MED-HUB

BIOCHEMISTRY

Molecular

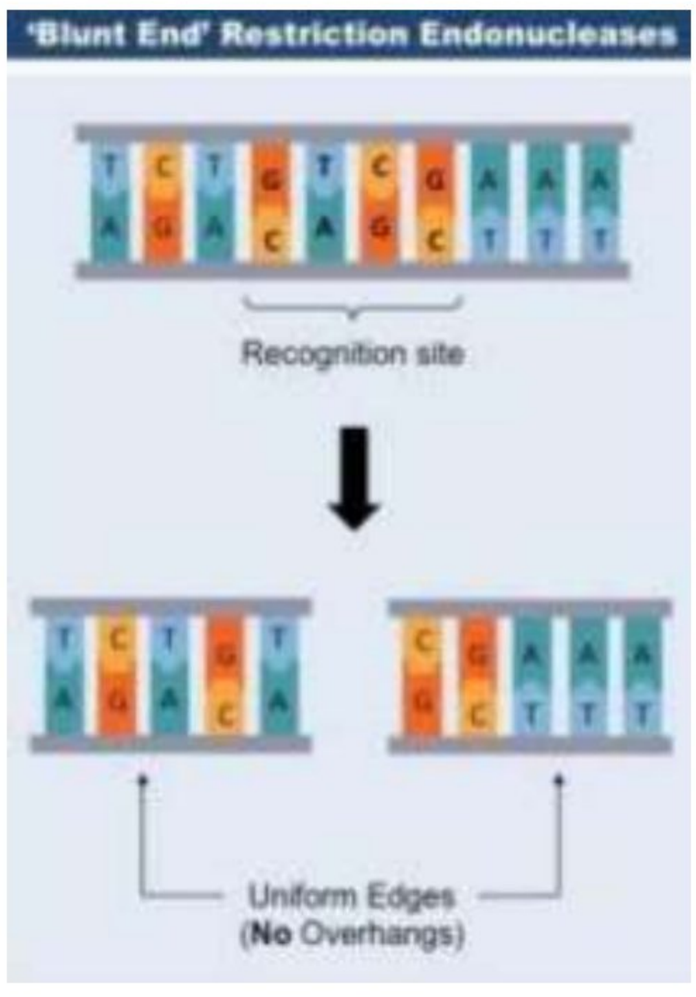
Molecular Lecture 1 Summary

- **Recombinant DNA**: A piece of DNA that is made from **2 or more different sources.**
- **Restriction endonucleases**: Bacterial enzymes that recognize and **break the phosphodiester bond** between nucleotides at specific sequences called **restriction sites**, generating **restriction fragments**.
- Bacteria produce these enzymes to **protect themselves from bacteriophages**, as these enzymes can degrade bacteriophages' DNA, and *that's why they are called restriction* (they restrict the growth of the bacteriophage).
- **Restriction sites** are **palindromic sequences**, meaning that they can be read the same from either strand, provided that the direction (5'-->3') or (3'-->5') is preserved.

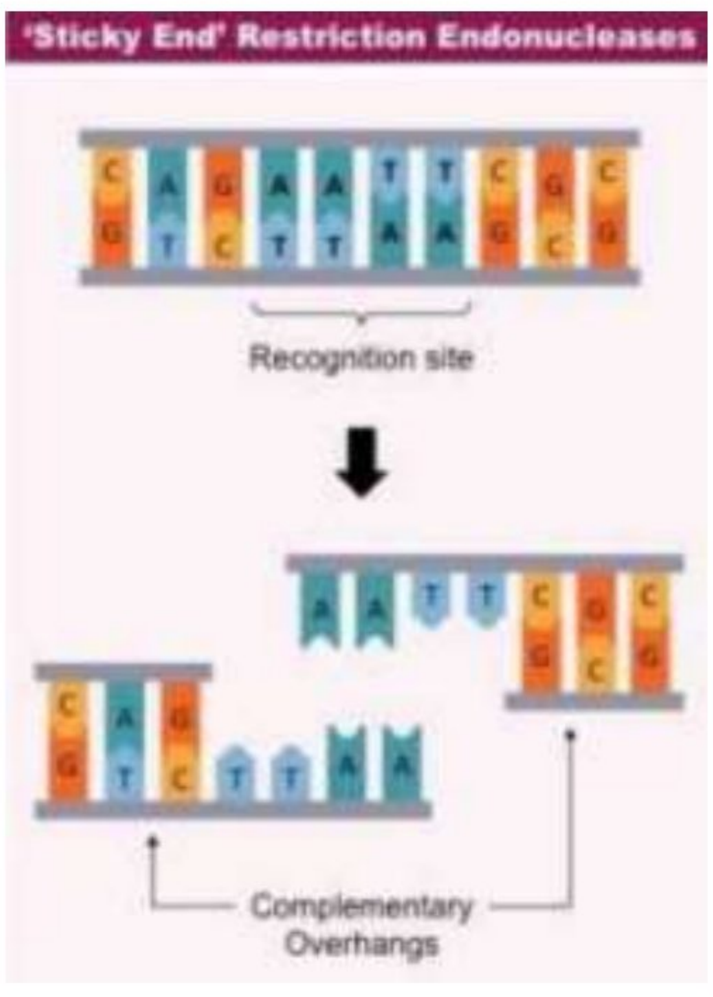
Restriction enzymes cut DNA in two different ways:

1. **Blunt**: enzymes cut at the **same position on both strands** giving blunt-ended (sharp) fragments.
2. **Staggered** (off-center): enzymes cut at **two different positions** generating **sticky or cohesive ends.**

Blunt



Staggered



How to make a recombinant DNA?

- You get DNA from 2 different sources.
- You expose them to the **same restriction endonuclease,** which results in the formation of **restriction fragments** that are **complementary** to each other.
- **They will combine** and form hydrogen bonds, except that they are **not stable,** because hydrogen bonds are non-covalent bonds.
- We use the enzyme **ligase** to make phosphodiester bonds, this process is called **ligation**.

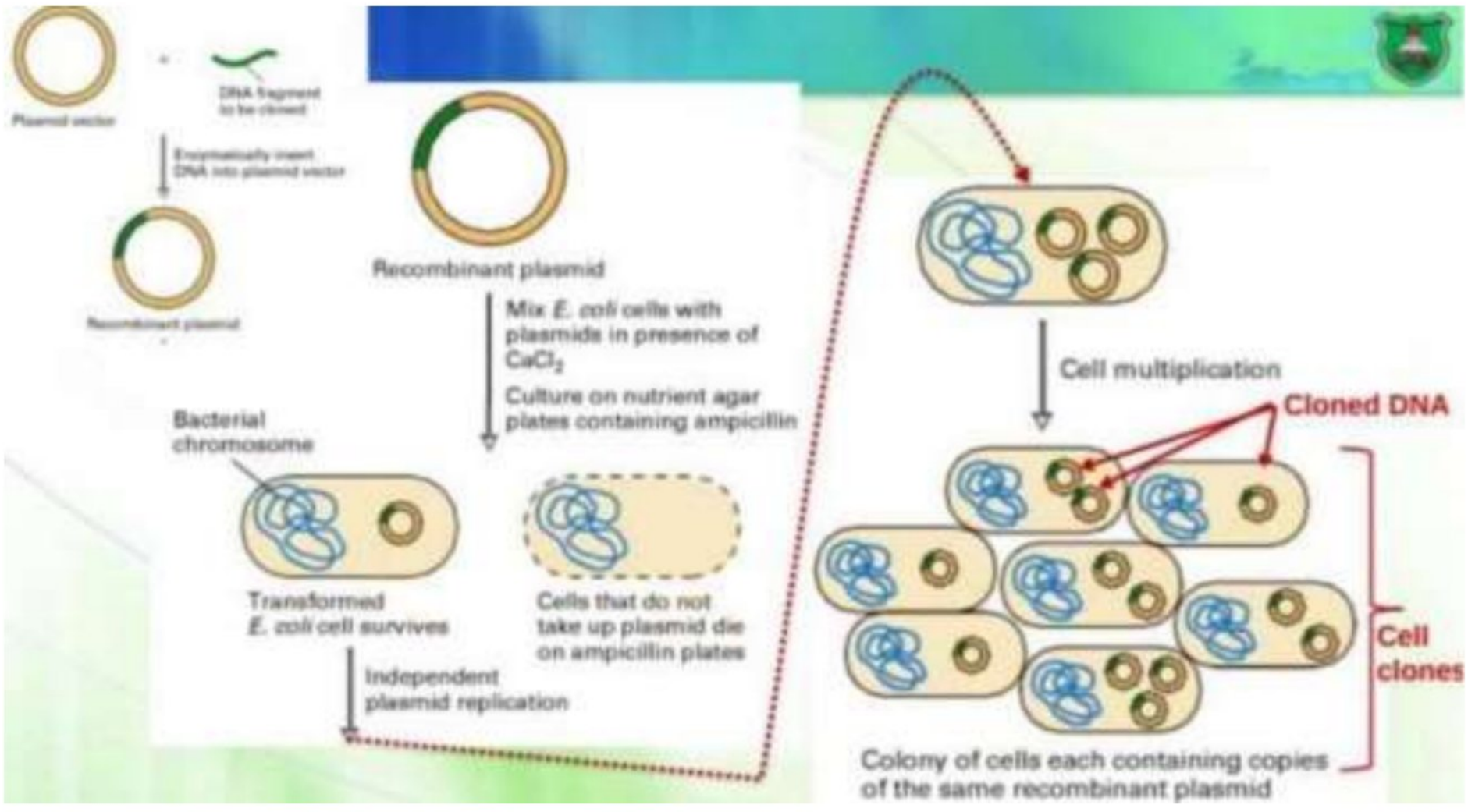
Notes:

☆ Ligase **covalently** joins DNA ends.

☆ Ligase uses **ATP**.

DNA cloning

- First, we make **recombinant DNA** composed of a **vector** (bacterial plasmid), and the DNA fragment of interest **(specific gene)**.
- Then, we insert the recombinant plasmid into bacterial cells.
- We add the antibiotic **to select the bacteria that have the plasmid**.
- Bacterial cells will make many, many copies of the plasmid, as the bacteria divide and the plasmid replicates **independently** inside bacteria.
- We isolate plasmids from bacterial cells and use the same restriction endonuclease to release the DNA fragment of interest.



- **In order to use a specific plasmid as a cloning vector, it should have the following three components:**

1. Their own **origin of replication** (OriC) that allows them to replicate independently of the bacterial chromosome.
2. **A selectable gene, like an antibiotic resistance gene**, to select the bacterial cells that have this plasmid.
3. **A restriction site** that allows for insertion of the DNA segment of interest into the plasmid.

- **In order to use the plasmid as an expression vector, it has to contain additional sequences** (in addition to the three components we have just mentioned):

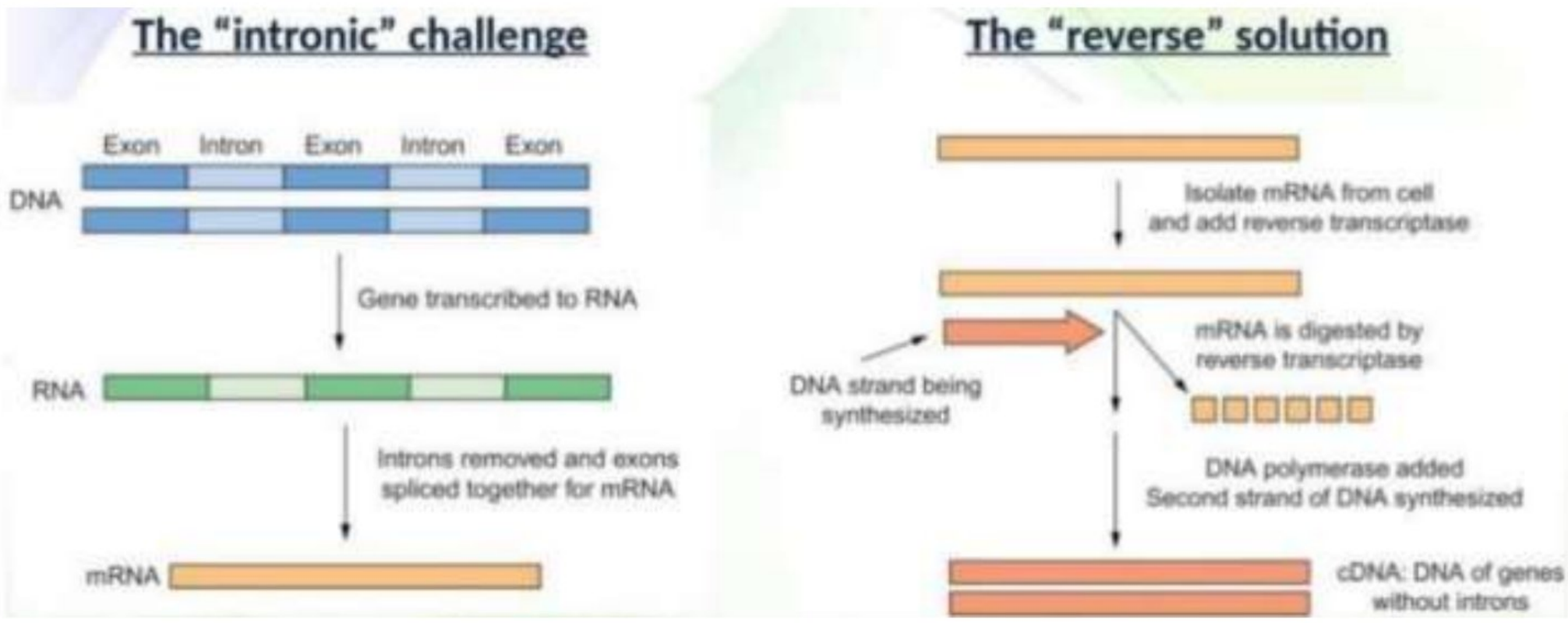
1. **Promoter sequence** upstream of gene to be inserted.
2. **Ribosomal binding sequences** (Shine-Dalgarno [SD] sequence in bacteria), the ribosome sits on this sequence initially and starts translation (from AUG) until it reaches a stop codon.
3. **A transcription termination sequence** at the end of the gene.

● We insert the expression vector into a bacterial cell, and the cell makes a lot of copies of the protein of interest, and this protein can then be purified.

Challenges and Solutions

● **First challenge: Bacteria don't have splicing techniques**, so introns would get translated.

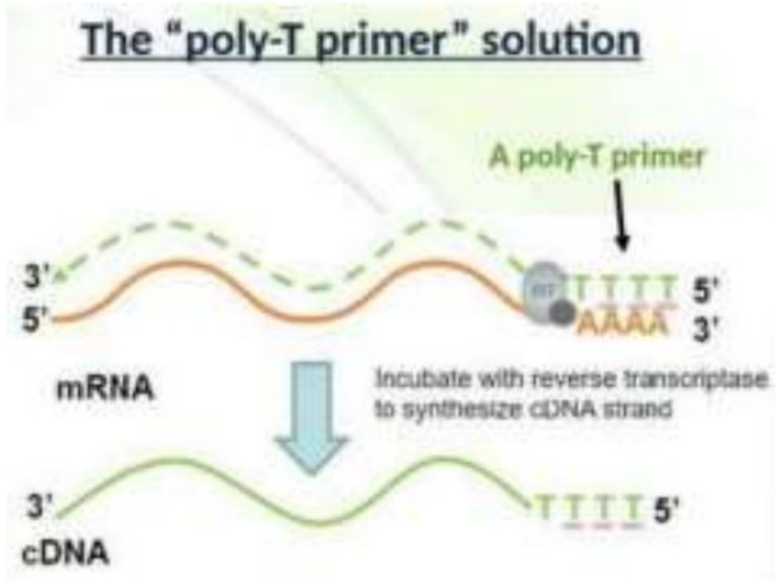
☆ **Solution**: We take the mature mRNA from human cells and convert it into cDNA using **reverse transcriptase**, and we insert this cDNA into a vector and continue the process.



Note: Reverse transcriptase requires a primer to function.

● **Second challenge:** There are many types of RNA molecules inside human cells, and we want to select the **protein-coding RNA** which is mRNA between other noncoding RNAs.

☆ **Solution**: mRNA has **poly-A tail,** so we target it with **poly-T primer** (DNA primer), **the reverse transcriptase** use the primer to start cDNA synthesis.



● Other challenges:

- **No internal disulfide bonds.**
- **No post-translational modification**, like glycosylation.
- **Protein misfolding**.
- **Protein degradation**.

☆ **Solution**: use a **eukaryotic system** like yeast.

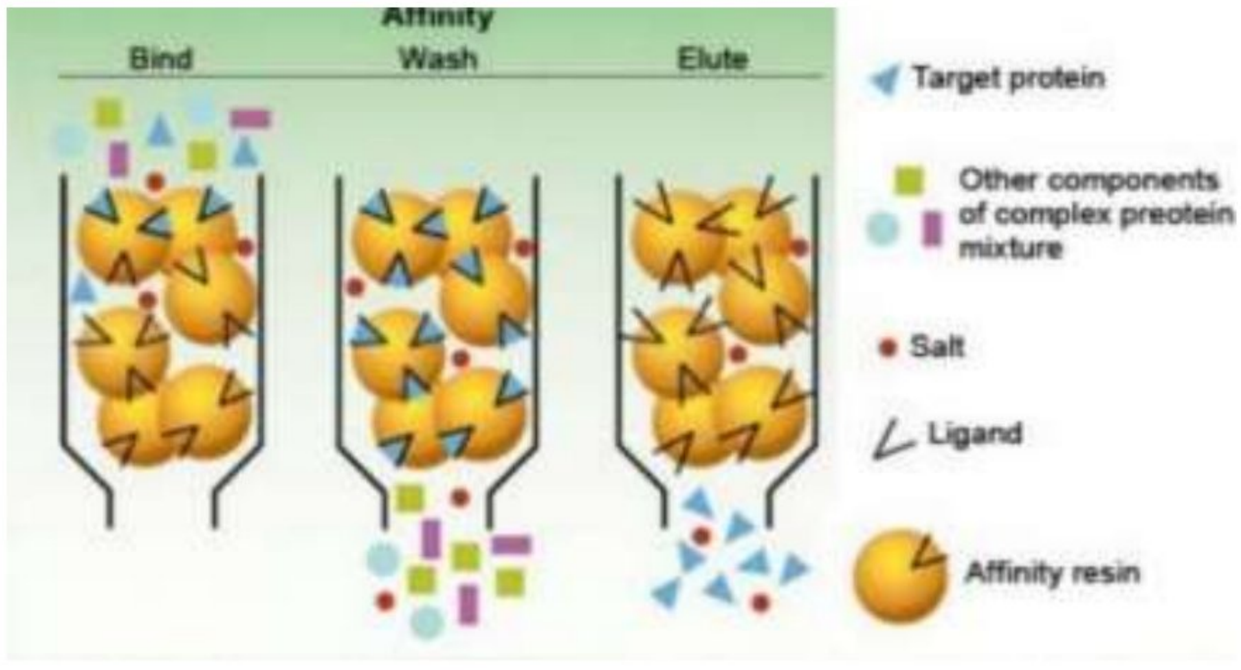
Protein tagging

● We use a vector that has the tag (a small piece of DNA that encodes amino acids sequence) as a part of it, and we insert the gene of interest next to this tag, so the tag becomes a part of the gene, and when the gene is expressed, the tag is going to be expressed with it and becomes a part of the protein.

Affinity chromatography

● It is used in **protein purification** (to separate a certain protein from other proteins), and here how it is done:

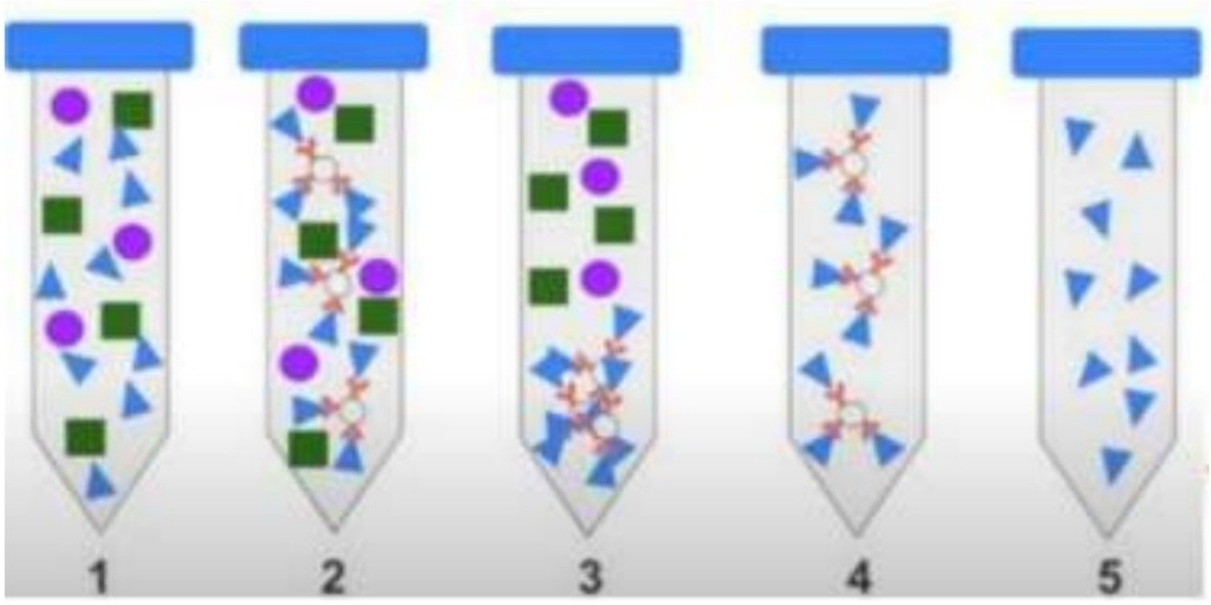
- We have beads (resins) inside the cylinder, and certain ligands are attached to these beads.
- We pass all proteins isolated from cells through the column, the protein that has affinity to the ligand will bind to it while other proteins would pass out.
- We isolate (release) the protein of interest from the column.



Immunoprecipitation

It is used in **protein purification** as well, and here how it is done:

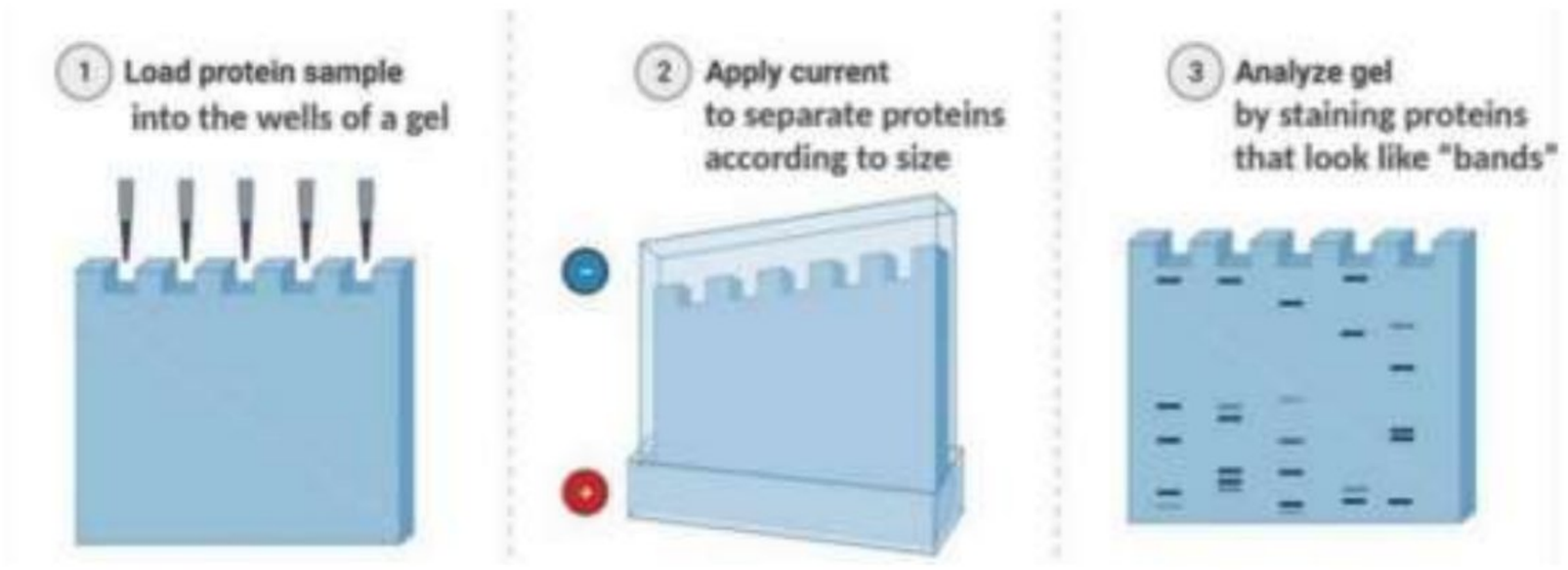
- We have a collection of proteins in the column.
- We add beads that have antibodies attached to their surface, so they will bind proteins that have affinity to these antibodies.
- The beads precipitate because they are heavy, dragging with them the protein that we want to isolate.
- We remove all other proteins, so now we only have the beads with the protein of interest bound to them.
- We release the protein of interest, so now we have it purified.



Gel electrophoresis (SDS-PAGE)

We use this technique to detect and **identify a protein**, and here how it is done:

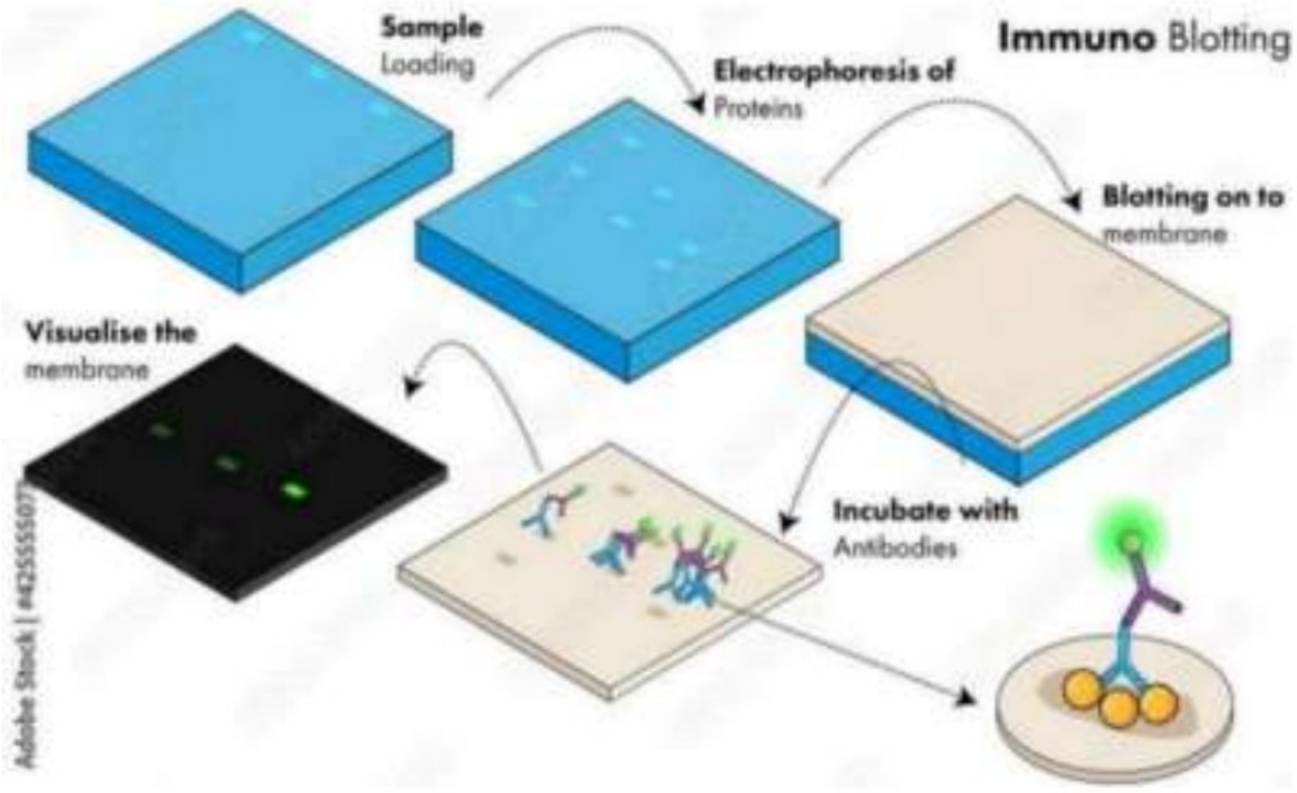
- We have a special gel and we apply our sample (collection of proteins) into wells.
- Electricity is applied, proteins migrate through the gel according to **size**.
- They look like bands, and every band represents many copies of the same protein or different proteins of the same size.



Immunoblotting (western blotting)

We use it in **protein detection**, and here how it is done (the same concept of southern blotting, but here we are dealing with proteins):

- After we have done SDS-PAGE and separated proteins according to size, we transfer the proteins into a membrane.
- We add **antibodies** that have signals attached to them, and these antibodies bind the protein of interest and give us signals, so the band lights up giving us a signal that the protein of interest is in this band.



Major protein tags

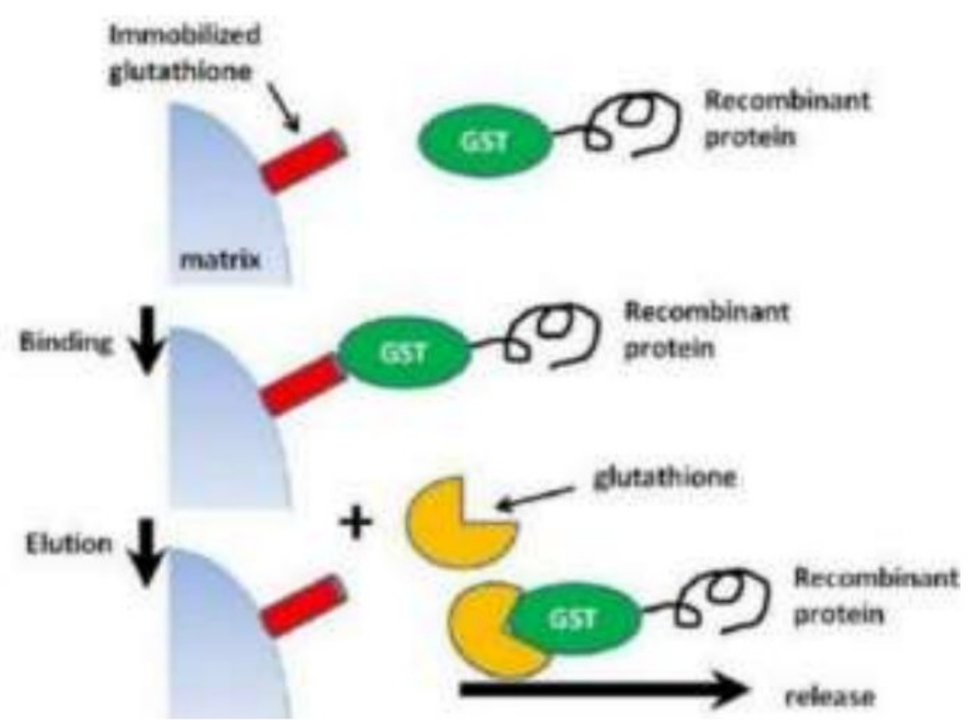
Poly-His	HHHHHH	antibody	nickel, imidazole
Glutathione S transferase (GST)	218 aa protein	antibody	glutathione
Green fluorescent proteins (GFP)	~220 aa protein	antibody or fluorescence	None

1. **Histidine tag (Poly-His)**

- It is composed of **six histidine** amino acids.
- The addition of His tag to a protein allows for **purification by affinity chromatography using beads with bound nickel ions,** as the nickel can bind to the six histidines.

2. **Glutathione S transferase (GST)**

- It is an enzyme, and its substrate is **glutathione**.
- It has the same concept as the Poly-His tag.

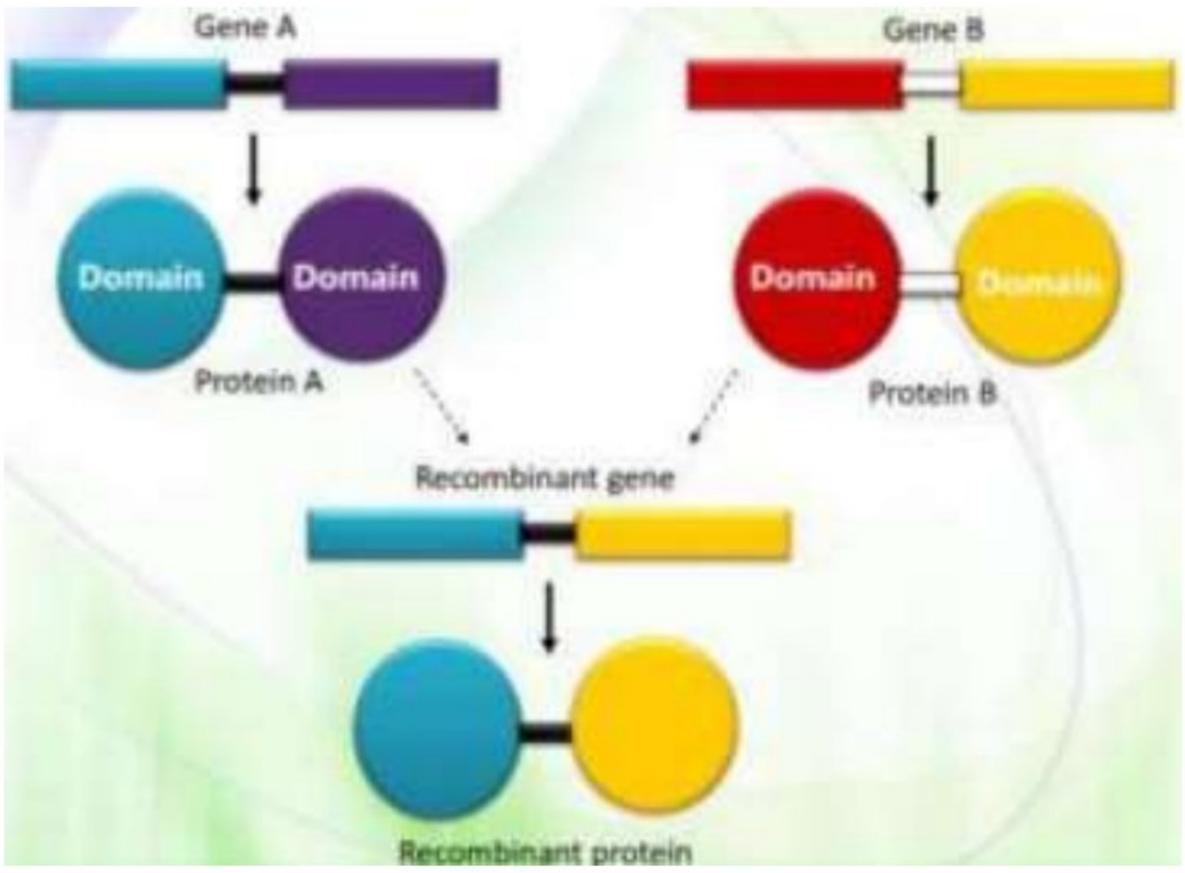


3. **Green fluorescent protein (GFP)**

- A protein that is produced from **jellyfish**.
- We use this protein for **protein detection not purification**, by attaching it to the protein of interest.
- Our protein of interest now **fluoresces**, not by itself, but because it is attached to GFP, so we can follow it.

Production of a recombinant protein

- We make a recombinant gene, and when the recombinant gene is expressed, the recombinant protein will be formed.
- **Domain**: A part of a protein that has **its own 3D-structure and function**, so **it can be cut from the protein and maintain its structure and function.**
- we can combine different domains from different proteins.

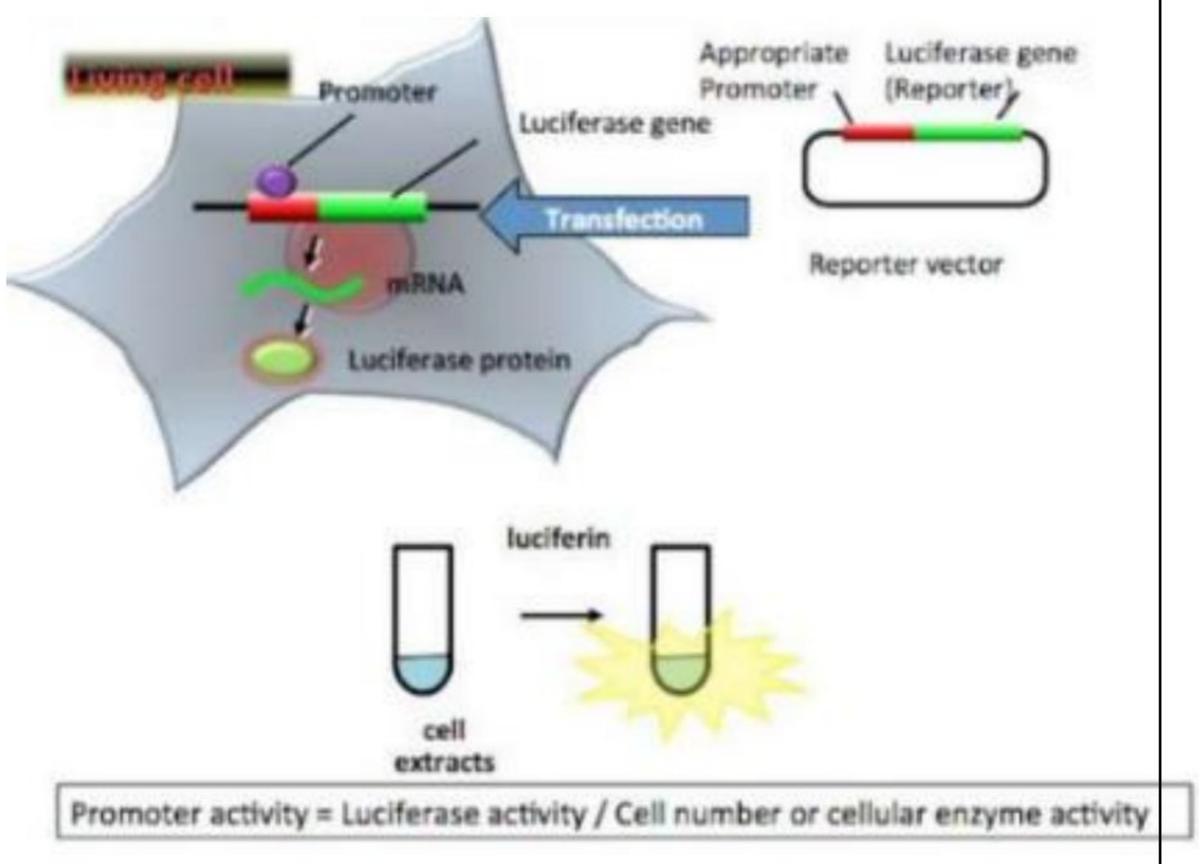


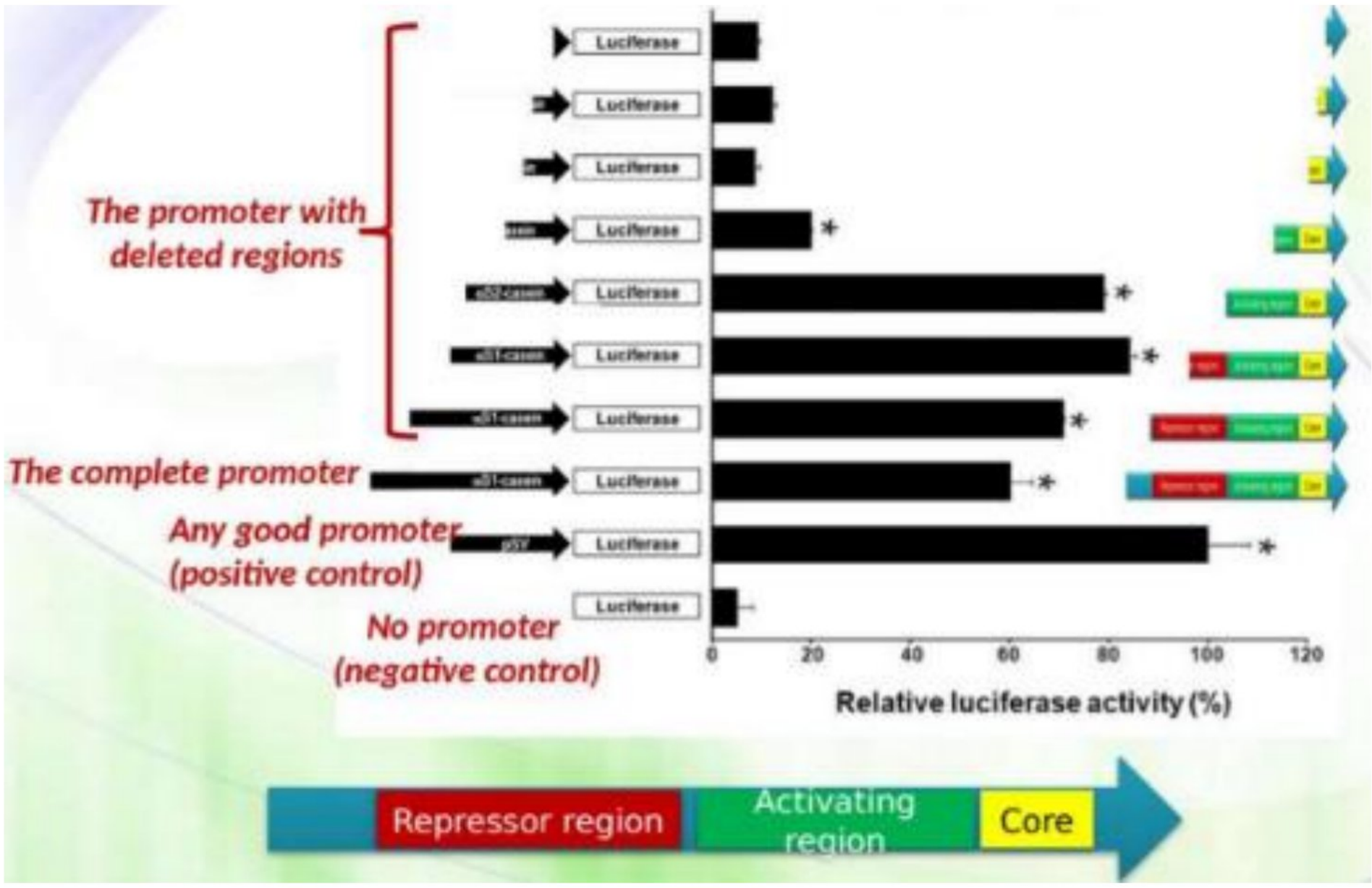
Molecular lecture 2 summary

Analysis of DNA regulatory sequences

- **Luciferase reporter assay** is used to study the function of certain regions of the promoter.
- **Luciferase**: An enzyme that is found in fireflies, that is used to produce light, through the conversion of luciferin into oxyluciferin.
  - **The process**:

- We take the promoter that we want to study and put it ahead (upstream) of the luciferase gene in a specific vector >> Now the luciferase gene is under control of the promoter that we want to study.
- We transfect (insert) this vector into living cell (biological system like yeast or cultured human cells)
- These cells will express luciferase gene under our promoter's control >> The production of luciferase and light is controlled by the promoter.
- We measure the amount of light produced:
  - *If the promoter does a positive control, the luciferase gene will be highly expressed, so there will be a high light production*.
  - *If the promoter does a negative control, the luciferase gene won't be expressed, so there won't be light production*.

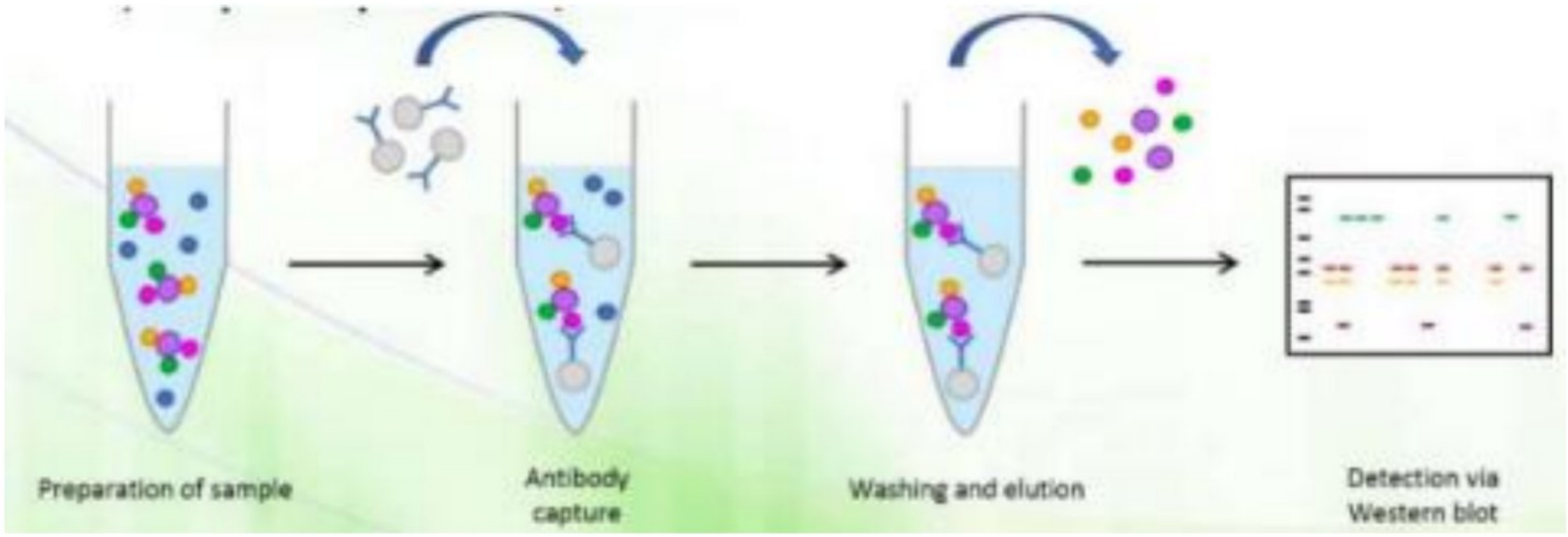




- **No promoter** >> **there is very low production of luciferase and light**. *This is called negative control*.
- **Any good promoter** >> **the amount of light produced is very high**. *This is called positive control*.
- **The complete promoter** >> to know the activity of our promoter.
- **The promoter with deleted** regions >> we delete a certain region and measure the amount of light produced. **If it increases, this means that the deleted piece does a negative control and if it decreases, this means that the piece does a positive control.**

Protein-protein interaction

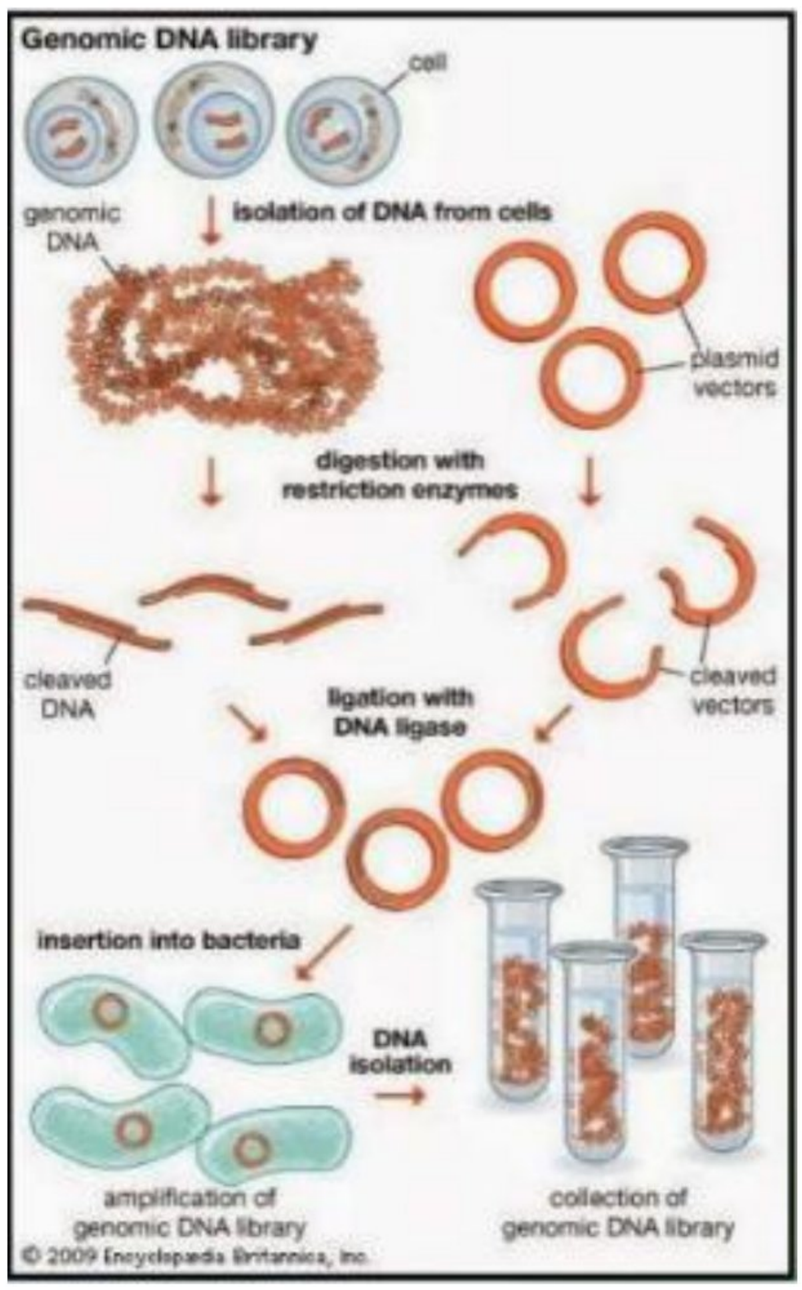
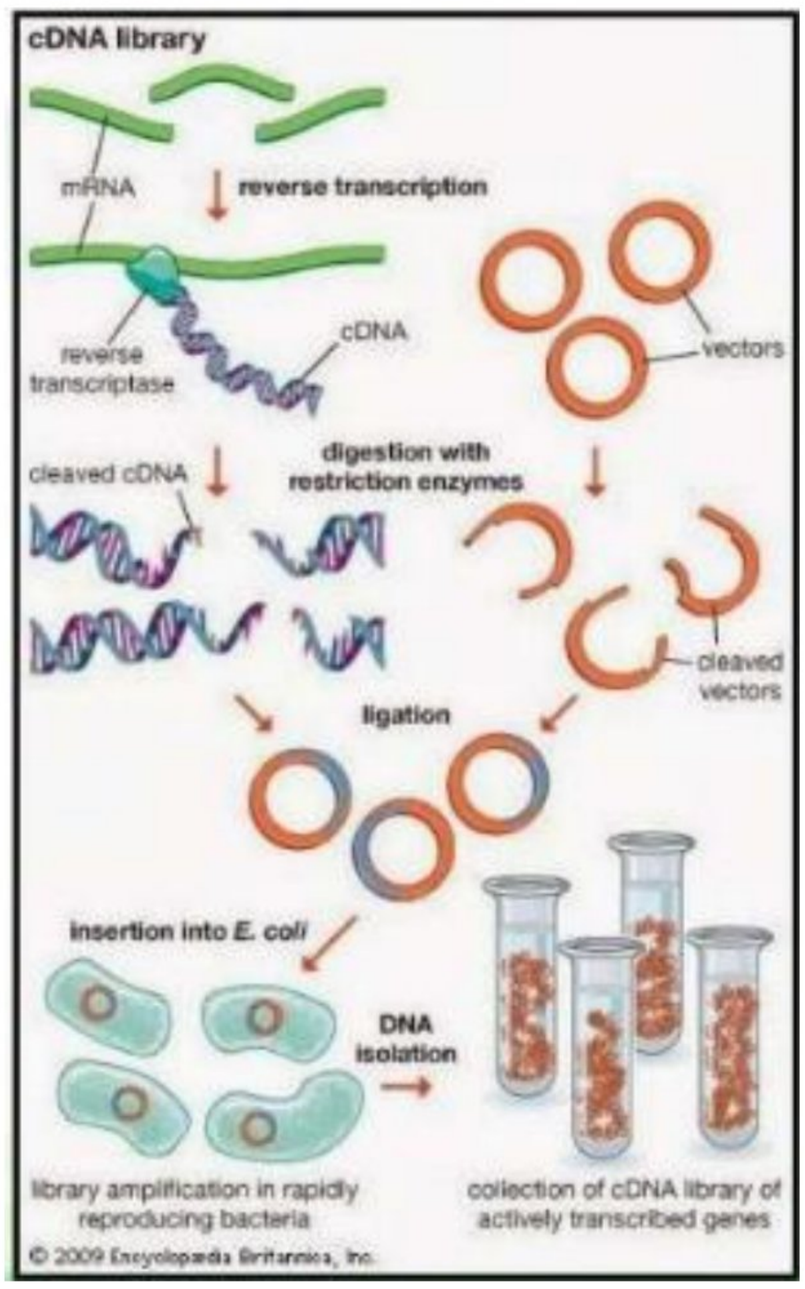
Co-immunoprecipitation



- **The difference between immunoprecipitation and co-immunoprecipitation is that here we don't pull down only our protein of interest, rather we pull down all of other proteins that interact with our protein,** then we release proteins and analyze them using SDS-PAGE or western blotting (immunoblotting).

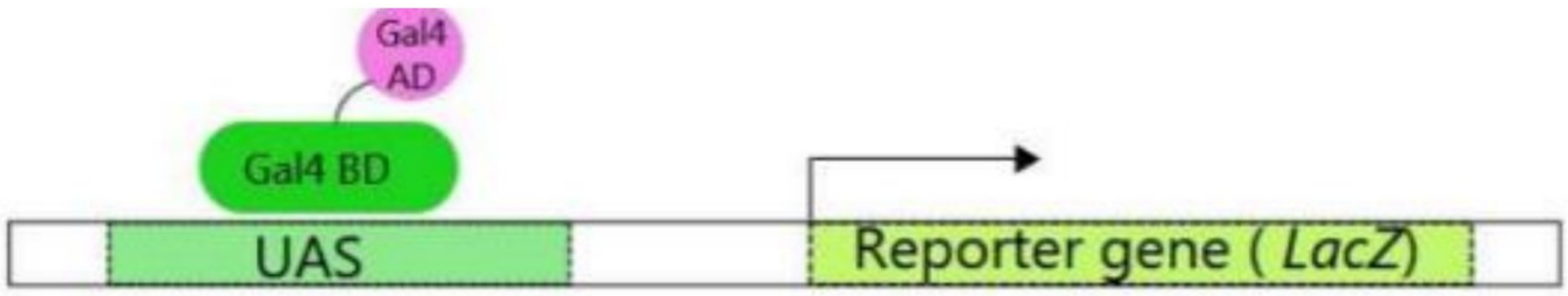
DNA library

- Collection of tubes containing different **DNA fragments** found in **plasmids** in **clones of bacteria** saved in **freezer**; every fragment has its specific position.

Genomic library	cDNA library
Genomic DNA **– the whole DNA** in a cell – is cut using restriction endonucleases into smaller fragments and these fragments are cloned	as we take **mRNAs** from cells and convert them into **cDNA** using **reverse transcriptase** and we insert these cDNA molecules inside plasmids and clone them.
we have the whole DNA (coding and non-coding regions)	**we have only coding regions (exons)**
The same among different types of body cells.	**Different** between different types of body cells.
	

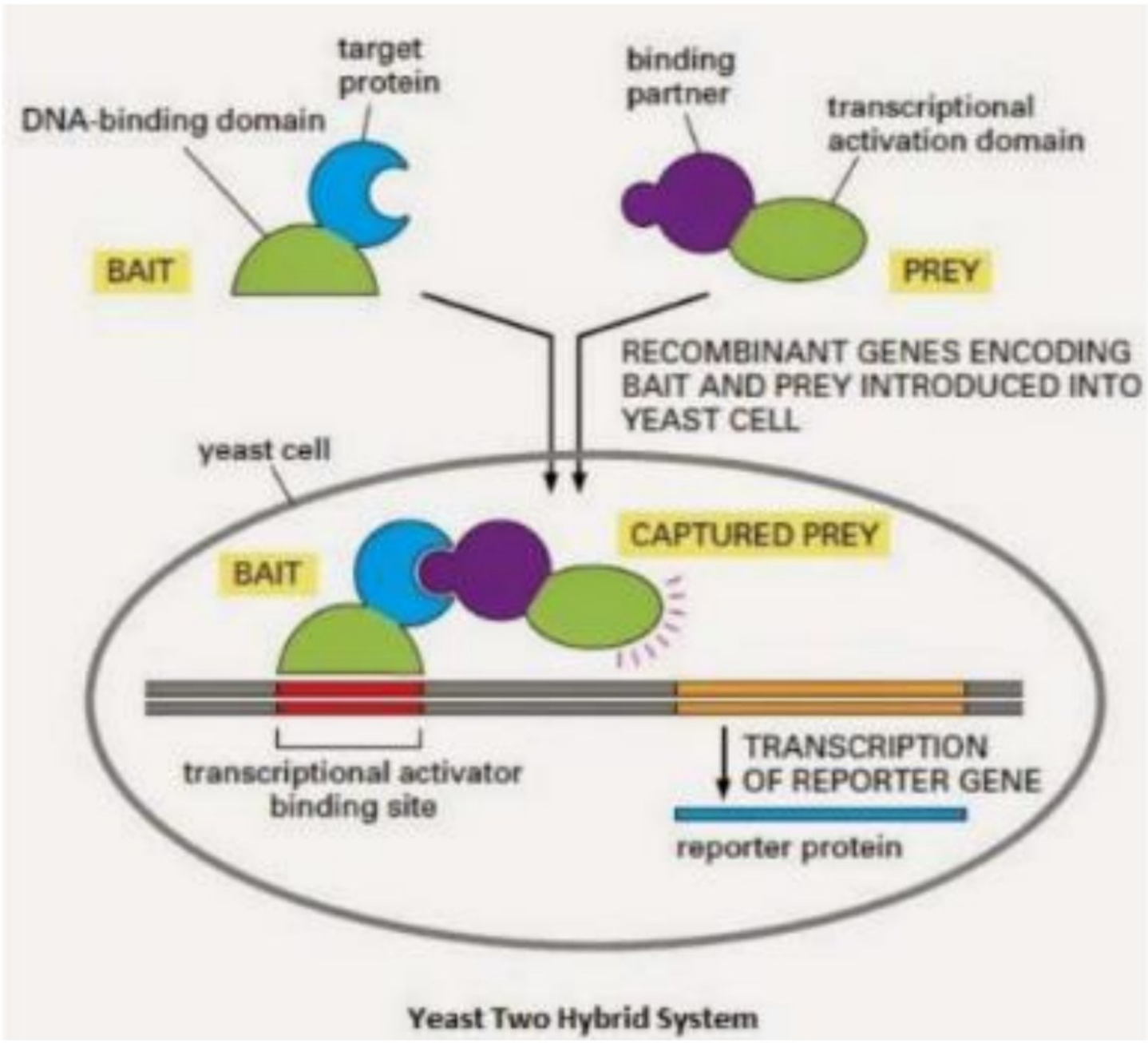
Yeast two-hybrid system

- **upstream activating sequence (UAS):** a regulatory sequence in yeast and it's the binding site of a transcription factor, known as **Gal4.** *When Gal4 binds UAS, gene expression is induced.*
- **Gal4 has 2 domains**: A DNA-binding domain (BD), that binds UAS and an activation domain (AD) that is responsible for the activation of transcription. **Both domains must be close to each other in order to transcribe the gene.**
- **LacZ gene** is a reporter gene that produces **β-galactosidase** that cleaves lactose.



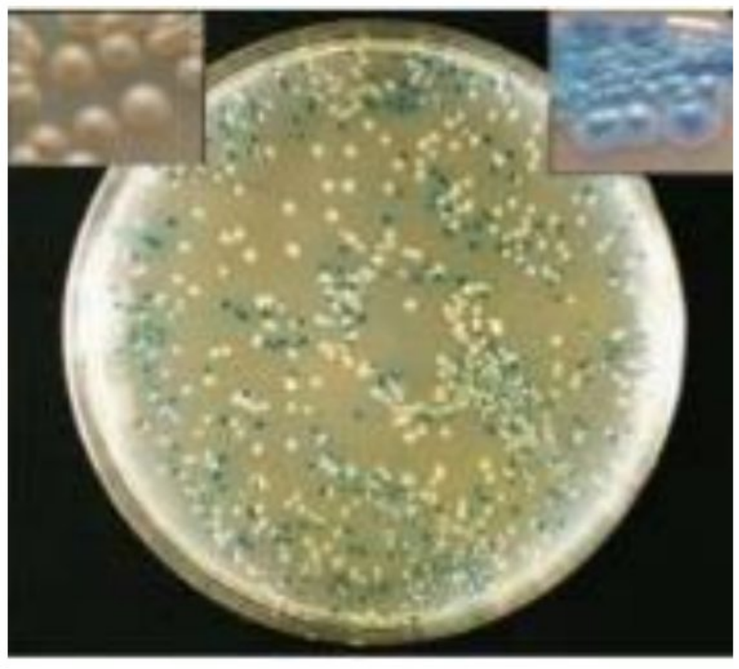
- We make 2 recombinant proteins in yeast cells:
  - The first one consists of two domains: **Gal4-BD** and **protein X**, which is the protein that we want to study.
  - The other protein also consists of two domains: **Gal4-AD** and **protein Y**, which is an unknown protein that we want to know if it interacts with protein X.
- **If protein X and protein Y interact with each other, Gal4-BD and Gal4-AD will be close to each other and the reporter gene (lacZ gene) will be expressed, and β-galactosidase will be produced.**
- We use a lactose analog, called **X-gal** **which generates a blue product when cleaved.**

- *<u>When the LacZ gene is activated, beta-galactosidase is produced, which cleaves X-gal generating blue colonies.</u>*



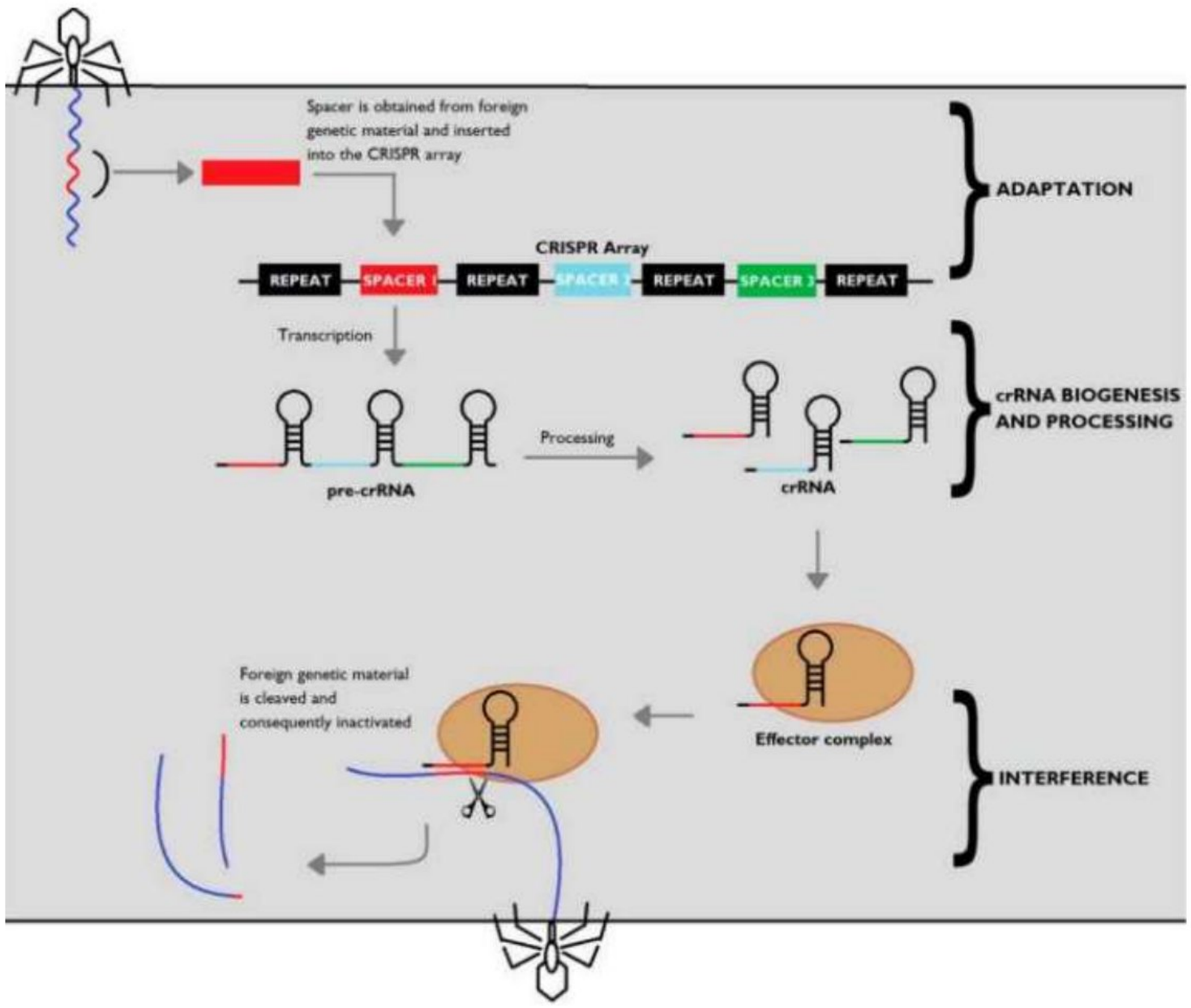
Yeast Two Hybrid System

- Blue yeast colonies are picked and plasmids are isolated to identify the unknown genes/proteins that interact with the known gene/protein.



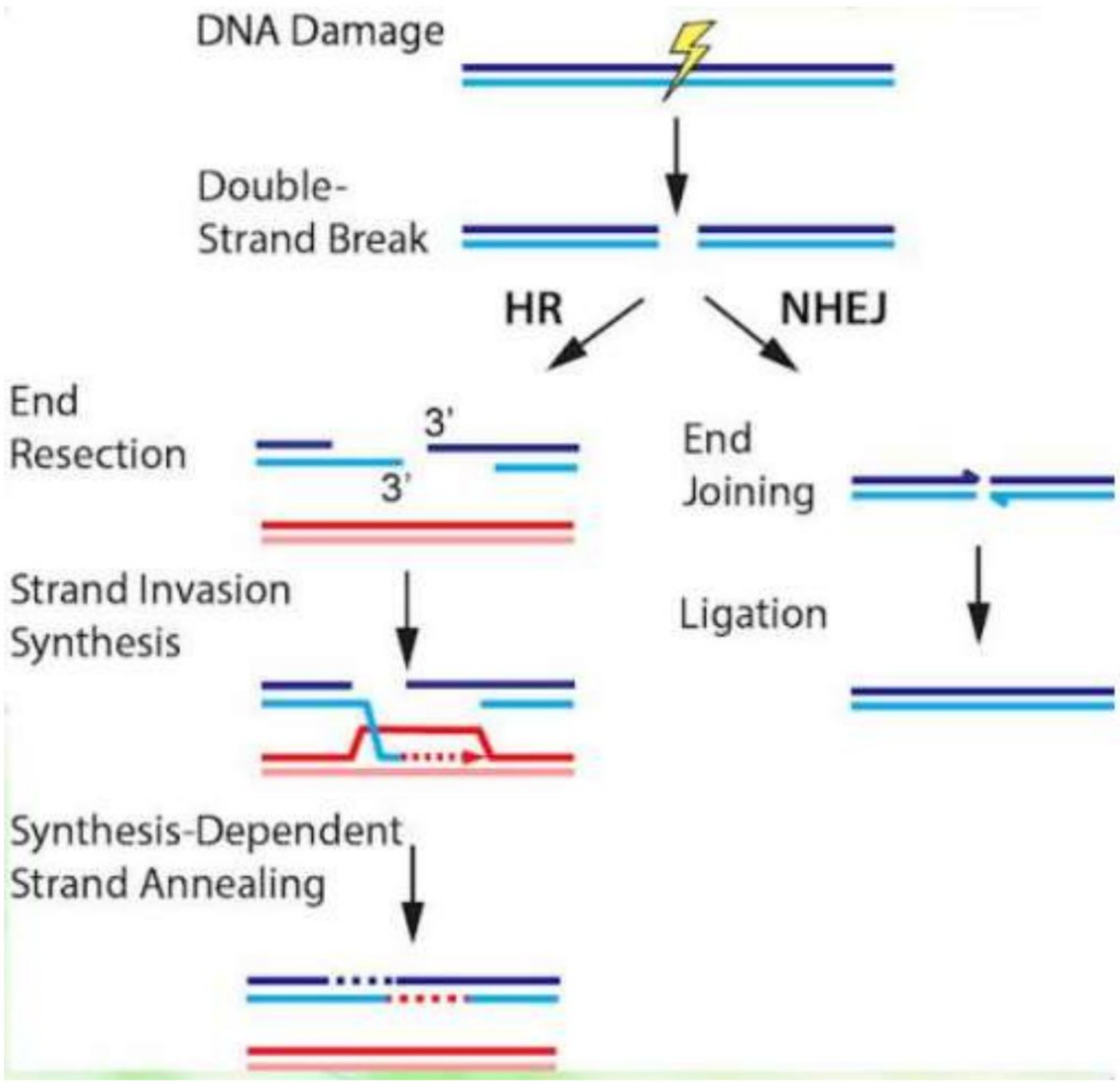
CRISPR / Cas9

- **CRISPR** stands for **c**lustered **r**egularly **i**nterspaced **s**hort **p**alindromic **r**epeats.
- It is a bacterial genetic system that constitutes **the immune system of bacteria** against phages.
- **Cas9** is an **RNA-guided nuclease** (**ribonuclease**) that can either **create DNA breaks**.
- The nuclease is directed to its target sequence by a short RNA fragment known as a **guide RNA (gRNA)**, which is complementary to the target segment of the genome.



- When a bacteriophage infects a bacterial cell. The cell **chops off** the phage DNA into smaller fragments and **integrates** a fragment into the CRISPR cluster as **a spacer**.
- When the phage infects the cell again, the cell transcribes the DNA into **pre-crispr RNA (pre-crRNA)** which will be cleaved into RNA molecules known as **crispr RNA (crRNA)**.
- crRNA is integrated into the Cas9 nuclease (as **guide RNA or gRNA**) and guides it to the phage DNA to degrade it.
- **NOTE**: crRNA consists of two parts: the transcribed part of **bateriophage's DNA** (the spacer) and the transcribed part of **bacterial DNA** – this part forms a loop because the bacterial DNA sequence is palindromic >> it can form hydrogen bonds –.

DNA Repair



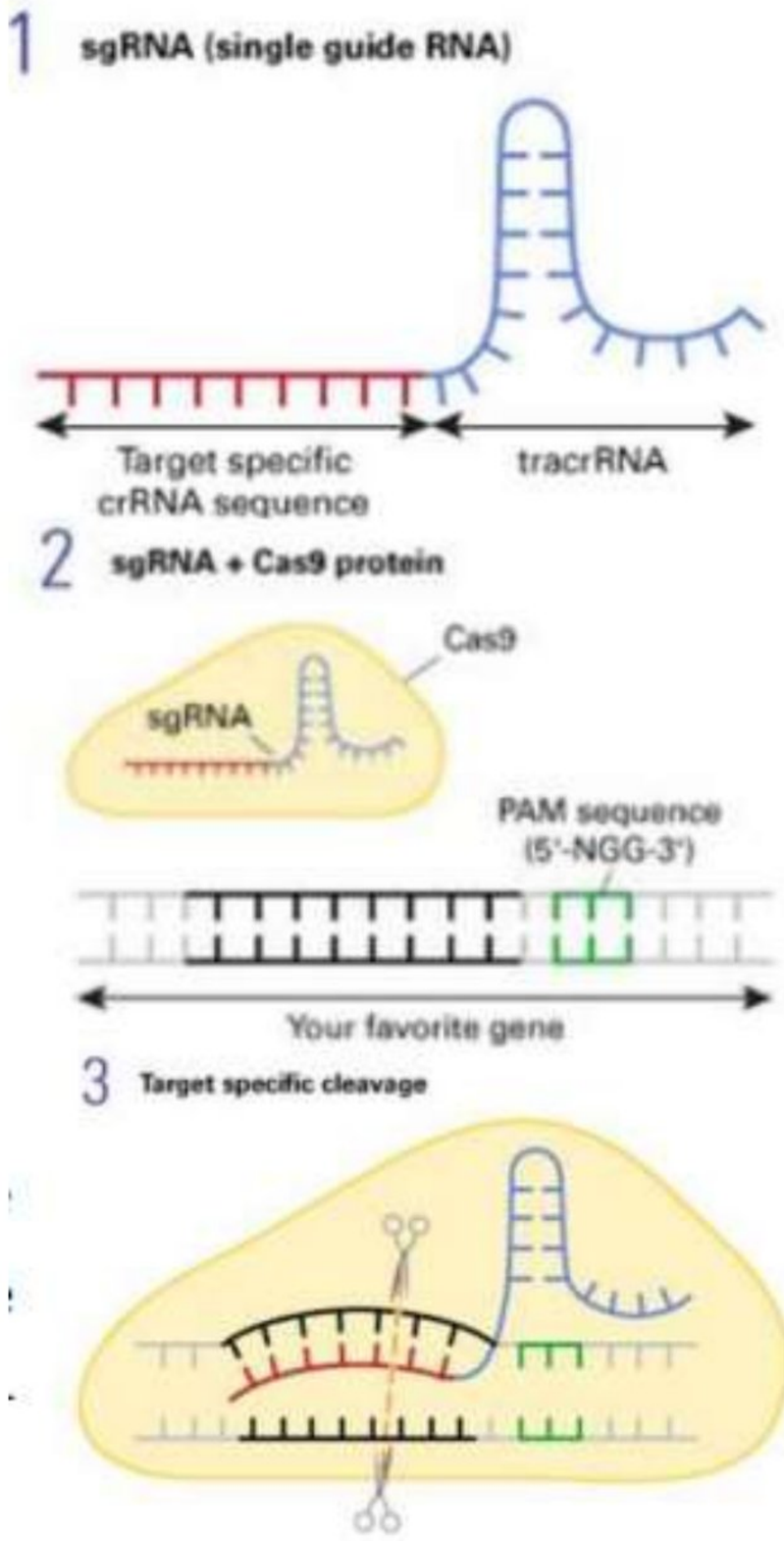
Non-homology end joining repair	Homology-directed repair
doesn't depend on the presence of a homologous chromosome	take advantage of the presence of **homologous DNA**
Ligation process: The two ends are glued together	**Exchange process**: cells take a part of one strand of the undamaged chromosome and put it in the other chromosome, filling the gap
It results in **indels** (insertion–deletion mutations), resulting in knocking out a gene.	It is like the replacement of a certain piece in the chromosome, meaning that it can correct mutations and introduce mutations. It can be also used to insert a fragment like a gene

Gene editing

- **To replace a gene:**

- We insert Cas9 gene into a plasmid along with an RNA molecule >> we introduce this plasmid into cells >> **cells produce Cas9 protein that is attached with RNA**.
- **Cas9 is directed by gRNA** into a region in the chromosome that is complementary to this RNA, and **it makes a single cut, causing DNA damage**.

  NOTE: Along with Cas9, we introduce a DNA piece that we want cells to use instead of the homologous chromosome in DNA homology-directed repair.
- **Homology-directed repair** mechanism is activated, and the system uses the fragment that we introduce not the homologous chromosome.
- We can fix genes or even mutate them to study them.



- **To guide proteins:**

- We use **dead Cas9** (A mutated Cas9 that can't cut DNA)
- We add the protein of interest to dCas9, so it will be guided to a specific DNA region.
- **Examples of proteins added to Cas9**:
  - **Transcriptional regulatory factors, like activators and repressors** to activate or inhibit certain DNA sequences.
  - **Green fluorescent protein (GFP)** to visualize genes.

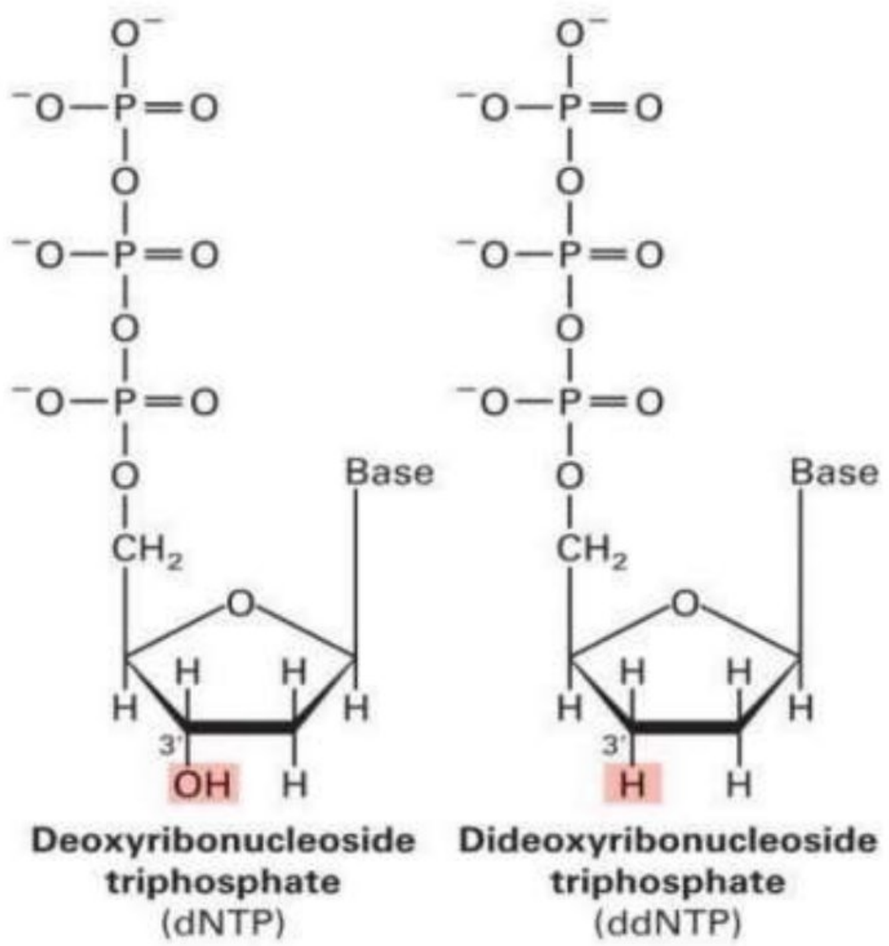
- **Other Cas9 enzymes:**
  - **Cas12a**: A smaller enzyme that introduces **staggered cuts**, rather than blunt cuts.
  - **Cas13a**: **An RNA endonuclease.**
  - **Cas9** makes **blunt ends.**
- **CRISPR-Cas9** will be used to fix DNA, treat diseases and modify human embryos.
- It can be used in **bioterrorism** and to do unethical things, like modifying human genome or making existing bacteria and viruses more dangerous.

DNA sequencing

- **DNA sequencing**: the process of determining the exact order of nucleotides in a genome or a DNA fragment.
- **The importance of DNA sequencing:**
  - ➢ Identification of genes and their localization**.**
  - ➢ Identification of protein structure and function.
  - ➢ Identification of DNA mutations.
  - ➢ Genetic variations among individuals in health and disease.
  - ➢ Prediction of disease-susceptibility and treatment efficiency.
  - ➢ Evolutionary conservation among organisms (Note: evolution is mythical)

- **Deoxynucleotides vs dideoxynucleotides**

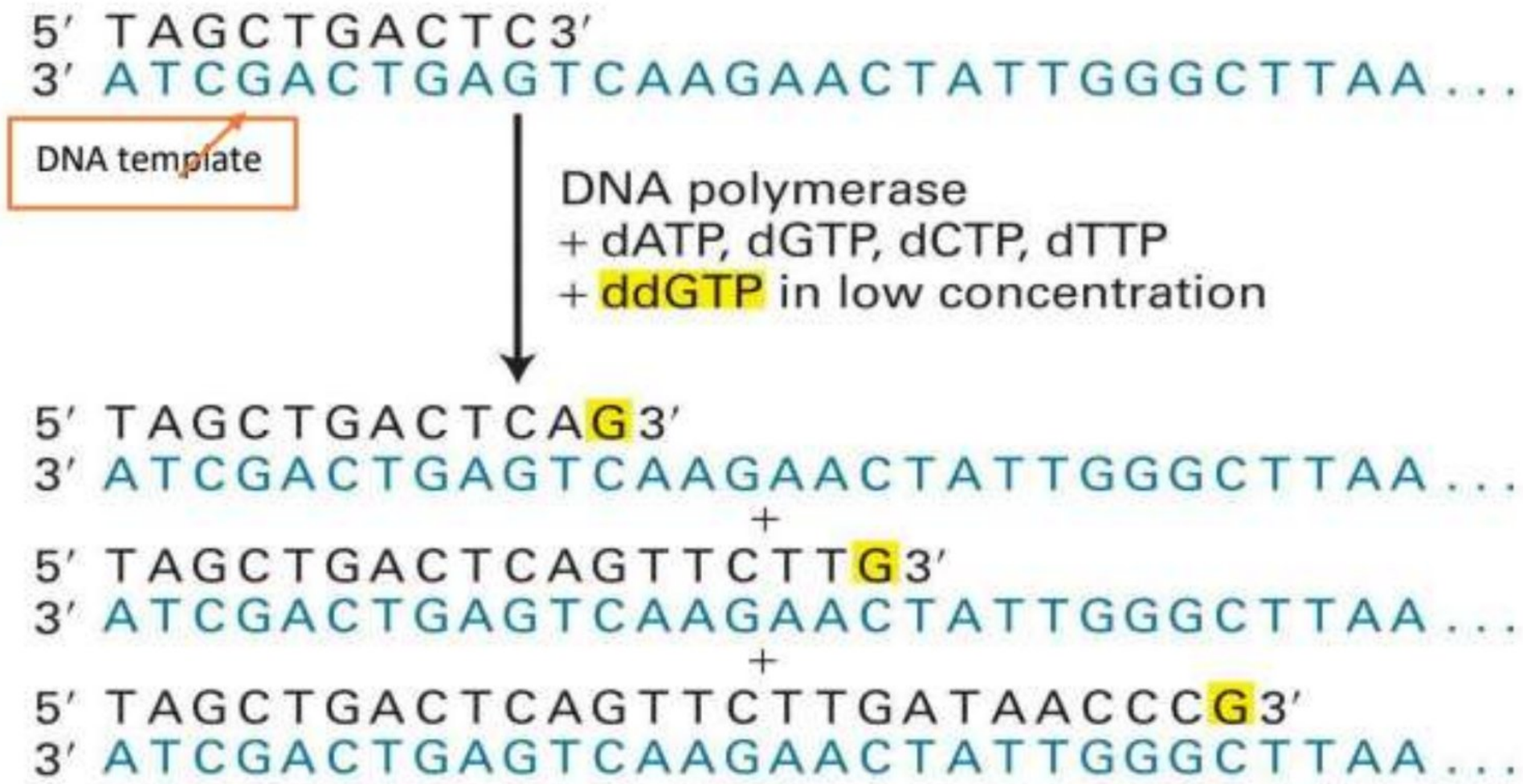
- **Deoxyribonucleoside triphosphate (dNTP)** is the normal substrate of DNA polymerase; it has a deoxygenated carbon (no.2), and the OH on carbon no.3 is the site where addition of another nucleotide occurs.
- **Dideoxyribonucleoside triphosphate (ddNTP)** has two deoxygenated groups on carbon no.2 and no.3; the abcence of OH on carbon no.3 prevents the addition of another nucleotide.

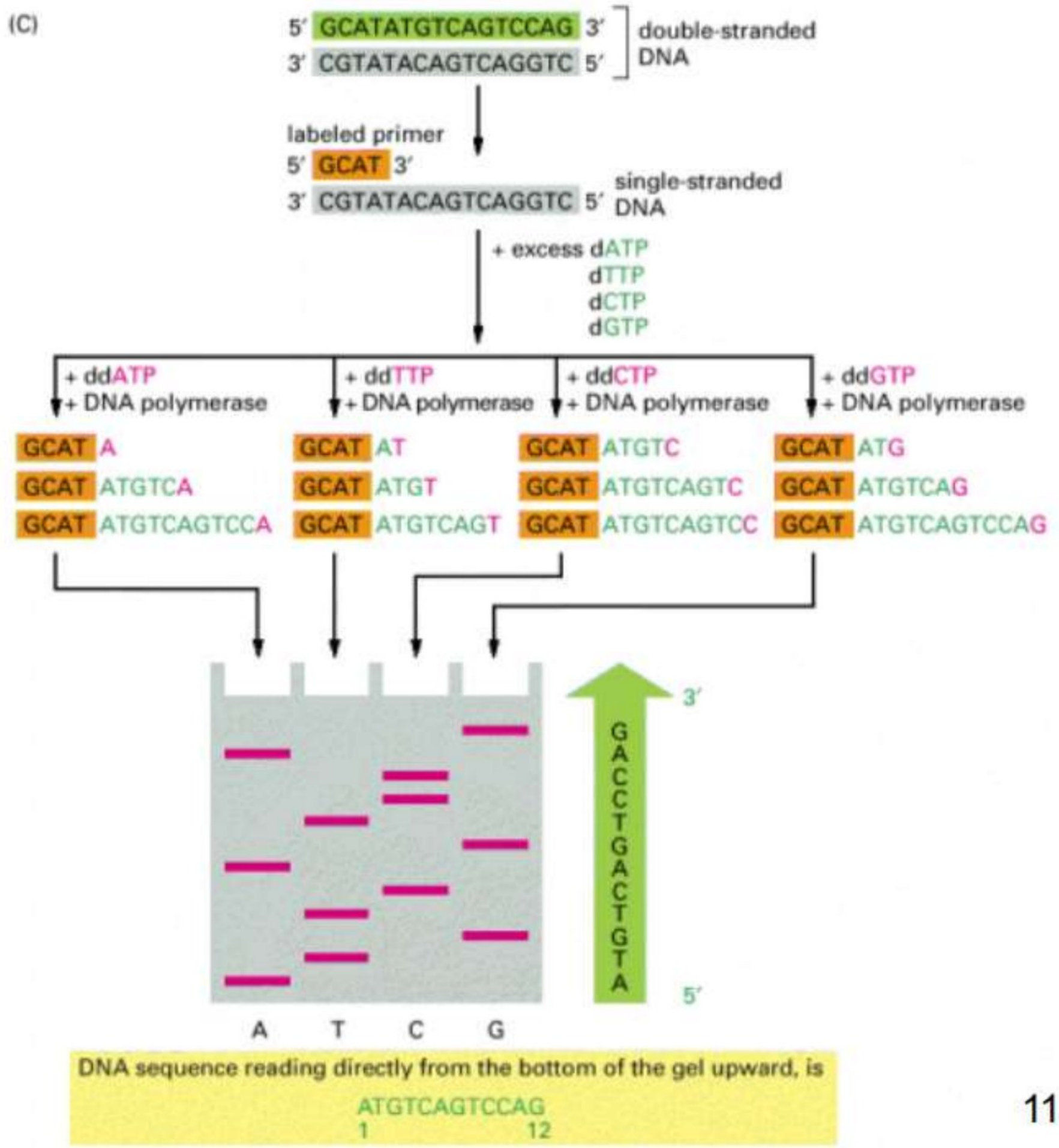


- **DNA sequencing method**

- DNA synthesis is initiated from **a radioactivity-labeled primer**.
- Four separate reactions are run, each including **the four deoxynucleotides** plus **one dideoxynucleotide (either A, C, G, or T).**
- Incorporation of a dideoxynucleotide **stops further DNA synthesis** >> a series of labeled DNA molecules are generated, each terminated by the dideoxynucleotide in each reaction.
- These DNA fragments are separated **according to size** by **gel electrophoresis** and detected by **exposure to X-ray film**.

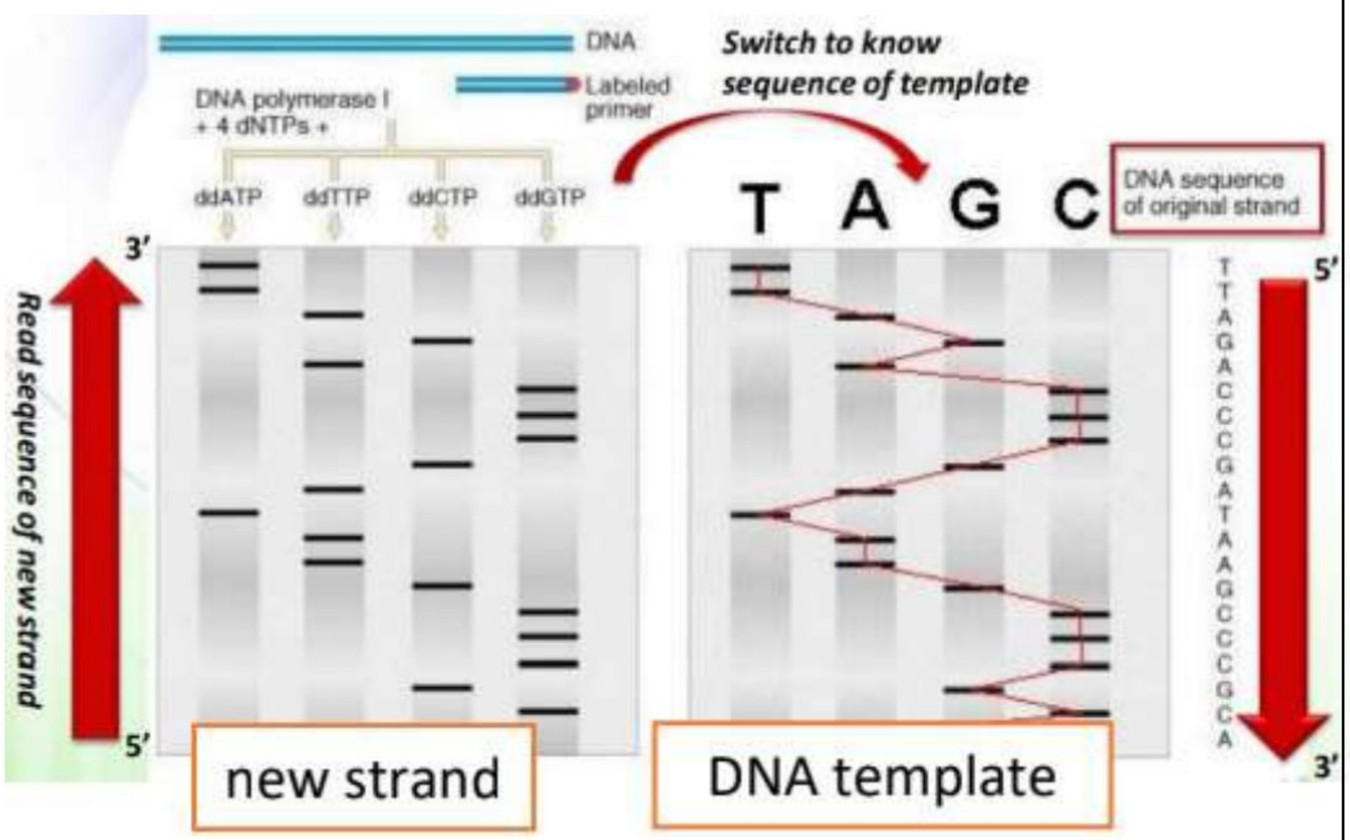
- So, we need DNA template, DNA polymerase, labeled primer, the 4 deoxynucleotides and 1 dideoxynucleotide in each tube.

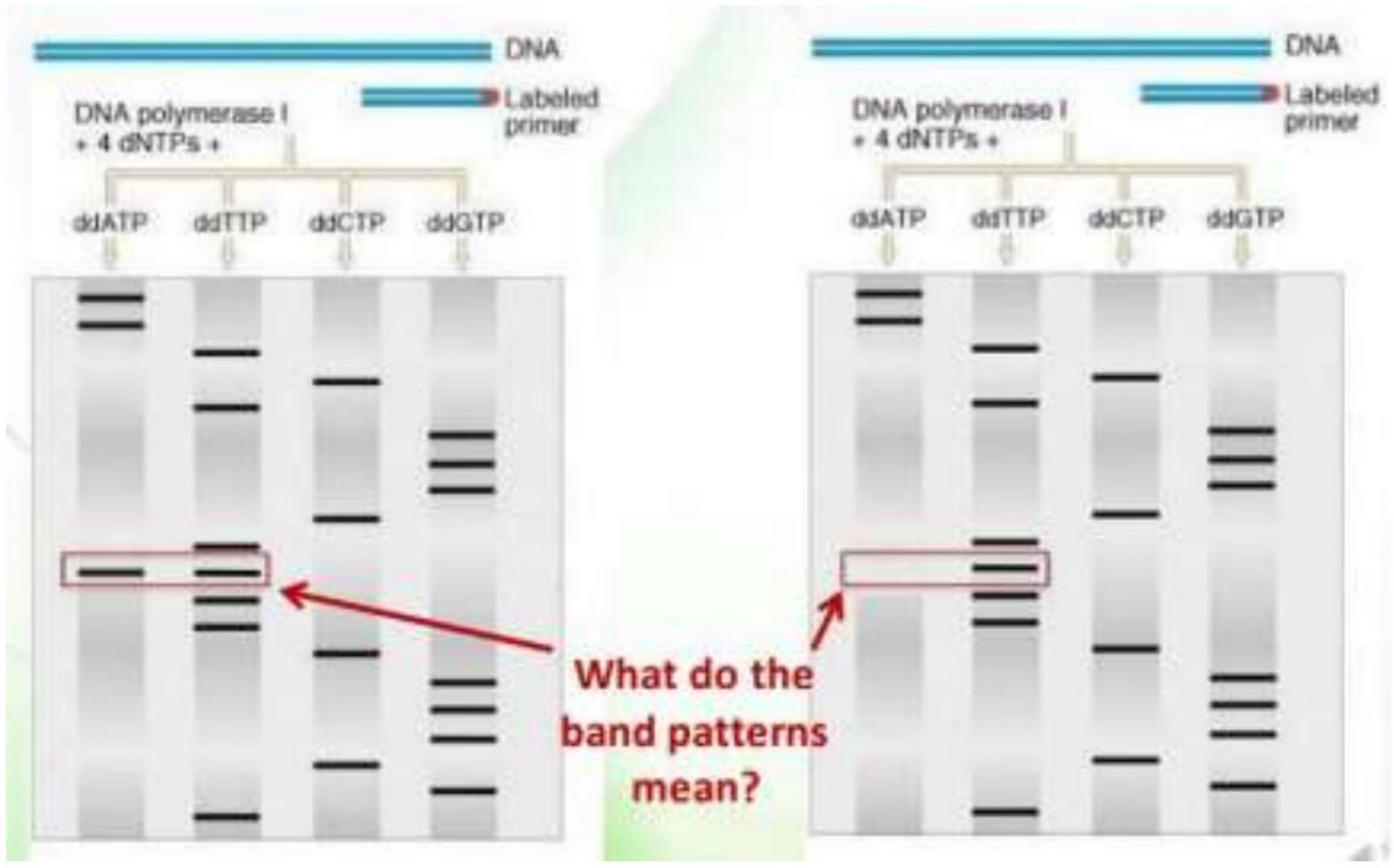




11

- **We read the new strand from bottom to top (5’ to 3’)**
- **To read the template strand, we have 2 ways**:
  - Write the new strand, and write the complementary strand.
  - Switch as the picture:

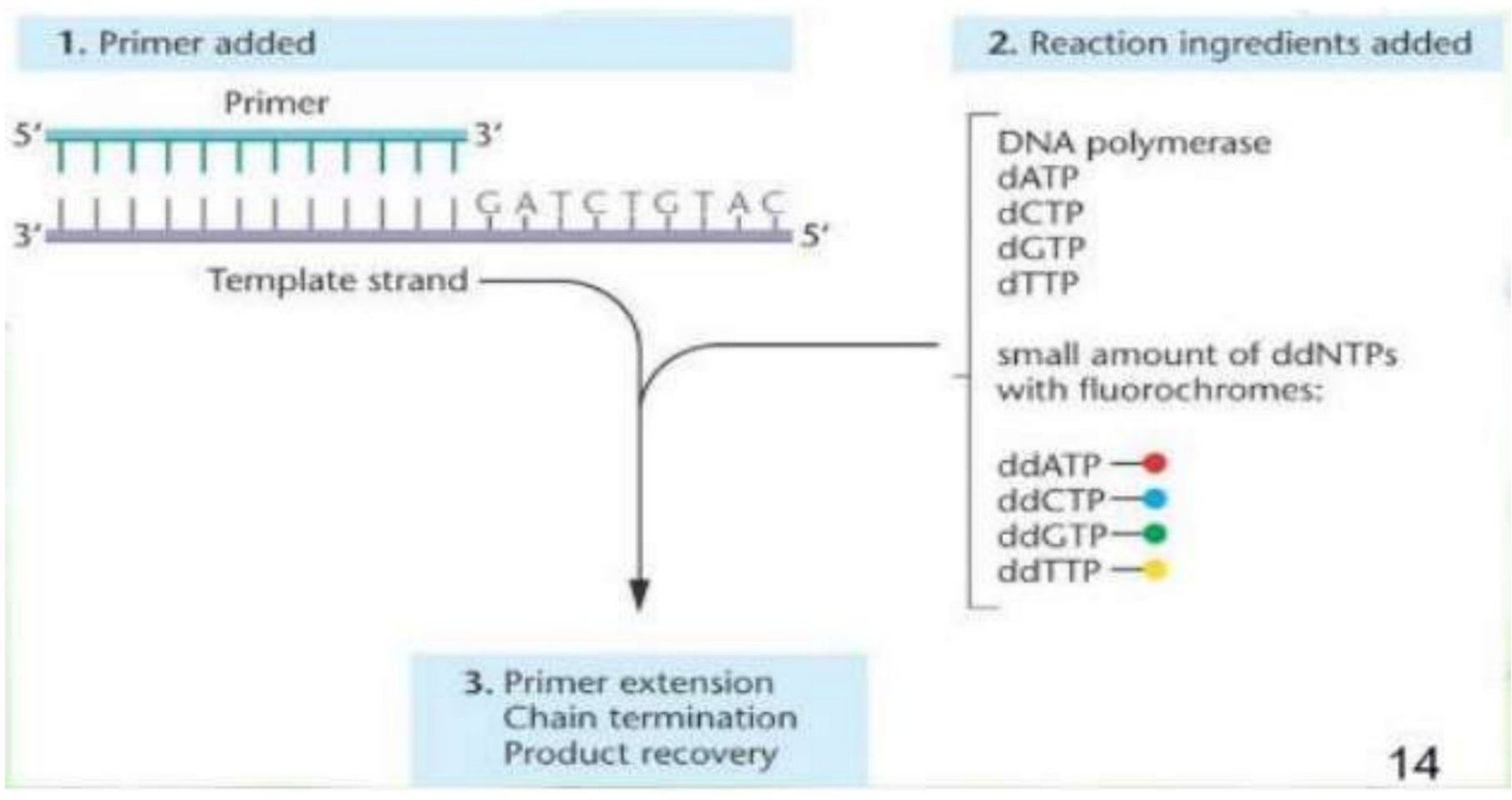




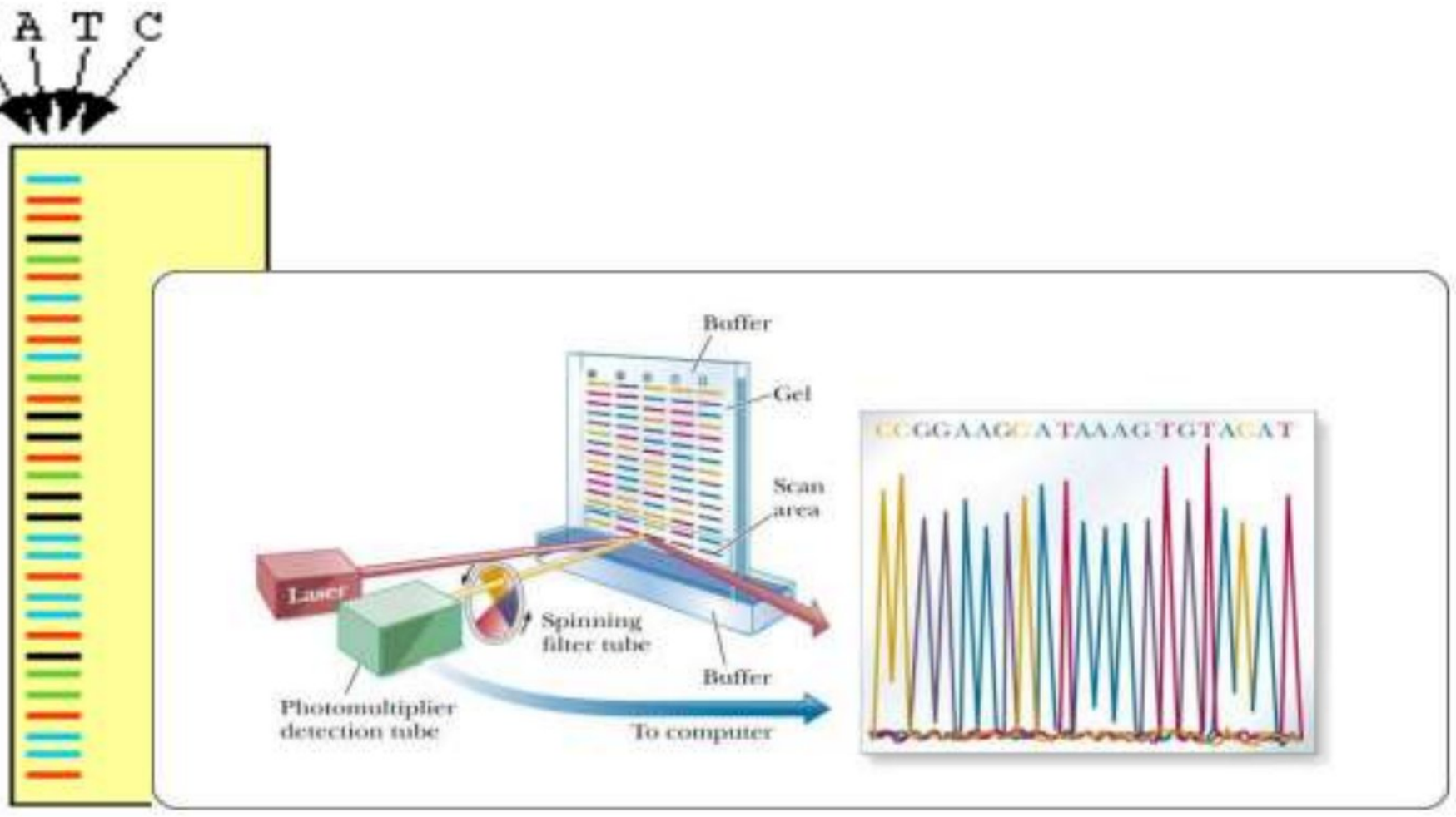
- If there are 2 bands at the same position, it means that the person is **heterozygous** at this nucleotide, it can be **a mutation or polymorphism like SNPs**.
  - **Mutation**, if it's present in less than 1% of population.
  - **Polymorphism**, if it's present in more than 1% of population.
- If the band is not in its normal position, it means that there is a **mutation on both chromosomes or normal variation on both chromosomes** (**homozygous**).

Fluorescence-based DNA sequencing

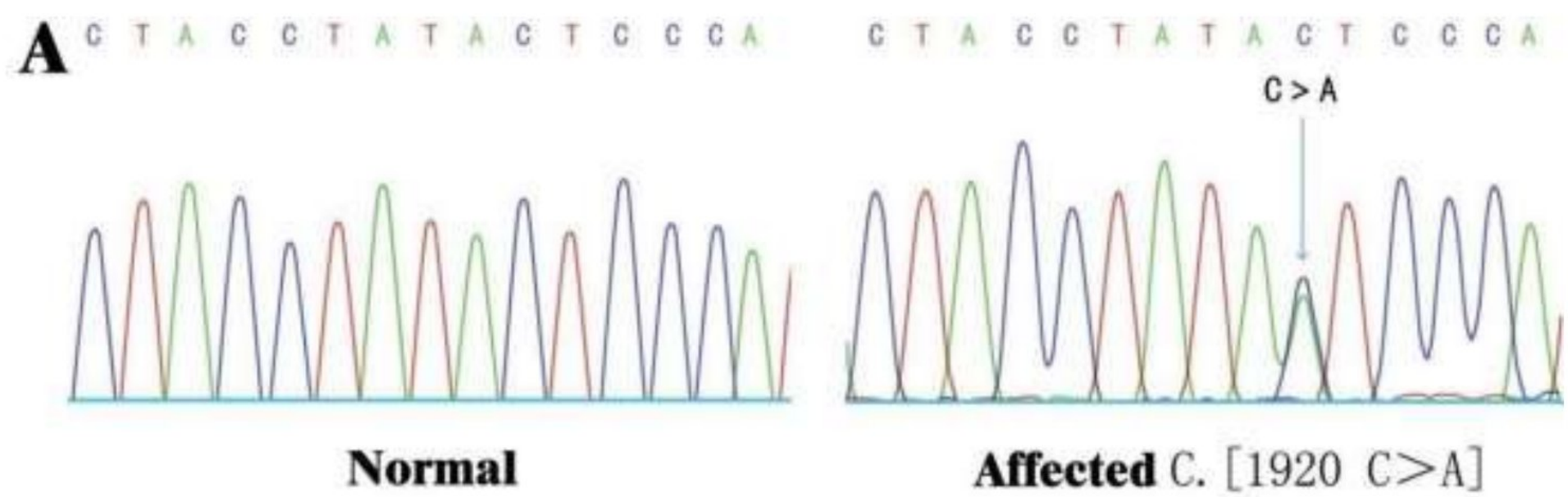
- Reactions include the four deoxynucleotides plus the four dideoxynucleotides in the same reaction with each ddNTP labeled with a unique fluorescent tag.



- the gel would be read by computers and the dideoxynucleotides would be labeled by fluorescence, instead of labeling primers, and each nucleotide would give a certain color throughout its own fluorescent signal.



- if there's a variation or a mutation, it will look like this: two peaks overlap (Heterozygosity).



- If there is a mutation on both alleles, we would still have a single peak but representing a different signal than what we would normally exist.

Next Generation sequencing

- The entire human genome could be sequenced in this method within 24 hours.

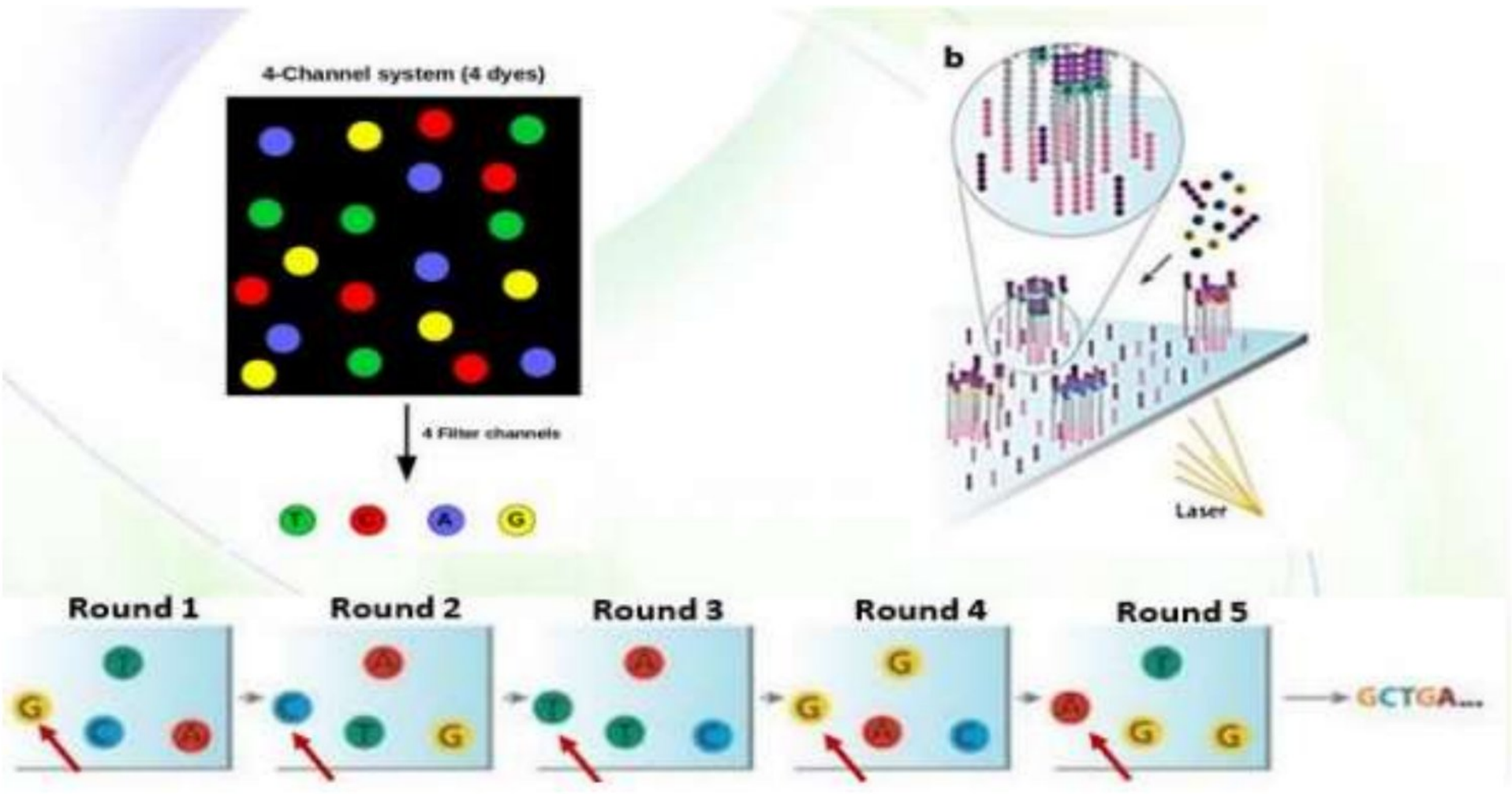
- **Process**:

- Cellular DNA is **fragmented randomly**.
- **Identical DNA adapters** are added enzymatically at the ends of each DNA fragment.
- Each DNA fragment is attached to a solid surface and **amplified** like PCR using primers that anneal to the adapter sequences.
- **Four-color nucleotides** with terminating ends are added

  NOTE: These nucleotides are special, as the nucleotide should be activated -chemically modified- and fluoresce (each nucleotide fluoresces with a different color) to allow the binding of the next nucleotide.
- The signal (fluorescence) is detected by a special camera and the cycle is repeated.

➢ Adapters importance:

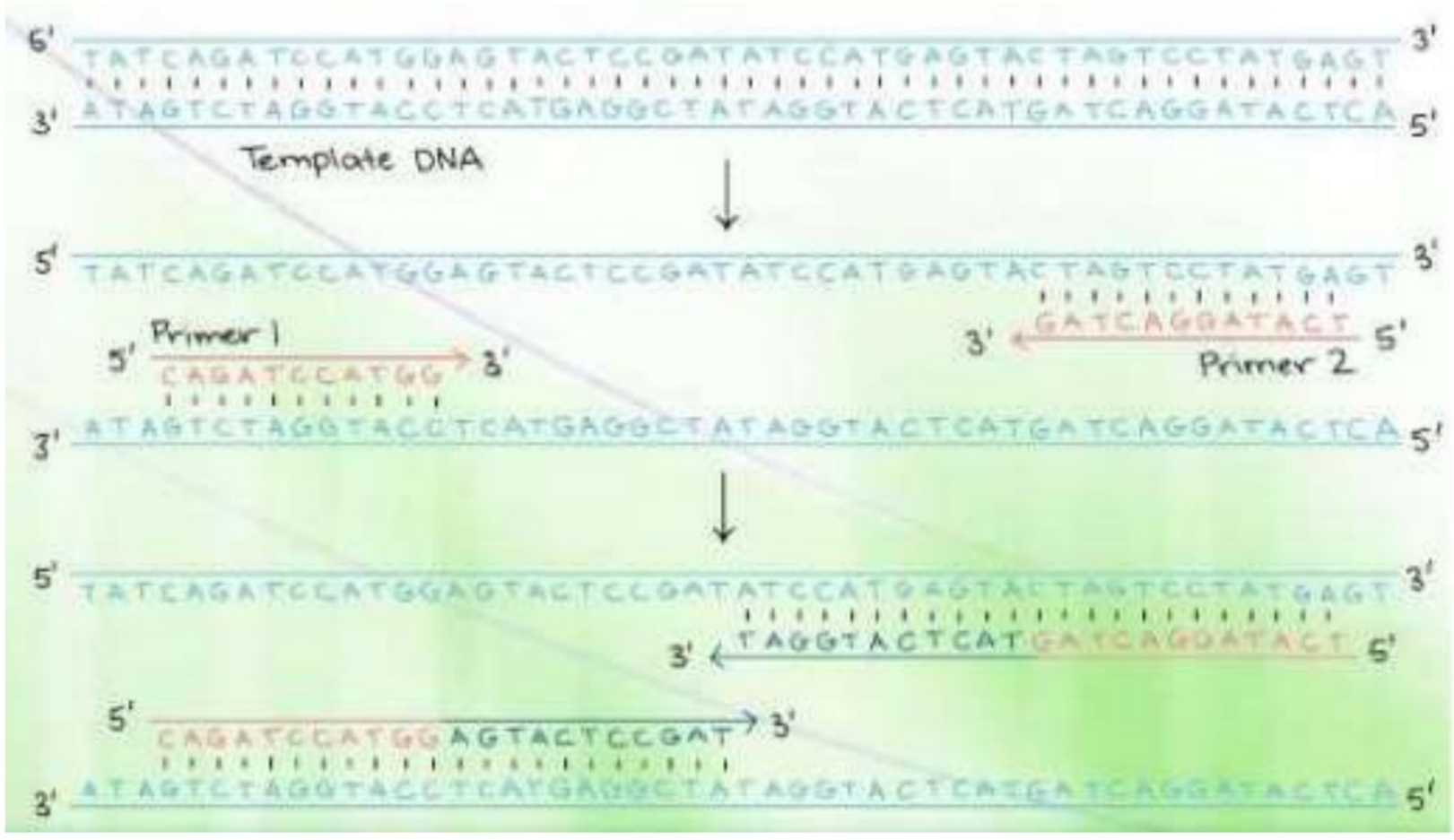
1. Attach DNA fragments to the surface.
2. Binding sites for primers.
3. adapters can be used as tags to differentiate in case we are sequencing 2 different samples.



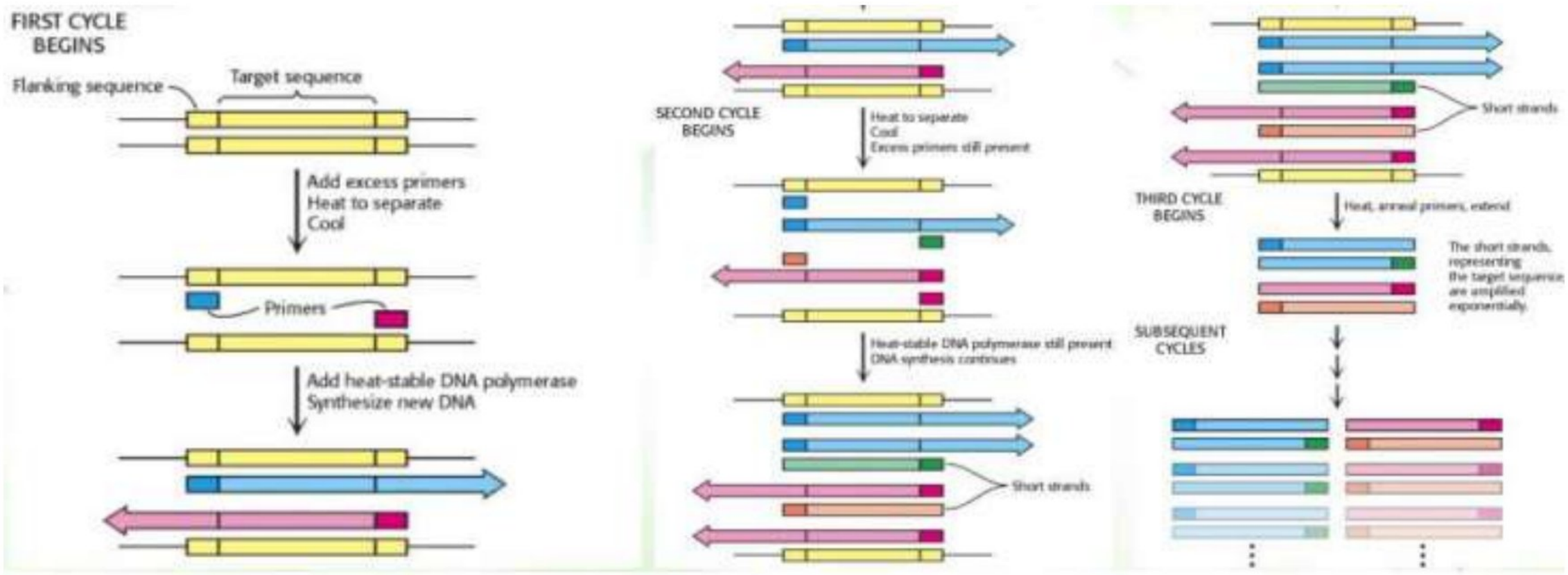
PCR-Polymerase Chain Reaction

- **Challenges we face when dealing with DNA:**
  - ➢ Genetic variation.
  - ➢ Minute (small) amount of genetic material.
  - ➢ Identification of organisms.
- **PCR** allows to **amplify** the DNA from a selected region of a genome a billionfold and it is extremely sensitive, as it can detect a single DNA molecule in a sample.
- **PCR** is a consecutive enzymatic reaction that consists of **repeated cycles (20-30 cycles)**
- **Each cycle consists of 3 phases:**

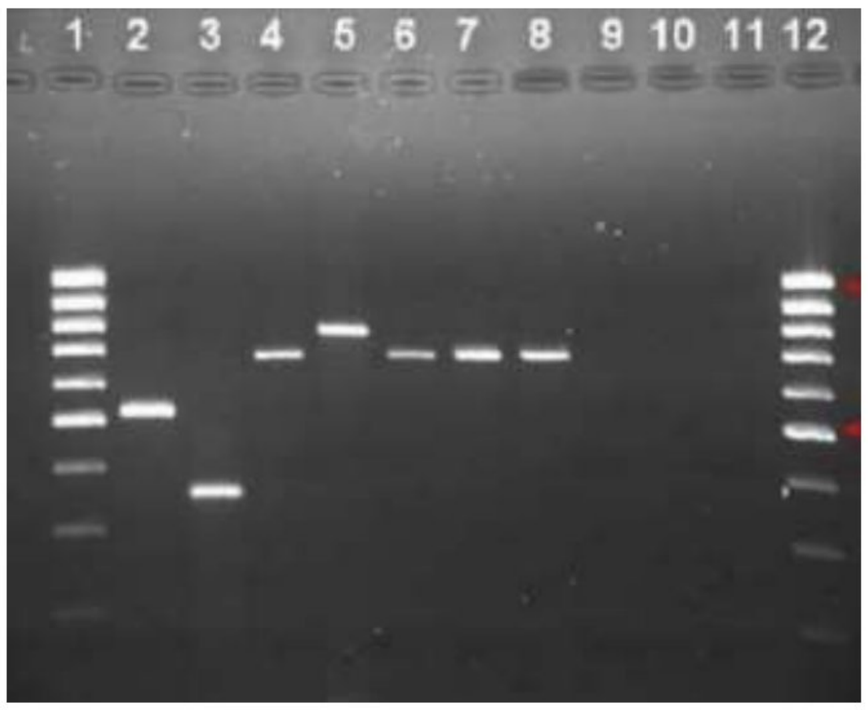
1. **Denaturation (at 95°C)**: The two strands are separated from each other.
2. **Annealing (50°C to 70°C**): the primers anneal (hybridize) to the DNA.
3. **Polymerization** or DNA synthesis **(at 72°C):** optimal for the polymerase.



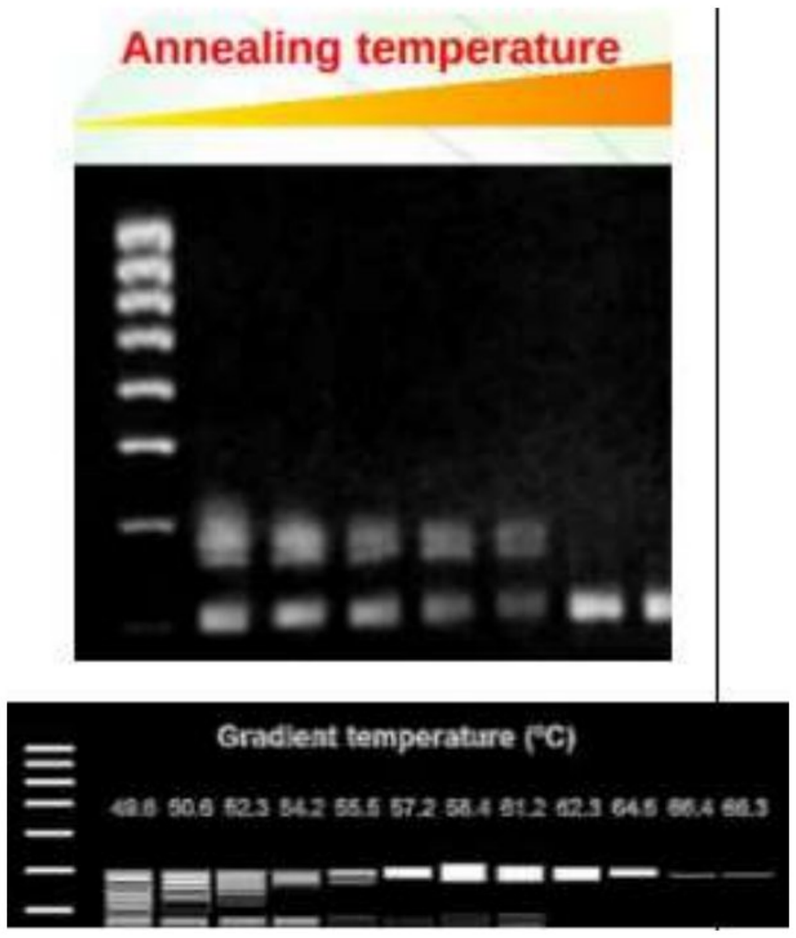
- We use **heat-stable** DNA polymerase, like **Taq polymerase**, which is extracted from thermostable bacteria that live in **hot springs**.



- The target product starts to appear at cycle 3.
- Every cycle **doubles** the amount of DNA.
- the products of each cycle serving as the **DNA templates** for the next-hence the term polymerase "*chain reaction*".
- At the end, we take our sample to analyze it using **gel electrophoresis**.



- This **specificity of amplification** depends on the **specificity of the primers** to not recognize and bind to sequences other than the intended target DNA sequences.
- *If you increase the temperature, more bonds will be broken and the DNA strands will be separated, and the DNA is denatured*.
- **melting point**: The temperature at which 50% of the DNA is double-stranded and 50% is single-stranded.
  - If the **annealing temperature is low**, this allows for non-specific hybridization of the primers, resulting in **non-specific amplification**.

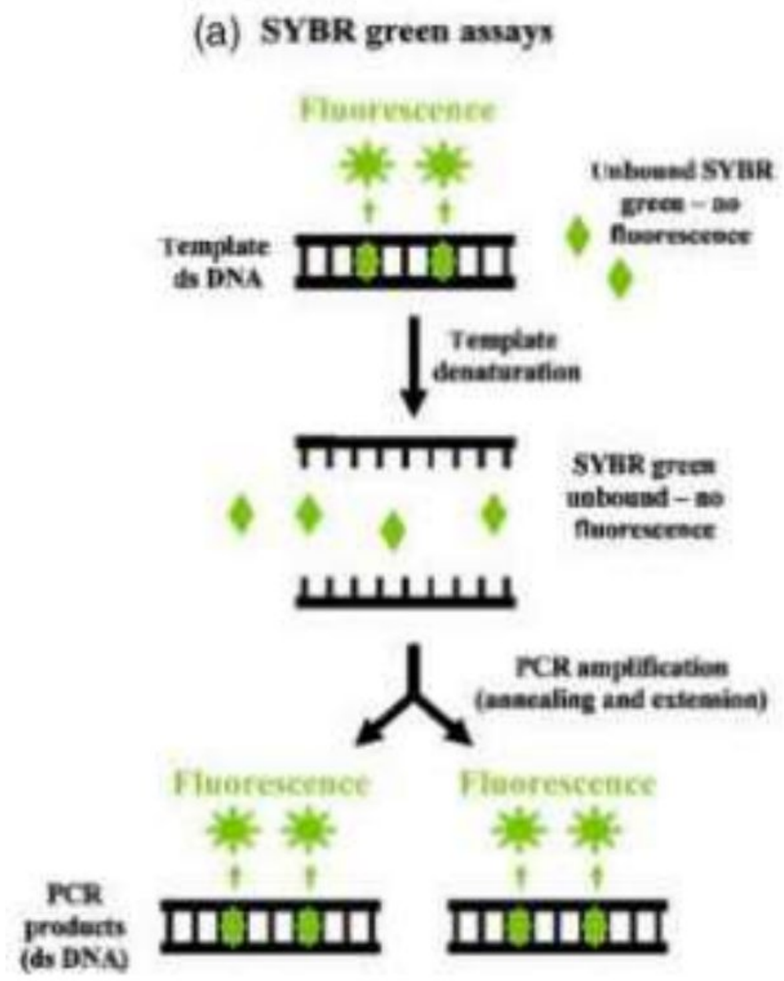


- **The higher the annealing temperature is, the more specific the annealing is and the more pure DNA is and it will appear as a single band**.

- **What if we raise the temp. more and more?** >> Nothing will happen, because primers wouldn't anneal with DNA.

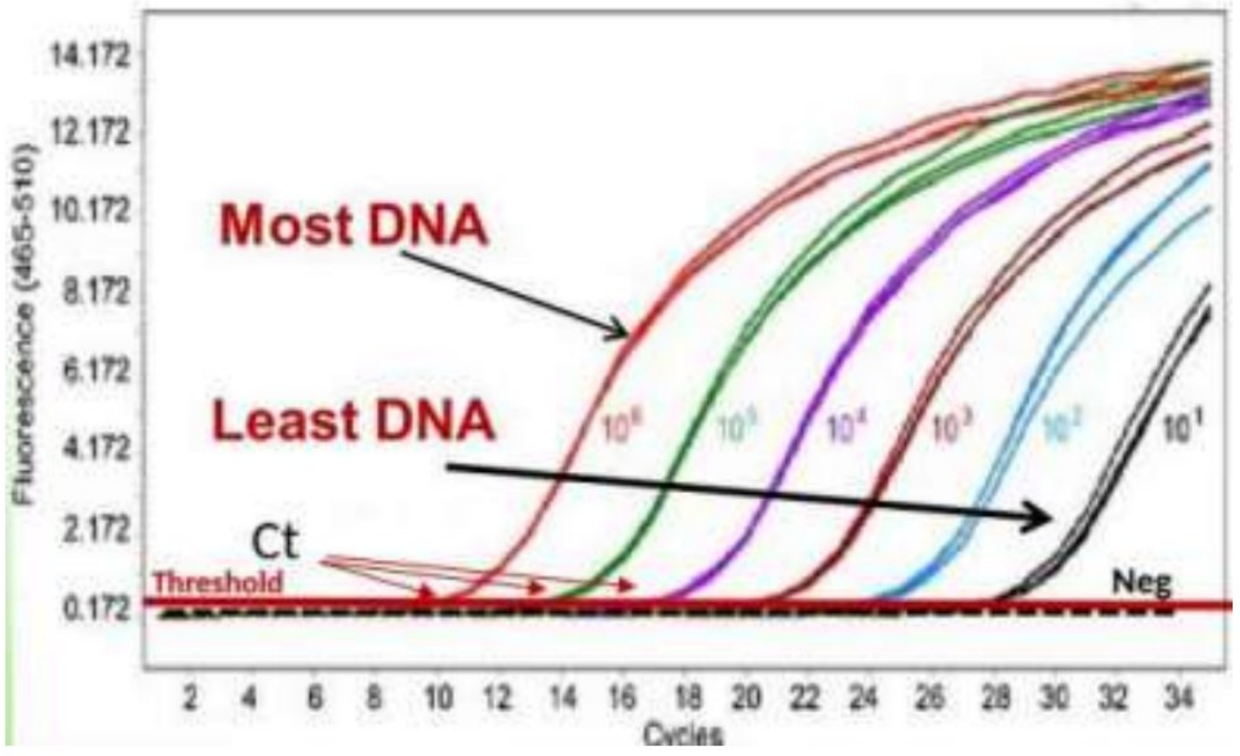
Real-time quantitative PCR (qPCR)

- **SYBR green binds to double-stranded DNA and fluoresces only when bound**, so *the more double-stranded DNA you have in a sample, the stronger the fluorescence is*.



- **If we have more DNA to start with, we will have the fluorescence early on, so the higher the amount of DNA, the sooner it is detected.**

- **Threshold cycle (Ct):** The cycle at which the signal is detected and is a measure of starting amount of DNA.



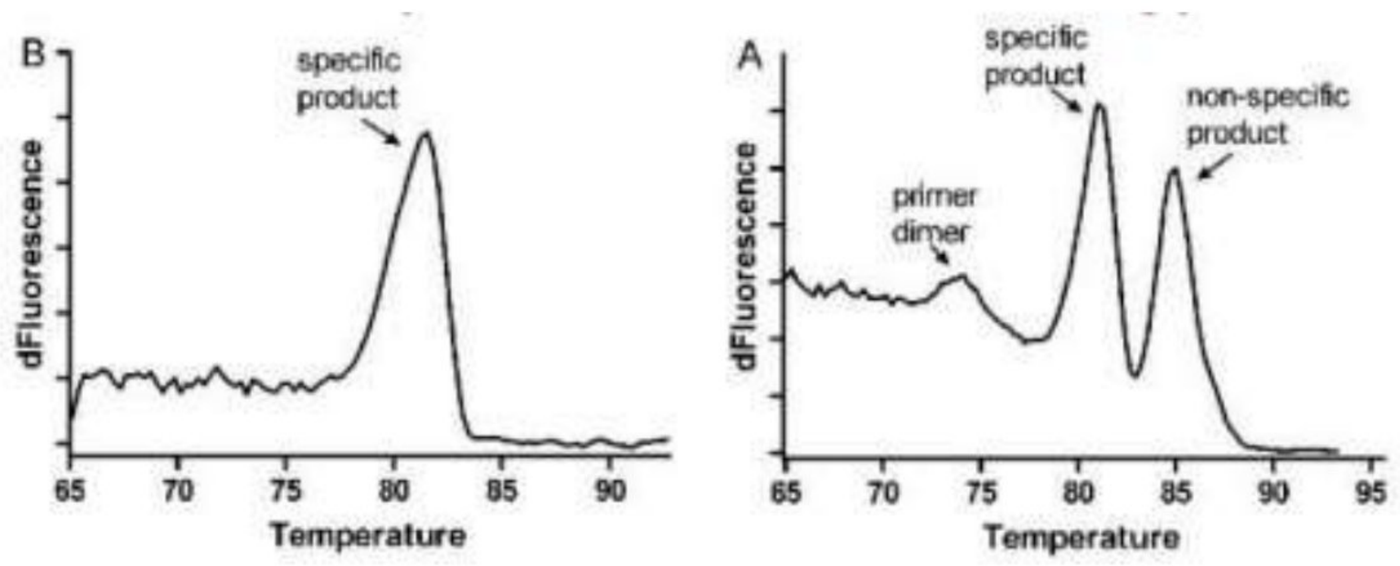
- **If the signal is detected early and the Ct is low, it means that the starting material amount is high.**

- **If the signal is detected lately and the Ct is high, it means that the starting material amount is low.**

Melting curve analysis of qPCR

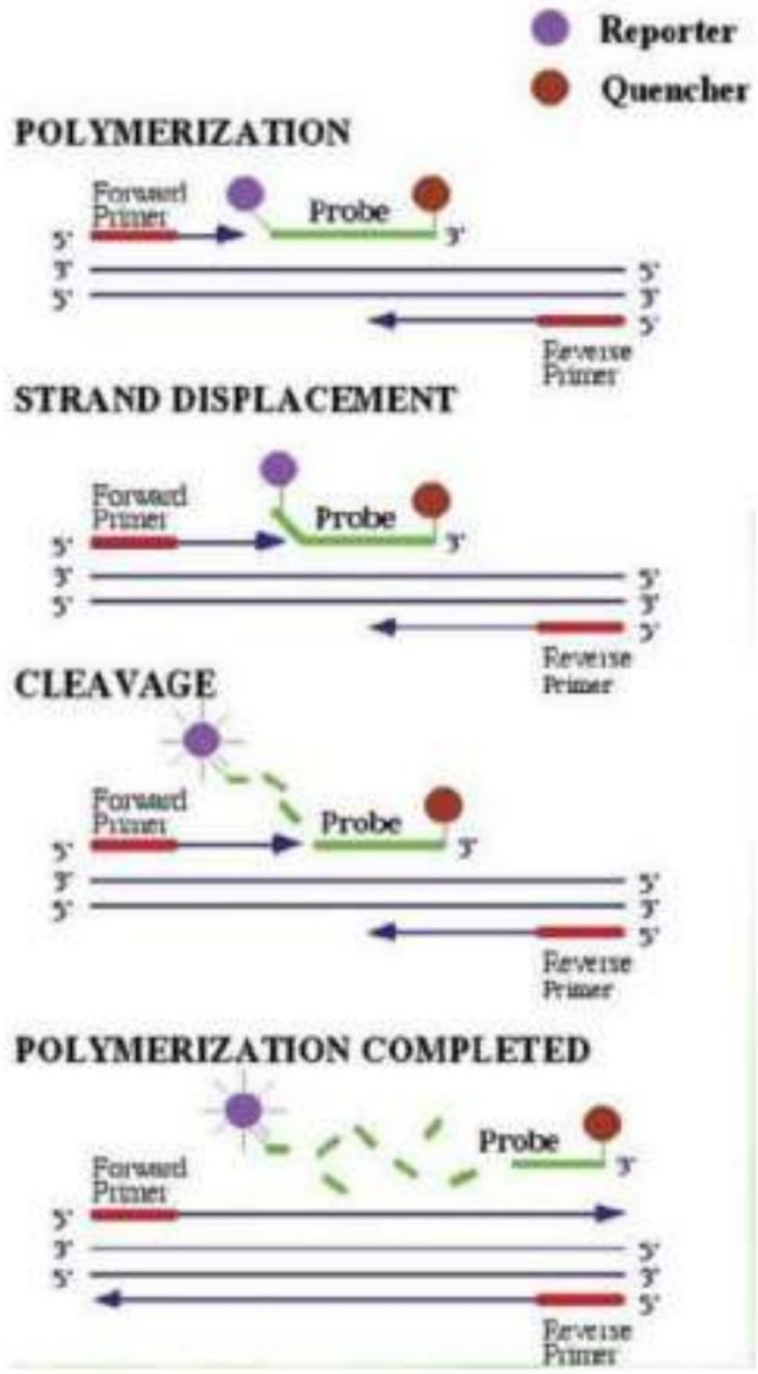
- **how can we be sure that we're amplifying our DNA, and there is no non-specific hybridization?** >> we can at the end do **gel electrophoresis** and see if we have a single band and where the bands are.
- But the instrument can do it by **melting curve analysis**:
- If we increase the temperature, the DNA denatures and becomes more single-stranded, so fluorescence decreases. It means that there is a change in fluorescence and **this change is mostly observed at melting point**.

- **If the amplification is not specific, there is 2 or more DNA molecules that are amplified and each molecule has its melting point, so there will be 2 or more temperatures at which change in fluorescence is mostly observed.**
- **If the amplification is specific, there will be a single temperature at which change in fluorescence is mostly observed, which is the melting point.**



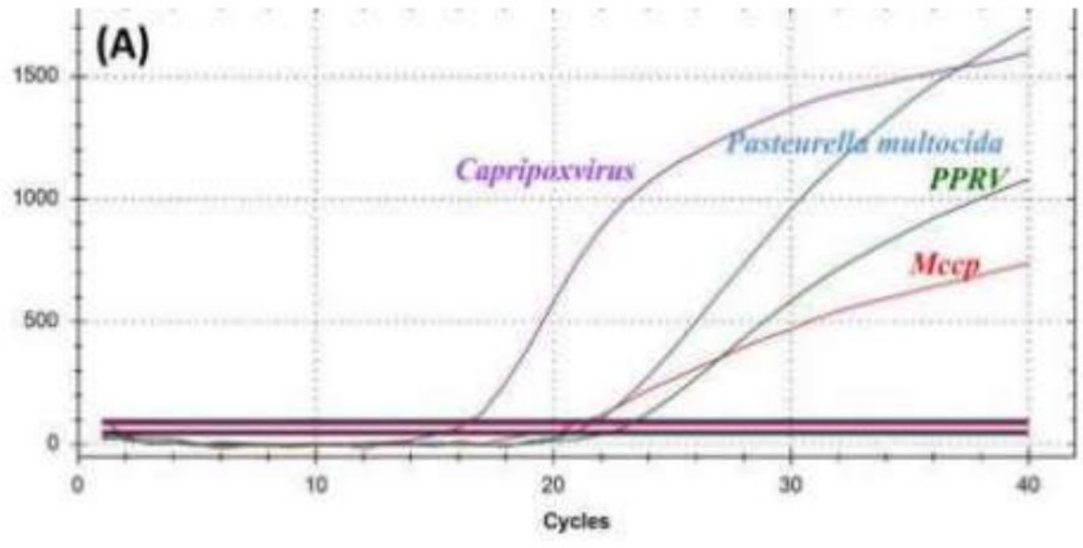
Taqman qPCR

- We use **probes** and each probe is attached to a **reporter** that fluoresces and a **quencher** that inhibits the reporter.
- If the quencher is close to the reporter, the reporter doesn't fluoresce, and if they get separated then the reporter can fluoresce.
- We use a special Taq DNA polymerase that has **5' to 3' exonuclease activity** along with its polymerase activity, so it synthesizes DNA and **once it reaches the probe, it activates the exonuclease activity and starts breaking the probe**, releasing the **reporter** and now it gets away from the **quencher** and fluoresces.



- **Advantages (vs SYBR green chemistry)**

- **More specific** - **More sensitive** - **More reproducible**
- **Multiplexing** (you can amplify different targets, you can use different sets of primers in the same exact reaction, so you can have a tube that contains the same template, Taq polymerase, same deoxyribonucleotides, but you also use different primers that can amplify different regions in the template, and use different probes that would bind to these amplified regions, and each probe would have a different reporter emitting different signals and you can quantify them because signals are emitted at different threshold cycles).

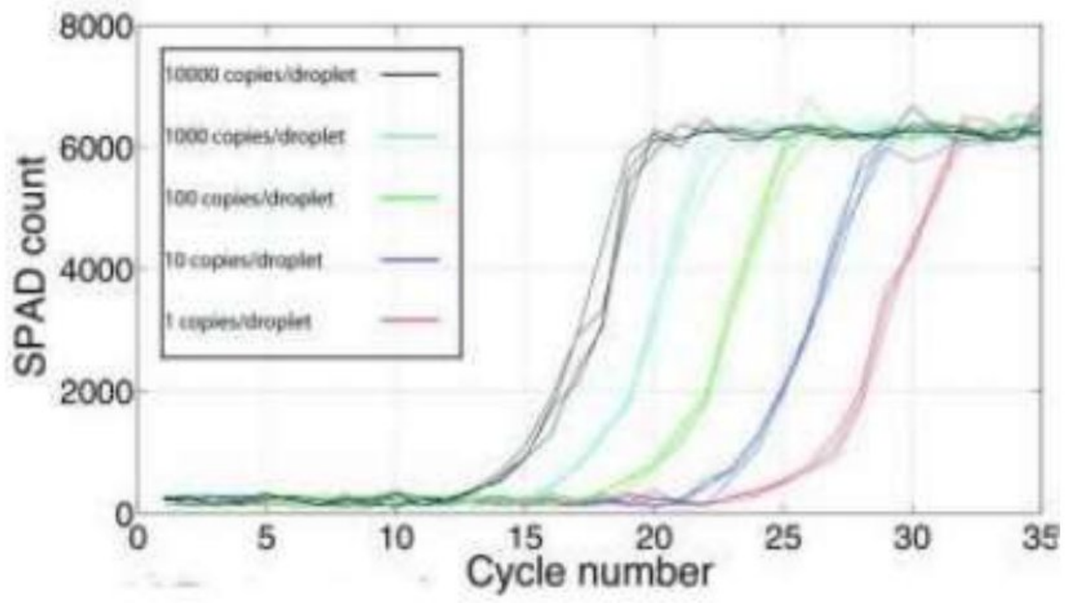


Analysis of gene expression and RNA levels

- How to make cDNA from mRNA?
  - A poly-T primer complementary to poly-A tail is added to mRNA and then **reverse transcriptase (RT)** binds to it.
  - **Reverse transcriptase** makes a complementary strand to mRNA using mRNA as a template, then it degrades the mRNA molecule and synthesizes the second strand, now we have double-stranded DNA (cDNA).

Real-time qPCR of mRNA

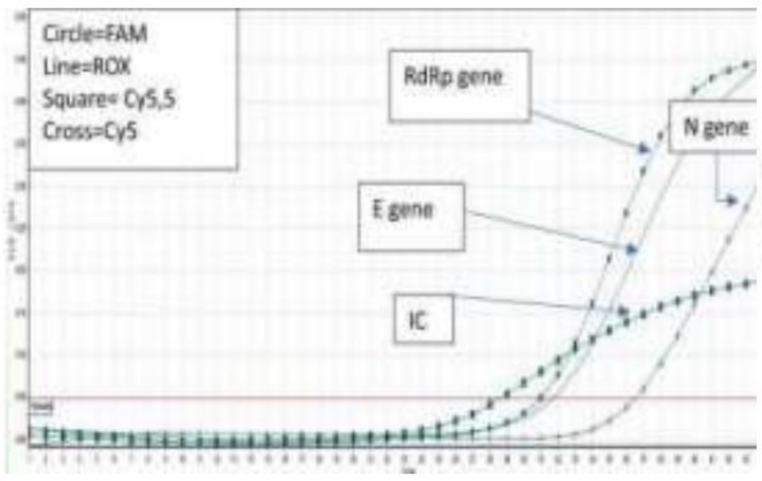
- **We convert RNA into cDNA**.
- We amplify DNA by **qPCR** in the presence of **SYBR green** and the amount of DNA can be quantified according to when the signal is detected.
- The higher amount of RNA >> The higher amount of cDNA >> The earlier it is detected.



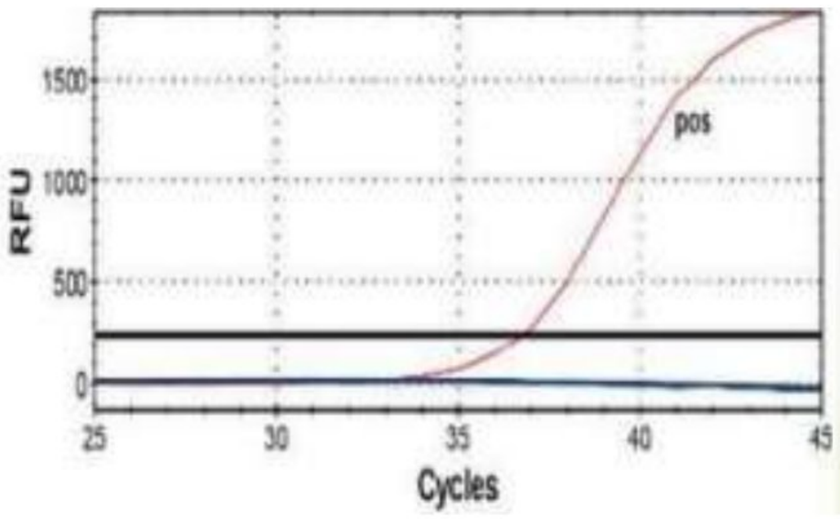
- **Housekeeping gene**: A gene that has constant expression between cells, like actin and tubulin.
- To compare 2 samples, we analyze the levels of expression of a housekeeping gene **to ensure that the amount of starting material (amount of total RNA) in both samples is the same**.
- **Expression of the housekeeping gene should be the same for both samples.**

- **Detection of SARS-CoV-2**

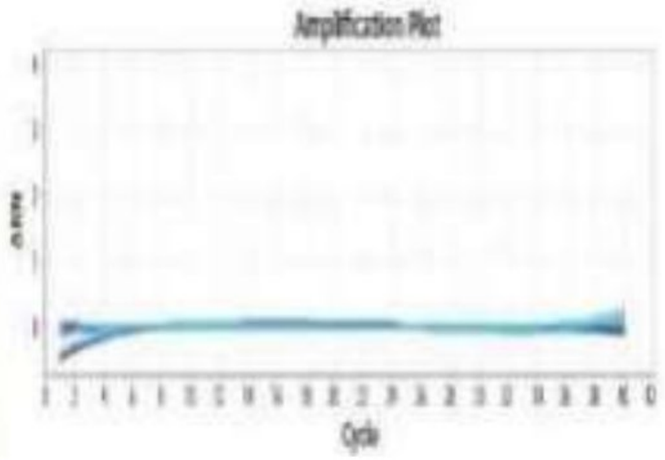
- We amplify **corona virus genes**, like E-gene and we also use an **internal control, IC,** which is an amplification of human RNA, to make sure that the sample is collected properly.



Positive

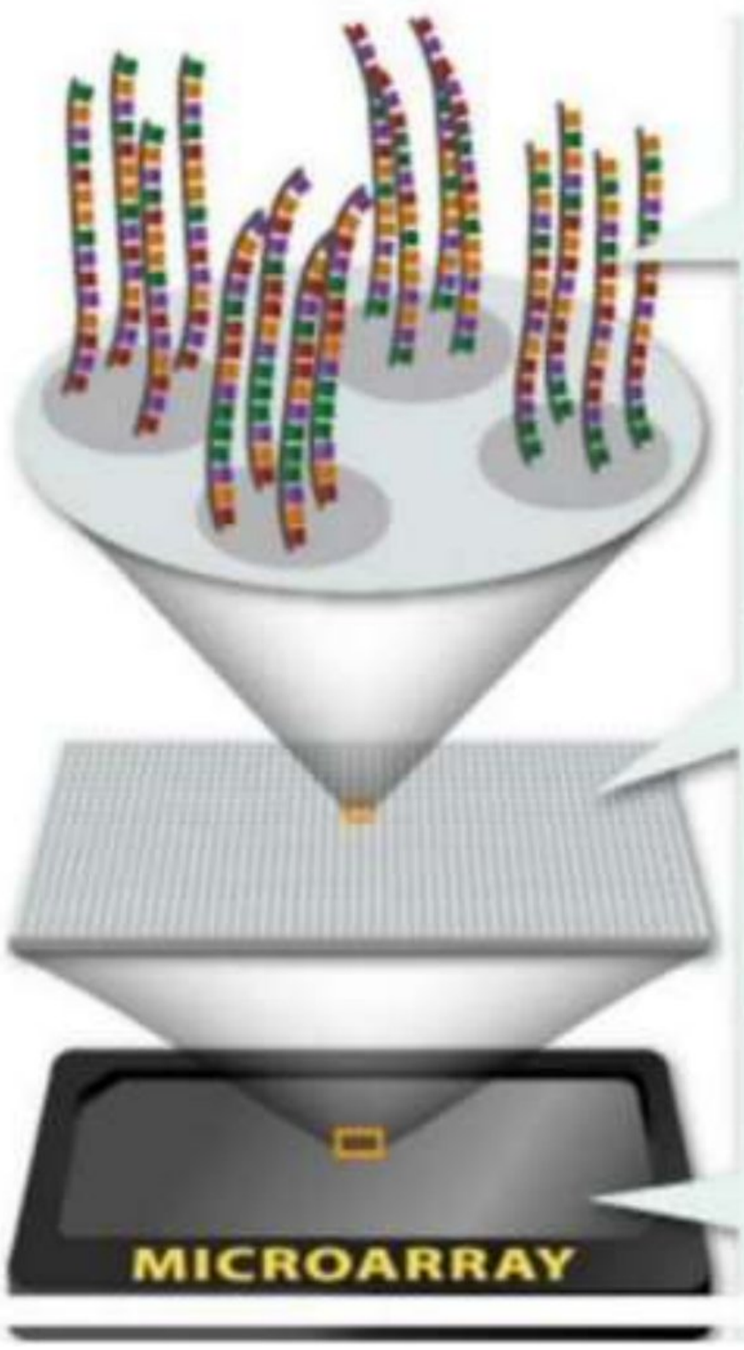


Negative



The sample is invalid

DNA microarray



A DNA micorarray allows scientists to perform an experiment on thousands of genes at the same time.

Each spot on a microarray contains multiple identical strands of DNA.

The DNA sequence on each spot is unique.

Each spot represents one gene.

Thousands of spots are arrayed in orderly rows and columns on a solid surface (usually glass).

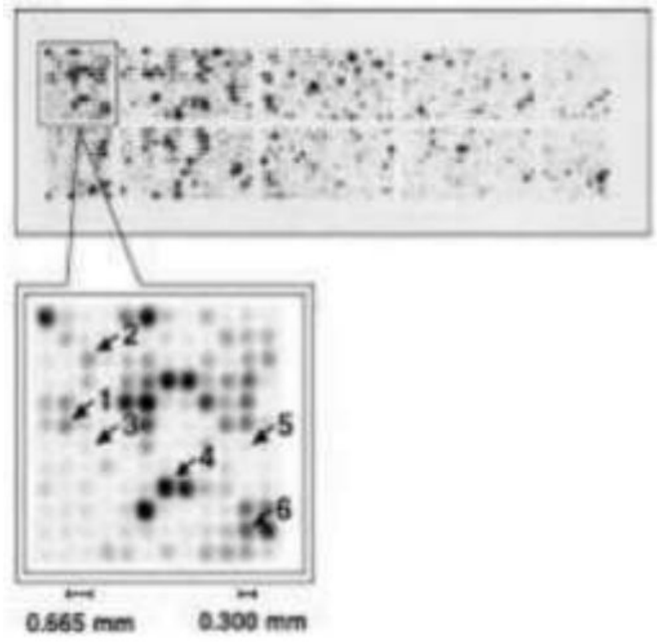
The precise location and sequence of each spot is recorded in a computer database.

Microarrays can be the size of a microscope slide, or even smaller.

- **The exact sequence and position of every DNA fragment on the array is known.**

- **The process:**

- mRNA from the cells being studied is first extracted and **converted to cDNA** and **cDNA is labeled with a radioactive probe**.
- We add cDNA molecules to the microarray and they will bind spots that are complementary to them and we can detect them because they are labeled.

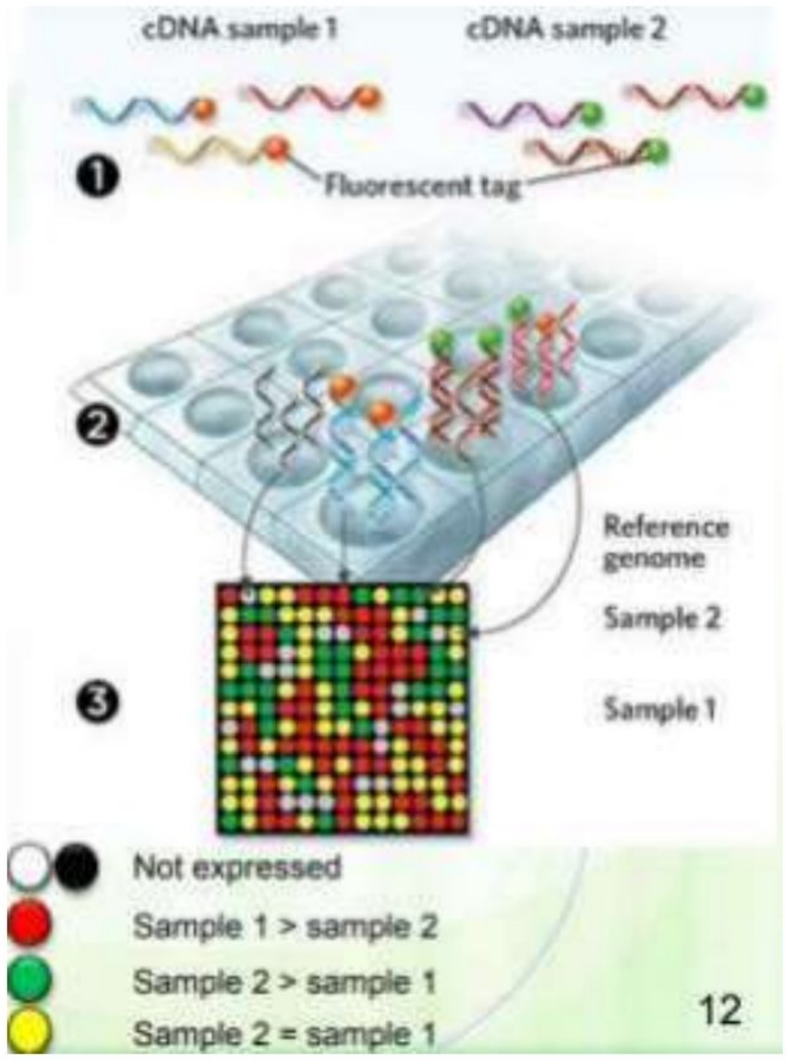


  - ➢ <u>The more cDNA molecules bound to the probes in a certain spot, we will get stronger signal</u>, **which reflects gene expression level – the activity of the gene.**
  - ➢ If there is no signal, it means that there is no cDNA (no mRNA) which indicates that **the gene is not expressed.**

- **Comparative expression**

- We can't compare the expression of 2 samples with radioactivity, because it has no distinct color.
  - ➢ We take RNA molecules from 2 different samples and convert them into cDNA.
  - ➢ We label DNA molecules from different samples with different fluorescence colors (red-sample 1 and green-sample2), then we mix them and add them to the slide.

- Red cDNA and green cDNA will compete to bind the probes in the clusters.
  - If gene expression in sample 1 is higher, red cDNA amount will be more >> there will be **red fluorescence**. (and vice versa)
  - If there are equal amounts of red and cDNA >> **yellow fluorescence**.
  - If the gene is not expressed in both samples ,there will be **no signal.**



12

- We use informatics, and computer makes a comparison between samples and it **eliminates similar genes** , because they aren't informative, and it puts similar samples together to ease the comparison, then we can say for example that a gene is **overexpressed** in one sample and **down regulated** in another.
- **Gene expression pattern can affect prognosis and treatment plan**.

RNA-seq

- RNA molecules are isolated from a cell and **reverse transcribed to cDNAs**, which are subjected to **next-generation sequencing.**
- **This technique tells us 2 things:**
  - The sequence of RNA molecules –what genes are expressed-.
  - The level of gene expression
- **RNA-seq advantages (vs microarrays):**
  - **Characterize novel transcripts (discover new genes)**
  - Identify splicing variants
  - Profile the expression levels of all transcripts.