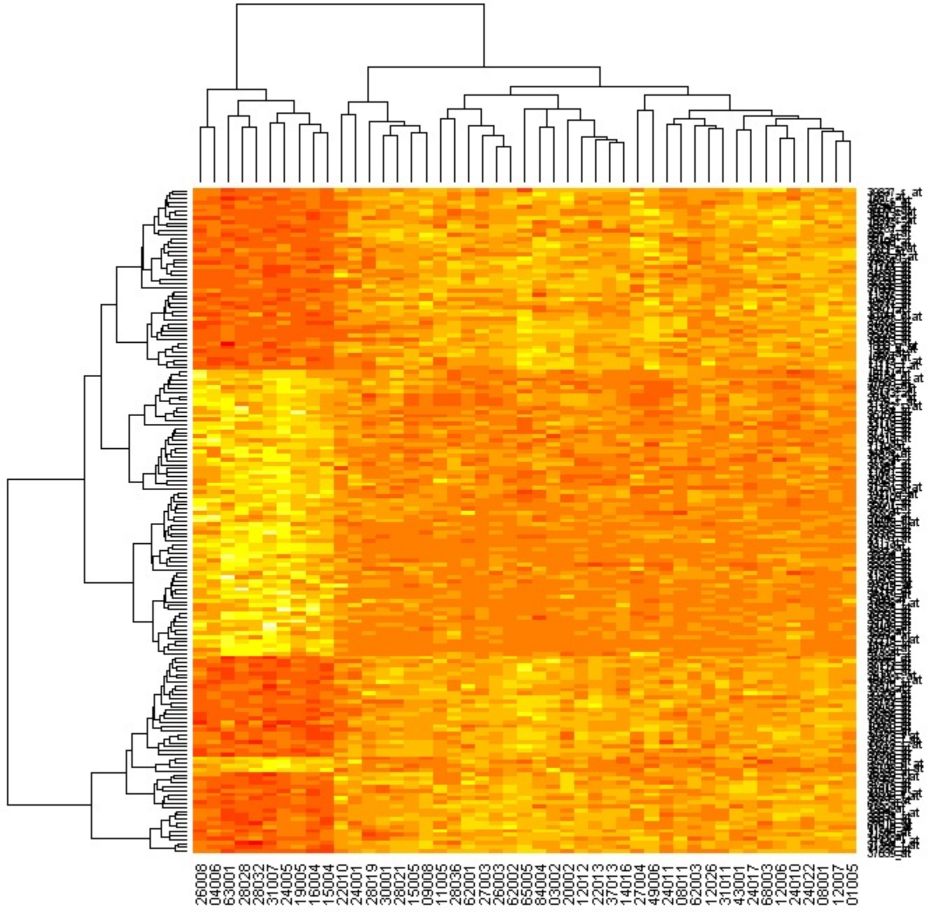
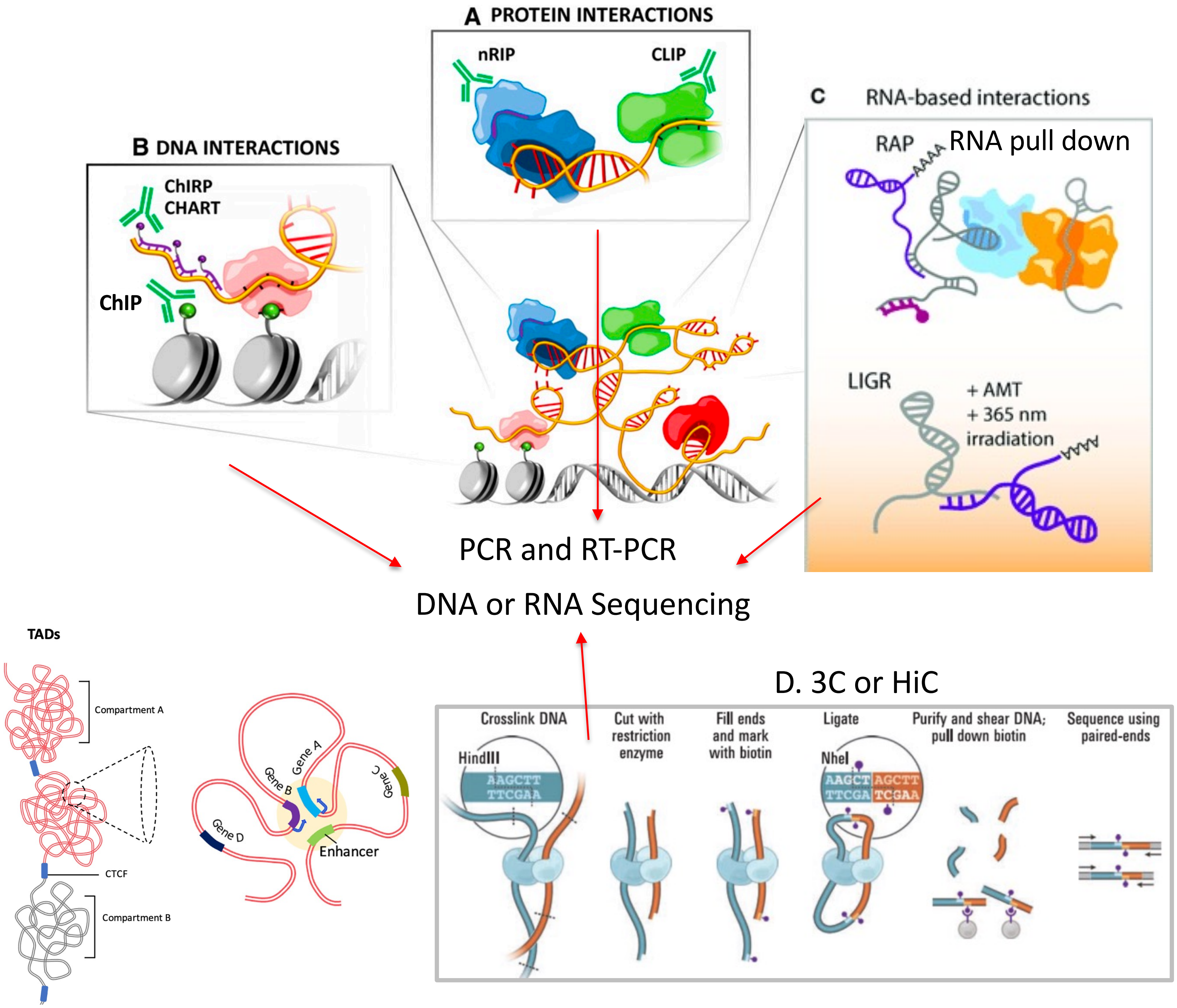
RNA-Seq: experimental procedures and data analysis





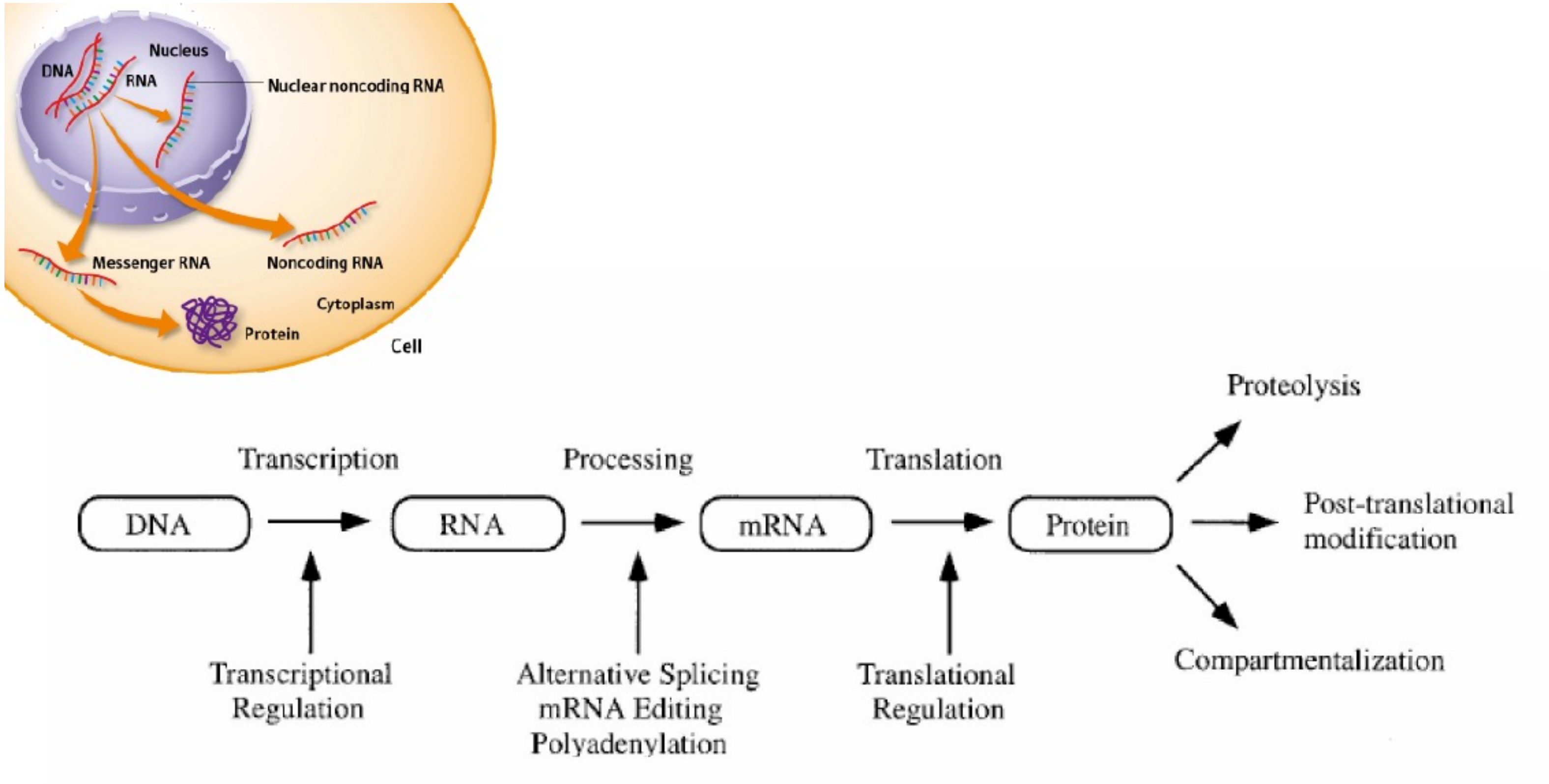
PCR and RT-PCR

DNA or RNA Sequencing

D. 3C or HiC

RNA pull down

STEPS THAT ARE ANALYZED BY -OMICS



Genome

Transcriptome

Proteome

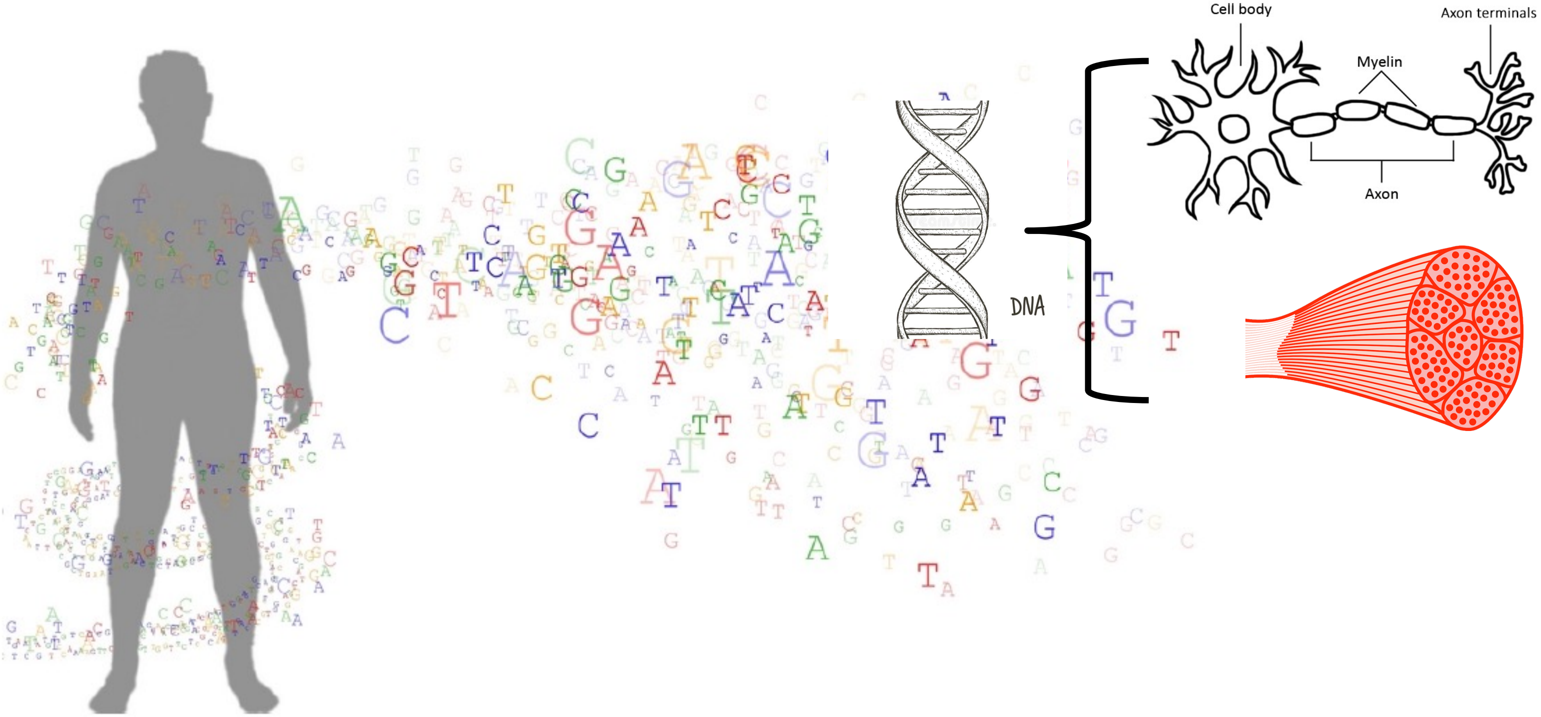
↑

RNA-Seq

TRANSCRIPTOME ANALYSIS: WHY?

Issues in the studies of Transcriptome

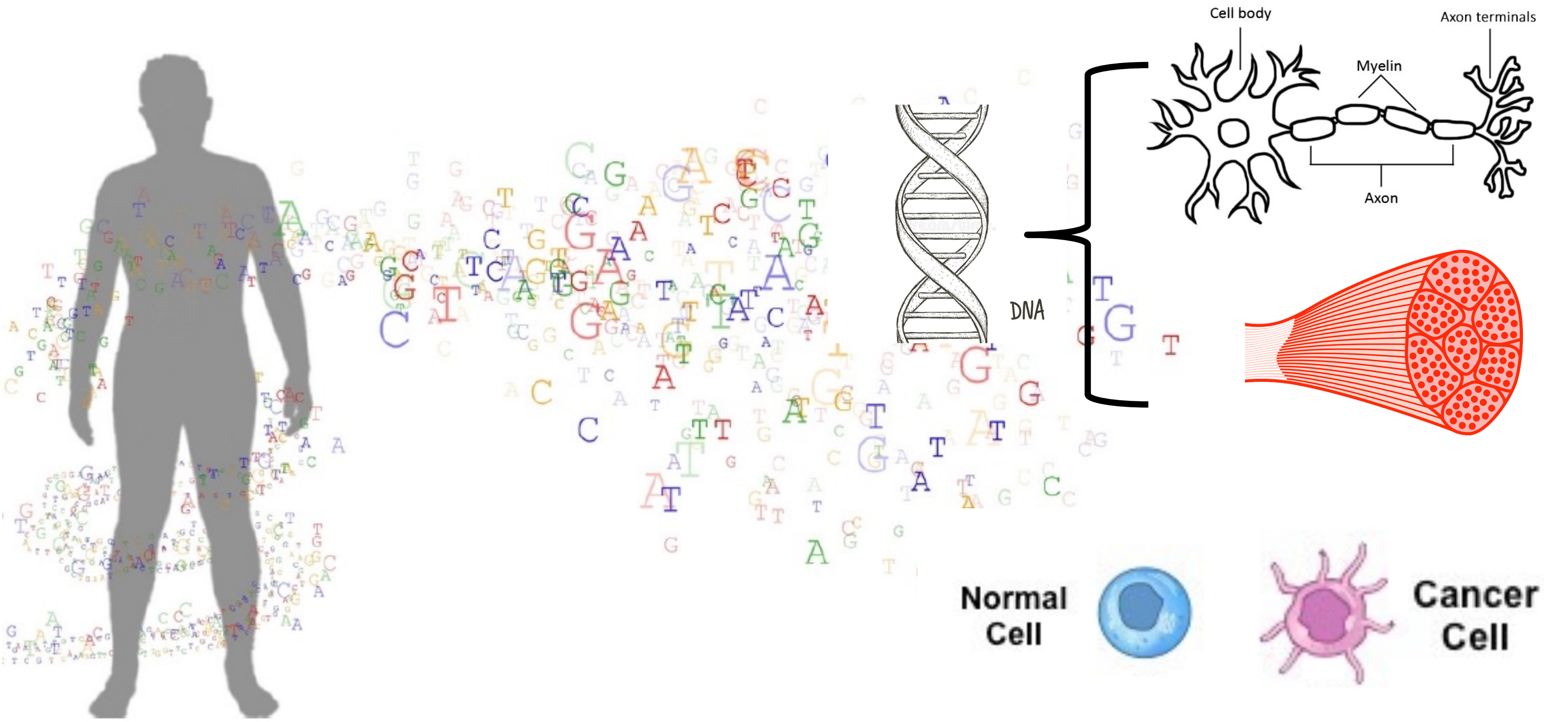
- The Transcriptome of a cell is a dynamic entity: unlike the Genome, it constantly changes.



TRANSCRIPTOME ANALYSIS: WHY?

Issues in the studies of Transcriptome

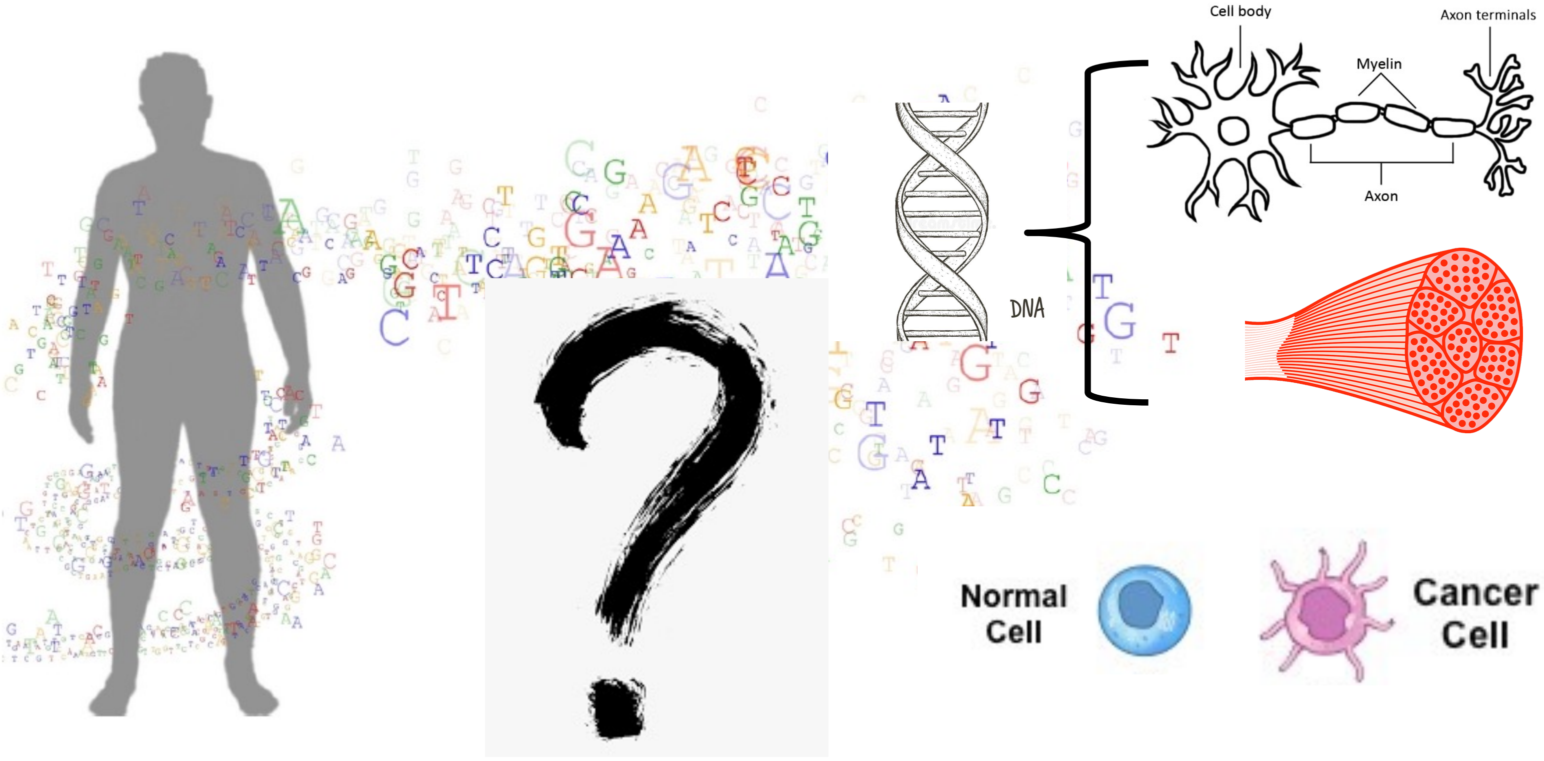
- The Transcriptome of a cell is a dynamic entity: unlike the Genome, it constantly changes.



TRANSCRIPTOME ANALYSIS: WHY?

Issues in the studies of Transcriptome

- The Transcriptome of a cell is a dynamic entity: unlike the Genome, it constantly changes.

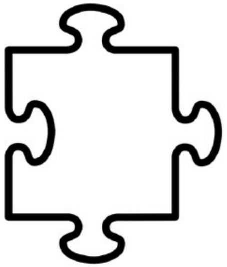


DIFFERENT WAYS TO APPROACH BIOLOGICAL QUESTIONS

⬆

BOTTOM-UP (Classical):

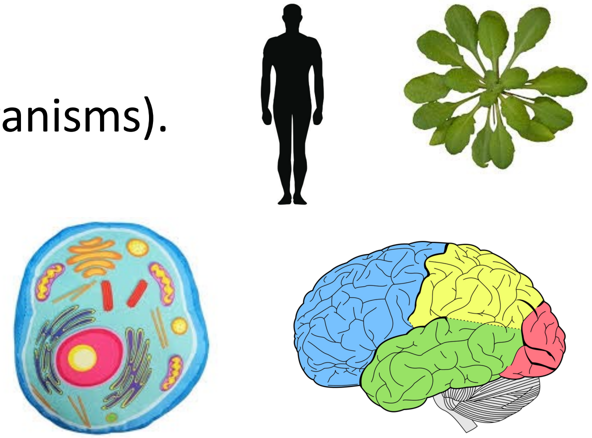
Detailed analysis of single gene/proteins. Step by step assembly of results to get an overview about processes within cells/organisms.



⬇

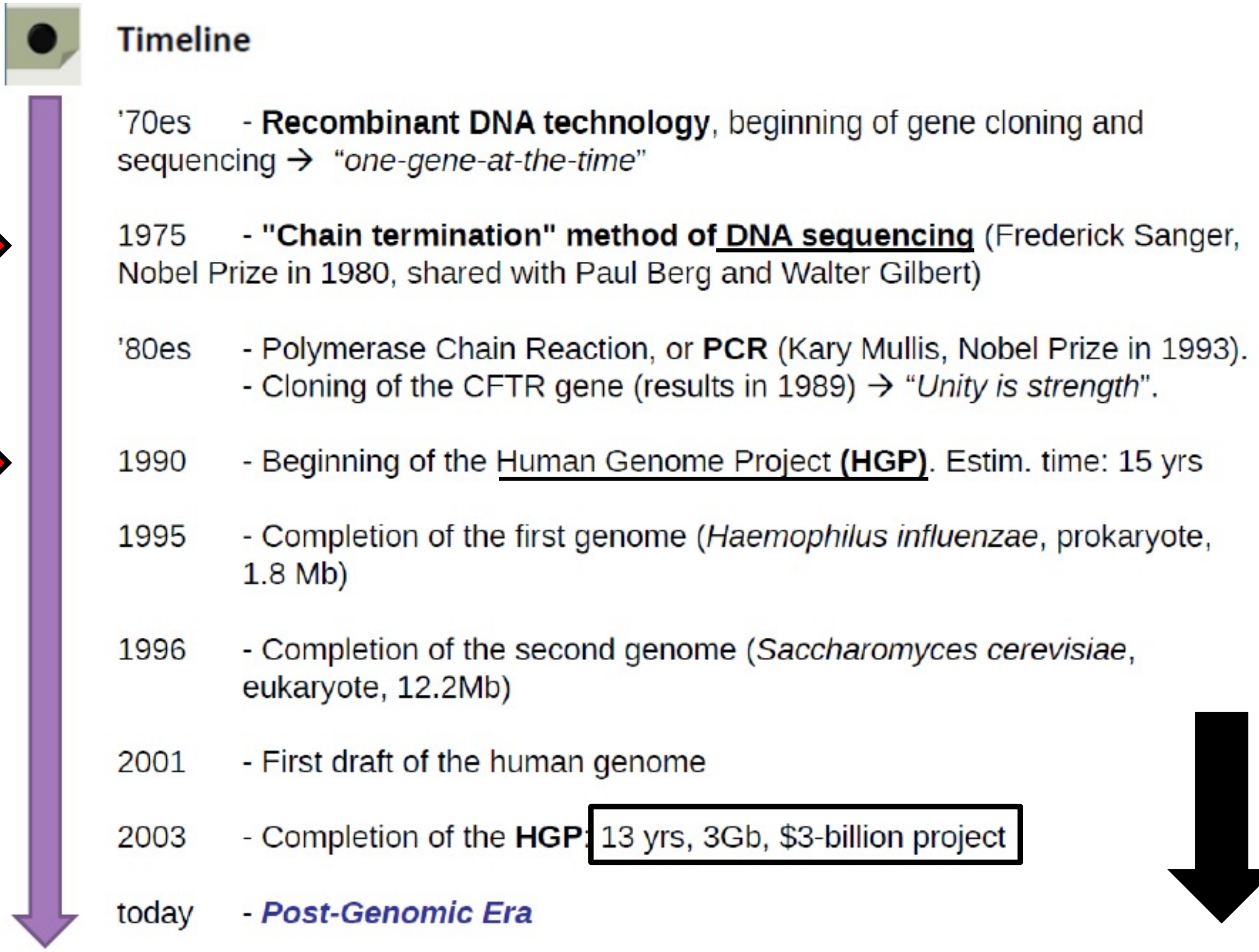
TOP-DOWN (Modern):

Analysis of complete systems (cells/tissues/organisms).



DIFFERENT WAYS TO APPROACH BIOLOGICAL QUESTIONS

Pre-NGS era



Timeline

'70es - **Recombinant DNA technology**, beginning of gene cloning and sequencing → *"one-gene-at-the-time"*

1975 - **"Chain termination" method of DNA sequencing** (Frederick Sanger, Nobel Prize in 1980, shared with Paul Berg and Walter Gilbert)

'80es - Polymerase Chain Reaction, or **PCR** (Kary Mullis, Nobel Prize in 1993).
- Cloning of the CFTR gene (results in 1989) → "*Unity is strength*".

1990 - Beginning of the Human Genome Project **(HGP)**. Estim. time: 15 yrs

1995 - Completion of the first genome (*Haemophilus influenzae*, prokaryote, 1.8 Mb)

1996 - Completion of the second genome (*Saccharomyces cerevisiae*, eukaryote, 12.2Mb)

2001 - First draft of the human genome

2003 - Completion of the **HGP**: 13 yrs, 3Gb, $3-billion project

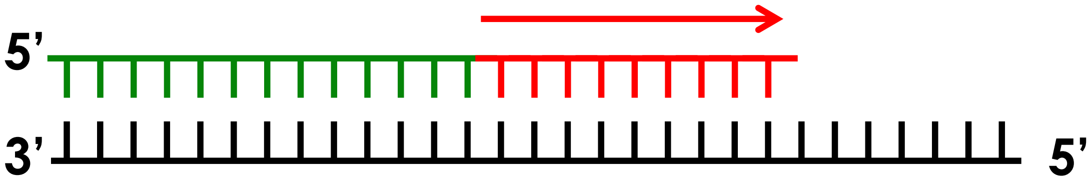
today - ***Post-Genomic Era***

Top Down

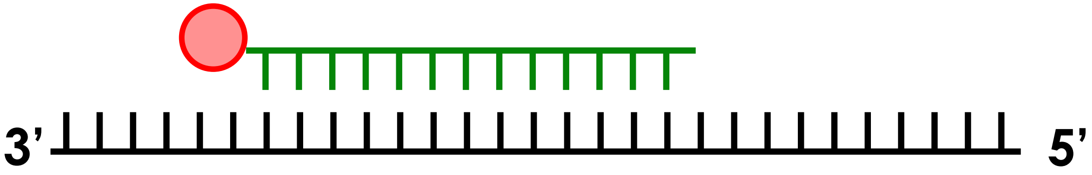
TRANSCRIPTOME ANALYSIS: HOW?

How to detect something that is unknown?

PCR / qPCR /
classic sequencing

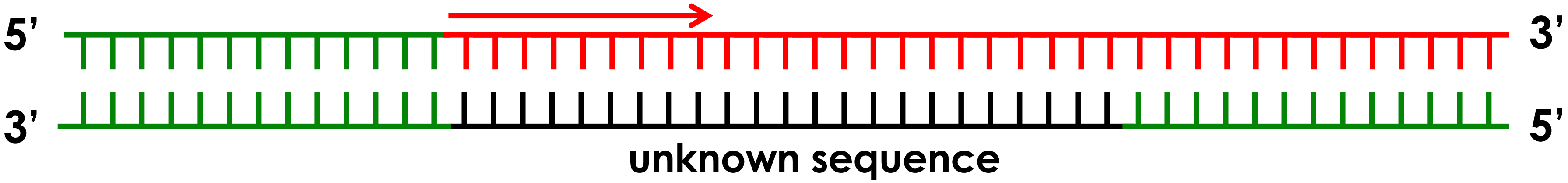


Northern blot /
Southern blot



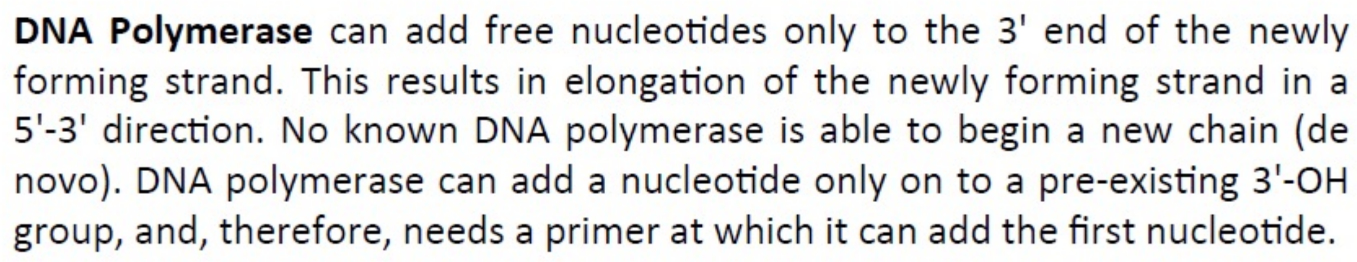
We need to make detectable something that is not known

Next-Generation Sequencing (NGS)

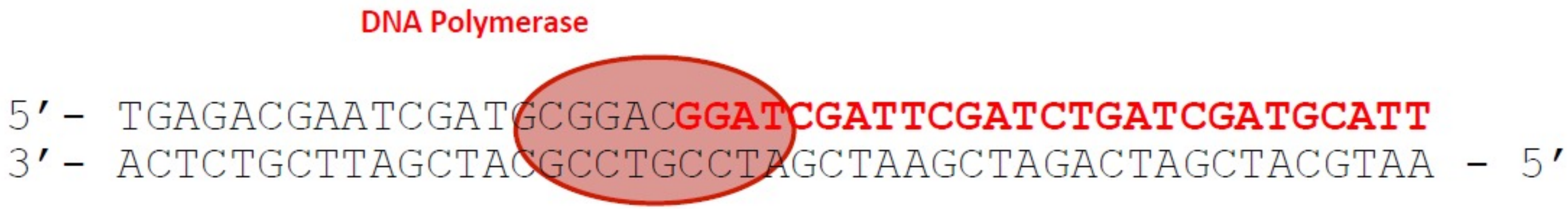


SANGER METHOD FOR DNA SEQUENCING

History of Sequencing: Sanger method for DNA sequencing



DNA Polymerase can add free nucleotides only to the 3' end of the newly forming strand. This results in elongation of the newly forming strand in a 5'-3' direction. No known DNA polymerase is able to begin a new chain (de novo). DNA polymerase can add a nucleotide only on to a pre-existing 3'-OH group, and, therefore, needs a primer at which it can add the first nucleotide.



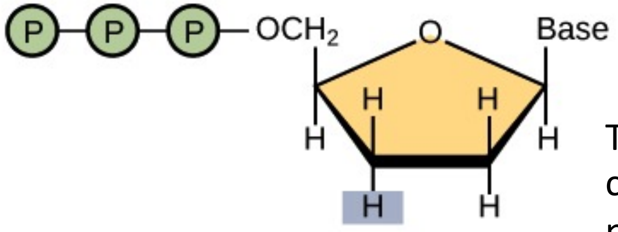
DNA Polymerase

5'- TGAGACGAATCGATGCGGAC**GGATCGATTCGATCTGATCGATGCATT**

3'- ACTCTGCTTAGCTACGCCTGCCTAGCTAAGCTAGACTAGCTACGTAA - 5'

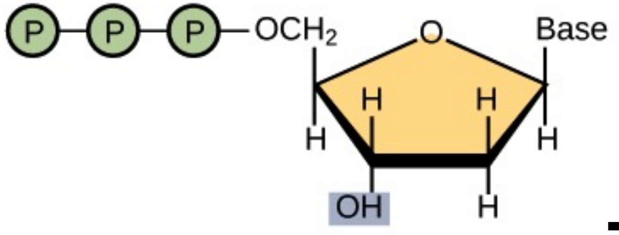
SANGER METHOD FOR DNA SEQUENCING

- "**Sanger Sequencing**" developed by Fred Sanger *et al.* in the mid 1970's
- Uses dideoxynucleotides for "chain termination", generating fragments of different lengths ending in ddATP, ddGTP, ddCTP or ddTTP



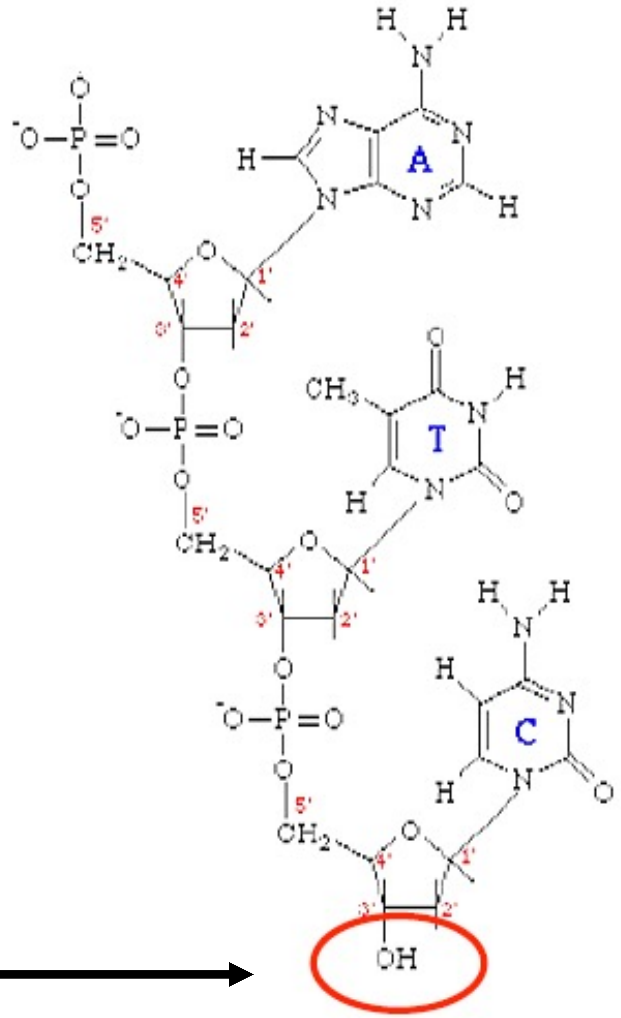
Dideoxynucleotide (ddNTP)

The dideoxynucleotide cannot form the phosphodiester bond with the next nucleotide

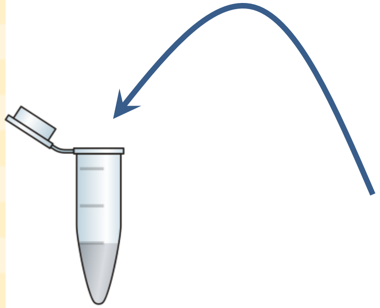


Deoxynucleotide (dNTP)

L'OH al 3' è richiesto per formare il legame fosfodiesterico con il nucleotide successivo



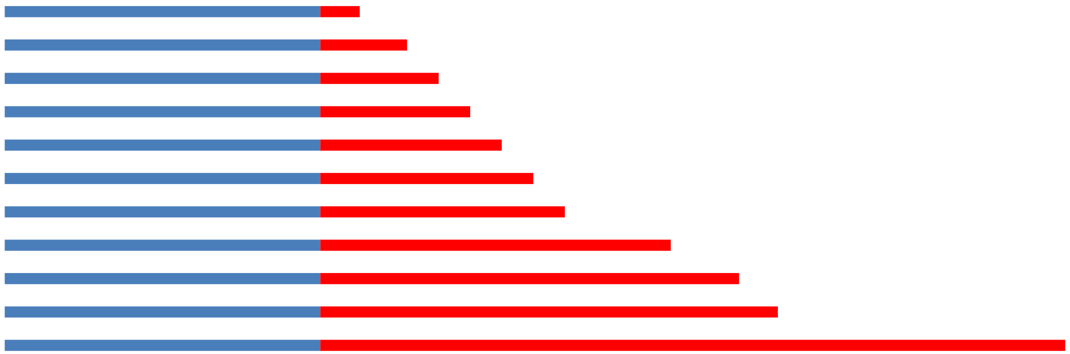
SANGER METHOD FOR DNA SEQUENCING

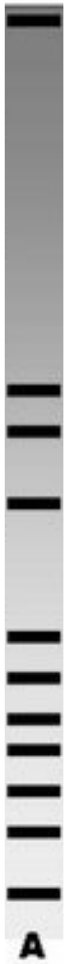


- Template DNA
- DNA Polymerase
- Primer
- dATP, dCTP, dGTP, dTTP
- **ddATP** (or ddCTP, ddGTP, ddTTP)

AT**AAAAAA**CTC**A**G**AA**CGGCTTCGT**A**

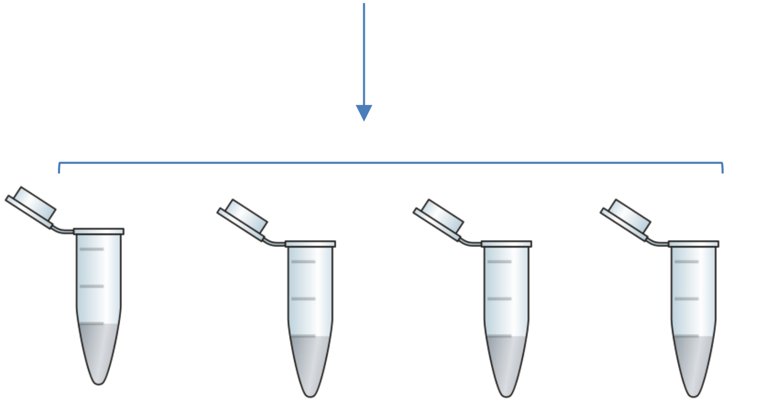
GACTGACTGACTATTTTTTGAGTCTTGCCGAAGCAT



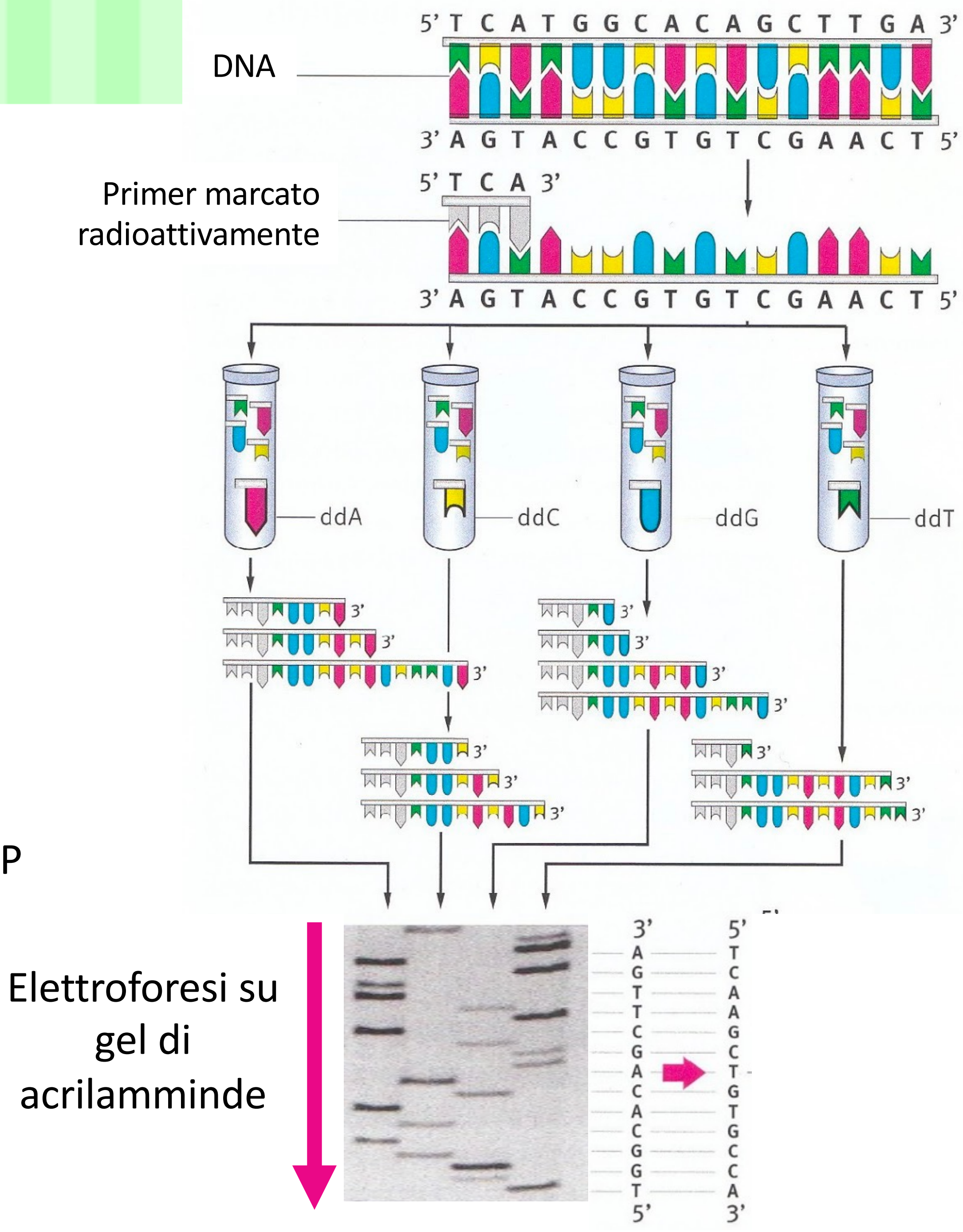


SANGER METHOD FOR DNA SEQUENCING

- Template DNA
- DNA Polymerase
- Primer
- dATP, dCTP, dGTP, dTTP



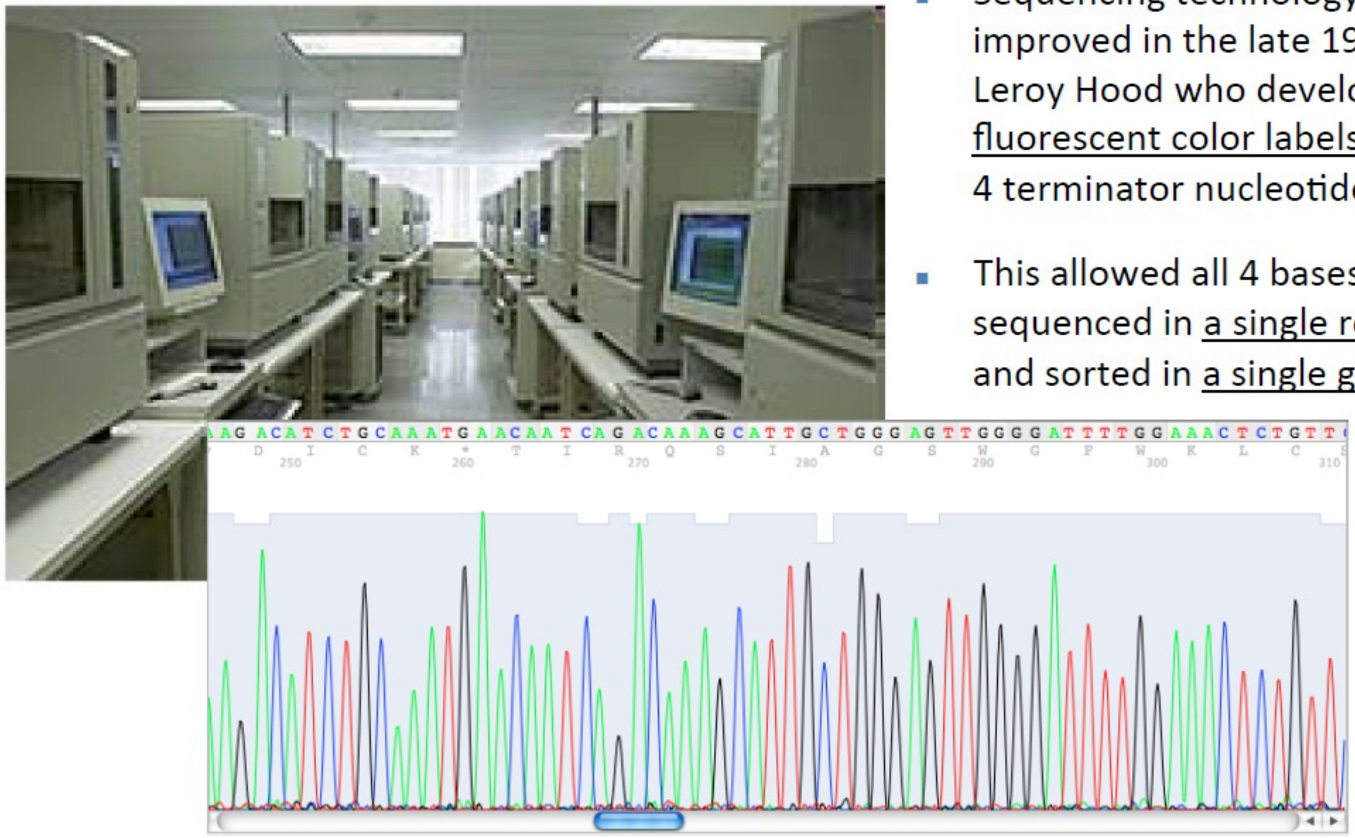
ddATP ddCTP ddGTP ddTTP


DNA
5' T C A T G G C A C A G C T T G A 3'
3' A G T A C C G T G T C G A A C T 5'
Primer marcato radioattivamente
5' T C A 3'
3' A G T A C C G T G T C G A A C T 5'
ddA
ddC
ddG
ddT
3'
Elettroforesi su gel di acrilamminde
3' A G T T C G A C A C G G T 5'
5' T C A A G C T G T G C C A 3'

SANGER METHOD FOR DNA SEQUENCING

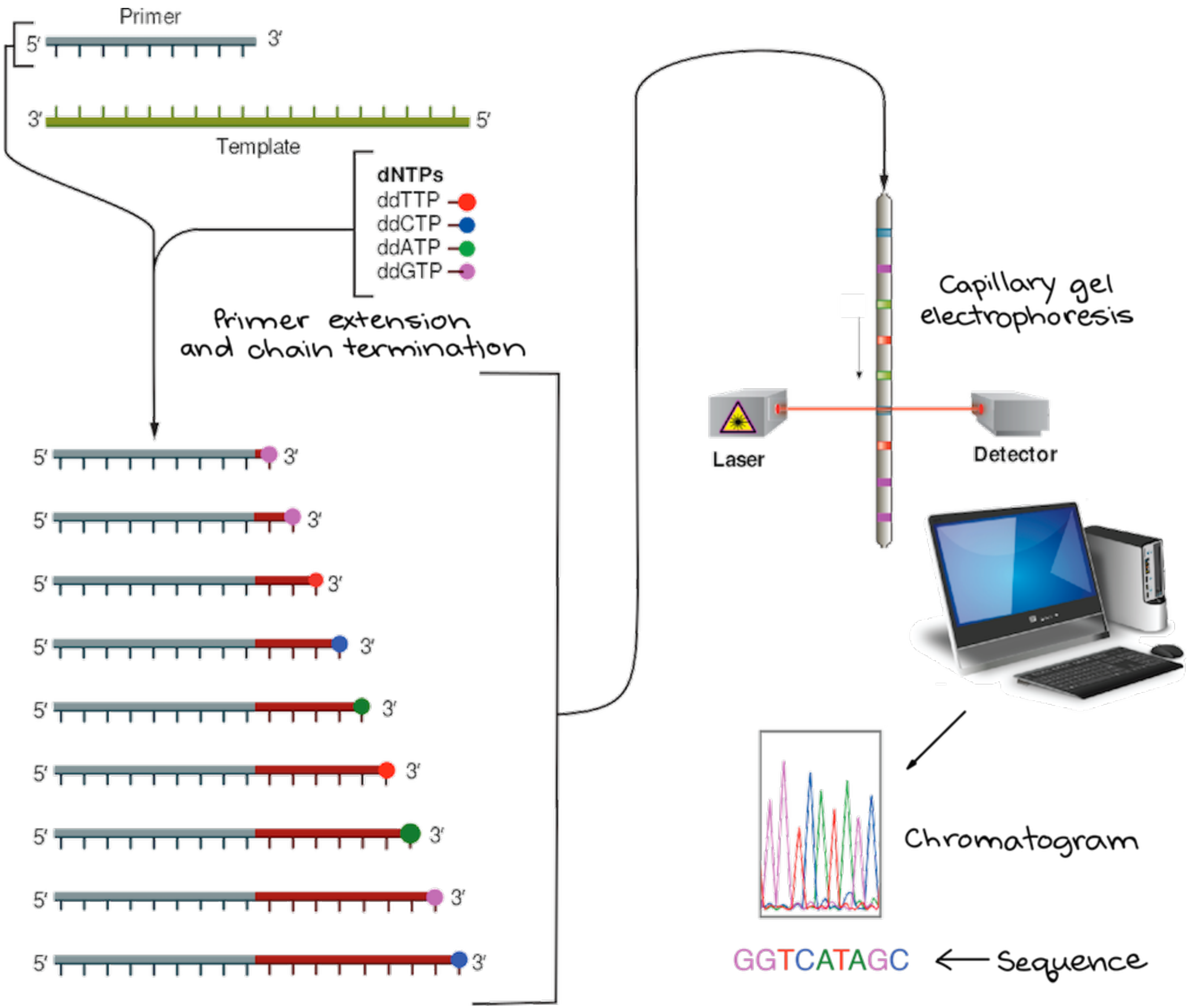
SANGER METHOD FOR DNA SEQUENCING

Automated Sequencing



- Sequencing technology was improved in the late 1980s by Leroy Hood who developed fluorescent color labels for the 4 terminator nucleotide bases.
- This allowed all 4 bases to be sequenced in a single reaction and sorted in a single gel lane

Metodo SANGER per il sequenziamento del DNA



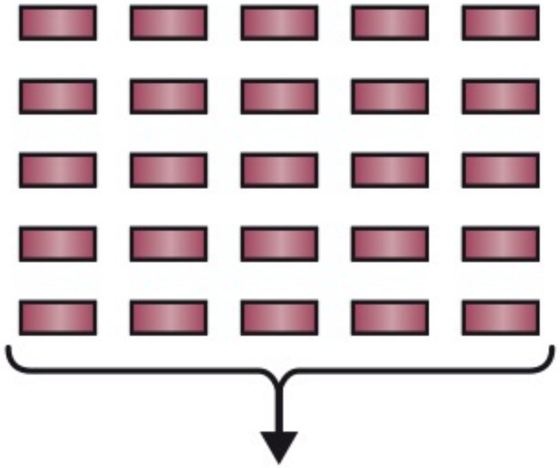
Chain-Termination Sequencing

Cloned insert or PCR product

TACGGATGGCATAGA

A single DNA sequence is generated

Second-Generation Sequencing



Library of DNA fragments

ACGTATCATGCGGATGG
TAGCATGACGTAGCGTT
GTAGCAGGTACGATGCC
GTAGACGATGCAGCATC
TAGGACCTAGCCGGACA

Many fragments are sequenced

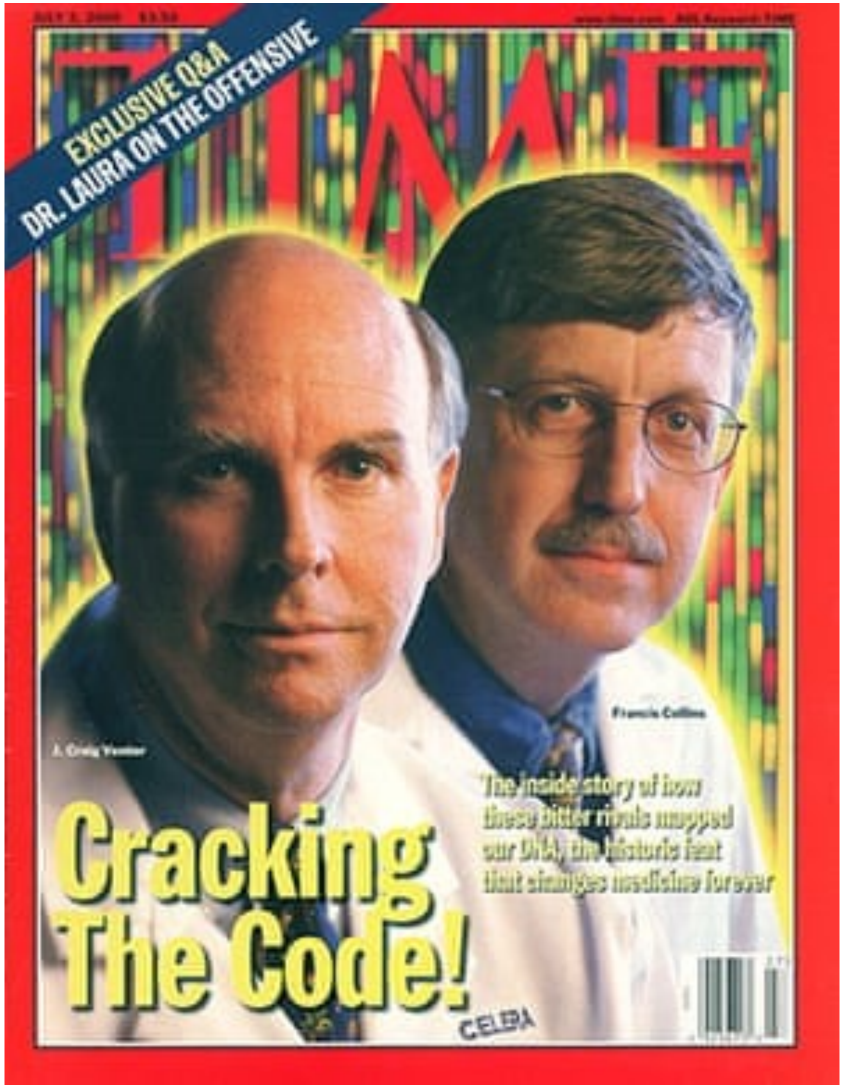
HUMAN GENOME PROJECT

Craig Venter

Celera Genomics

- Private company
- start in 1998
- 300 Milion $

No public access to data



TIME
EXCLUSIVE Q&A
DR. LAURA ON THE OFFENSIVE
J. Craig Venter
Francis Collins
Cracking The Code!
The inside story of how these bitter rivals mapped our DNA, the historic feat that changes medicine forever
CELERA

Francis Collins

International Consortium

- 20 groups from USA, UK, China, Japan, Germany and France
- more than 1000 scientists
- start in 1990
- 2.7 billion $

Public access to data

Strategies

- Hierarchical shotgun approach
    - International Human Genome Sequencing Consortium (IHGSC)
- Whole-genome shotgun approach
    - Celera Genomics

HUMAN GENOME PROJECT: SHOTGUN SEQUENCING

- Sequencing technology allows for obtaining a sequence of about 800 bp at a time.

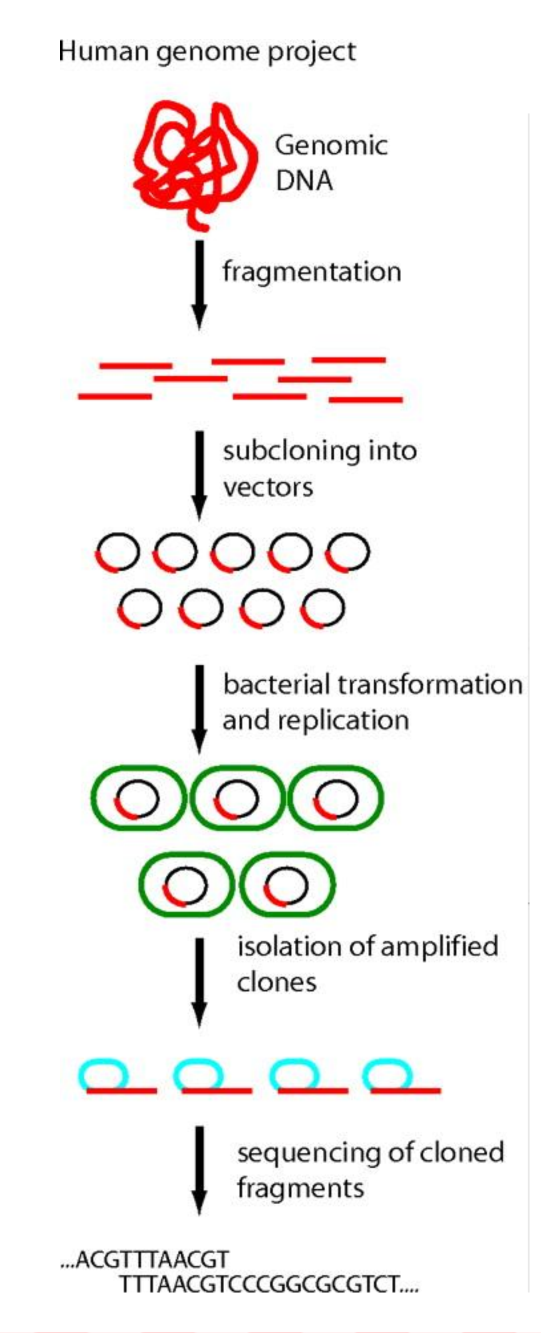
- Genomic DNA must be fragmented into small pieces for sequencing and then reassembled like a giant puzzle.

- Fragments of 150–350 kb are inserted into bacterial artificial chromosomes (BACs), which are then transformed into bacterial cells and replicated.

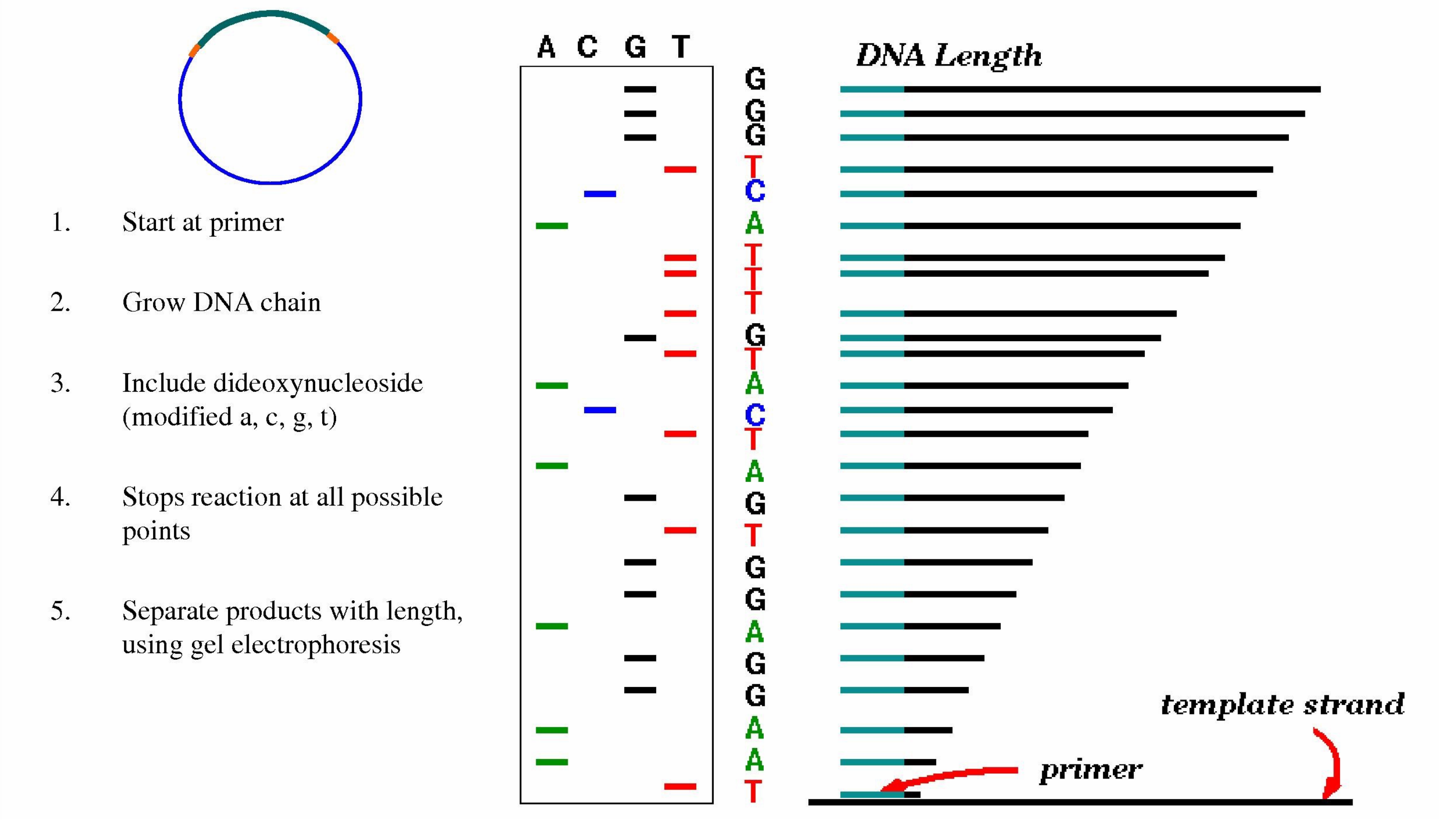
- The clones are fragmented into subclones of smaller sizes (4,000–6,000 bp) and reinserted into bacteria for amplification.

- DNA is extracted from the colonies.

- Sequenced using the Sanger method



HUMAN GENOME PROJECT: SHOTGUN SEQUENCING

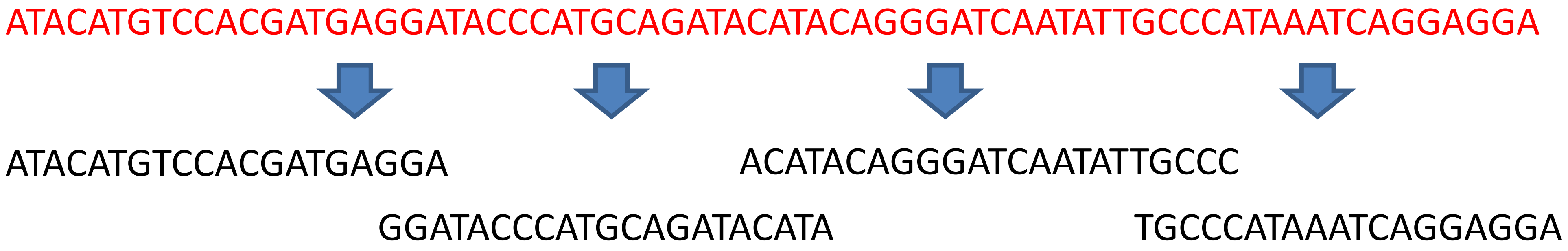


1. Start at primer
2. Grow DNA chain
3. Include dideoxynucleoside (modified a, c, g, t)
4. Stops reaction at all possible points
5. Separate products with length, using gel electrophoresis

- Can produce DNA fragments 700-900bp long, but it's slow
- Lots of other problems including clone library generation and low-throughput
- The Human Genome Project used Sanger sequencing, completion took over 10 years

HUMAN GENOME PROJECT: SHOTGUN SEQUENCING

The principle is to obtain a series of overlapping DNA fragments that can be connected into a continuous map.



ATACATGTCCACGATGAGGATACCCATGCAGATACATACAGGGATCAATATTGCCCATAAATCAGGAGGA

ATACATGTCCACGATGAGGA

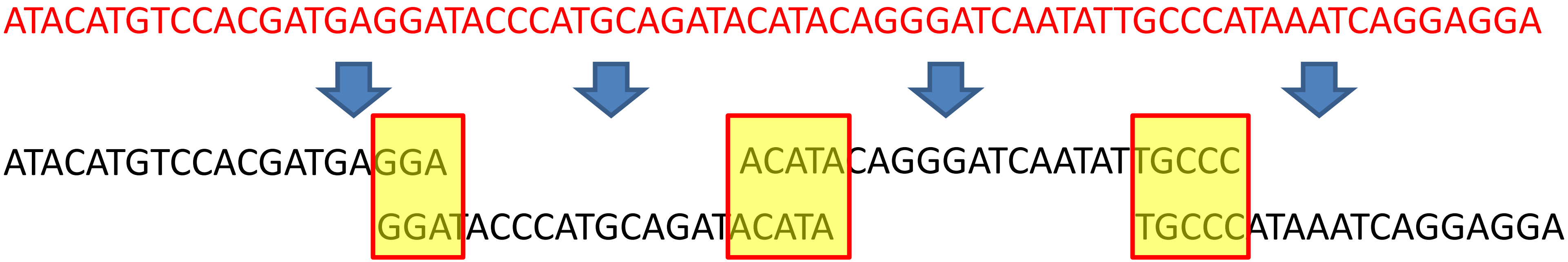
GGATACCCATGCAGATACATA

ACATACAGGGATCAATATTGCCC

TGCCCATAAATCAGGAGGA

HUMAN GENOME PROJECT: SHOTGUN SEQUENCING

The principle is to obtain a series of overlapping DNA fragments that can be connected into a continuous map.



ATACATGTCCACGATGAGGATACCCATGCAGATACATACAGGGATCAATATTGCCCATAAATCAGGAGGA

ATACATGTCCACGATGAGGA

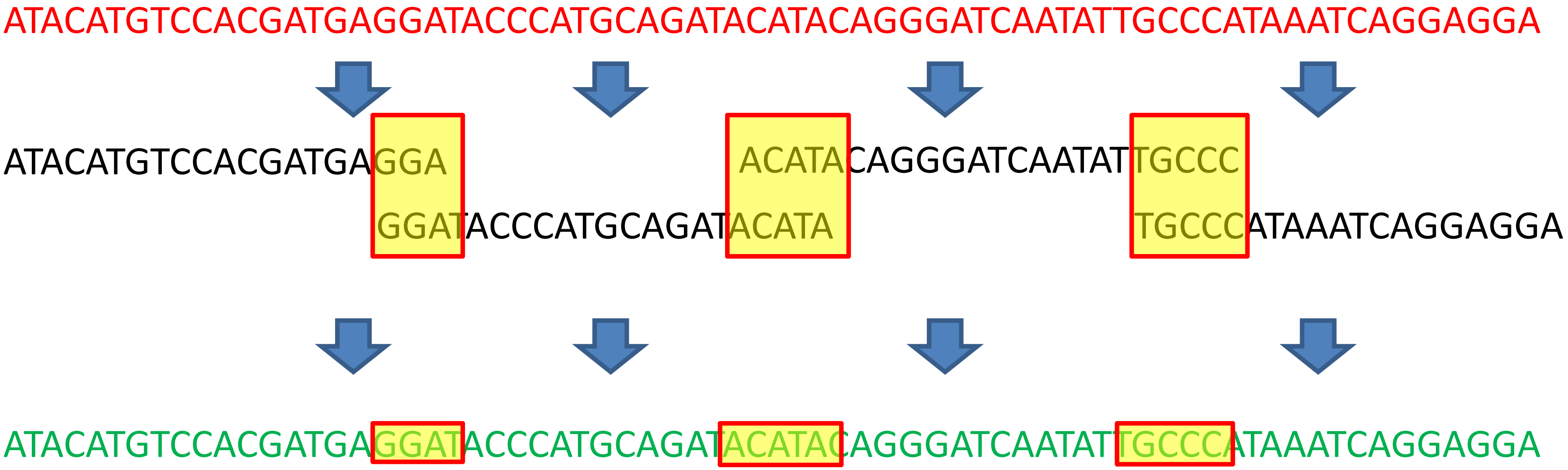
GGATACCCATGCAGATACATA

ACATACAGGGATCAATATTGCCC

TGCCCATAAATCAGGAGGA

HUMAN GENOME PROJECT: SHOTGUN SEQUENCING

The principle is to obtain a series of overlapping DNA fragments that can be connected into a continuous map.



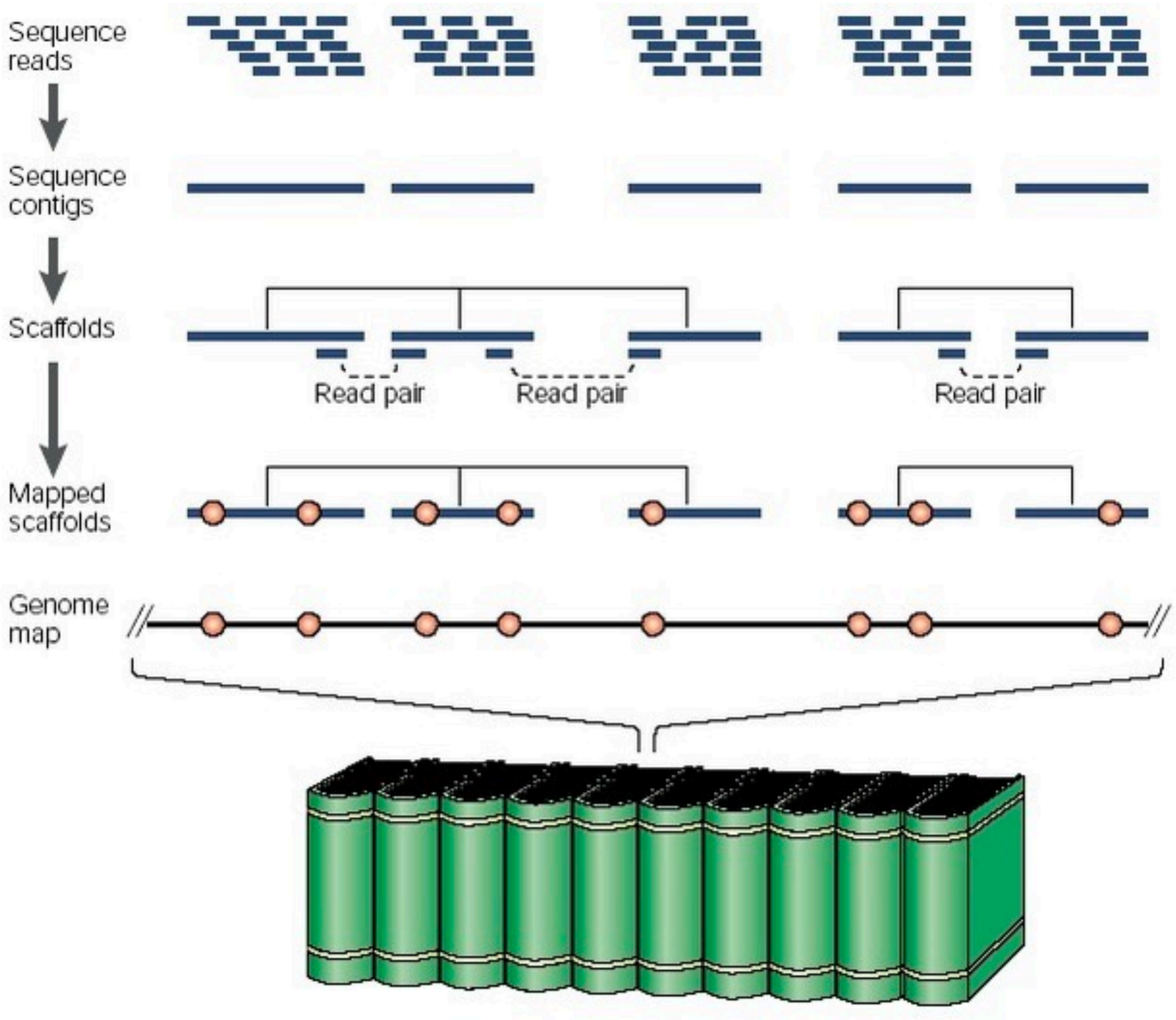
ATACATGTCCACGATGAGGATACCCATGCAGATACATACAGGGATCAATATTGCCCATAAATCAGGAGGA

ATACATGTCCACGATGAGGA ACATACAGGGATCAATATTGCCC

GGATACCCATGCAGATACATA TGCCCATAAATCAGGAGGA

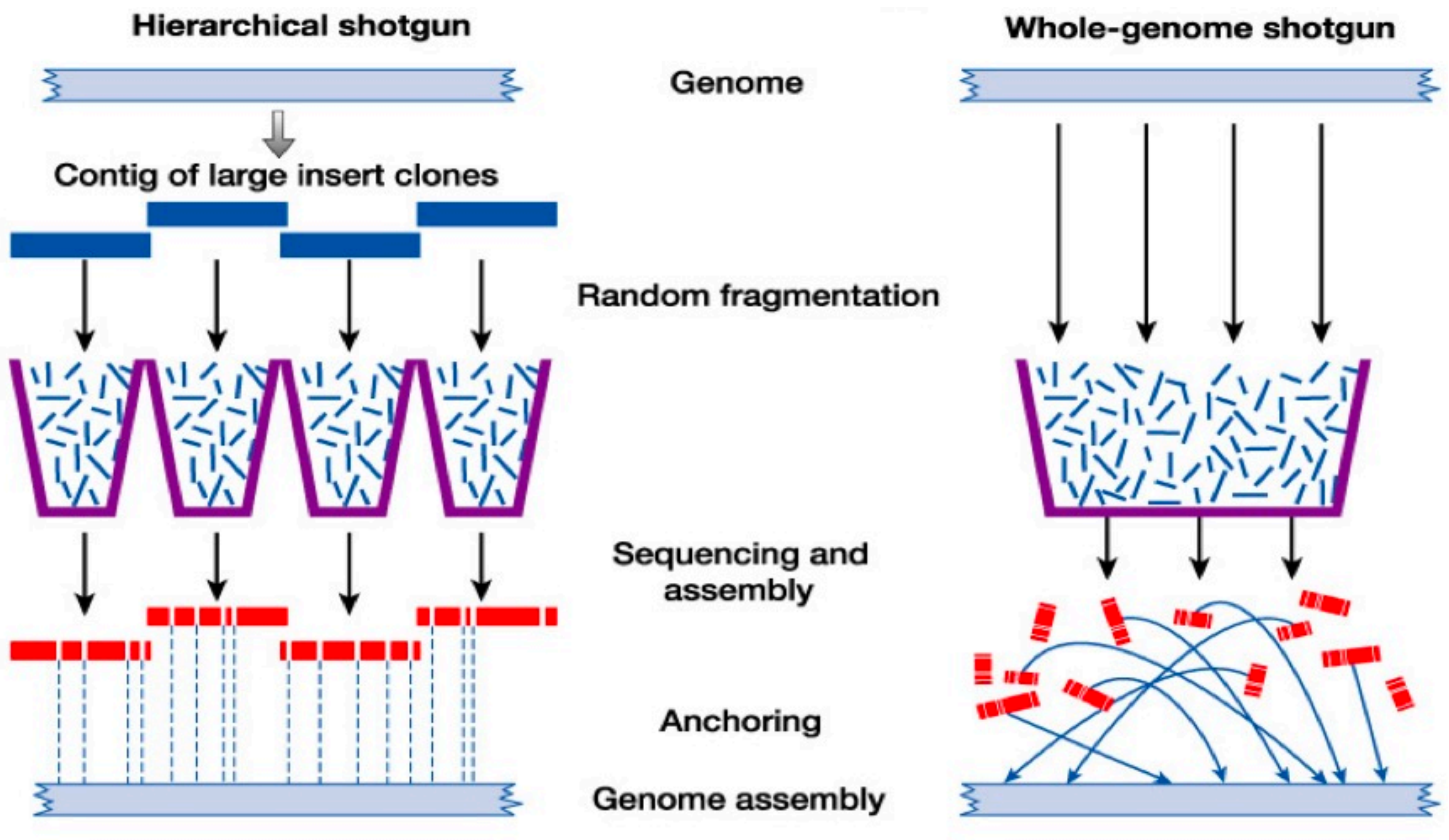
ATACATGTCCACGATGAGGATACCCATGCAGATACATACAGGGATCAATATTGCCCATAAATCAGGAGGA

HUMAN GENOME PROJECT: SHOTGUN SEQUENCING



- **"paired ends" sequencing**
- Sequence contigs from computational homology search
- "Scaffolds" use information from paired-end sequencing (not clone maps)
- More suitable for small genomes and/or those with few repetitive elements.

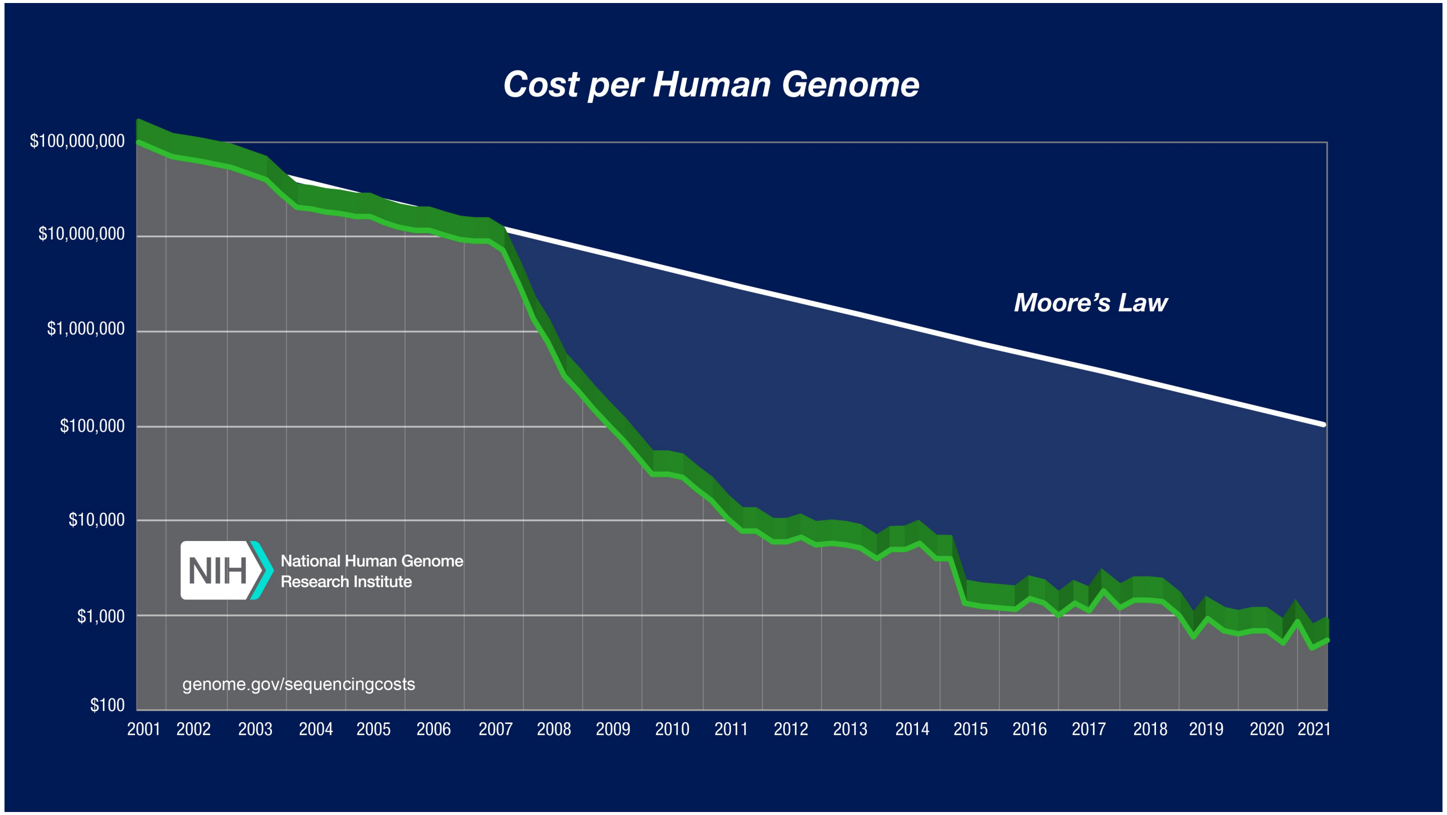
HUMAN GENOME PROJECT



The whole-genome shotgun approach simplifies and speeds up the preparation of a genomic library, making it more cost-effective. However, it requires more intensive computational processing. This has become feasible due to advancements in bioinformatic techniques and increased computational power.

SEQUENCING A HUMAN GENOME (3,2 BILLION BP)

300 million $



1000 $/genome

SEQUENCING A HUMAN GENOME (3,2 BILLION BP)

Costs and time for sequencing a human genome (3,2 billion bp)

2001	First human genome	13 years	300 million $
2005	Technology review	6 months	20-30 million $
2005	**454 Roche**	**1 month**	**900’000 $ (1X coverage)**
2009	Solexa (Illumina)	6 months	50’000 $ (30X coverage)
2010	Illumina		19’500 $ (30X coverage)
Today	**Personalized medicine**	Today 300$ (30x coverage)	

https://www.longdom.org/open-access/generations-of-sequencing-technologies-from-first-to-next-generation-0974-8369-1000395.pdf

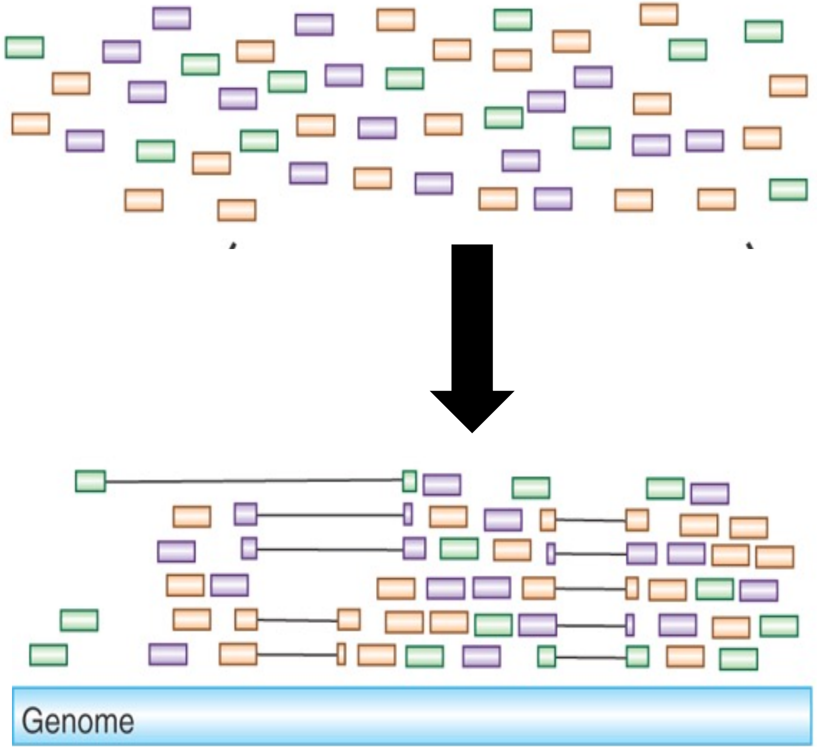
NEXT GENERATION SEQUENCING

What is it?

Set of new high throughput technologies:

- Allow millions of short DNA sequences from a biological sample to be "read" or sequenced in a rapid manner

- Computational power is then used to assemble or align the "reads" to a reference genome, allowing biologists to make comparisons and interpret various biological phenomena



- Due to high depth of coverage (30-100x), accurate sequencing is obtained much faster and cheaper compared to traditional Sanger/Shotgun sequencing

NEXT GENERATION SEQUENCING

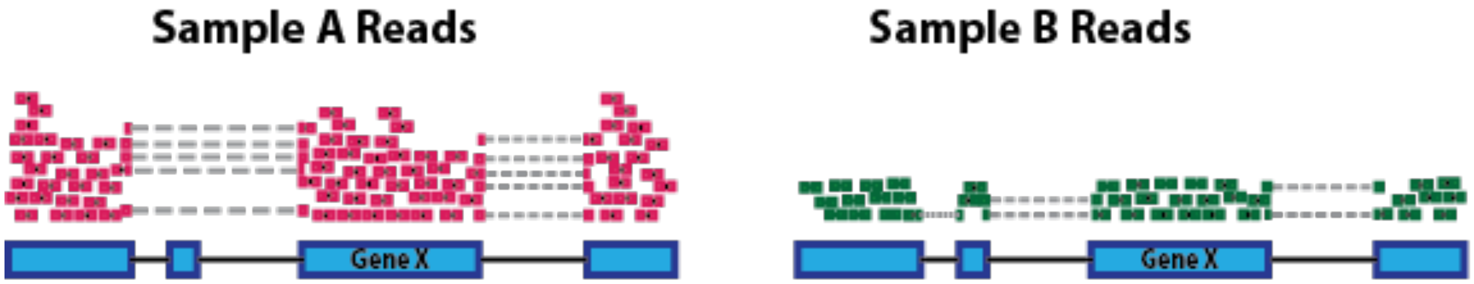
Just DNA sequencing or something more…

- **Mutation and SNP** identification or analysis (genome re-sequencing)
- Gene/Disease Linkage (genome re-sequencing)
- Pathogen identification (de novo sequence assembly or re-sequencing)
- DNA methylation study (medip-seq)
- Chromatin study (**ChIPseq**)
- Transcription factor study (ChIPseq)
- Genome structure (HiC)
- Transcriptome analysis (**RNAseq**)
- miRNAs, siRNA, piRNA, tRF, etc... (**small RNA seq**)
- Single cell transcriptome analysis

NEXT GENERATION SEQUENCING

Deep sequencing → Qualitative information

Deep sequencing → Quantitative information



Example: RNA-Seq

RNA-Seq

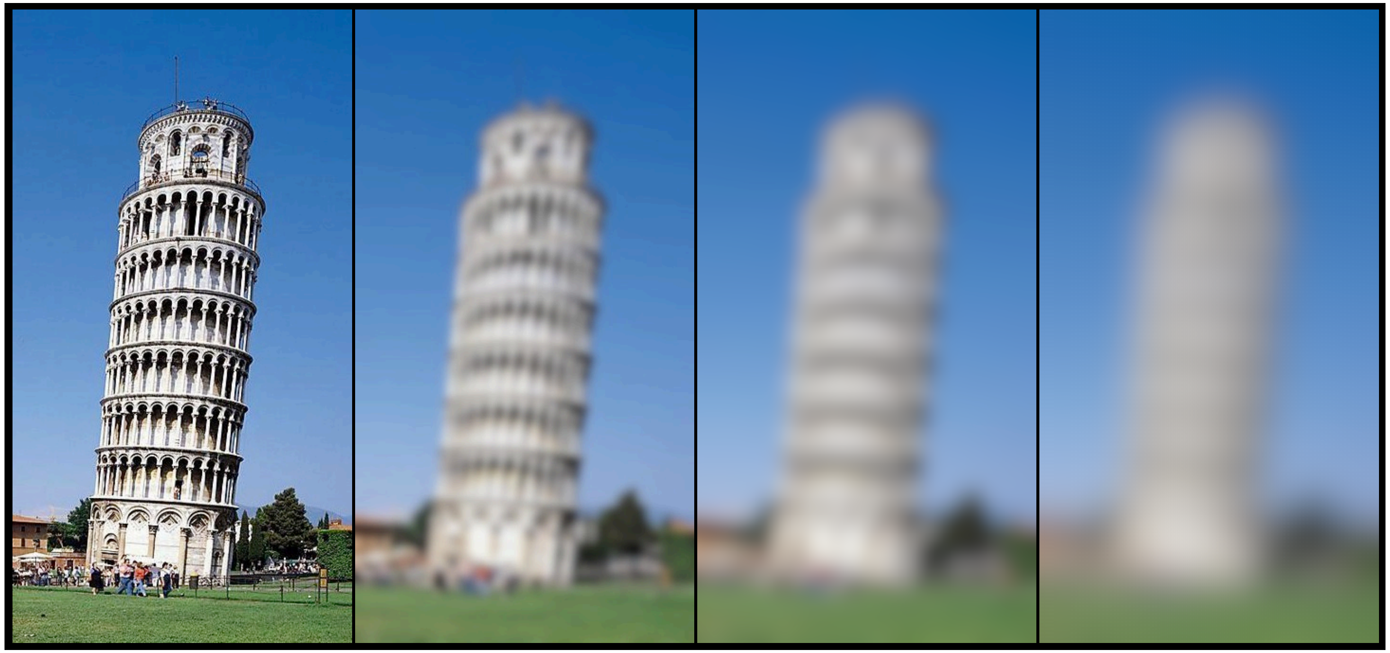
What is RNA-seq?

- RNA-seq is essentially **massively parallel sequencing of RNA** (or, in fact, the corresponding cDNA) and has heralded the second technical revolution in transcriptomics.
- It is **based on next-generation sequencing (NGS) platforms** that were initially developed for high-throughput sequencing of genomic DNA.
- Typically, **all the RNA molecules in a sample are reverse transcribed into cDNA,** and depending on the platform to be used, the **cDNA molecules may (amplification-based sequencing) or may not (single-molecule sequencing (SMS)) be amplified before deep sequencing.**
- After the sequencing reaction has taken place, **the obtained sequence stretches (reads) are mapped onto a reference genome** to deduce the structure and/or expression state of any given transcript in the sample.

NEXT GENERATION SEQUENCING

Sequencing Depth

How many reads to produce from a sample

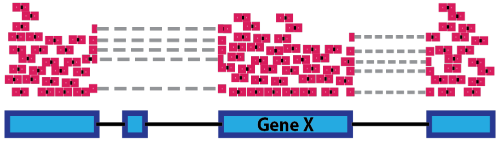


High resolution

↓

many information

many published human RNA-Seq experiments have been sequenced with a sequencing depth **between 20 M - 50 M reads per sample**



Low resolution

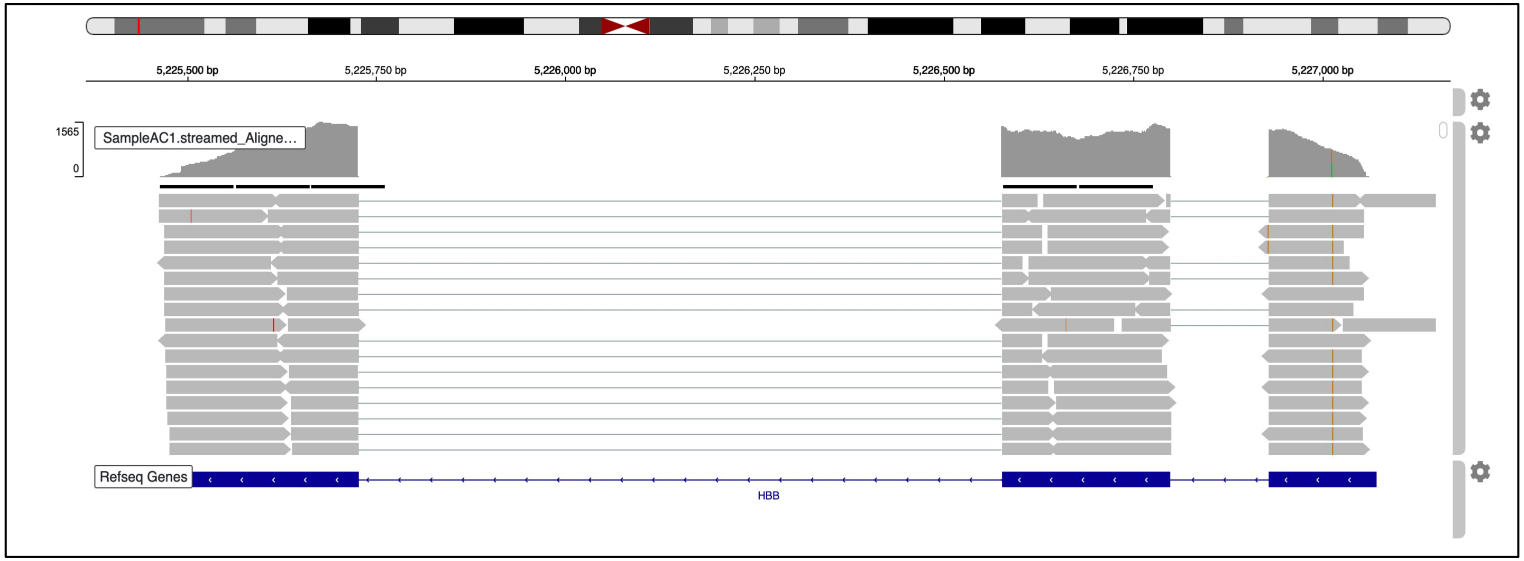
↓

few information

RNA-Seq

- Example of reads aligned to the reference genome

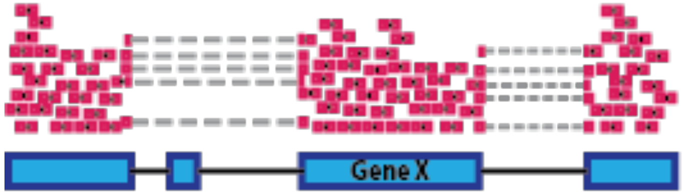
```
                                 AACAAATGAGACGCTGTGCAATTGCTGA
                                 AACAAATGAGACGCTGTGCAATTGCTGAGTAC
                                 AACAAATGAGACGCTGTGCAATTGCTGAGTACCGTA
                                     ATGAGACGCTGTGCAATTGCTGAGTACCGT
                                              CTGTGCAATTGCTGAGTACCGTAGGTAGAAC
                                              CTGTGCAATTGCTGAGTACCGTAGGTAGAAC                            reference genome
GCATCCATCTTGGGGCGTCCCAATTGCTGAGTAACAAATGAGACGCTGTGCAATTGCTGAGTACCGTAGGTAGAACCCCGCAGGGTTAACGACTCATTGTTTACTCTGCGAC
```



RNA-Seq

RNA-Seq provides the ability to look at:

- changes in **gene expression**
- **alternatively spliced transcripts**, alternative promoters and polyA sites
- **post-transcriptional changes**
- **gene fusions**
- In addition to mRNA transcripts, RNA-Seq can look at **different populations of RNA (tRNA, miRNA)**
- **exon/intron** boundaries
- verify or amend previously annotated 5’ and 3’ gene boundaries.

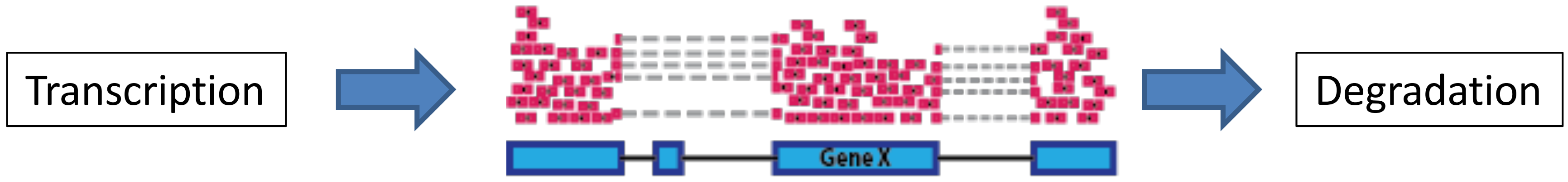


RNA-Seq

Deep sequencing → Qualitative information

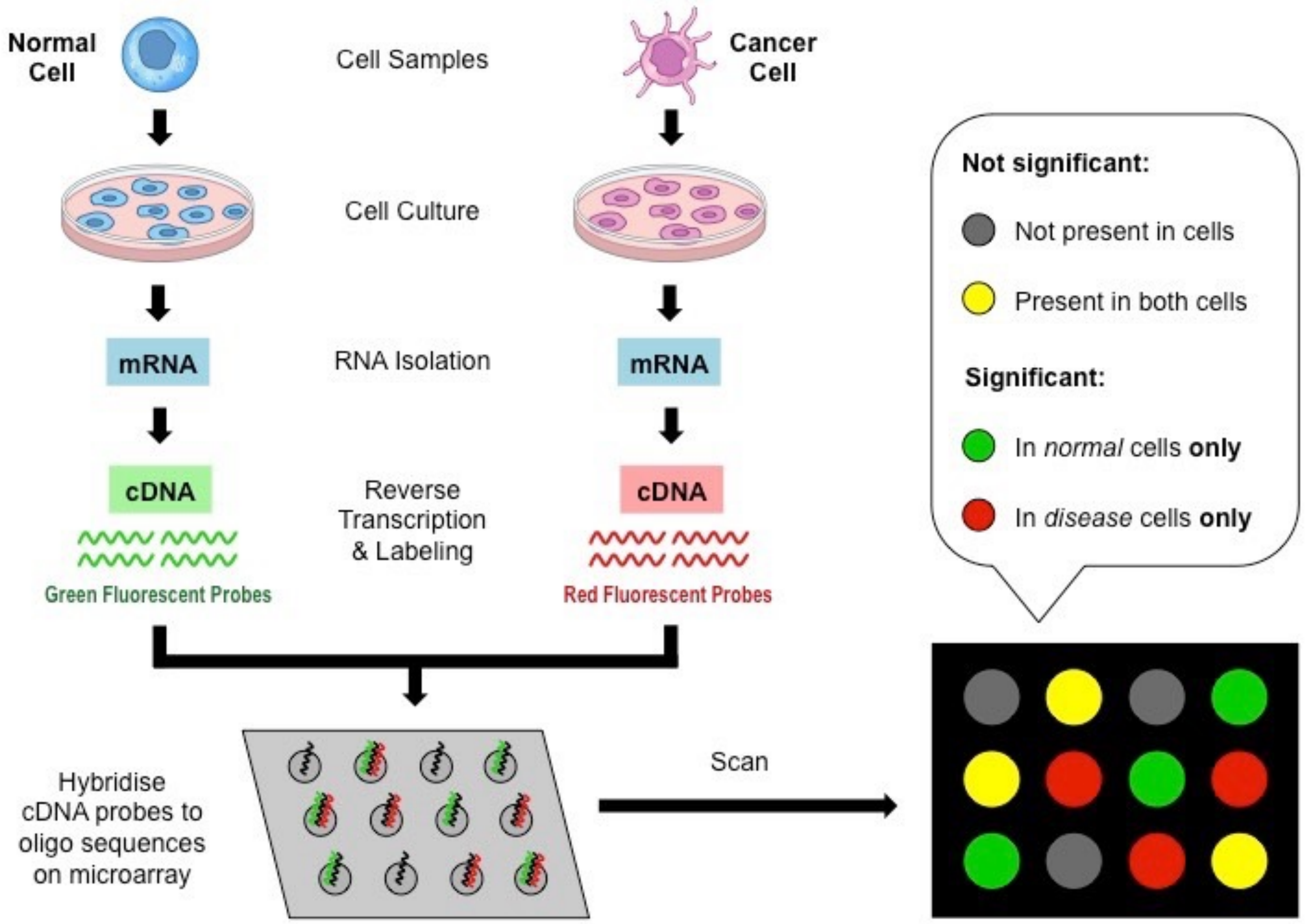
Deep sequencing → Quantitative information

RNA expression detected in a standard RNA-Seq is a «Steady state»



Transcription → → Degradation

RNA-Seq VS Microarray (what is it?)

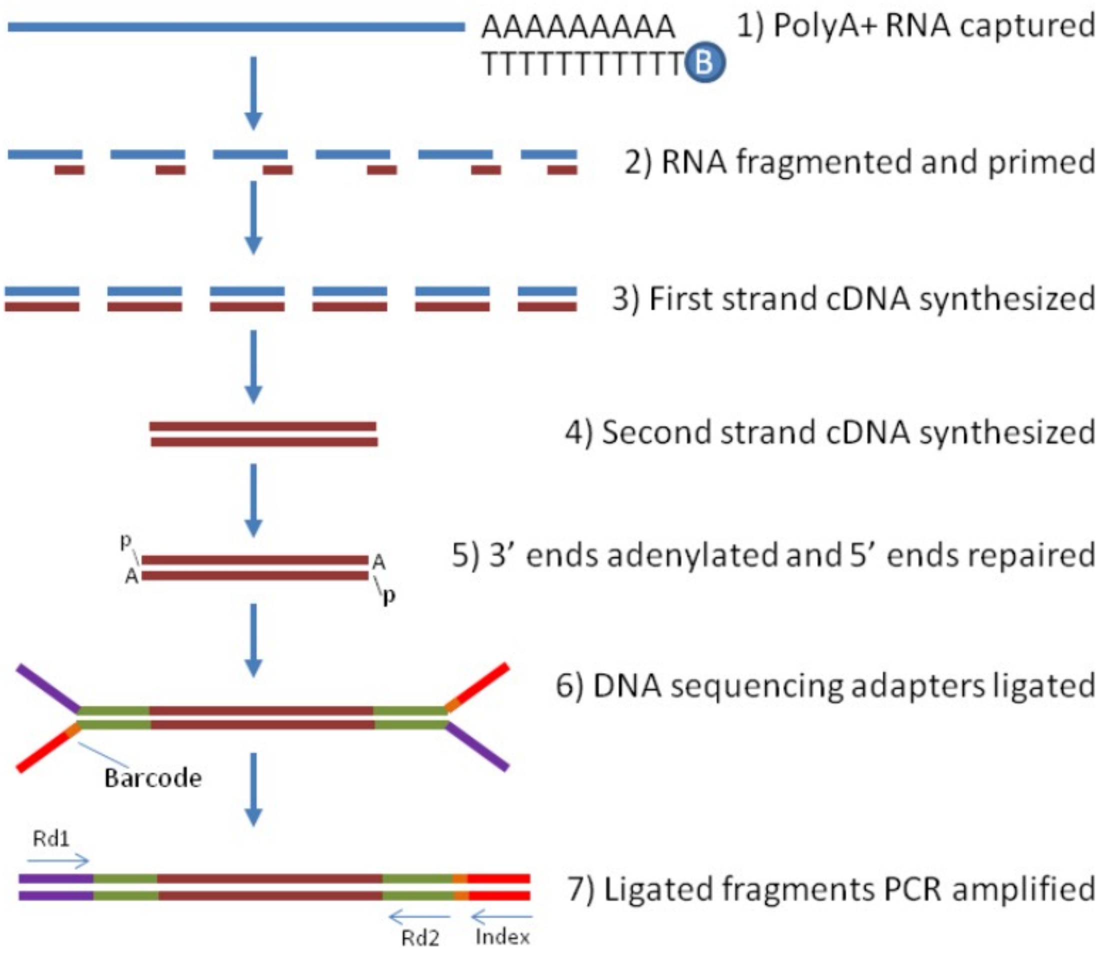


RNA-Seq VS Microarray

- **RNA-Seq has a wider dynamic range**, which depends on the sequencing depth. Microarrays show saturation at high expression levels and loss of signal at low expression levels.
- **RNA-Seq is more sensitive than microarrays**: it is able to identify more genes.
- RNA-Seq is able to **identify and quantify novel splicing variants**.
- RNA-Seq allows to **identifiy new SNPs and editing.**
- Microarray are cheaper and easier to analyze.
- Arrays still have a place for targeted identification of already known common allele variants, making them ideal for regulatory diagnostics.
- https://bioinfomagician.wordpress.com/2014/01/28/rna-seq-vs-microarray-what-is-the-take/comment-page-1/

RNA-Seq: LIBRARY PREPARATION

Example of library preparation: Illumina Truseq


AAAAAAAAA
TTTTTTTTTTT
B
1) PolyA+ RNA captured
2) RNA fragmented and primed
3) First strand cDNA synthesized
4) Second strand cDNA synthesized
p
A
A
p
5) 3’ ends adenylated and 5’ ends repaired
6) DNA sequencing adapters ligated
Barcode
Rd1
7) Ligated fragments PCR amplified
Rd2
Index

RNA-Seq: LIBRARY PREPARATION

coding RNAs		**mRNA**		**RNA-seq**
non-coding RNAs	large	**rRNA**		
		Xist		
		lincRNA		
		Pseudogenes		
		circular RNAs		
		……………		
	small	**tRNA**	translation	
		snRNAs	splicing	
		snoRNAs	modification	
		scRNAs	transl. control	
		gRNAs	editing	
		miRNAs	transl. control	**Small RNA-seq**
		siRNAs	RNA stability	
		rasiRNAs	chromatin	
		piRNAs	genome stability	
		……………		

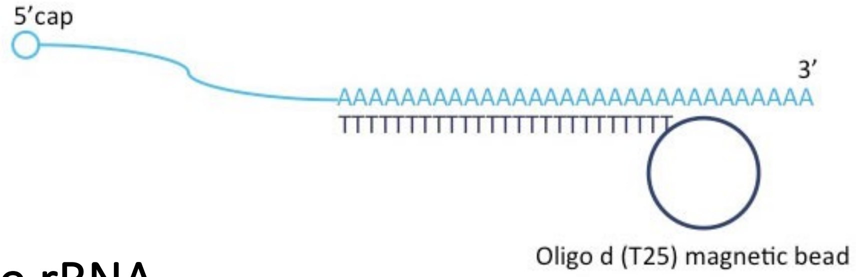
rRNA + tRNA → ~ 95%

RNA-Seq: LIBRARY PREPARATION

Two ways to isolate long RNA molecules:

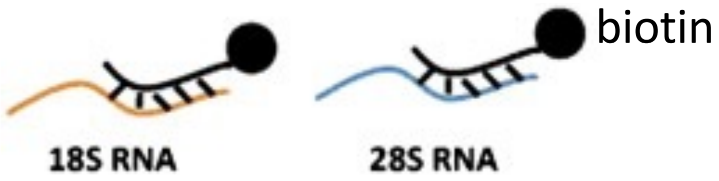
1a - Purify and Fragment mRNA

This process purifies the poly-A containing RNA molecules (mainly mRNA) using poly-T oligo-attached magnetic beads.


5'cap
3'
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
TTTTTTTTTTTTTTTTTTTTTTTTT
Oligo d (T25) magnetic bead

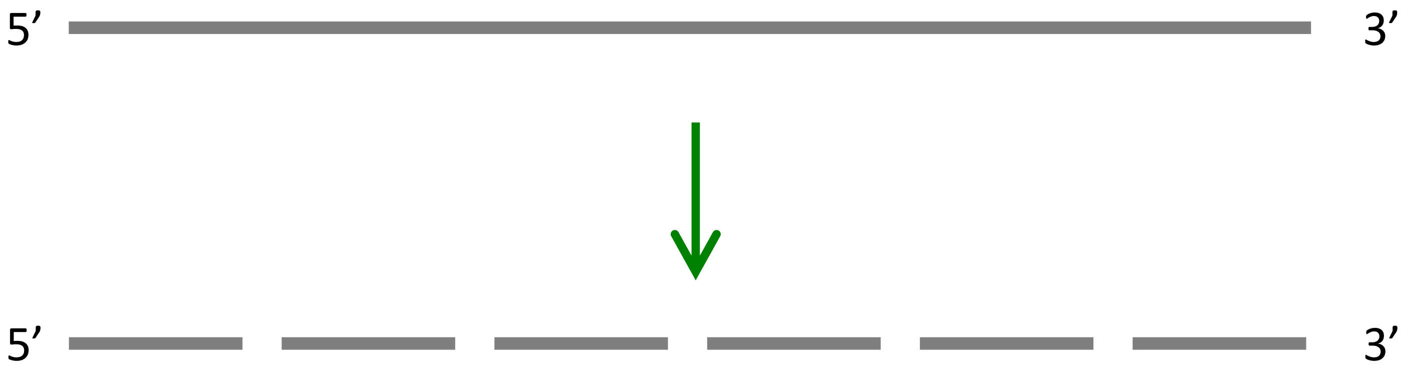
1b - Remove rRNA

After the ribosomal RNA is depleted, the remaining RNA (not only mRNA) is purified, fragmented and primed for cDNA synthesis. rRNA is removed using a hybridization-based technique.



RNA-Seq: LIBRARY PREPARATION

2 - RNA fragmentation
RNA molecules are fragmented into small pieces using divalent cations under elevated temperature

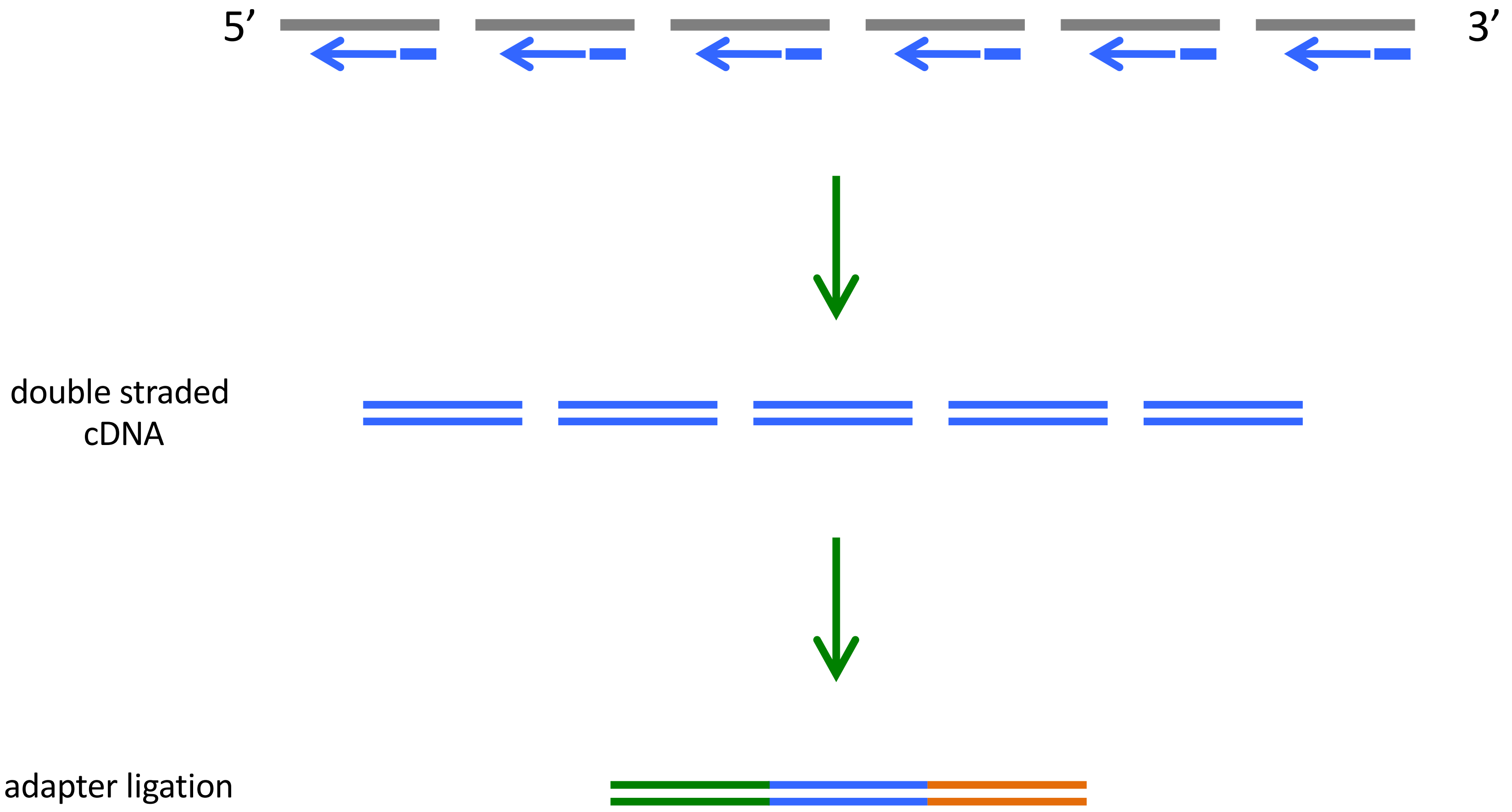


Range of fragments length: **120-225 bp**

RNA-Seq: LIBRARY PREPARATION

3 - Synthesize First Strand cDNA

This process reverse transcribes the cleaved RNA fragments that were primed with random hexamers into first strand cDNA using reverse transcriptase and random primers.

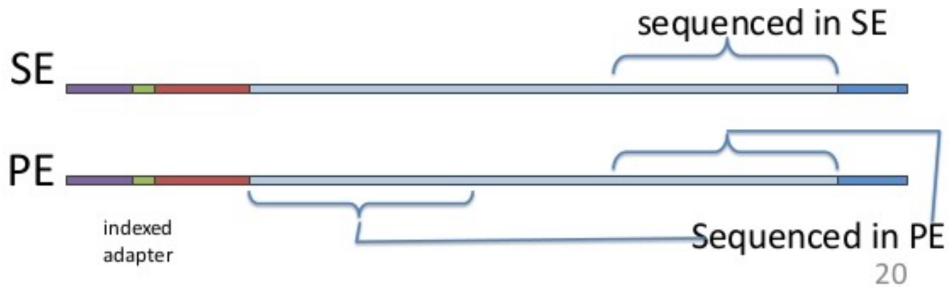


RNA-Seq: LIBRARY PREPARATION

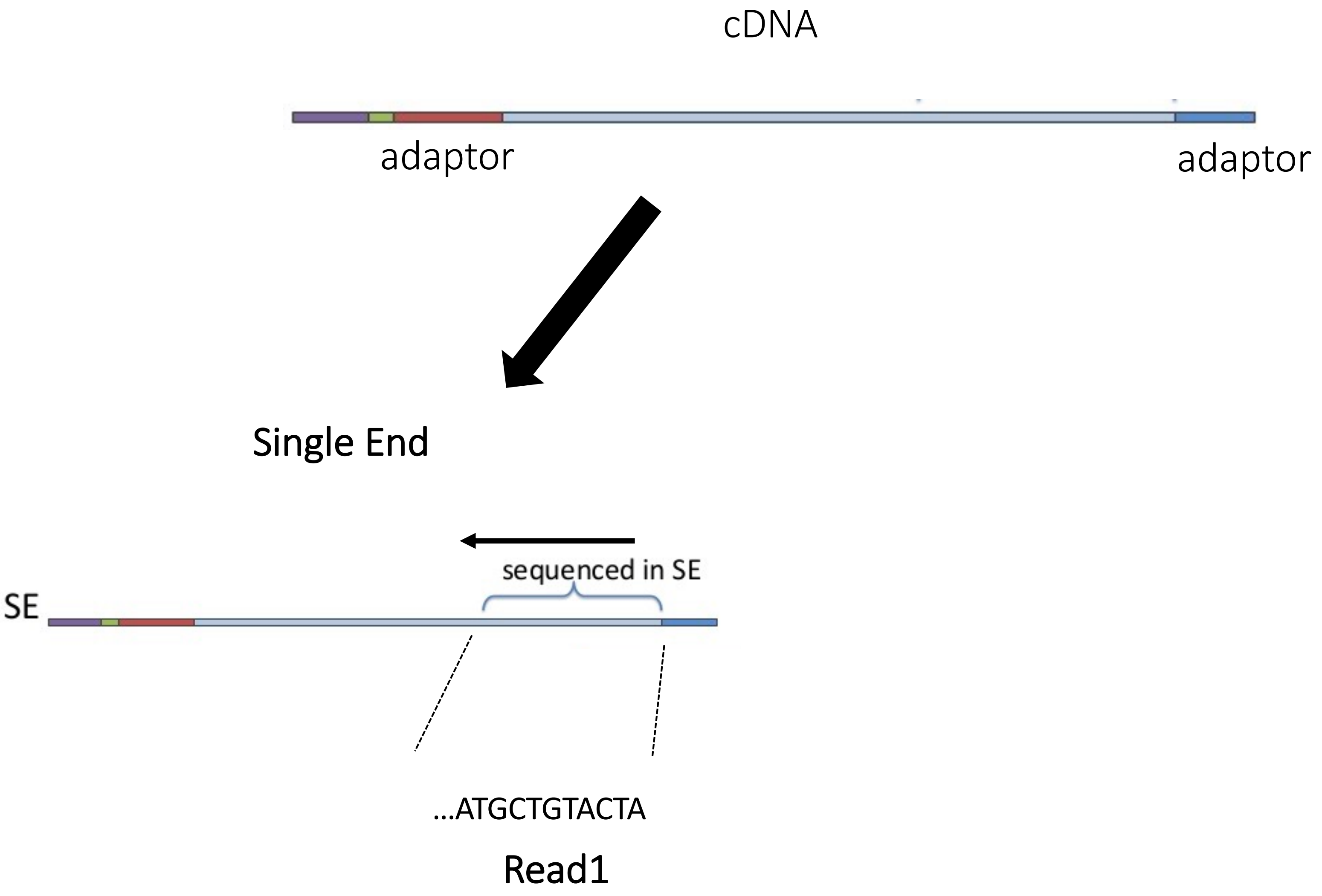
Single-end VS paired-end sequencing

- **Single-end sequencing (SE)**, involves sequecing of the fragment from only one end.

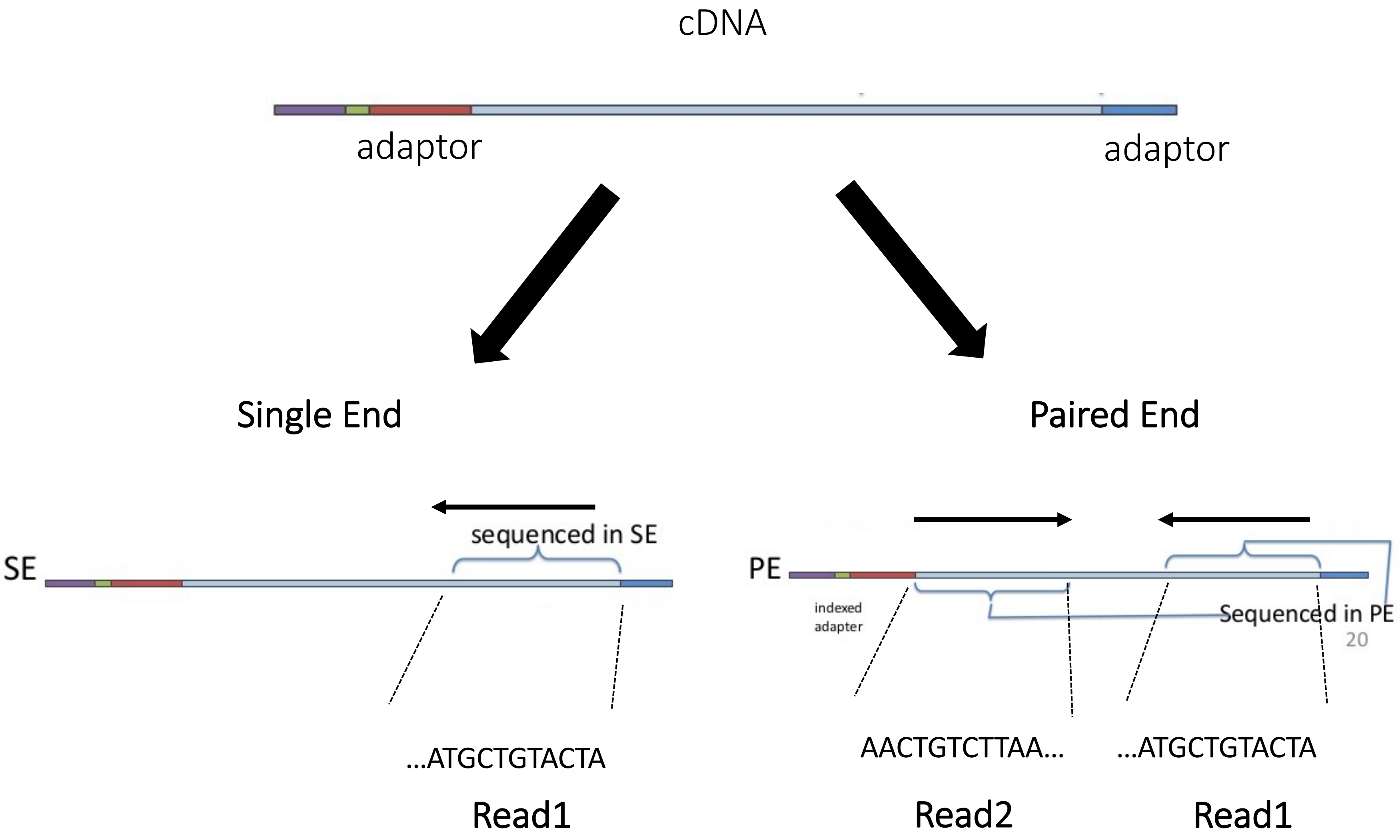
- **Paired-end sequecing (PE)**, involves sequencing both ends of a fragment, resulting in the production of read pairs. This allows to improve the alignment, to better identify and quantify splicing variants, and to detect rearrangements such as insertions, deletions, and inversions.



RNA-Seq: LIBRARY PREPARATION

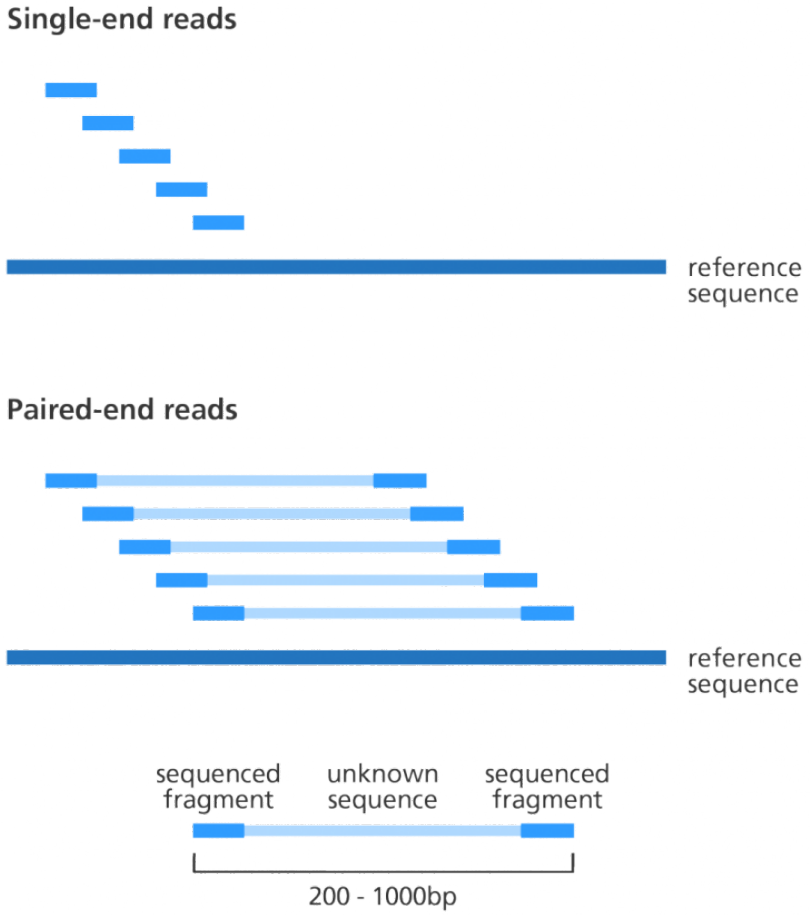


RNA-Seq: LIBRARY PREPARATION

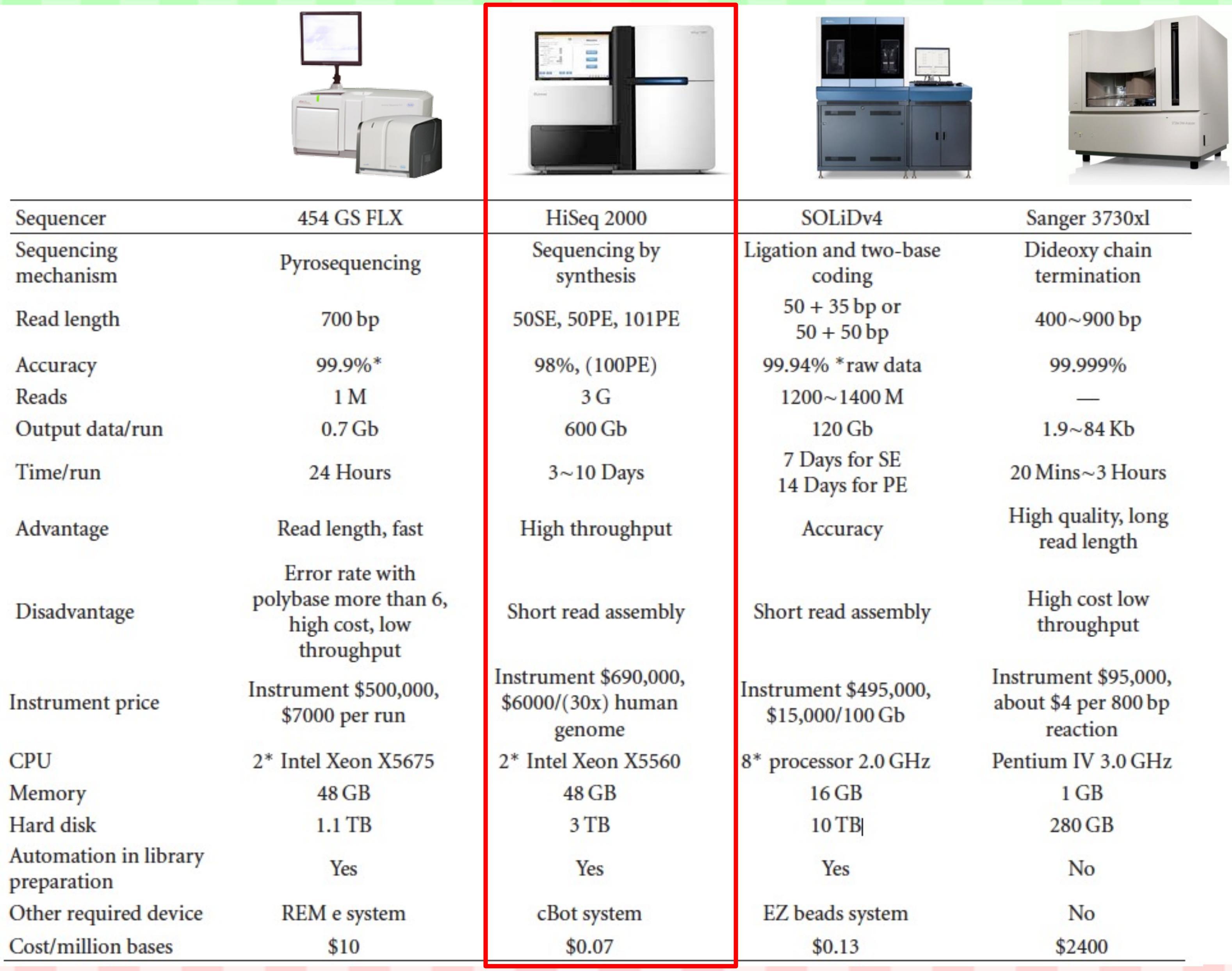


RNA-Seq: LIBRARY PREPARATION

Single-end VS paired-end sequencing



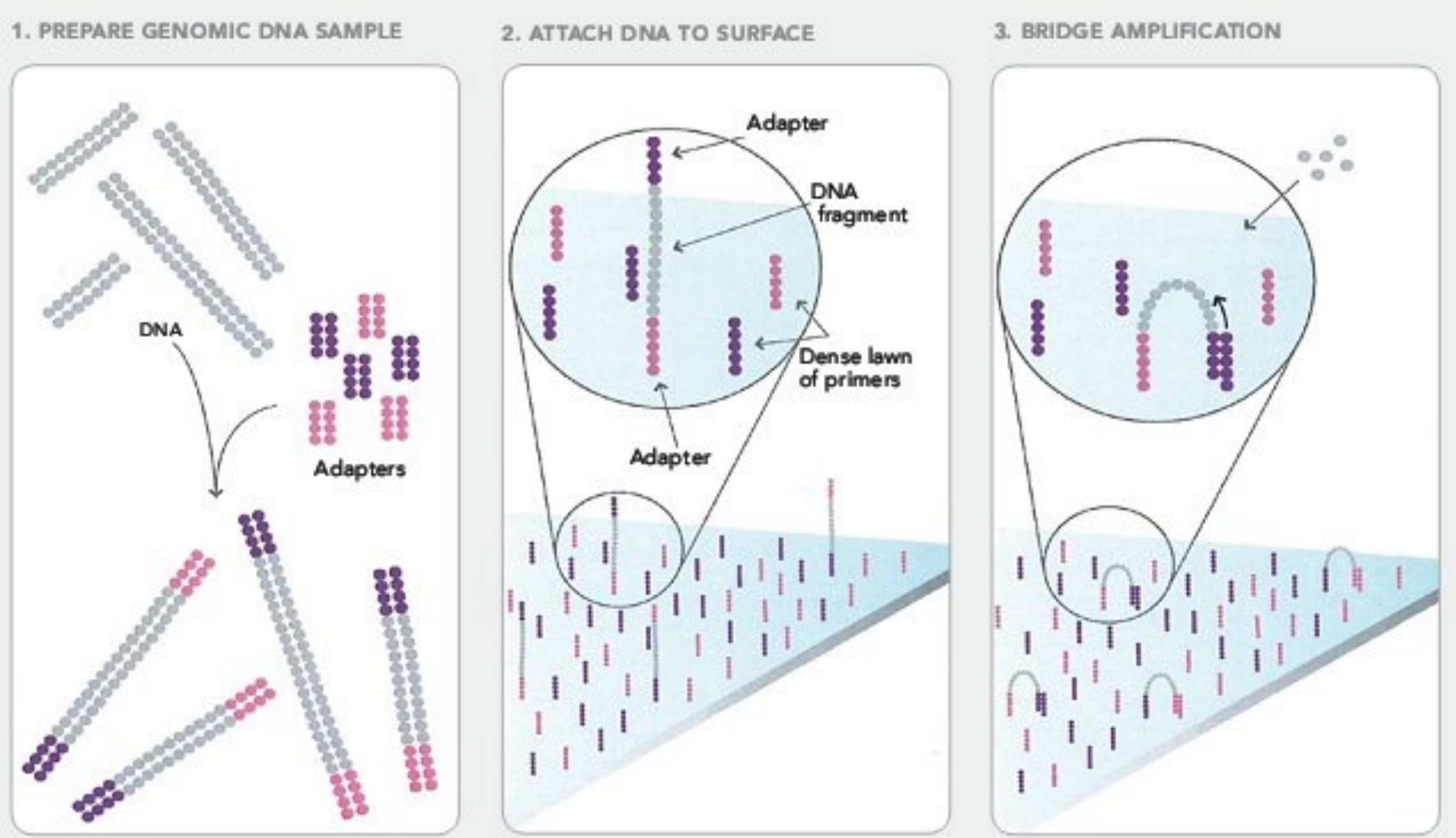
RNA-Seq: SEQUENCING REACTION



Sequencer	454 GS FLX	HiSeq 2000	SOLiDv4	Sanger 3730xl
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Ligation and two-base coding	Dideoxy chain termination
Read length	700 bp	50SE, 50PE, 101PE	50 + 35 bp or 50 + 50 bp	400~900 bp
Accuracy	99.9%*	98%, (100PE)	99.94% *raw data	99.999%
Reads	1 M	3 G	1200~1400 M	—
Output data/run	0.7 Gb	600 Gb	120 Gb	1.9~84 Kb
Time/run	24 Hours	3~10 Days	7 Days for SE 14 Days for PE	20 Mins~3 Hours
Advantage	Read length, fast	High throughput	Accuracy	High quality, long read length
Disadvantage	Error rate with polybase more than 6, high cost, low throughput	Short read assembly	Short read assembly	High cost low throughput
Instrument price	Instrument $500,000, $7000 per run	Instrument $690,000, $6000/(30x) human genome	Instrument $495,000, $15,000/100 Gb	Instrument $95,000, about $4 per 800 bp reaction
CPU	2* Intel Xeon X5675	2* Intel Xeon X5560	8* processor 2.0 GHz	Pentium IV 3.0 GHz
Memory	48 GB	48 GB	16 GB	1 GB
Hard disk	1.1 TB	3 TB	10 TB	280 GB
Automation in library preparation	Yes	Yes	Yes	No
Other required device	REM e system	cBot system	EZ beads system	No
Cost/million bases	$10	$0.07	$0.13	$2400

RNA-Seq: SEQUENCING REACTION

Illumina platform: Sequencing by Synthesis



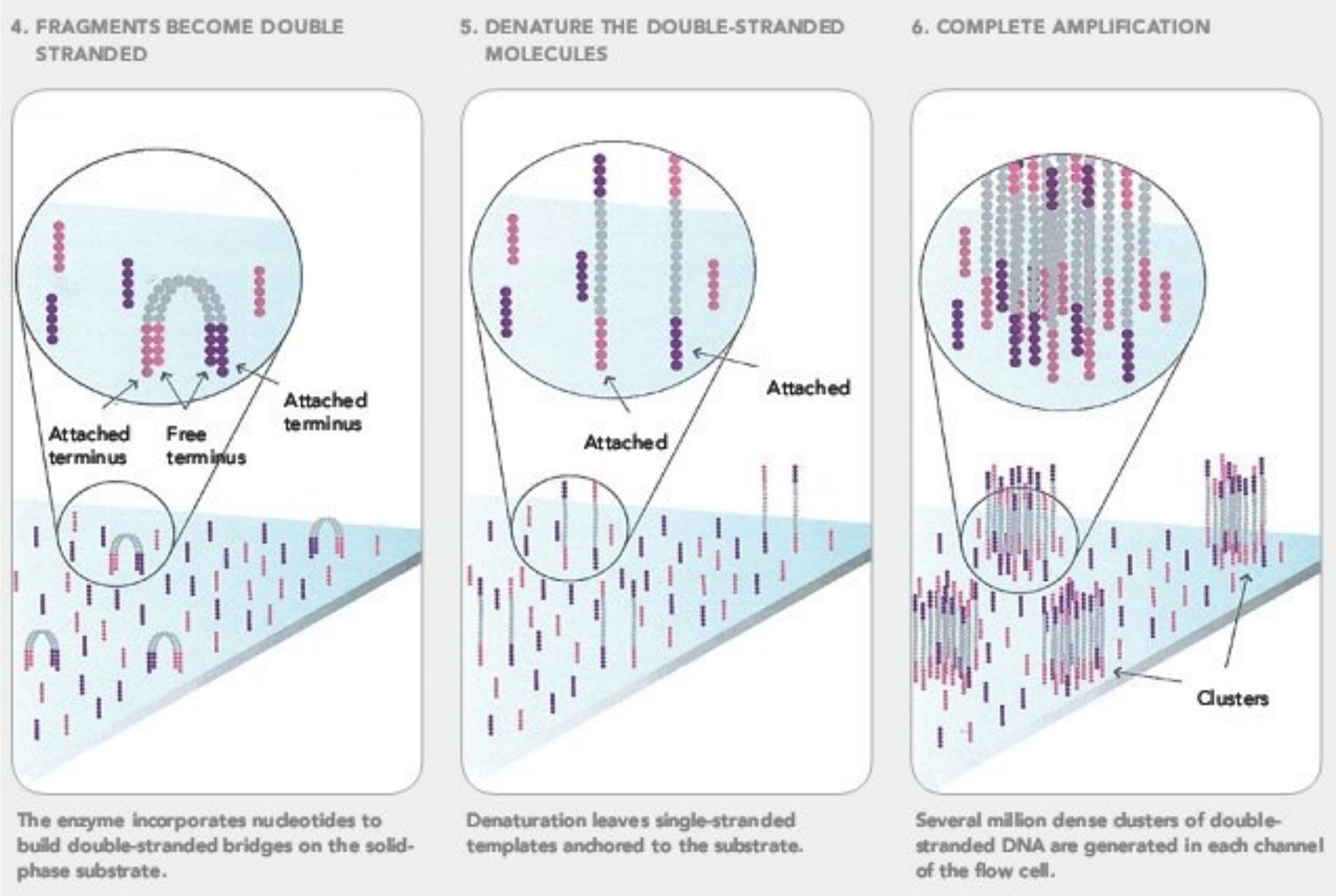
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

RNA-Seq: SEQUENCING REACTION

Illumina platform: Sequencing by Synthesis



4. FRAGMENTS BECOME DOUBLE STRANDED

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

5. DENATURE THE DOUBLE-STRANDED MOLECULES

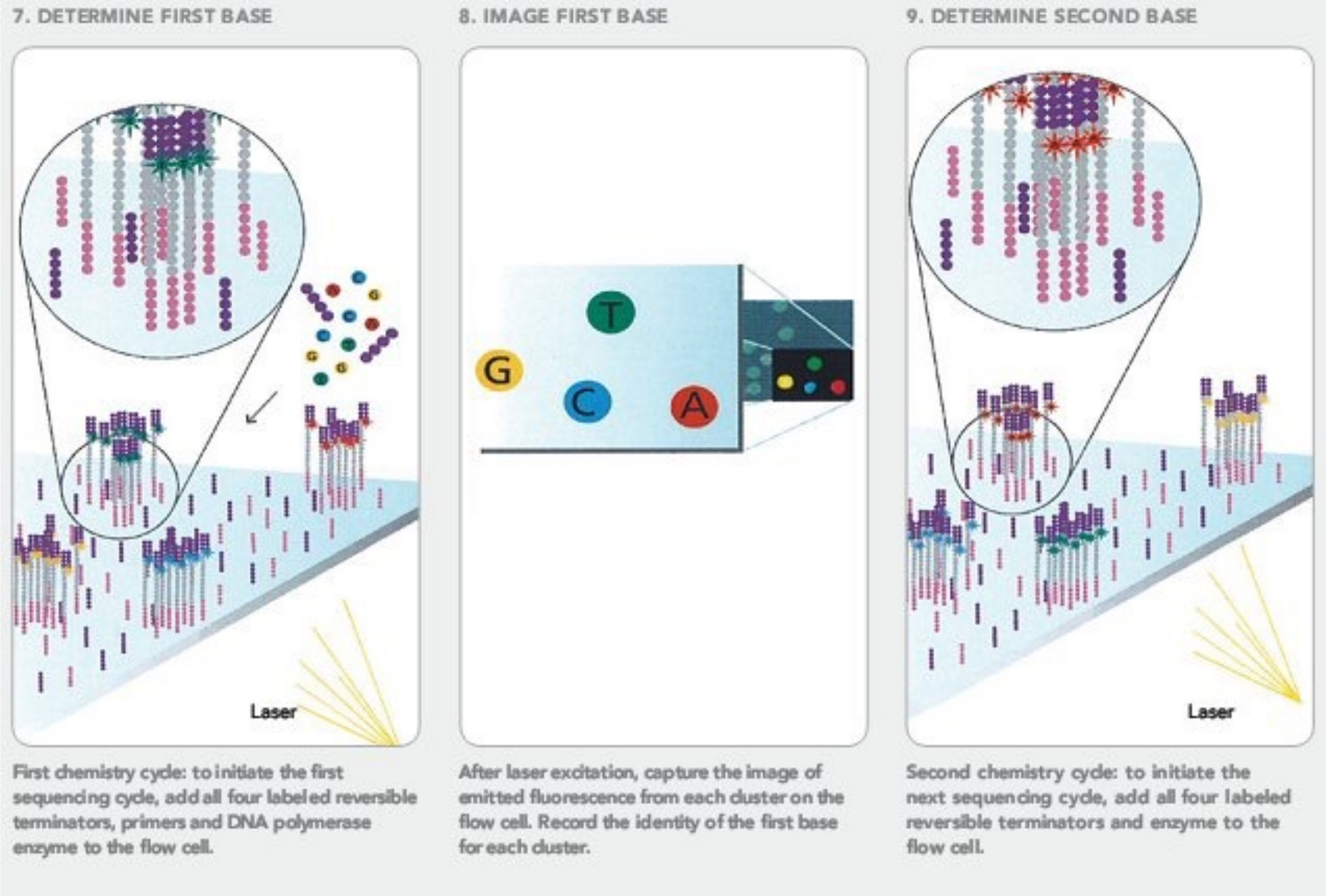
Denaturation leaves single-stranded templates anchored to the substrate.

6. COMPLETE AMPLIFICATION

Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

RNA-Seq: SEQUENCING REACTION

Illumina platform: Sequencing by Synthesis



7. DETERMINE FIRST BASE

First chemistry cycle: to initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.

8. IMAGE FIRST BASE

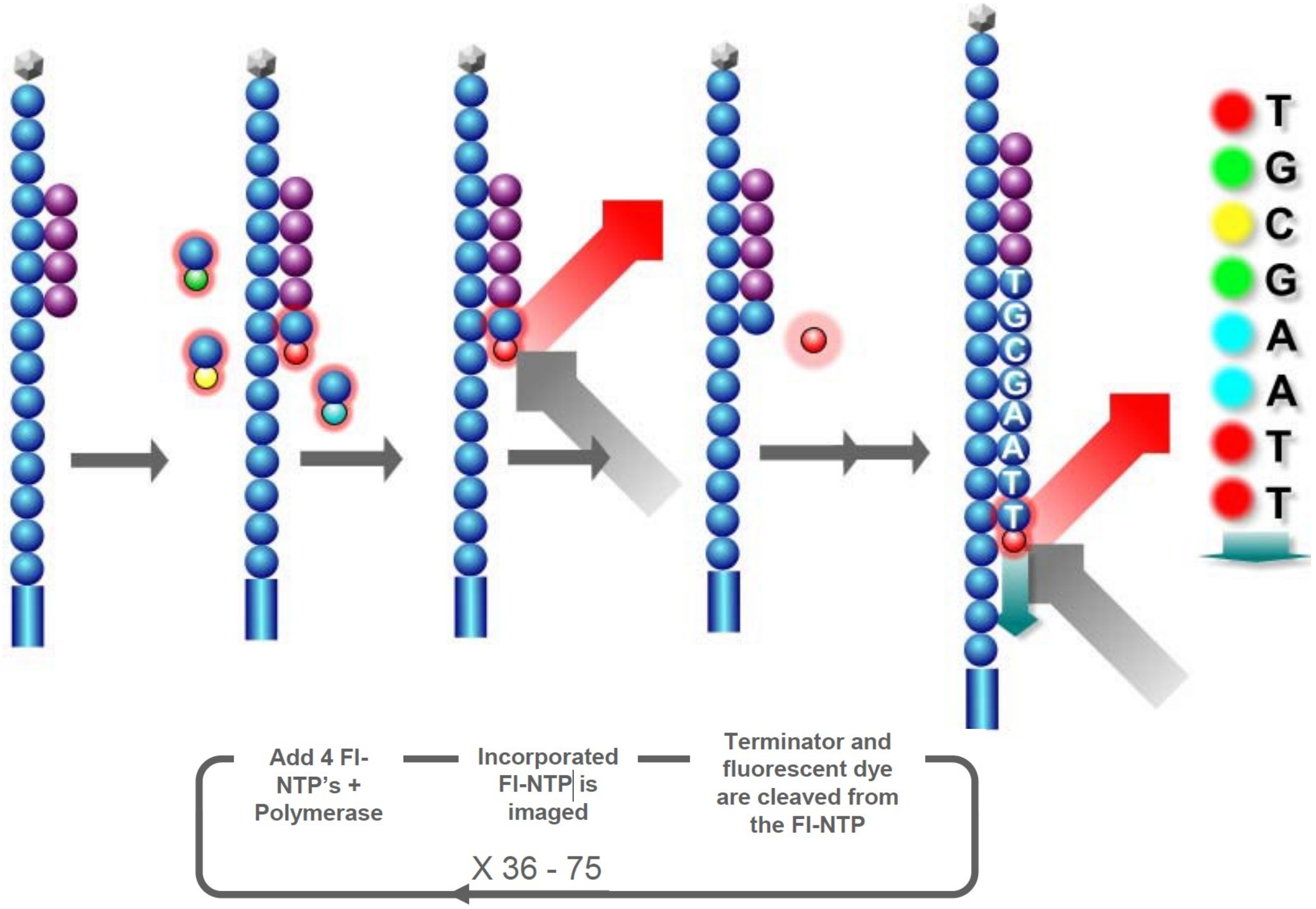
After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

9. DETERMINE SECOND BASE

Second chemistry cycle: to initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.

TOTAL READS NUMBER = Number of clusters in flow cell

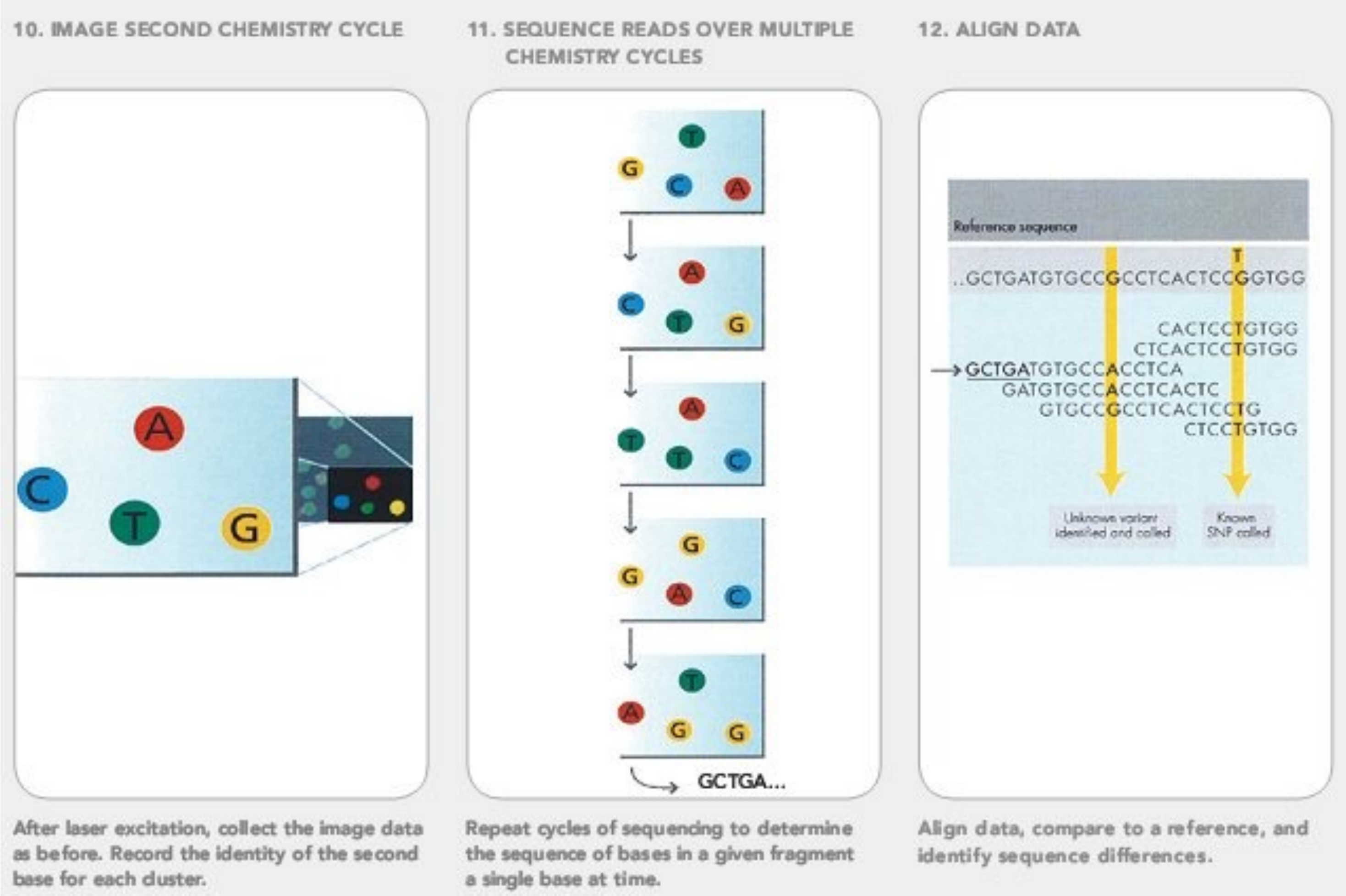
RNA-Seq: SEQUENCING REACTION


T
G
C
G
A
A
T
T
T
G
C
G
A
A
T
T
Add 4 Fl-NTP's + Polymerase
Incorporated Fl-NTP is imaged
Terminator and fluorescent dye are cleaved from the Fl-NTP
X 36 - 75

READ LENGTH = Number of reaction cycles

WET LAB: Reazione di Sequenziamento

Illumina platform: Sequencing by Synthesis



10. IMAGE SECOND CHEMISTRY CYCLE

After laser excitation, collect the image data as before. Record the identity of the second base for each cluster.

11. SEQUENCE READS OVER MULTIPLE CHEMISTRY CYCLES

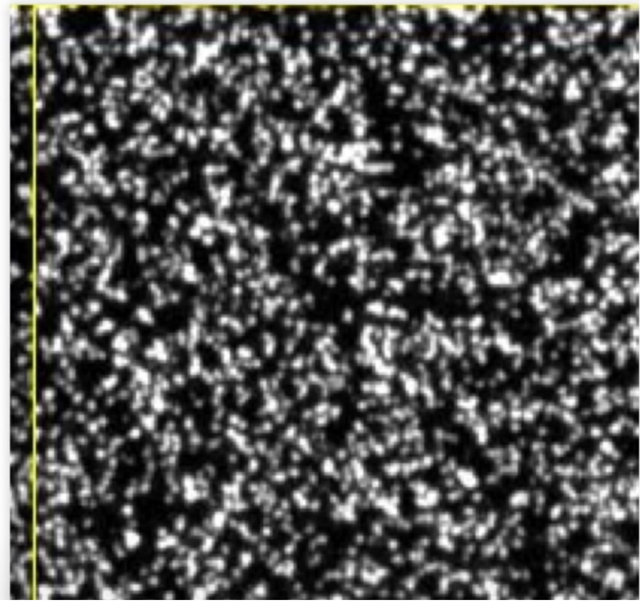
Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at time.

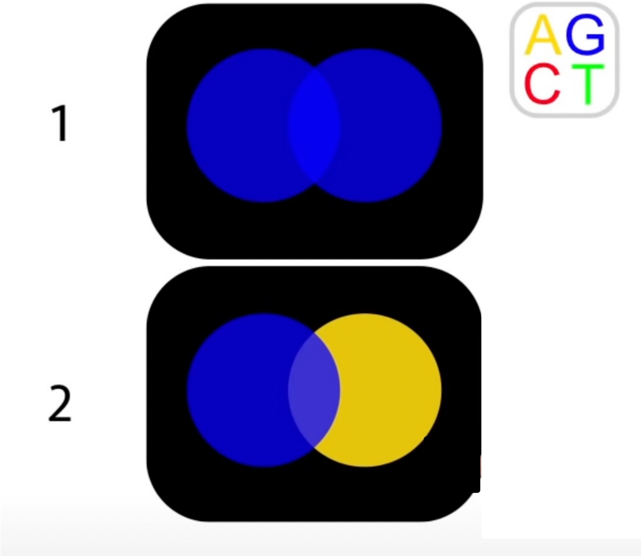
12. ALIGN DATA

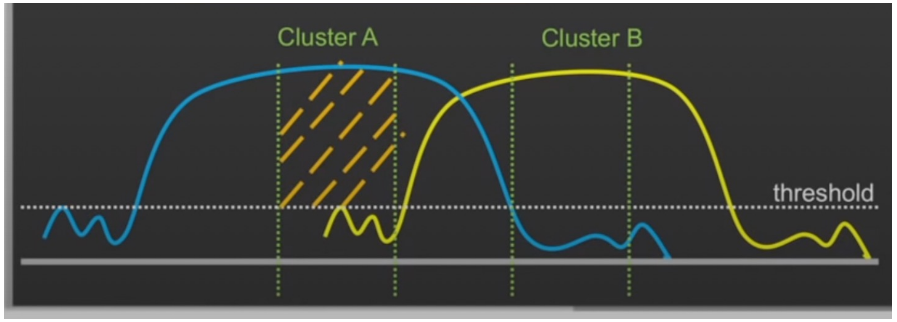
Align data, compare to a reference, and identify sequence differences.

RNA-Seq: SEQUENCING REACTION

Illumina platform: Sequencing by Synthesis

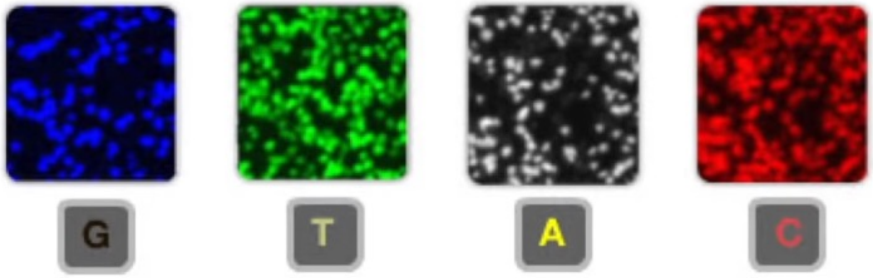


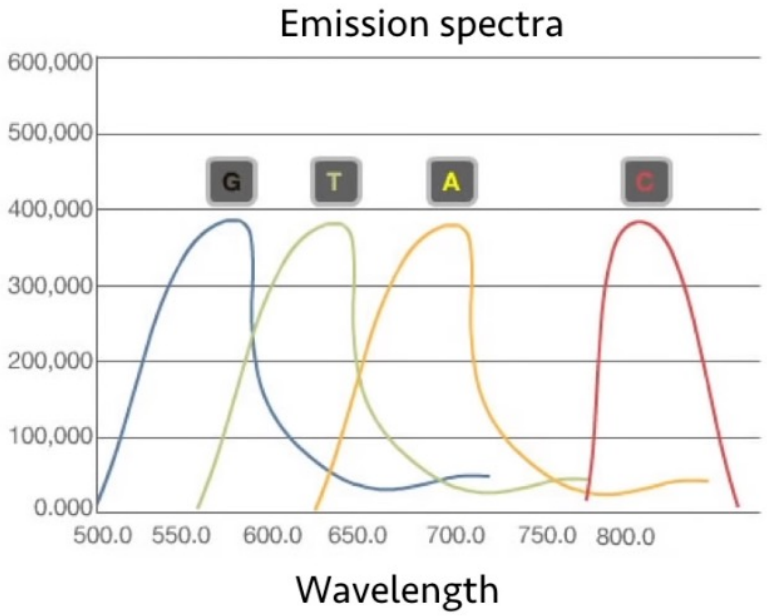




RNA-Seq: SEQUENCING REACTION

4 colour chemistry



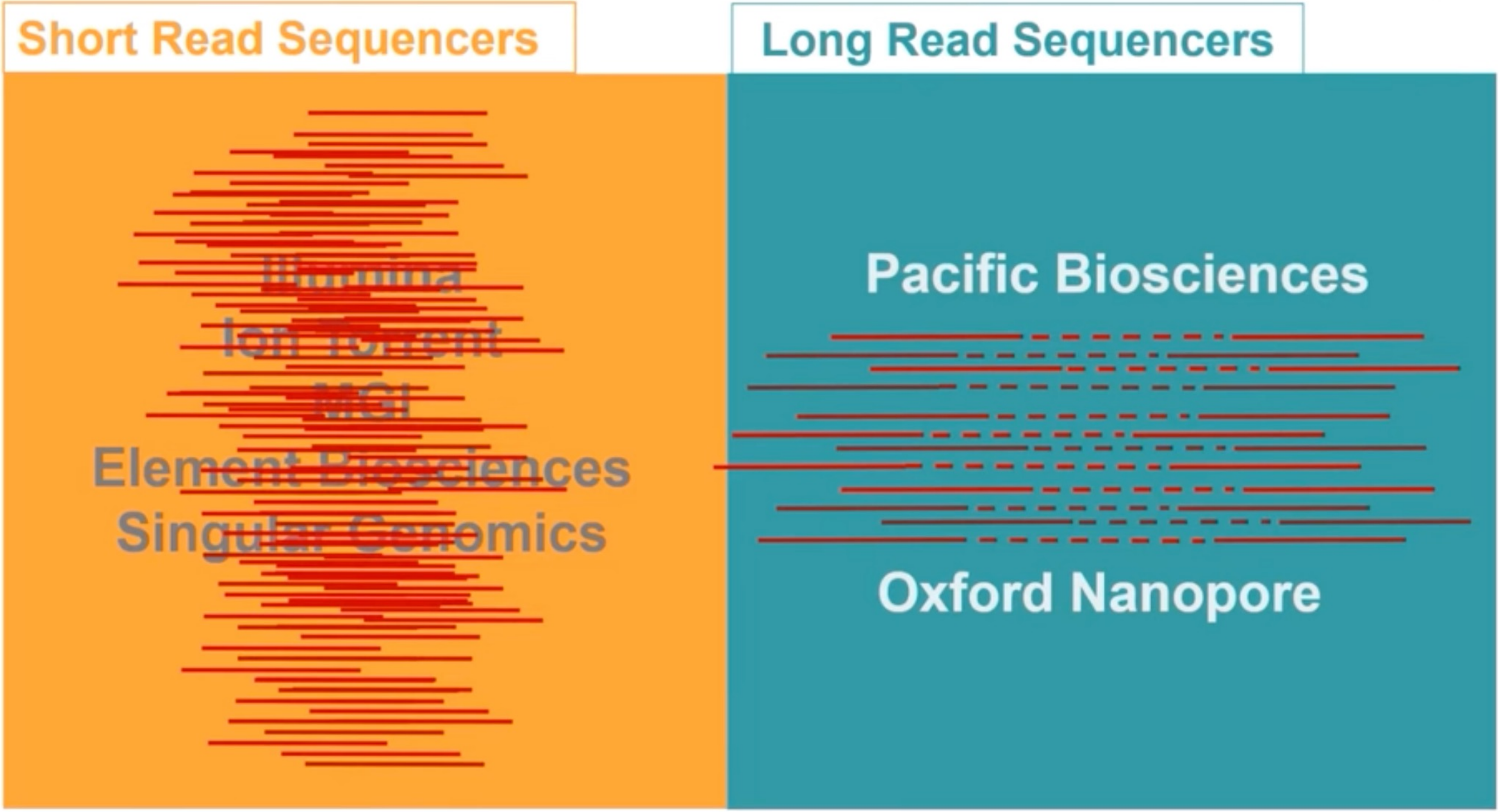


Illumina sequencing

Video:

https://www.youtube.com/watch?annotation_id=annotation_228575861&feature=iv&src_vid=womKfikWlxM&v=fCd6B5HRaZ8

Short reads vs Long reads



Short Read Sequencers

Illumina

Ion Torrent

MGI

Element Biosciences

Singular Genomics

Long Read Sequencers

Pacific Biosciences

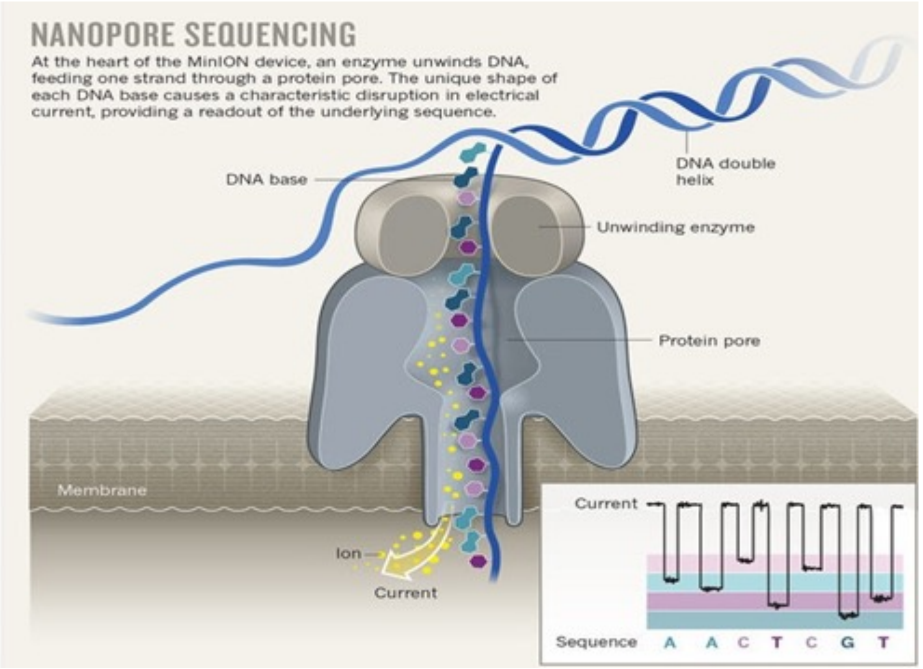
Oxford Nanopore

short but many reads

extremely loo...ong but not many reads

NANOPORE

- **MinIon** Oxford Nanopore



NANOPORE SEQUENCING

At the heart of the MinION device, an enzyme unwinds DNA, feeding one strand through a protein pore. The unique shape of each DNA base causes a characteristic disruption in electrical current, providing a readout of the underlying sequence.

- High error rates (10-15%)
  - Biased errors
- Really long reads (2 Mb)
- Can directly sequence RNA
- Maybe proteins in the future?

Easy sample preparation

Fast (450bases/sec) and cheap

Realtime data

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

NANOPORE

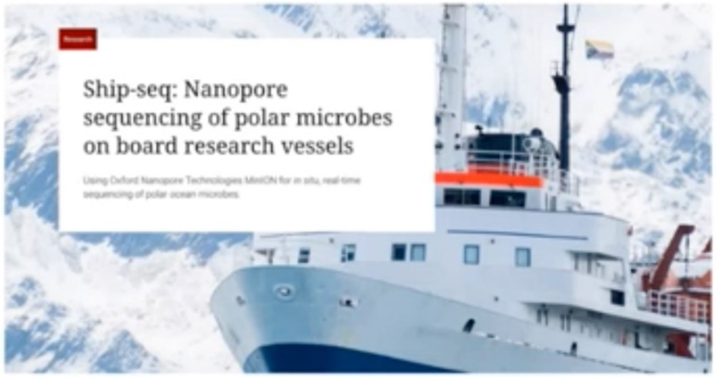
Nanopore is extremely portable

Nature **521,** 15–16 (07 May 2015)

NANOPORE

Nanopore is extremely portable

MinION: field deployable



Research

Ship-seq: Nanopore sequencing of polar microbes on board research vessels

Using Oxford Nanopore Technologies MinION for in situ, real-time sequencing of polar ocean microbes

MUSE/Science Museum of Trento

The MinION device can sequence small genomes, such as those of bacteria and viruses, displaying the results as they are generated.

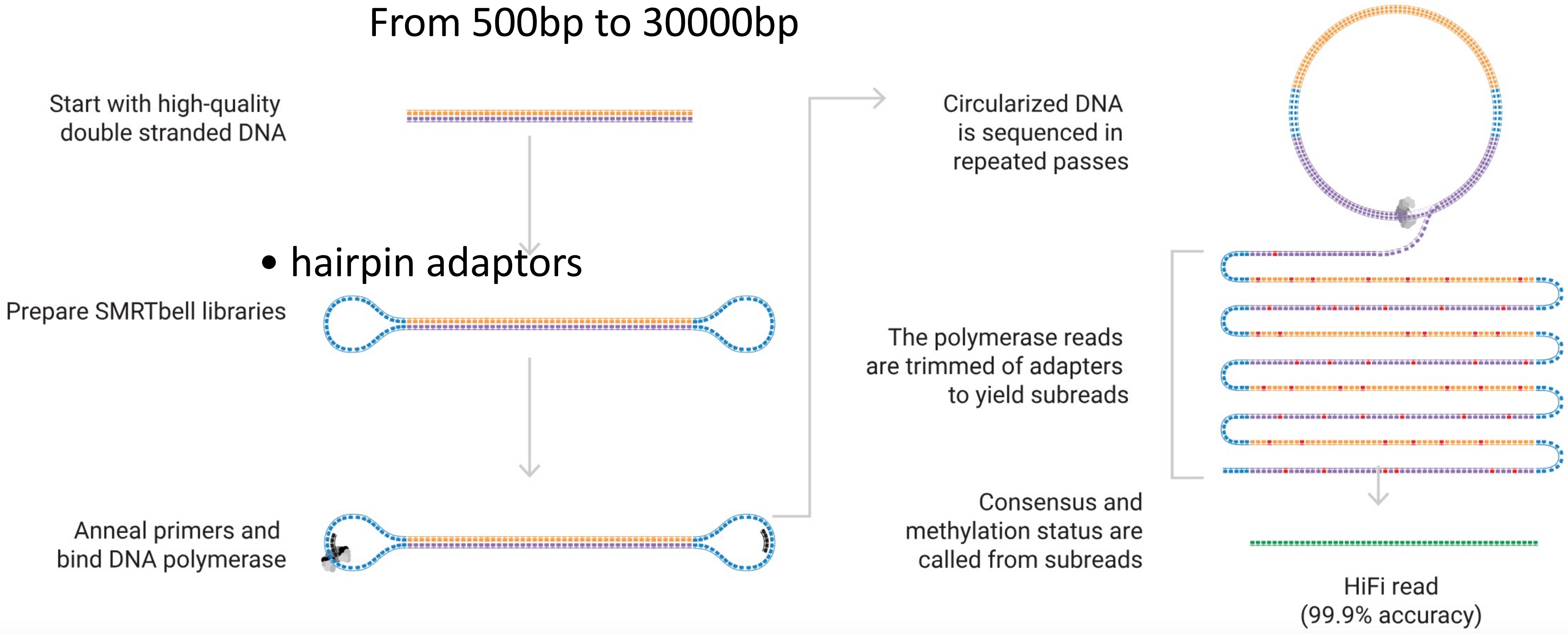
NASA Astronaut Kate Rubins sequenced DNA in space for the first time ever for the Biomolecule Sequencer investigation, using the MinION sequencing device.
Credits: NASA

PACIFIC BIOSCIENCE SEQUENCING

- many sequencing cycles

From 500bp to 30000bp

- hairpin adaptors



Start with high-quality double stranded DNA

Prepare SMRTbell libraries

Anneal primers and bind DNA polymerase

Circularized DNA is sequenced in repeated passes

The polymerase reads are trimmed of adapters to yield subreads

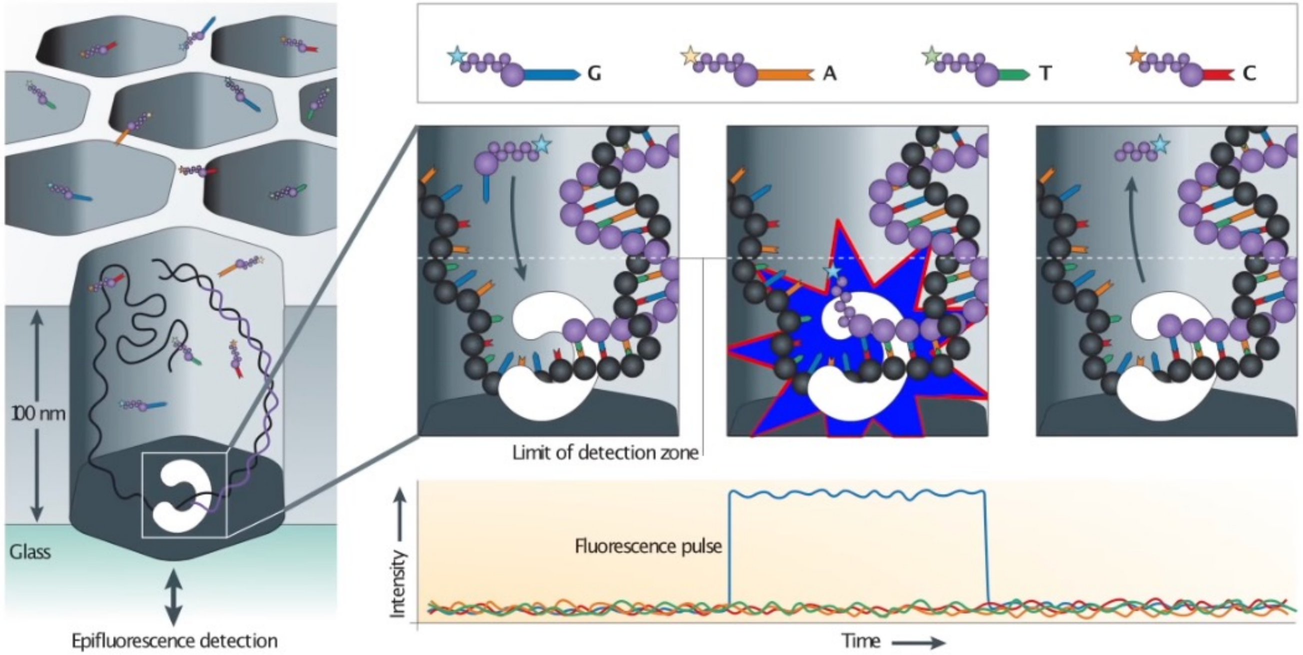
Consensus and methylation status are called from subreads

HiFi read (99.9% accuracy)

- distinguish mutations and random errors

PACIFIC BIOSCIENCE SEQUENCING

fluorescently labeled nucleotides, the unique fluorescent signal of each base (A, T, C, or G)

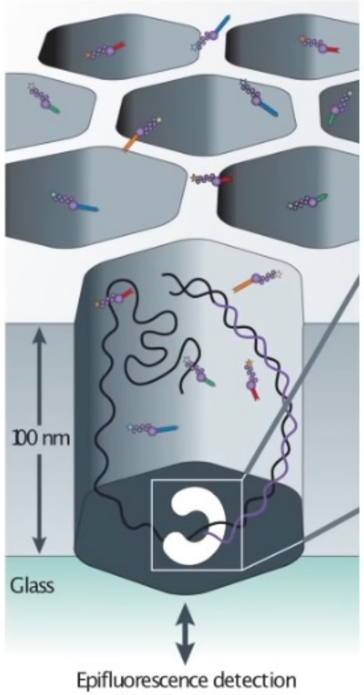


Zero-Mode Waveguides (ZMWs). Each ZMW contains a single DNA polymerase, which synthesizes the complementary strand of the DNA template.

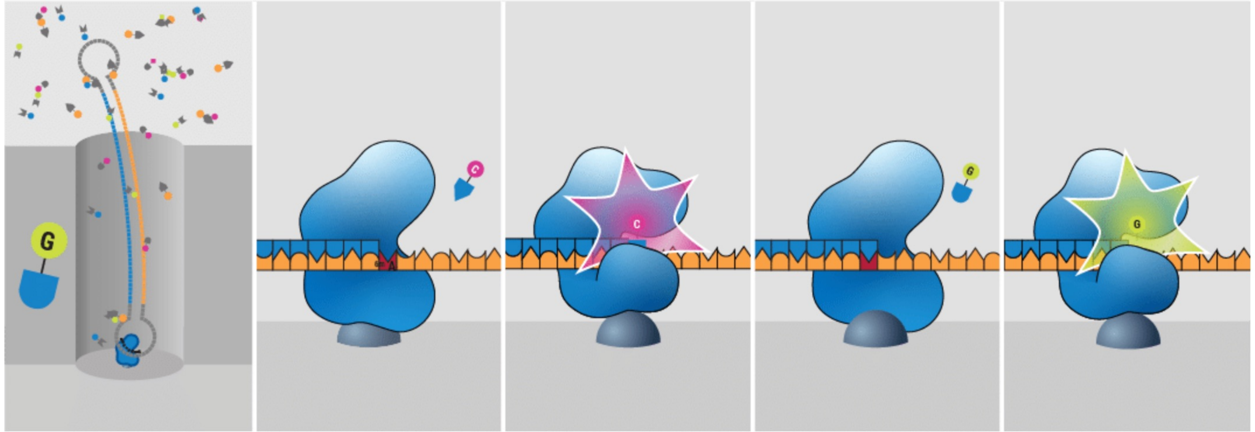
Nature Reviews Genetics **11**, 31–46 (2010)

PACIFIC BIOSCIENCE SEQUENCING

Pacific Bioscience sequencing



- Long reads (100kB)
- High error rates (10-15%)
  - Errors are random - Good thing!

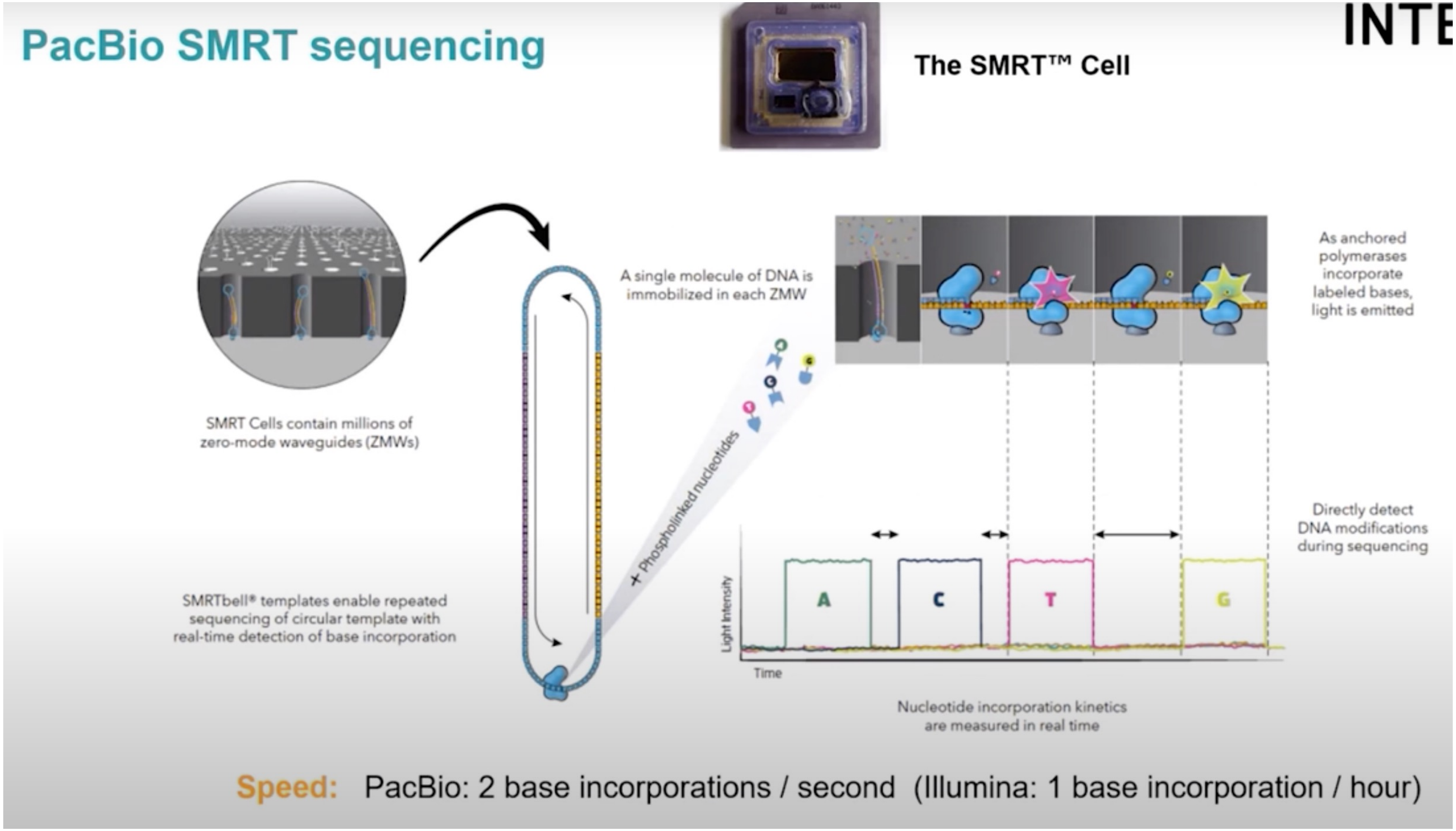


As the polymerase incorporates fluorescently labeled nucleotides, the unique fluorescent signal of each base (A, T, C, or G) is detected and recorded. The continuous observation of the polymerase allows for long reads, making it particularly useful for sequencing large and complex genomes and for detecting structural variants

PACIFIC BIOSCIENCE SEQUENCING

https://www.google.com/search?sca_esv=5b4135254b5cefab&q=pacbio+sequencing&tbm=vid&s
ource=lnms&fbs=AEQNm0Be9hsxO5zOUoY5v2srYNPRwAZKm6L2wMvuJQea-
bATJFvYWVldac53RWY9UFAkudUlgOpsSf_UFsWgSudHjf7uA2fiCym9xNHPZUFwoQkURK9ZPhYbTRj
pdA_O1eEDHd5Y23L13-
8v4Ajf7EIAvj8YPVKoTvsMQ6TlpMMJVks3fSrLkE&sa=X&ved=2ahUKEwjfhKL2tsOJAxXbhf0HHbeUKM
EQ0pQJegQIHBAB&biw=1357&bih=716&dpr=2#fpstate=ive&vld=cid:c1e82dd7,vid:_lD8JyAbwEo,s
:0

PACIFIC BIOSCIENCE SEQUENCING vs ILLUMINA



PacBio SMRT sequencing

The SMRT™ Cell

INTE

SMRT Cells contain millions of zero-mode waveguides (ZMWs)

SMRTbell® templates enable repeated sequencing of circular template with real-time detection of base incorporation

A single molecule of DNA is immobilized in each ZMW

As anchored polymerases incorporate labeled bases, light is emitted

Directly detect DNA modifications during sequencing

Nucleotide incorporation kinetics are measured in real time

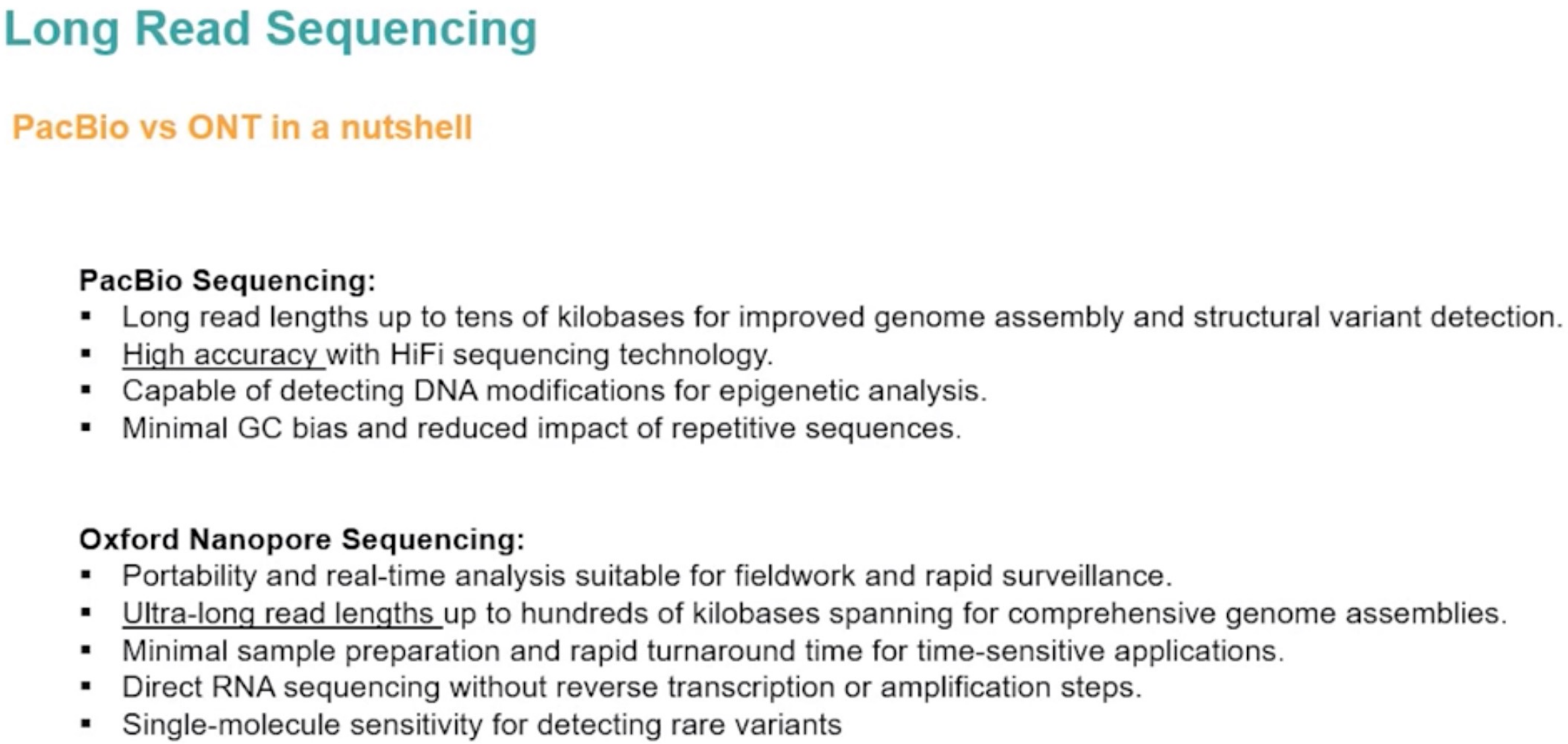
Speed: PacBio: 2 base incorporations / second (Illumina: 1 base incorporation / hour)

PACIFIC BIOSCIENCE SEQUENCING vs OXFORD NANOPORE

PacBio Sequel IIe

MinIONs

PACIFIC BIOSCIENCE SEQUENCING vs OXFORD NANOPORE



Long Read Sequencing

PacBio vs ONT in a nutshell

PacBio Sequencing:

- Long read lengths up to tens of kilobases for improved genome assembly and structural variant detection.
- High accuracy with HiFi sequencing technology.
- Capable of detecting DNA modifications for epigenetic analysis.
- Minimal GC bias and reduced impact of repetitive sequences.

Oxford Nanopore Sequencing:

- Portability and real-time analysis suitable for fieldwork and rapid surveillance.
- Ultra-long read lengths up to hundreds of kilobases spanning for comprehensive genome assemblies.
- Minimal sample preparation and rapid turnaround time for time-sensitive applications.
- Direct RNA sequencing without reverse transcription or amplification steps.
- Single-molecule sensitivity for detecting rare variants