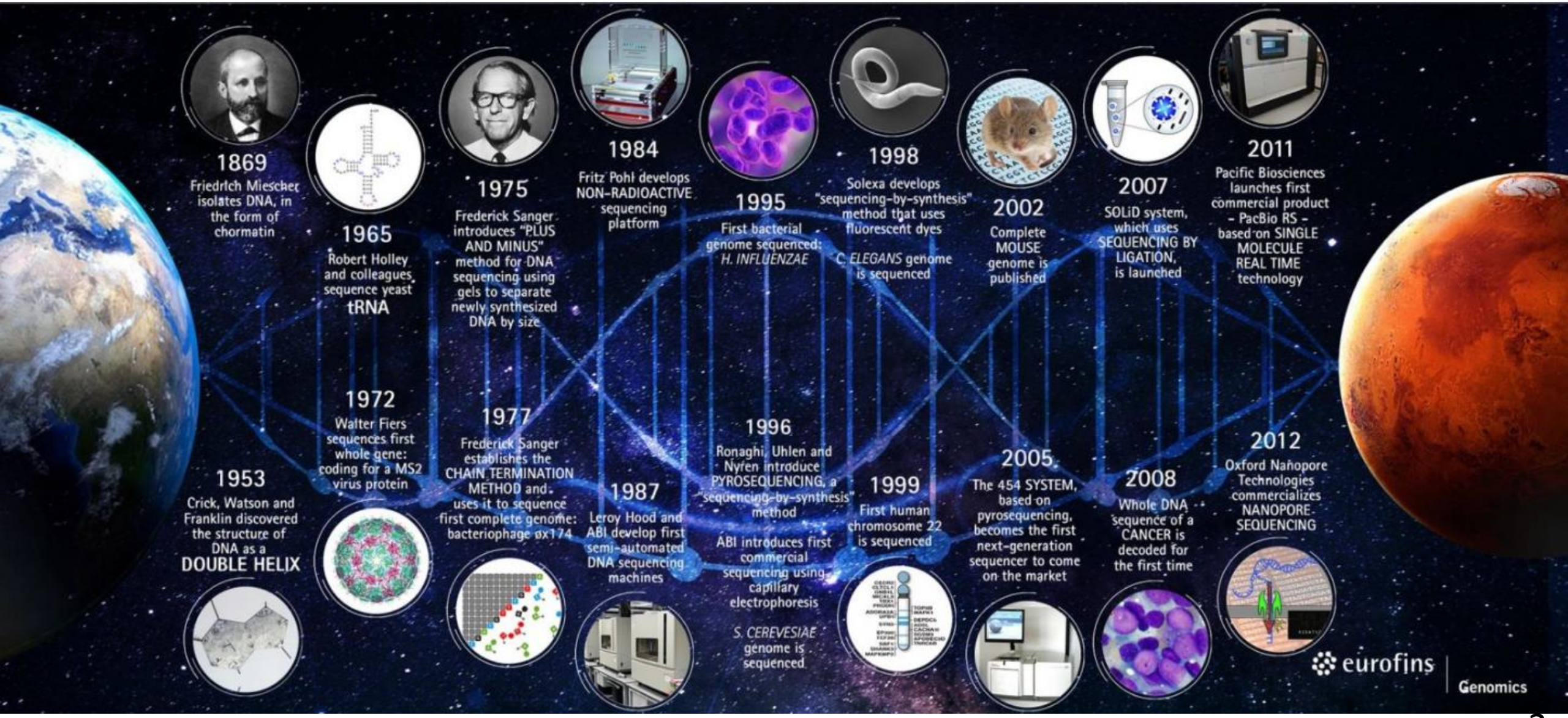
The next generation sequencing technology

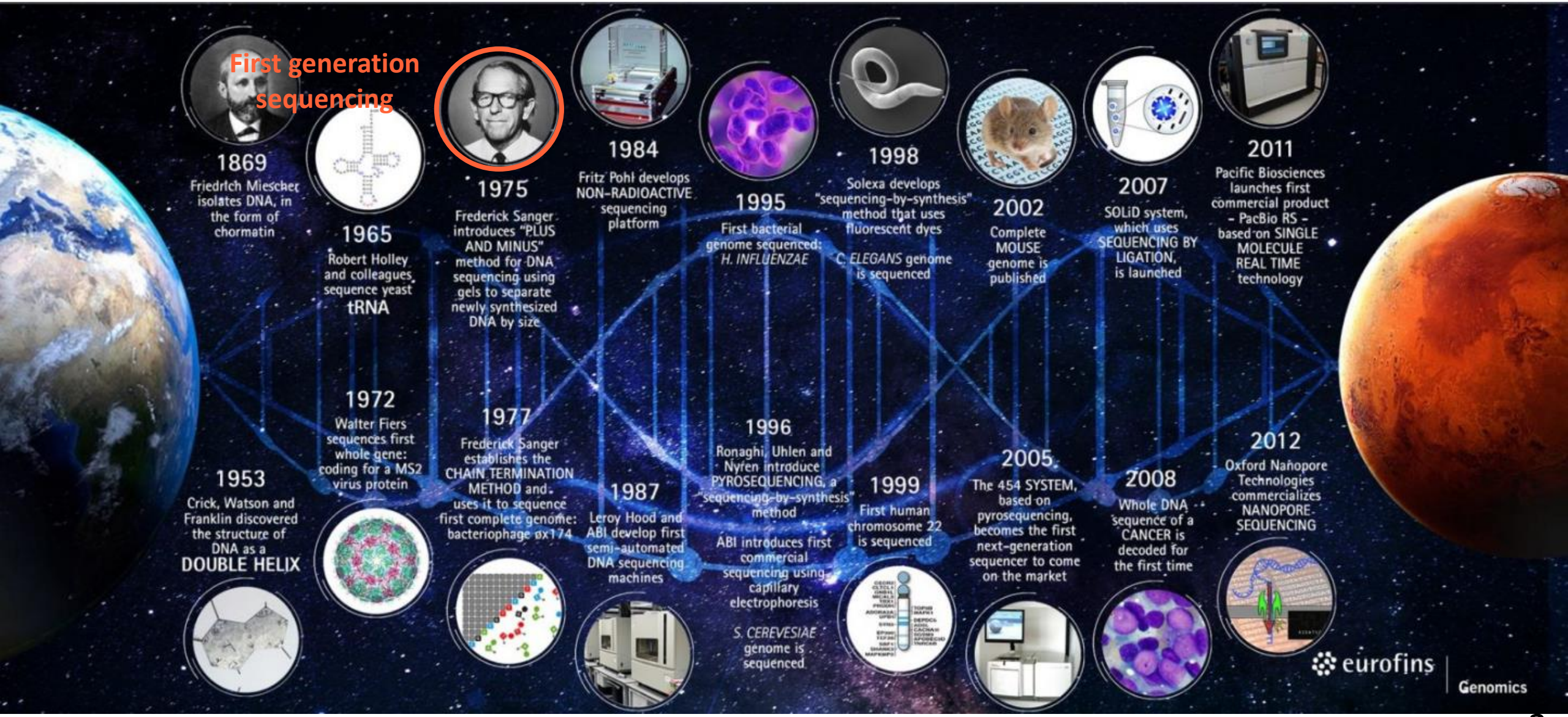
GENOMICS

BIGNAUD AMAURY
20/09/2023

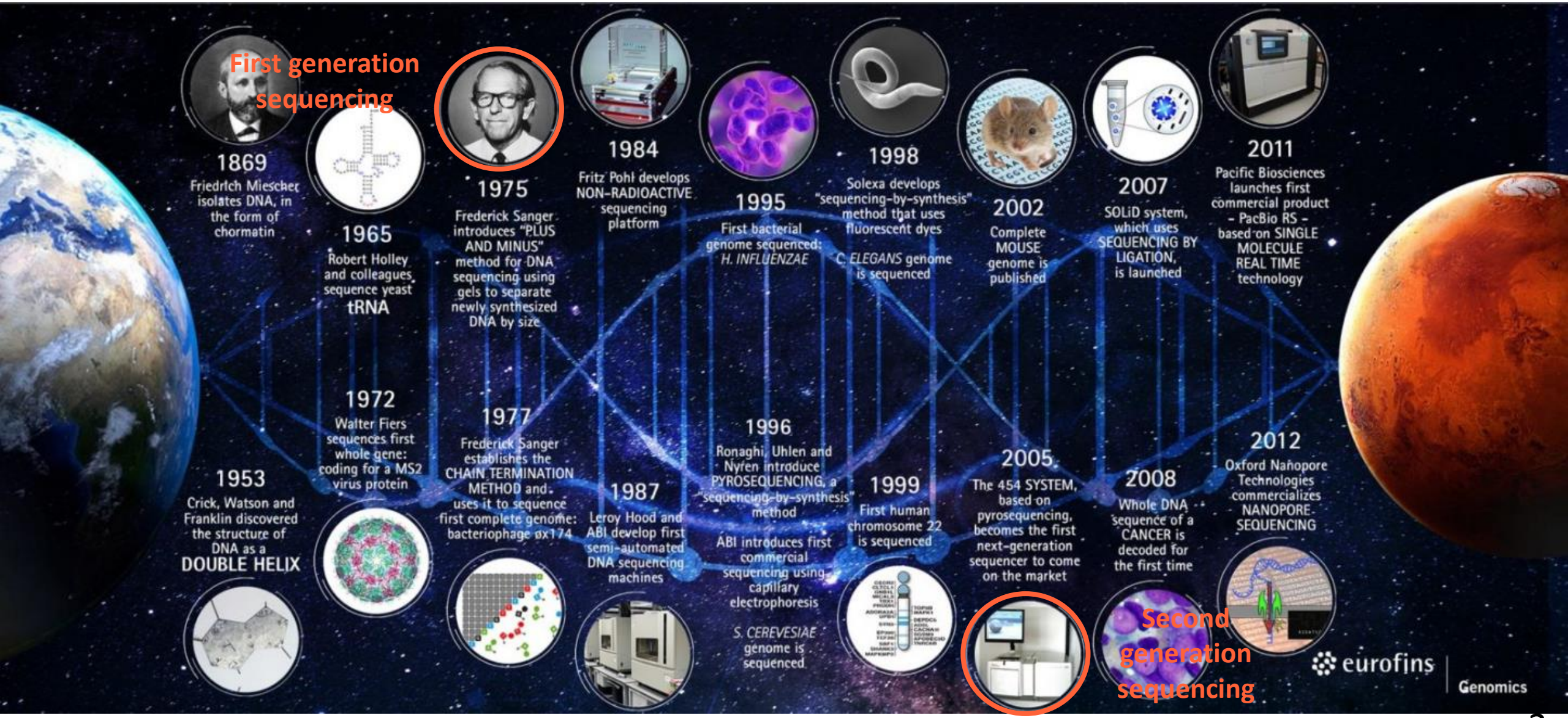
The history of sequencing



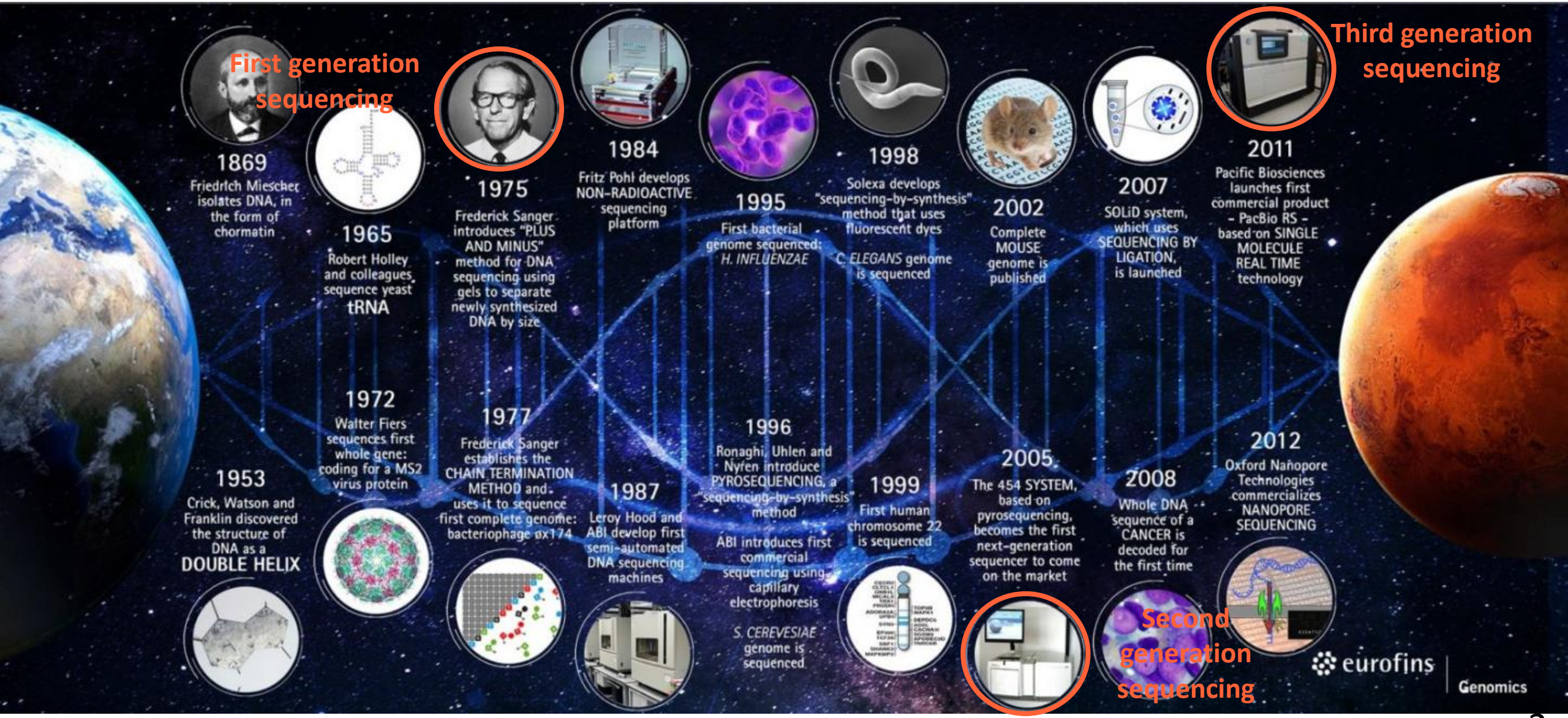
The history of sequencing



The history of sequencing



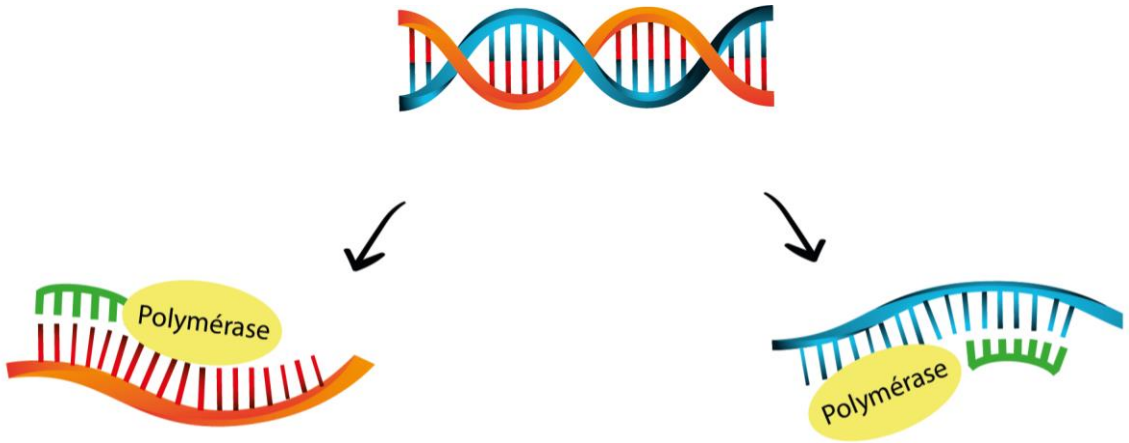
The history of sequencing



Sanger sequencing

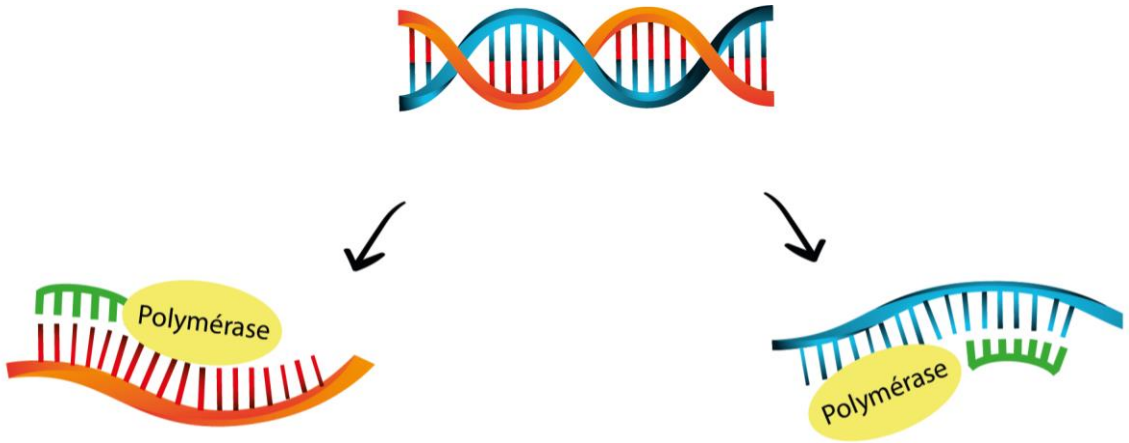
Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR)



Denaturation – 95°C

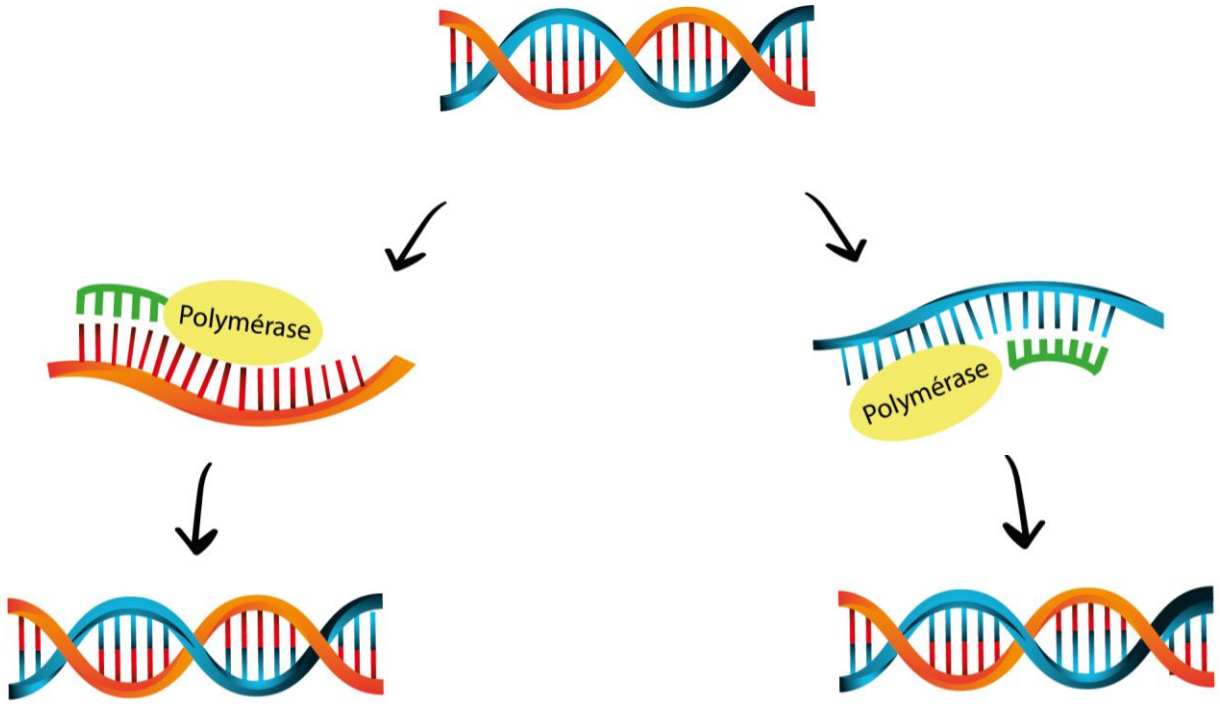
Polymerase chain reaction (PCR)



Denaturation – 95°C

Annealing – 68°C

Polymerase chain reaction (PCR)

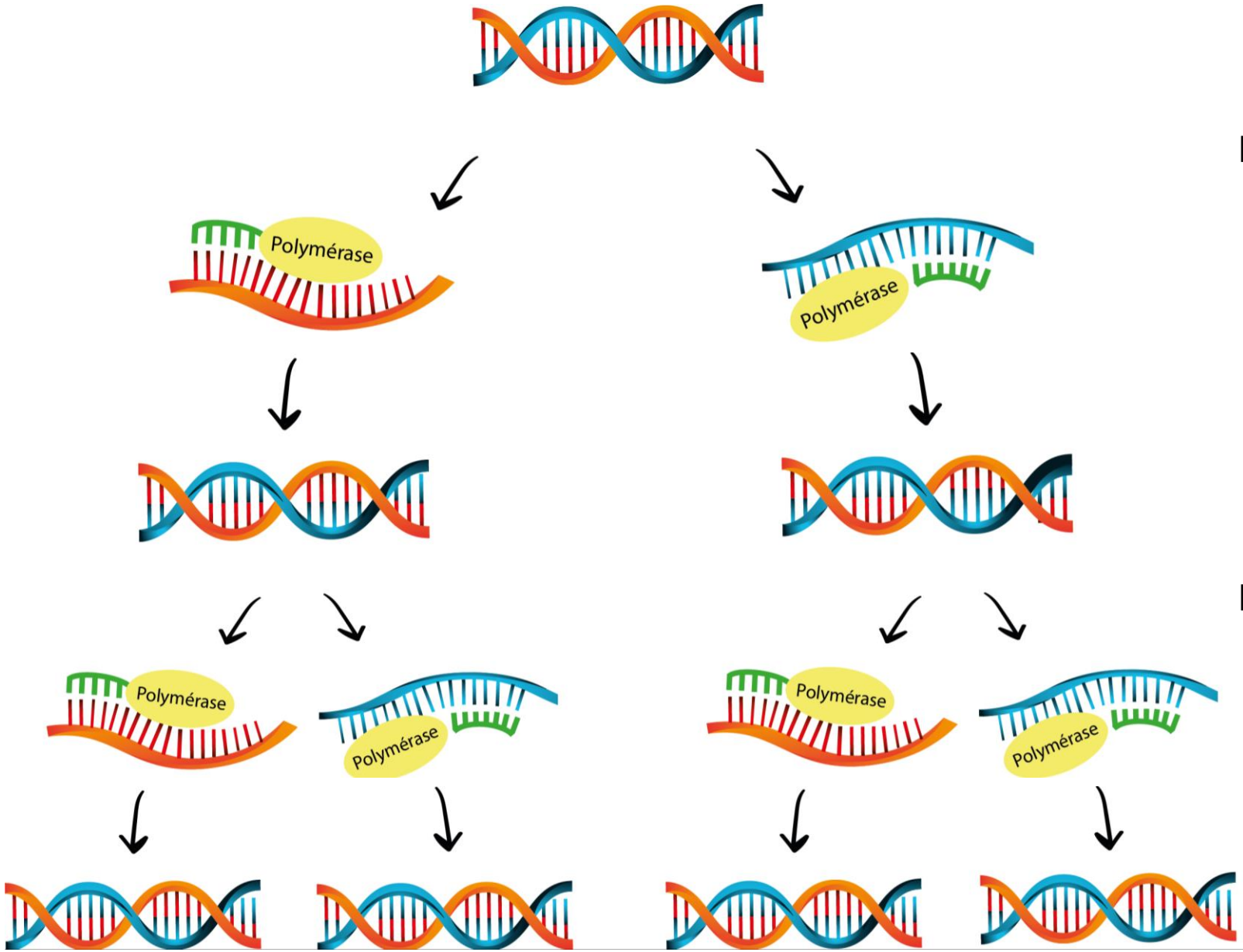


Denaturation – 95°C

Annealing – 68°C

Elongation – 72°C

Polymerase chain reaction (PCR)



Denaturation – 95°C

Annealing – 68°C

Elongation – 72°C

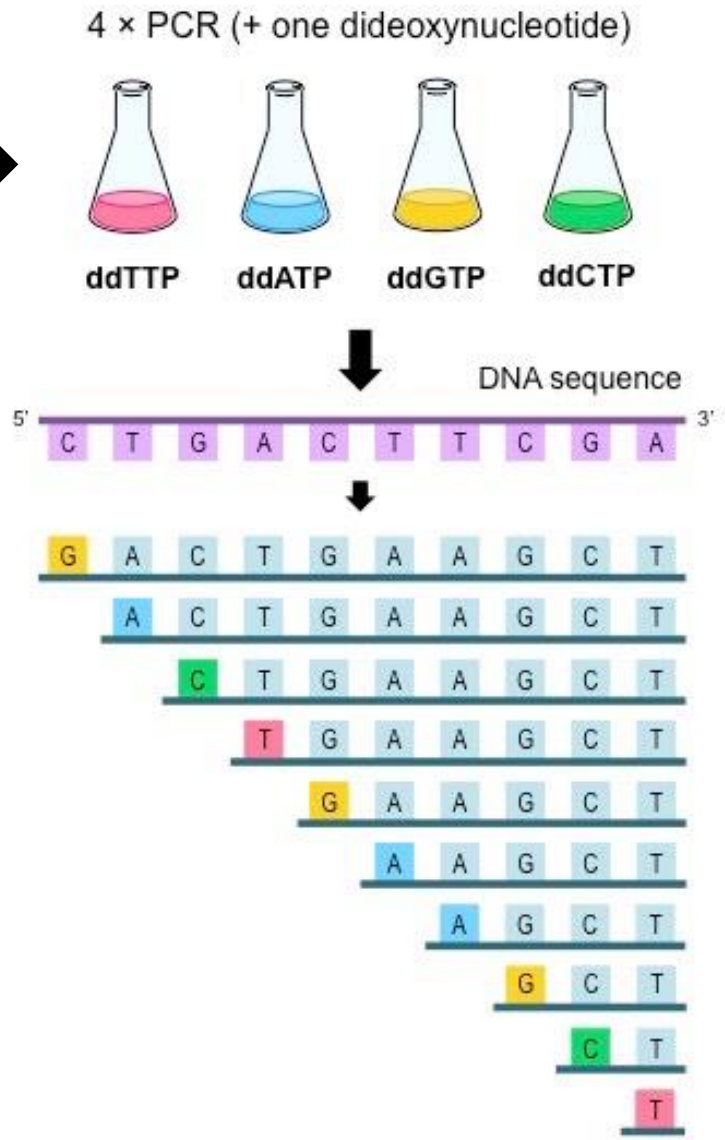
Denaturation – 95°C

Annealing – 68°C

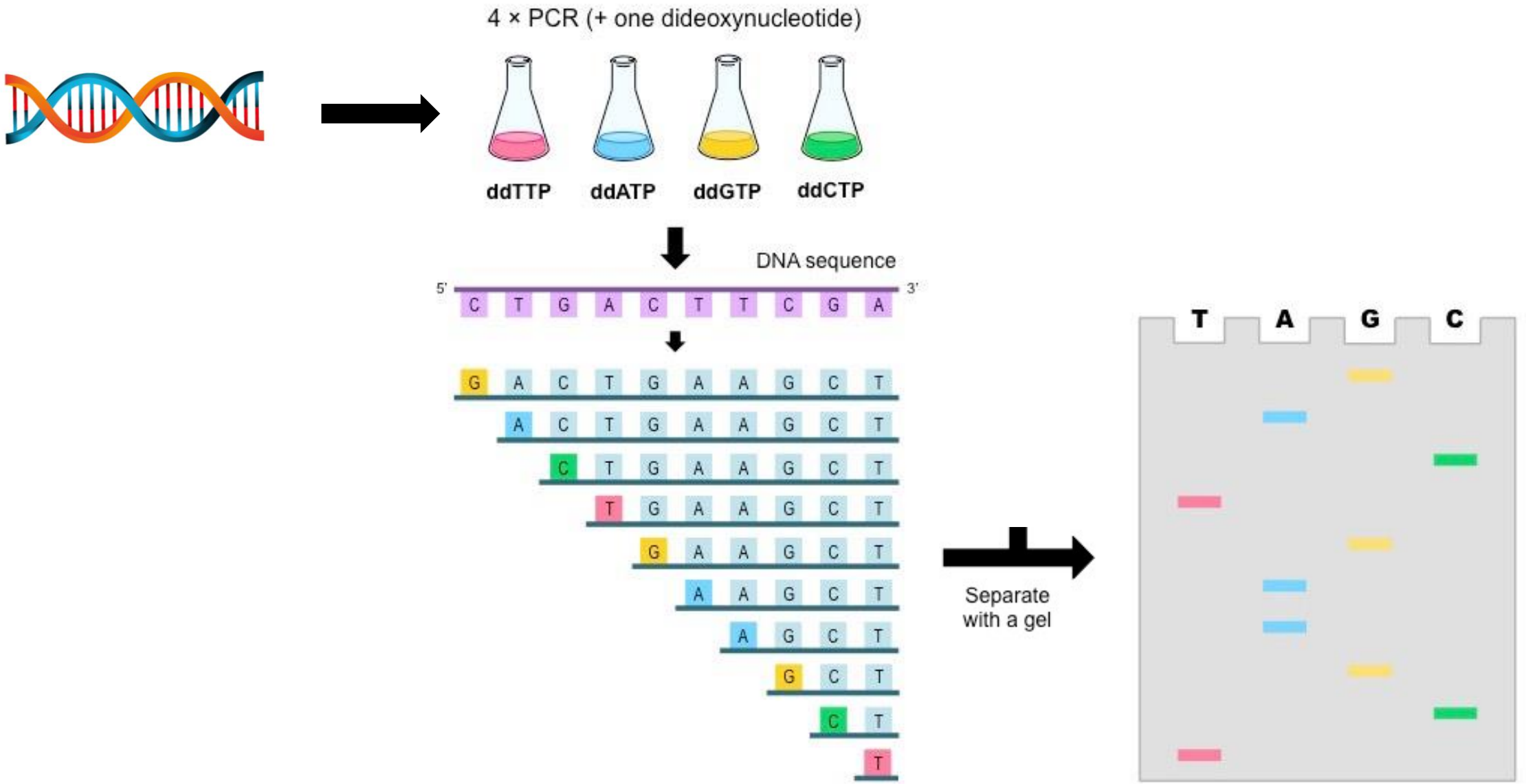
Elongation – 72°C

Sanger sequencing

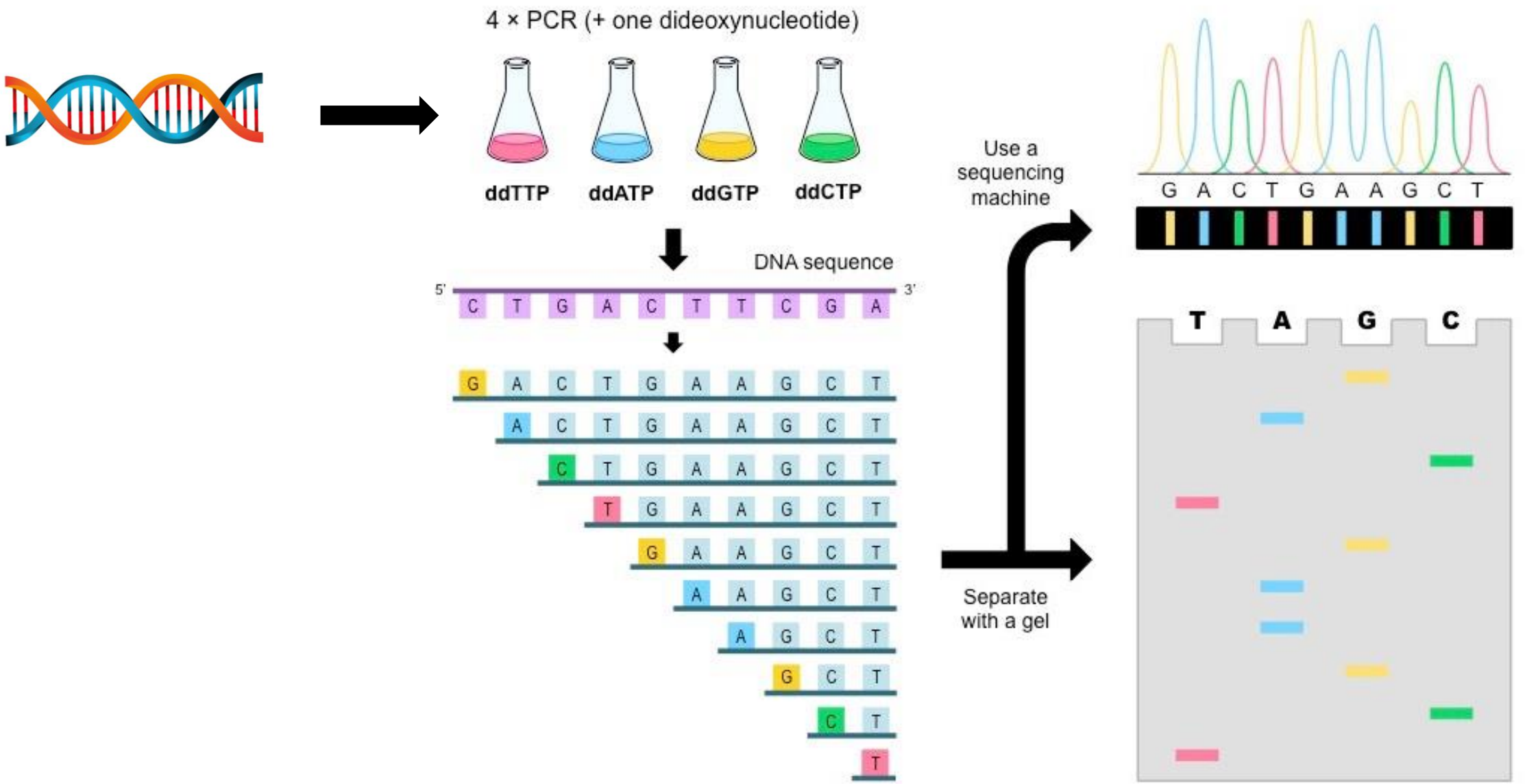
Sanger sequencing



Sanger sequencing



Sanger sequencing



Limitations

At the origin, only electrophoresis gene reading:

- Fragmented DNA were first needed to be clone in one bacteria
- 1kb of DNA per run of 6-8 hours.
- Radioactive label on the primer to read the gel.
- You have to manually read the gel.

➢ It took two days to sequence one kilobase of DNA.

Now Sanger sequencing is still used:

- No need to clone it in bacteria first, we can directly do PCR on it.
- 300kb of DNA per run of 3 hours (Fragments of 1-3kb maximum).
- Radioactive label have been replaced by four different fluorescent dice (one PCR and one migration instead of four in glass capillary)
- Machine read automatically the sequence.

➢ Useful to sequence few kilobases sequences used to genetically modified genomes or in synthetic biology.

Next Generation sequencing (NGS)

The Illumina sequencing machine

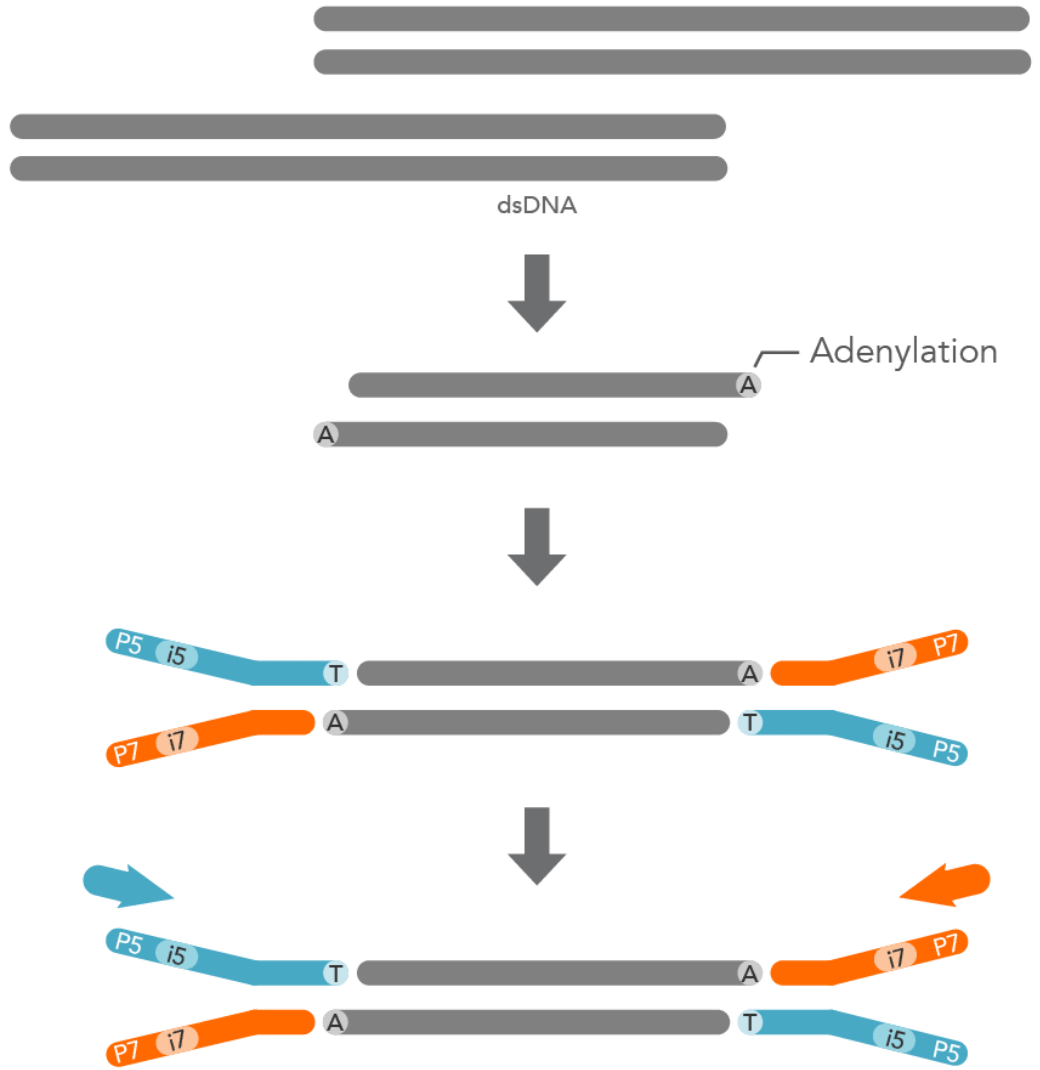
NGS – DNA library preparation

Fragmentation

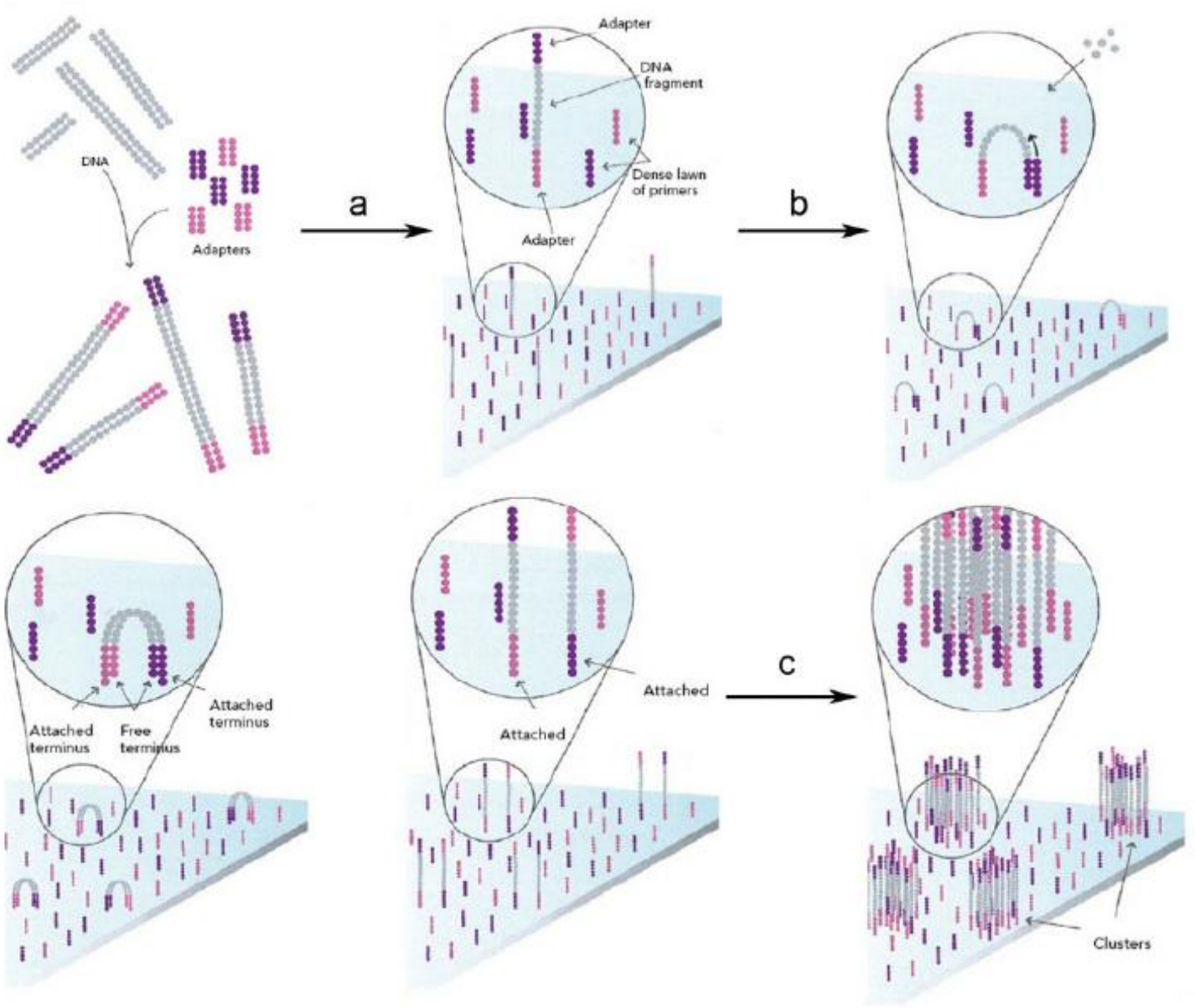
End repair and A-tailing

Ligation

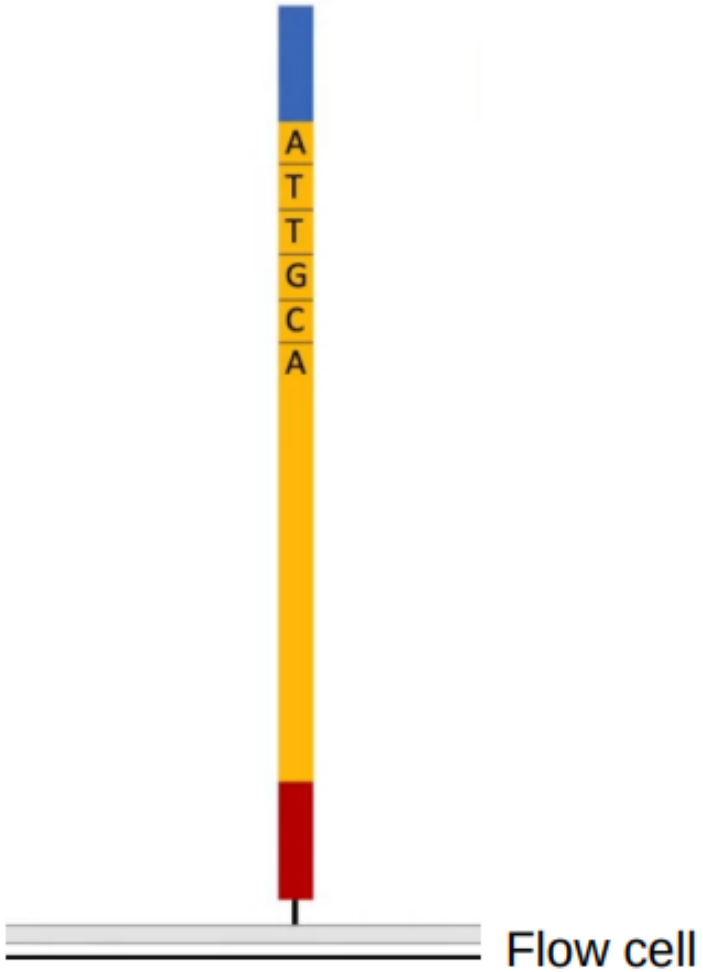
PCR amplification



NGS – Cluster amplification

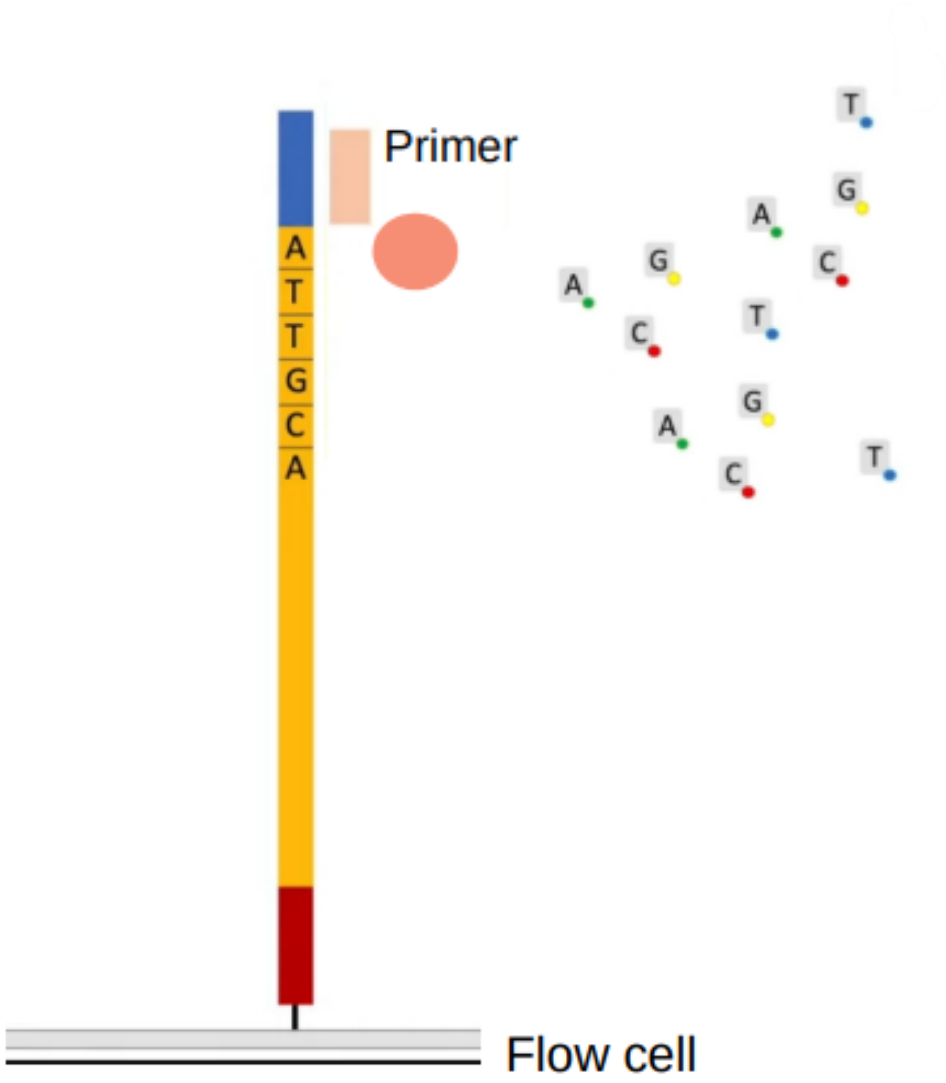


NGS – DNA sequencing



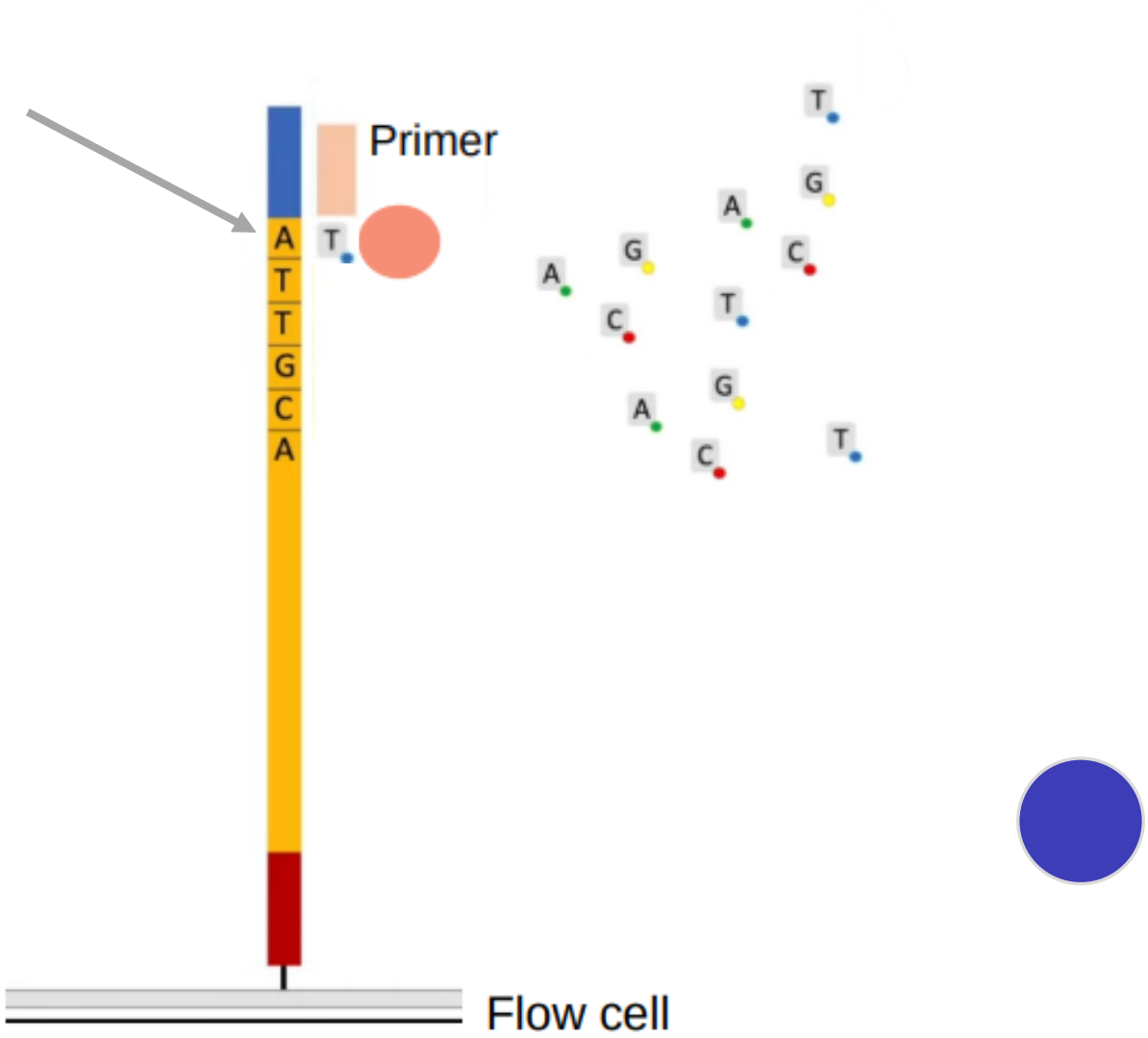
- Simultaneously for millions of cluster, yielding millions of reads.

NGS – DNA sequencing



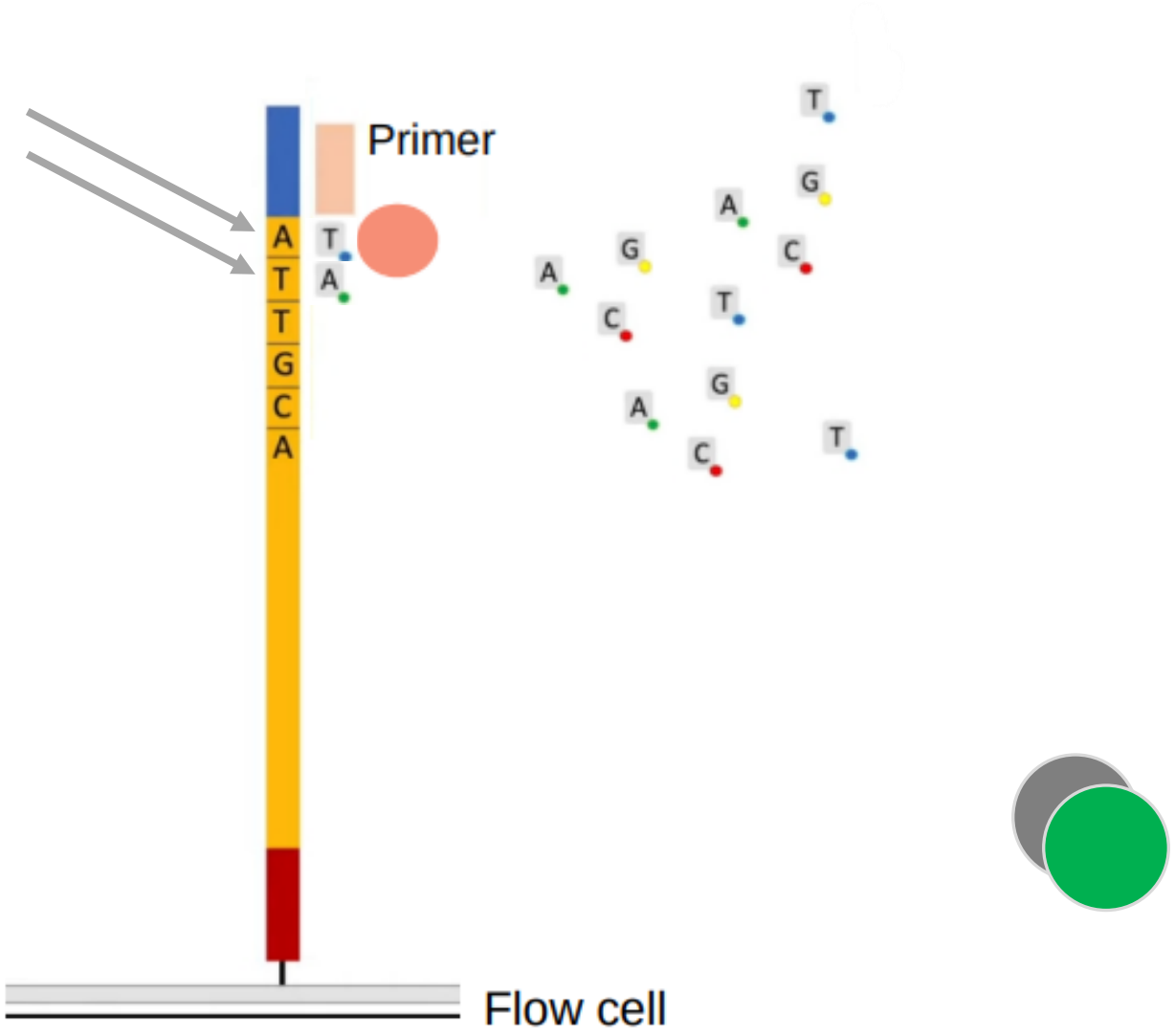
- Simultaneously for millions of cluster, yielding millions of reads.

NGS – DNA sequencing



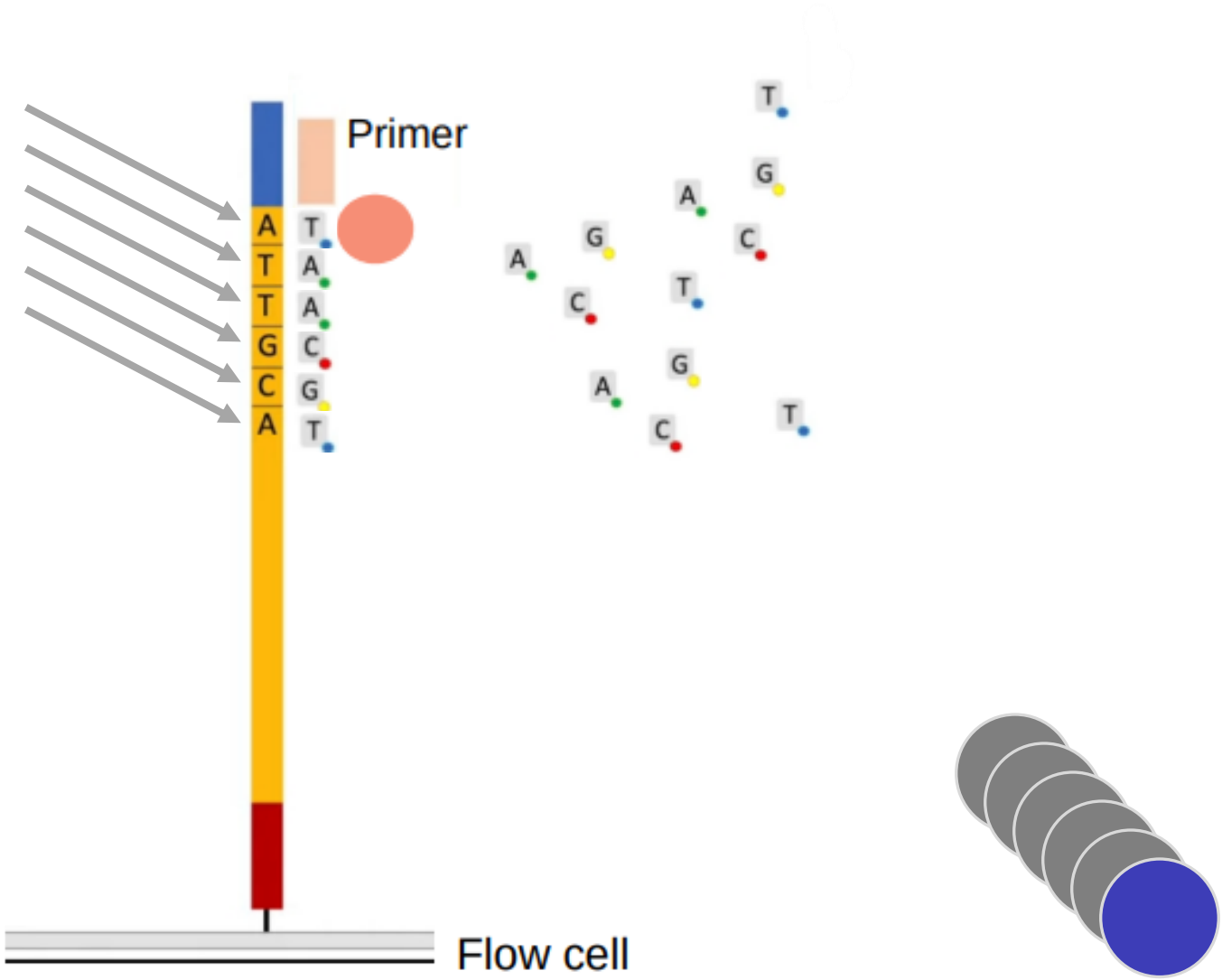
- Simultaneously for millions of cluster, yielding millions of reads.

NGS – DNA sequencing



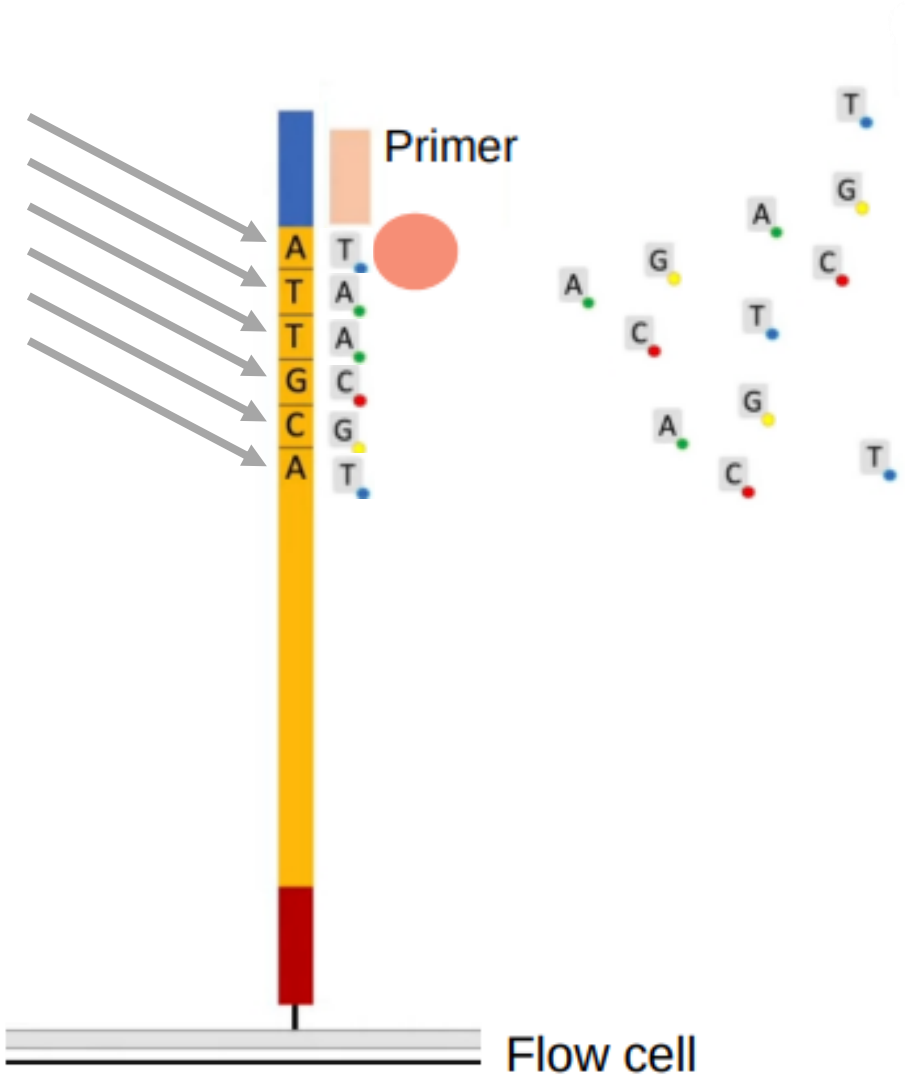
- Simultaneously for millions of cluster, yielding millions of reads.

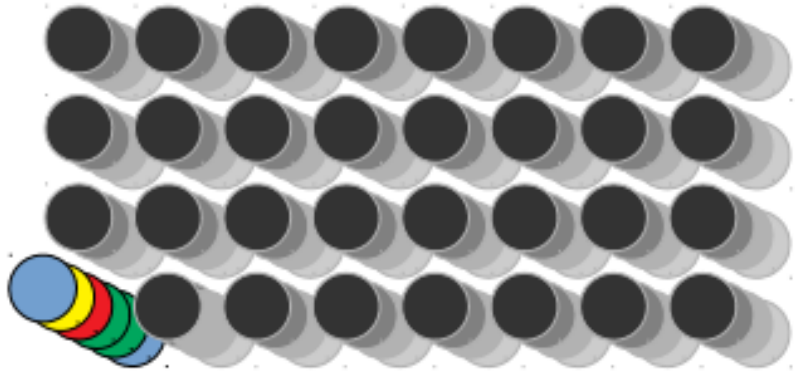
NGS – DNA sequencing

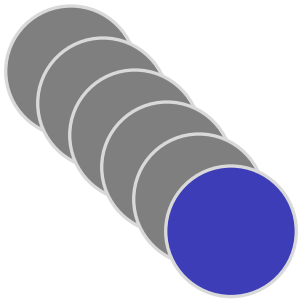


- Simultaneously for millions of cluster, yielding millions of reads.

NGS – DNA sequencing

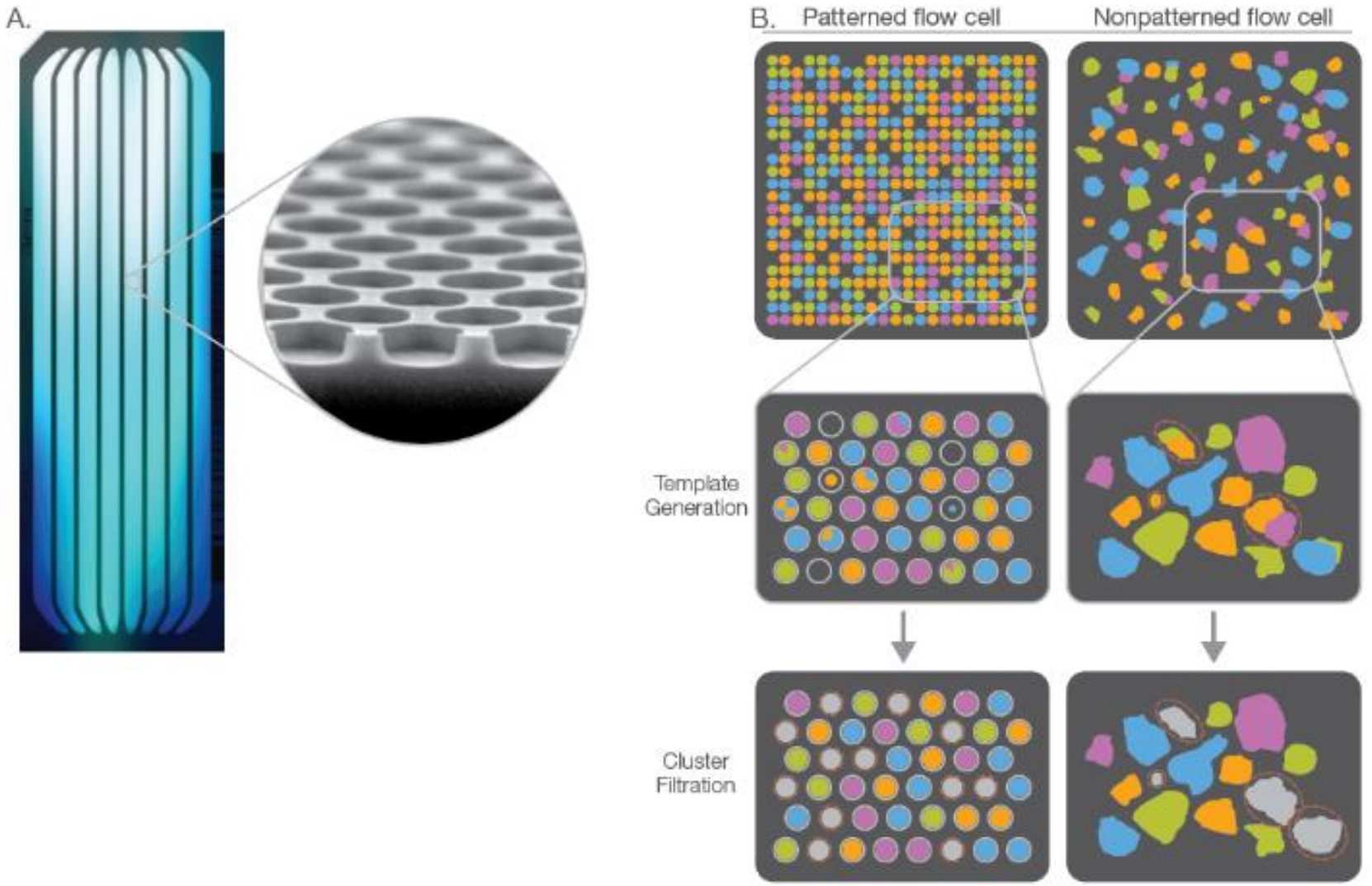




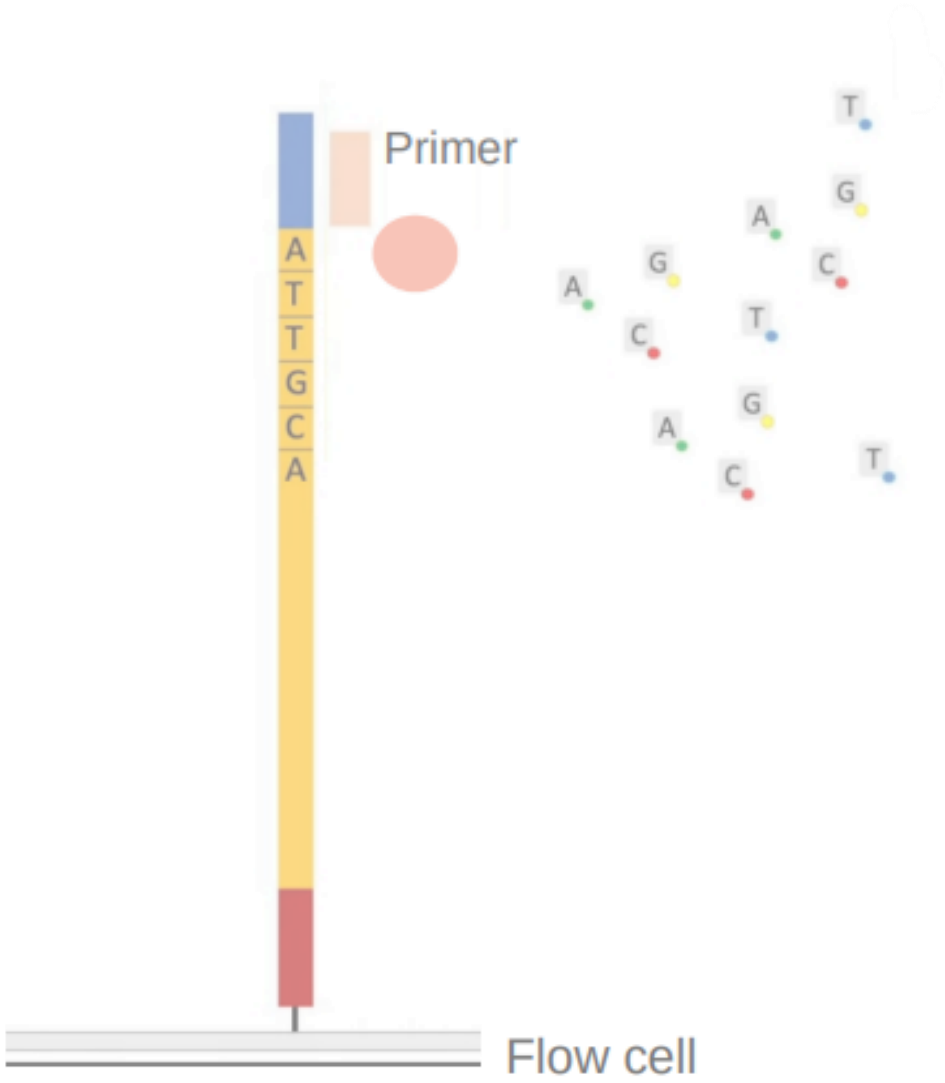


- Simultaneously for millions of cluster, yielding millions of reads.

Flow cell and cluster detection

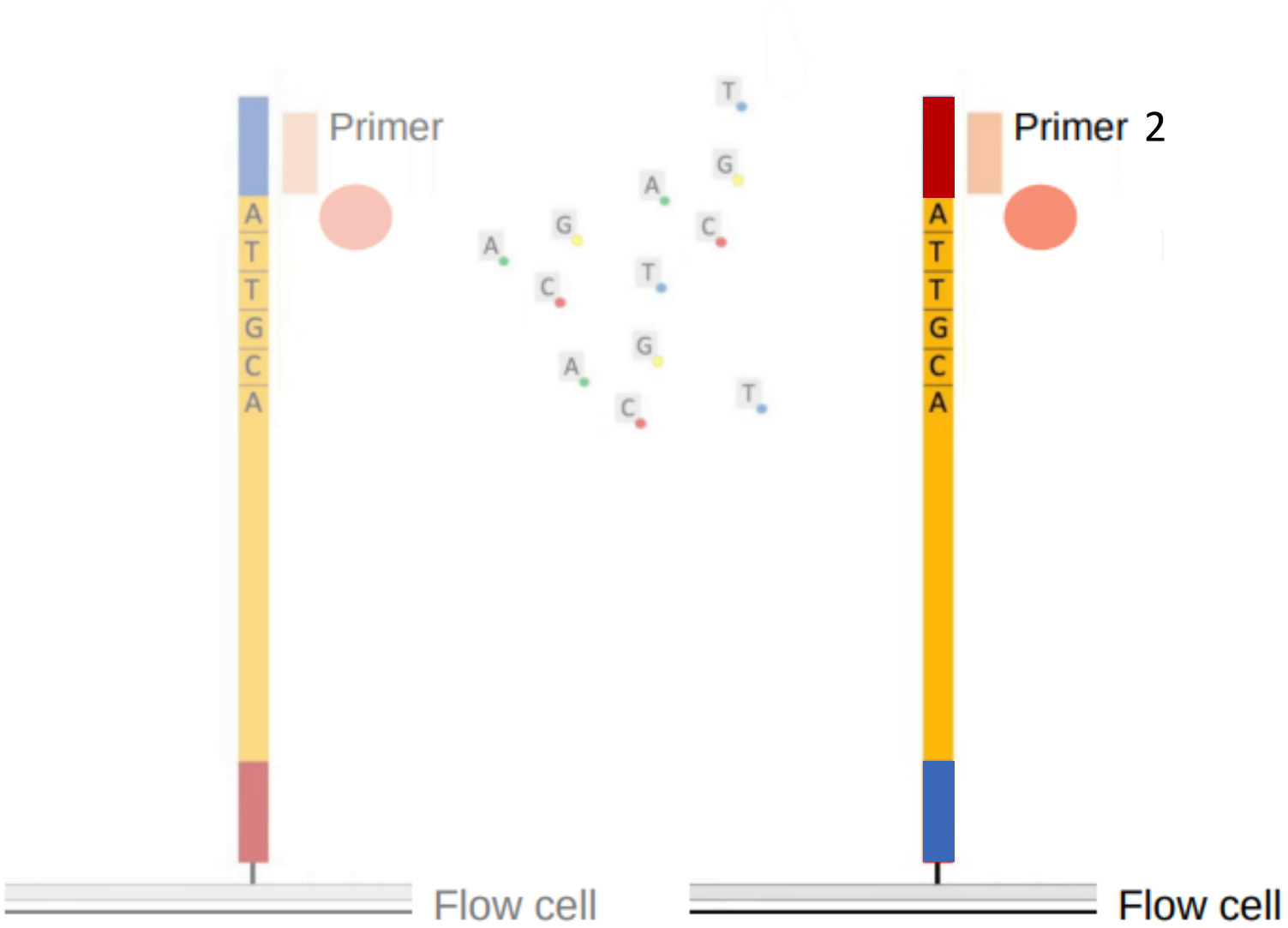


NGS – Paired end sequencing



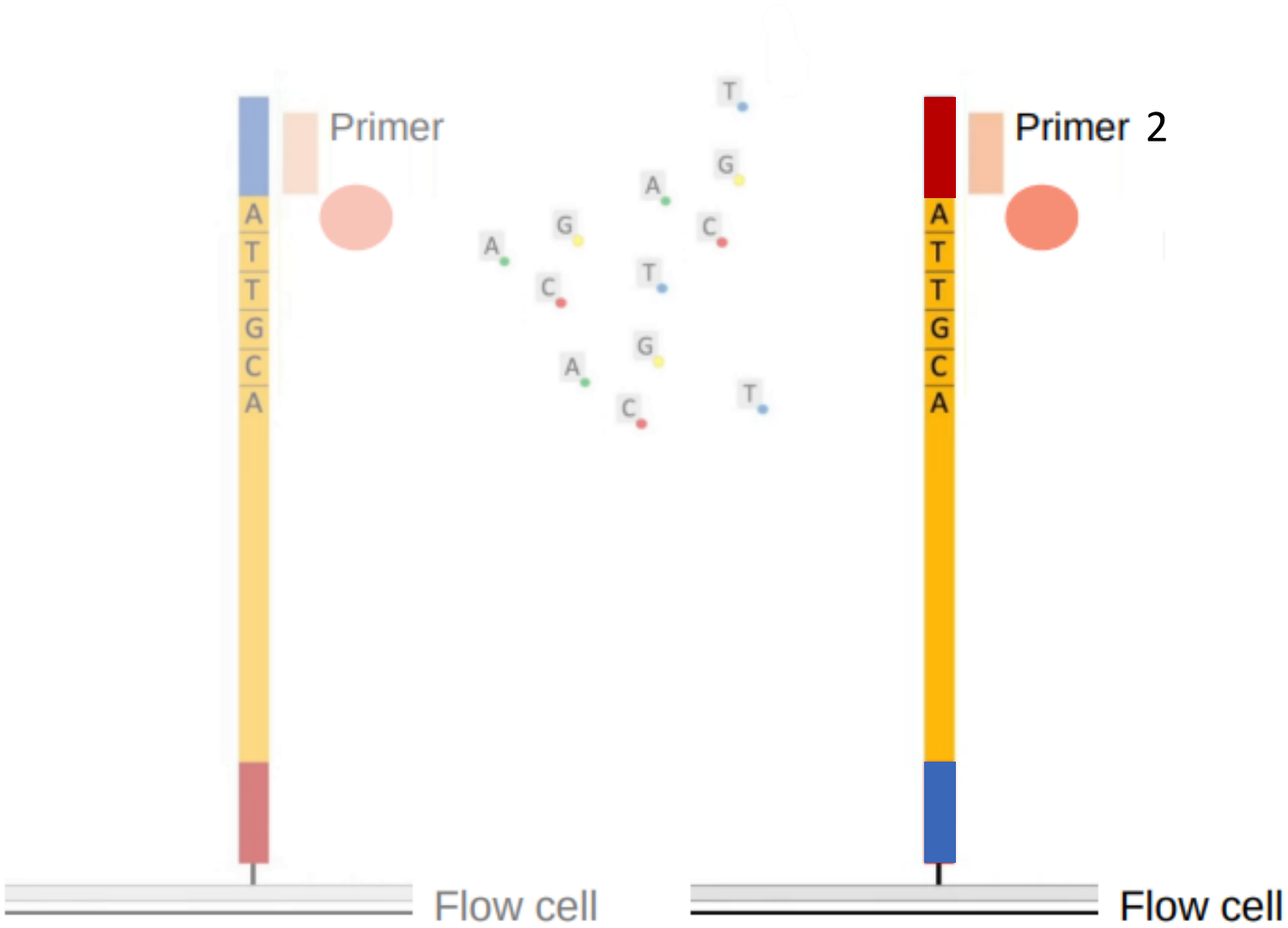
- Pair-end sequencing allows to sequence both extremities of your fragment.

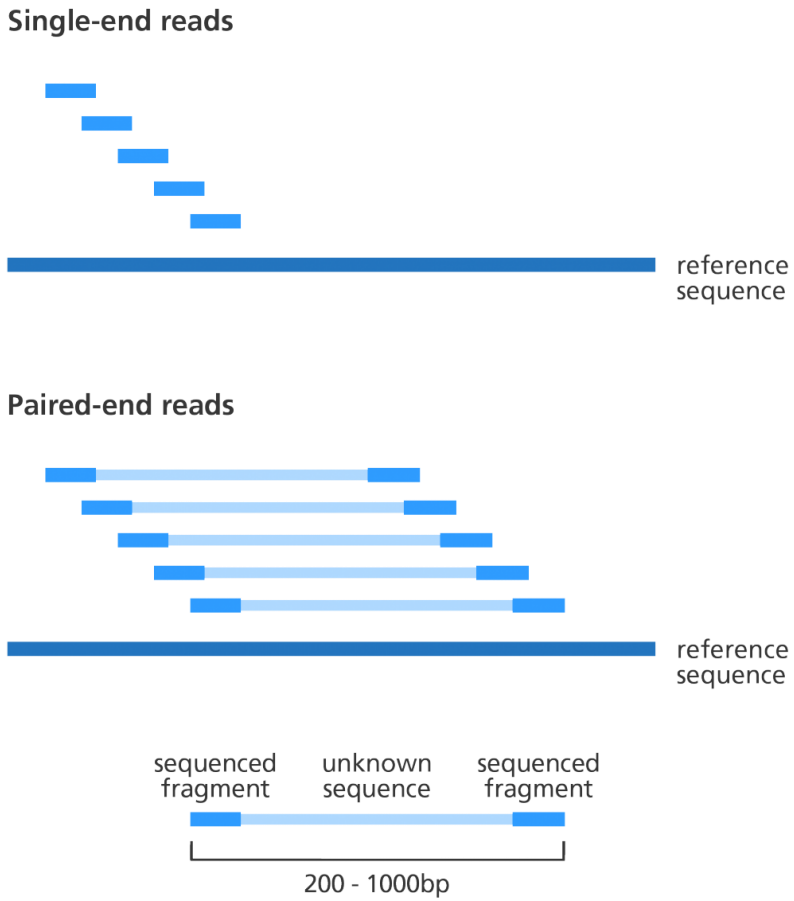
NGS – Paired end sequencing



- Pair-end sequencing allows to sequence both extremities of your fragment.

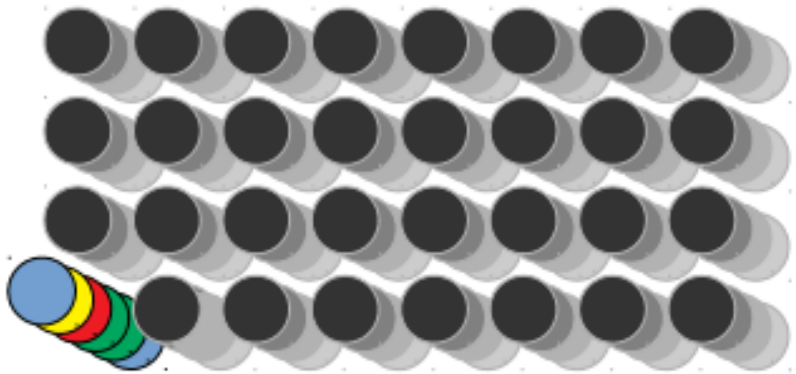
NGS – Paired end sequencing





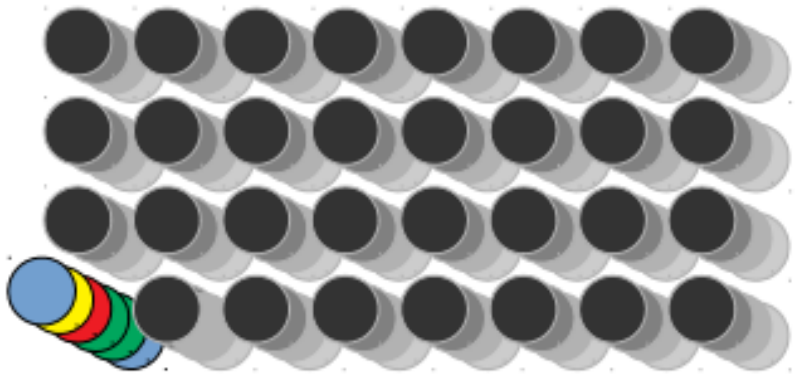
- Pair-end sequencing allows to sequence both extremities of your fragment.

NGS – the fastq file format



```
@NS500150:681:HWJNLBGXK:1:11101:3526:1073 1:N:0:CGGCTATG+NTTCGCCT
AAGGCAGCGACGAGTCTGACGATGCGAAAACTGAA
+
AAAAAEEEE/E/EEEEEE6EE/EEA<EEE/EEEEE
@NS500150:681:HWJNLBGXK:1:11101:20163:1075 1:N:0:CGGCTATG+NTTCGCCT
GATCCCATACGTACCTCAAGTGGTTTAGCAGTGTA
+
AAAAAEEEEEEEEEEEEEEEE/EEEEEEEEEEEE
```

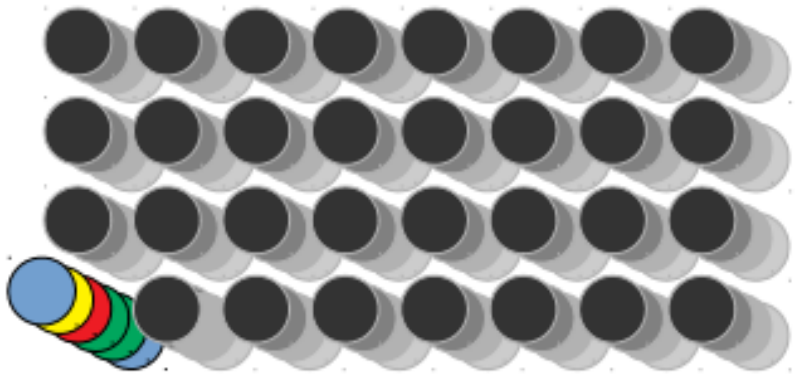
NGS – the fastq file format



Sequence ID

```
@NS500150:681:HWJNLBGXK:1:11101:3526:1073 1:N:0:CGGCTATG+NTTCGCCT
AAGGCAGCGACGAGTCTGACGATGCGAAAACTGAA
+
AAAAAEEEE/E/EEEEEE6EE/EEA<EEE/EEEEE
@NS500150:681:HWJNLBGXK:1:11101:20163:1075 1:N:0:CGGCTATG+NTTCGCCT
GATCCCATACGTACCTCAAGTGGTTTAGCAGTGTA
+
AAAAAEEEEEEEEEEEEEEE/EEEEEEEEEEEE
```

NGS – the fastq file format

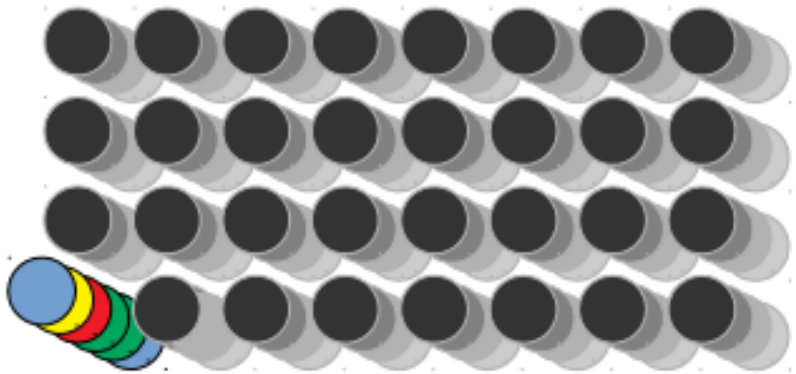


Sequence ID

DNA sequence

```
@NS500150:681:HWJNLBGXK:1:11101:3526:1073 1:N:0:CGGCTATG+NTTCGCCT
AAGGCAGCGACGAGTCTGACGATGCGAAAACTGAA
+
AAAAAEEEE/E/EEEEEE6EE/EEA<EEE/EEEEE
@NS500150:681:HWJNLBGXK:1:11101:20163:1075 1:N:0:CGGCTATG+NTTCGCCT
GATCCCATACGTACCTCAAGTGGTTTAGCAGTGTA
+
AAAAAEEEEEEEEEEEEEEE/EEEEEEEEEEEE
```

NGS – the fastq file format



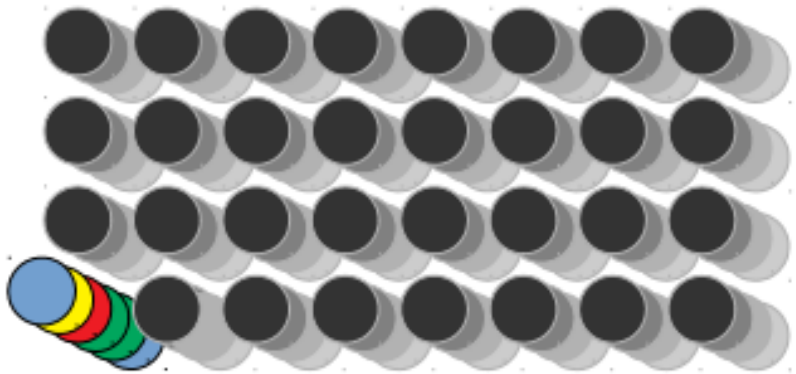
Sequence ID

DNA sequence

A useless plus

```
@NS500150:681:HWJNLBGXK:1:11101:3526:1073 1:N:0:CGGCTATG+NTTCGCCT
AAGGCAGCGACGAGTCTGACGATGCGAAAACTGAA
+
AAAAAEEEE/E/EEEEEE6EE/EEA<EEE/EEEEE
@NS500150:681:HWJNLBGXK:1:11101:20163:1075 1:N:0:CGGCTATG+NTTCGCCT
GATCCCATACGTACCTCAAGTGGTTTAGCAGTGTA
+
AAAAAEEEEEEEEEEEEEEE/EEEEEEEEEEEE
```

NGS – the fastq file format



Sequence ID

```
@NS500150:681:HWJNLBGXK:1:11101:3526:1073 1:N:0:CGGCTATG+NTTCGCCT
AAGGCAGCGACGAGTCTGACGATGCGAAAACTGAA
+
AAAAAEEEE/E/EEEEEE6EE/EEA<EEE/EEEEE
@NS500150:681:HWJNLBGXK:1:11101:20163:1075 1:N:0:CGGCTATG+NTTCGCCT
GATCCCATACGTACCTCAAGTGGTTTAGCAGTGTA
+
AAAAAEEEEEEEEEEEEEEEE/EEEEEEEEEEEE
```

DNA sequence

A useless plus

A quality score

Sequencing output and price



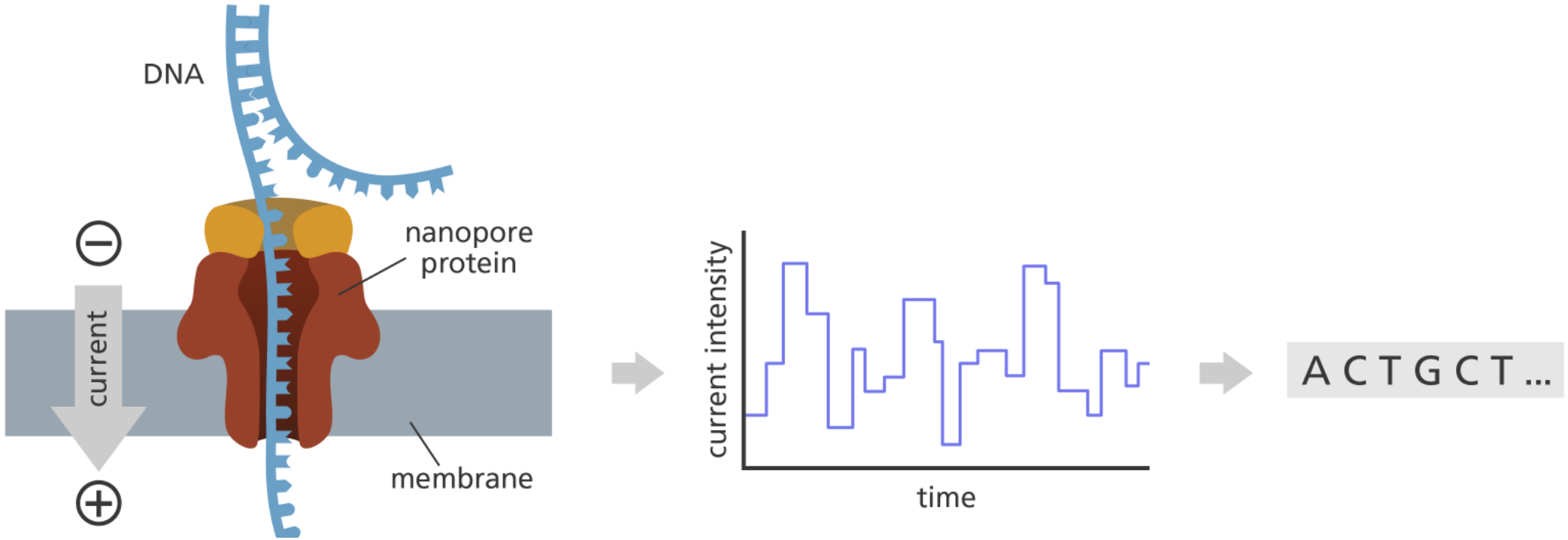


	iSeq 100	MiniSeq	MiSeq Series	NextSeq 550 Series	NextSeq 1000 & 2000
Run Time	9.5–19 hrs	4–24 hours	4–55 hours	12–30 hours	11-48 hours
Maximum Output	1.2 Gb	7.5 Gb	15 Gb	120 Gb	360 Gb *
Maximum Reads Per Run	4 million	25 million	25 million †	400 million	1.2 billion *
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 300 bp
	800$/Gb	80$/Gb	80$/Gb	25$/Gb	15$/Gb

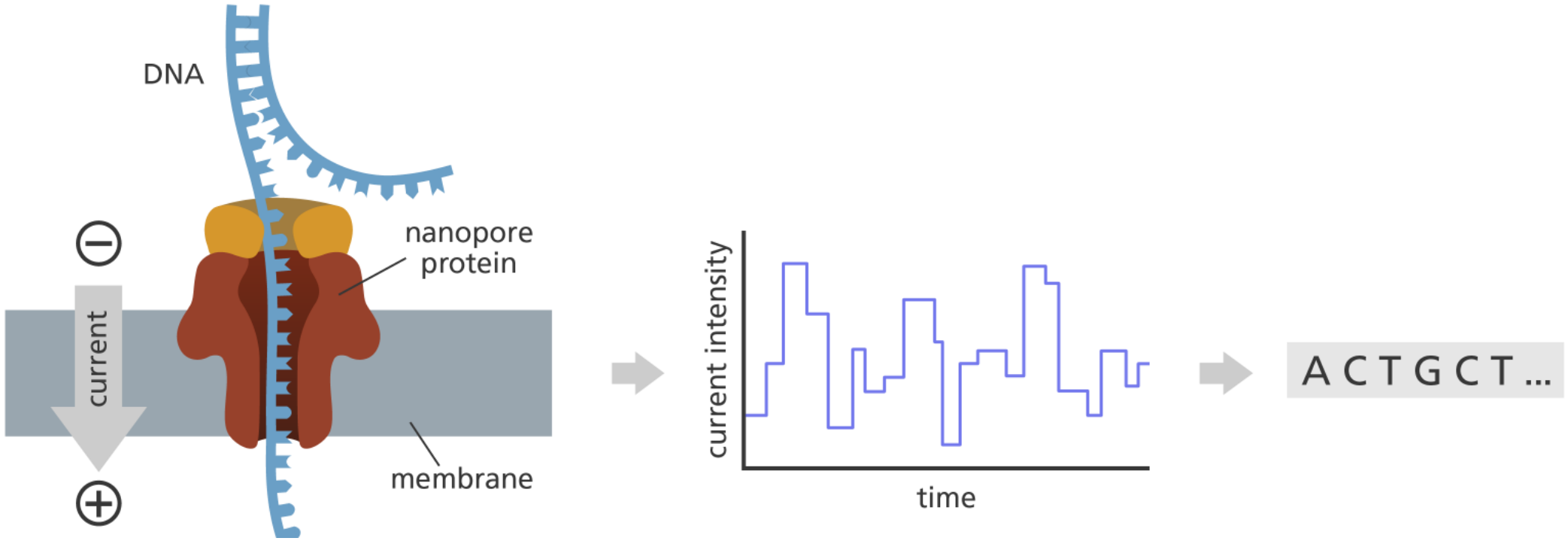
- Others sequencing companies now (MGI).
- Others sequencing technologies with long reads output with PacBio and Nanopore.

Opening on long reads sequencing

Oxford Nanopore Sequencing

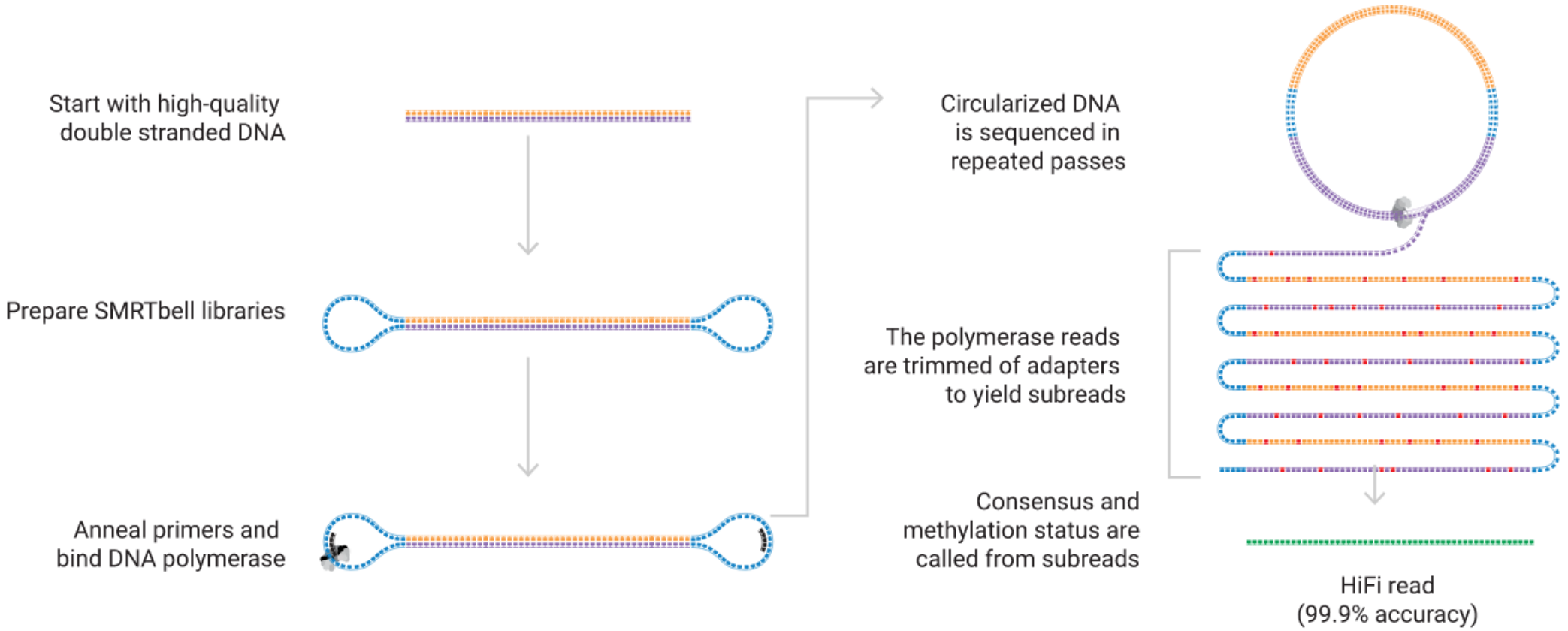


Oxford Nanopore Sequencing



- Long reads (around 25kb and up to 2Mb)
- Low fidelity (1 error every 10bp – 1 error every 10,000bp for Illumina)
- Around 50$/Gb

Pacific Bioscience technology



- Long reads (around 10kb)
- Higher fidelity (1 error every 1000bp – 1 error every 10,000bp for Illumina)
- Quite expensive 200$/Gb