# DNA SEQUENCING

Techniques:

## 1. Maxam e Gilbert:first method

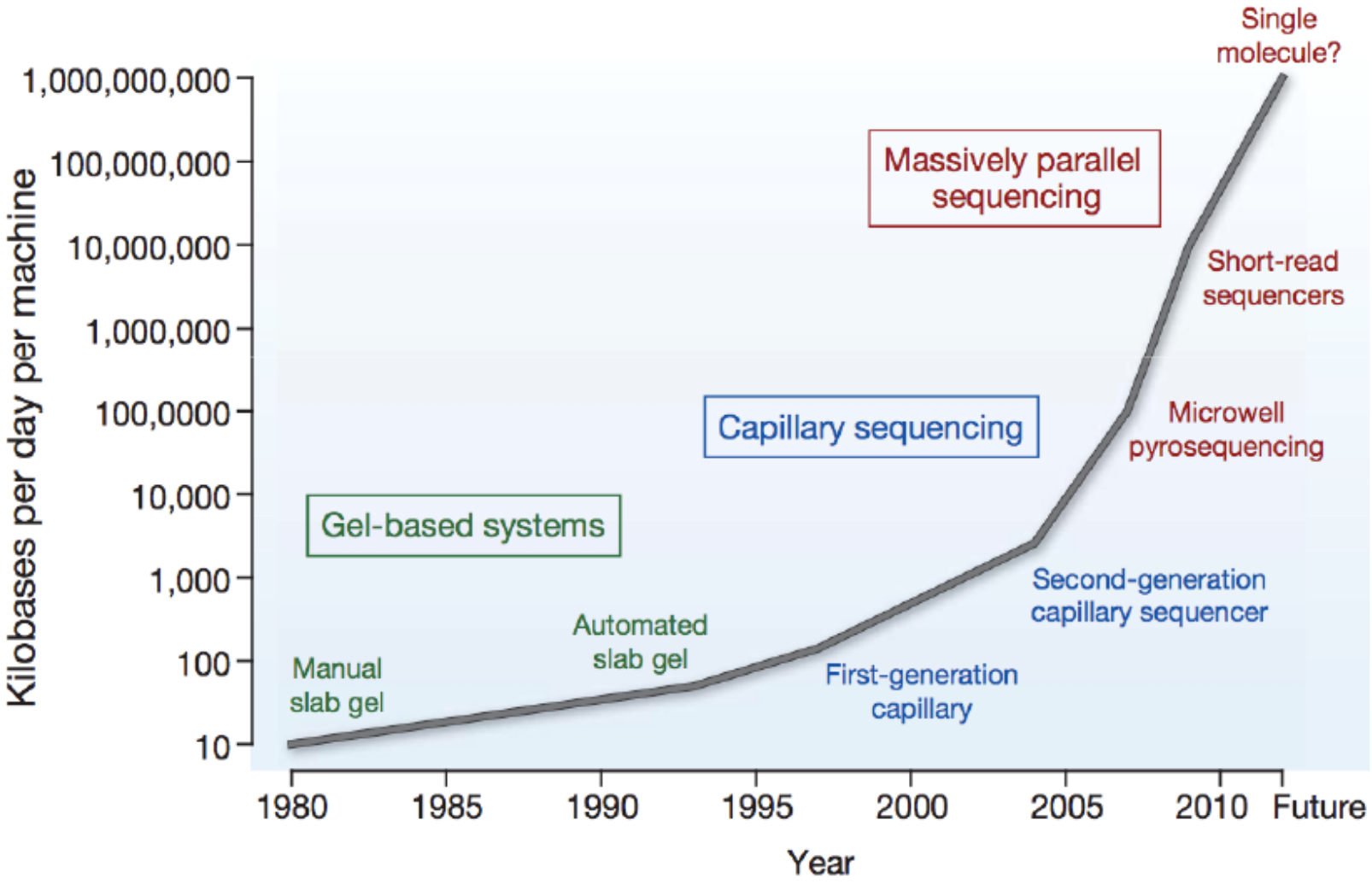
## 2. Sanger Sequencing:

**basis for all seqeuncing tecniques**

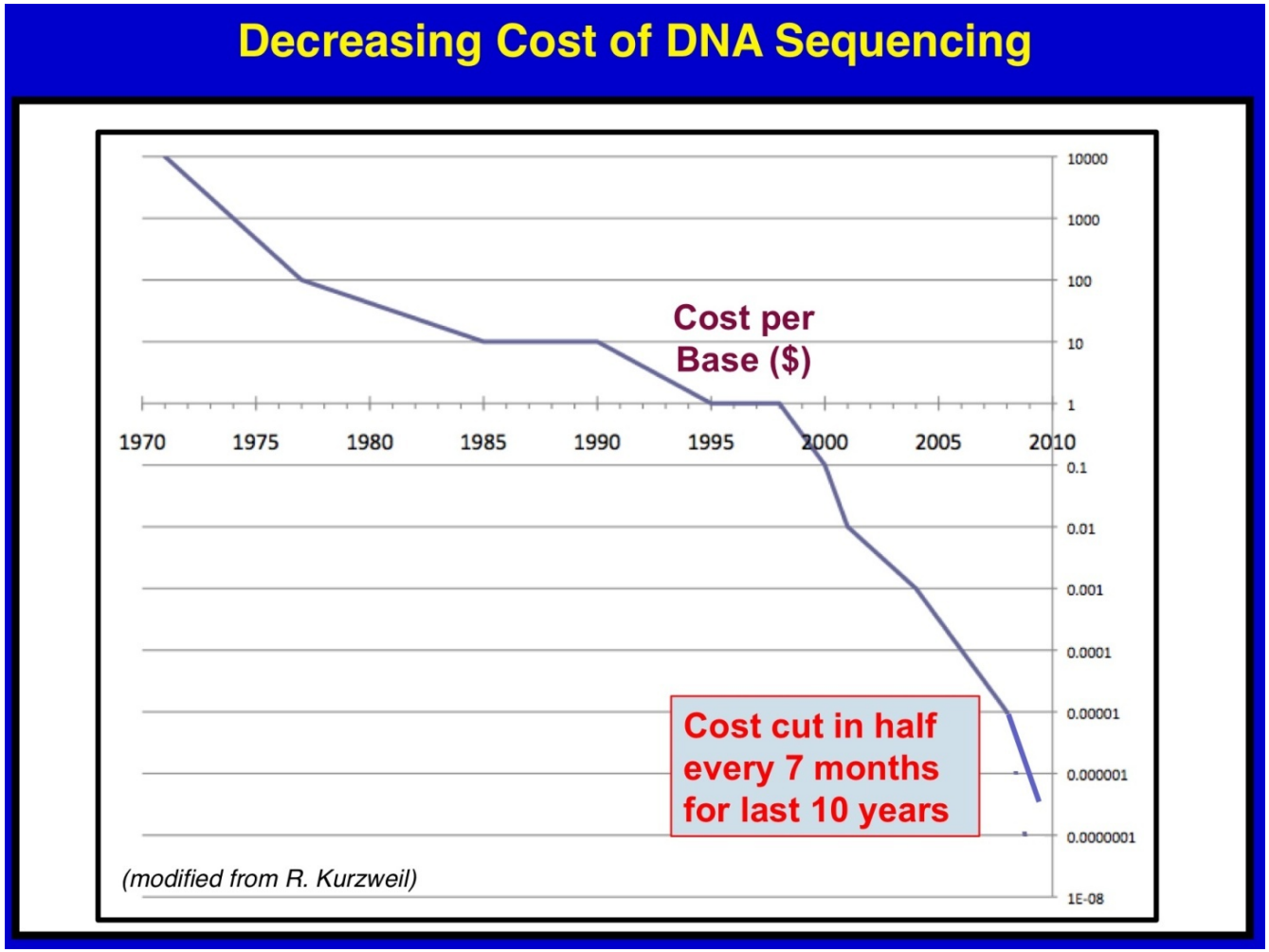
## 3. Massive Parallel Seqeuncing

**DNA sequencing** includes several methods and technologies that are used for determining the order of the nucleotide bases—adenine, guanine, cytosine, and thymine—in a molecule of DNA.

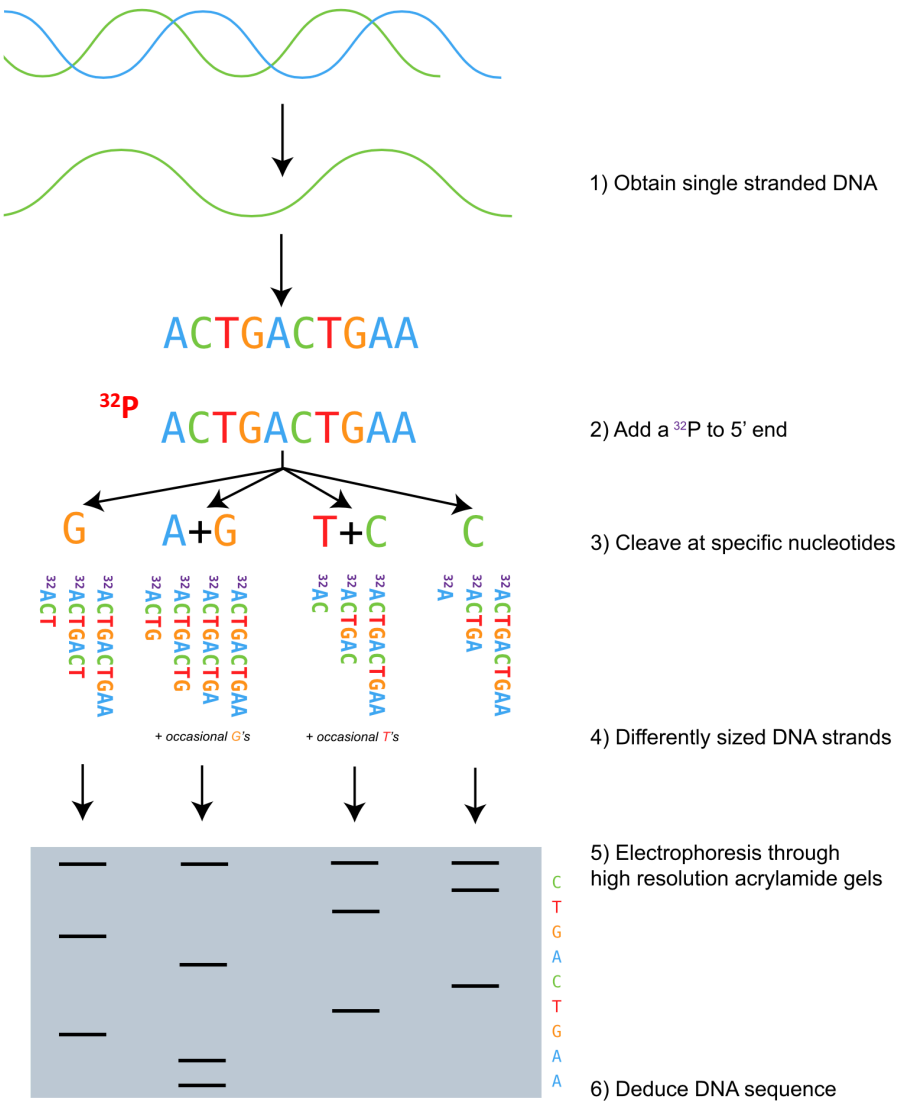
# The History of DNA Sequencing Technology



# Quickly reduced Cost



# 1. Maxam-Gilbert Method
# chemical sequencing




1) Obtain single stranded DNA

2) Add a $^{32}P$ to 5' end

3) Cleave at specific nucleotides

4) Differently sized DNA strands

5) Electrophoresis through high resolution acrylamide gels

6) Deduce DNA sequence

1. Denature a double-stranded DNA to single-stranded by increasing temperature.
2. Radioactively label one 5' end of the DNA fragment to be sequenced by a kinase reaction using gamma-$^{32}P$-ATP.
3. Cleave DNA strand at specific positions using chemical reactions.

- Reaction 1: Guanines (and to some extent the adenines) are methylated by dimethyl sulfate
- Reaction 2: Purines (A+G) are depurinated using formic acid,
- Reaction 3: Pyrimidines (C+T) are hydrolysed using hydrazine.
- Reaction 4: Hydrazine + salt (sodium chloride) inhibits the reaction of thymine for the C-only reaction.

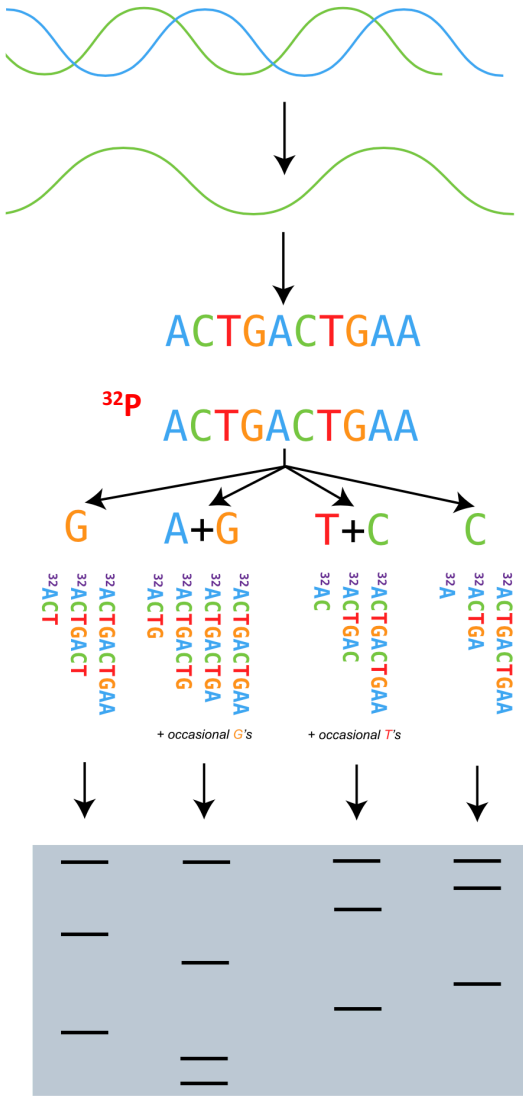
- → NOTE: concentration of chemicals is chosen to only cause 1 modification in a molecule of interest

- **→ The modified DNAs may then be cleaved by hot piperidine**

- Now in four reaction tubes, we will have several differently sized DNA strands that carry $^{32}P$ at 5'end

- Fragments are electrophoresed in high-resolution acrylamide gels for size separation.

- These gels are placed under X-ray film, which then yields a series of dark bands which show the location of radiolabeled DNA molecules. The fragments are ordered by size and so we can deduce the sequence of the DNA molecule.




1) Obtain single stranded DNA

2) Add a $^{32}P$ to 5' end

3) Cleave at specific nucleotides

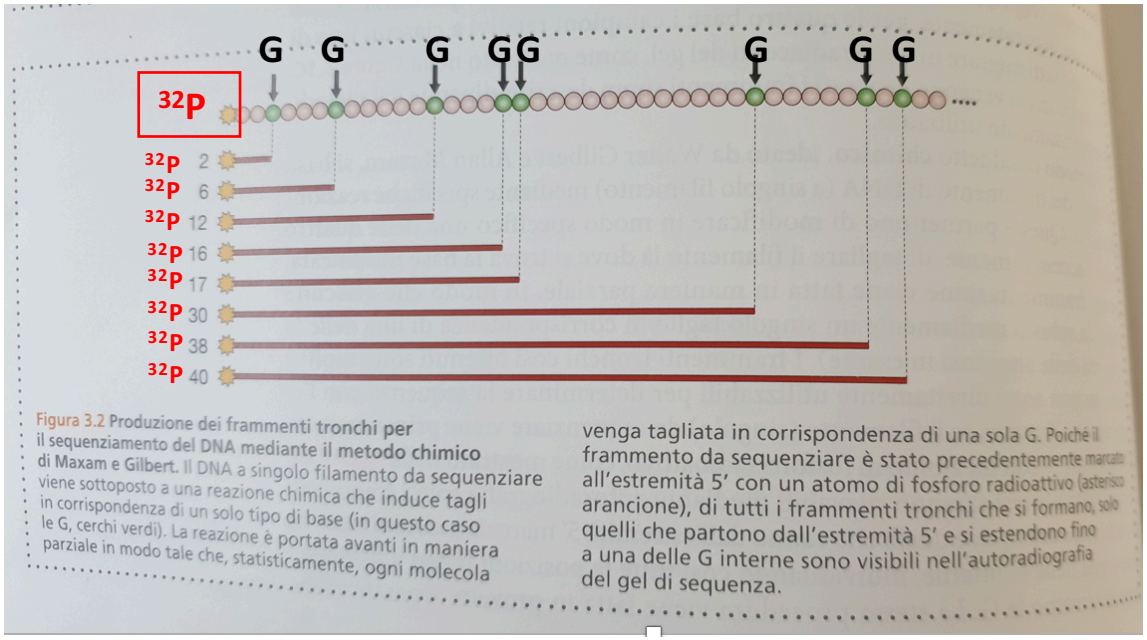
4) Differently sized DNA strands

5) Electrophoresis through high resolution acrylamide gels

6) Deduce DNA sequence

**Reaction 1: Guanines are methylated by dimethyl sulfate**

- Set up reaction to induce only one modification per DNA molecule
- Cut with piperidine at modified site
- Release of specfic 5' labelled DNA fragmetns



# 1. Maxam-Gilbert Method chemical sequencing

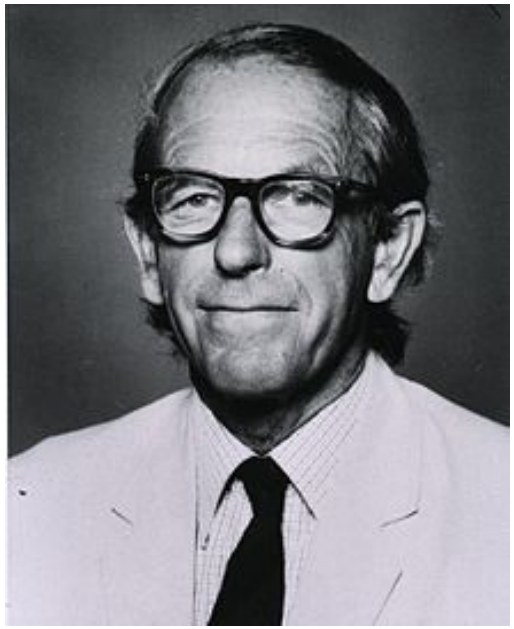
Pros

Maxam-Gilbert sequencing was at one point more popular than the Sanger method. Purified DNA could be used directly, while the Sanger method required that each read start be cloned for production of single-stranded DNA.

Cons

Cons included difficulties scaling up, and the handling of X-rays and radiolabeling, which were harmful to technicians.

# 2. Sequenziamento di DNA mediante il metodo di Sanger



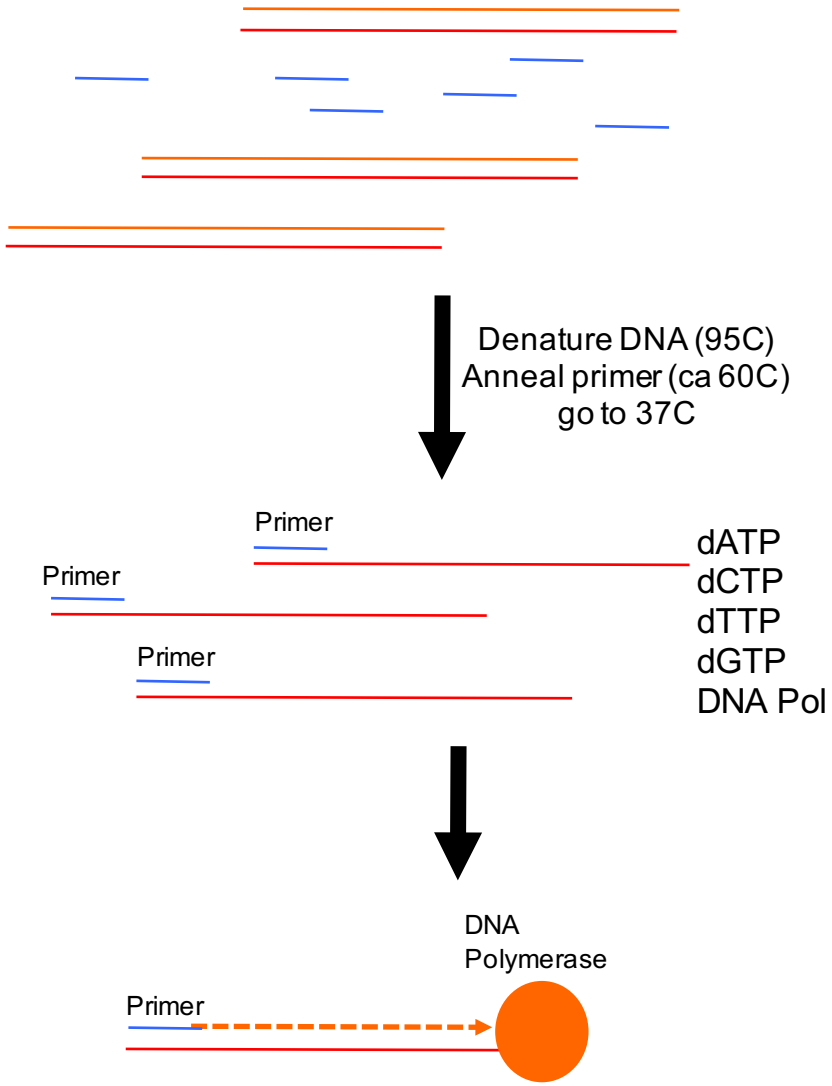
## Sequenziamento con il metodo dei **dideossinucleotidi**

F. Sanger 13 agosto 1918 – 19 novembre 2013

**Due** premi Nobel. Uno per il sequenziamento dell'insulina ed uno per il sequenziamento del genoma del fago $\phi$–X174

wellcome trust sanger institute

# General concept in a Sanger sequencing reaction: The synthesis of a new strand of DNA from a ss template DNA



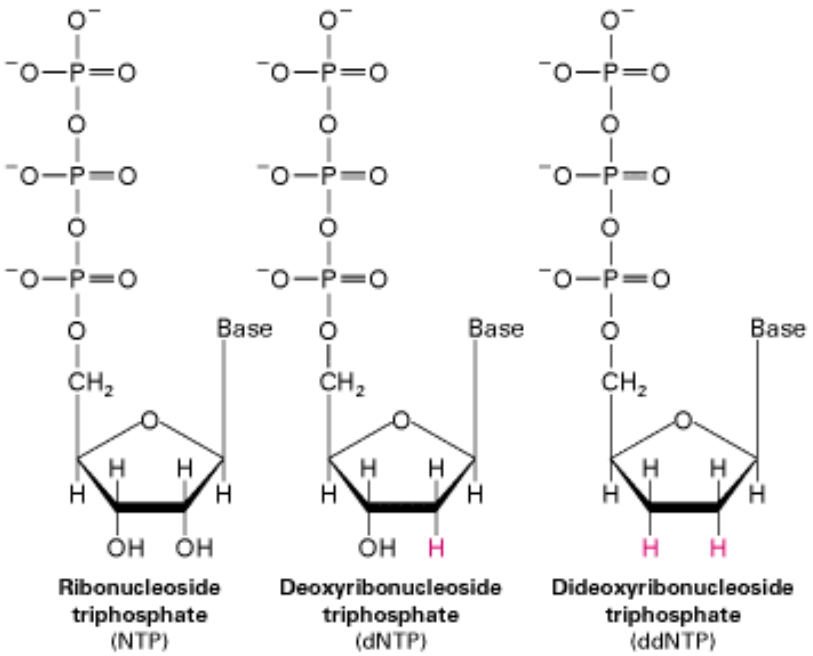

- Primer oligonucleotides
Identical dsDNA molcules with known primer target site

add dATP, dTTP, dCTP, dGTP and DNA polymerase

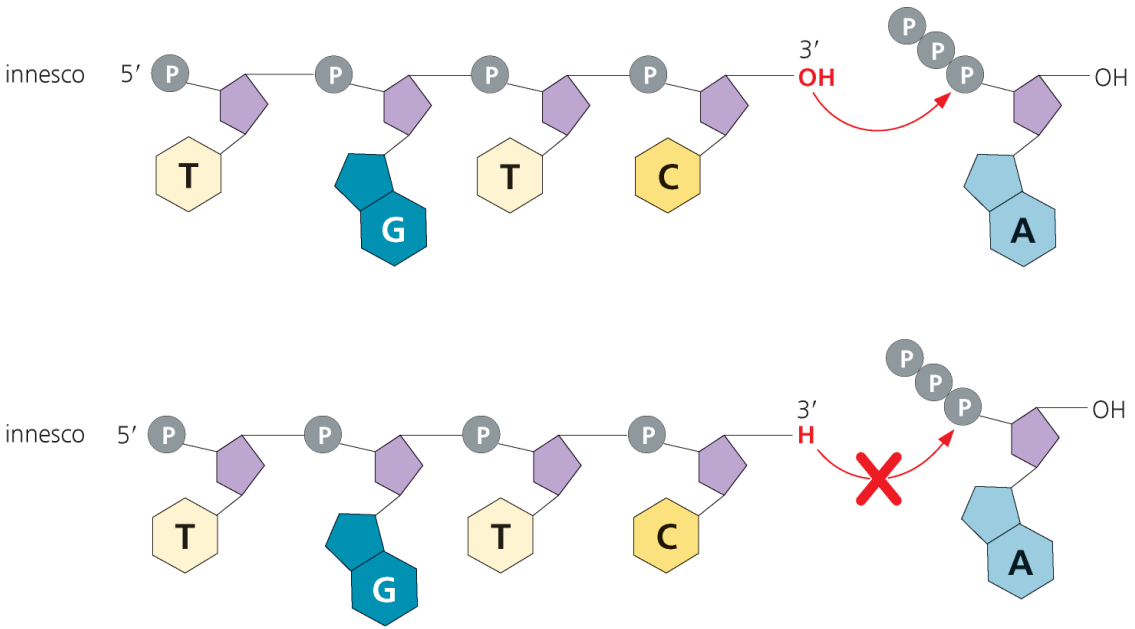
Polymerase elongates template DNA

**PROBLEM: HOW CAN WE READ THE NEWLY SYNTHEZISED DNA SEQUENCE??**

# A great trick: using di-deoxyribonucleoside triphosphates to terminate the synthesis of DNA molecules




**deoxyadenine triphosphate**



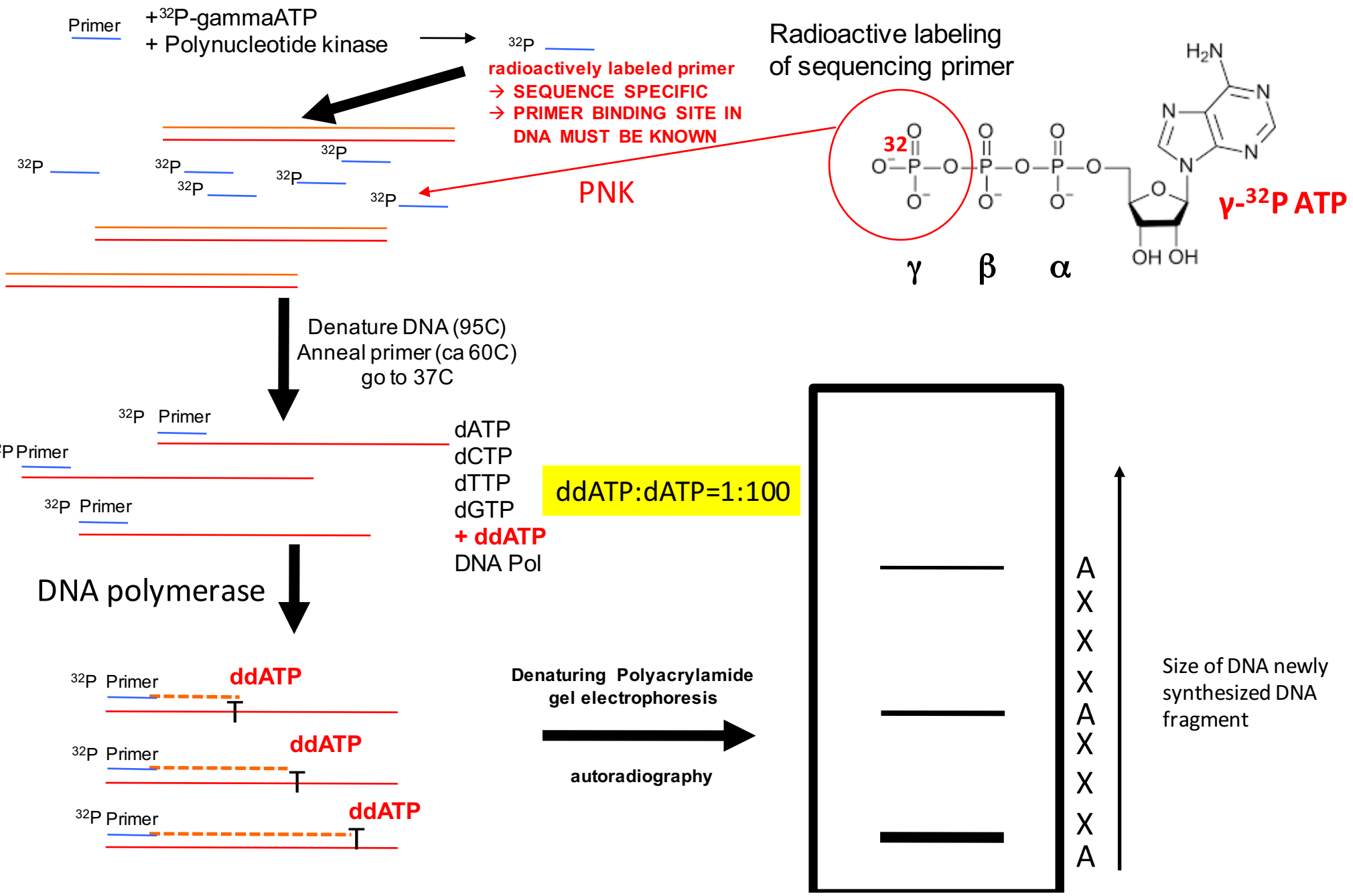

**di-deoxyadenine triphosphate**

Concept: mixing a low amount of ddATPs into a high amount of dATP (**ca 1:100**): A pool of DNA molecules will be generated in which DNA molecules terminate at all possible A sites.

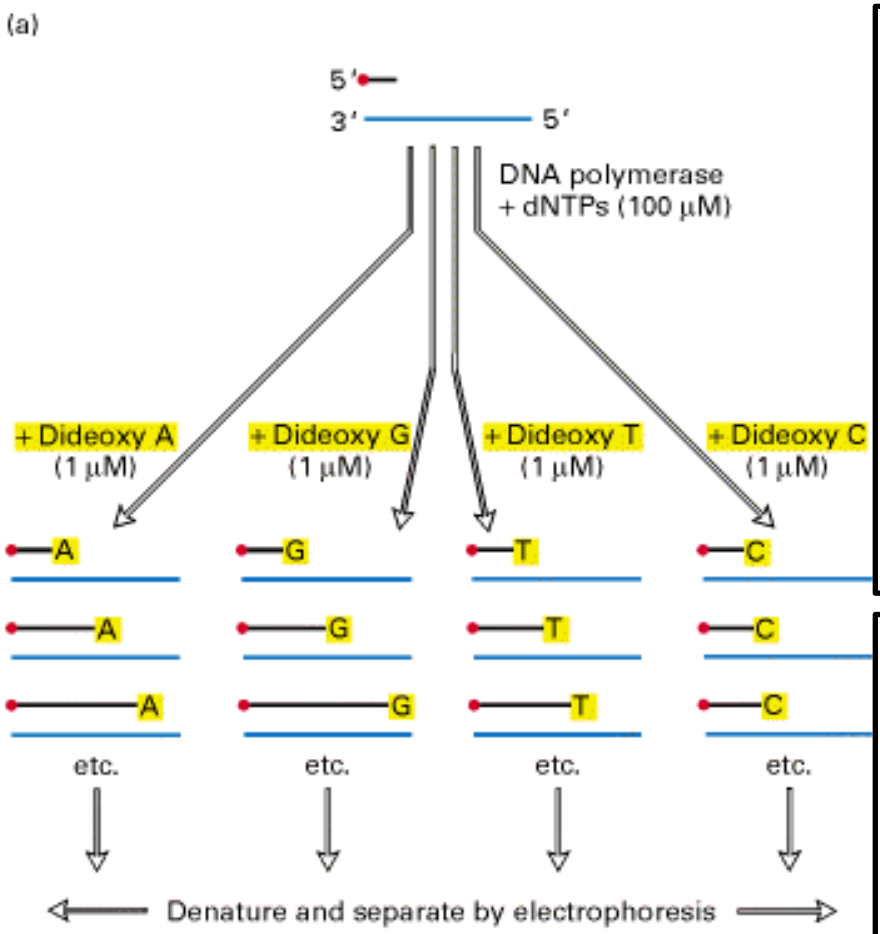
**ddNTP enable me to terminate sequencing at a defined position in the newly synthesized DNA molecule**

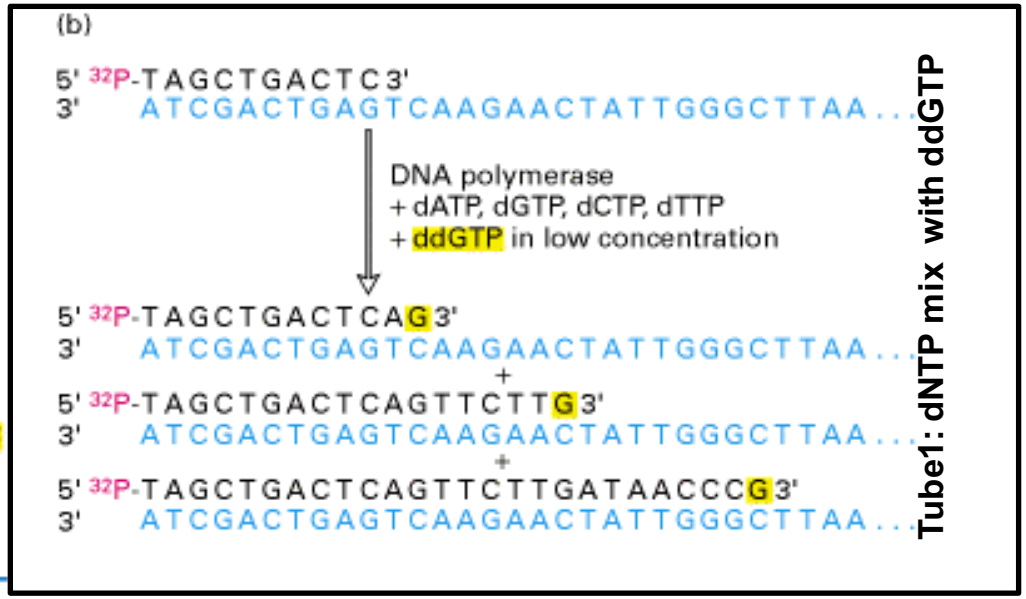
**PROBLEM: How can detect sequencing products??**

# General concept in a sequencing reaction:
# The synthesis of a new strand of DNA from a ss template DNA



# Classic Sanger sequencing of a DNA fragment requires 4 parallel sequencing reactions





**Tube3: dNTP mix with ddCTP**

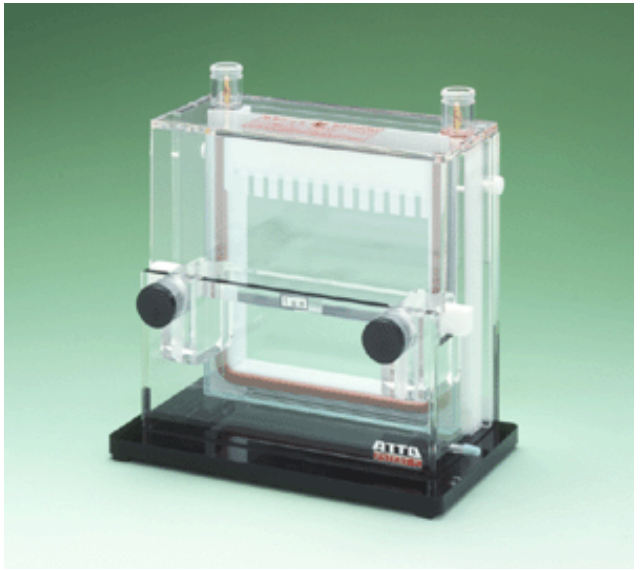
**Tube4: dNTP mix with ddTTP**

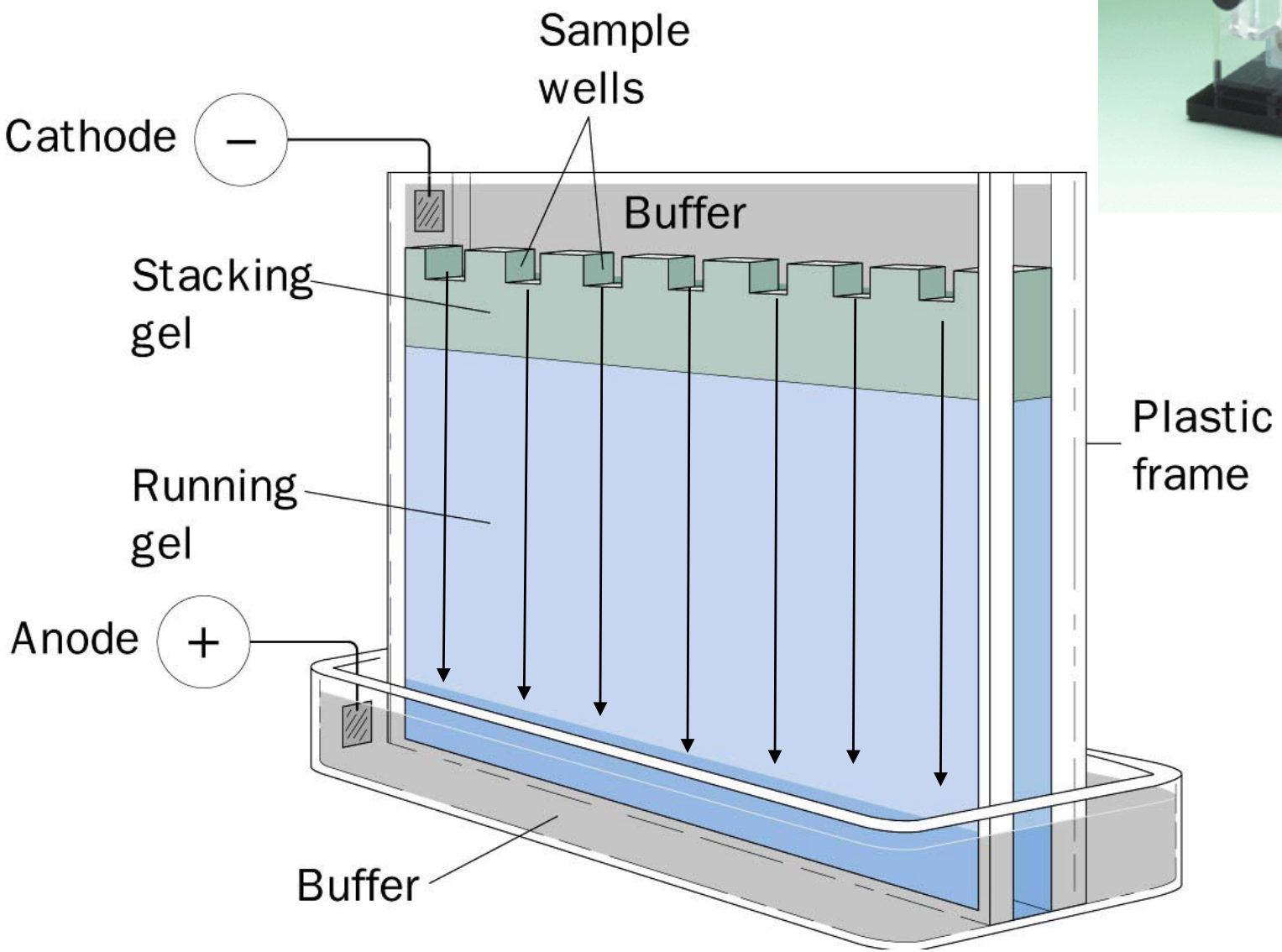
# Le due basilari tecniche elettroforetiche per la separazione di frammenti di DNA (e di RNA).

- **Elettroforesi su gel di poliacrilammide (PAGE)**:
- il gel è ottenuto per polimerizzazione in soluzione acquosa tamponata di acrilammide con una piccola percentuale di bisacrilammide tra due vetri con intercapedine di 0,5–2 mm mantenuti verticalmente. Il gel è costituito da una rete tridimensionale covalente del polimero ed è sostanzialmente un gel irreversibile.

# PolyAcrylamide Gel Electrophoresis

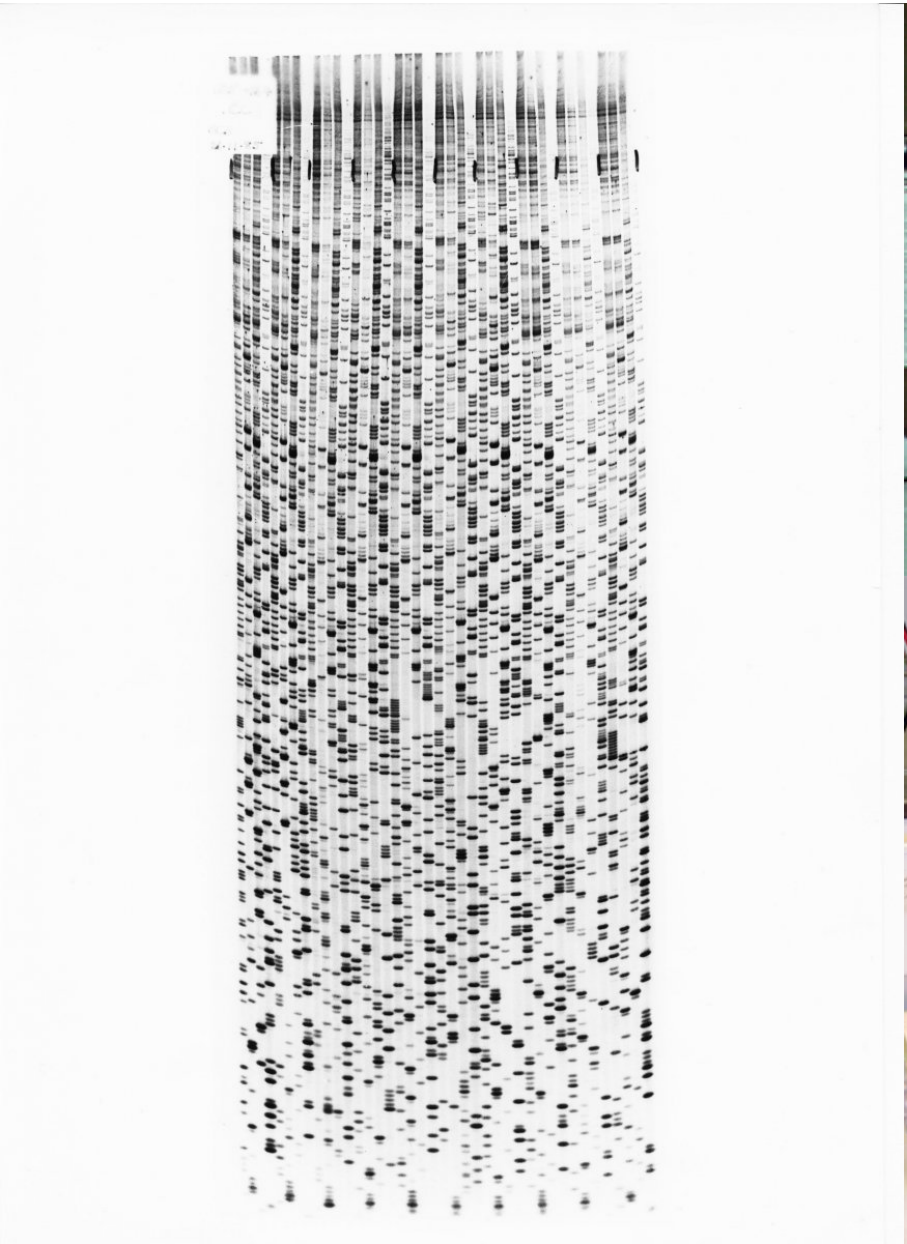
il gel è ottenuto per polimerizzazione in soluzione acquosa tamponata di acrilammide con una piccola percentuale di bisacrilammide tra due vetri con intercapedine di 0,5–2 mm mantenuti verticalmente. Il gel è costituito da una rete tridimensionale covalente del polimero ed è sostanzialmente un gel irreversibile.

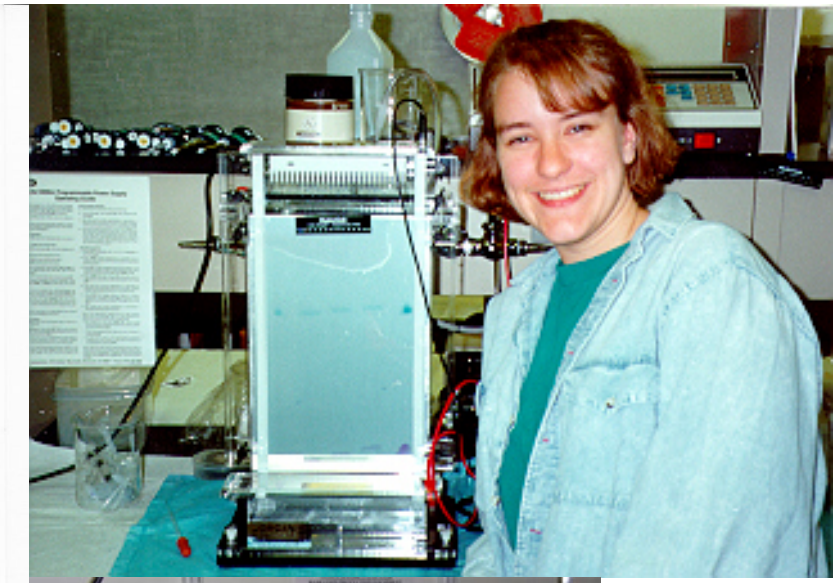


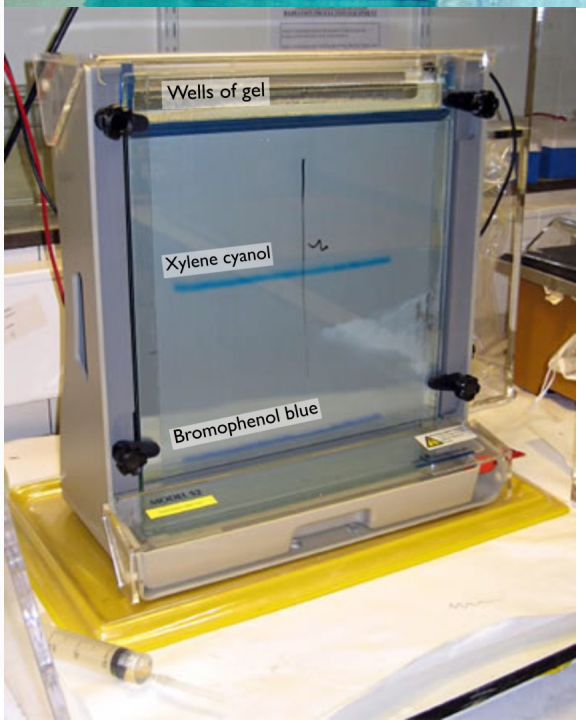



**PAGE**

**Read length:**
**500 – 700 nucleotdi**

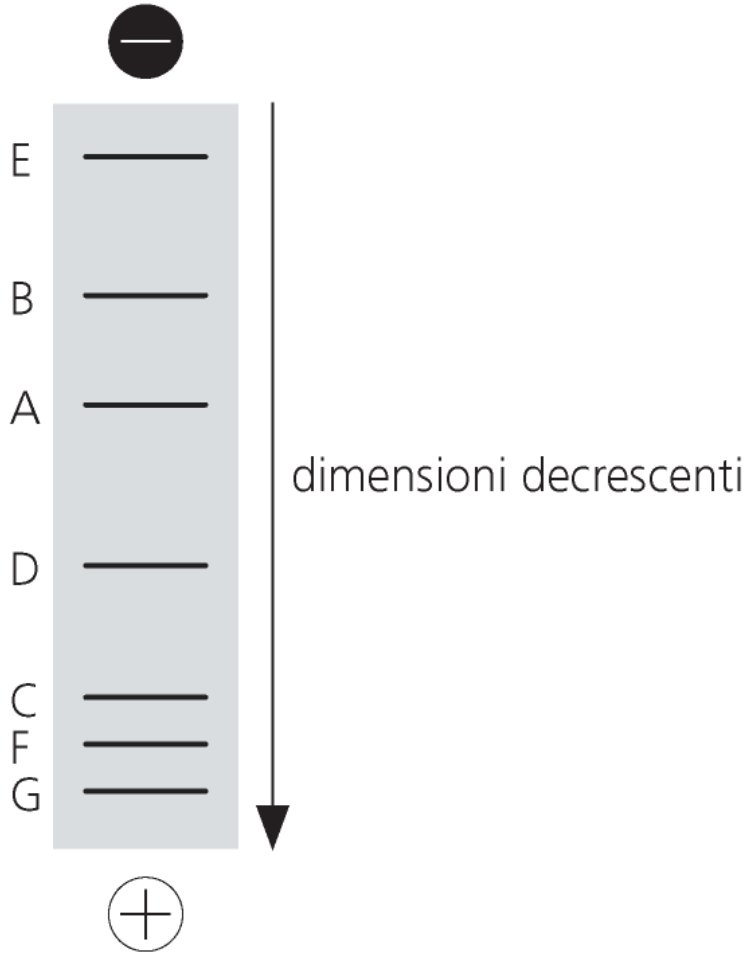


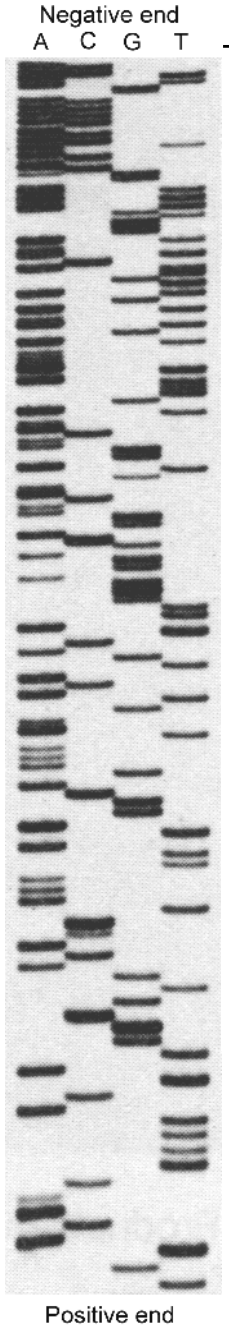




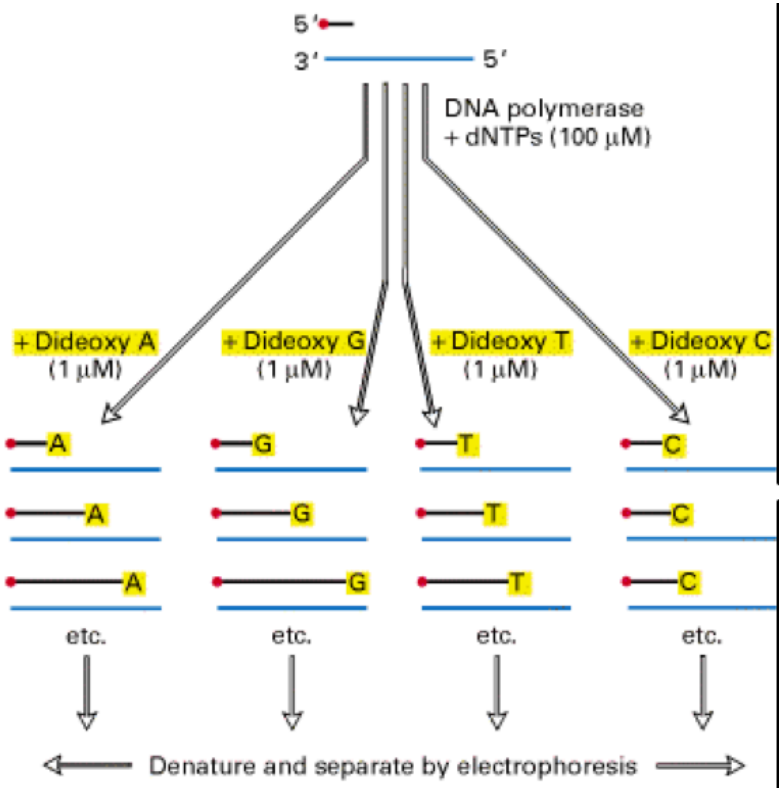

Standard in lab
until ca. 1995

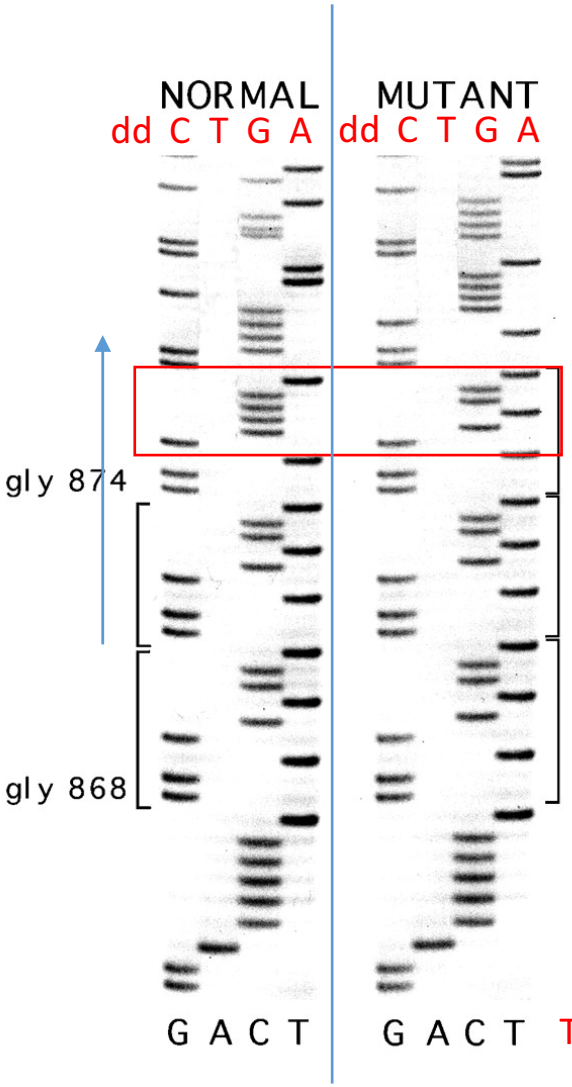
Primer is radioactively labelled!!
All fragments produced by DNA Polymerase can be visualized by autoradiography




ddNTPs






# Can you read the DNA sequence?

NORMAL: GGT GCT CCT GGT GCT CCT GGT GCC CCT GGC CCC GTT GGC CCT GCT

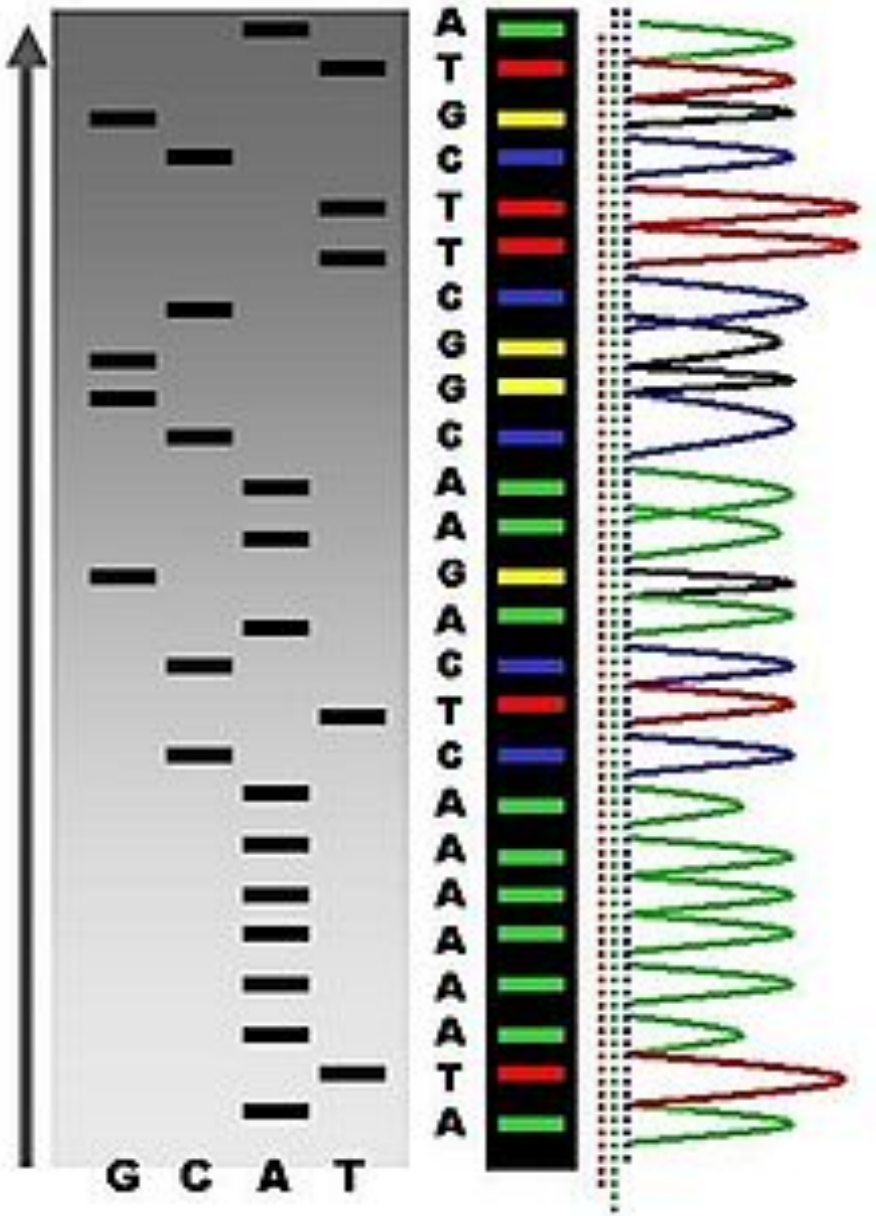
AMMINO ACID SEQEUNCE: G A P G A P G A P G P V G P A

MUTANT: GGT GCT CCT GGT GCT CCT GGT GCT CCT GGT GCC CCT GGC CCC GTT

AMMINO ACID SEQEUNCE: G A P G A P G A P G A P G P V

# 2.1. Automated sequencing based on Sanger technique

# Dye-terminator Sanger sequencing




Classic radioactive

Dye-terminator Sanger sequencing

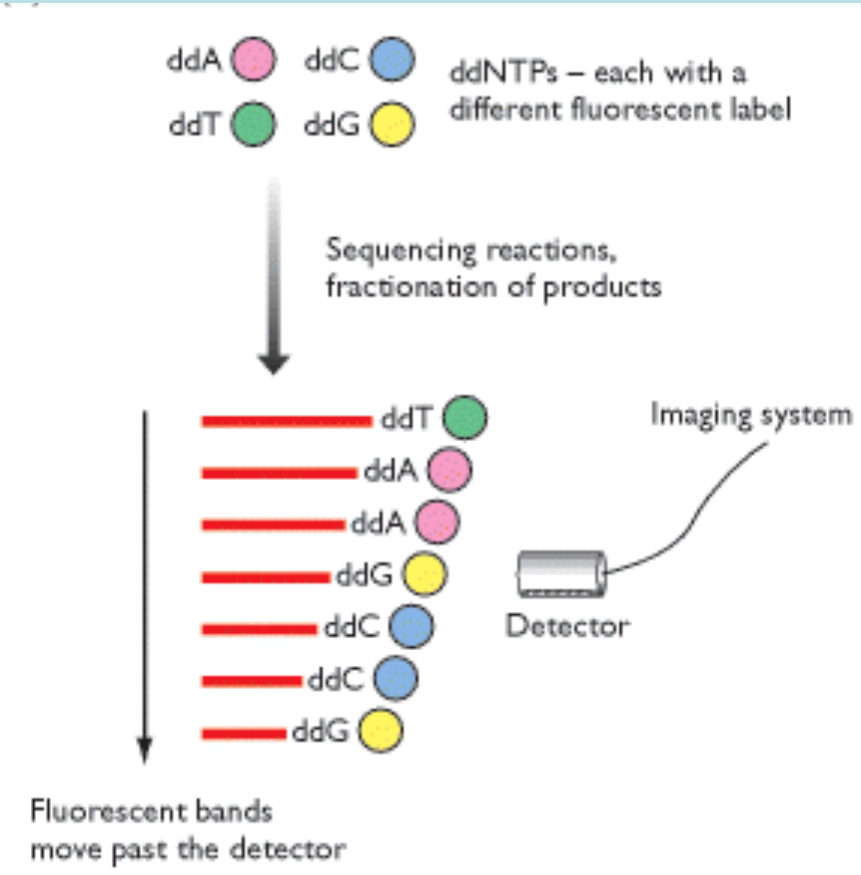
# Dye-terminator Sanger sequencing

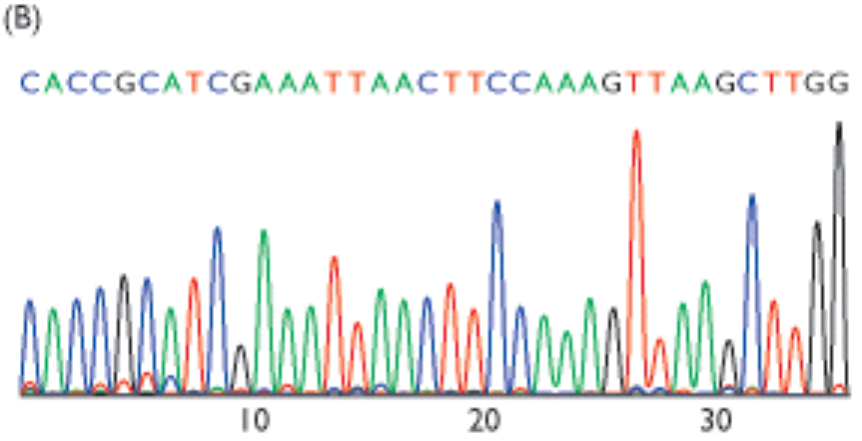
**Tre diversi modi di marcare i frammenti di Sanger:**

1) I frammenti di Sanger sono resi radioattivi per incorporazione di $\alpha$-dNTP marcato

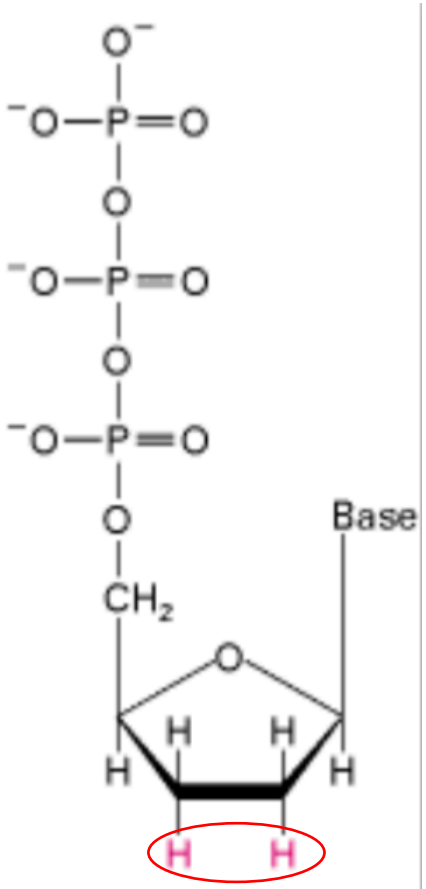
Questo metodo richiede quattro reazioni di polimerizzazione separate e quattro corsie elettroforetiche.

2) Ciascun ddNTP è reso fluorescente con un fluoroforo diverso. Questo metodo consente anche di effettuare tutte le reazioni in un' unica provetta ed unica corsie elettroforeticha..

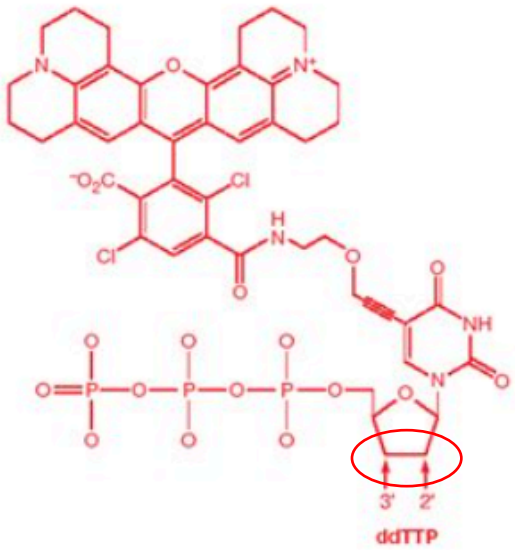


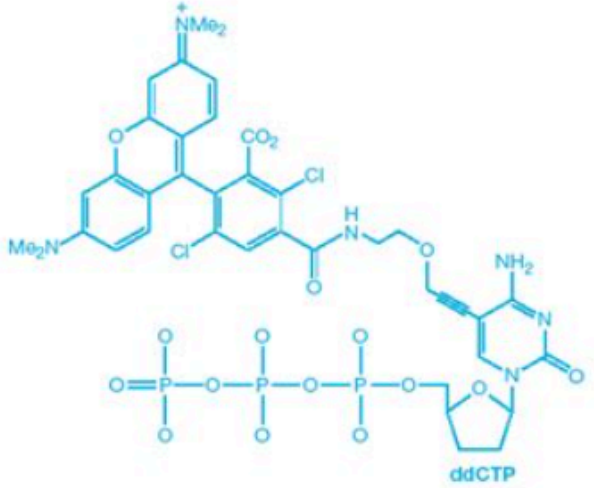
# Dye-terminator ddNTPs



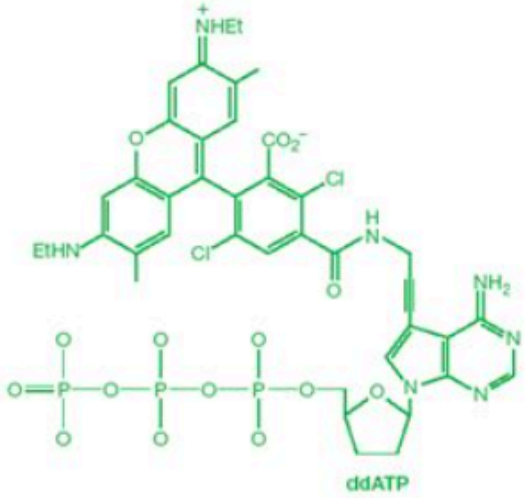

**Dideoxyribonucleoside triphosphate (ddNTP)**



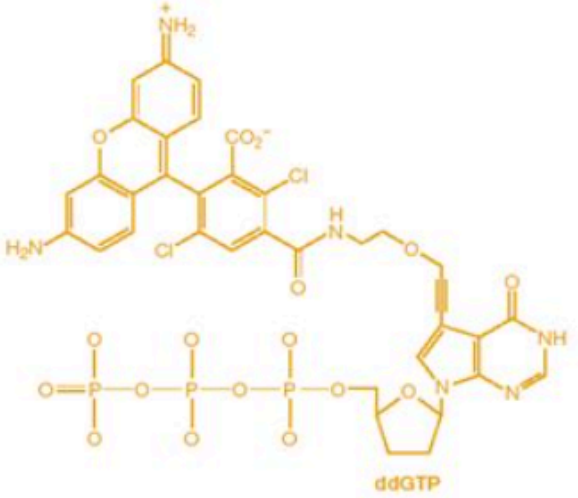

ddTTP



ddCTP

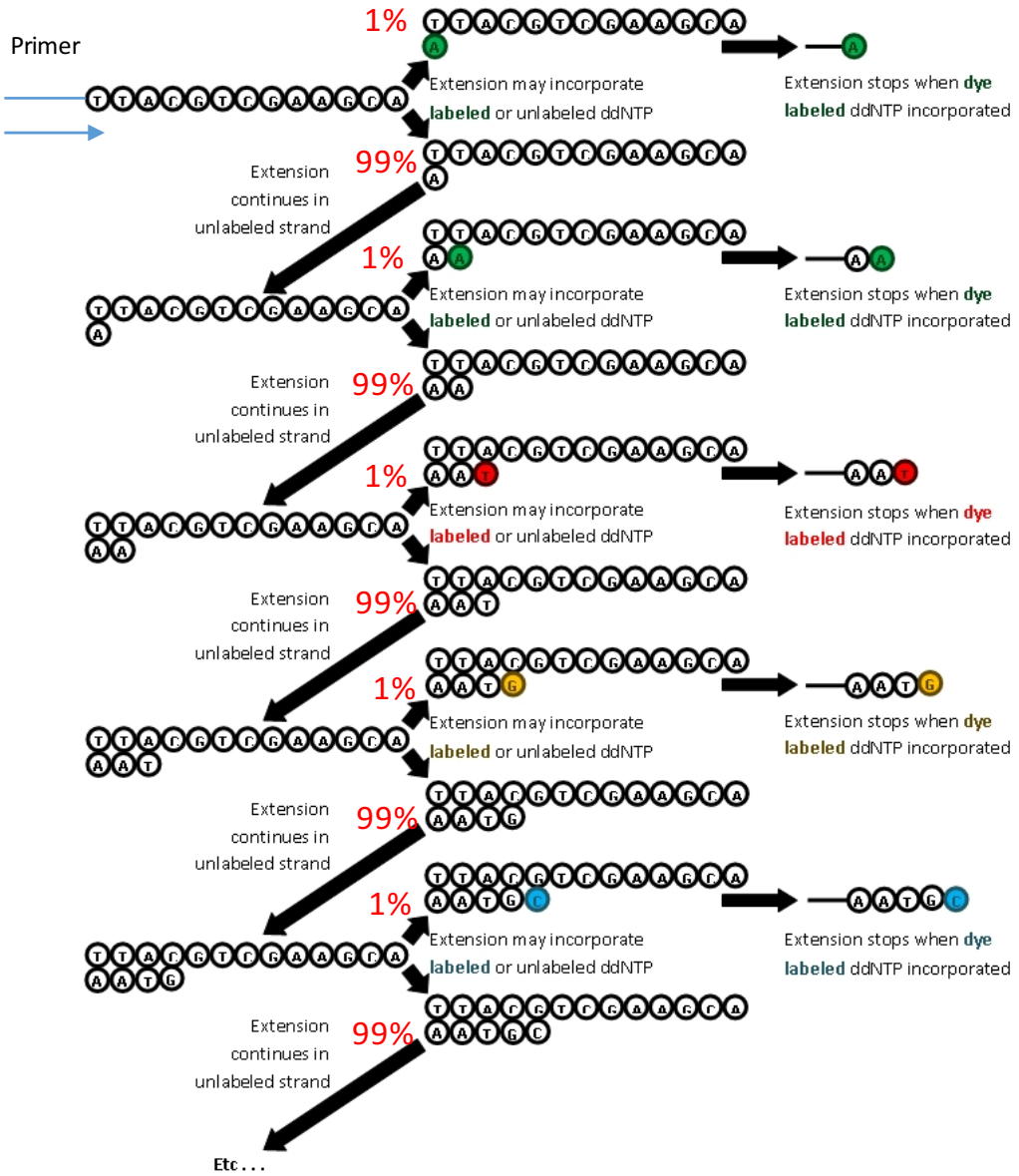


ddATP



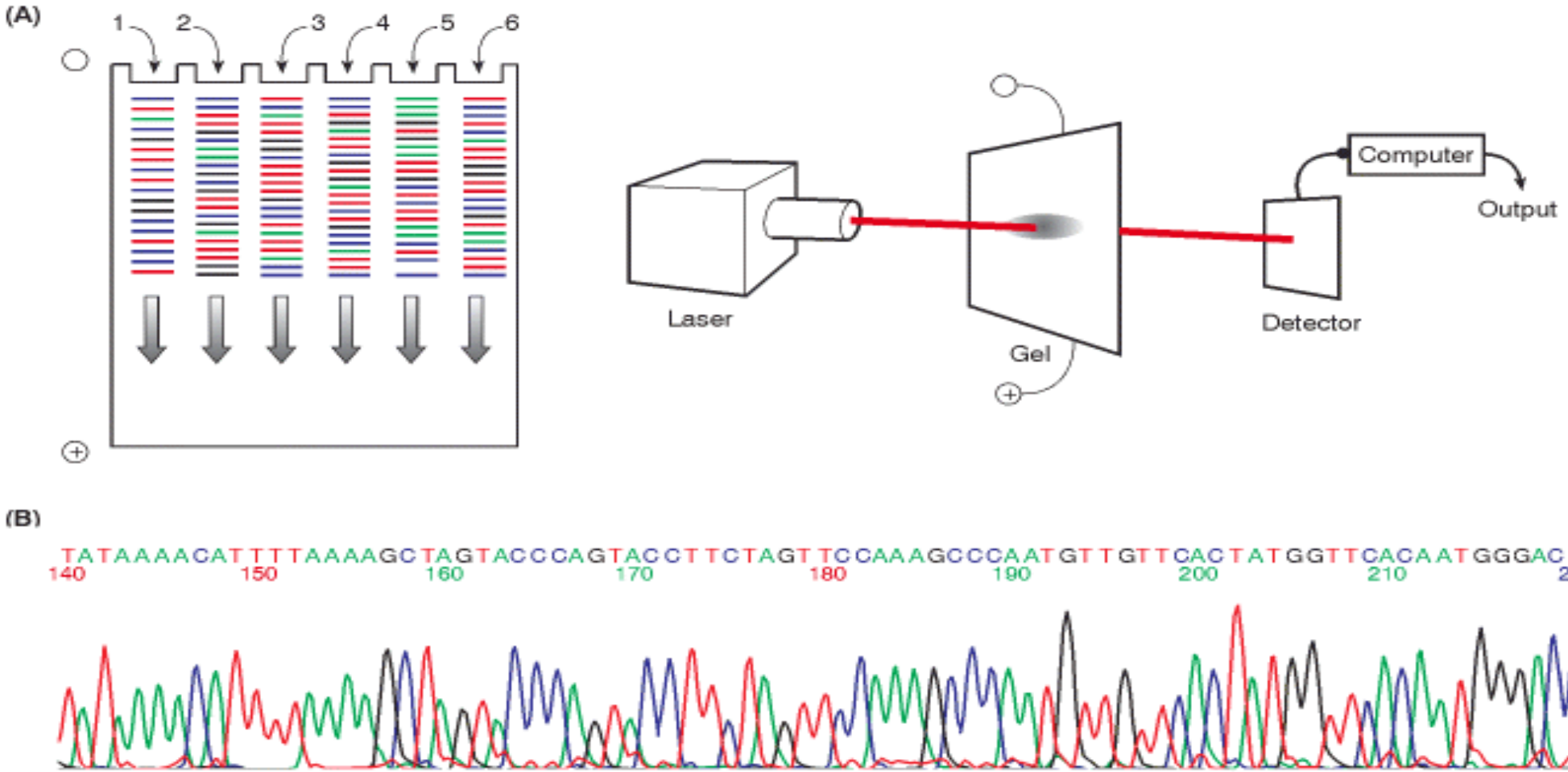
ddGTP

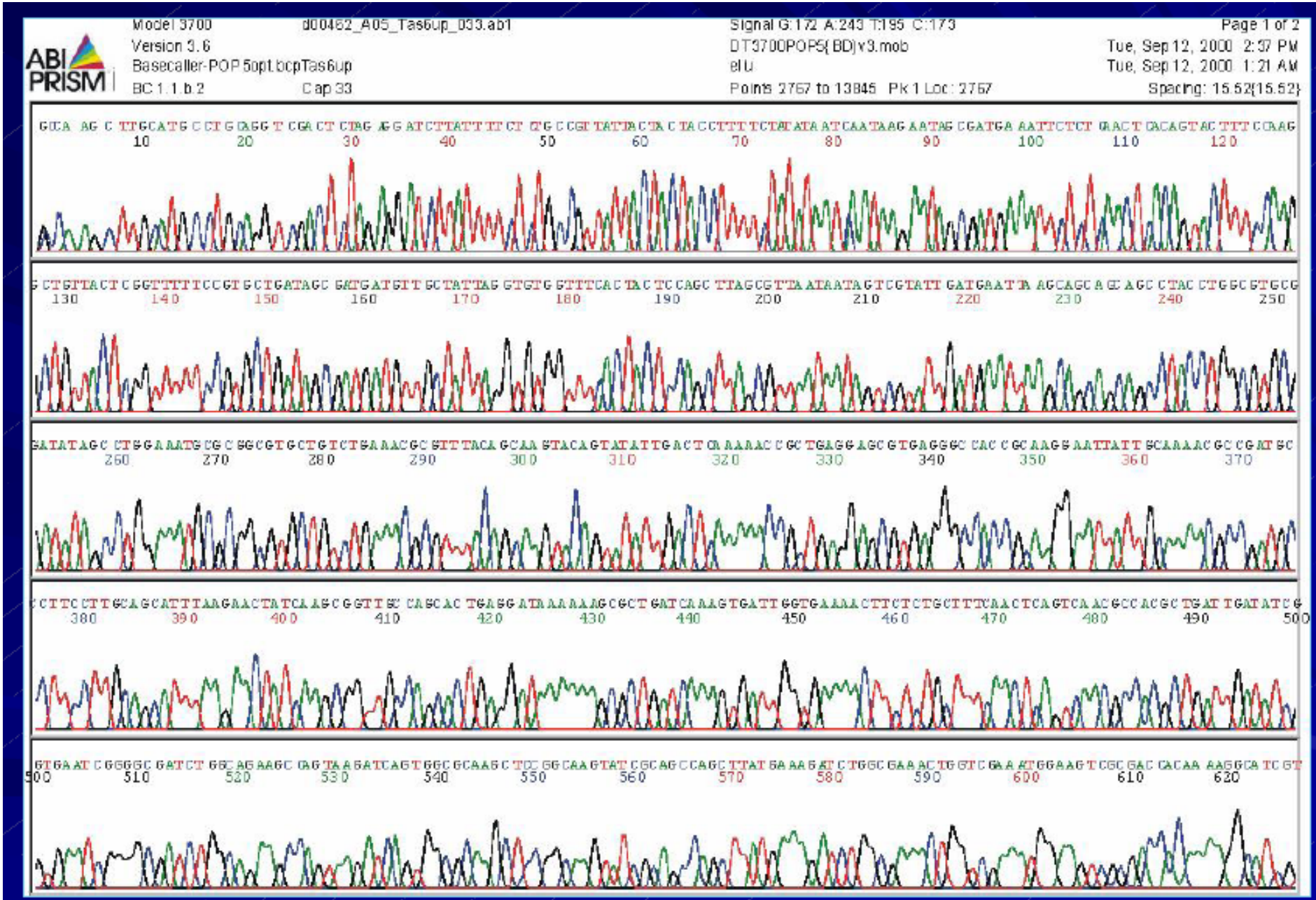
# Dye-terminator Sanger sequencing




- Modified ddNTPs are incorporated into the produced DNA, however they also stop the extension of the chain (hence they are called terminators). NOTE: ddNTP:dNTP = 1:100). The use of these terminating ddNTPs creates a selection of DNA fragments of differing size, each of which ends with a particular nucleotide which is labeled with a different coloured dye.
- The fragments can then be separated according to size. In conventional agarose gel electrophoresis, a sample of DNA is loaded into a well in the agarose and an electric current applied. Because the conditions within the gel give the DNA an overall negative charge, the electric field pushes the DNA through the gel from the negative terminal to the positive terminal. Smaller fragments of DNA can move more quickly through the gel than larger fragments, and so the DNA separates out into regions of the gel which contain fragments of a similar size. Dye terminator sequencing can be performed using a conventional gel. However in more modern automated systems, the electrophoresis is performed in a thin tube called a capillary. No dye needs to be incorporated into the gel as the DNA fragments are already fluorescently labeled using the dye terminators.
- As the soup of differently sized DNA fragments separates out in the capillary, it produces a series of coloured bands. Each band represents fragments of DNA of a particular size, and each colour represents the base at which the fragment terminates. The shorter fragments, representing the bases at the beginning of the sequence will move through the capillary first.
- An excitation laser shines through the capillary and the light emitted by the fluorescent dye is it returns to a lower energy level is detected by a detector system. As each coloured band is detected, it creates a signal which is processed by the sequencer and presented as a peak on a graph. Each peak represents a different base

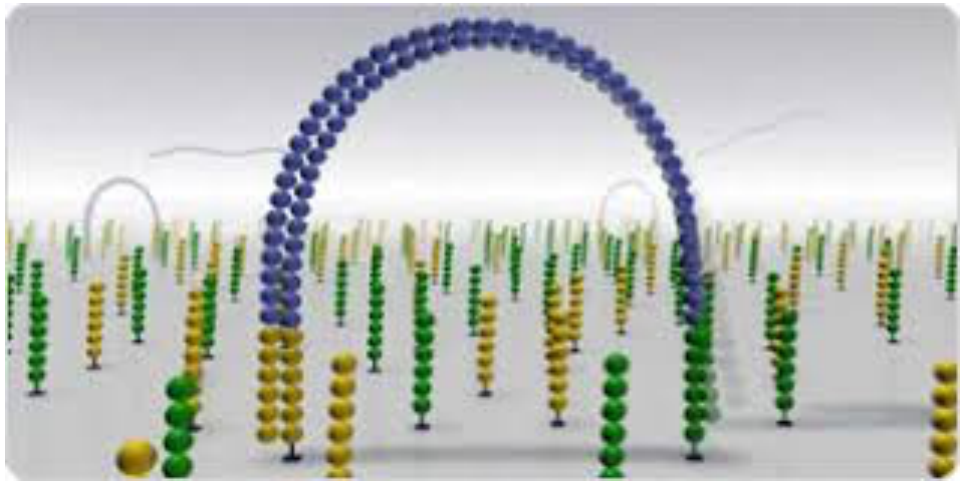
# Labeling of each dideoxy-type enables performing sequencing of 4 nucleotide types in only one lane






Lars Paulin Institute of Biotechnology University of Helsinki

# 3. Massive parallel seqeuncing

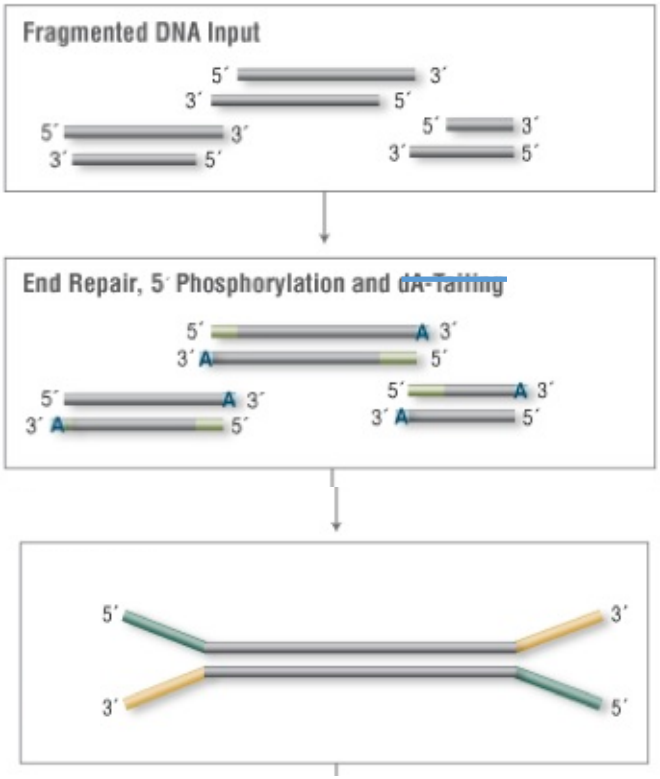


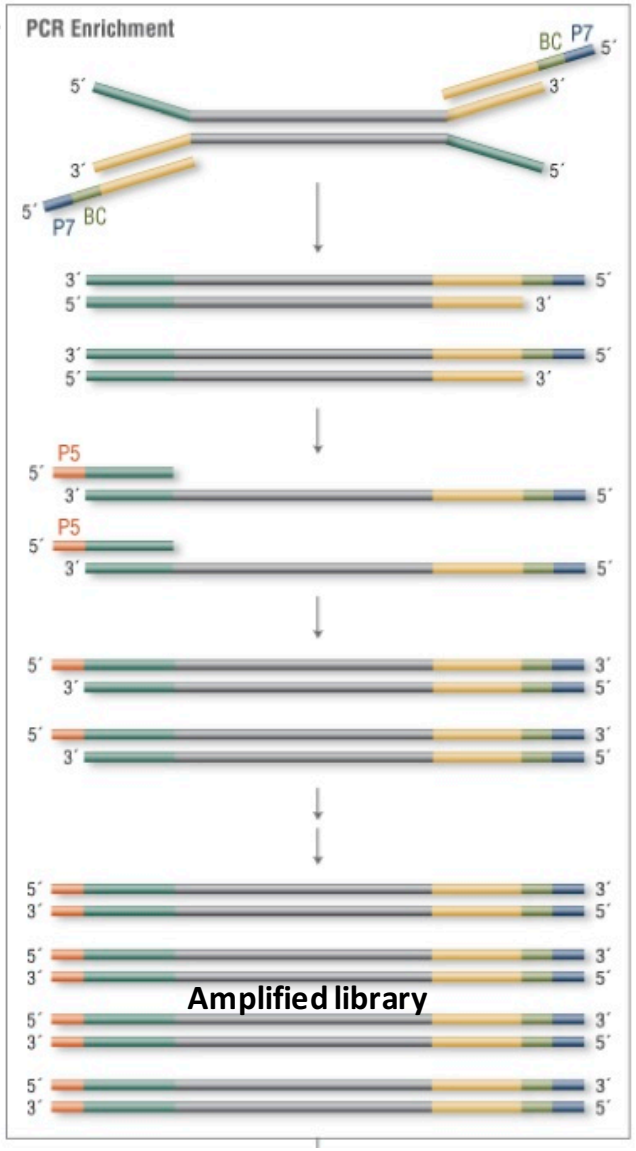
# Next generation sequencing of pools of DNAs

## Use DNA to generate DNA libraries:

→ Genomic DNA (fragemented)

→ Other libraries (cDNA, ChIP,…)




Linkers serve as uniform primer binding sites.
This allows the amplification of the entire DNA library using only 2 types of oligonucleotides

BC: barcode.
Each biological sample has common P7 oligos (blue and yellow) and P5 oligos (red/green); however for each biological sample a defined BC sequence is chosen.
This links the sequencing result to the biological sample
→ Many samples can be sequenced at the same time
→ (libraries are prepared separately)

**READY FOR MASSIVE PARALLEL SEQEUNCING**

# ChIP seq: Analysis of epigenetic information on the single nucleotide level → GENERASTION OF GENOME WIDE EPIGENTIC MAPS

## Illumina Massively Parallel Sequencing

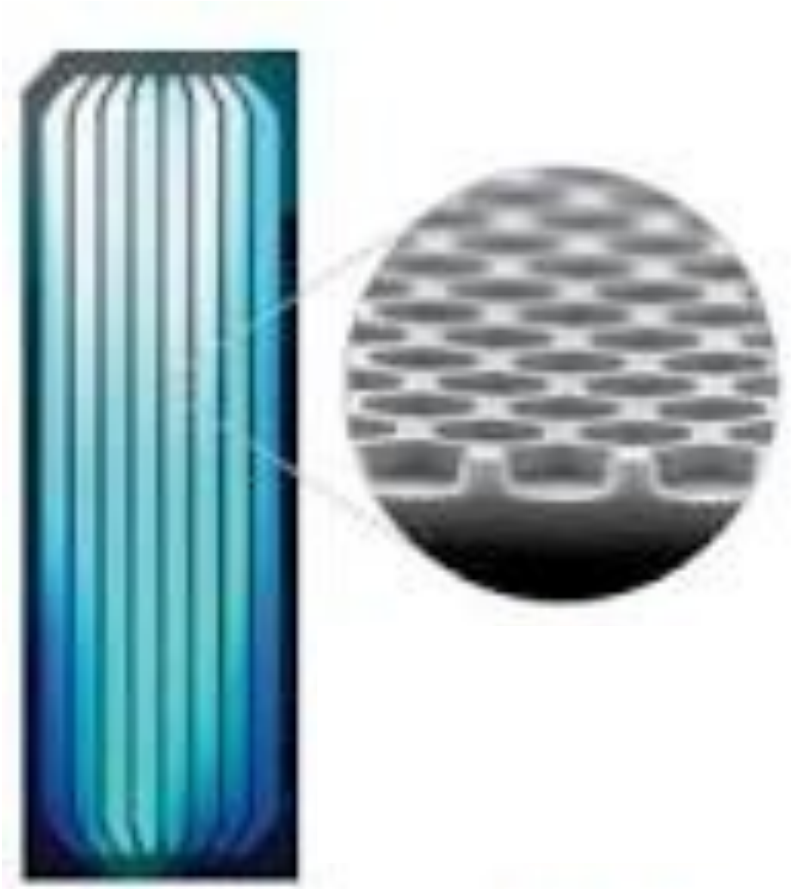
https://www.illumina.com/company/video-hub/pfZp5Vgsbw0.html


**The heart of the Illumina Massive Parallel Sequencer is the "FLOW-CELL". A surface with millions of small wells that allow thousands of Sanger-sequencing reaction In parallel = "massive parallel sequencing". In each well a SINGLE MOLECULE of DNA Is amplified and sequenced**

**Illumina offers the most potent massive sequencing instruments – leader on the market**

https://www.youtube.com/watch?v=pfZp5Vgsbw0

# CLUSTER AMPLIFICATION:



**Flow cell contains surface with millions of wells**

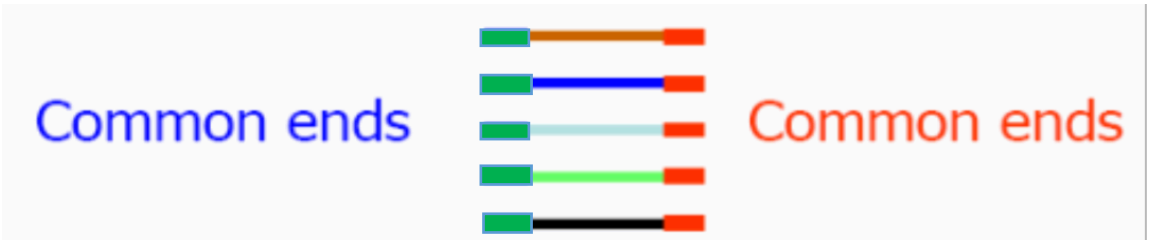
**→Each well contains beads mounted with 2 species of oligonucleotides that hybridize with adaptor oligos of DNA library**

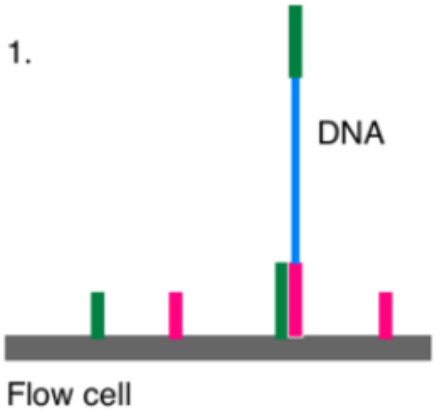
**→DNA library will be loaded onto the flow cell in a determined concentration:**

**ONLY ONE MOLECULE OF DNA WILL BE PROCESSED FOR SEQUENCING IN A SINGLE WELL**

# CLUSTER AMPLIFICATION:

-making DNA library (~300bp fragments)
-ligation of adapters **A** and **B** to the fragments



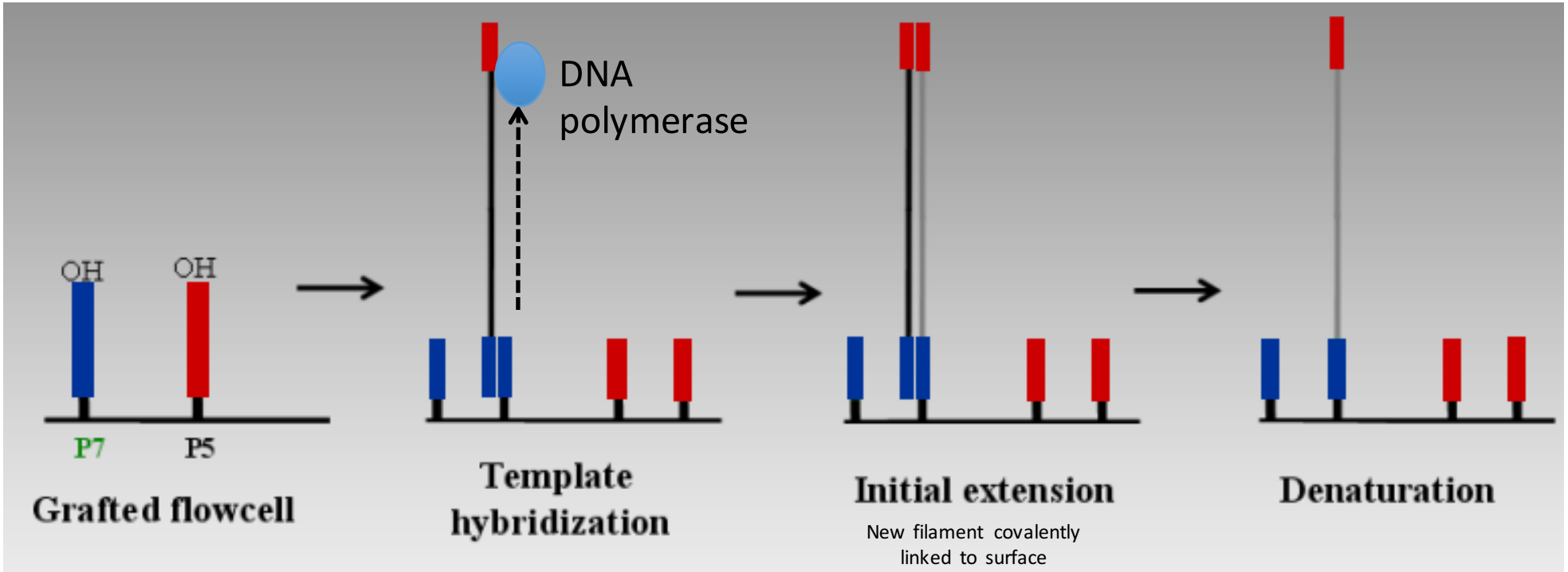

1 well, covered with
millions of 2 types of oligos

1 well in a flow-cell with
billions of wells

- ***complementary*** primers are ligated to the surface
- pairing with ChiP ed ssDNA at random position in the well of the flow cell

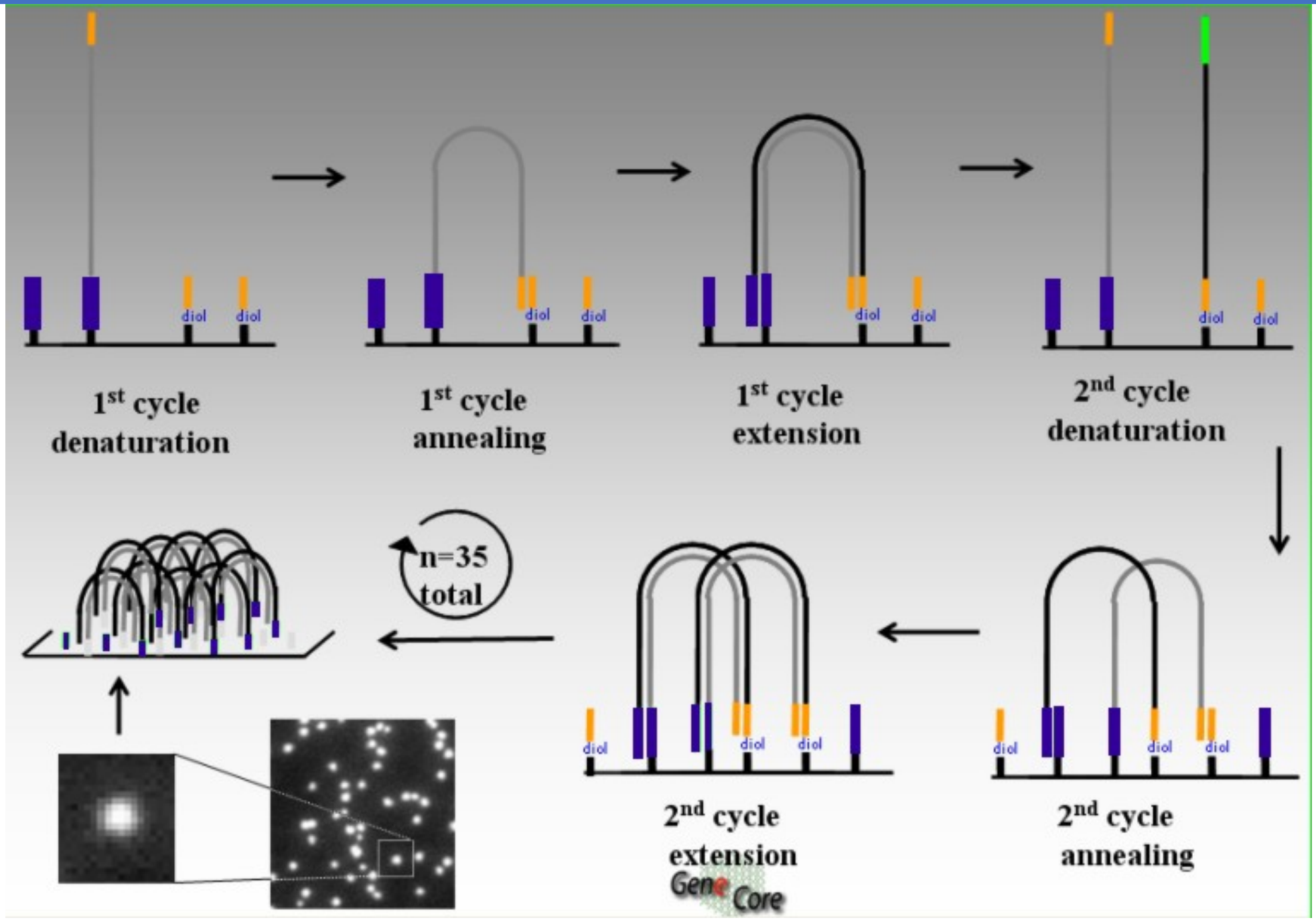
# CLUSTER AMPLIFICATION:

**Bridge amplification**: takes place on surface of beads (each bead is mounted with 2 species of oligos; each oligo can hybridize to a DNA library fragment): initiation

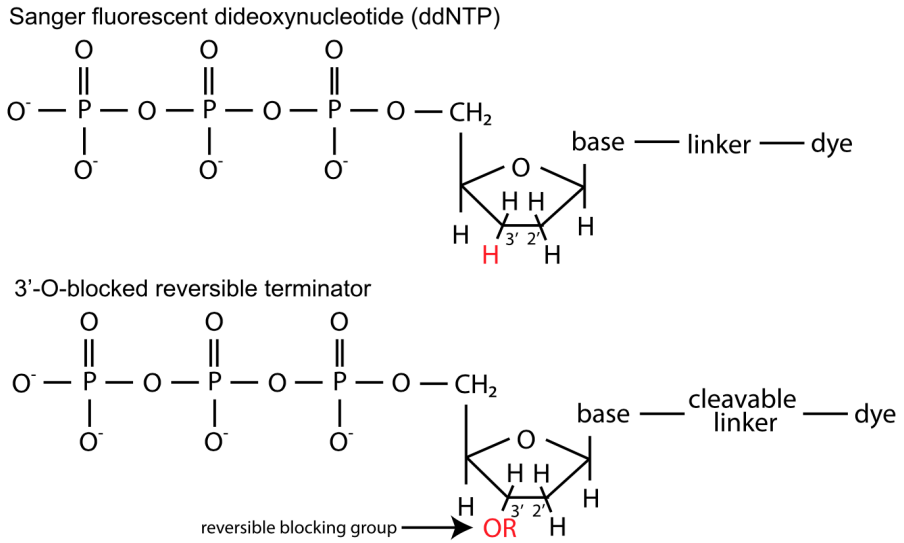



On the surface: complementary oligos

# CLUSTER AMPLIFICATION:



# REVERSIBLE CHAIN TERMINATORS:




Instead of promoting irreversible primer extension like the Sanger dye terminator method, the reversible chain terminators method uses a cyclic method that consists of nucleotide incorporation, fluorescence imaging and cleavage.
The figure shows a modified nucleotide with a **cleavable dye** and **reversible blocking group. Once the blocking group is removed, a 3'OH is formed and a new nucleotide may come in.**

NOTE: no classic dNTPs are used for sequencing!!!!

## Procedure

The steps for such a process can be outlined as follows:

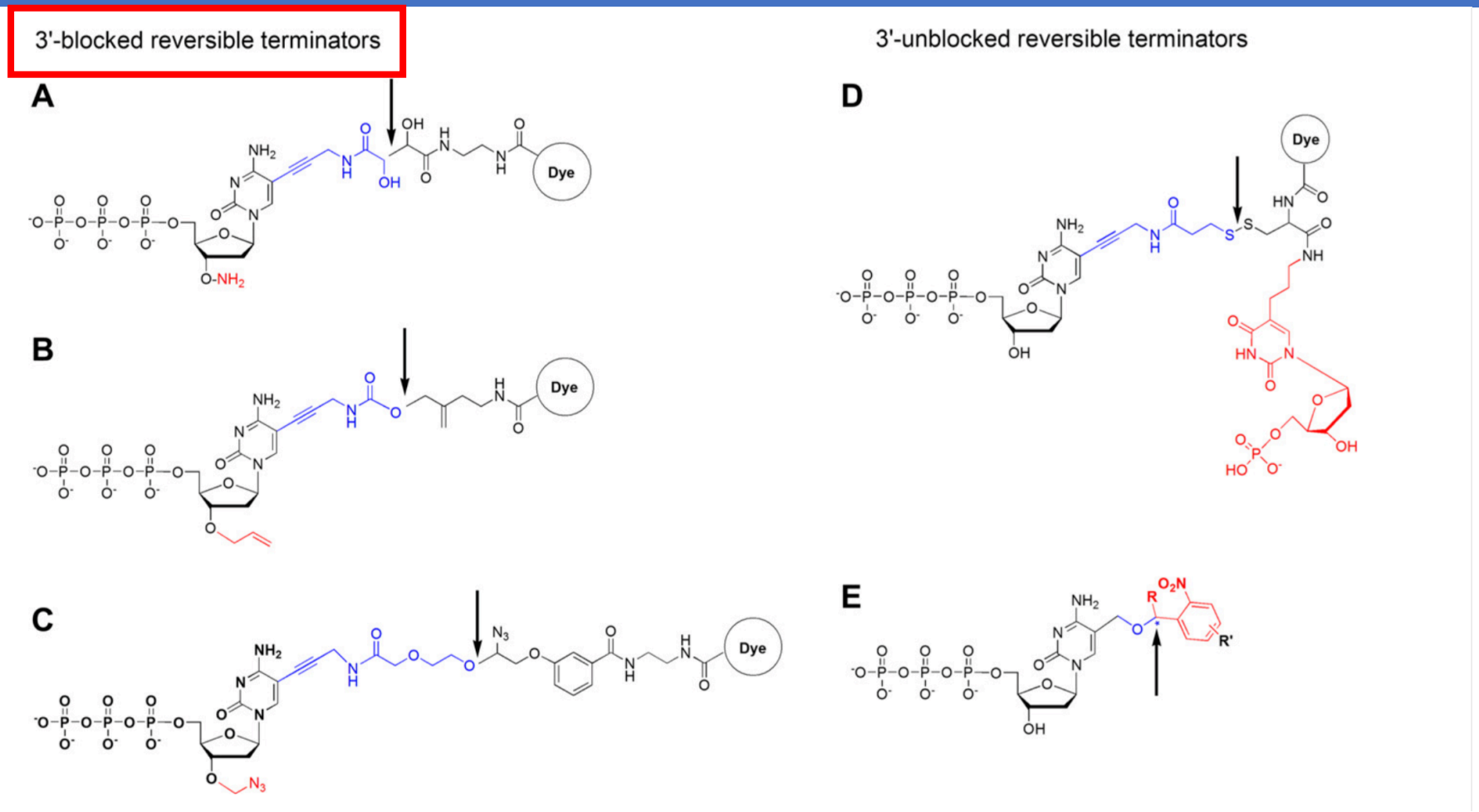
1. Have four dNTP's, each with a different fluorescent marking. These markings should not interfere with base pairing or phosphodiester bond formation.
2. Each dNTP should terminate DNA elongation temporarily with a blocking group on the 3' carbon of the sugar moiety.
3. Upon each cycle, have just one dNTP bind to the elongating strand and emit a fluorescent dye color.
4. Depending on the color emitted, record the particular nucleotide.
5. Cleave the blocking group and fluorescent dye with a palladium-catalyst.
6. Restore a 3' hydroxyl so that the growing strand can now elongate.
7. Repeat from step 1.

## Cons

There are some limitations to this method which include:

- Incomplete cleavage of blocking groups.
- Difficulties incorporating fluorescent nucleotides.

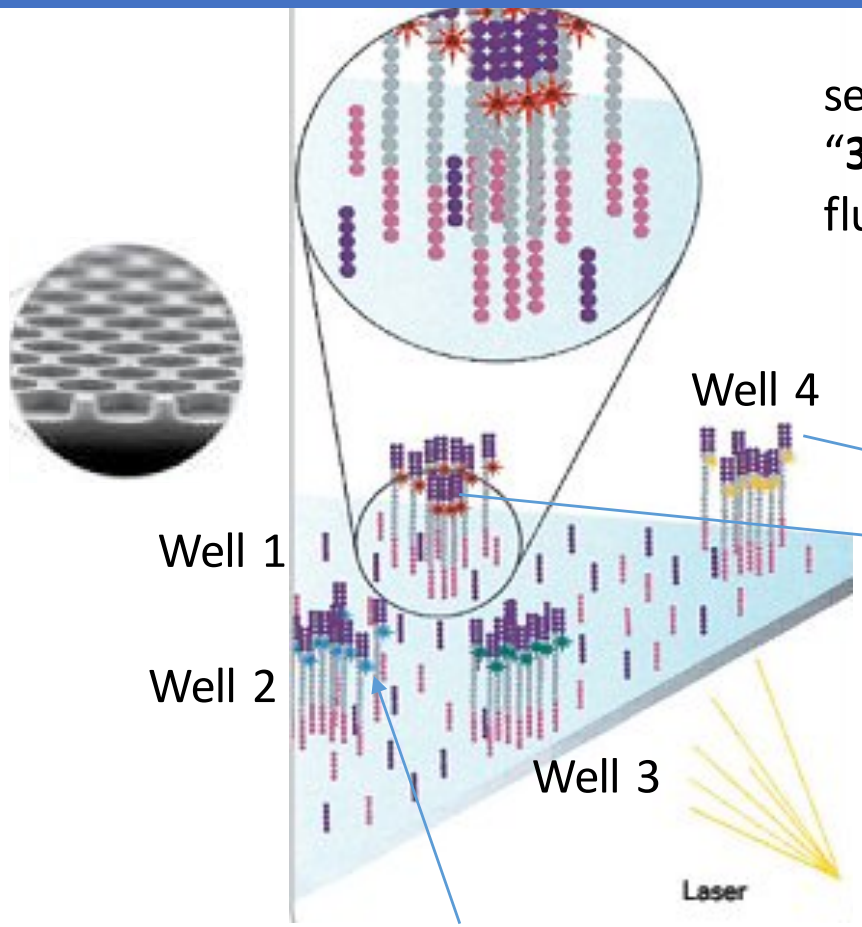
# CLUSTER AMPLIFICATION:



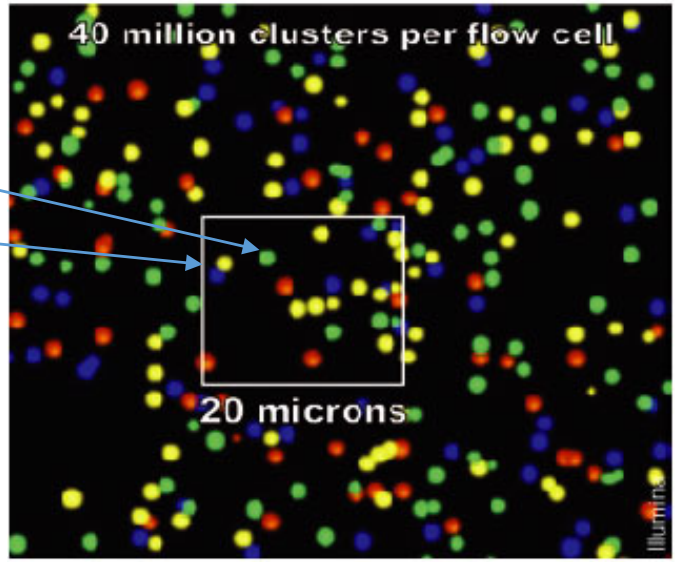

Three different 3’-blocked reversible terminators were shown on the left (A–C) and two 3’-unblocked reversible terminators were shown on the right (D–E).

The chemical structures in red denote the reversible terminating groups. Arrows indicate the site of cleavage separating the fluorescent groups from the nucleotide, and the chemical structures in blue denote the molecular scars that are attached to the base.

# Illumina: massive parallel sequencing:




sequencing by synthesis with
"**3' blocked reversible terminator**" + cleavable fluorescently label (for each nucleotide)




1. **Start of synthesis using primer = incorporation of fluorescent 3'blocked reversible terminator: synthesis blocked**
2. **Scanning of fluorescent signals of all wells of flow-cell with laser (image)**
3. **Dye cleavage + elimination of reversible blocking group**
4. **wash step**
1. **Repeat steps 1-4 ca. 150x**

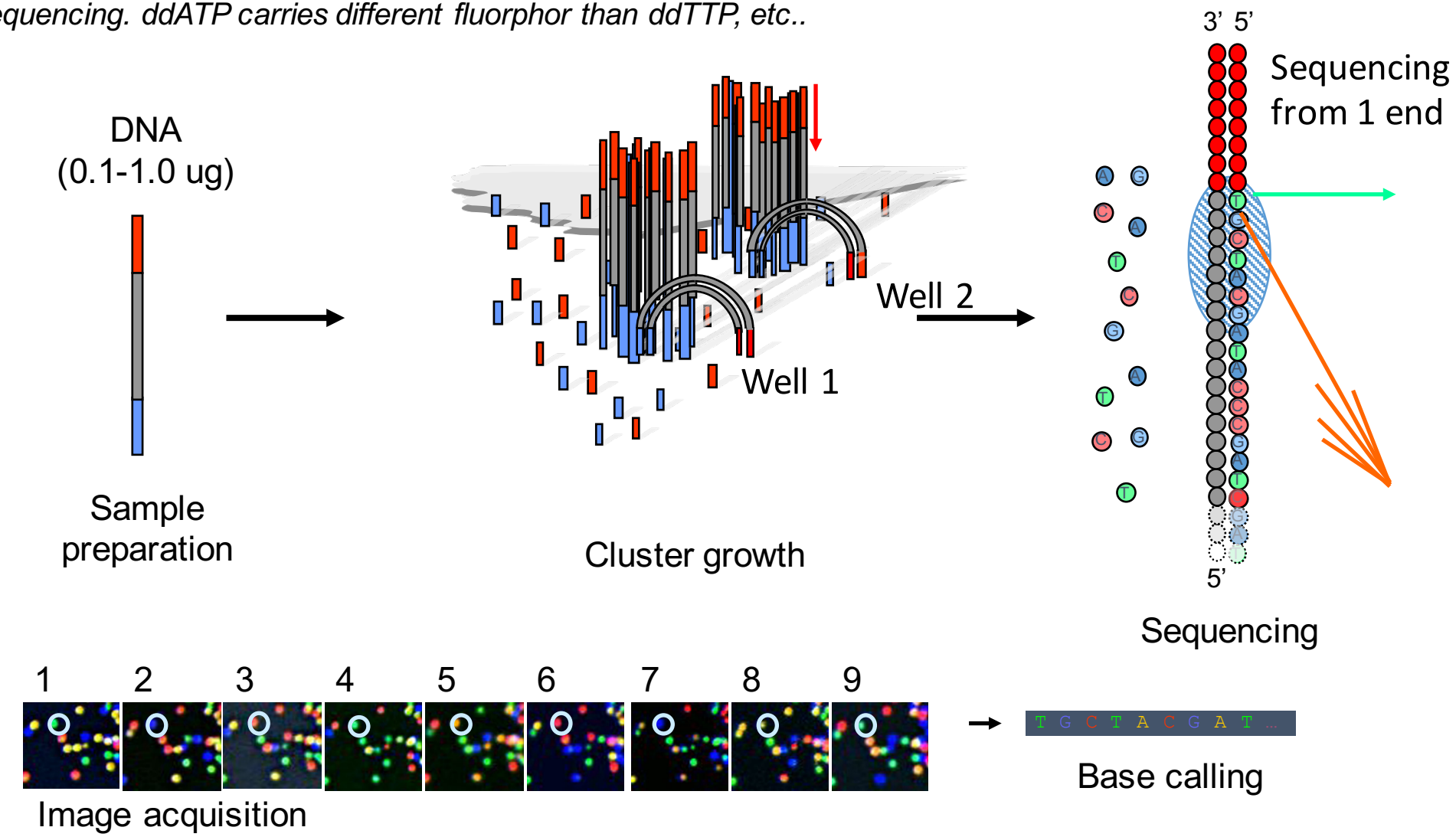
**READ LENGTH: ca: 150nt from each primer (2x150nt = 300nt)**

illumina.com

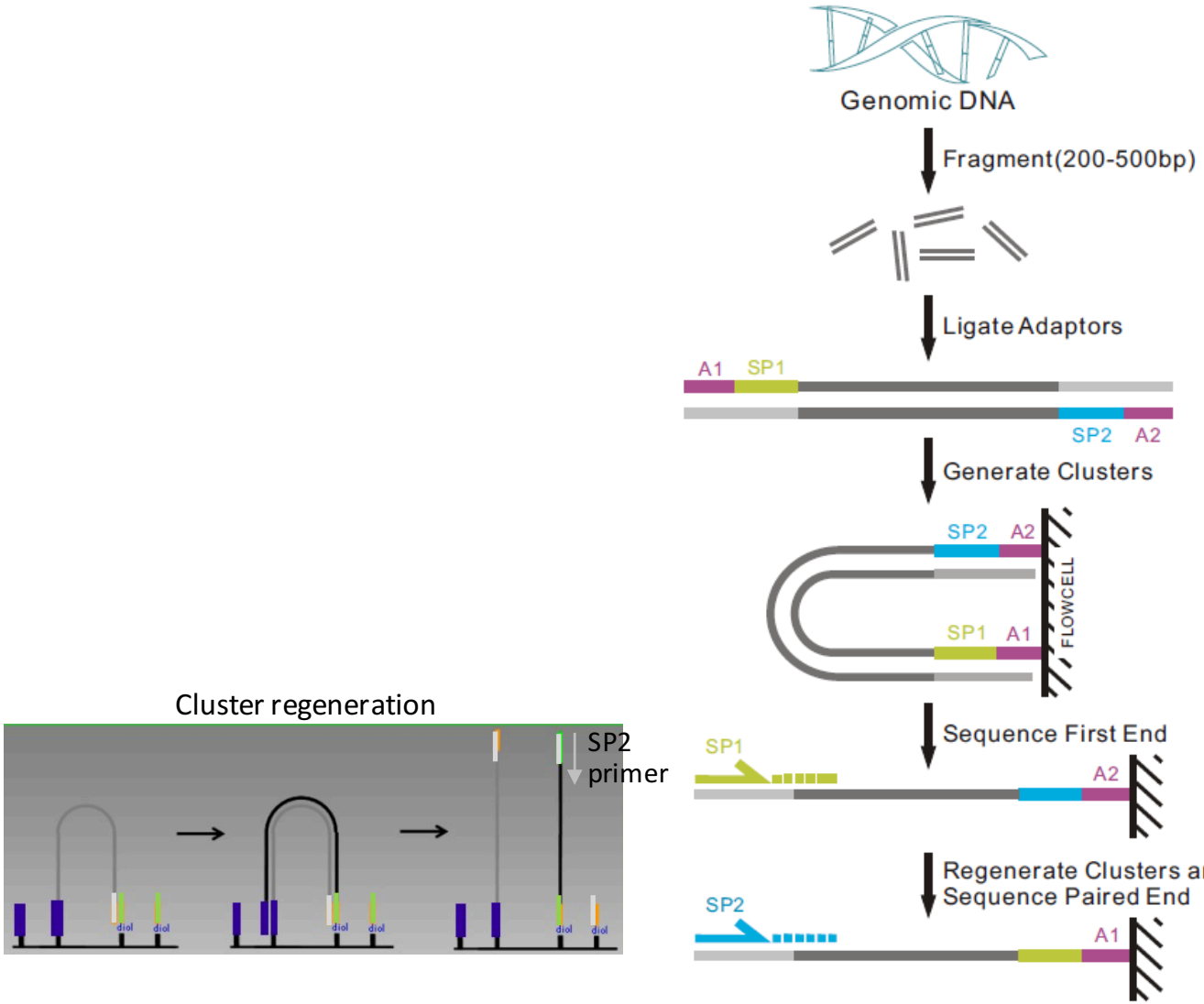
# Illumina Sequencing Technology

## *Robust Reversible Terminator Chemistry Foundation*

*In each round of sequencing a fluorescently labelled ddNTP will be used for sequencing. ddATP carries different fluorphor than ddTTP, etc..*



# Illumina: paired end sequencing increases information content




1° strand sequencing by SP1

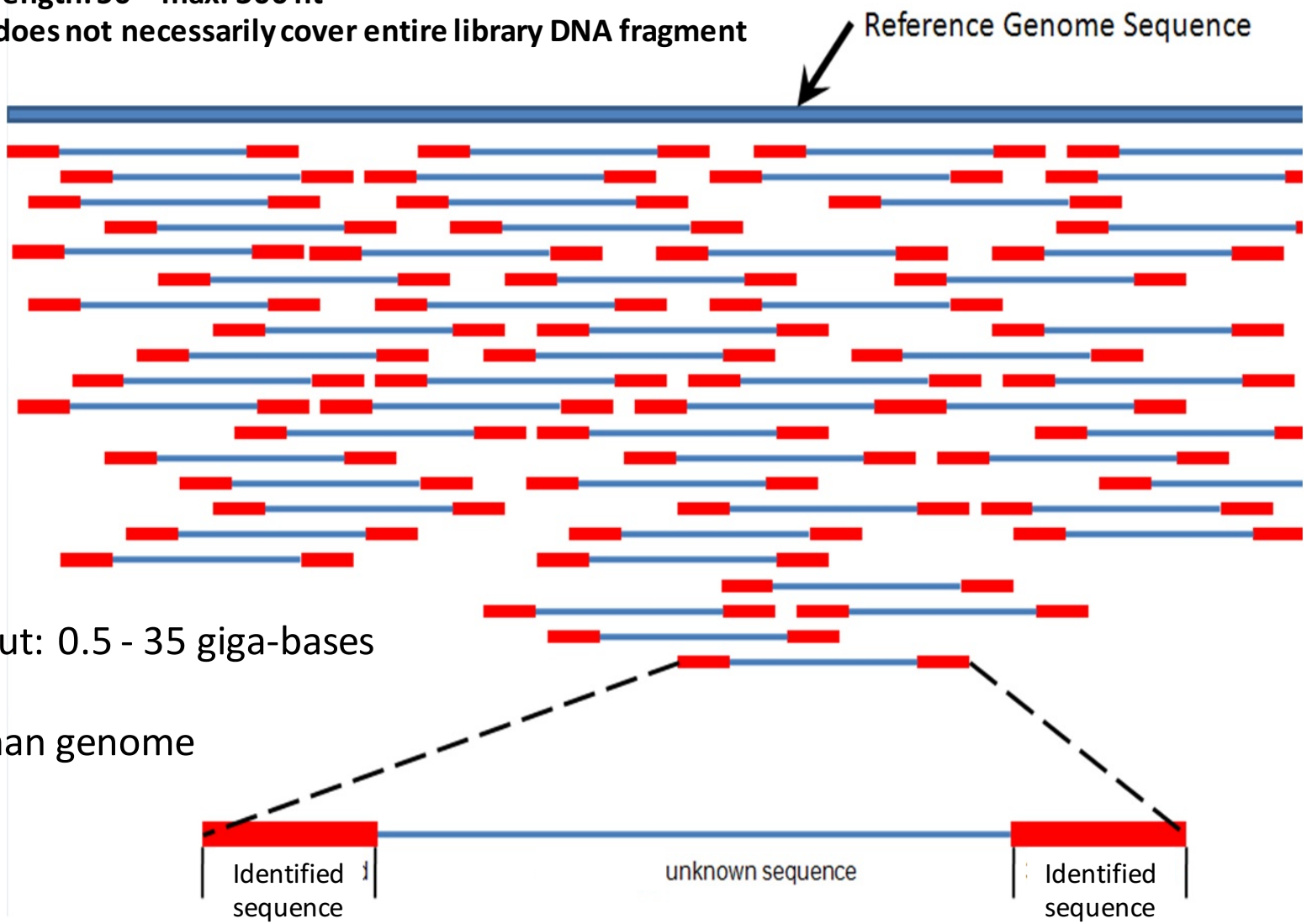
2° strand sequencing by SP2

Figure 1-2-1 Pipeline of paired-end sequencing (www.illumina.com)

https://www.youtube.com/watch?v=9YxExTSwgPM

# Data analysis: obtained sequence reads are aligned along genomic DNA sequence → high number of reads necessary to obtain full sequence coverage

**Read length: 50 – max. 300 nt**
**Read does not necessarily cover entire library DNA fragment**




Max. output: 0.5 - 35 giga-bases
$=3.5*10^{10}$
= 10x human genome

***Sequence derived from one amplified cluster***

# PILE – UP ALIGNEMENT ACROSS THE REFERENCE GENOME