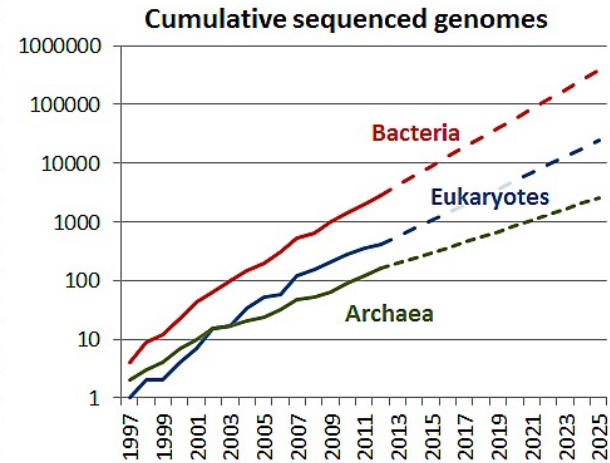
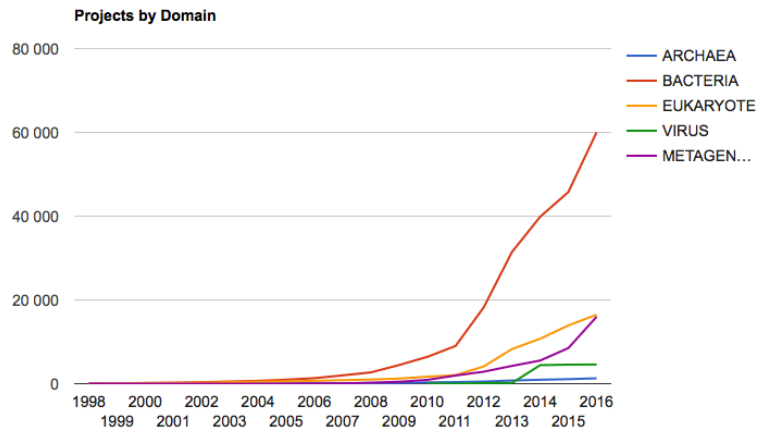
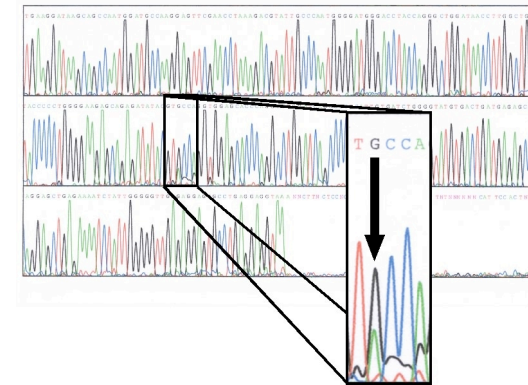


# **GENOME SEQUENCING**

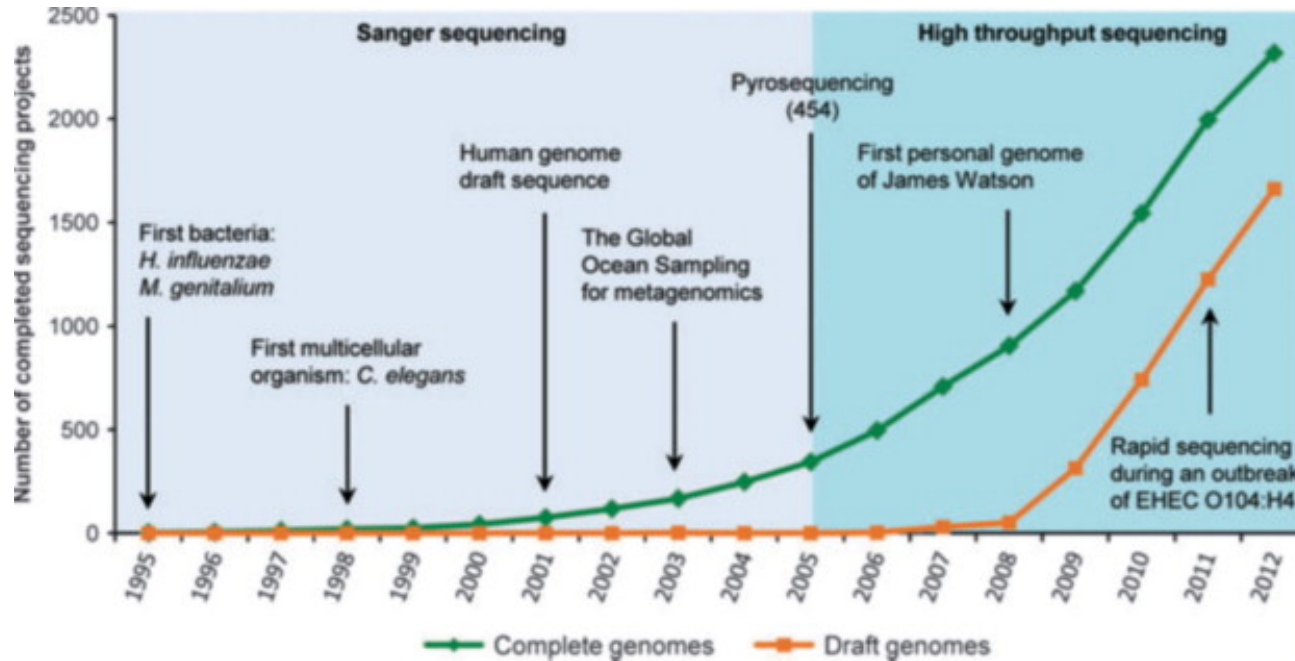
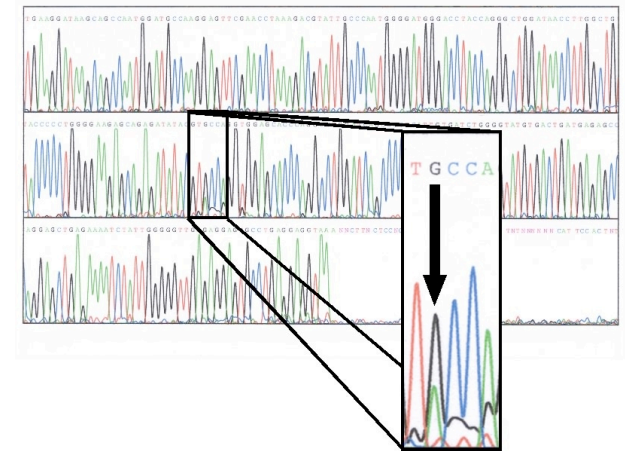
# Introduction

## Sequencing the genomes



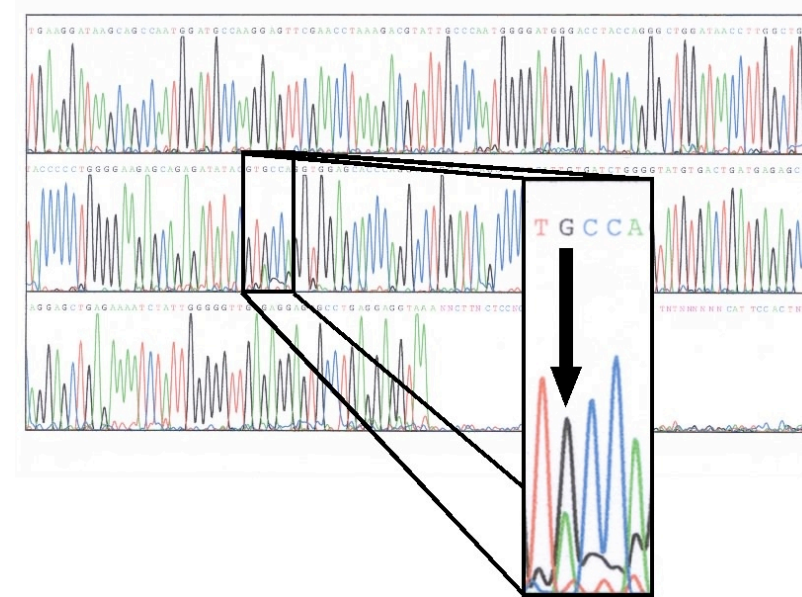
# Introduction

## Sequencing the genomes



# Introduction

## Sequencing the genomes



### Why sequencing these genomes?

Complete inventory of the genes of various organisms : # of genes, % of coding genome

Information about the sequence of every polypeptides

Basis for new technologies and medical applications

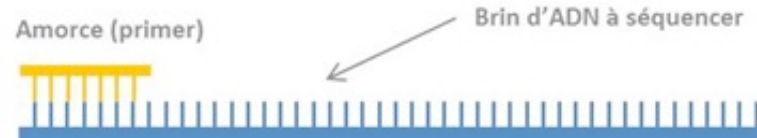
Global data at the level of the whole cell/tissue (tumor heterogeneity)

New data for evolutionary and species classification analysis

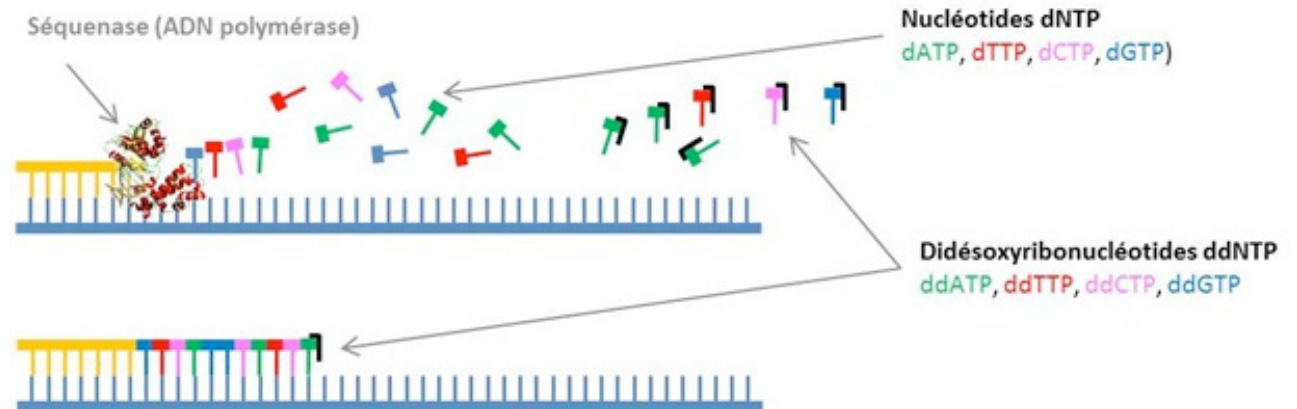
# I. Genome sequencing

## The Sanger Method (1977)

Denaturation  
+priming



Polymerization



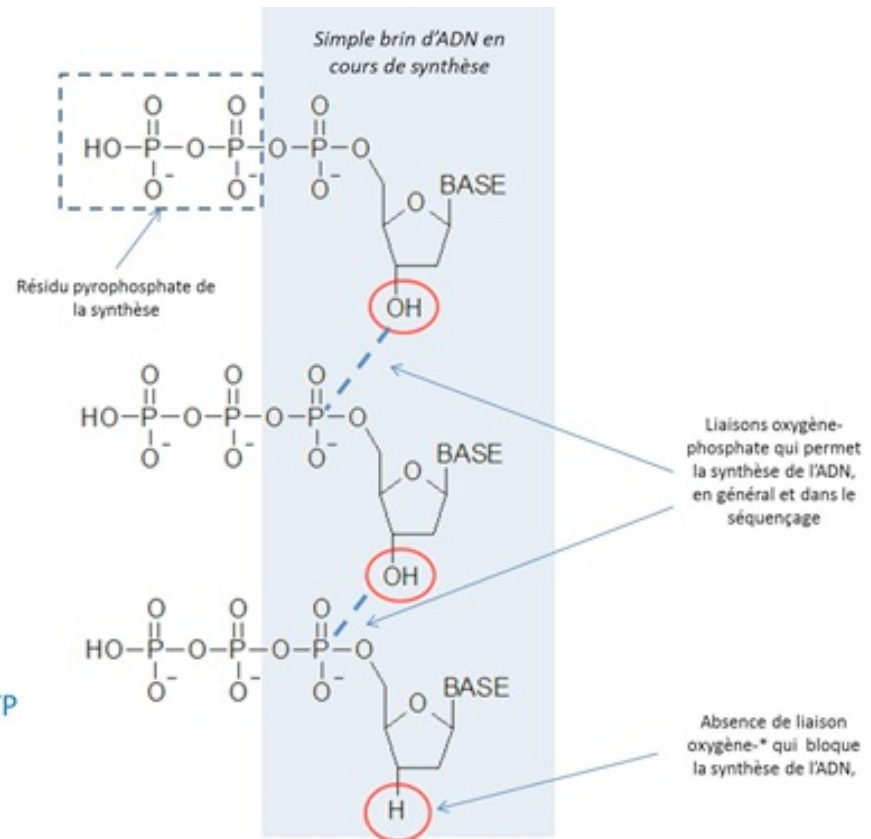
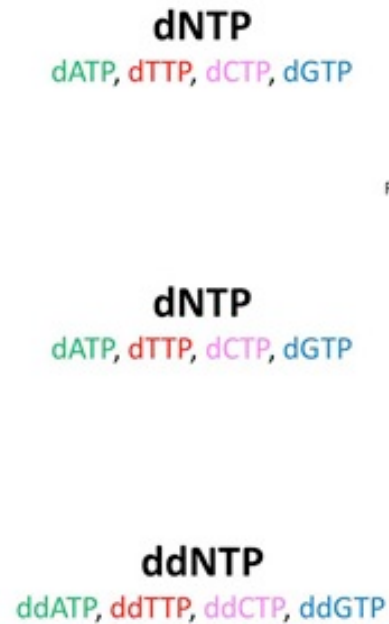
Arrest



# I. Genome sequencing

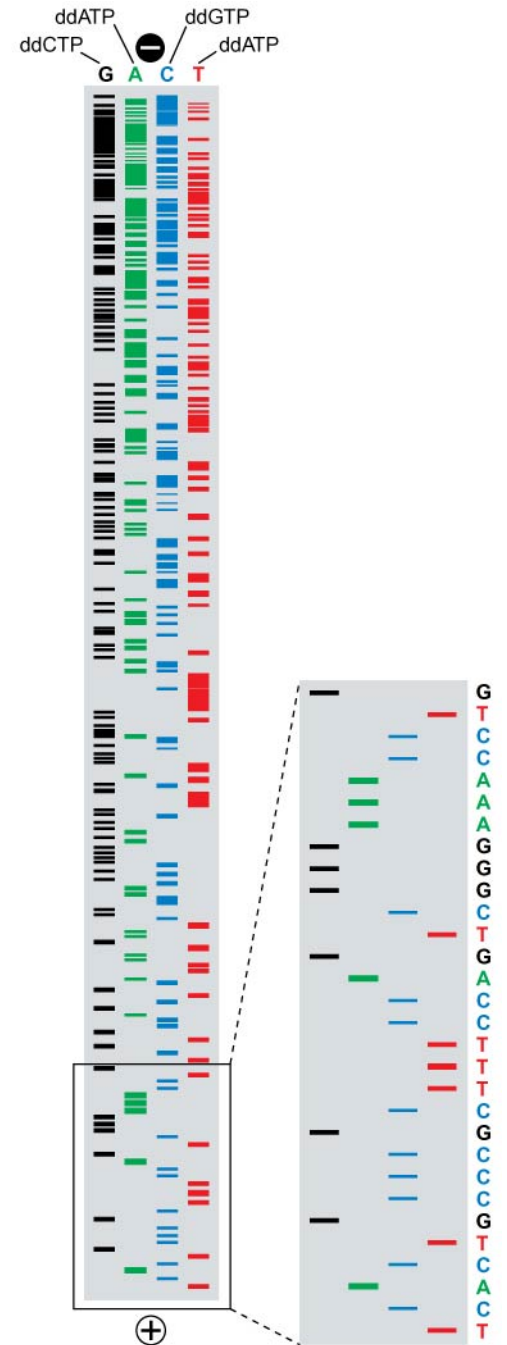
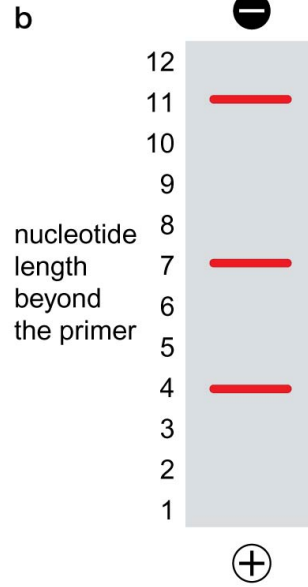
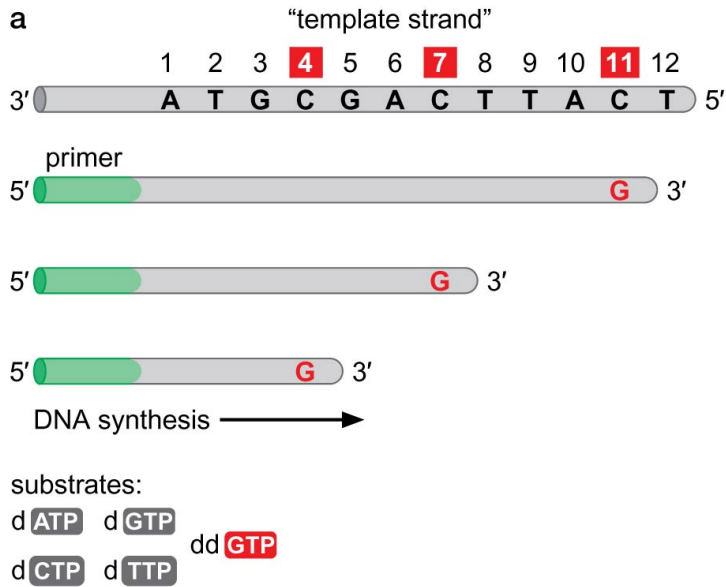
## The Sanger Method (1977)

End of  
polymerization



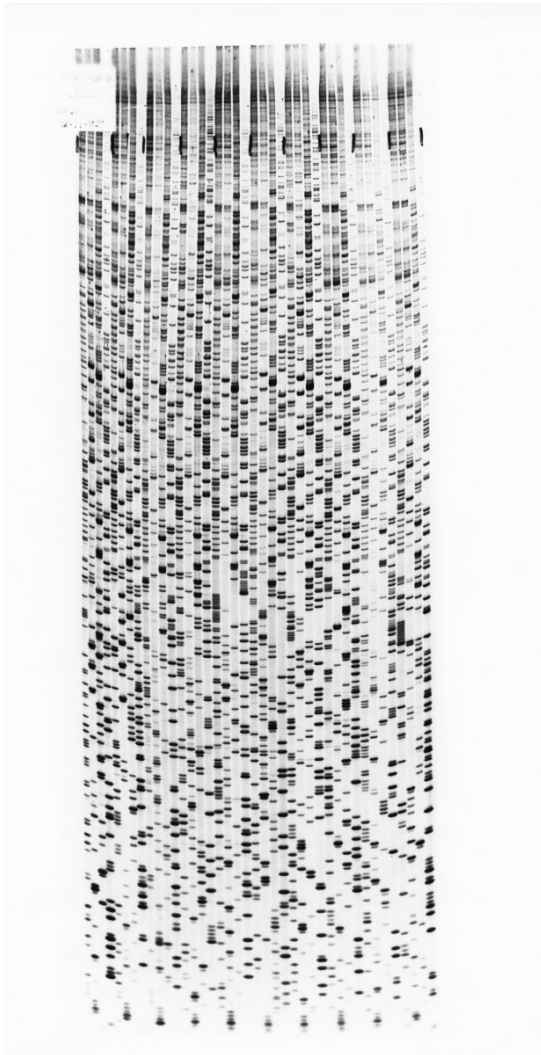
# I. Genome sequencing

## The Sanger Method (1977)



# I. Genome sequencing

## The caveats of the Sanger Method



Technically cumbersome and time consuming (limited samples processing)

High amounts of starting material

Some knowledge about the sequence is required (primer)

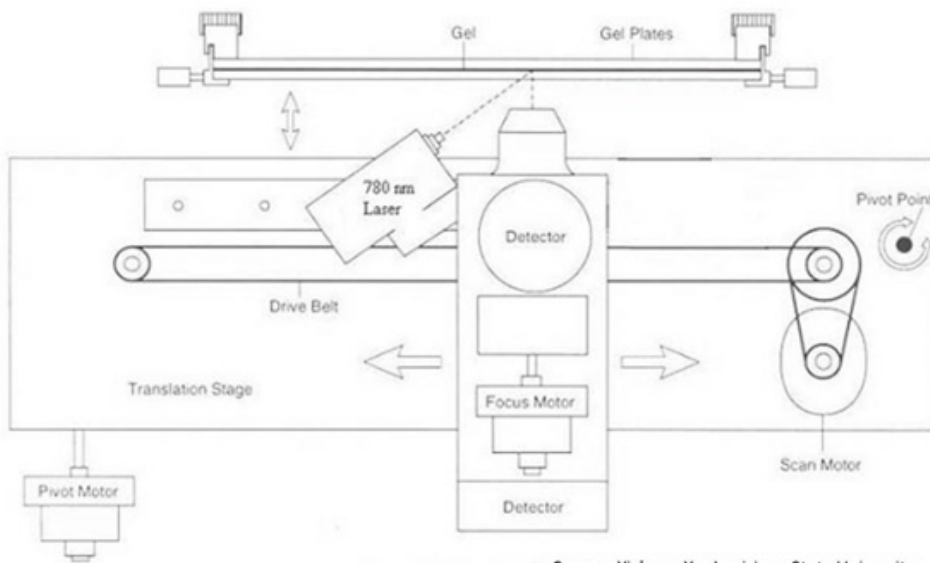
-> Not well-suited for high-throughput analyses (genome sequencing)



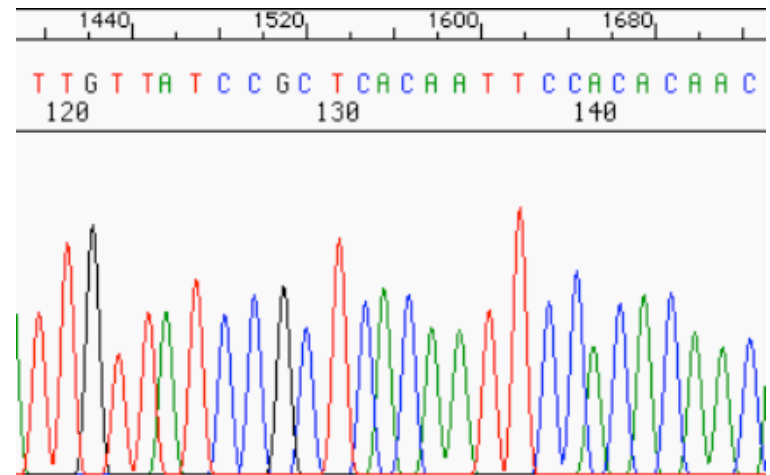
# I. Genome sequencing

## The Sanger Method (1977)

1st generation sequencing : Improvement of sequence analysis (fluorescent ddNTP + column electrophoresis)



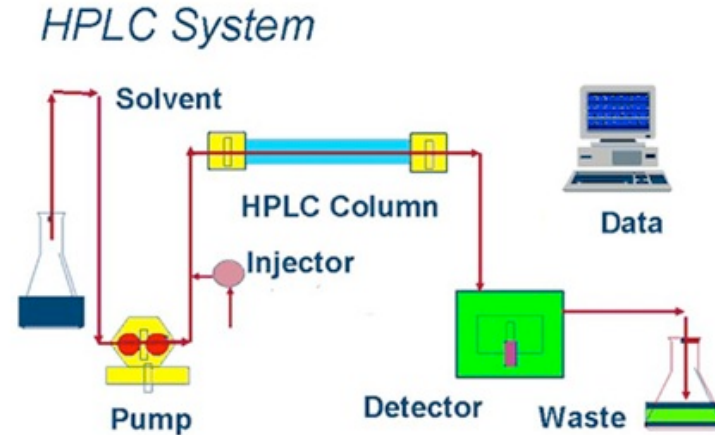
Source: Yichuan Xu, Louisiana State University



# I. Genome sequencing

## The Sanger Method (1977)

1st generation sequencing : Improvement of sequence analysis (capillary electrophoresis)



Source: College of Arts and Sciences, New Mexico State University



Agilent 1290 Infinity Quaternary LC System



PerkinElmer Flexar FX-15 UHPLC



Nexera UHPLC



Dionex UltiMate 3000 Standard LC Systems

# I. Genome sequencing

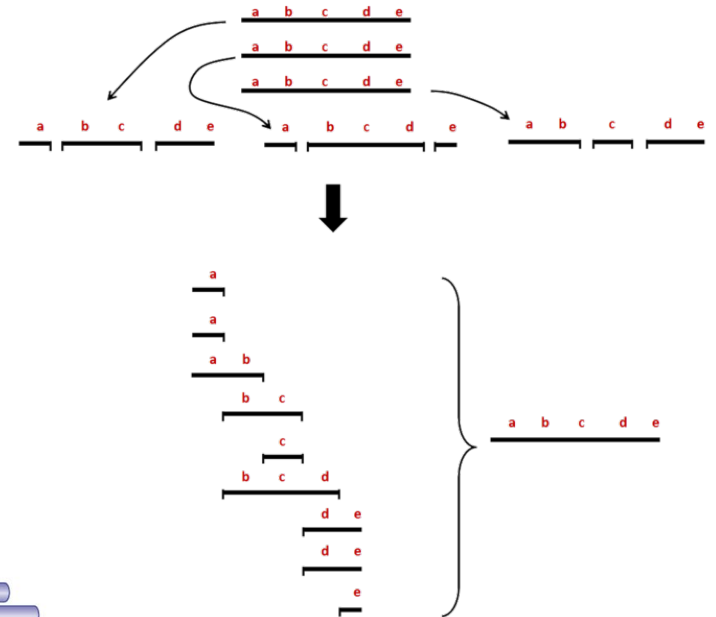
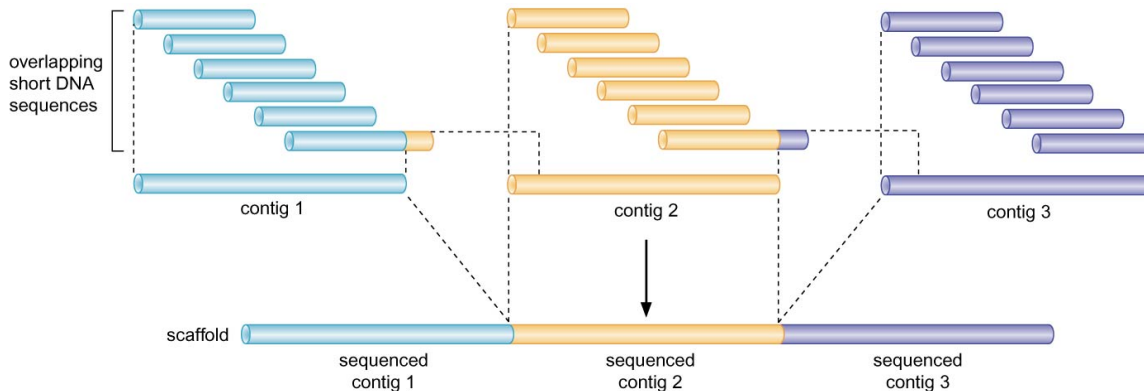
1st generation sequencing : Improvement of the methodology

Parallel random sequencing : after fragmentation (shotgun sequencing) and cloning

- genomic DNA is fragmented
- cloned to a plasmid vector -> transform

*E. coli*

- a single bacterial colony is picked -> plasmid DNA isolated

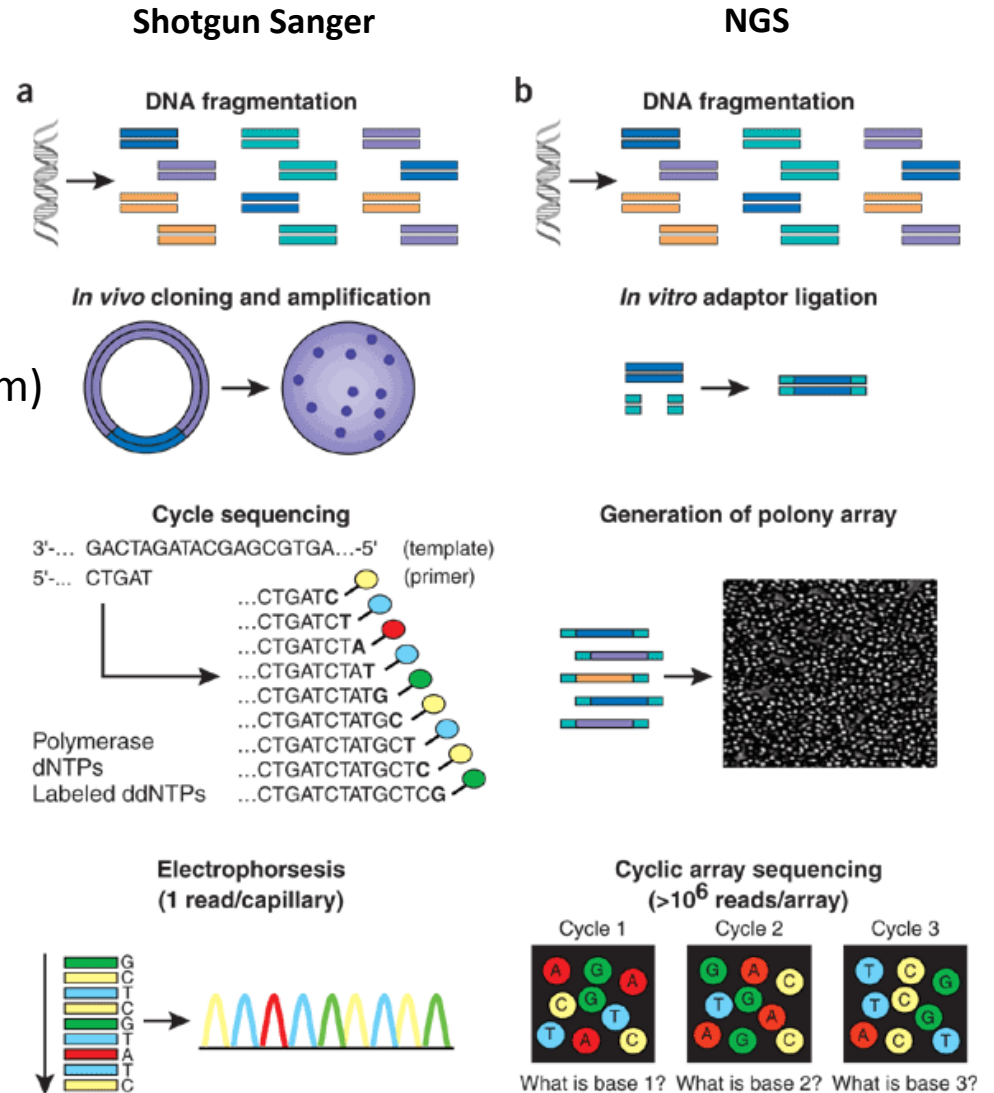


! Redundancy (8x or 20x)

# I. Genome sequencing

## 2nd generation and next generation (NGS) sequencing : Areas of improvement

- Template preparation  
amplification by [PCR](#) (Cell free system)
- Sequencing and imaging
- Data analysis

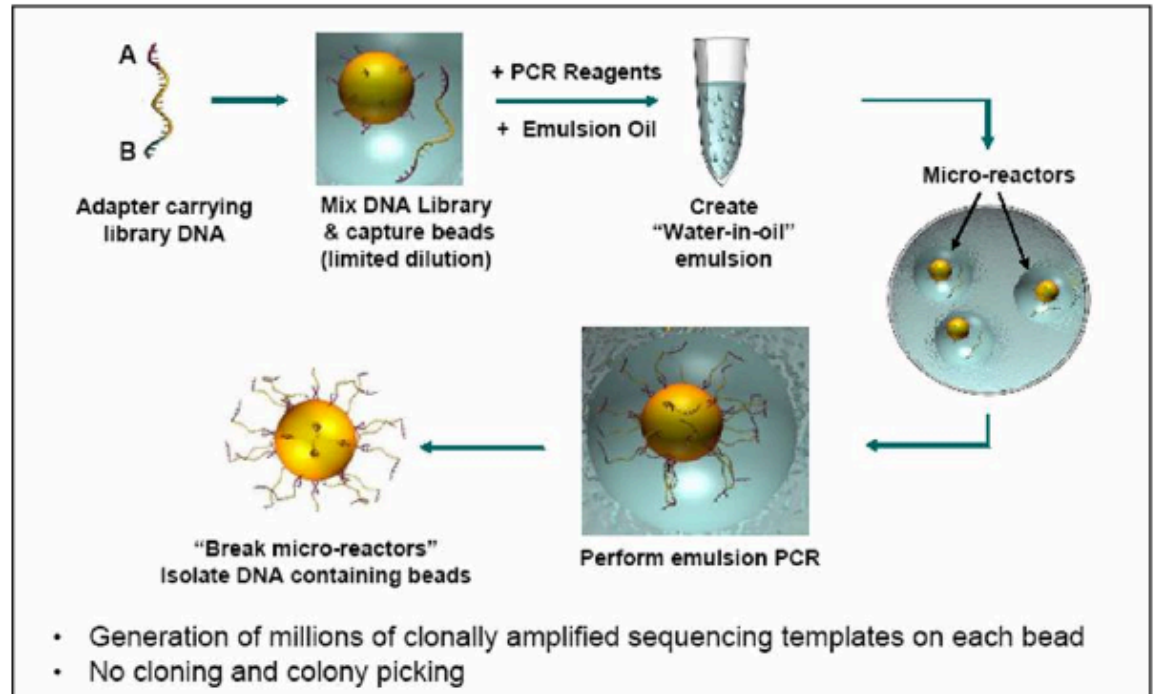


# I. Genome sequencing

2nd generation sequencing/NGS :

Improvement of template preparation : amplification by PCR

Emulsion PCR (emPCR)



Used by 454 (life science), Polonator, Ion torrent and Solid

# I. Genome sequencing

2nd generation sequencing/NGS :

Improvement of template preparation : amplification by PCR

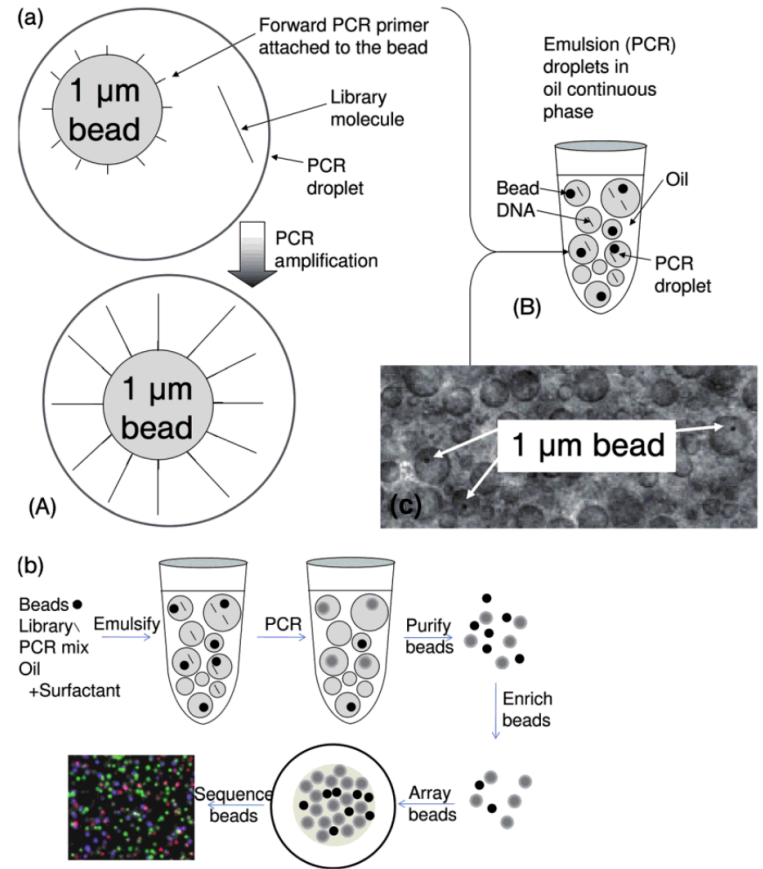
Emulsion PCR (emPCR)

Beads are immobilized on:

Microscope slides (Polonator)

PicoTiterPlates (PTP) (Roche/454)

Microchip sensor (Ion Torrent)

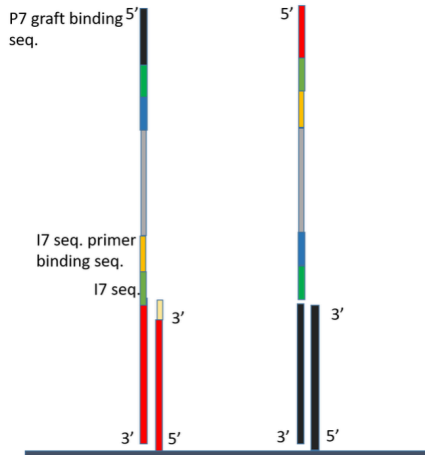
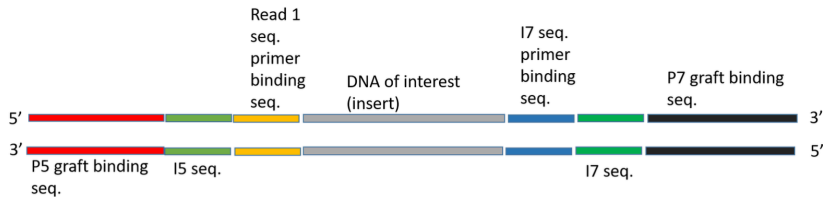


# I. Genome sequencing

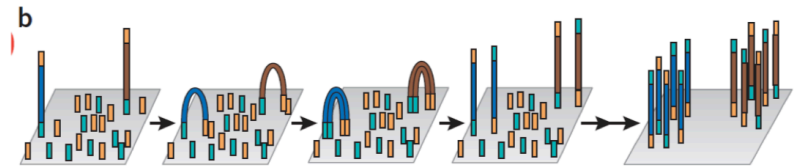
2nd generation sequencing :

Improvement of template preparation : amplification by [PCR](#)

Bridge PCR ([illumina platform](#))



*Fragments bound to flow cell*



- DNA fragments are flanked with adaptors.
- A flat surface coated with two types of primers, corresponding to the adaptors.
- Amplification proceeds in cycles, with one end of each bridge tethered to the surface.
- Used by Solexa.

# I. Genome sequencing

2nd generation sequencing :

Improvement of template preparation : amplification by [PCR](#)

- Still require large amount of template
- Sequencing of short length fragments only
- PCR amplification induces biases



# I. Genome sequencing

3rd generation sequencing :

Improvement of template preparation : Single Molecule Sequencing (SMRT)

- Requires less starting material
- Provides longer sequence fragments

■ Immobilized on the solid surface by

Primers: Helicos BioSciences

Template: Helicos BioSciences

Polymerase: Pacific Biosciences, Life/Visigen, LI-COR Biosciences

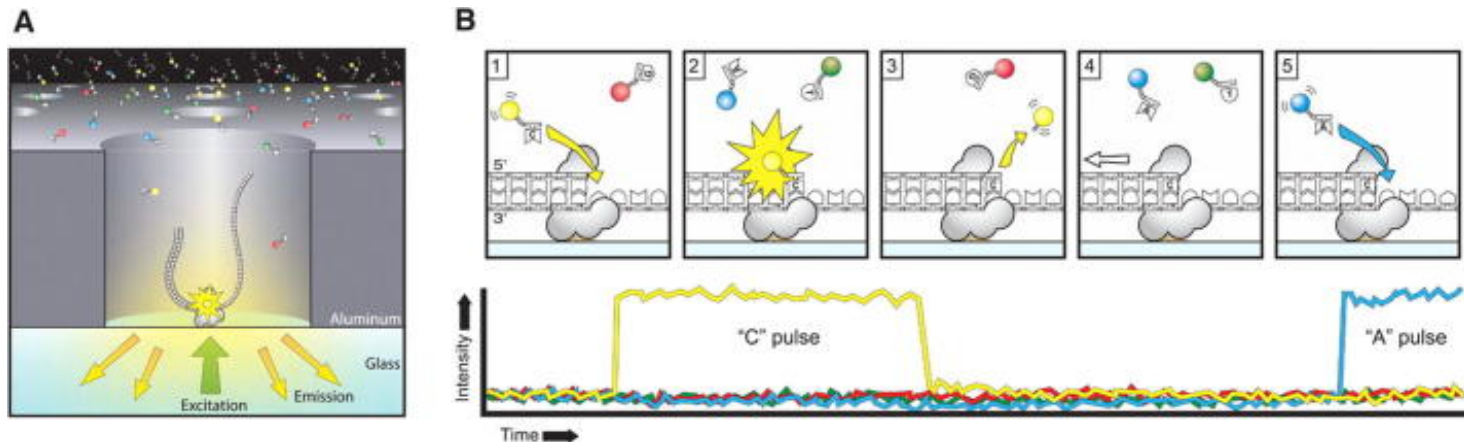
# I. Genome sequencing

3rd generation sequencing :

Improvement of template preparation : Single Molecule Sequencing (SMRT)

SMRT Sequencing Technology : **Single-Molecule** Real-Time Sequencing (3rd generation)

A true continuous imaging of dye-labelled nucleotides incorporation



[Pacific Biosciences](#) : SMRT : Single Molecule Real-Time Sequencing

# I. Genome sequencing

3rd generation sequencing :

Improvement of template preparation : Single Molecule Sequencing

## Oxford Nanopore sequencing

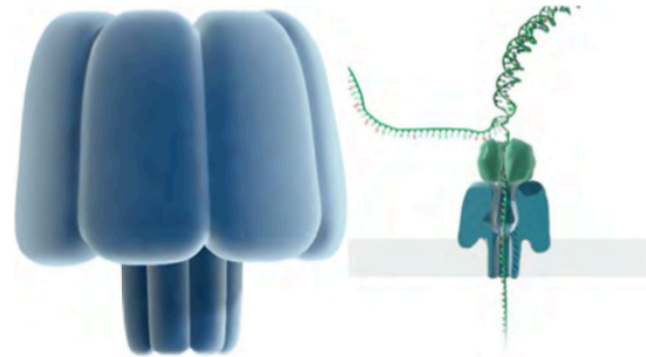


Tether keeps DNA fragment on the membrane leading to a ~20K fold higher DNA concentration close to the pore.

Motor protein unwinds DNA and ratchets it through the pore.

Abasic nucleotides in the hairpin are a recognition point.

Brake protein prevents the motor protein from zipping through the complement strand.

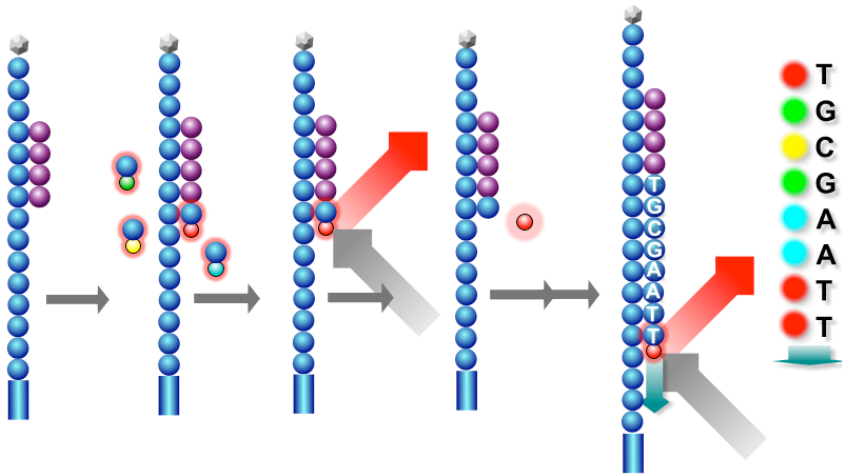


<https://www.youtube.com/watch?v=RcP85JHLmnl>

# I. Genome sequencing

2/3rd generation sequencing : real-time sequencing

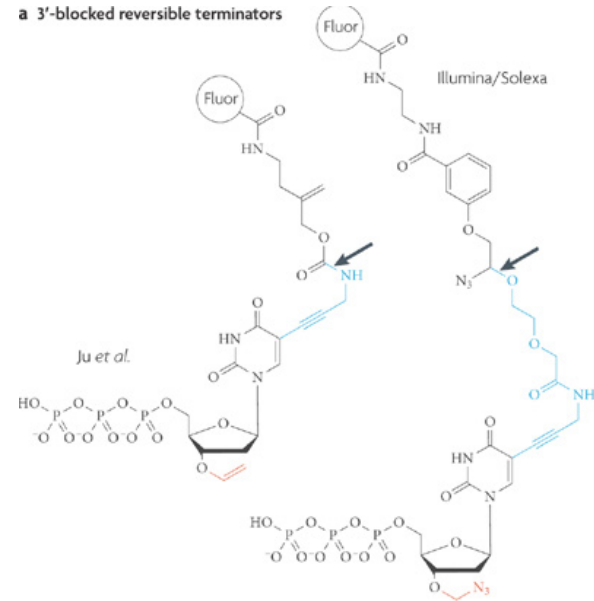
Sequencing **by reversible terminator** (illumina/Solexa genome Analyzer)



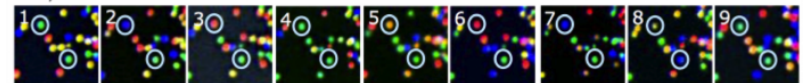
Modified polymerase incorporates nucleotides

- after each nucleotide incorporation process stops
- camera reads fluorophore signal (filter for each nucleotide type)
- terminator and labeling is removed and cycle starts again

a 3'-blocked reversible terminators



TGCTACGAT...



TTTTTTTGT...

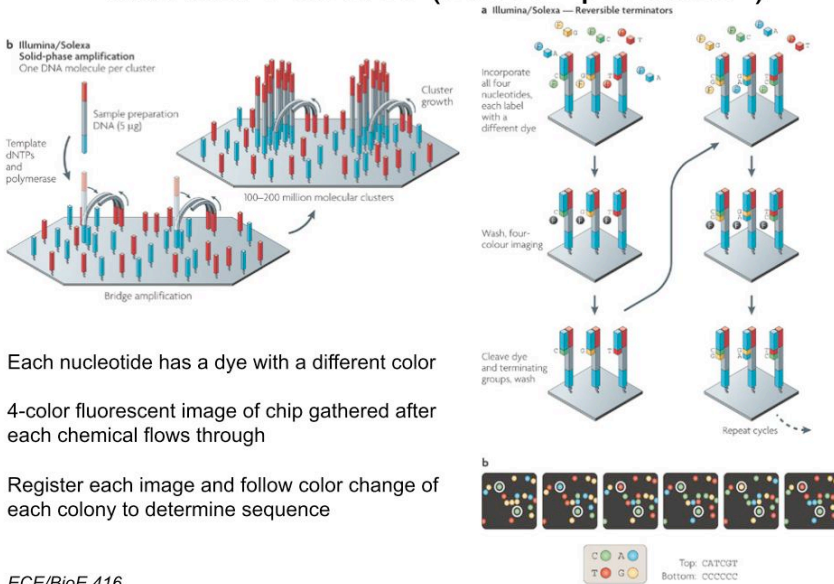
# I. Genome sequencing

2nd generation sequencing : real-time sequencing

Sequencing by reversible terminator

Used in combination with bridge PCR in the illumina/Solexa genome Analyzer

## Illumina Colonies (called “polonies”)



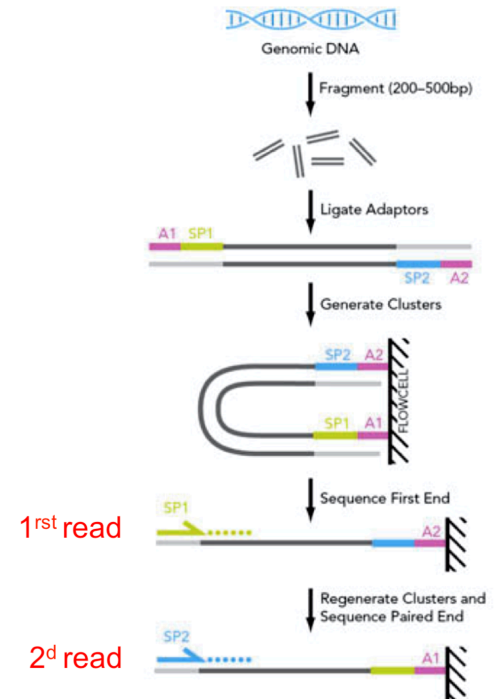
Each nucleotide has a dye with a different color

4-color fluorescent image of chip gathered after each chemical flows through

Register each image and follow color change of each colony to determine sequence

FCF/RioF 416

## Paired-end sequencing



# I. Genome sequencing

2nd generation sequencing : real-time sequencing

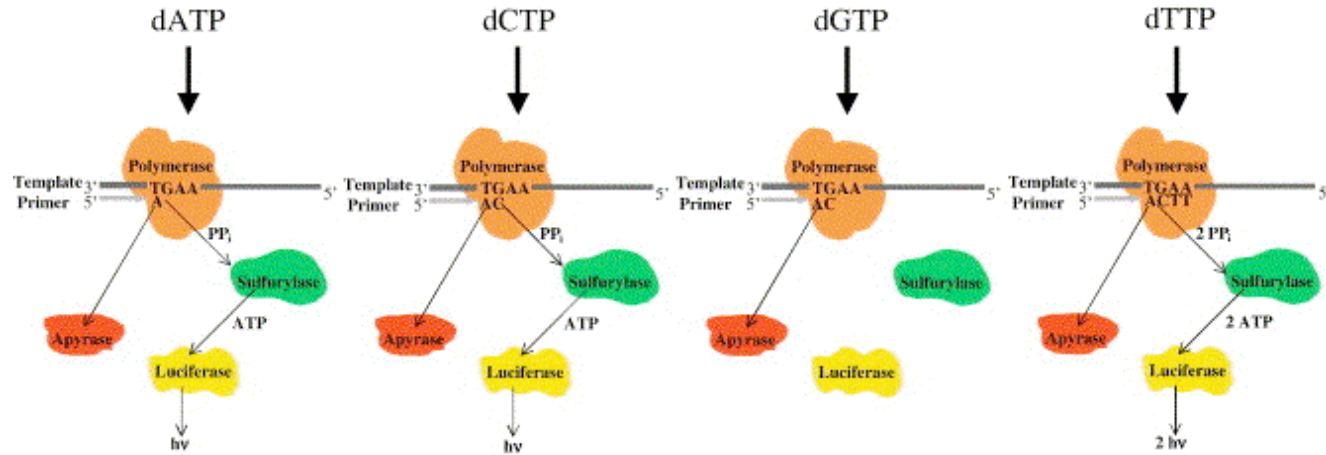
Sequencing **by reversible terminator**



# I. Genome sequencing

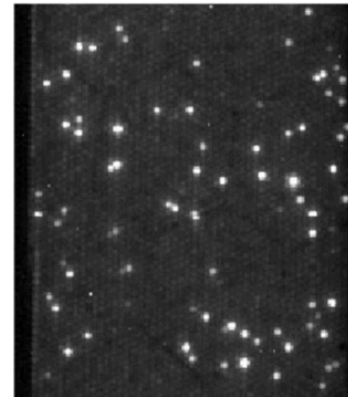
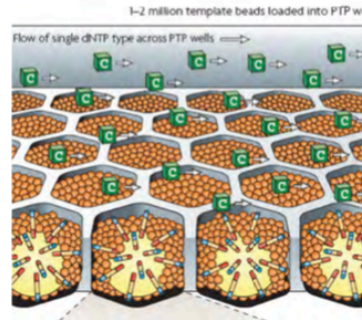
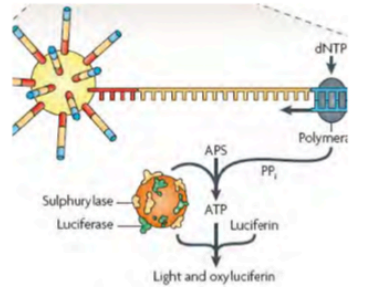
2nd generation sequencing : real-time sequencing

## Pyrosequencing



Used in the 454 platform (Life Sciences)

emPCR+Pyrosequencing



# I. Genome sequencing

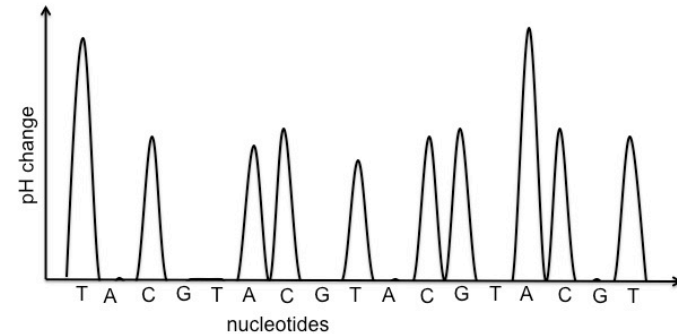
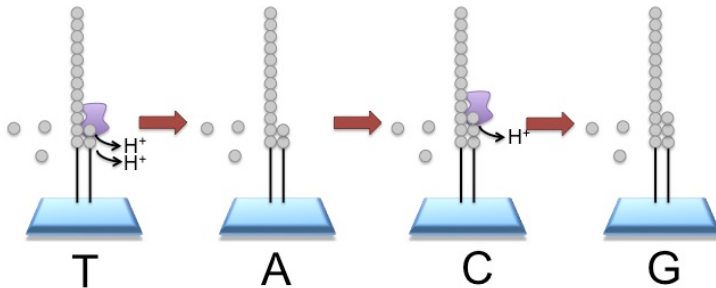
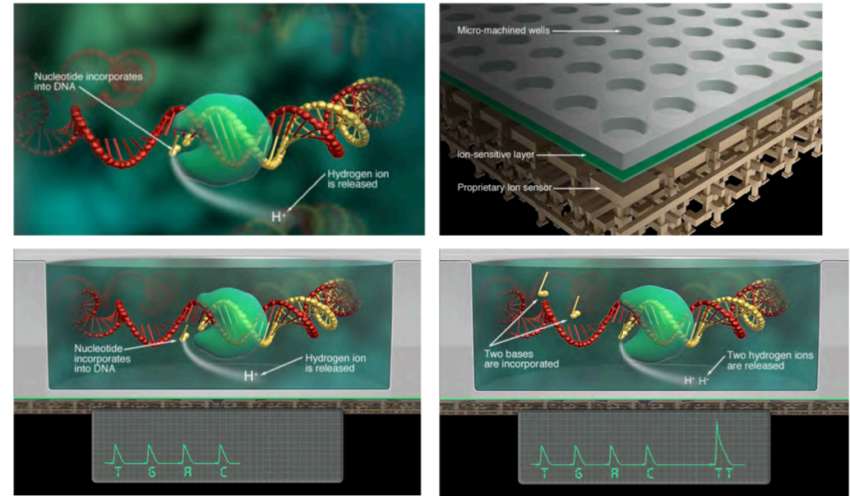
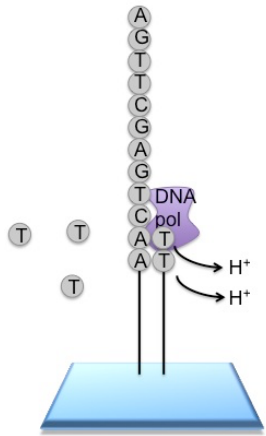
2nd generation sequencing : real-time sequencing



## Ion torrent sequencing

Life Technologies: Ion Proton & Ion PGM

**Ion proton sequencing (Ion torrent):** emPCR + measurement of changes in pH

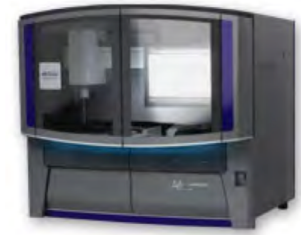




# I. Genome sequencing

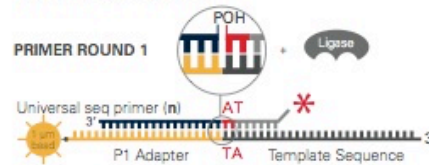
2nd generation sequencing : real-time sequencing

Ligase-based sequencing (ABI's Solid technology)



- Difference – DNA ligase
- Hybridization of a fluorescently labelled probe
- SOLiD cycle of 1,2-probe hybridization

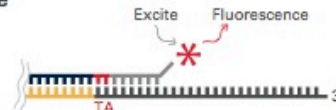
## 1. Prime and Ligate



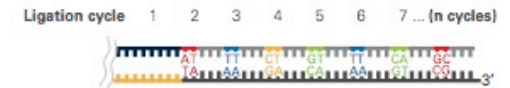
## 4. Cleave off FLUOR



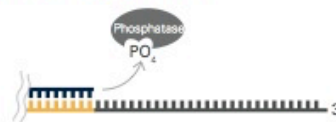
## 2. Image



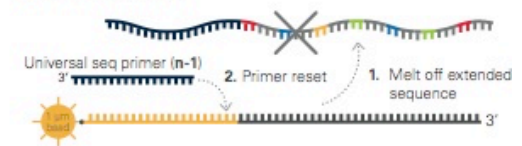
## 5. Repeat steps 1-4 to Extend Sequence



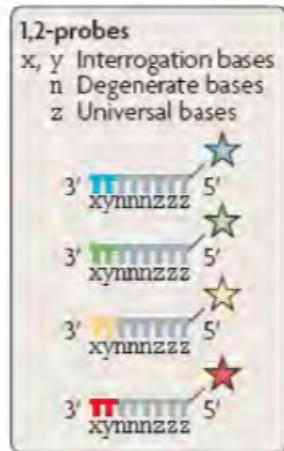
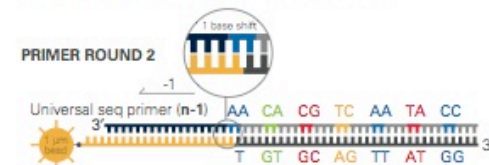
## 3. Cap Unextended Strands



## 6. Primer Reset



## 7. Repeat steps 1-5 with new primer

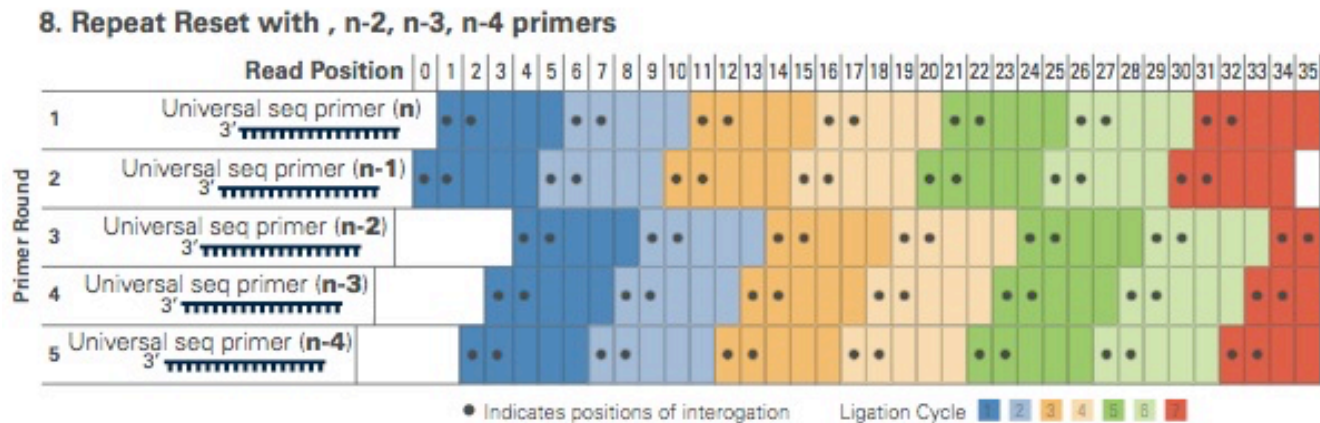


# I. Genome sequencing

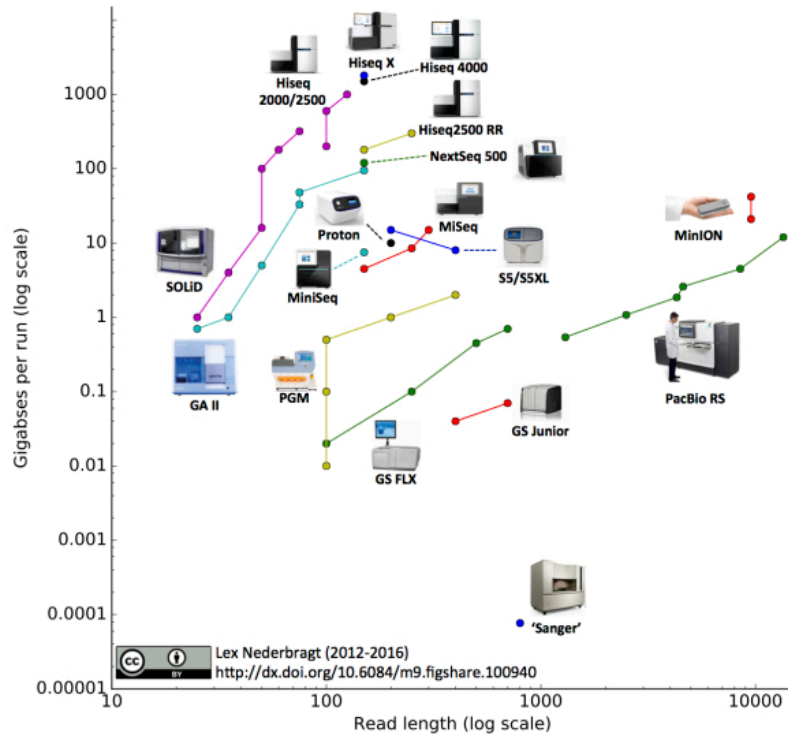
2nd generation sequencing : real-time sequencing

Ligase-based sequencing

## [SOLiD DNA Sequencing Technology \(Applied Biosystems\)](#)



## II. Genome sequencing : the good choice



Performance comparison of sequencing platforms of various generations

Method	Generation	Read length (bp)	Single pass error rate (%)	No. of reads per run	Time per run	Cost per million bases (USD)	Refs.
Sanger ABI 3730x1	1st	600–1000	0.001	96	0.5–3 h	500	[14], [18], [19], [20], [21]
Ion Torrent	2nd	200	1	$8.2 \times 10^7$	2–4 h	0.1	[15], [25]
454 (Roche) GS FLX+	2nd	700	1	$1 \times 10^6$	23 h	8.57	[14], [17], [27]
Illumina HiSeq 2500 (High Output)	2nd	$2 \times 125$	0.1	$8 \times 10^9$ (paired)	7–60 h	0.03	[9], [16], [26]
Illumina HiSeq 2500 (Rapid Run)	2nd	$2 \times 250$	0.1	$1.2 \times 10^9$ (paired)	1–6 days	0.04	[9], [16], [26]
SOLiD 5500x1	2nd	$2 \times 60$	5	$8 \times 10^8$	6 days	0.11	[14], [24]
PacBio RS II: P6-C4	3rd	$1.0\text{--}1.5 \times 10^4$ on average	13	$3.5\text{--}7.5 \times 10^4$	0.5–4 h	0.40–0.80	[5], [12], [15]
Oxford Nanopore MinION	3rd	$2\text{--}5 \times 10^3$ on average	38	$1.1\text{--}4.7 \times 10^4$	50 h	6.44–17.90	[22], [23]

## II. Genome sequencing : the good choice

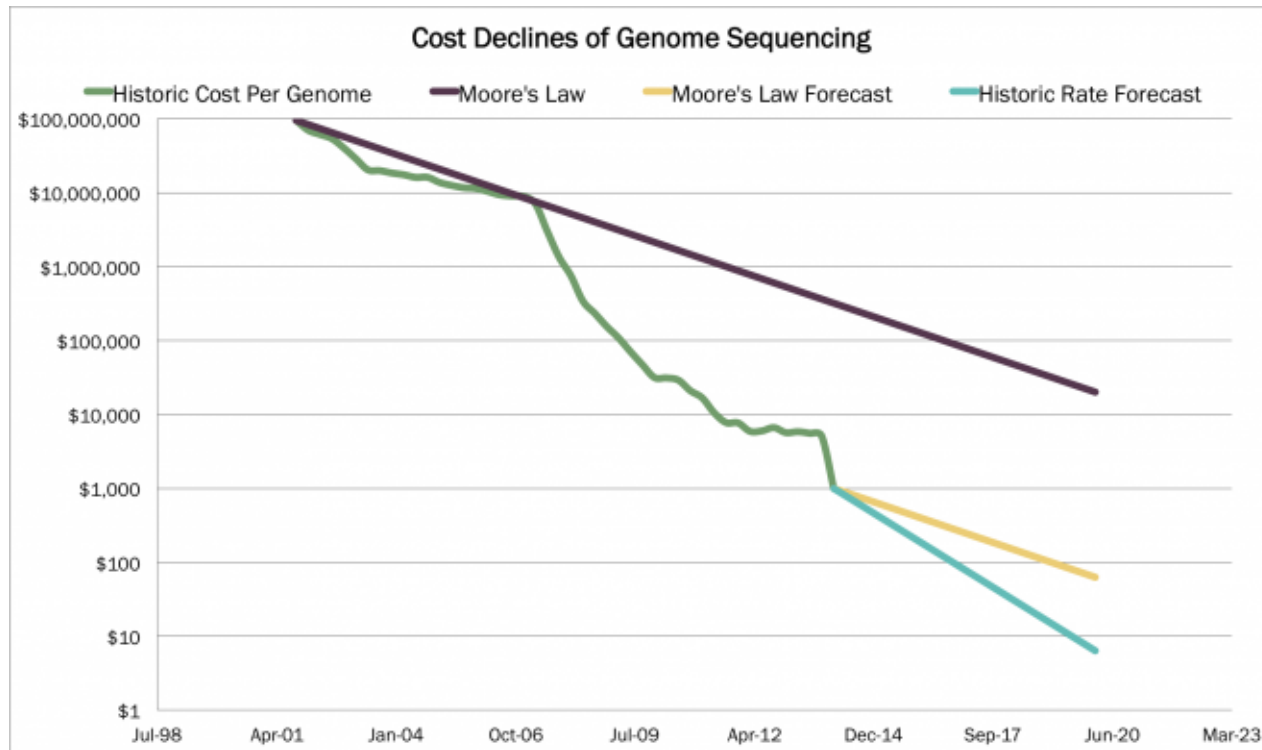
### Number of sequencing machines by country

Name	Number of machines
United States	818
China	200
United Kingdom	137
Germany	135
Australia	79
Canada	74
Spain	56
Netherlands	41
France	38
Japan	34

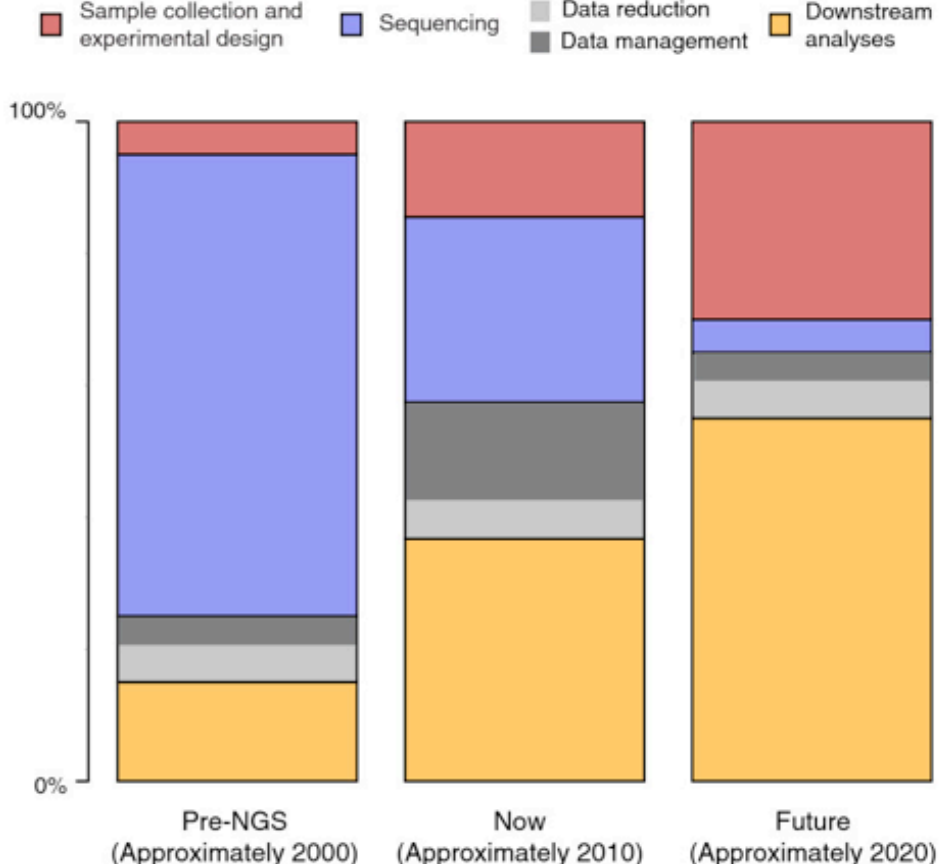
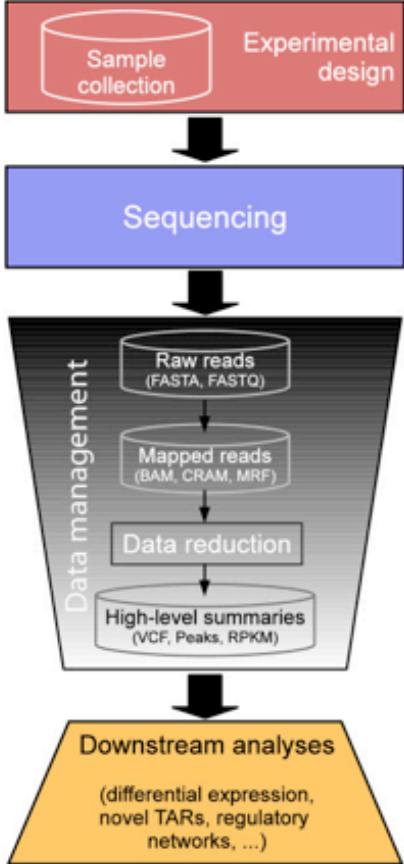
### Centres with platform

Name	Number of centres
Illumina Genome Analyser 2x	279
Roche 454	265
Illumina HiSeq 2000	178
ABI SOLiD	173
Ion Torrent	101
Pacific Biosciences	26
Illumina MiSeq	23
Polonator	5

# I. Genome sequencing : the price



# II. Data analysis



## II. Data analysis

### ■ Bioinformatics tools for:

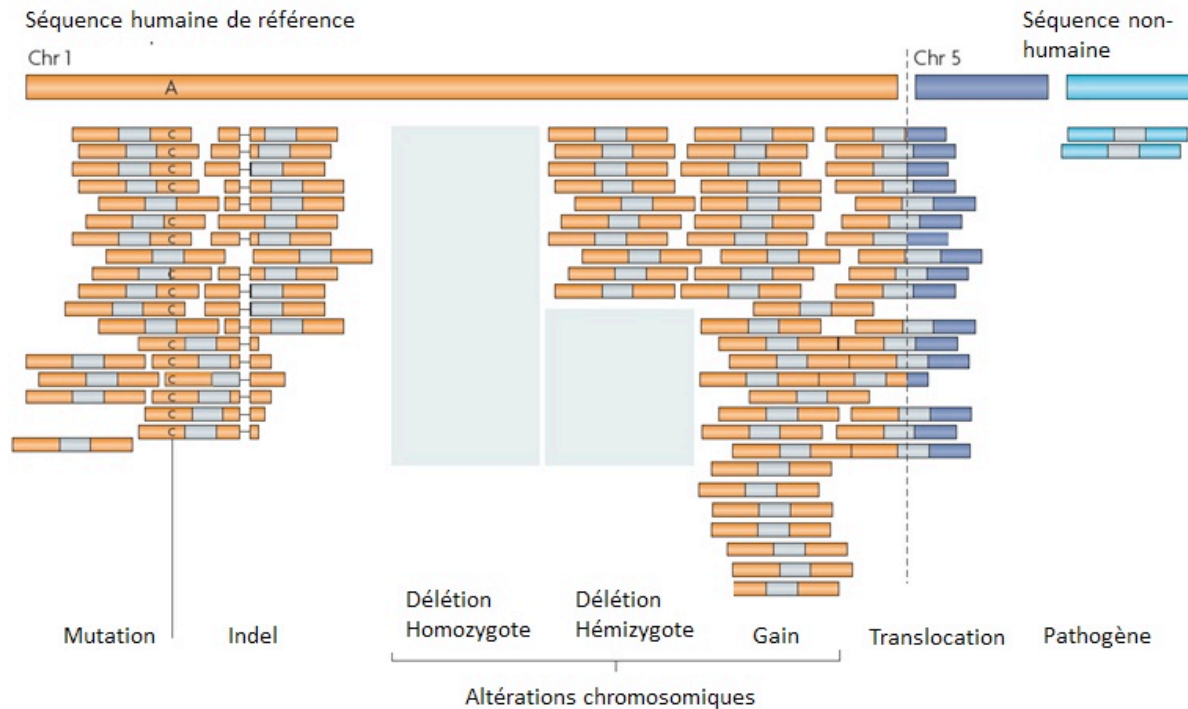
- Alignment
- Base calling/polymorphism detection
- *De novo* assembly
- Genome browsing or annotation

### ■ Challenging problems:

- *De novo* assembly of short reads -> mate-paired libraries required
- Reads in repetitive regions

### III. NGS Applications

- Genome resequencing: polymorphism and mutation discovery in humans (1000 Genomes Project)





### III. NGS Applications

#### ■ Transcriptome sequencing:

Gene expression

Alternative splicing

Transcript annotation

