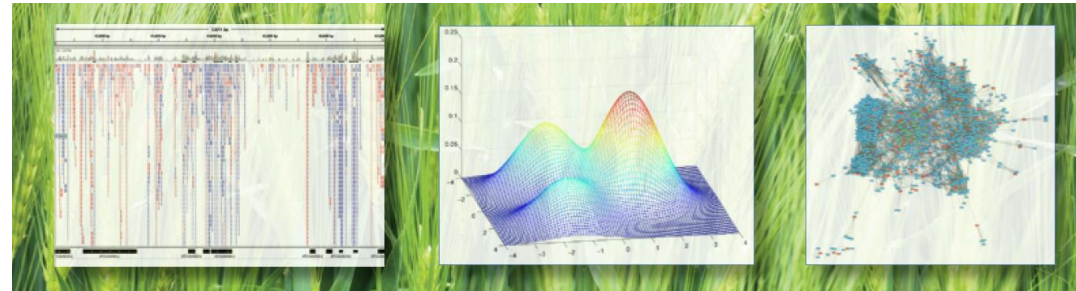


# Ecole Chercheur 2017

De l'expression des gènes aux réseaux

Module 1 : High Troughput Sequencing Technologies, available and futures...



Mathilde CLEMENT  
mathilde.clement@inra.fr

- I. What is transcriptomic analysis?
- II. DNA Sequencing history
- III. Available High Throughput Sequencing technologies
- IV. What is France Génomique?

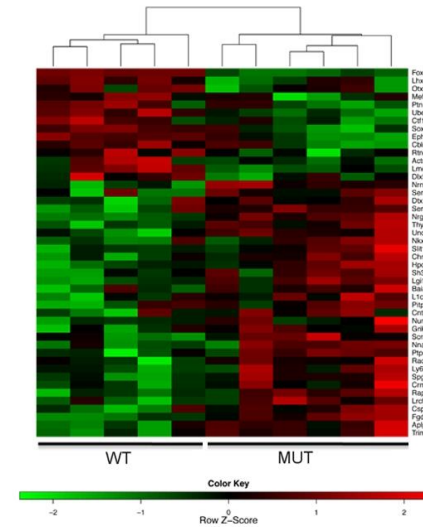
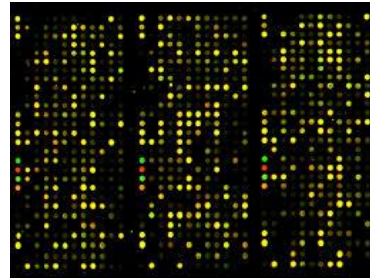
# What is a transcriptomic analysis ?

**Quantification of the changing expression levels of each transcript during development and under different conditions**

# What is a transcriptomic analysis ?

**Quantification of the changing expression levels of each transcript during development and under different conditions**

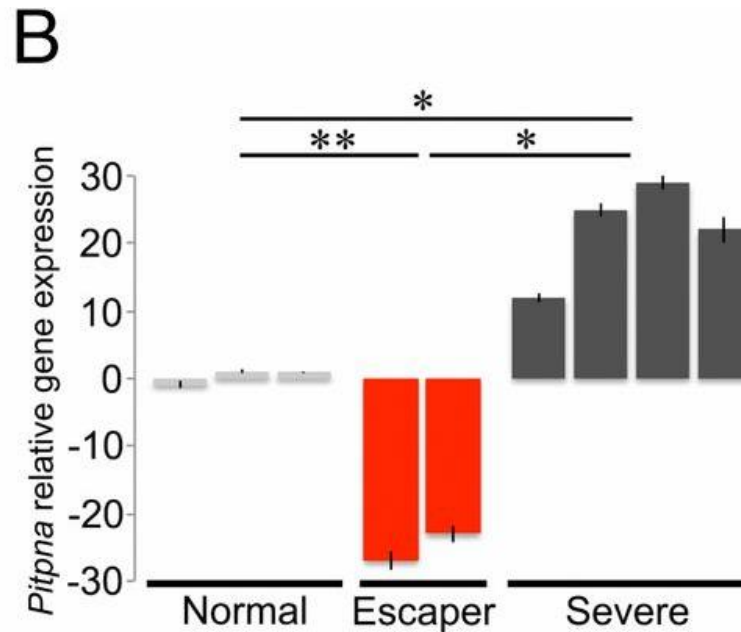
**Microarray technology in the mid-1990s**



Foxp2  
Lhx8  
Chaz  
Meis2  
Pit1  
Lhx3a  
Ctcf  
Sox11  
Eno2  
Ctcf1  
Pit4  
Acta2  
Dlx5  
Nes1  
Sema4f  
Ctcf  
Sema3b  
Ngn1  
Thy1  
Lmo2  
Nkx2-2  
Sox1  
Chrm3  
Hes6  
Sox9  
Lgl1  
Wt1  
Lmo2  
Pipem1  
Ctcf  
Numb  
Ctcf  
Scn8a  
Nes1  
Pipem1  
Pipem1  
Lgl1  
Sox7  
Ctcf1  
Nes1  
Lmo2  
Chrm3  
Foxp2  
Acta1  
Tm6

# What is a transcriptomic analysis ?

## Repression of phosphatidylinositol transfer protein $\alpha$ ameliorates the pathology of Duchenne muscular dystrophy

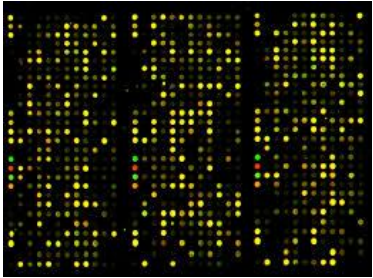


Agilent Technologies  
Canine 4 x 44K

# What is a transcriptomic analysis ?

Limitation of microarray analysis :

Why did we need another technique of gene expression analysis?

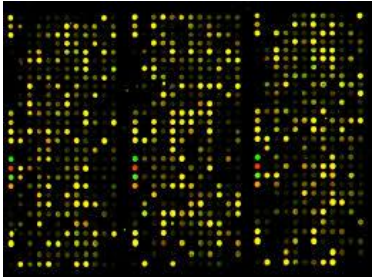


- **Arrays had to be designed and manufactured for each species**
- **Detection of only « known » genes**
- **No detection of genes weakly differentially expressed**

# What is a transcriptomic analysis ?

Limitation of microarray analysis :

Why did we need another technique of gene expression analysis?



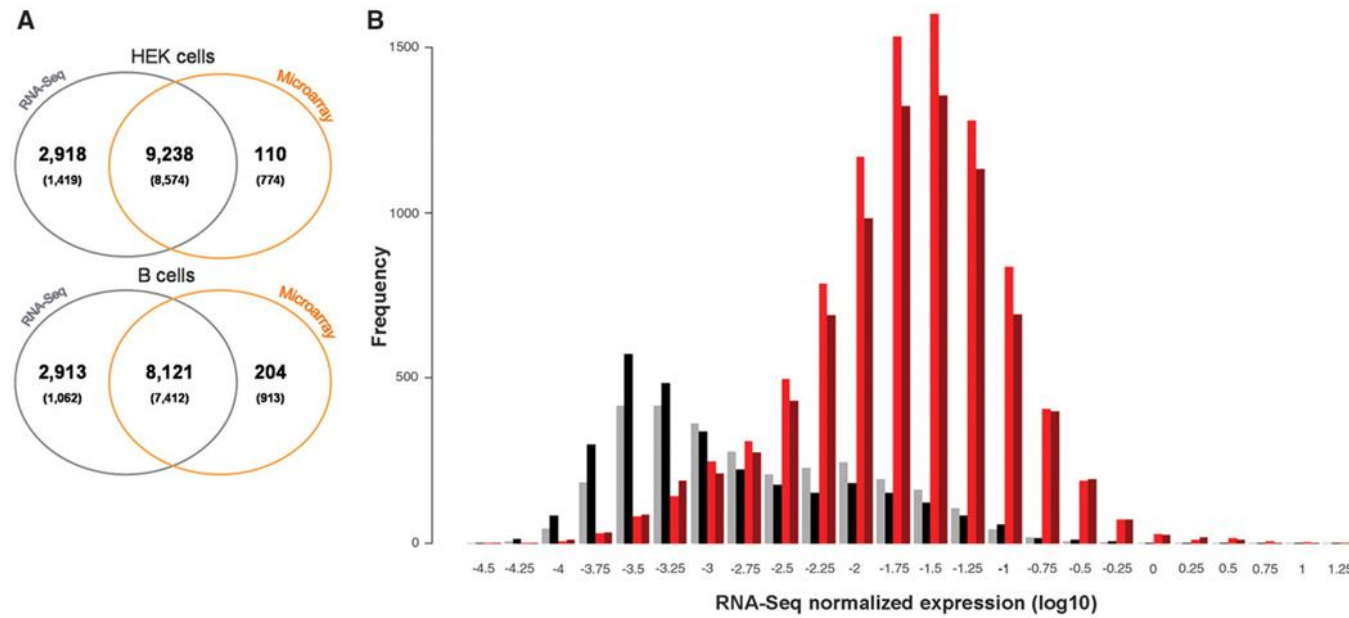
- **Arrays had to be designed and manufactured for each species**
- **Detection of only « known » genes**
- **No detection of genes weakly differentially expressed**



**In 2008 , the first RNA-Seq protocol appering**

- **measuring transcriptomes at base-par resolution**
- **using essentially the same protocol for any species**
- **minimal noise level**

# What is a transcriptomic analysis ?



Sultan et al., (2008) Sciences

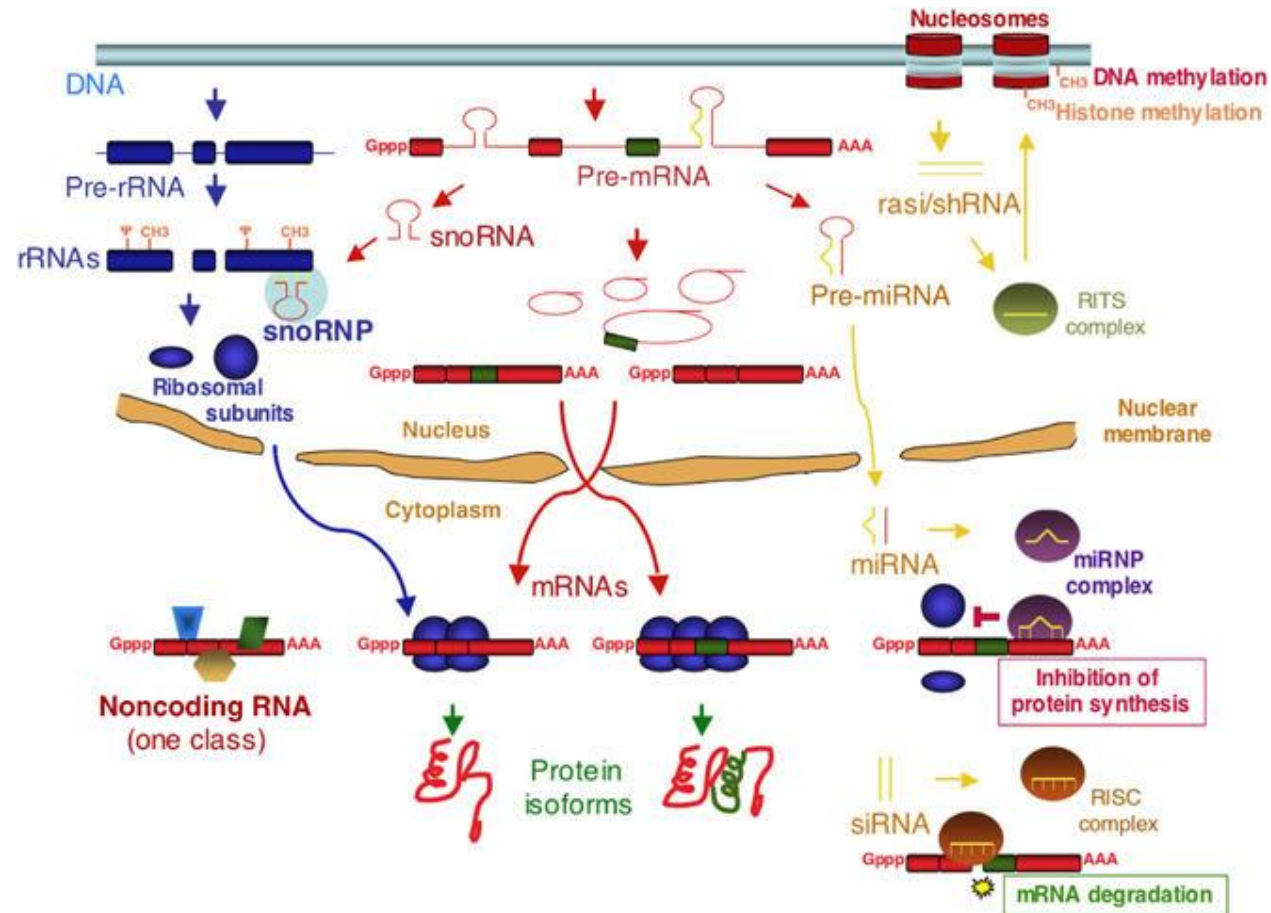


RNA-Seq can detect 25% more genes than can microarrays



# What is a transcriptomic analysis ?

An overview of the eukaryotic transcriptome...



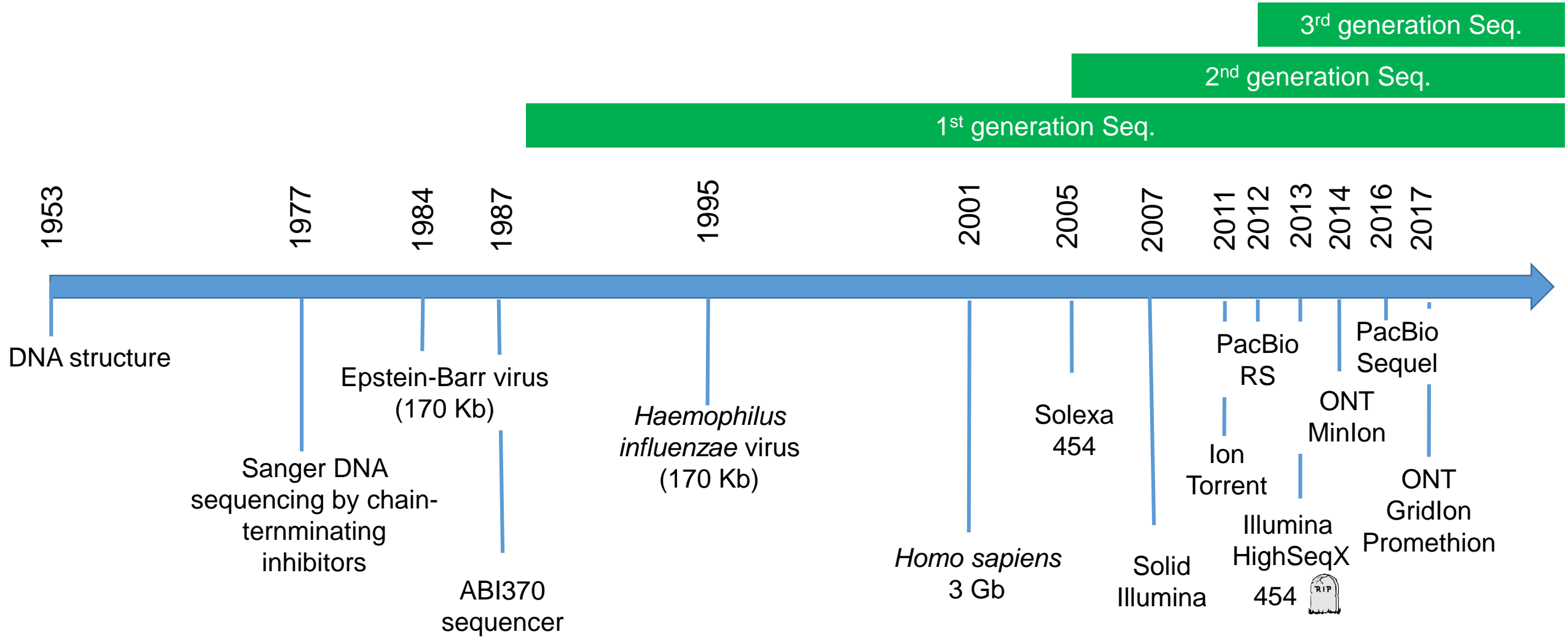
# What is a transcriptomic analysis ?

An overview of the eukaryotic transcriptome...

- Catalogue all species of transcript, including mRNAs, non-coding RNAs and small RNAs;
- Determine the transcriptional structure of genes, in terms of their start sites, 5' and 3' ends, splicing patterns and other post-transcriptional modifications;
- and to quantify the changing expression levels of each transcript during development and under different conditions

# DNA Sequencing History

## Sequencing over the Ages



# DNA Sequencing History

## 1<sup>st</sup> generation sequencing \_ Sanger sequencing

Frederick Sanger

13 Aug 1918 – 19 Nov 2013

Won the Nobel Prize for Chemistry in **1958** and **1980**. Published the dideoxy chain termination method or “Sanger method” in 1977



<http://daily.aj/1f1XeTB>

Proc. Natl. Acad. Sci. USA  
Vol. 74, No. 12, pp. 5463-5467, December 1977  
Biochemistry

### DNA sequencing with chain-terminating inhibitors

[DNA polymerase; nucleotide sequences; bacteriophage  $\phi$ X174]

F. SANGER, S. NICKLEN, AND A. E. COULSON

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QR, England

Contributed by F. Sanger, October 6, 1977

**ABSTRACT** A new method for determining nucleotide sequences in DNA is described. It is similar to the “plus and minus” method [Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* **94**, 441-448] but makes use of the 2',3'-dideoxy and 2-thiothymine nucleoside analogues of the normal deoxynucleoside triphosphates, which act as specific chain-terminating inhibitors of DNA polymerase. The technique has been applied to the DNA of bacteriophage  $\phi$ X174 and in some rapid acid minisequencing that differs from the plus and minus method.

The “plus and minus” method (1) is a relatively rapid and simple technique that has made possible the determination of the sequence of the genome of bacteriophage  $\phi$ X174 (2). It depends on the use of DNA polymerase to transcribe specific regions of the DNA under controlled conditions. Although the method is considerably more rapid and simple than other available techniques, neither the “plus” nor the “minus” method is completely accurate, and in order to establish a sequence both must be used together, and sometimes confirmed

a determination of sites at which the 3' hydroxyl group is substituted in these positions with respect to the 3' hydroxyl group. The thymine (ara) nucleoside act as chain-terminating inhibitors of *Escherichia coli* DNA polymerase I in a manner comparable to dIT (3), although synthesized chains ending in 3' araC can be further extended by some mammalian DNA polymerases (3). In order to obtain a suitable pattern of bands from which an extensive sequence can be read it is necessary to have a ratio of terminating triphosphate to normal triphosphate such that only partial incorporation of the terminator occurs. For the dideoxy derivatives this ratio is about 100, and for the arabinoside derivatives about 5000.

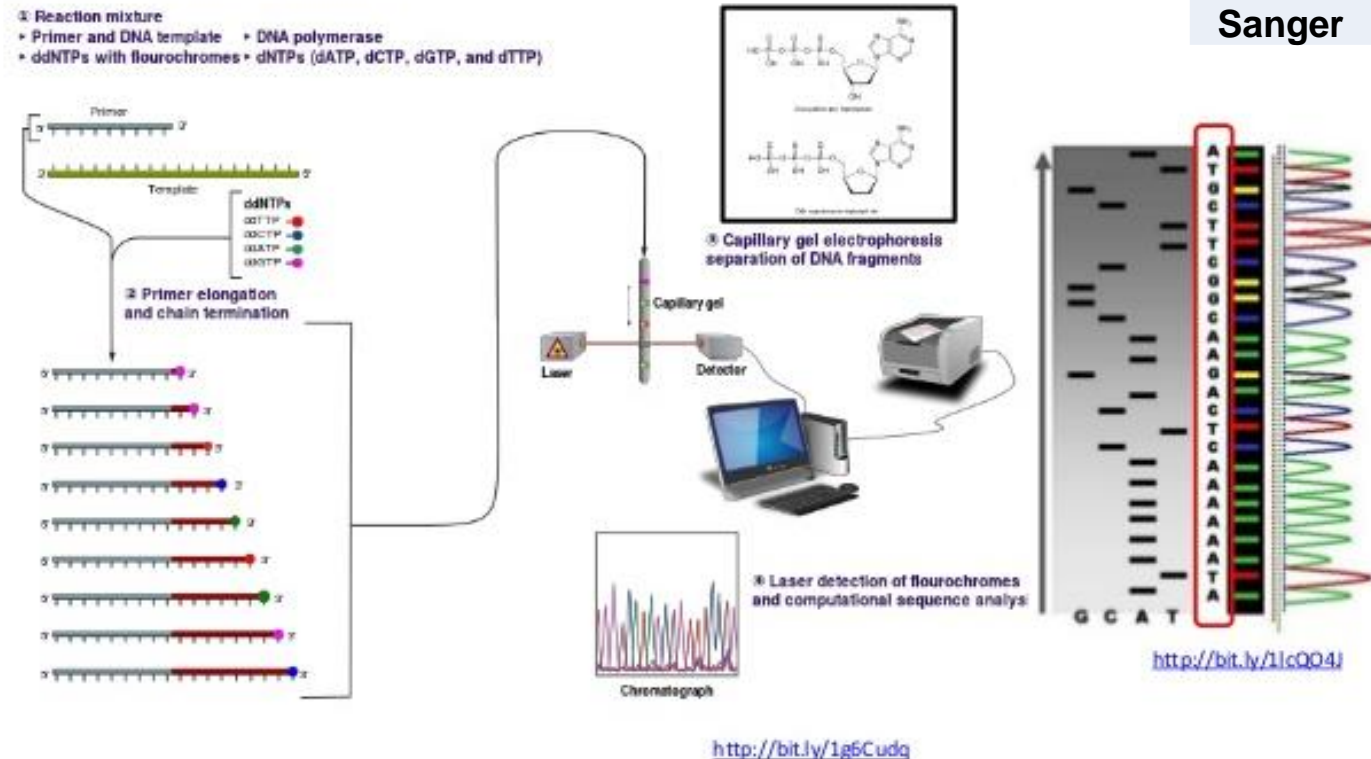
#### METHODS

**Preparation of the Triphosphate Analogues.** The preparation of dITP has been described (3, 7), and the material is now commercially available. dItA has been prepared by

# DNA Sequencing History

## 1<sup>st</sup> generation sequencing \_ Sanger sequencing

Platform	Reads per run	Read length	Bases per run (gigabases)
ABI Sanger	96	800	0.0000768

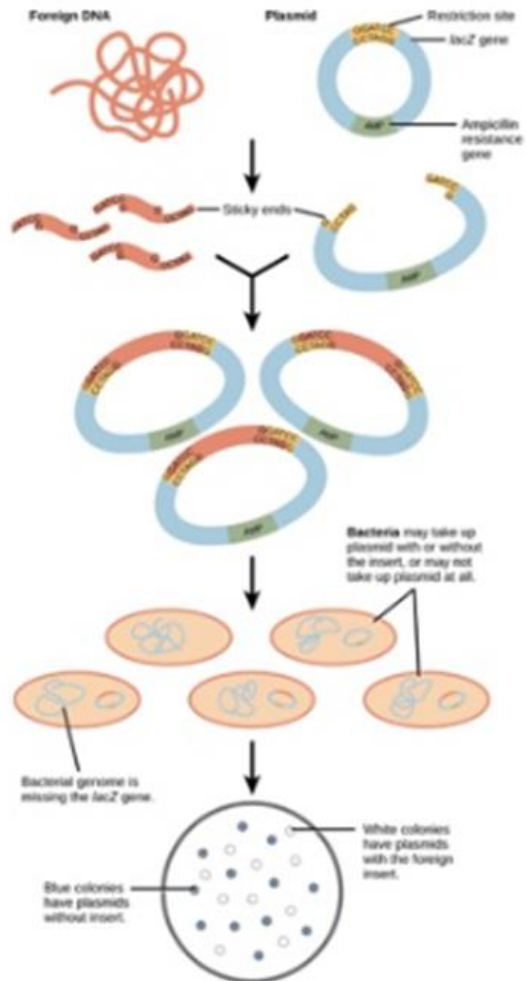


The method relies on the use of (dideoxynucleotide) ddNTP which will terminate the polymerization

# DNA Sequencing History

## 1<sup>st</sup> generation sequencing \_ Sanger sequencing

Platform	Reads per run	Read length	Bases per run (gigabases)
ABI Sanger	96	800	0.0000768

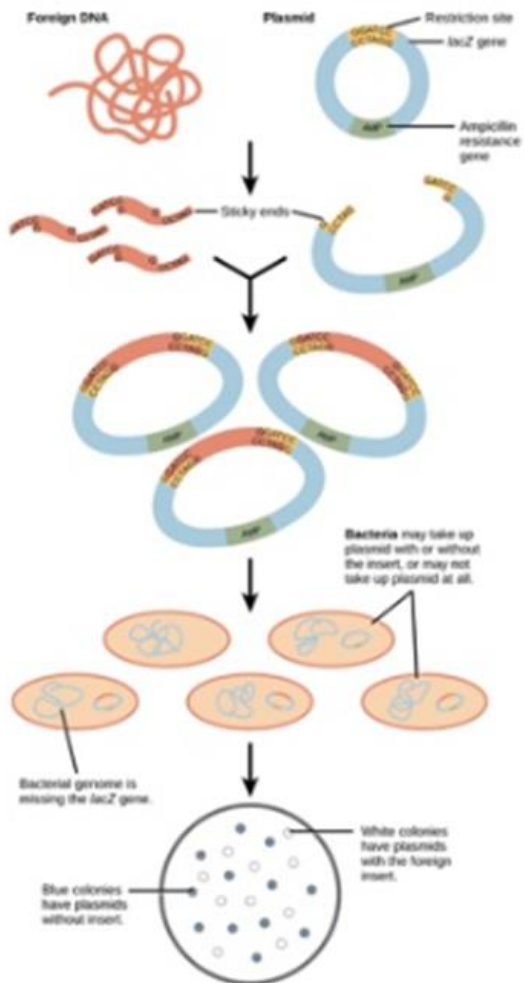




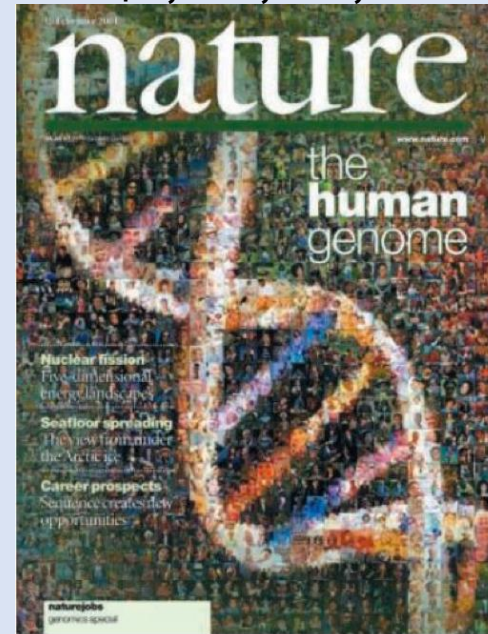
# DNA Sequencing History

## 1<sup>st</sup> generation sequencing \_ Sanger sequencing

Platform	Reads per run	Read length	Bases per run (gigabases)
ABI Sanger	96	800	0.0000768



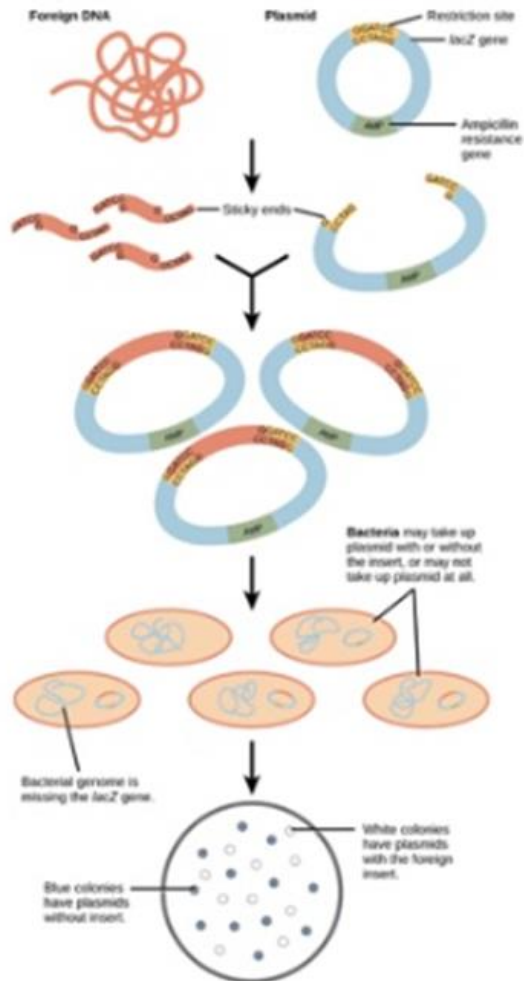
**The International Human Genome Consortium**  
**13 years (1990-2003)**  
**~ \$3,000,000,000**



*Nature*. 2001 Feb 15;409(6822):860-921.  
 Initial sequencing and analysis of the human genome.

# DNA Sequencing History

## 1<sup>st</sup> generation sequencing \_ Sanger sequencing



Very high quality sequences (99.999%), 800 bp

Sanger sequencing is not an high throughput technology



# DNA Sequencing History

## 1<sup>st</sup> generation sequencing \_ Sanger sequencing

Very high quality sequences (99.999%)

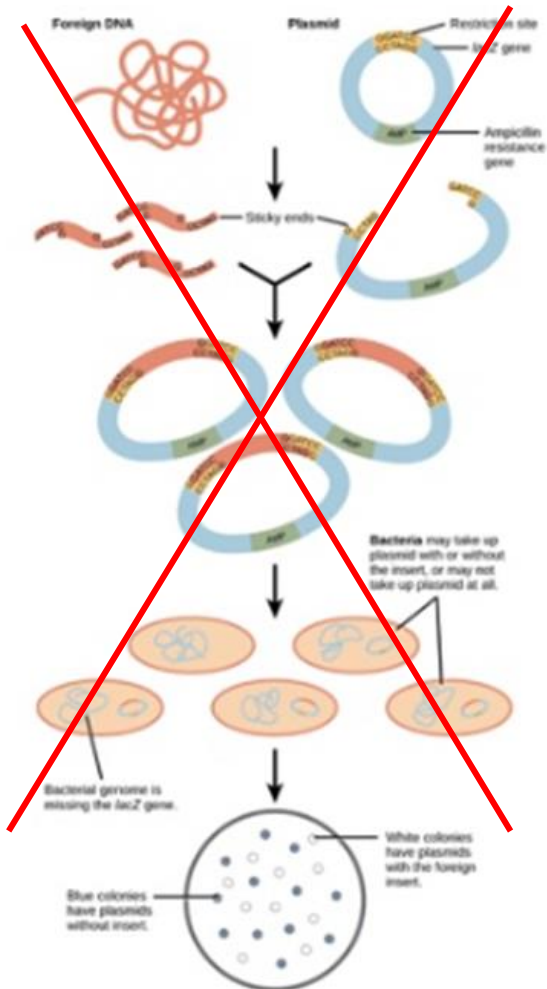
Sanger sequencing is not an high throughput technology



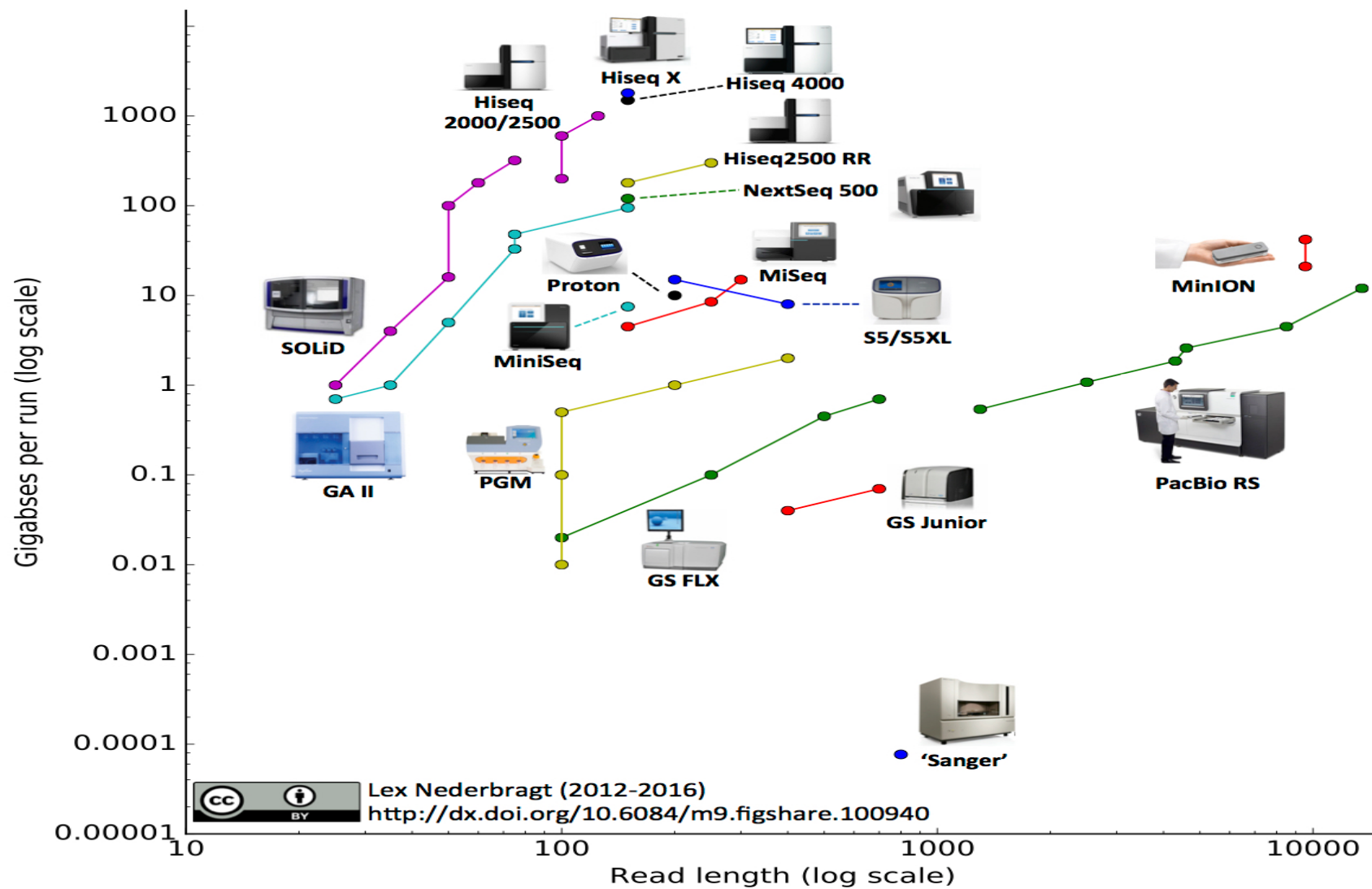
**Massive parallel sequencing**  
**No need to clone sequences and performe libraries of plasmids in bacteria**



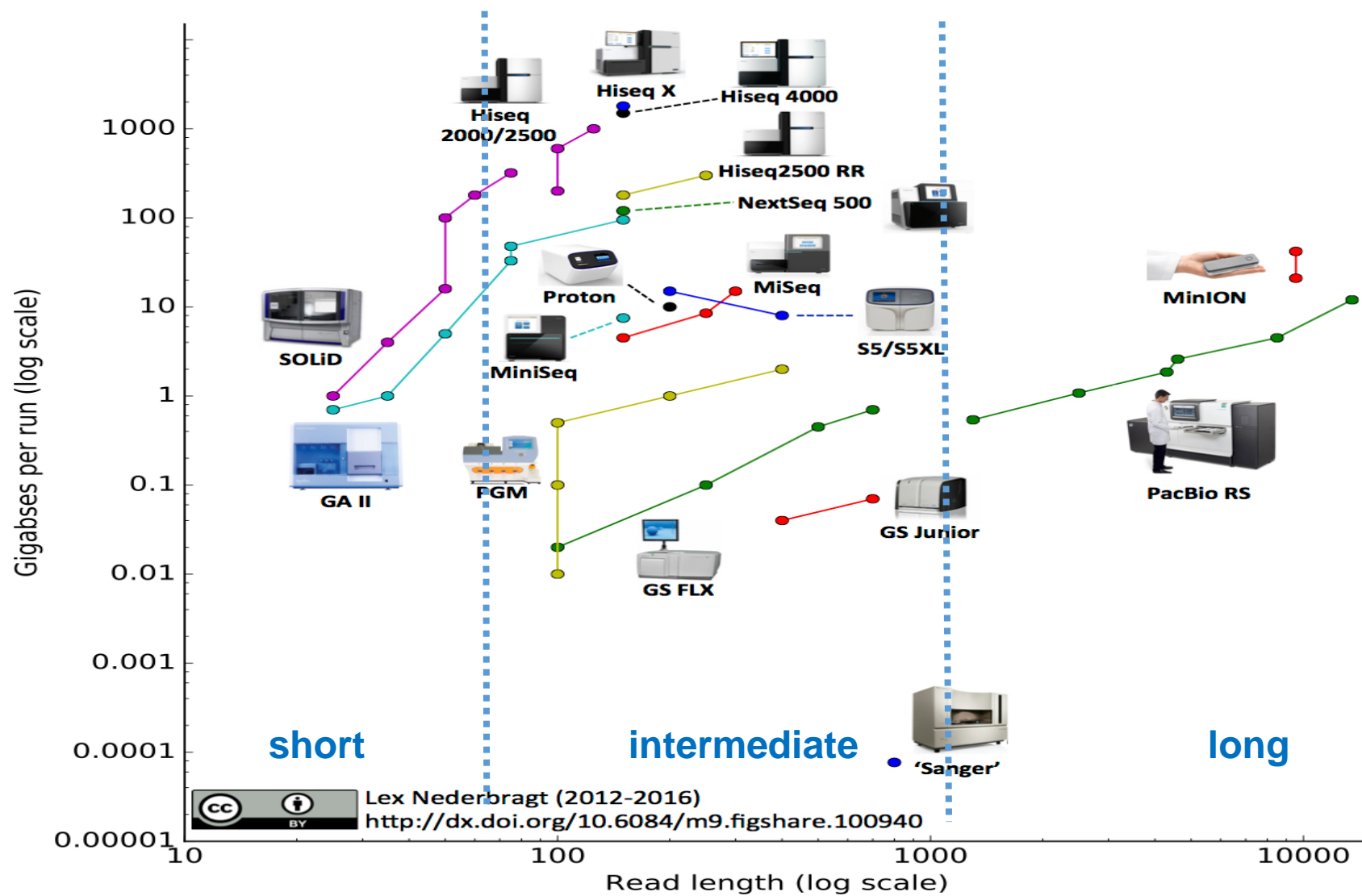
2<sup>nd</sup> /3<sup>rd</sup> generation sequencing technologie\_Next Generation Sequencing\_NGS



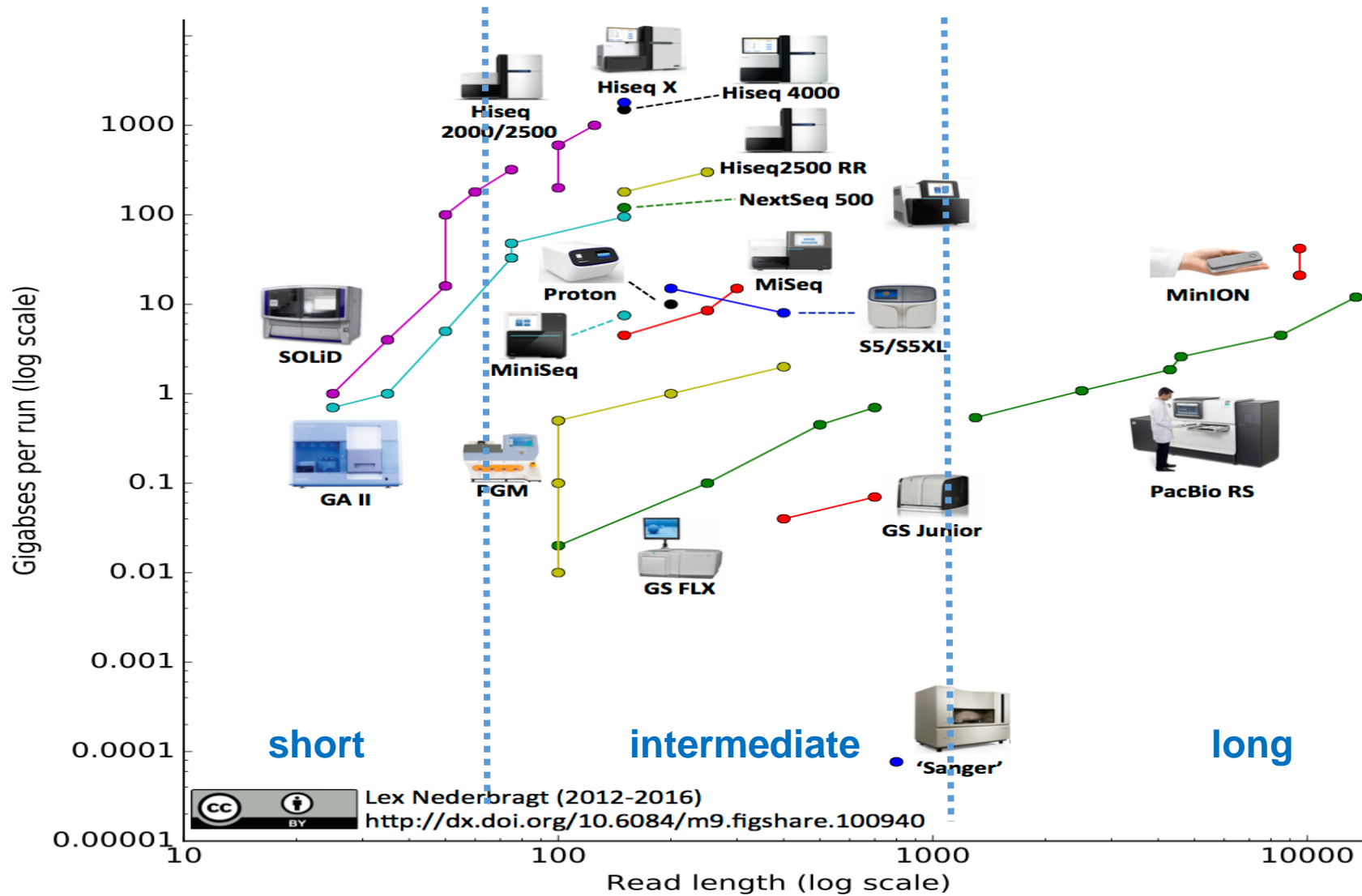
# Available High Throughput Sequencing Technologies...



# Available High Throughput Sequencing Technologies...



# Available High Throughput Sequencing Technologies...



....the arms race....

Human genome project:  
13 years (1990-2003)  
~ \$3,000,000,000



With Illumina XTen  
platform (2014):  
< 1 week  
~\$1,000



Lex Nederbragt (2012-2016)  
<http://dx.doi.org/10.6084/m9.figshare.100940>

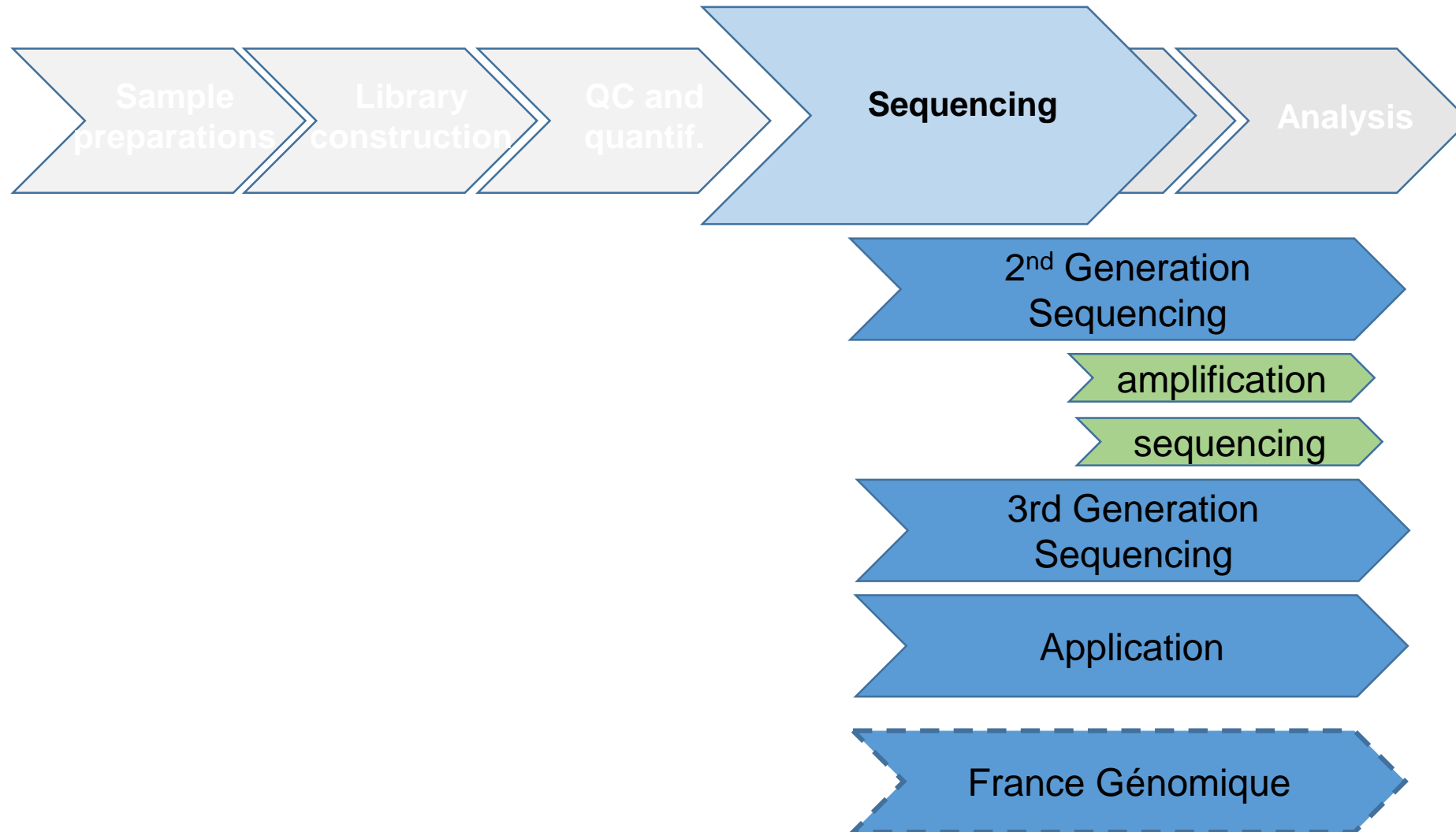
# Available High Throughput Sequencing Technologies...

The general principle of NGS is to sequence several DNA fragments in the same sample



# Available High Throughput Sequencing Technologies...

The general principle of NGS is to sequence several DNA fragments in the same sample



## Amplification

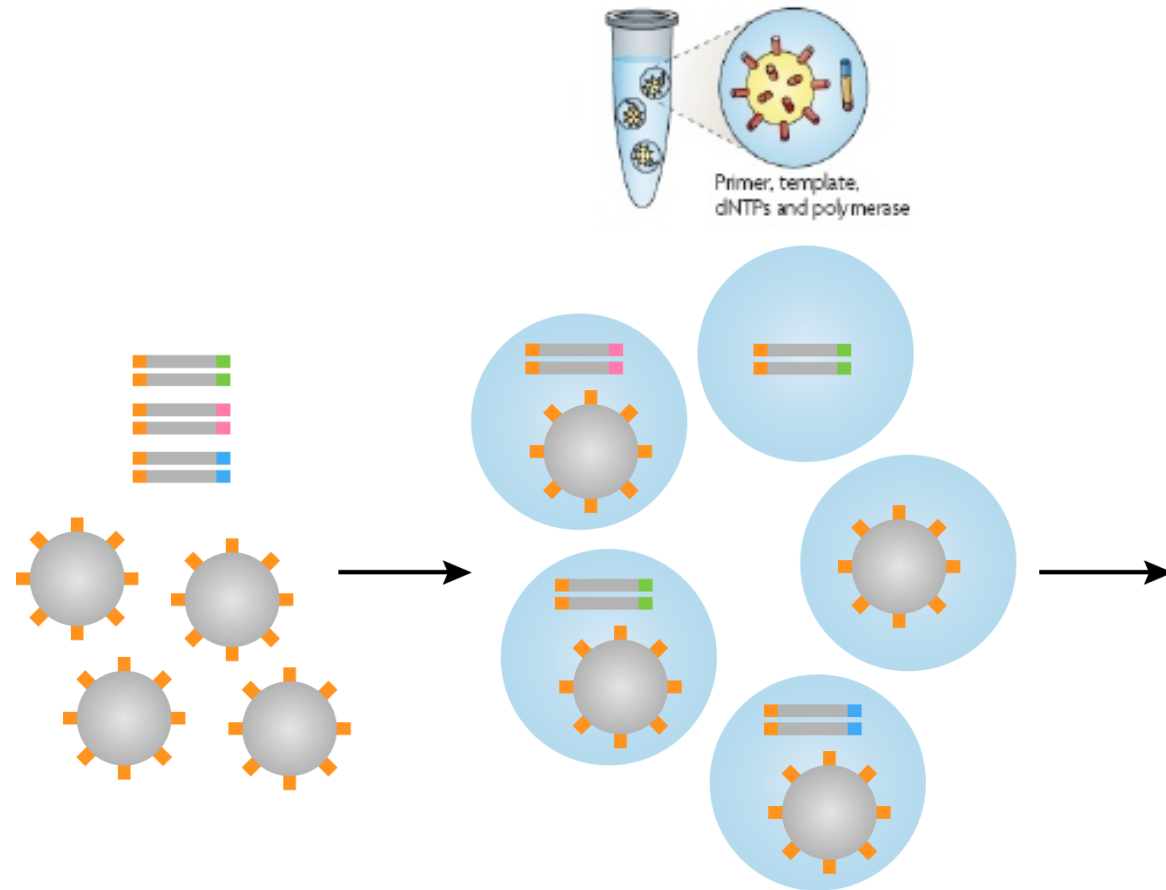
Library amplification is required to obtain a sufficient signal from the sequencer.  
Two types of amplification:

1- the oil/ water emulsion ePCR emulsion PCR  
**Roche 454, SOLID, Ion PGM**

2- immobilisation on solid phase with an oligonucleotide (primer) and bridge amplification  
**Illumina**

## Amplification

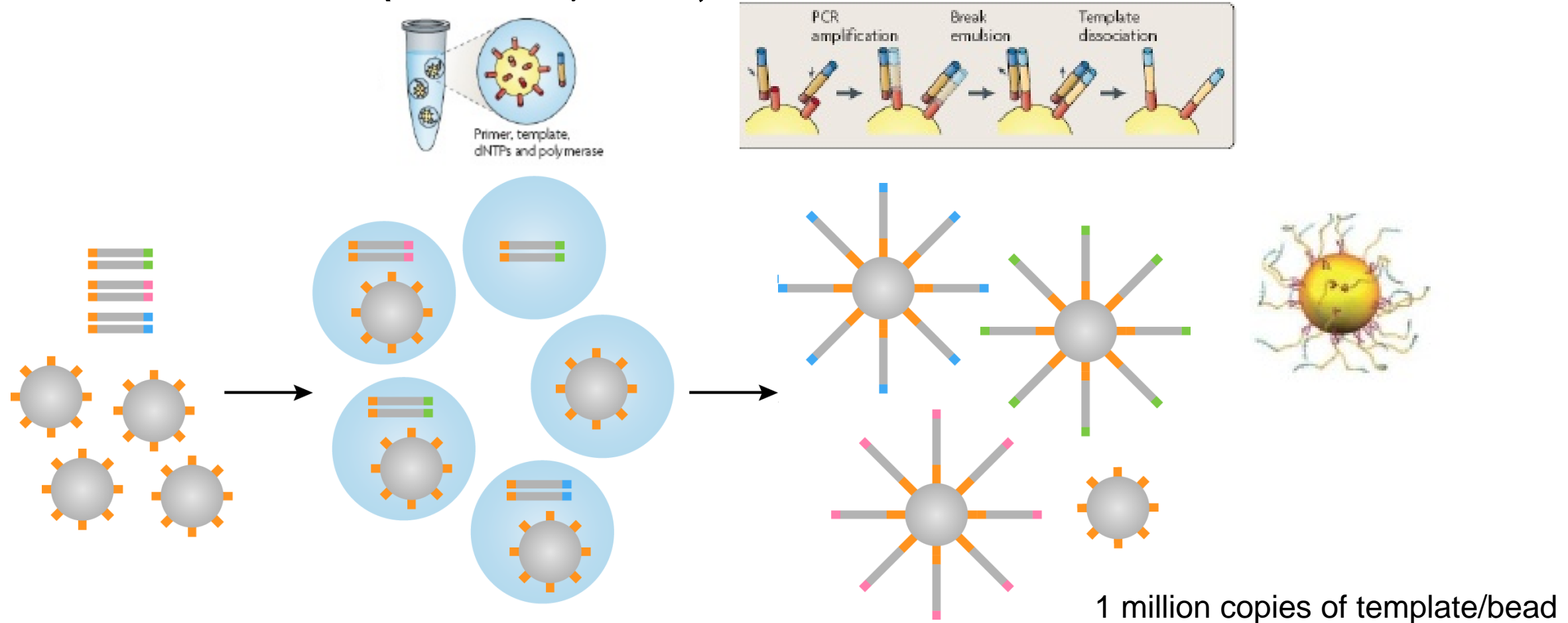
### 1- Emulsion PCR (Roche 454, SOLID, Ion PGM)





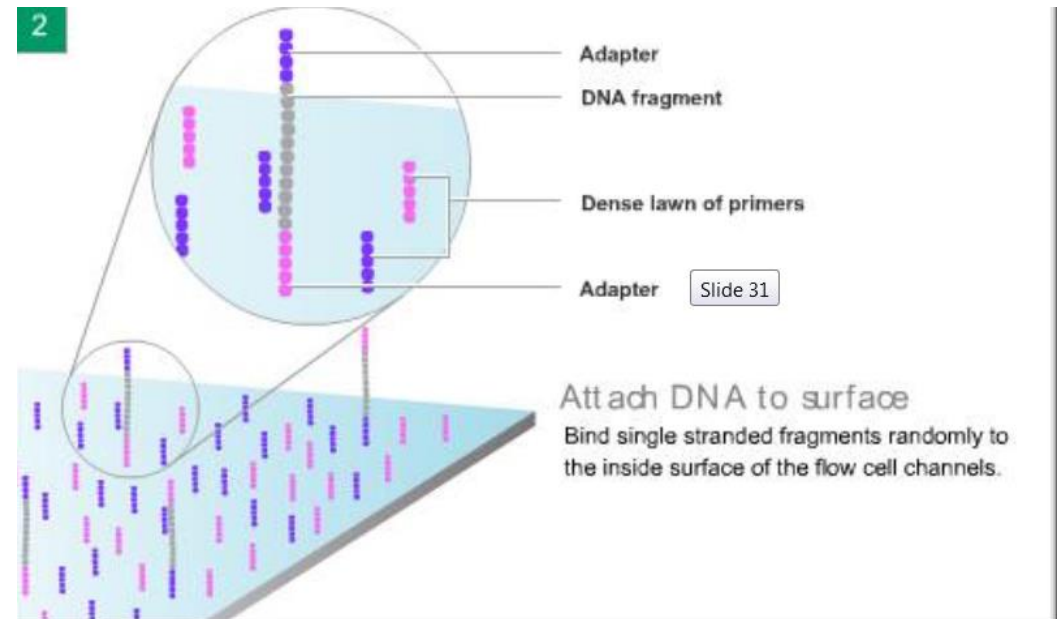
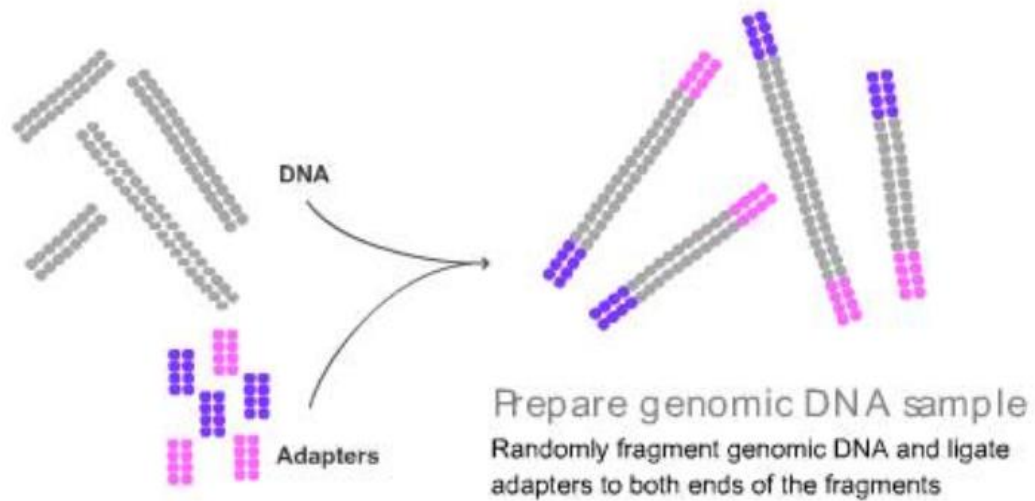
## Amplification

### 1- Emulsion PCR (Roche 454, SOLID, Ion PGM)



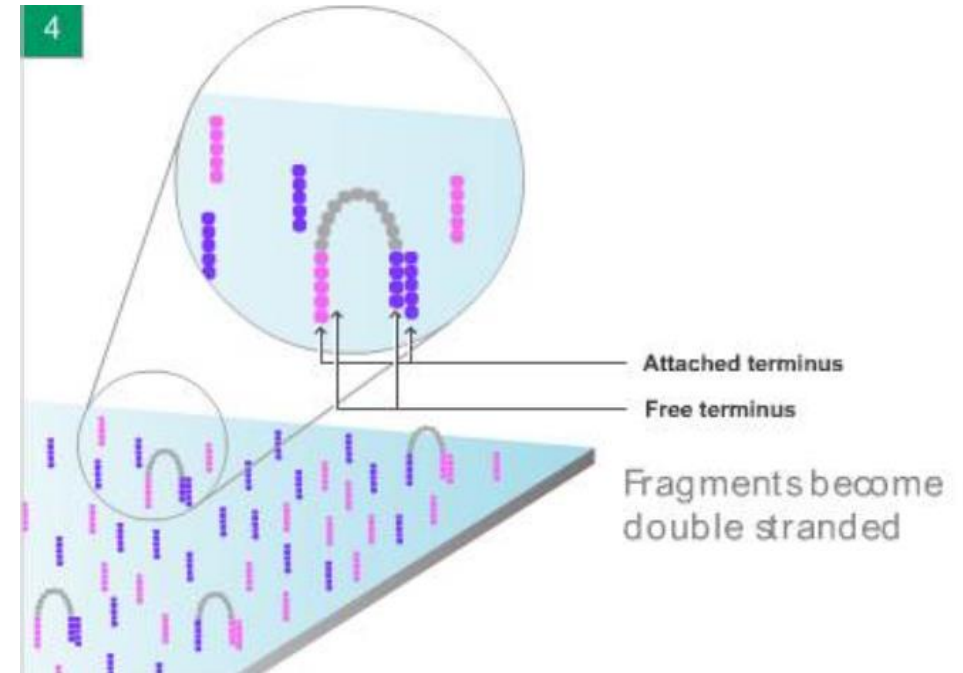
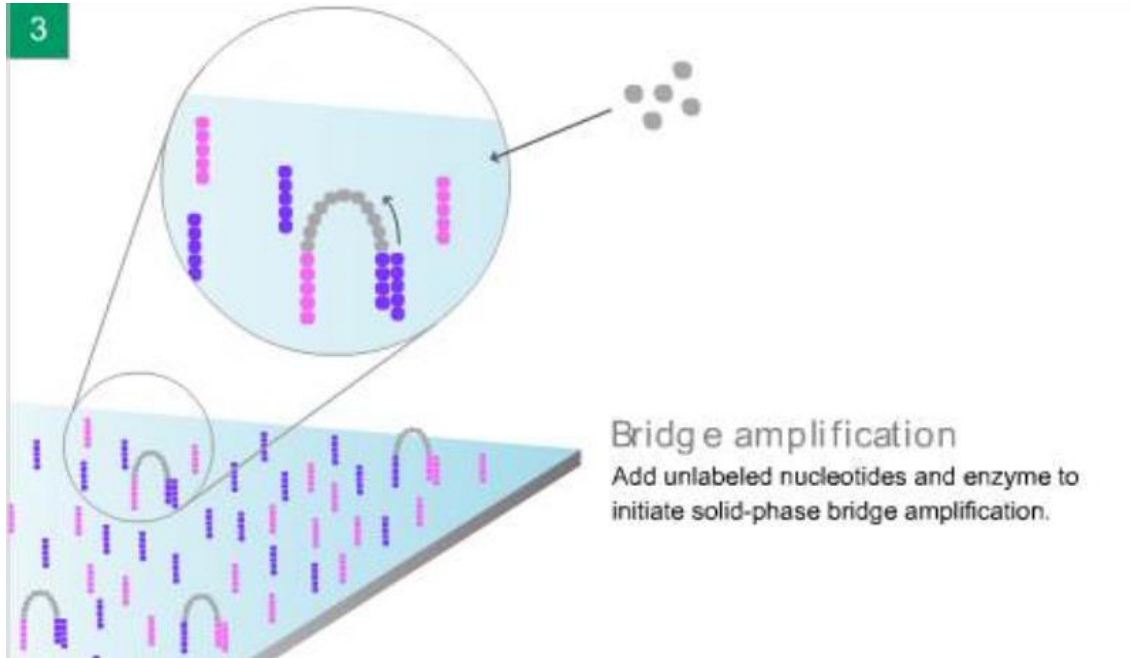
## Amplification

1- immobilisation on solid phase with an oligonucleotide (primer) and bridge amplification (Illumina)



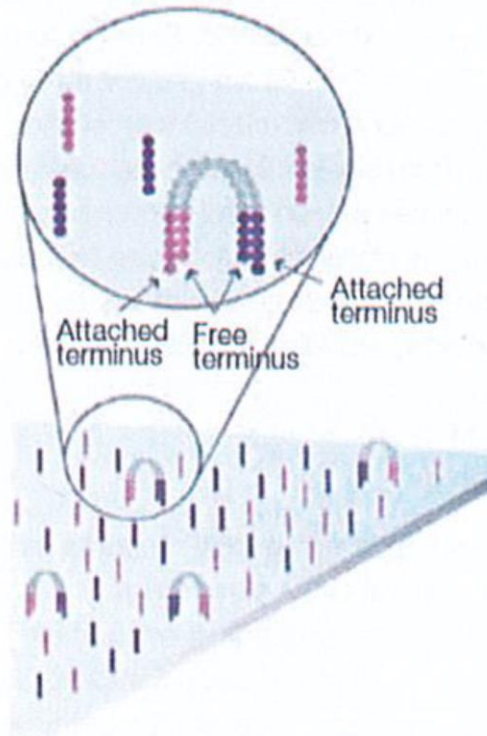
## Amplification

1- immobilisation on solid phase with an oligonucleotide (primer) and bridge amplification (Illumina)

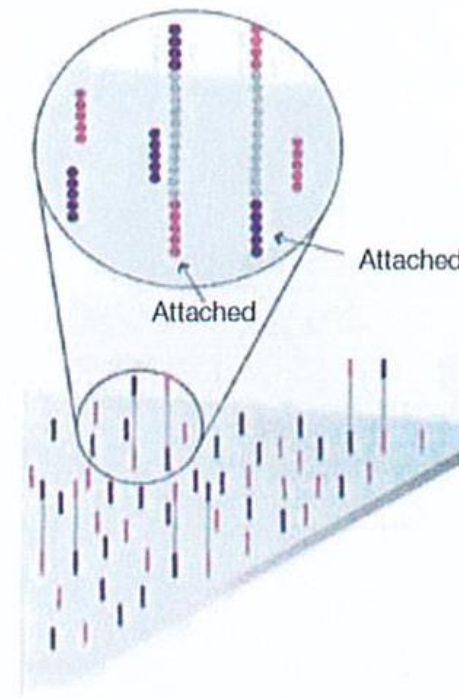


## Amplification

1- immobilisation on solid phase with an oligonucleotide (primer) and bridge amplification (Illumina)



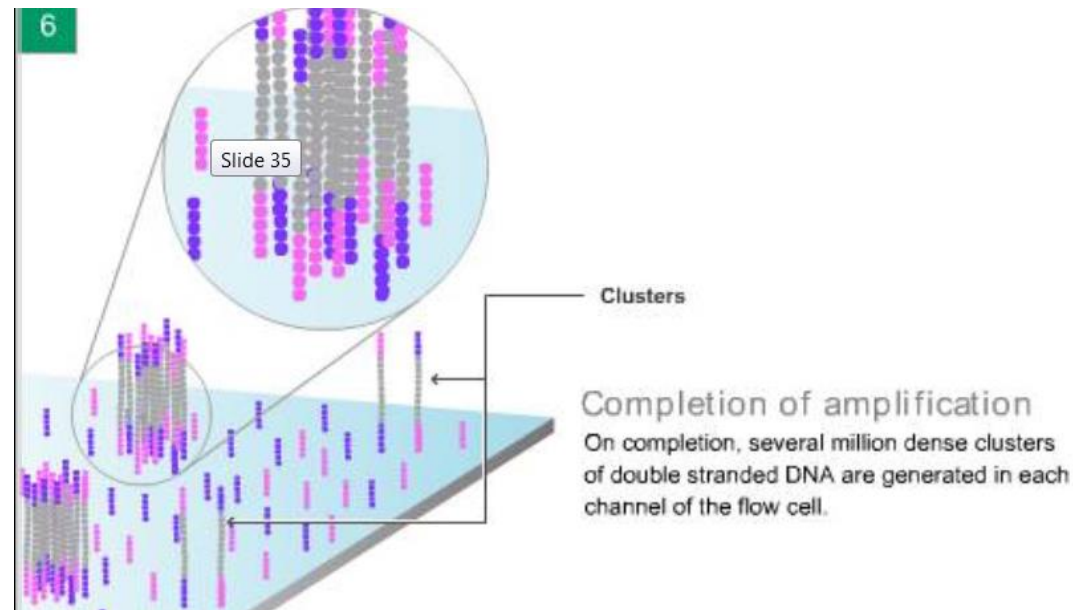
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.



Denaturation leaves single-stranded templates anchored to the substrate.

## Amplification

1- immobilisation on solid phase with an oligonucleotide (primer) and bridge amplification (Illumina)



## Sequencing

Several types of sequencing dependent on the sequencer.

1- Pyrosequencing

**Roche 454**

2- sequencing by ionic detection

**Ion PGM\_Ion Proton\_IonS5 by Ion Torrent / Thermo Fisher**

3- sequencing by ligation

**SOLID**

4- sequencing by four-color reversible termination

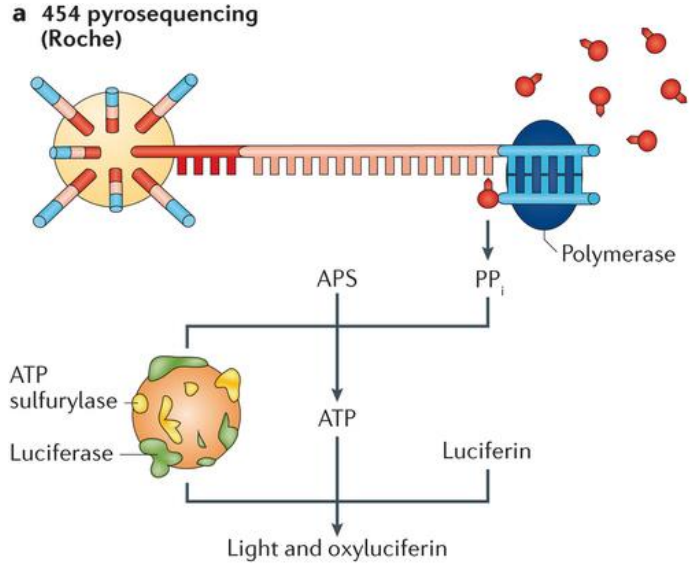
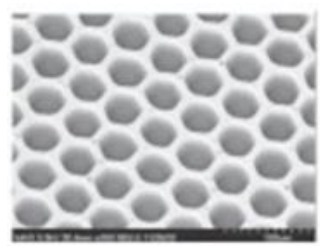
**Illumina**

# 2<sup>nd</sup> Generation sequencing

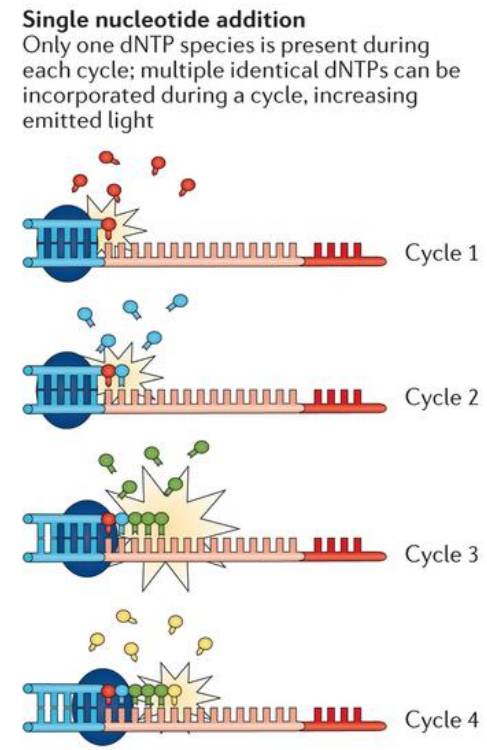
## Sequencing

### 1- Pyrosequencing \_ Roche 454

Platform	Label	Reads per run	Read length (mode or average)	Bases per run (gigabases)
ABI Sanger	ABI Sanger 3730xl	96	800	0.0000768
<b>454</b>	<b>454 GS FLX</b>	<b>1 million</b>	<b>700</b>	<b>0.7</b>



**Pyrosequencing**  
As a base is incorporated, the release of an inorganic pyrophosphate triggers an enzyme cascade, resulting in light

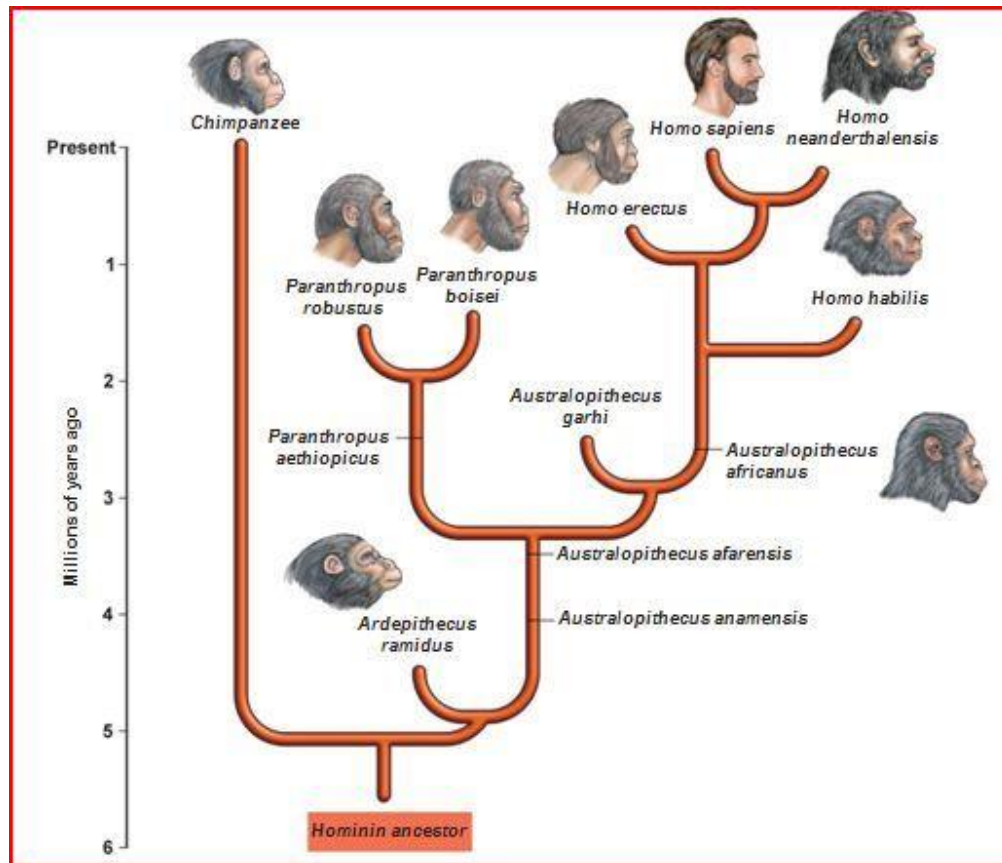


450,000 \$



## Sequencing

### 1- Pyrosequencing \_ Roche 454



#### Analysis of one million base pairs of Neanderthal DNA

*Nature* **444**, 330-336 (16 November 2006) | doi:10.1038/nature05336

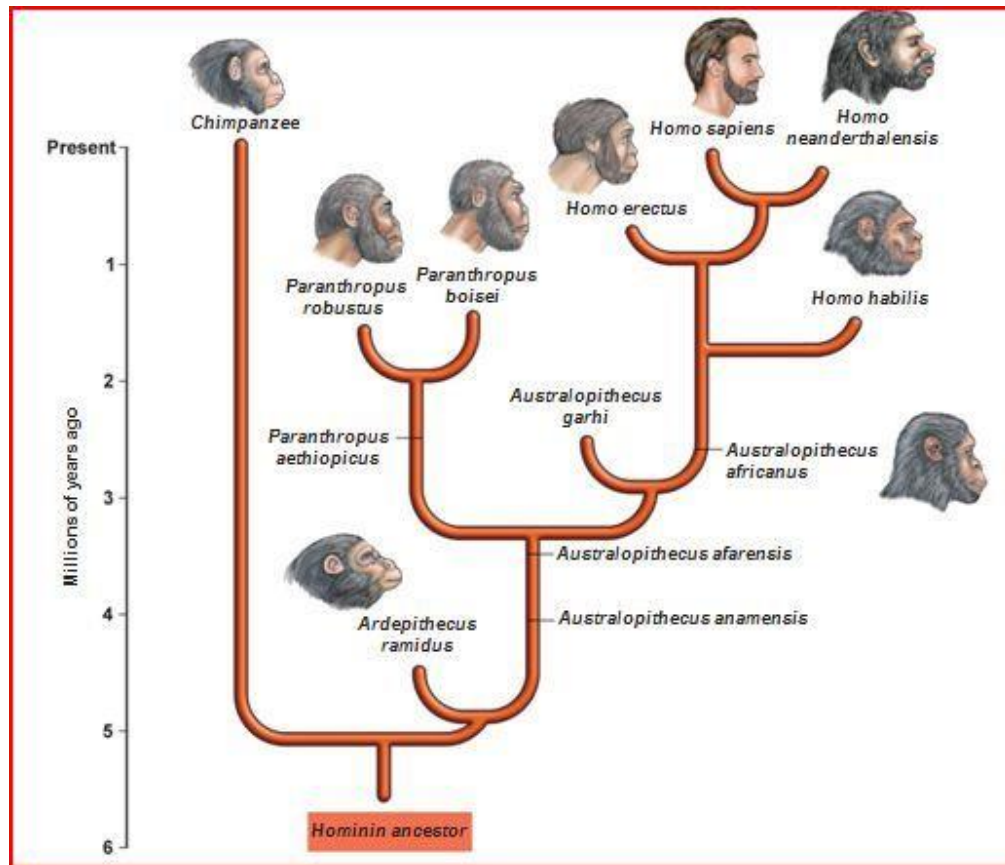
#### A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing

*Cell*. 2008 August 8; 134(3): 416–426 | doi:10.1016/j.cell.2008.06.021



## Sequencing

### 1- Pyrosequencing \_ Roche 454

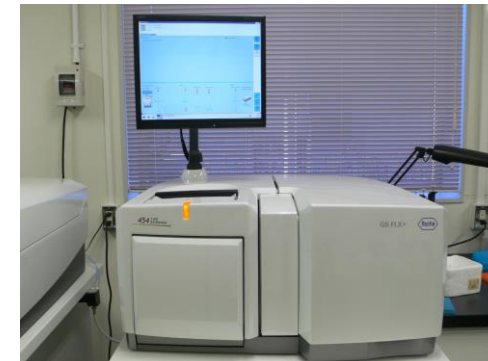


#### Analysis of one million base pairs of Neanderthal DNA

*Nature* **444**, 330-336 (16 November 2006) | doi:10.1038/nature05336

#### A complete Neanderthal mitochondrial genome sequence determined by high-throughput sequencing

*Cell*. 2008 August 8; 134(3): 416–426 | doi:10.1016/j.cell.2008.06.021

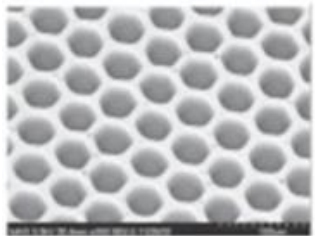


# 2<sup>nd</sup> Generation sequencing

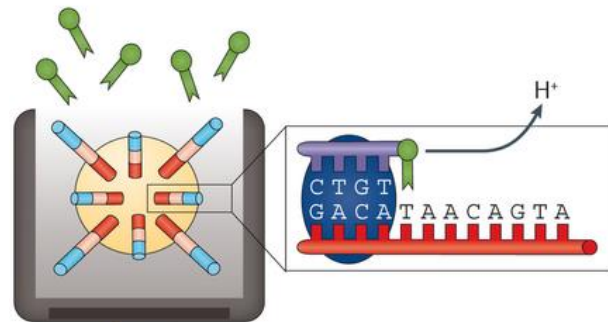
## Sequencing

### 2- Sequencing by ionic detection \_Ion Torrent

Platform	Reads per run	Read length (mode or average)	Bases per run (gigabases)
ABI Sanger	96	800	0.0000768
454	1 millions	700	0.7
<b>Ion Torrent</b>	<b>75 millions</b>	<b>200</b>	<b>15</b>

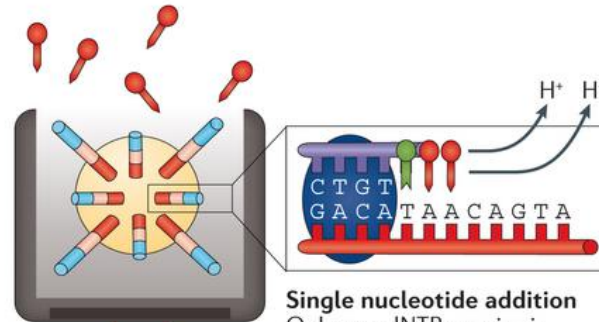
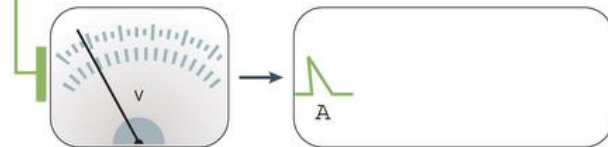


**b** Ion Torrent (Thermo Fisher)



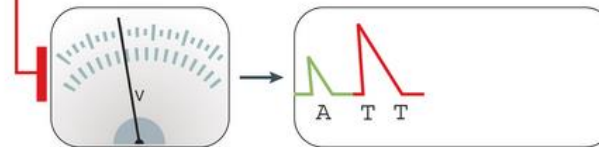
**Semiconductor sequencing**

As a base is incorporated, a single H<sup>+</sup> ion is released, which is detected by a CMOS-ISFET sensor



**Single nucleotide addition**

Only one dNTP species is present during each cycle; several identical dNTPs can be incorporated during a cycle, increasing the emitted ions



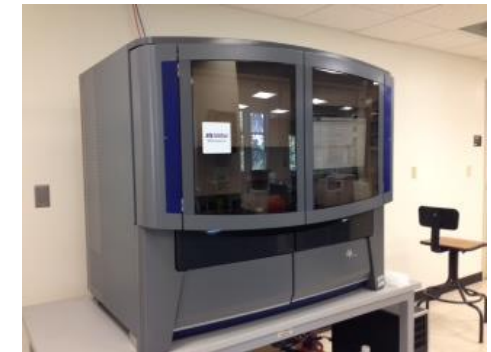
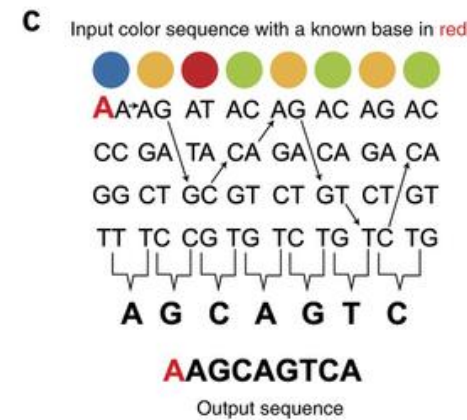
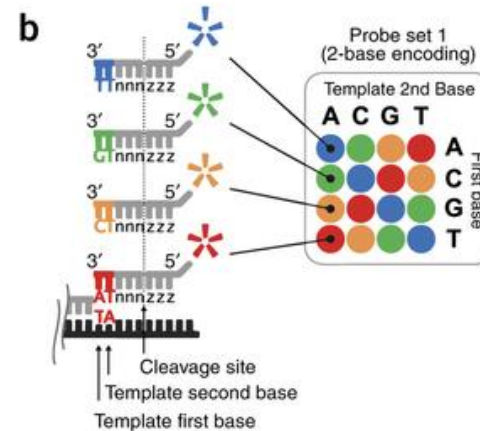
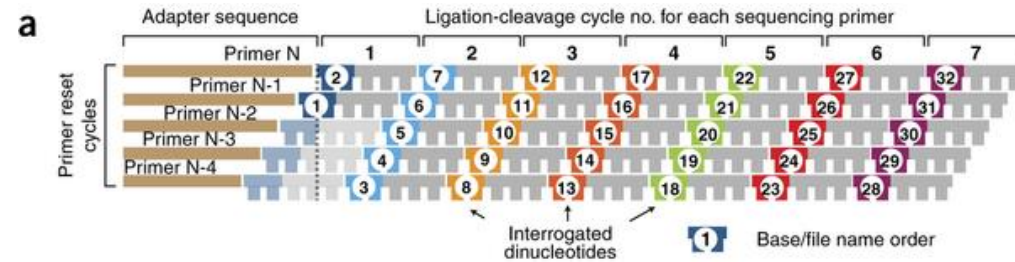
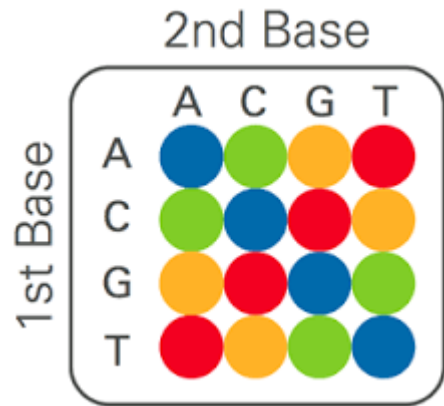
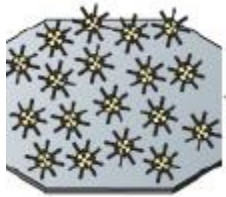
49- 65,000\$

# 2<sup>nd</sup> Generation sequencing

## Sequencing

### 3- Sequencing by ligation

Platform	Reads per run	Read length (mode or average)	Bases per run (gigabases)
ABI Sanger	96	800	0.0000768
454	1 millions	700	0.7
IonTorrent	75 millions	200	15
<b>SOLiD</b>	<b>3 billions</b>	<b>75</b>	<b>320</b>

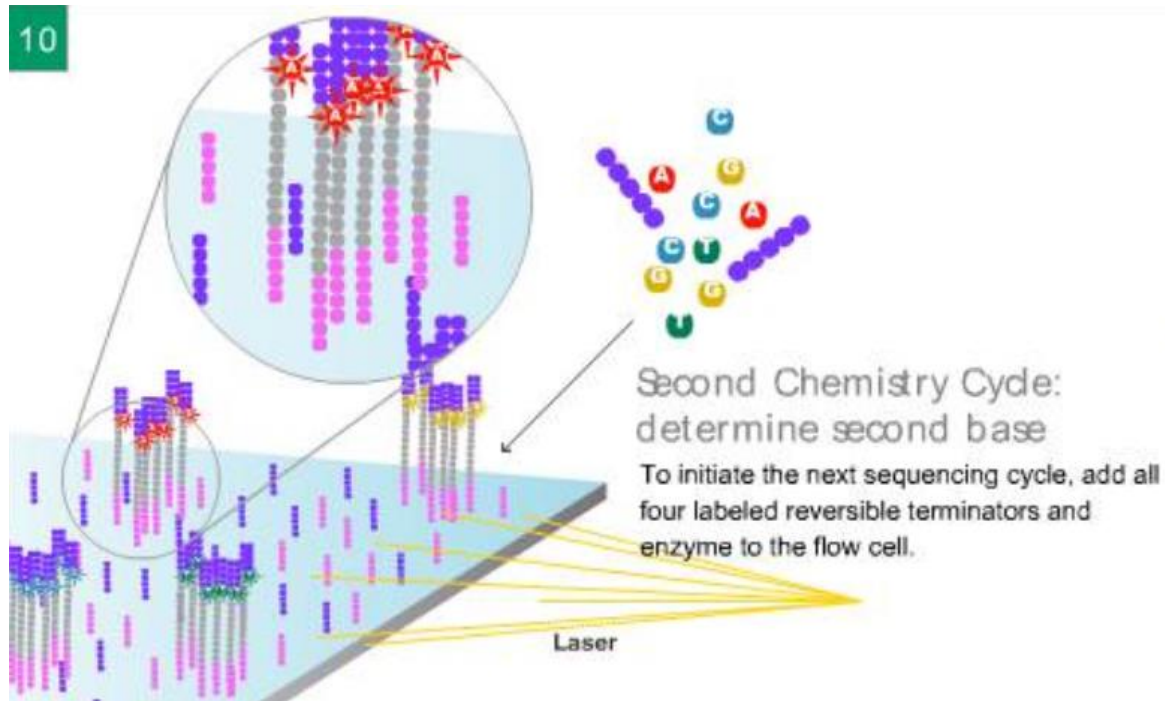


251,000 \$

## Sequencing



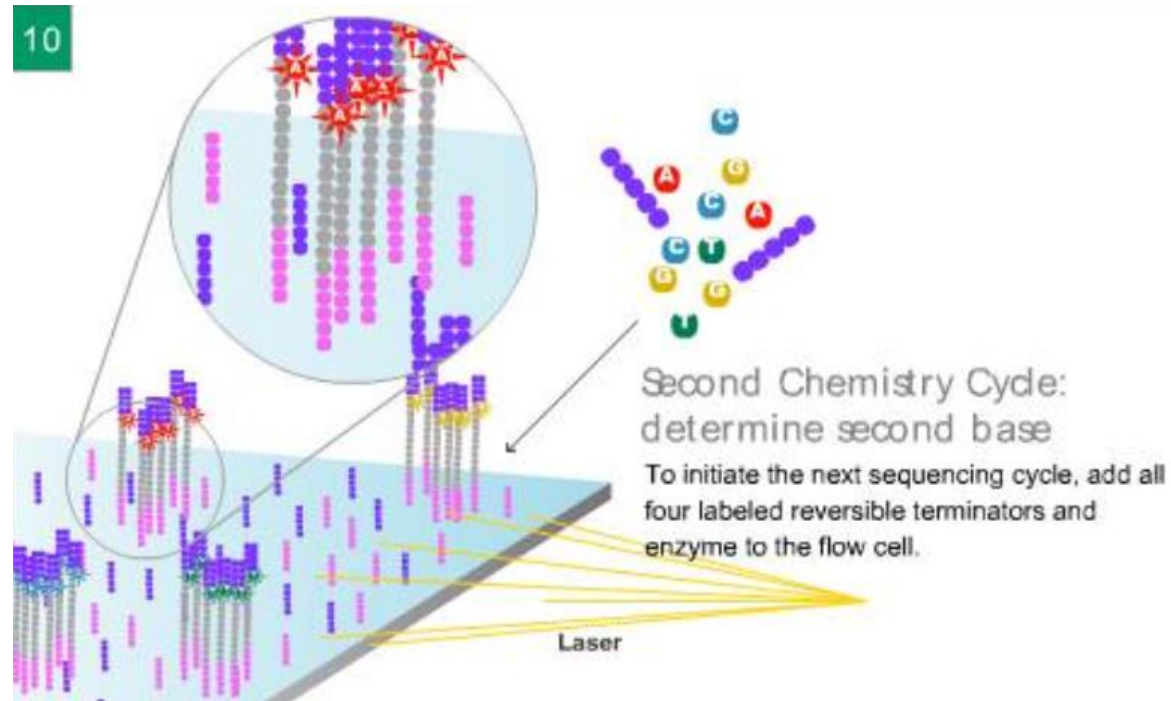
### 3- Sequencing by four-color reversible termination\_Illumina



## Sequencing



### 3- Sequencing by four-color reversible termination\_Illumina





## Sequencing



### 3- Sequencing by four-color reversible termination\_Illumina

9

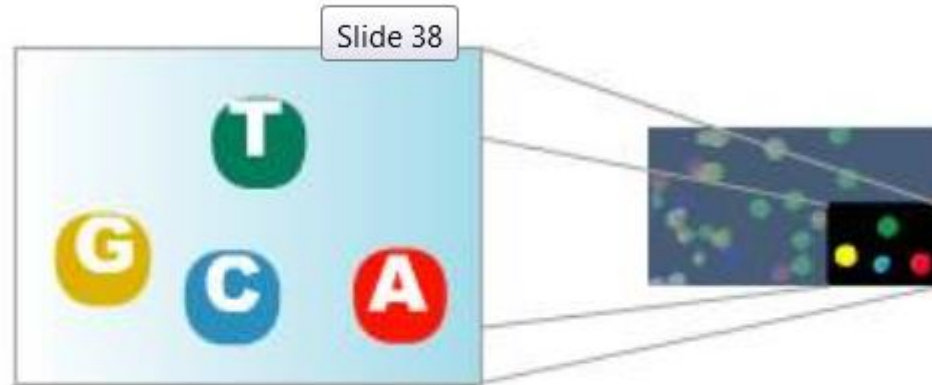


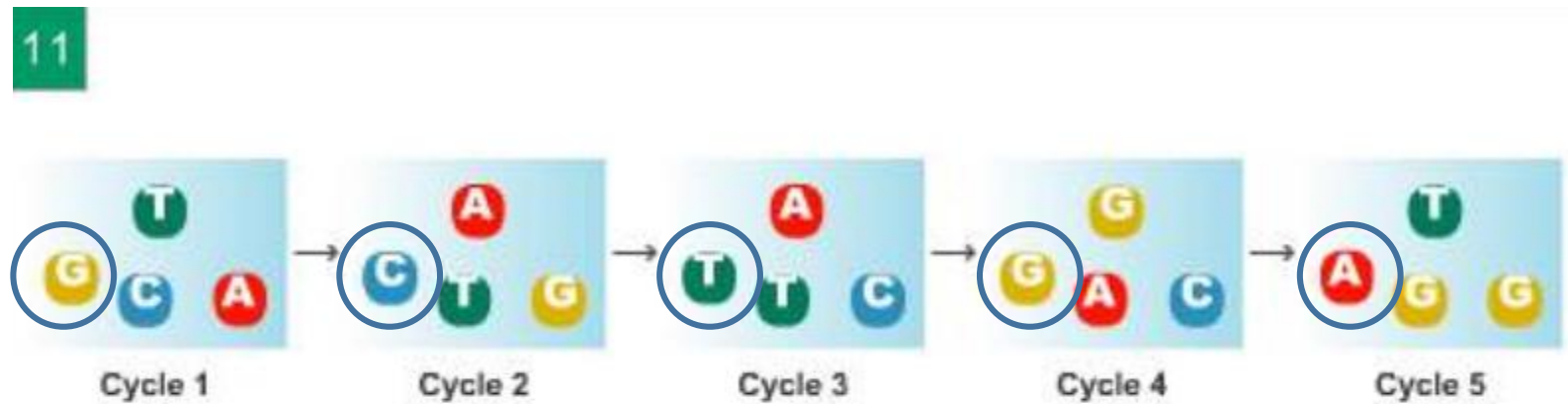
Image of first chemistry cycle  
After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

Before initiating the next chemistry cycle  
The blocked 3' terminus and the fluorophore from each incorporated base are removed.

## Sequencing



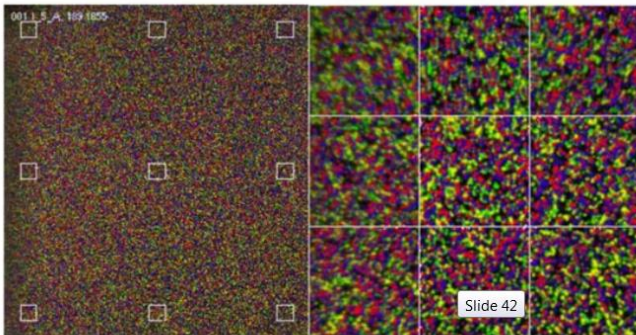
### 3- Sequencing by four-color reversible termination\_Illumina



GCTGA...

Sequence read over multiple chemistry cycles

Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time.



# 2<sup>nd</sup> Generation sequencing

Sequencing  
illumina<sup>®</sup>

Platform	Reads per run	Read length (mode or average)	Bases per run (gigabases)
ABI Sanger	96	800	0.0000768
454	1 millions	700	0.7
IonTorrent	75 millions	200	15
SOLiD	3 billions	75	320
<b>Illumina</b>	<b>600 millions to 6 milliards</b>	<b>100 to 300</b>	<b>7.5 to 2 000</b>



MiniSeq System  
Up to 7.5 Gb



MiSeq Series +  
Up to 15 Gb



NextSeq Series +  
Up to 120 Gb



NextSeq Series +  
Up to 120 Gb



HiSeq Series +  
Up to 750 Gb



HiSeq X Series †  
Up to 800 Gb



NovaSeq Series +  
Up to 2 Tb



# 2<sup>nd</sup> Generation sequencing

## Sequencing

illumina<sup>®</sup>

Platform	Reads per run	Read length (mode or average)	Bases per run (gigabases)
ABI Sanger	96	800	0.0000768
454	1 millions	700	0.7
IonTorrent	75 millions	200	15
SOLiD	3 billions	75	320
<b>Illumina</b>	<b>600 millions to 6 milliards</b>	<b>100 to 300</b>	<b>7.5 to 2 000</b>



MiniSeq System  
50,000 \$



MiSeq Series +  
99,000 \$



NextSeq Series +  
250,000 \$



NextSeq Series +  
250,000 \$



HiSeq Series +  
690,000 \$-  
900,000 \$



HiSeq X Series †  
1,000,000 \$  
(X10= 10,000,000 \$)



NovaSeq Series +  
850,000 \$-  
985,000 \$

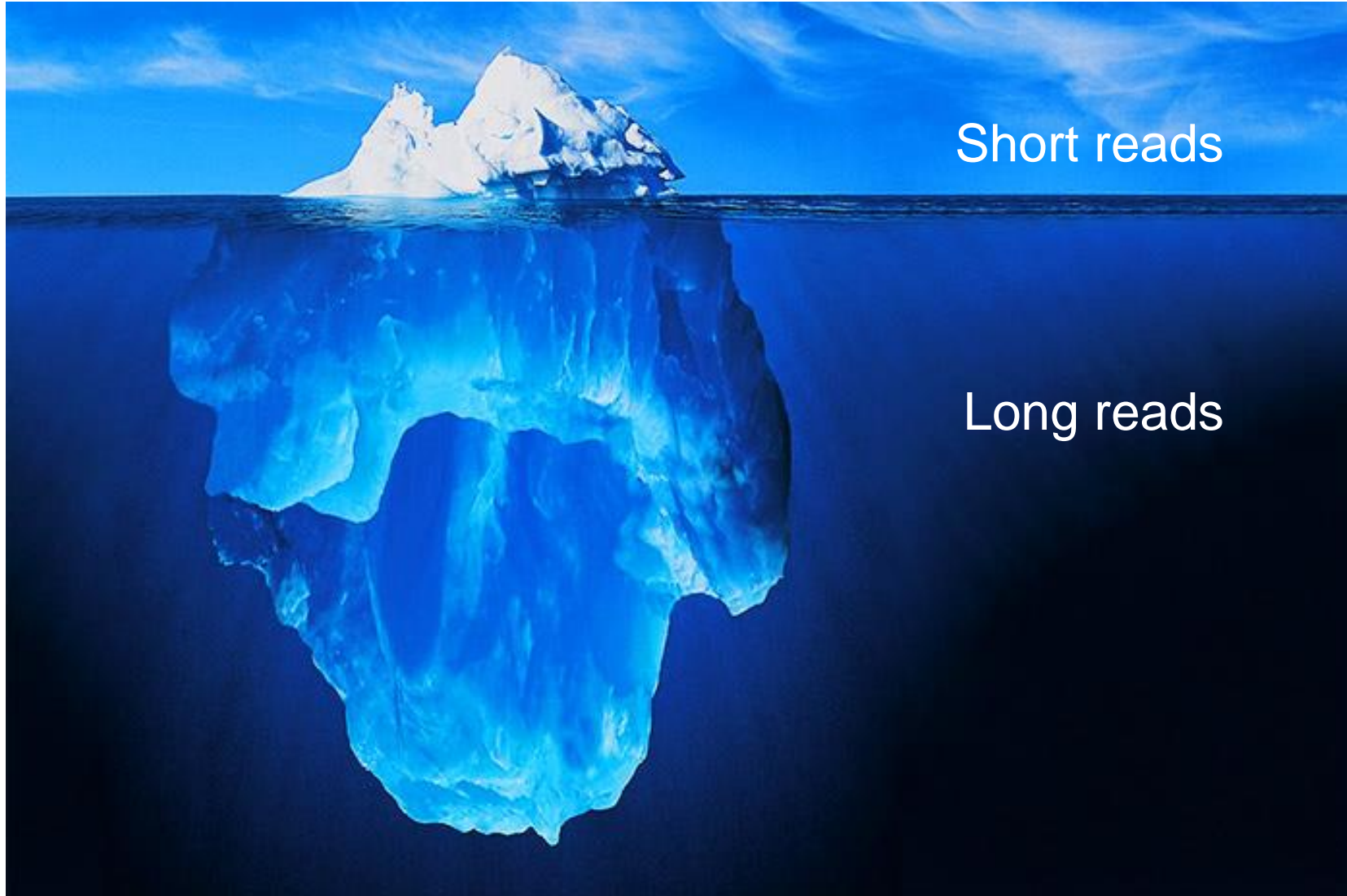
## +Advantages

- High throughput technology
- Very good quality of the sequences, error rate < 0.1 %

## -Limitations

- Process of reverse transcription
- Secondary structure of RNA
- PCR-biais
- Sequences with high AT do not amplify as well as GC-normal one
- Short reads

# 2<sup>nd</sup> Generation sequencing



# 3<sup>rd</sup> Generation sequencing

Since 2012, a new cohort of techniques has been developed:

- using single molecule sequencing
- single real time sequencing
- removing the need for clonal amplification.

This :

- reduces errors caused by PCR
- simplifies library preparation
- and, most importantly, gives a much higher read length



## Single molecule Long Read Sequencing



PACIFIC  
BIOSCIENCES®



Oxford  
**NANOPORE**  
Technologies®

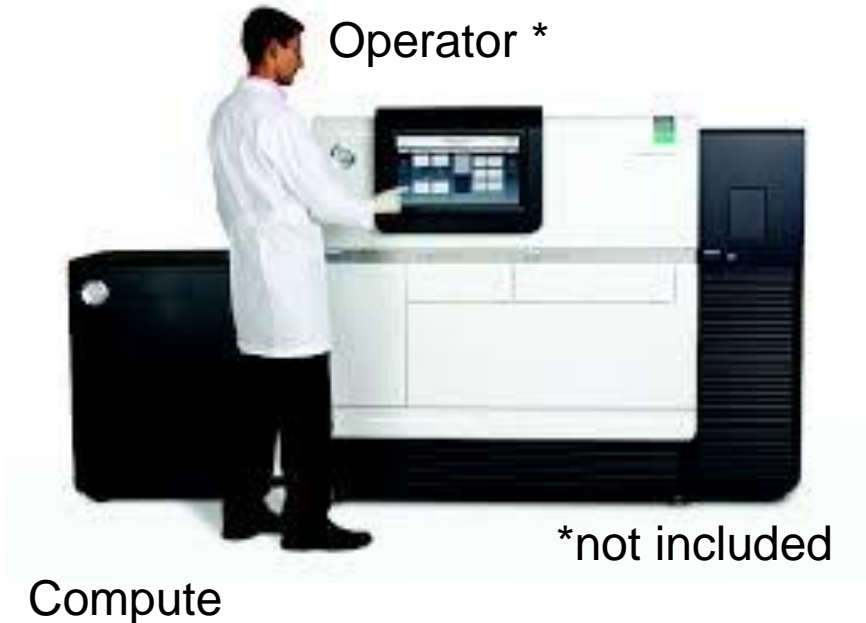
# 3<sup>rd</sup> Generation sequencing

## Single molecule Long Read Sequencing



Platform	Year	Reads per run	Read length (mode or average)	Bases per run (gigabases)
ABI Sanger	2002	96	800	0.0000768
454	2011	1 millions	700	0.7
SOLiD	2013	3 milliards	75	320
IonTorrent	2015	75000000	200	15
Illumina	2016	600 millions to 6 milliards	100 to 300	7.5 to 2 000
<b>PacBio</b>	<b>2014</b>	<b>660000</b>	<b>13500</b>	<b>20</b>

RSII 700,000 \$

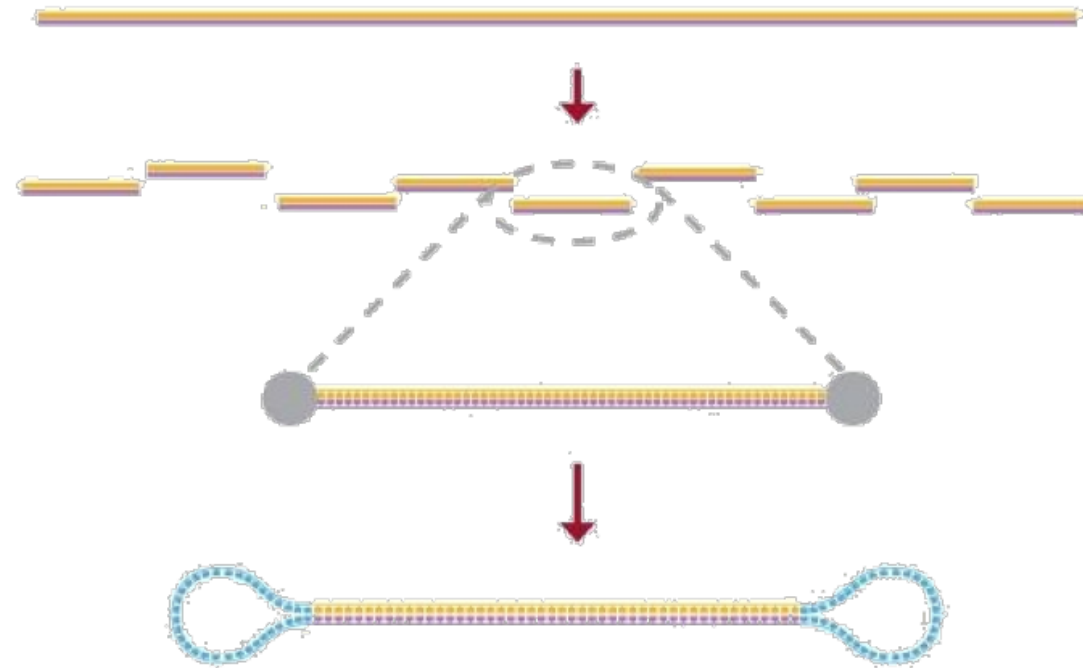


Sequel 350,000 \$



# 3<sup>rd</sup> Generation sequencing

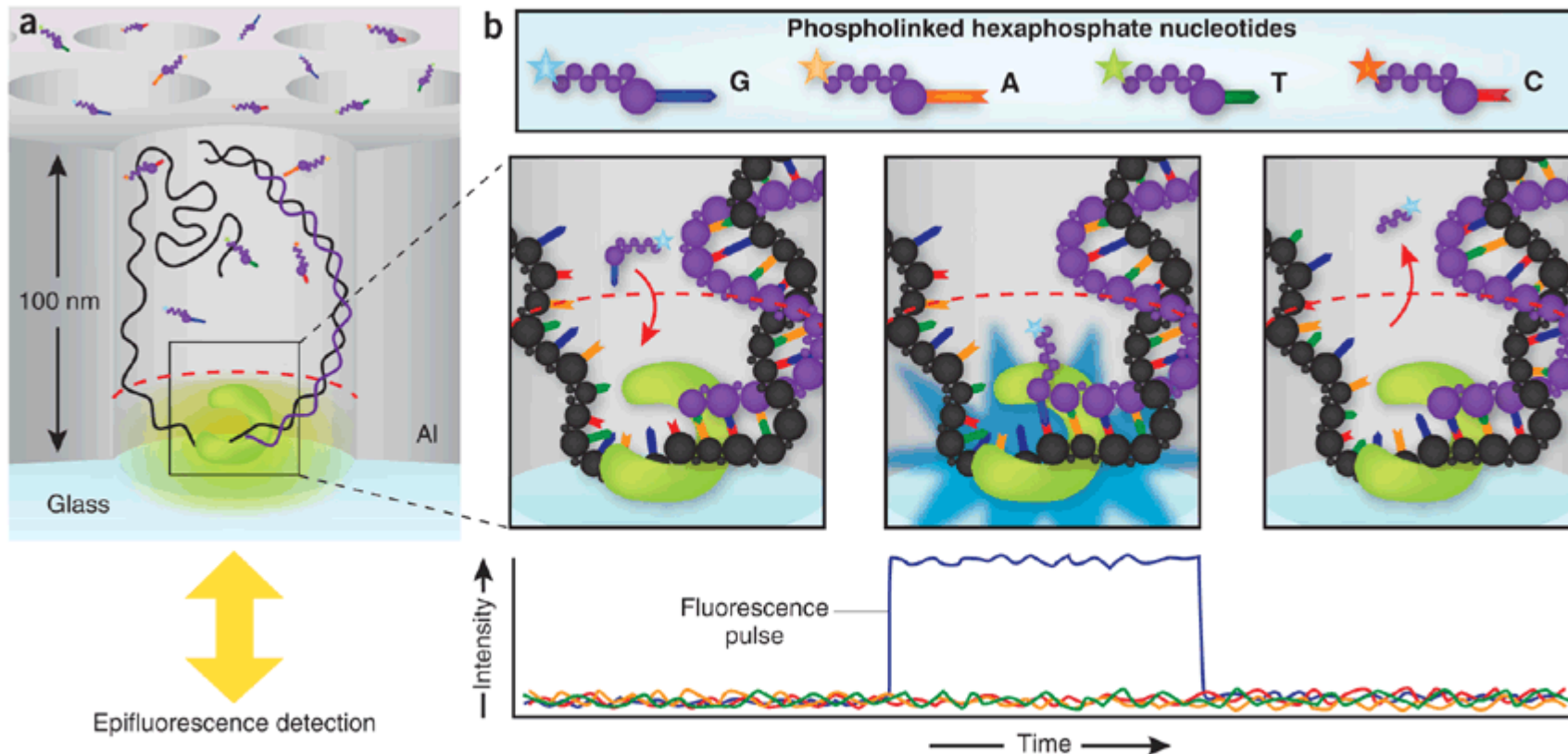
## Single molecule Long Read Sequencing



**SMRTbell™** Template preparation can be used to create libraries of various insert sizes from 250 bp to 20,000 bp depending on the needs of the application.

# 3<sup>rd</sup> Generation sequencing

## Single molecule Long Read Sequencing





# 3<sup>rd</sup> Generation sequencing

## Single molecule Long Read Sequencing



Application: **Whole genome**



Jarvis et al. (2017) Nature



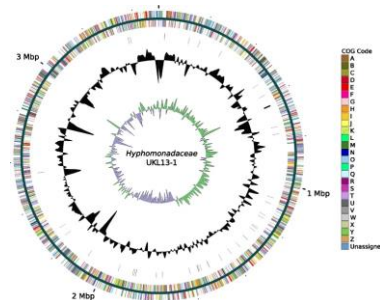
Bickhart et al. (2017) Nature genetics



Brozynska et al. (2016) Plant Biotechnology Journal



Seo et al. (2016) Nature



Driscoll et al. (2017) Standards in Genomic Sciences



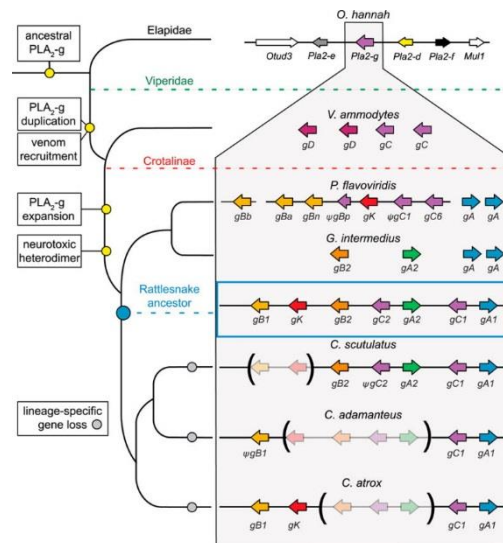
INRA TOULOUSE/SUNRISE PROJECT

## Single molecule Long Read Sequencing



### Application: Targeted sequencing

to fully characterize genetic complexity — including structural variation, rare SNPs, indels, copy number variation, microsatellites, haplotypes, and phasing



Evolution of PLA2 related toxins



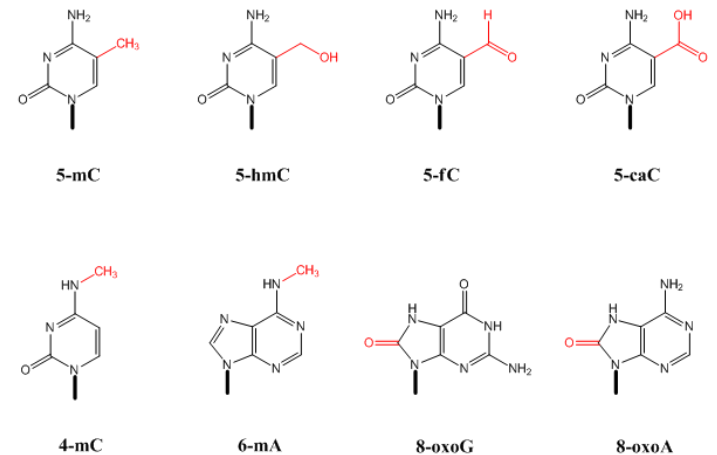
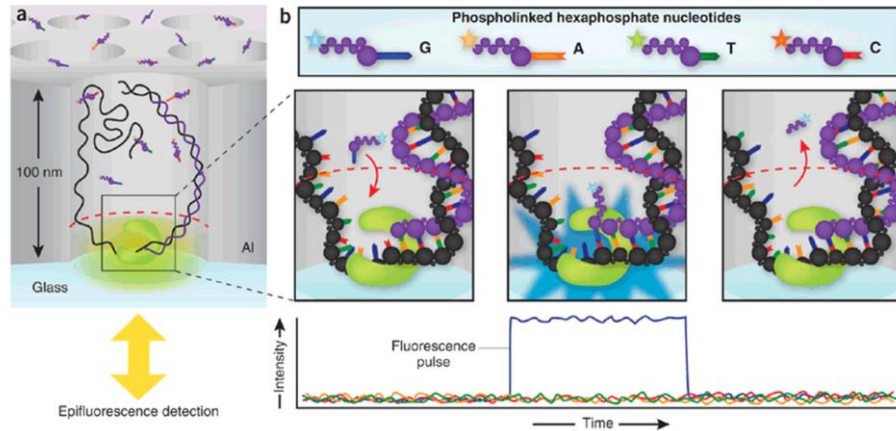
Dowell et al. (2016) Current Biology

## Single molecule Long Read Sequencing



### Application: Epigenetics

Single Molecule, Real-Time (SMRT) Sequencing directly detects epigenetic modifications by measuring kinetic variation during base incorporation. By capturing these modifications simultaneously with sequence data, this method eliminates the need for special sample preparation and additional sequencing.

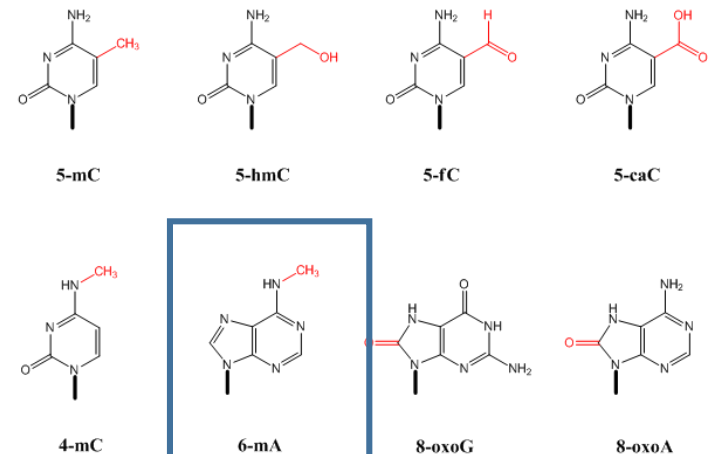
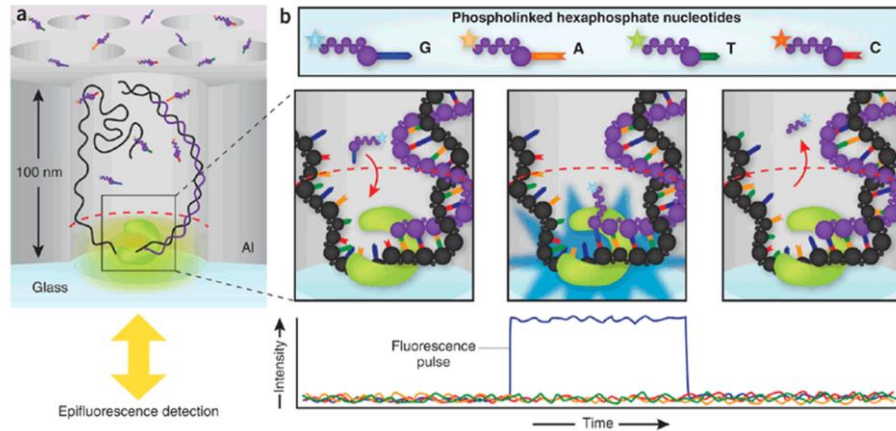


## Single molecule Long Read Sequencing



### Application: Epigenetics

Single Molecule, Real-Time (SMRT) Sequencing directly detects epigenetic modifications by measuring kinetic variation during base incorporation. By capturing these modifications simultaneously with sequence data, this method eliminates the need for special sample preparation and additional sequencing.



Role in gene activation

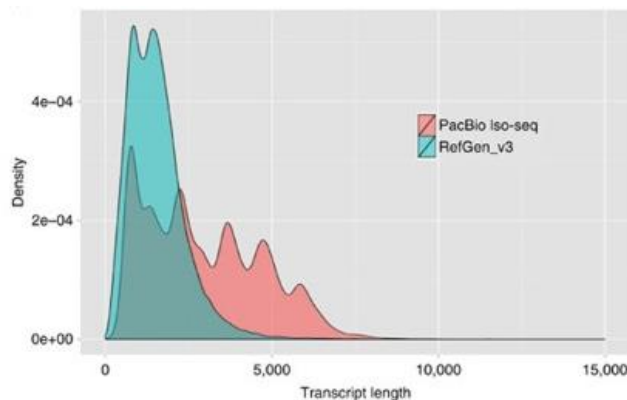
Wu et al. (2016) Nature

## Single molecule Long Read Sequencing



### Application: RNA Sequencing

- The isoform sequencing (Iso-Seq) application generates full-length cDNA sequences — from the 5' end of transcripts to the poly-A tail — transcriptome reconstruction
- The Iso-Seq method generates information without about alternatively spliced exons and transcriptional start sites. It also delivers information about poly-adenylation sites for transcripts up to 10 kb.



Wang et al. (2016) Nature Communications



Workman et al. (2017) *bioRxiv*



# 3<sup>rd</sup> Generation sequencing

## Single molecule Long Read Sequencing



Platform	Reads per run	Read length (mode or average)	Bases per run (gigabases)
ABI Sanger	96	800	0.0000768
454	1 millions	700	0.7
SOLiD	3 milliards	75	320
IonTorrent	75000000	200	15
Illumina	600 millions to 6 milliards	100 to 300	7.5 to 2 000
PacBio	660000	13500	12.000
<b>Oxford Nanopore</b>	<b>4400000</b>	<b>9545</b>	<b>42</b>



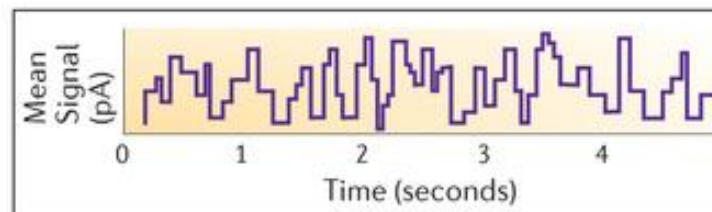
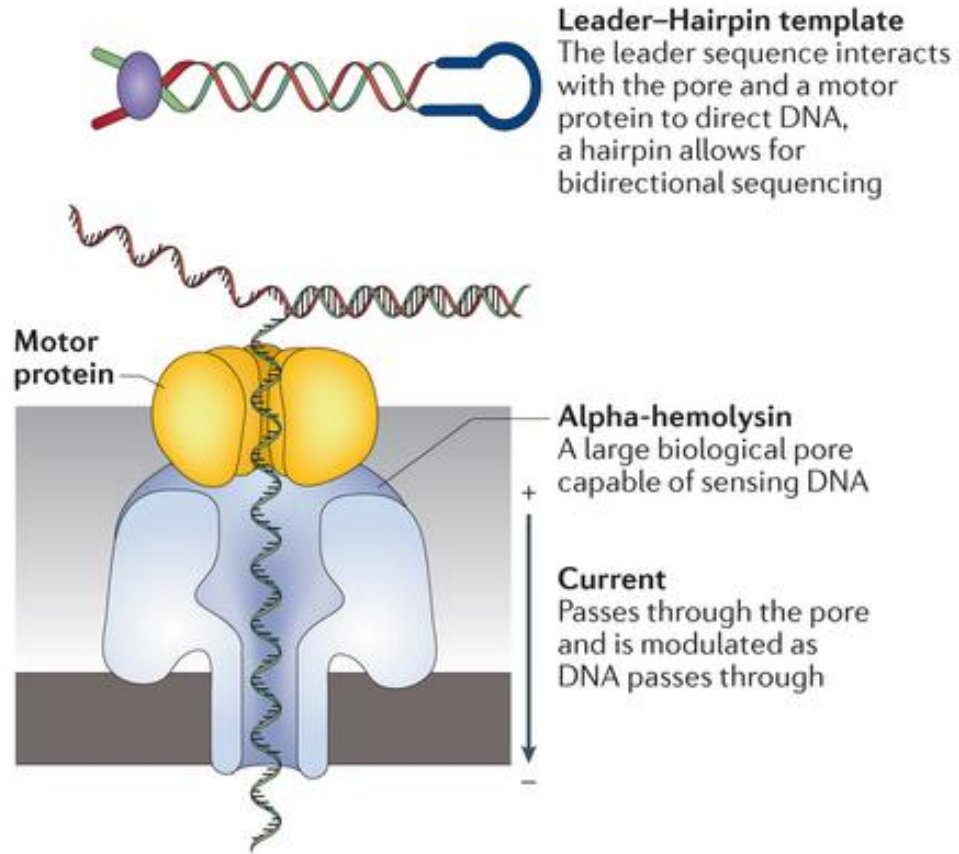
MinION Mk1: portable, real time biological analyses

MinION 1,000 \$



# 3<sup>rd</sup> Generation sequencing

## Single molecule Long Read Sequencing



**ONT output (squiggles)**  
Each current shift as DNA translocates through the pore corresponds to a particular k-mer



# 3<sup>rd</sup> Generation sequencing

## Single molecule Long Read Sequencing

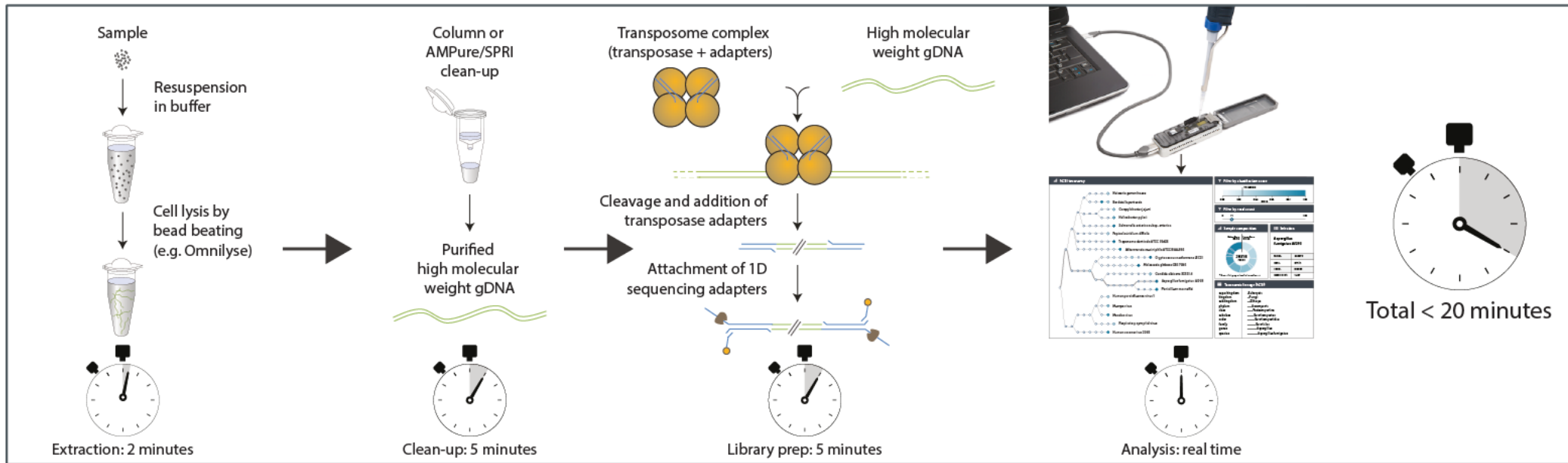


Fig. 1 Identification of bacterial, archaeal, fungal and viral species using rapid gDNA library preparation and the real-time WIMP analysis workflow











# 3<sup>rd</sup> Generation sequencing

## Single molecule Long Read Sequencing



The screenshot shows the Oxford Nanopore Technologies website with a dark teal header. The header contains the company logo and navigation links: PRODUCTS, HOW IT WORKS, APPLICATIONS (highlighted), GET STARTED, and PUBLICATIONS. Below the header, a light grey section features the text "Nanopore sequencing offers advantages in all areas of research..." followed by a grid of 10 research application icons and their corresponding labels.

Nanopore sequencing offers advantages in all areas of research...

 Pathogens / Microbiology / Antimicrobial resistance	 Environmental research	 Microbiome	 Basic genome research	 Human genetics	 Cancer research
 Clinical research	 Plant research	 Transcriptome analysis	 Population genomics		

# 3<sup>rd</sup> Generation sequencing

## Single molecule Long Read Sequencing



Application: **Microbiome**



# 3<sup>rd</sup> Generation sequencing

## Single molecule Long Read Sequencing



Application: **Microbiome**



Bork et al., (2015) Science



# 3<sup>rd</sup> Generation sequencing

## Single molecule Long Read Sequencing



### Application: RNA Sequencing

**Nanopore technology is the only available sequencing technology which can sequence RNA directly**

A screenshot of a bioRxiv preprint page. The page header includes the Cold Spring Harbor Laboratory logo and the bioRxiv logo with the tagline "THE PREPRINT SERVER FOR BIOLOGY". Navigation links for HOME, ABOUT, SUBMIT, ALERTS / RSS, and CHANNELS are visible. A search bar is present. The main content area shows the title "Highly parallel direct RNA sequencing on an array of nanopores" and the authors: Daniel R. Garalde, Elizabeth A. Snell, Daniel Jachimowicz, Andrew J. Heron, Mark Bruce, Joseph Lloyd, Anthony Warland, Nadia Pantic, Tigris Admassu, Jonah Ciccone, Sabrina Serra, Jemma Keenan, Samuel Martin, Luke McNeill, Jayne Wallace, Lakmal Jayasinghe, Chris Wright, Javier Blasco, Botond Sipos, Stephen Young, Sisse Juul, James Clarke, Daniel J. Turner. The DOI is https://doi.org/10.1101/068809. Below the title and authors, there are buttons for "Download PDF", "Email", "Share", and "Citation Tools". There are also social media sharing buttons for Tweet, Facebook, and Google+. The "Subject Area" section is set to "Genomics". A list of "Subject Areas" is shown, including Animal Behavior and Cognition, Biochemistry, Bioengineering, Bioinformatics, Biophysics, and Cancer Biology. The "Abstract" section is partially visible, starting with "Ribonucleic acid sequencing can allow us to monitor the RNAs present in a sample. This enables us to detect the presence and nucleotide sequence of viruses, or to build a picture of how active transcriptional processes are changing -- information that is useful for understanding the status and function of a sample. Nanopore-based sequencing technology is capable of electronically analysing a sample's DNA directly, and in real-time. In this manuscript we demonstrate the ability of an array of nanopores to sequence RNA directly, and we apply it to a range of biological situations. Nanopore technology is the only available sequencing technology which..."

Garalde et al., (2016) bioRxiv



# 3<sup>rd</sup> Generation sequencing

## Single molecule Long Read Sequencing



Next....

GridION



MinION X5

PromethION



MinION X48

SmidgION



Always more....

**RNA Sequencing provides insight into :**

- **gene expression analysis**
- **discovery of novel transcripts**
- **identification of alternatively spliced genes**
- **detection of allele specific expression**
- **de novo RNA sequencing (without reference transcriptome)**
- **single cell RNA sequencing**

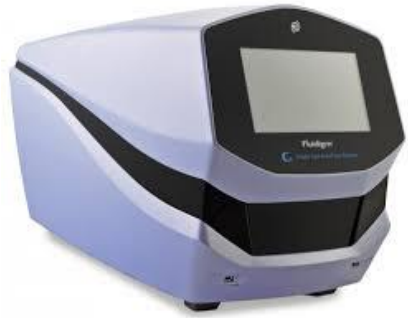
**In addition to polyadenylated messenger RNA (mRNA) Transcripts , analysis of different population of RNAs :**

- **Total RNA**
- **Pre-mRNA**
- **Non coding RNA : micro RNA, long ncRNA**
- **tRNA sequencing**

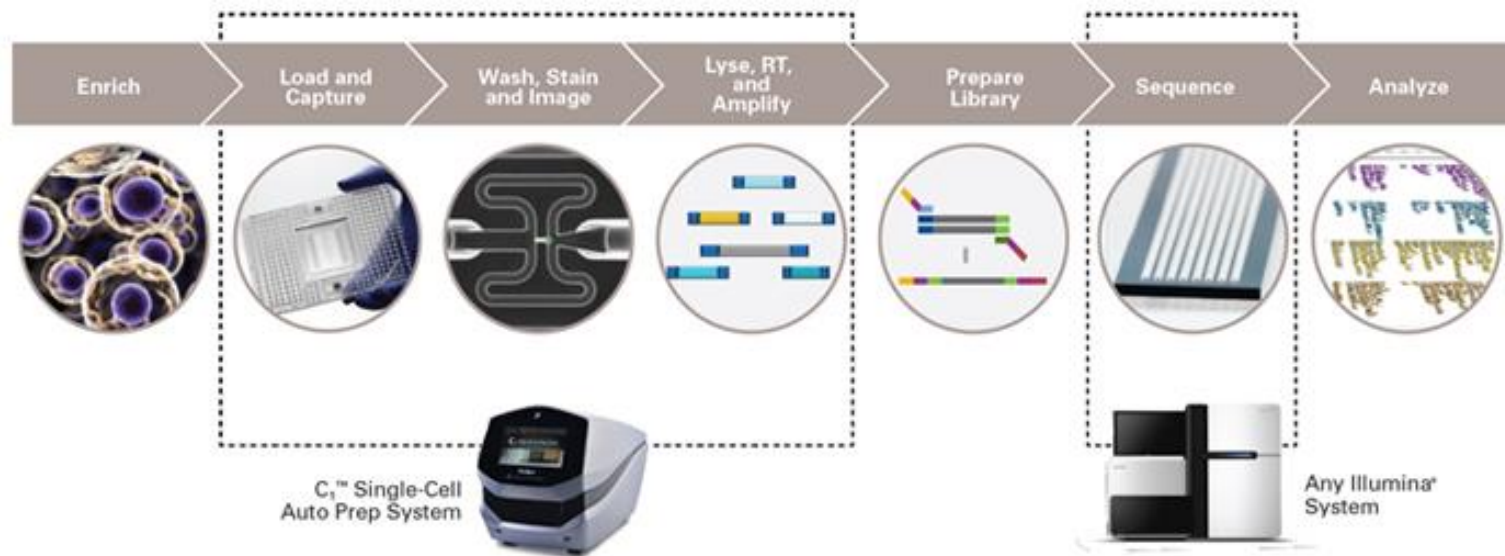


# Application of RNA sequencing

## Single Cell RNA-Seq

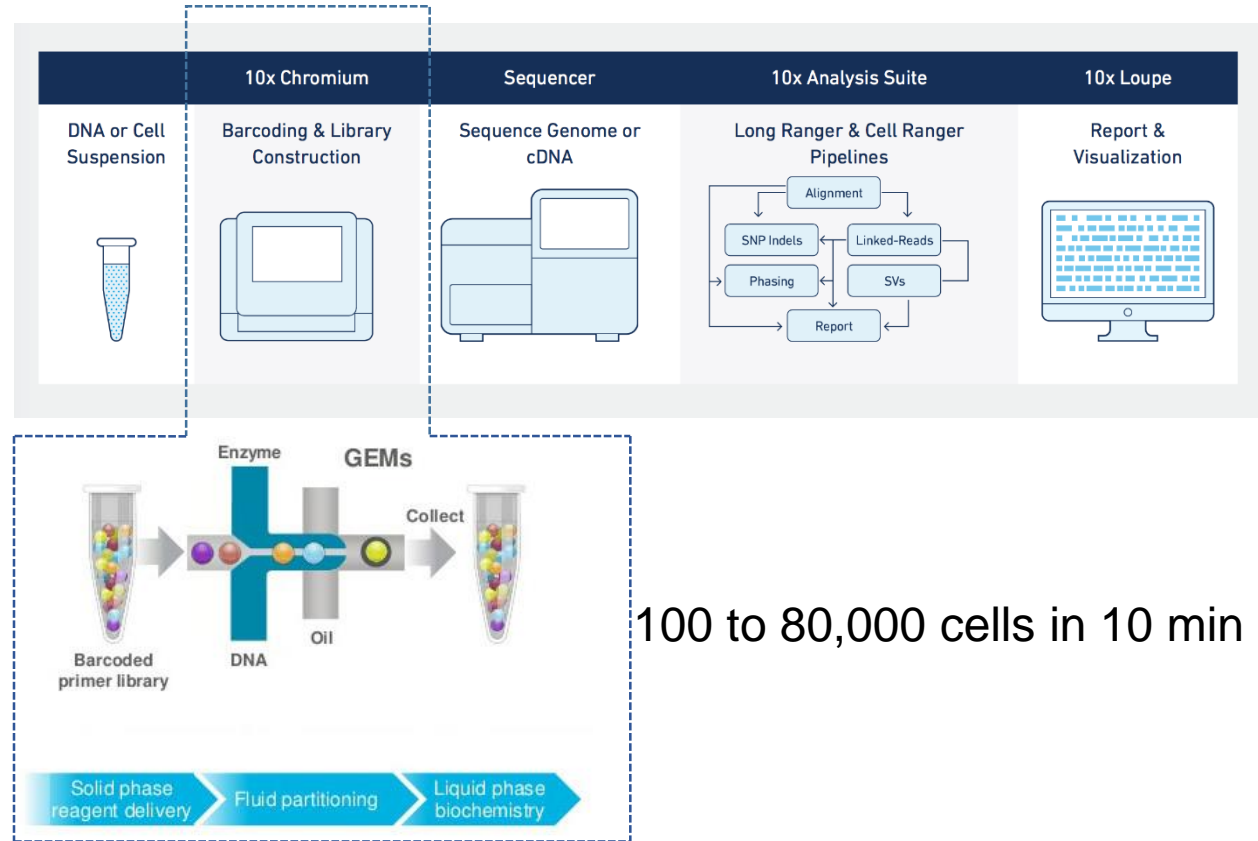


C<sub>1</sub>™ Single-Cell Auto Prep System  
96 cells in 5 min



# Application of RNA sequencing

## Single Cell RNA-Seq



# Application of RNA sequencing

## tRNA Sequencing

### BRIEF COMMUNICATIONS

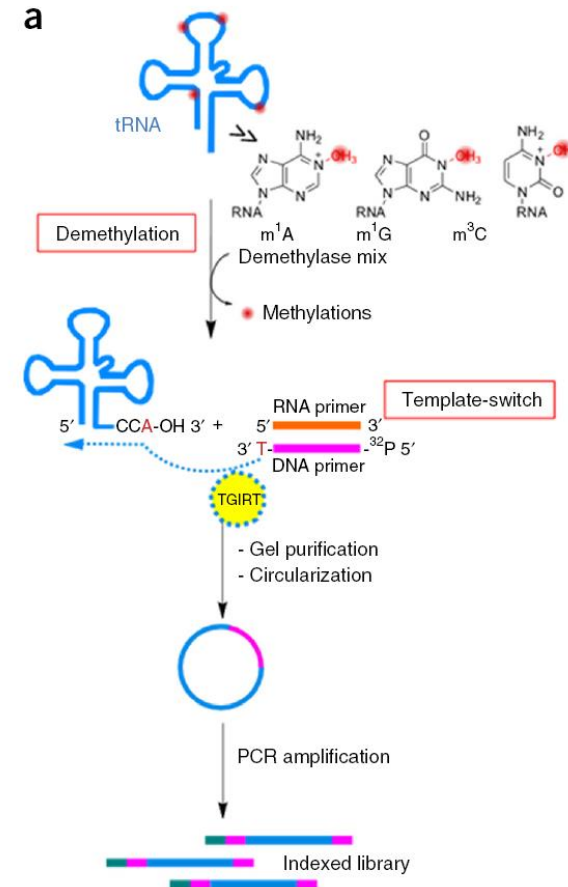
## Efficient and quantitative high-throughput tRNA sequencing

Guanqun Zheng<sup>1,7</sup>, Yidan Qin<sup>2,7</sup>, Wesley C Clark<sup>1</sup>, Qing Dai<sup>3</sup>, Chengqi Yi<sup>1,6</sup>, Chuan He<sup>1,3-5</sup>, Alan M Lambowitz<sup>2,8</sup> & Tao Pan<sup>1,4,8</sup>

Despite its biological importance, tRNA has not been adequately sequenced by standard methods because of its abundant post-transcriptional modifications and stable structure, which

and are particularly problematic for reverse transcriptases (RTs), causing cDNA synthesis to stop or incorporate a wrong nucleotide. In mammals, *N*<sup>1</sup>-methyladenosine (m<sup>1</sup>A) is present in all tRNAs at position 58, *N*<sup>3</sup>-methylcytosine (m<sup>3</sup>C) is present in five tRNAs at position 32 and the variable loop, and *N*<sup>1</sup>-methylguanosine (m<sup>1</sup>G) is present in about half of all tRNAs at position 37 or 9. We applied two recombinant enzymes as a mixture to remove these three methylations in human tRNAs. The first was the wild-type enzyme AlkB from *Escherichia coli*, which is known to efficiently demethylate m<sup>1</sup>A and m<sup>3</sup>C in single-stranded nucleic acids as its DNA and RNA repair function<sup>5,6</sup>. Wild-type AlkB, however, works very poorly on m<sup>1</sup>G modification<sup>7</sup>. On the basis of its known three-dimensional structure complexed with nucleic acids, we engineered AlkB to generate a specific mutant, D135S, that

Zheng et al., (2015) Nature Methods



# Application of RNA sequencing

## Dual RNA-Sequencing

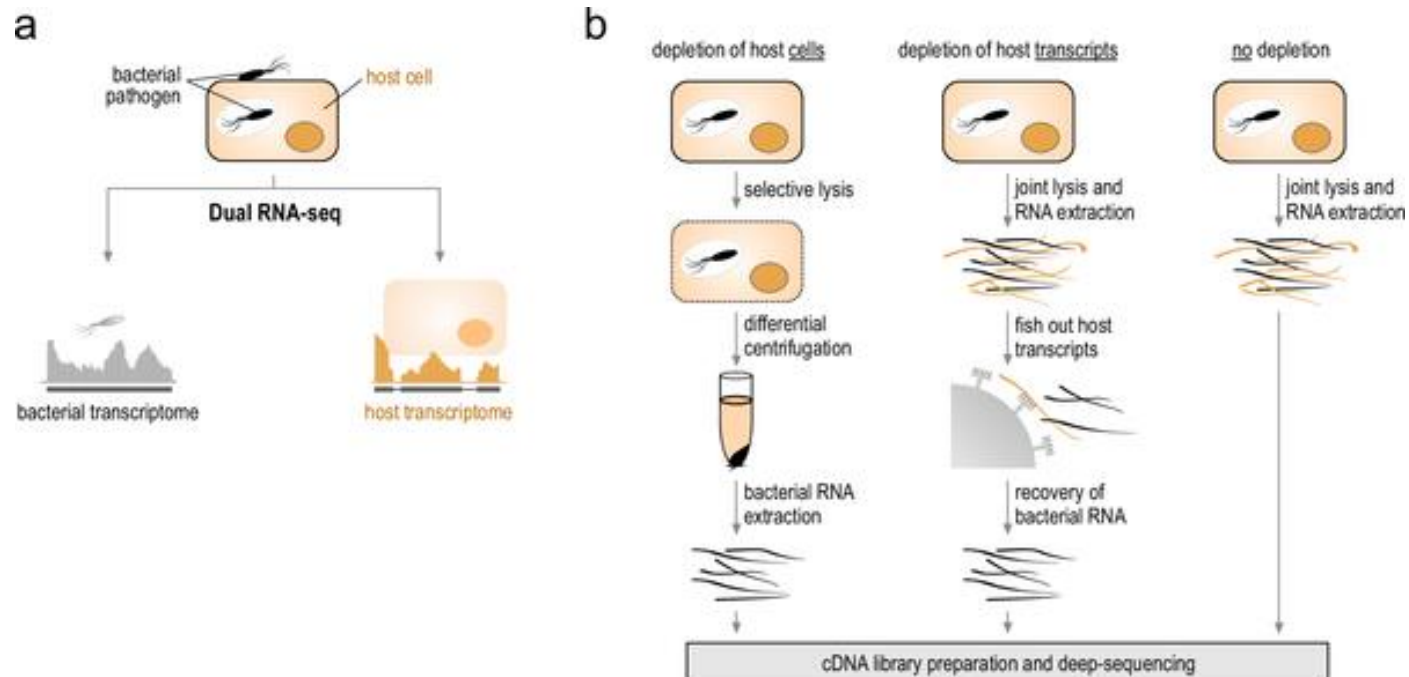


Westermann et al., (2017) Plos Pathogens

REVIEW

### Resolving host-pathogen interactions by dual RNA-seq

Alexander J. Westermann<sup>1</sup>\*, Lars Barquist<sup>1</sup>\*, Jörg Vogel<sup>1,2\*</sup>



# Application of RNA sequencing

## In situ RNA-Sequencing

BRIEF COMMUNICATIONS

Ke et al., (2013) Nature methods

### FISSEQ, Fluorescent In Situ RNA SEQuencing

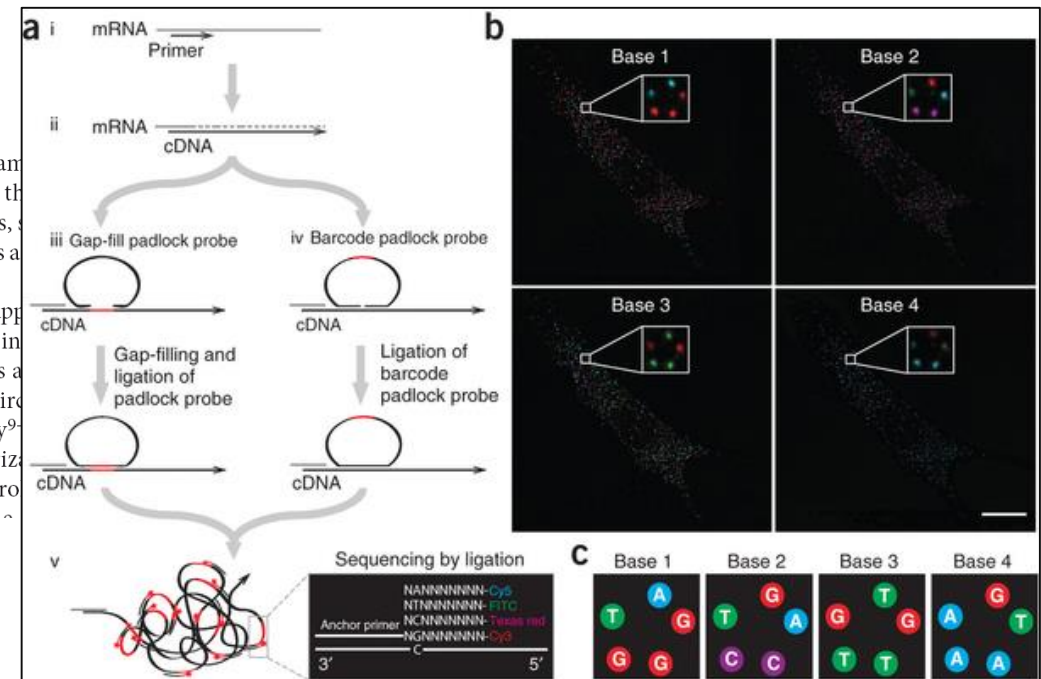
#### *In situ* sequencing for RNA analysis in preserved tissue and cells

Rongqin Ke<sup>1,2,5</sup>, Marco Mignardi<sup>1,2,5</sup>,  
Alexandra Pacureanu<sup>3</sup>, Jessica Svedlund<sup>1</sup>,  
Johan Botling<sup>2</sup>, Carolina Wählby<sup>3,4</sup> & Mats Nilsson<sup>1,2</sup>

Tissue gene expression profiling is performed on homogenates or on populations of isolated single cells to resolve molecular

are limited in spatial resolution. Also, because of sampling the collected cells might not reflect the nature of the compartment targeted for expression profiling. Thus, sequencing single molecules directly in the tissue environment is a challenge for single-cell analysis technology.

Here we show that sequencing chemistry can be applied for analysis of up to four-base-pair fragments in single molecules in the unperturbed context of fixed cells and tissues. Our method is based on padlock probing, rolling-circle amplification (RCA) and sequencing-by-ligation chemistry<sup>9</sup>. RCA in combination with padlock-probe circularization has been used to produce clonally amplified rolling-circle products (RCRs) at high density in cells and tissues.



# Conclusions

NGS technologies are now routine part of biological research.

To date more than 14,000 genomes have been deposited within NCBI

In 2013 Schatz and Langmead reported that the world can generate ~15 petabytes per year!!

The new challenge... to provide infrastructure for analyses and storage.

The NGS arm race is not finished...

- GenapSys (Sigma-Aldrich)
- Genia (Roche)
- new nanopore technology with Firefly (Illumina)
- ...





# What is France Génomique



**A national infrastructure that brings together the capacities and expertise of genomic and bioinformatic platforms**





# What is France Génomique

<https://www.france-genomique.org>

**FRANCE GÉNOMIQUE**

MUTUALISATION DES COMPÉTENCES ET DES ÉQUIPEMENTS FRANÇAIS POUR L'ANALYSE GÉNOMIQUE ET LA BIO-INFORMATIQUE

ACCUEIL À PROPOS DE FRANCE GÉNOMIQUE L'OFFRE DE FRANCE GÉNOMIQUE DOMAINES D'EXPERTISES MODALITÉS DE COLLABORATION GRANDS PROJETS FORMATION ET E-LEARNING PUBLICATIONS CONTACTS

**ACTUALITÉS ET FAITS MARQUANTS**

*3 mai 2017*  
**France Génomique membre de "GA4GH"**  
Le 17 mars 2017, France Génomique est devenu "Organizational Member" de l'Alliance Mondiale pour la Génomique et la santé (Global Alliance for Genomic (...))

*27 avril 2017*  
**Summer School Génopole**  
Bioinformatic and biostatistic tools in medical genomics 4-7 Juillet 2017 Chateaufort/Les Berges de Seine (77)  
Ateliers sur la génomique et la métagénomique d'intérêt médical, l'éthique (...)

**FRANCE GÉNOMIQUE** rassemble et mutualise les ressources des principales plateformes françaises de génomique et de bio-informatique. Créée grâce à un financement « Investissements d'Avenir », elle a pour ambition de maintenir la recherche française au plus haut niveau de compétitivité et de performance dans la production et l'analyse des données de génomique, à la pointe de l'état de l'art à l'échelle internationale. Elle offre à la communauté scientifique publique et privée l'accès aux plateformes françaises les plus performantes, un accompagnement des projets, le plus haut niveau d'expertise et de compétences mutualisées ainsi que la possibilité de participer à des projets ambitieux au niveau national et international.

**SOUMETTRE UN PROJET**  
OUVERT EN CONTINU

**PLATE-FORMES OPÉRATIONNELLES EN GÉNOMIQUE & BIOINFORMATIQUE**

**DÉVELOPPEMENT & VEILLE TECHNOLOGIQUE**

**FRANCE GÉNOMIQUE RECRUTE**  
POSTES À POURVOIR

**INTRANET**

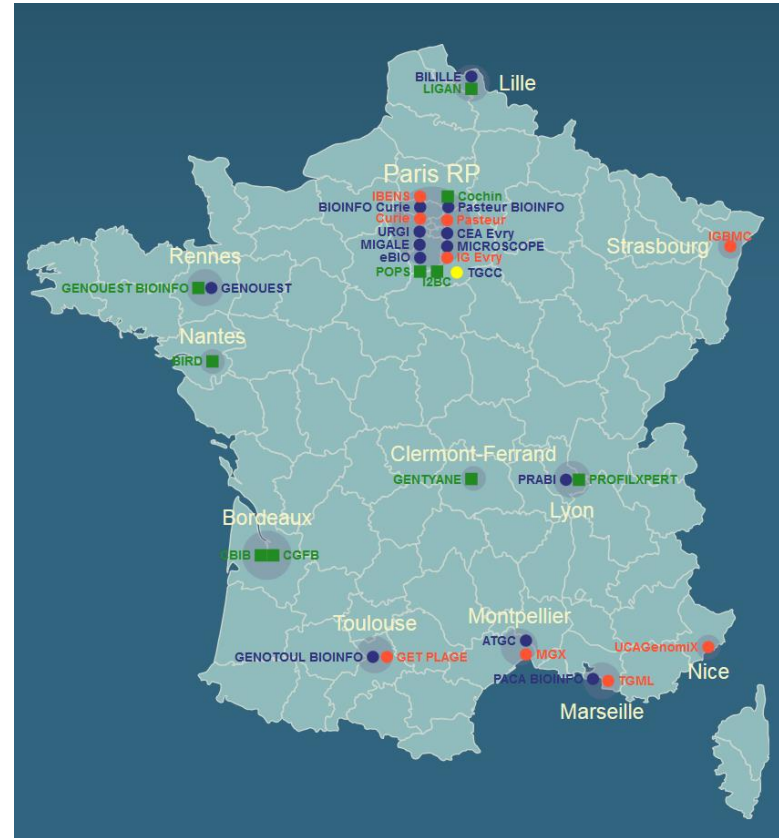
**FORUM**

**ACCÈS TABLEAUX DE BORD**

Tweets de @fr\_genomics  
France Génomique a retweeté  
John Tyson @JTyson

# What is France Génomique

## The France Génomique plateformes



# What is France Génomique

Technologies, Equipments and Expertise sharing



illumina®



Oxford  
**NANOPORE**  
Technologies



 **PACIFIC  
BIOSCIENCES®**

 **FLUIDIGM®**

# What is France Génomique

Call for proposals: « Large-scale sequencing projects »

Environmental and agronomical genomics

**TARA  
OCEANS**

Plankton diversity



<http://oceans.taraexpeditions.org/>

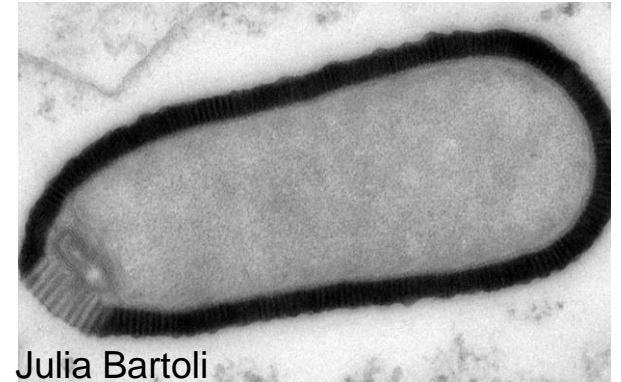
## Discovery of giant virus

Vostochny Mine, East Siberia



Naymushin / REUTERS

Giant Virus



Julia Bartoli

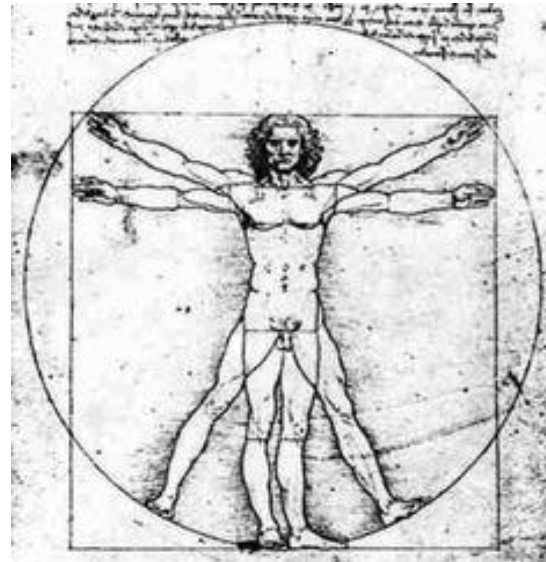


# What is France Génomique

Call for proposals: « Large-scale sequencing projects »

## Health and Human Genetics

**Cancer**



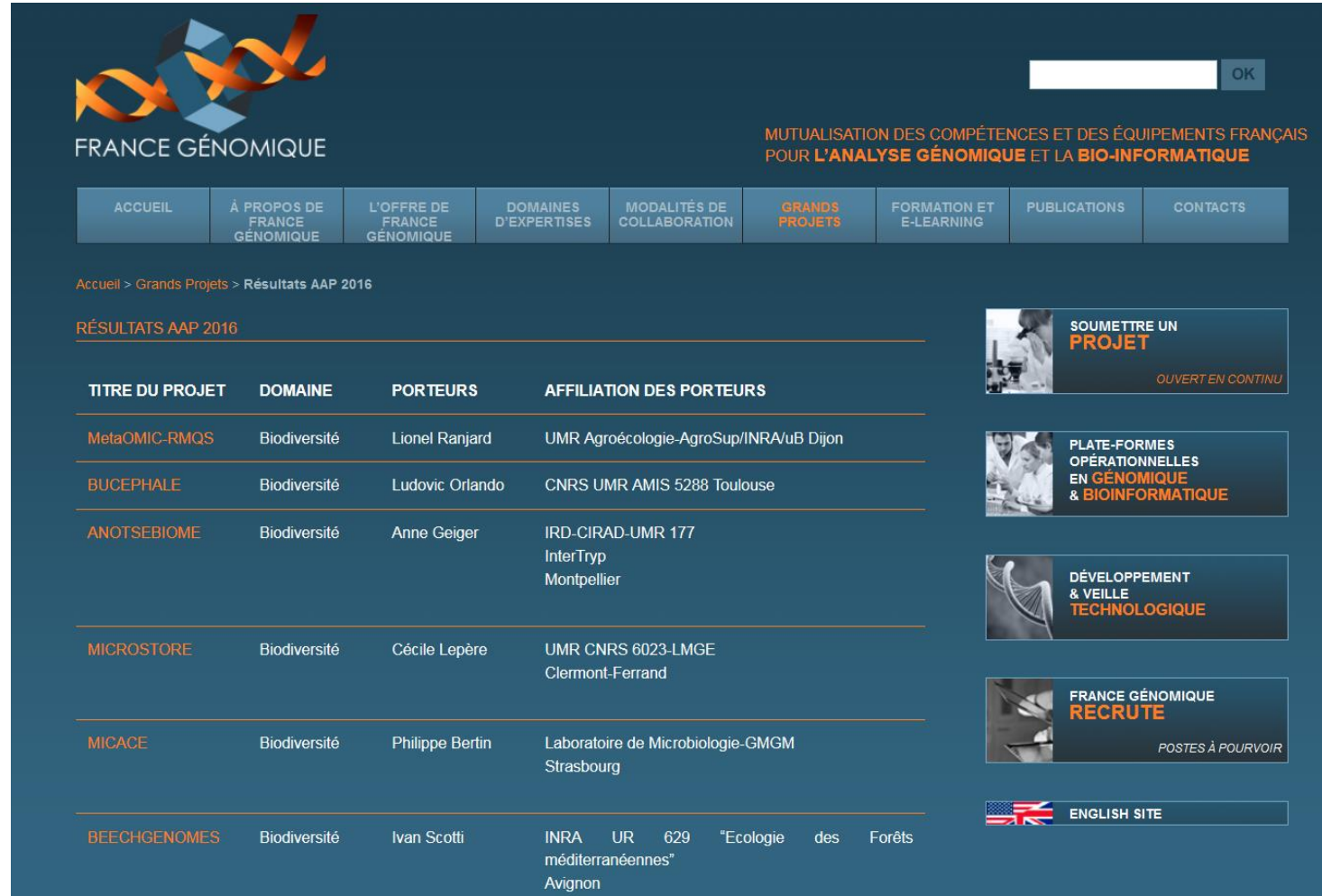
**Myopathies**

**Schizophrenia**

**Muscular dystrophy**

# What is France Génomique

## Call for proposals: « Large-scale sequencing projects »



The screenshot displays the France Génomique website interface. At the top left is the logo, a stylized DNA double helix in orange and blue, with the text "FRANCE GÉNOMIQUE" below it. To the right of the logo is a search bar with an "OK" button. Below the logo, a navigation menu contains several tabs: ACCUEIL, À PROPOS DE FRANCE GÉNOMIQUE, L'OFFRE DE FRANCE GÉNOMIQUE, DOMAINES D'EXPERTISES, MODALITÉS DE COLLABORATION, **GRANDS PROJETS** (highlighted), FORMATION ET E-LEARNING, PUBLICATIONS, and CONTACTS. Below the navigation menu, a breadcrumb trail reads "Accueil > Grands Projets > Résultats AAP 2016". The main content area is titled "RÉSULTATS AAP 2016" and features a table with the following columns: TITRE DU PROJET, DOMAINE, PORTEURS, and AFFILIATION DES PORTEURS. The table lists six projects: MetaOMIC-RMQS, BUCEPHALE, ANOTSEBIOME, MICROSTORE, MICACE, and BEECHGENOMES. To the right of the table are four vertical panels: "SOUMETTRE UN PROJET" (OUVERT EN CONTINU), "PLATE-FORMES OPÉRATIONNELLES EN GÉNOMIQUE & BIOINFORMATIQUE", "DÉVELOPPEMENT & VEILLE TECHNOLOGIQUE", and "FRANCE GÉNOMIQUE RECRUTE" (POSTES À POURVOIR). At the bottom right, there is a button for "ENGLISH SITE" with a flag icon.

MUTUALISATION DES COMPÉTENCES ET DES ÉQUIPEMENTS FRANÇAIS POUR L'ANALYSE GÉNOMIQUE ET LA BIO-INFORMATIQUE

ACCUEIL À PROPOS DE FRANCE GÉNOMIQUE L'OFFRE DE FRANCE GÉNOMIQUE DOMAINES D'EXPERTISES MODALITÉS DE COLLABORATION **GRANDS PROJETS** FORMATION ET E-LEARNING PUBLICATIONS CONTACTS

Accueil > Grands Projets > Résultats AAP 2016

RÉSULTATS AAP 2016

TITRE DU PROJET	DOMAINE	PORTEURS	AFFILIATION DES PORTEURS
MetaOMIC-RMQS	Biodiversité	Lionel Ranjard	UMR Agroécologie-AgroSup/INRA/uB Dijon
BUCEPHALE	Biodiversité	Ludovic Orlando	CNRS UMR AMIS 5288 Toulouse
ANOTSEBIOME	Biodiversité	Anne Geiger	IRD-CIRAD-UMR 177 InterTryp Montpellier
MICROSTORE	Biodiversité	Cécile Lepère	UMR CNRS 6023-LMGE Clermont-Ferrand
MICACE	Biodiversité	Philippe Bertin	Laboratoire de Microbiologie-GMGM Strasbourg
BEECHGENOMES	Biodiversité	Ivan Scotti	INRA UR 629 "Ecologie des Forêts méditerranéennes" Avignon

SOUMETTRE UN PROJET  
OUVERT EN CONTINU

PLATE-FORMES OPÉRATIONNELLES EN GÉNOMIQUE & BIOINFORMATIQUE

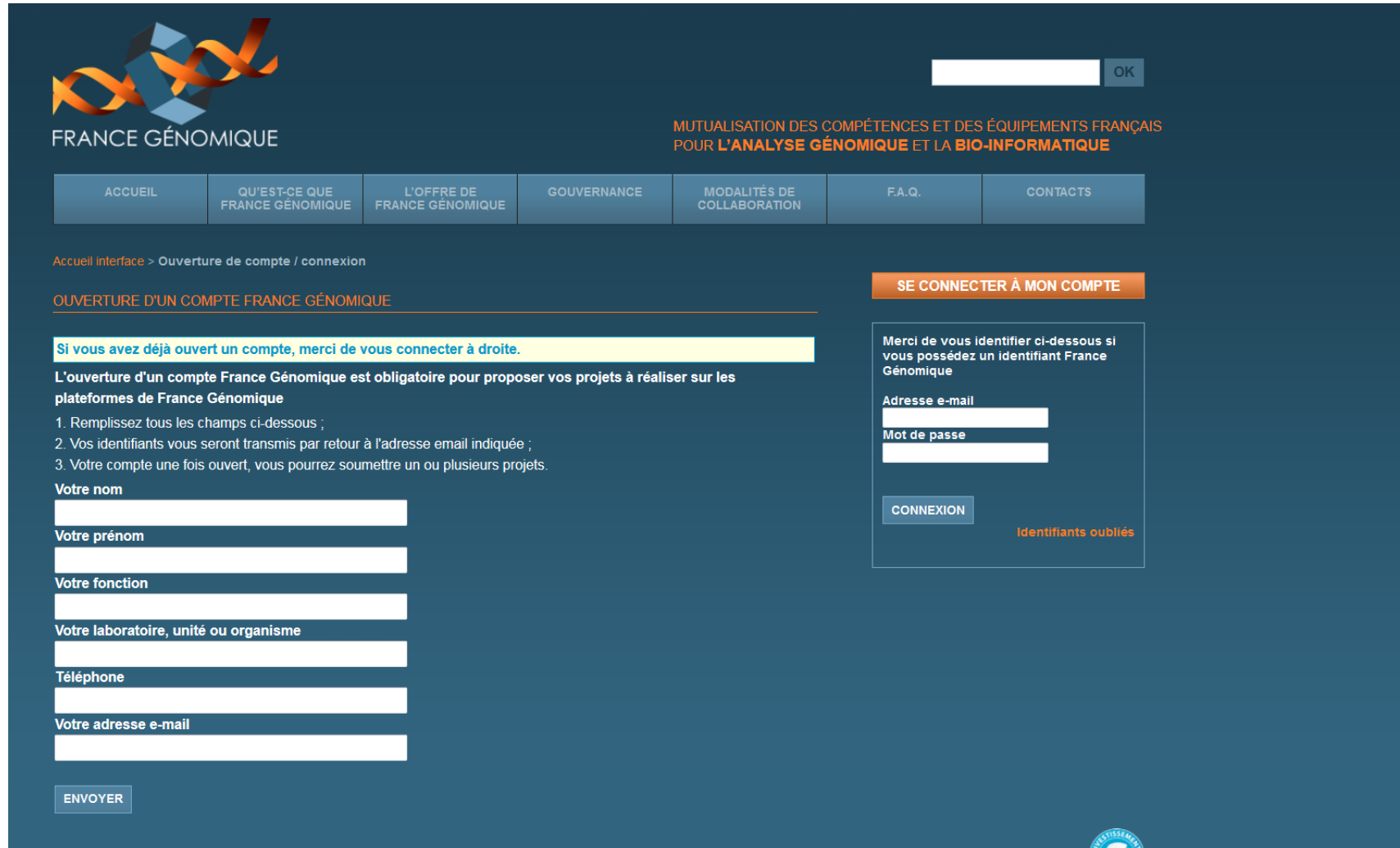
DÉVELOPPEMENT & VEILLE TECHNOLOGIQUE

FRANCE GÉNOMIQUE RECRUTE  
POSTES À POURVOIR

ENGLISH SITE

# What is France Génomique

## Continuous project submission



The screenshot displays the France Génomique website interface. At the top left is the logo, a stylized DNA double helix in orange and blue, with the text "FRANCE GÉNOMIQUE" below it. To the right of the logo is a search bar with an "OK" button. Below the logo and search bar is a navigation menu with seven items: ACCUEIL, QU'EST-CE QUE FRANCE GÉNOMIQUE, L'OFFRE DE FRANCE GÉNOMIQUE, GOUVERNANCE, MODALITÉS DE COLLABORATION, F.A.Q., and CONTACTS. Below the navigation menu is a breadcrumb trail: "Accueil interface > Ouverture de compte / connexion". The main content area is titled "OUVERTURE D'UN COMPTE FRANCE GÉNOMIQUE". A yellow highlighted box contains the text: "Si vous avez déjà ouvert un compte, merci de vous connecter à droite." Below this, a paragraph states: "L'ouverture d'un compte France Génomique est obligatoire pour proposer vos projets à réaliser sur les plateformes de France Génomique". A list of three instructions follows: 1. Remplissez tous les champs ci-dessous ; 2. Vos identifiants vous seront transmis par retour à l'adresse email indiquée ; 3. Votre compte une fois ouvert, vous pourrez soumettre un ou plusieurs projets. Below the instructions are several input fields: "Votre nom", "Votre prénom", "Votre fonction", "Votre laboratoire, unité ou organisme", "Téléphone", and "Votre adresse e-mail". At the bottom left of this section is an "ENVOYER" button. On the right side of the page, there is a box titled "SE CONNECTER À MON COMPTE". Inside this box, it says "Merci de vous identifier ci-dessous si vous possédez un identifiant France Génomique". Below this text are two input fields: "Adresse e-mail" and "Mot de passe". At the bottom of this box is a "CONNEXION" button and a link labeled "Identifiants oubliés". At the bottom right of the page, there is a small circular logo for "UNIVERSITÉ DE BORDEAUX" and the text "FRANCE GÉNOMIQUE EST FINANCÉ PAR".

FRANCE GÉNOMIQUE

MUTUALISATION DES COMPÉTENCES ET DES ÉQUIPEMENTS FRANÇAIS  
POUR L'ANALYSE GÉNOMIQUE ET LA BIO-INFORMATIQUE

ACCUEIL QU'EST-CE QUE FRANCE GÉNOMIQUE L'OFFRE DE FRANCE GÉNOMIQUE GOUVERNANCE MODALITÉS DE COLLABORATION F.A.Q. CONTACTS

Accueil interface > Ouverture de compte / connexion

OUVERTURE D'UN COMPTE FRANCE GÉNOMIQUE

Si vous avez déjà ouvert un compte, merci de vous connecter à droite.

L'ouverture d'un compte France Génomique est obligatoire pour proposer vos projets à réaliser sur les plateformes de France Génomique

1. Remplissez tous les champs ci-dessous ;
2. Vos identifiants vous seront transmis par retour à l'adresse email indiquée ;
3. Votre compte une fois ouvert, vous pourrez soumettre un ou plusieurs projets.

Votre nom

Votre prénom

Votre fonction

Votre laboratoire, unité ou organisme

Téléphone

Votre adresse e-mail

ENVOYER

SE CONNECTER À MON COMPTE

Merci de vous identifier ci-dessous si vous possédez un identifiant France Génomique

Adresse e-mail

Mot de passe

CONNEXION

Identifiants oubliés

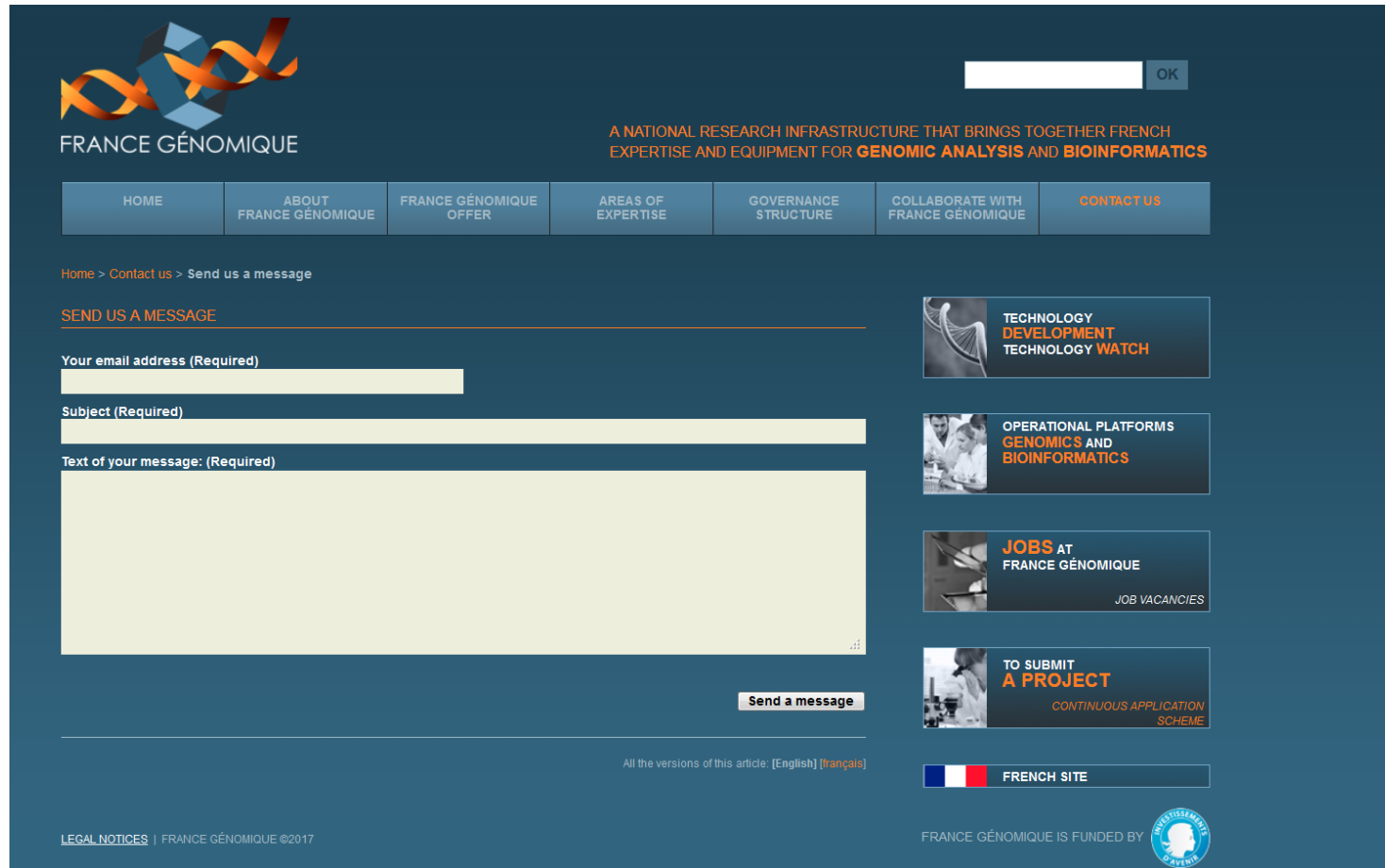
UNIVERSITÉ DE BORDEAUX

FRANCE GÉNOMIQUE EST FINANCÉ PAR



# What is France Génomique

## To contact us



The screenshot shows the contact page of the France Génomique website. At the top left is the logo, a stylized DNA double helix in orange and blue, with the text "FRANCE GÉNOMIQUE" below it. To the right of the logo is a search bar with an "OK" button. Below the logo is a navigation menu with seven items: HOME, ABOUT FRANCE GÉNOMIQUE, FRANCE GÉNOMIQUE OFFER, AREAS OF EXPERTISE, GOVERNANCE STRUCTURE, COLLABORATE WITH FRANCE GÉNOMIQUE, and CONTACT US. The "CONTACT US" item is highlighted. Below the navigation menu is a breadcrumb trail: "Home > Contact us > Send us a message". The main content area is titled "SEND US A MESSAGE" and contains three input fields: "Your email address (Required)", "Subject (Required)", and "Text of your message: (Required)". A "Send a message" button is located at the bottom right of the form. To the right of the form are four promotional boxes: "TECHNOLOGY DEVELOPMENT TECHNOLOGY WATCH" (with a DNA helix icon), "OPERATIONAL PLATFORMS GENOMICS AND BIOINFORMATICS" (with a photo of people working), "JOBS AT FRANCE GÉNOMIQUE JOB VACANCIES" (with a photo of a person), and "TO SUBMIT A PROJECT CONTINUOUS APPLICATION SCHEME" (with a photo of a person). At the bottom of the page, there is a link for "All the versions of this article: [English] [français]", a "FRENCH SITE" button with the French flag, and a footer with "LEGAL NOTICES | FRANCE GÉNOMIQUE ©2017" and "FRANCE GÉNOMIQUE IS FUNDED BY" followed by the logo of the French Government (Gouvernement de la France).

@: mathilde.clement@inra.fr

**Thank you for your attention!**