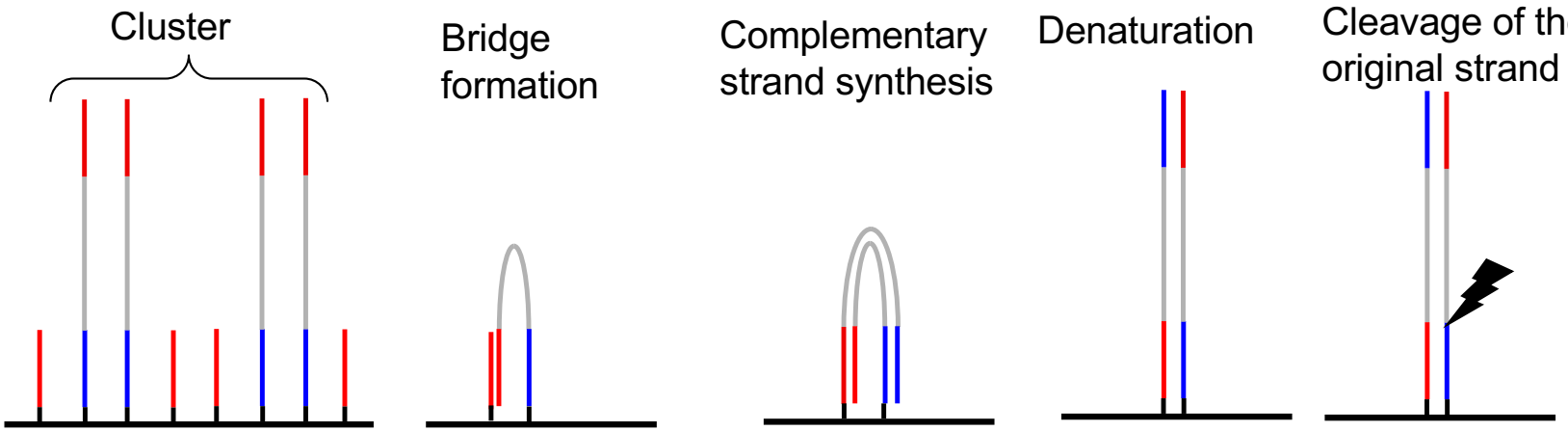


Paired-end sequencing



- Sequencing of read 1 : previously described method
- Then sequencing of read 2 :

For each molecule on the surface of the flow cell :



➔ Sequence the other end of the original molecule

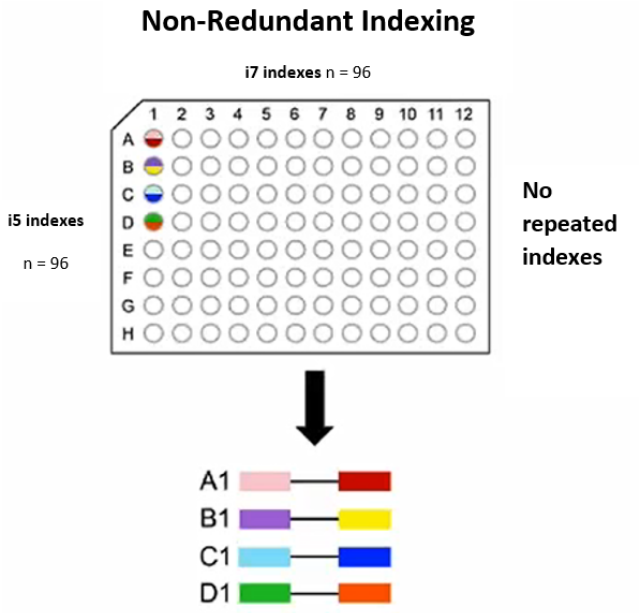
➔ Step performed on the flow cell in the sequencer :
keep the position of clusters
This information allows to link pairs of sequences

Multiplexing

- Add a barcode (index) specific to each sample
- Sequencing of several samples together
- Single indexing : 1 barcode

- Dual indexing : 2 barcodes

  - Unique dual indexes : distinct index sequences for each of the two indexes

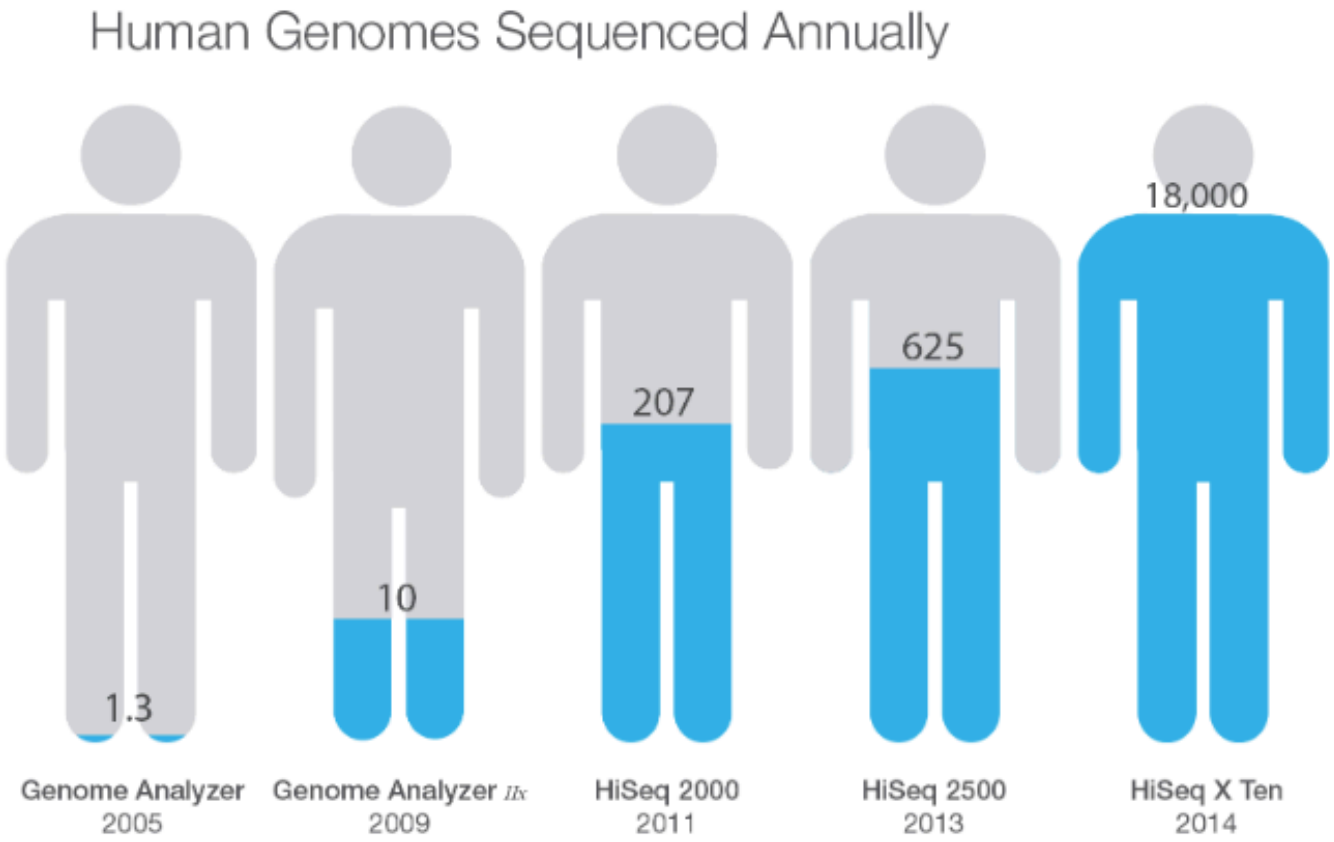


https://support.illumina.com/

Illumina sequencing technology

- Introduction to second generation sequencing
- Library preparation
- Amplification
- Sequencing
- Illumina sequencers and throughput
- Comparison between different generations of sequencing technologies

Illumina throughput over time



~ 48 human genomes in 44h

NovaSeq6000
2017

Capacity to sequence a 30X human genome annually
https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf

Illumina sequencers

Increasing system price and output
Decreasing price per Gb


iSeq 100 System


MiniSeq System


MiSeq Series


NextSeq Series


HiSeq 4000 System


HiSeq X Series‡


NovaSeq 6000 System

Benchtop sequencers

Production-scale sequencers

https://emea.illumina.com/systems/sequencing-platforms.html

Illumina sequencers



	iSeq 100	MiniSeq	MiSeq Series
Run Time	9.5–19 hrs	4–24 hours	4–55 hours
Maximum Output	1.2 Gb	7.5 Gb	15 Gb
Maximum Reads Per Run	4 million	25 million	25 million †
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp

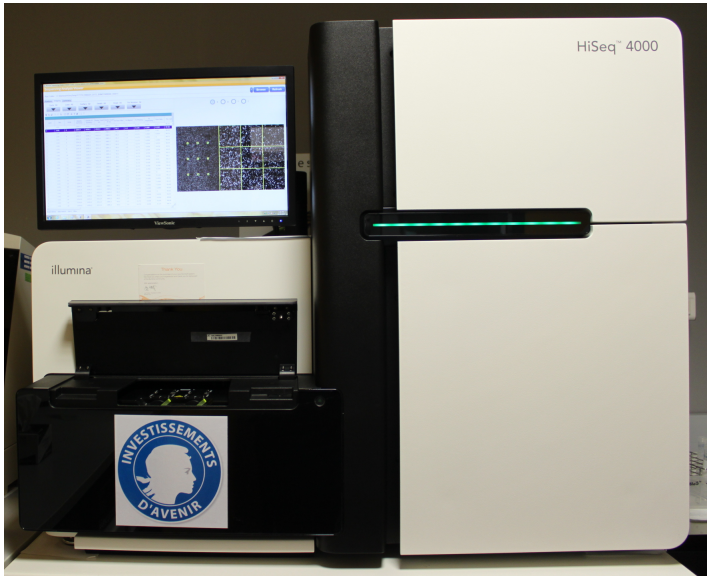


	NextSeq 550 Series	NextSeq 1000 & 2000	NovaSeq 6000
Run Time	12–30 hours	11-48 hours	~13 - 38 hours (dual SP flow cells) ~13–25 hours (dual S1 flow cells) ~16–36 hours (dual S2 flow cells) ~44 hours (dual S4 flow cells)
Maximum Output	120 Gb	330 Gb*	6000 Gb
Maximum Reads Per Run	400 million	1.1 billion*	20 billion
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 250**

From https://emea.illumina.com/systems/sequencing-platforms.html

Illumina Hiseq4000 throughput

- Up to 2x150 bp
- ~ 330 million single-end reads per lane
- 1 or 2 flow cells per run
- 8 lanes per flow cell
- Throughput per run with 2 flow cells
  - ~ 5 billion single-end reads
  - ~ 10 billion paired-end reads
  - ~ 1.5 Tbases per 2x150bp run
  - ~ 12 30X human genomes
  - ~ 128 RNA-seq (polyA+) or ChIP-seq (~ 30 million single-end reads)
  - ~ 128 human exomes per 2x100 run (> 60X mean coverage)



Illumina sequencing technology

- Introduction to second generation sequencing
- Library preparation
- Amplification
- Sequencing
- Illumina sequencers and throughput
- Comparison between different generations of sequencing technologies

Comparison between different generations of sequencing technologies

- 1st generation
  - Sequencing pre-amplified molecules one by one
- 2nd generation
  - Clonal amplification and sequencing of million of molecules at the same time
  - Amplification, RT needed ➔ bias
- 3rd generation
  - Nanopore sequencing, Pacific Biosciences
  - Main improvements
    - Long reads
    - No amplification
    - Direct RNA sequencing
  - Current drawbacks
    - Lower per read accuracy and number of reads than short-read sequencing