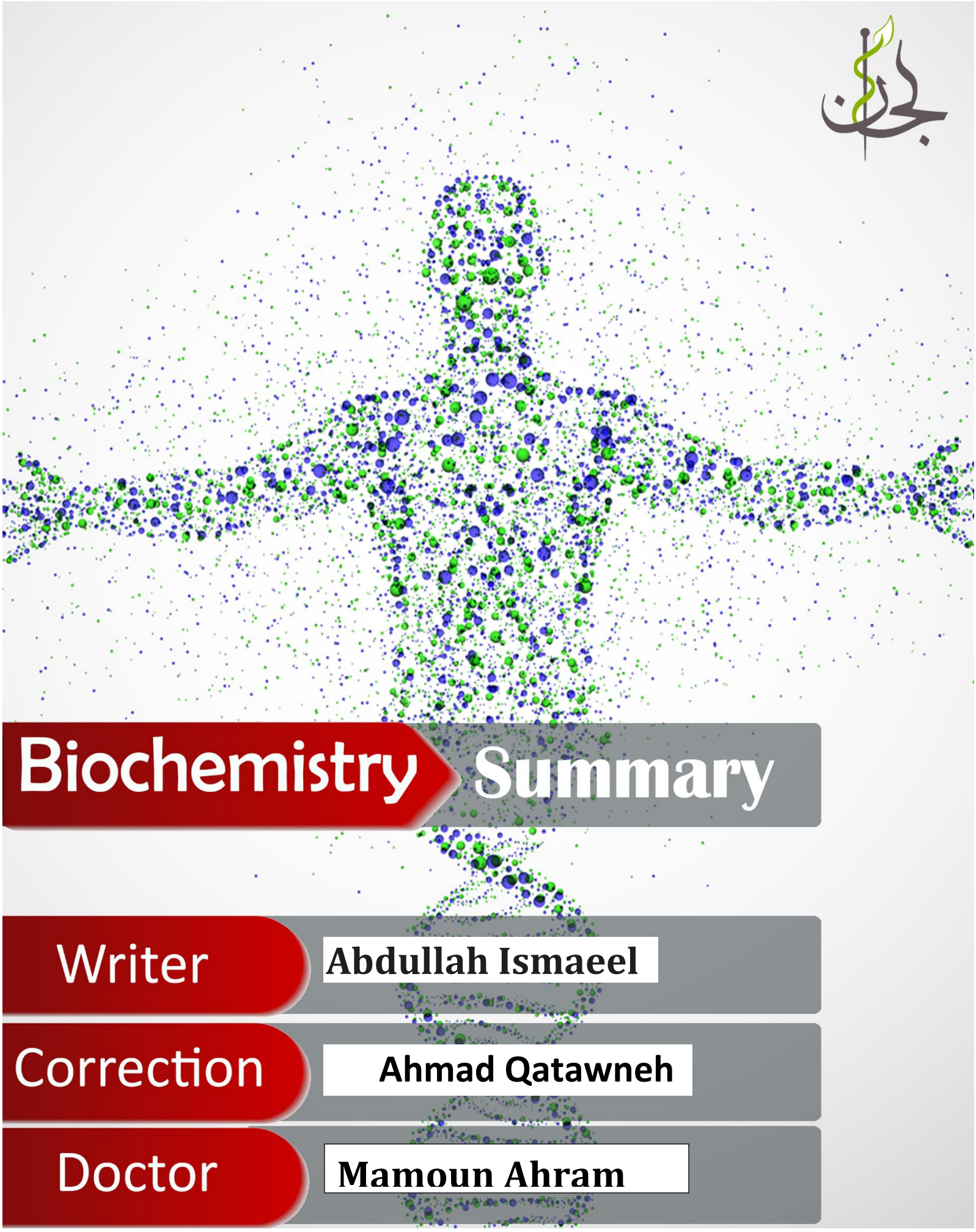


Biochemistry Summary

Writer **Abdullah Ismaeel**

Correction **Ahmad Qatawneh**

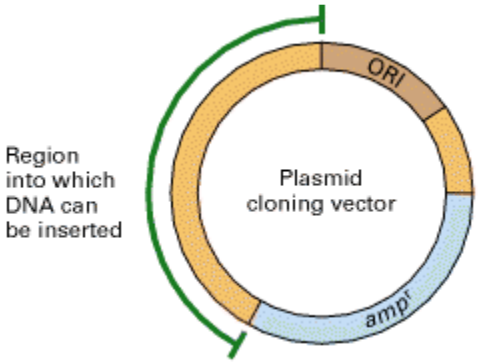
Doctor **Mamoun Ahram**

Biochemistry (Summary)

DNA cloning: is a technique that allows for amplifying a DNA segment, usually in a biological system (such as: bacteria)

A Recombinant DNA is composed of 2 or more DNA molecules from **different** sources.

In order to synthesize it we need: **vector, gene of interest, ligase& restriction endonucleases**



Vector: a carrier; usually a bacterial plasmid. A plasmid is not a part of the bacterial genome (Chromosome).

Must contain at least 3 parts :

it can replicate independently of the genome, can be selected for/against by an internal drug resistance gene (selectable marker). also it must contains restriction site(s).

Gene of interest: protein coding gene

Restriction endonuclease: a bacterial enzyme that cleave (restrict the growth of) the DNA (break the phosphodiester bond), it recognizes specific palindromic sequence (4- to 8-bp Called restriction sites) generating **restriction fragments**.

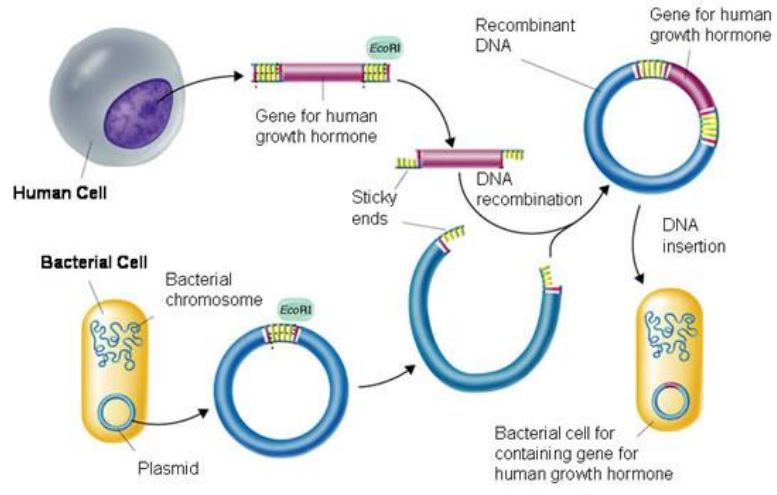
Palindromic: means that it can be read the same 5→3 on both strands.

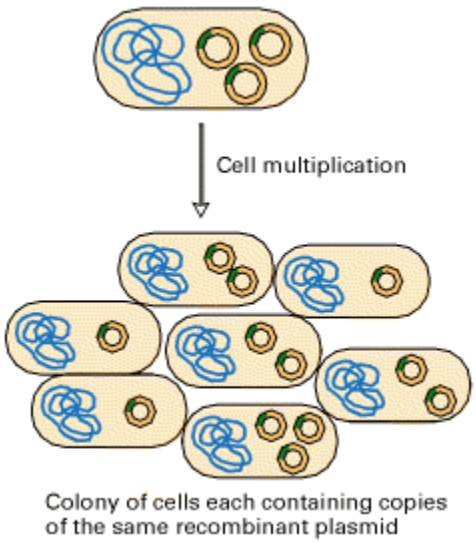
The cuts can be: A) **blunt**: cut at the same position on both strands

B**) staggered**: cut the strands at different positions generating sticky or cohesive ends.

Ligase: Catalyzes the ATP-dependent forming of phosphodiester bond.

Cloning (making copies) mechanism:





NOTE: Both DNA fragments (the DNA to be cloned and a vector) are cut

by the **same restriction endonuclease**.

After inserting the recombinant DNA in the bacteria, it duplicates with it, then we can collect the clones by (**lysing the cell → using the same endonuclease→ purification of the solution**)

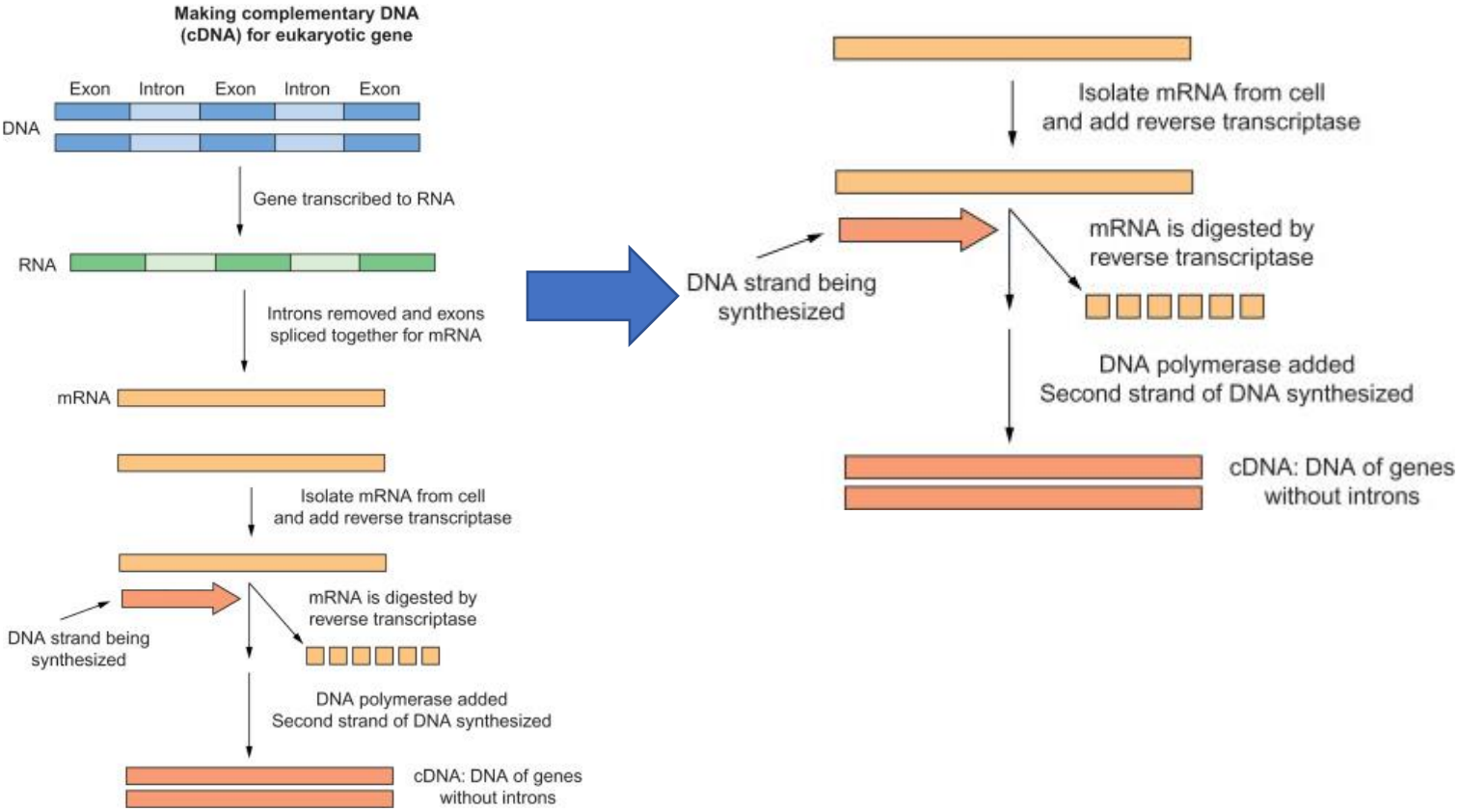
For Expression purposes:

The vector must have another 3 parts:

Promoter sequences upstream of gene to be inserted, **Ribosomal binding sequences** (Shine-Dalgarno [SD] sequences) & **Transcription termination sequence**.

Then the vector is inserted into the bacteria, after protein expression (translation process) the protein is purified (so that we can use it for humans. e.g.: insulin)

But what if the human gene have introns in it? We have to choose the gene from cDNA library (cDNA is produced from mature and spliced m-RNA *converted by reverse transcriptase* enzyme)



Other challenges: **Internal disulfide bonds in bacteria(not exist)**, **No post-translational modification** (example: glycosylation), **Misfolding& Degradation**.

Solutions: using a eukaryotic system such as yeast.

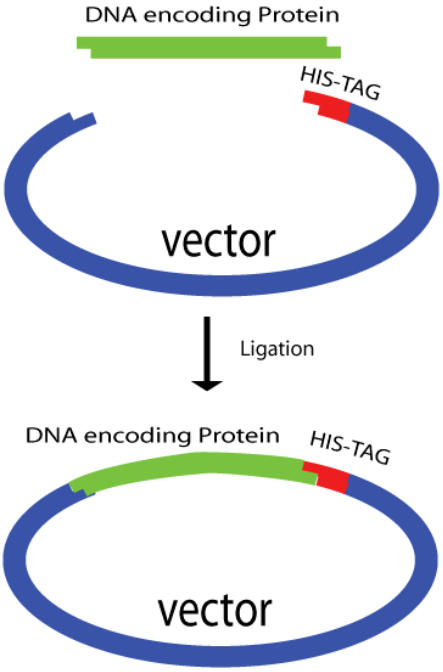
But how can we determine or purify the protein after the cell express it?

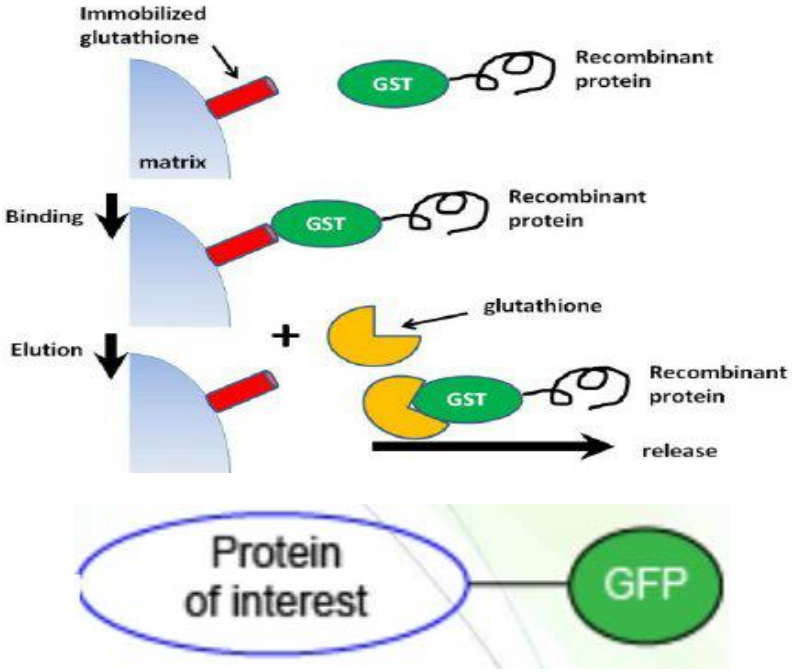
By tagging it with **His-tag** (sequence of 6 histidines), **GST** (helps more in purification) & **GFP** (helps more in detection not purification).

How can we tag the protein? By **synthesizing a recombinant DNA that has the gene of interest next to the tag agent.**

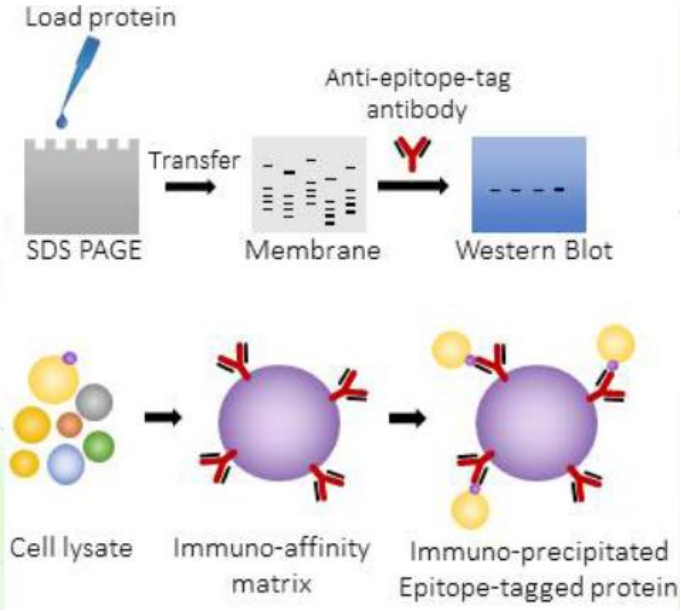
Does this affect the folding process? No due to the fact that the **His-tag** is very small, so it doesn't affect the folding process, GST& GFP (was taken from Jelly fish) **are domains so they fold independently** from the rest of the recombinant protein.

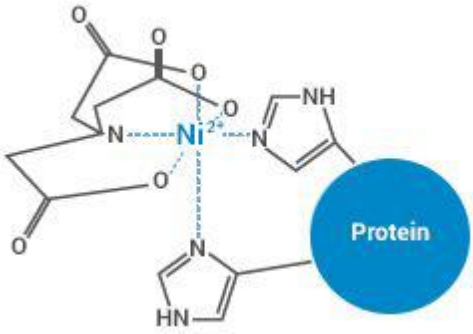
(You are required to memorize the red words in the table in slide 22)





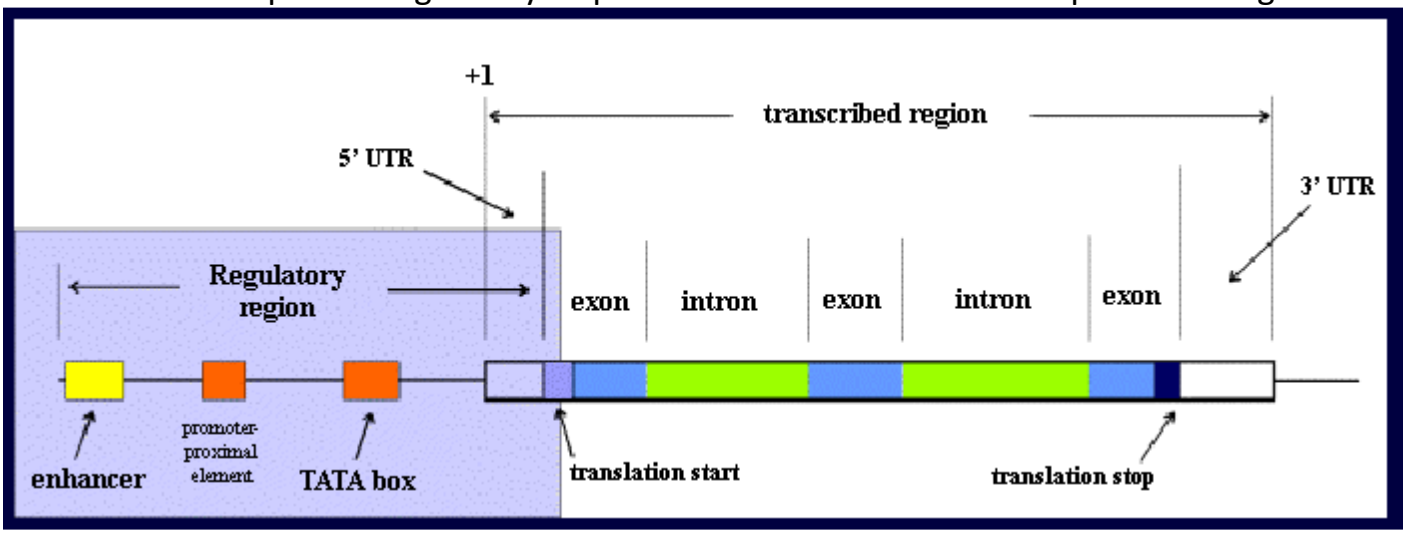
Uses of protein tags:





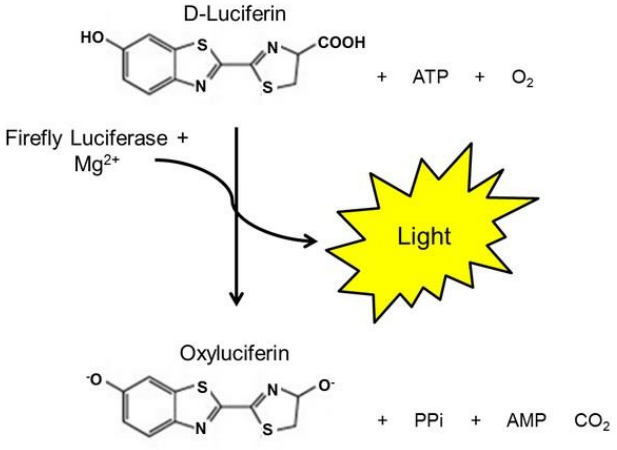
Analysis of transcriptional regulatory sequences:

What are transcriptional regulatory sequences? It controls the transcription of the gene.



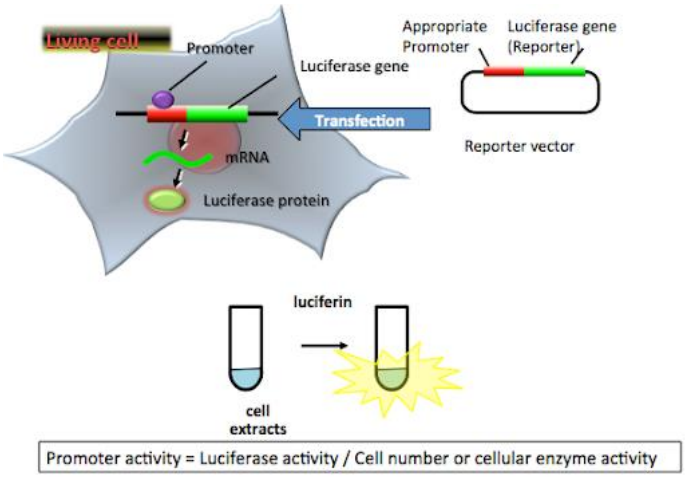
TATA box (core promoter) , Other activating/inhibiting sequences could be present.

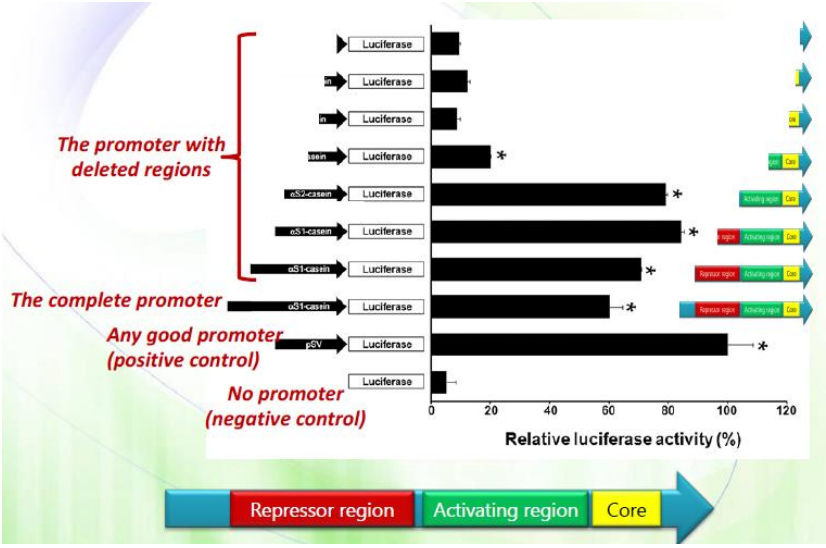
We can use **Luciferase enzyme** as a reporter (which can tell us whether the gene is transcribed or not). This enzyme was taken from fireflies is referred to as **mono oxygenase (oxidoreductase**)



Luciferase assay:

We synthesize a recombinant DNA (from **the promoter of interest+ Luciferase** (as a transcribed reporter gene)) then we put it into expression vector, so the intensity of light represents the activation degree of promoter.

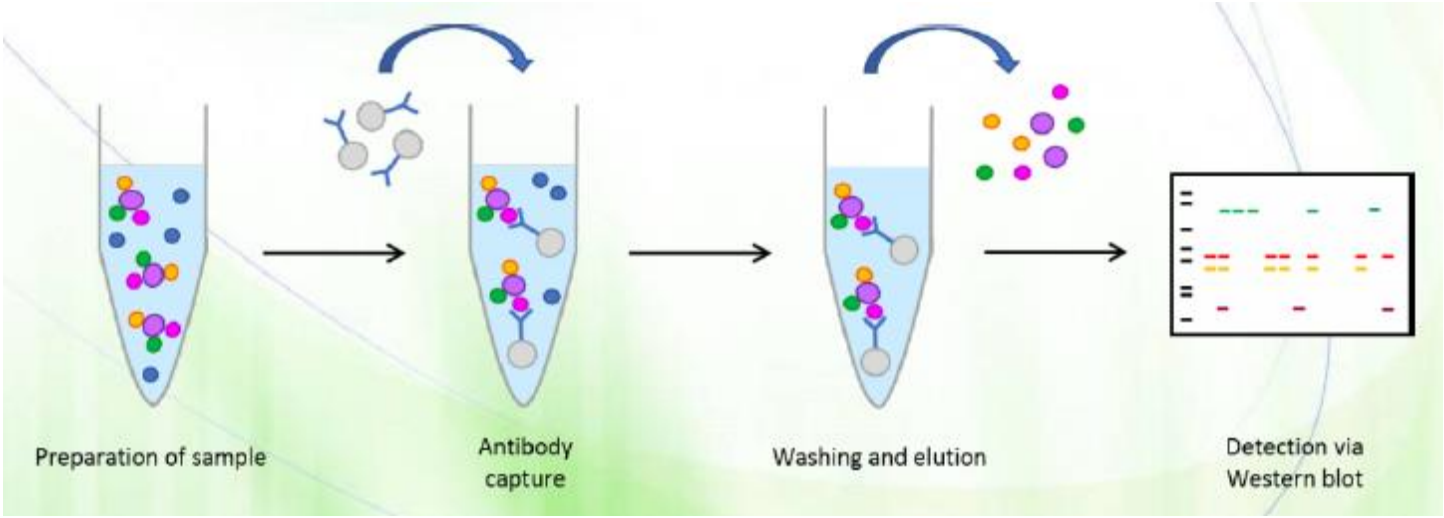




Protein-protein interaction:

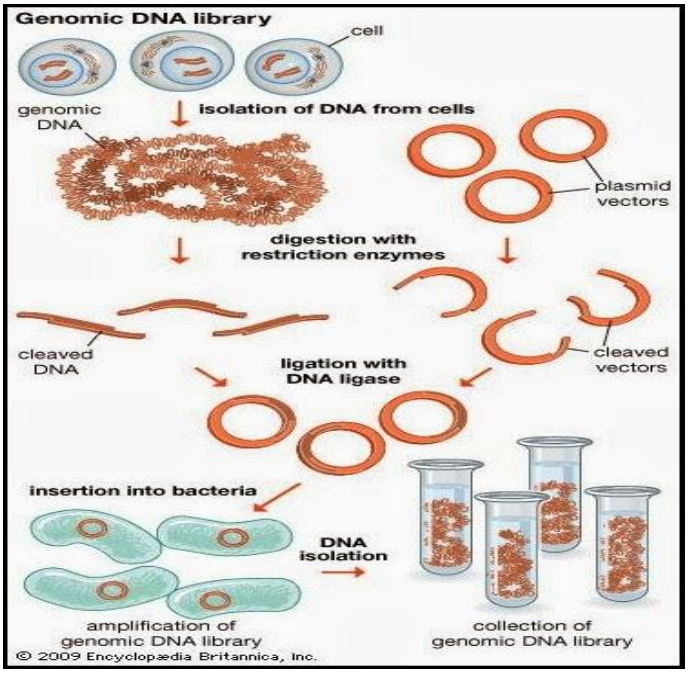
Determined by two different techniques:-

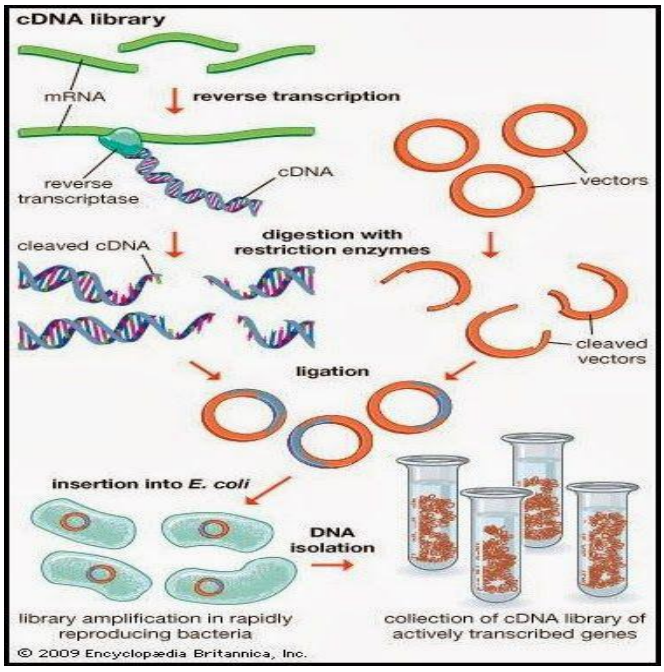
1- (Co)-Immunoprecipitation: if we have proteins A&B and we want to check if they can interact & forming complex. We do this:
We conjugate specific **antibodies for protein A with beads**, so that only the protein of interest (A) is bound-precipitated along with other proteins (B) that interact with protein A , followed by washing and elution of the proteins so that we can use western plotting to identify if there are other proteins that interacted with our PROTEIN (A).



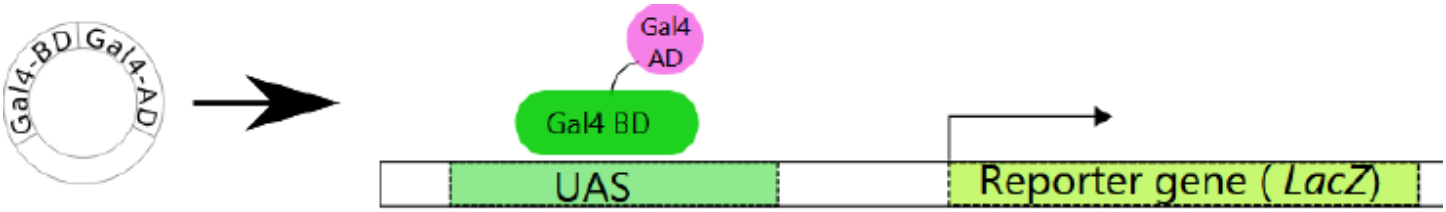
2- Yeast two-hybrid system:

In order to avoid the presence of introns we depend on cDNA library





In the normal situations, the transcriptional factor composed of 2 domains (binding & activating). If the domains are close to each other the promoter will activated → the gene of interest will be transcribed , but if they were far away from each other or one of them doesnt exist then nothing will happen.

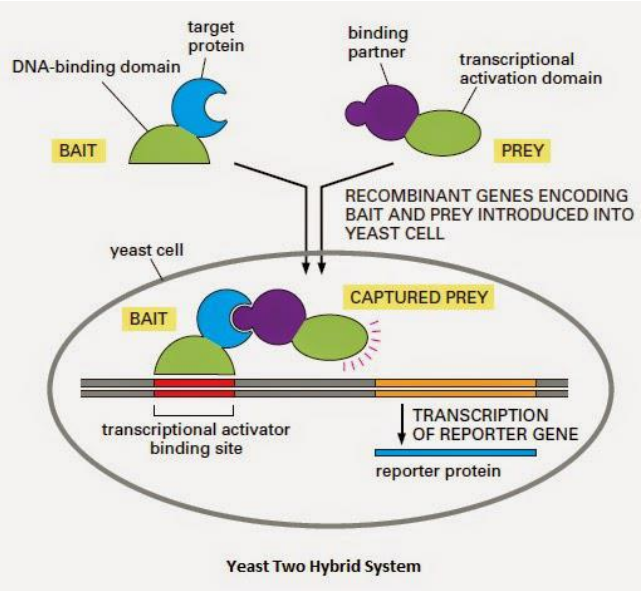


A. Regular transcription of the reporter gene

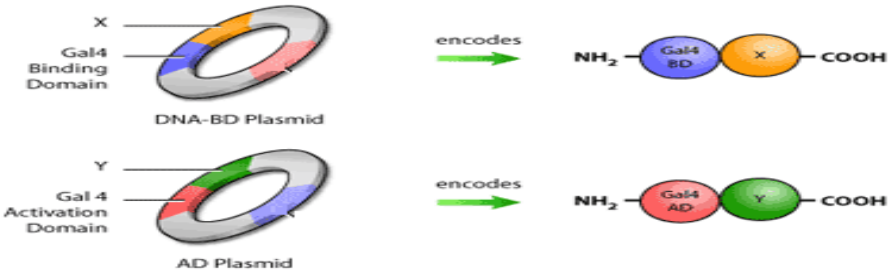
UAS= upstream activating sequence.

The reporter gene here is (lacZ) which is transcribed to give B-galactosidase.

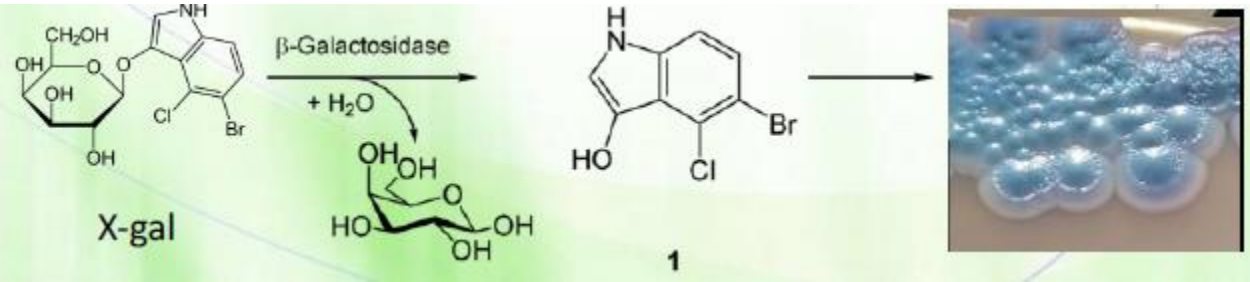
We can take advantage from this by binding one domain to one known protein & other domain to another unknown protein so if they interact, the domains will be close to each other and the activating and transcription will happen. But if they don't , nothing will happen indicating that the known protein doesn't actually interact with the other unknown protein we used.



How can we bind the domain with the protein of interest? By synthesizing recombinant DNA which has the genes of both proteins of interest (before transcription) & the domain (B or A) and put them into expression vector. Then we collect the released proteins & do the experiment.



How can we know if the reporter gene went into transcription? If it goes well then, we will have the production of the **B-galactosidase** protein & instead of cleaving lactose we introduce another molecule (**X-gal**) which produce a blue color when it is cleaved indicating by **B-galactosidase** production of the reporter gene (beta-galactosidase), as we observe the formation of blue bacterial colonies. We can then isolate these colonies to find out the properties of the protein that interacted with our known protein.

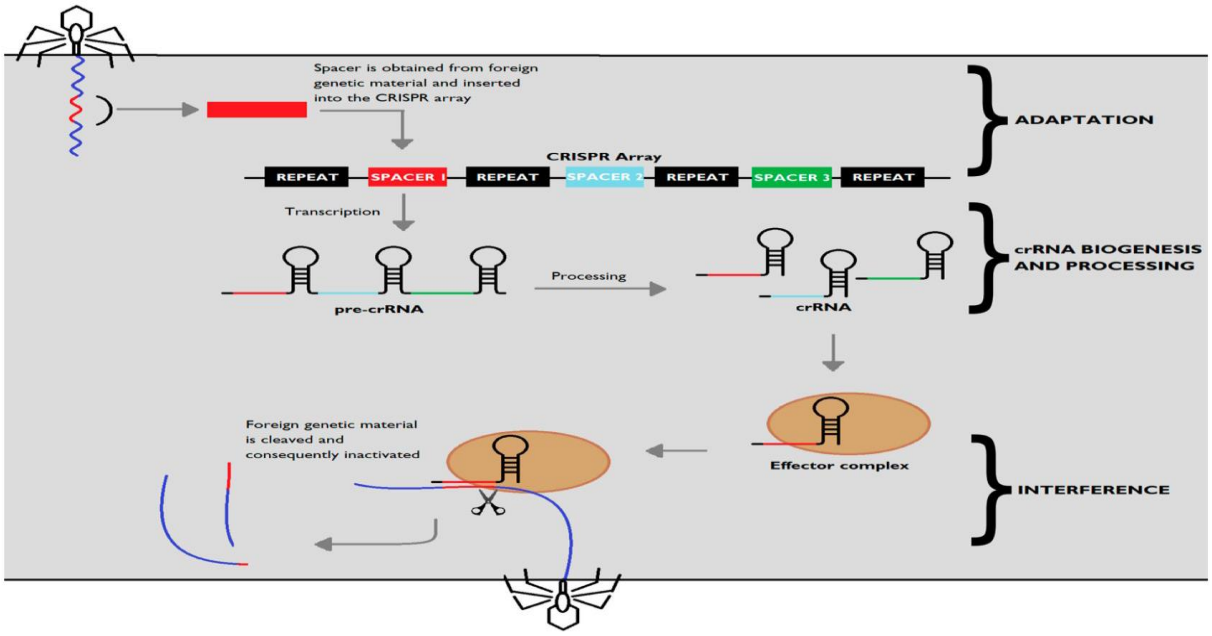


CRISPR-CAS9 and gene editing:

CRISPR: is clustered regularly interspaced short palindromic repeats (bacterial immune system)

Cas9: is a RNA guided nuclease that can either create single or double strand breaks. It's guided to its target by guide RNA (gRNA) or single guide RNA (sgRNA), which is complementary to the target sequence.

Mechanism of work: starts with infection by phage, the bacteria chops fragments of it & put it in CRISPR system, then it will be used to create gRNA that bind to CAS9 protein and guide it to the sequence of the same phage genome in the next time it tries to infect the bacterial system. When the Cas9-nuclease is guided to the viral-phage DNA, it cleaves it effectively preventing infection.



How can we take advantage from this? We can use this system to cause break in specific place in the DNA of human

Therefore, we need to Recall the mechanism of DNA repairing.

HR=homologous recombination-repair

(it can: correct mutation, introduce mutation& insert gene)

NHEJ=Non- homologous End Joining

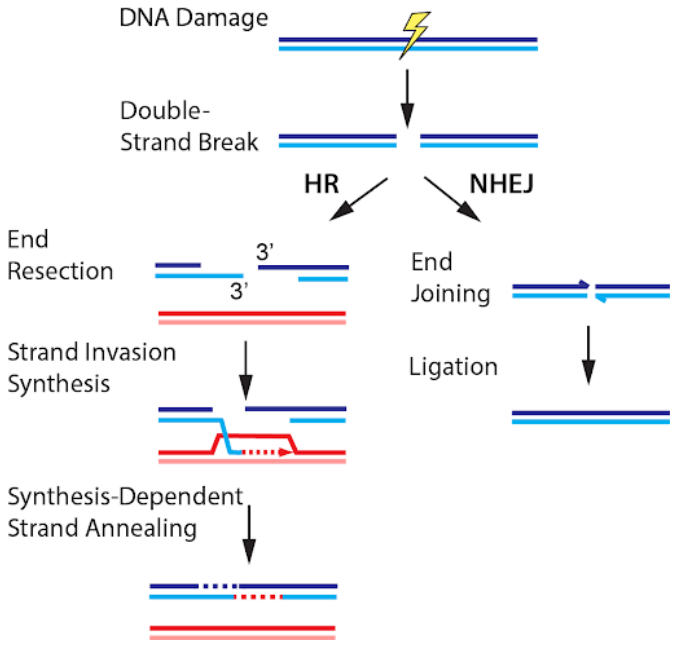
(it can: knock out the gene)

(It could be prone to errors leading to mutations)

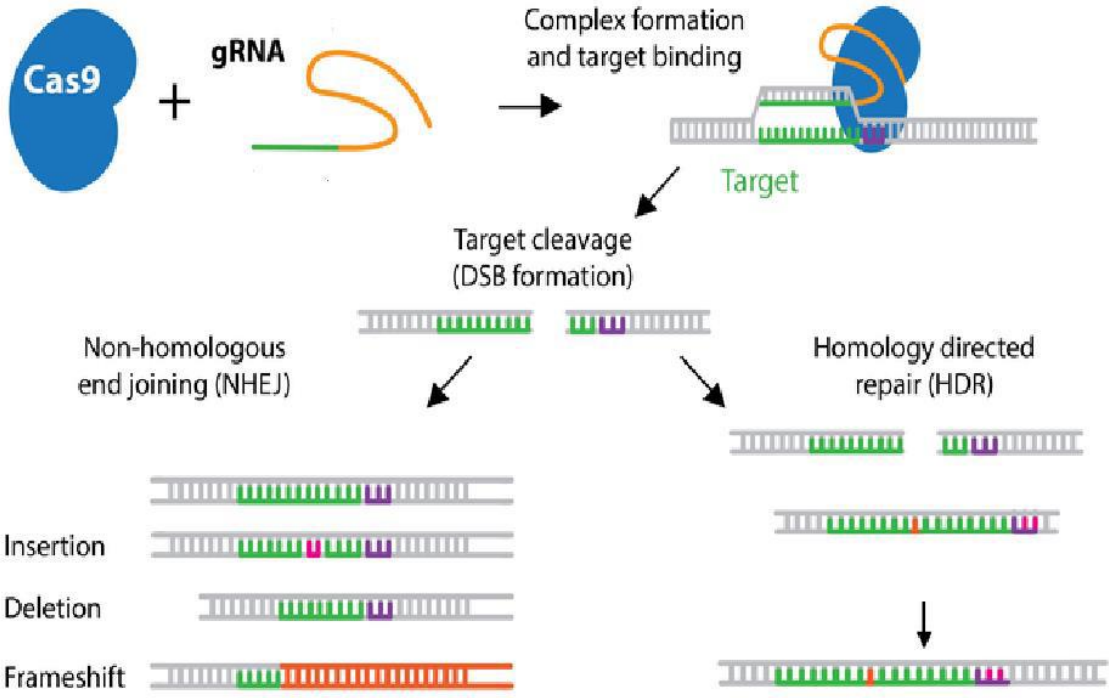
Therefore, if we put a specific chromosome in HR

Mechanism, we can control the repairing sequence

By using the 2 systems together we can control the place of break & how to fix it



(Both the gRNA and Cas9 gene can be introduced into human cells as genes cloned into plasmid vectors.)



DNA sequencing: (knowing the order of nucleotides in the genomic-DNA or piece of it)

Importance:

- 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

THE scientist did DNA sequencing for:

Viruses& prokaryotes→ human mtDNA→ yeast→ multicellular nematode→ human genome

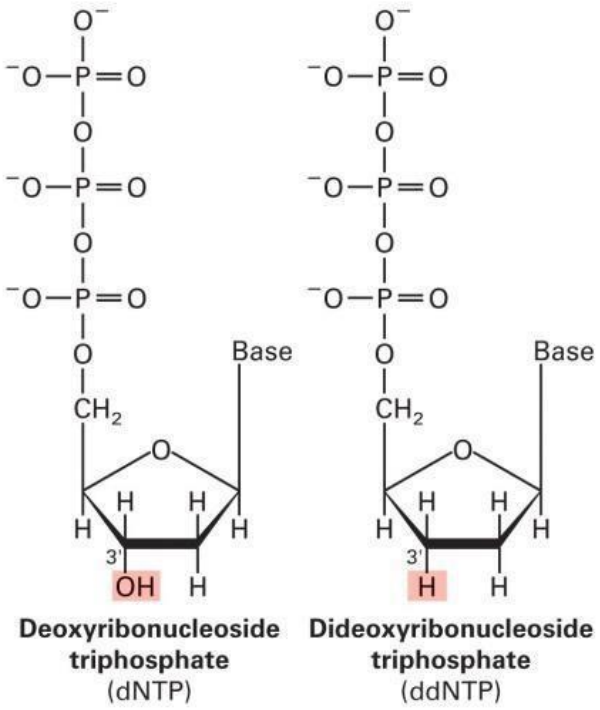
Organisms have variations in the number of genes& base pairs.

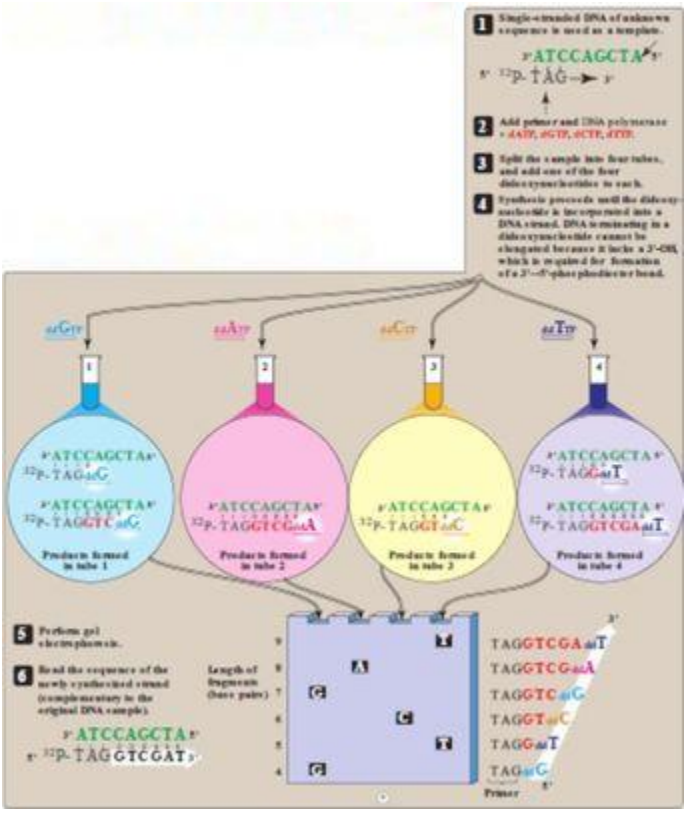
How can we do that? There are 2 ways (old & modern)

Old ways: 1- **Sanger ddNTP method**: (ddNTP→ terminates DNA growing)

We need: **DNA template to be sequenced, DNA polymerase, dNTP, labelled primer and ddNTP.**

NOTE: the new DNA is the complementary one





2-Fluorescence based DNA sequencing

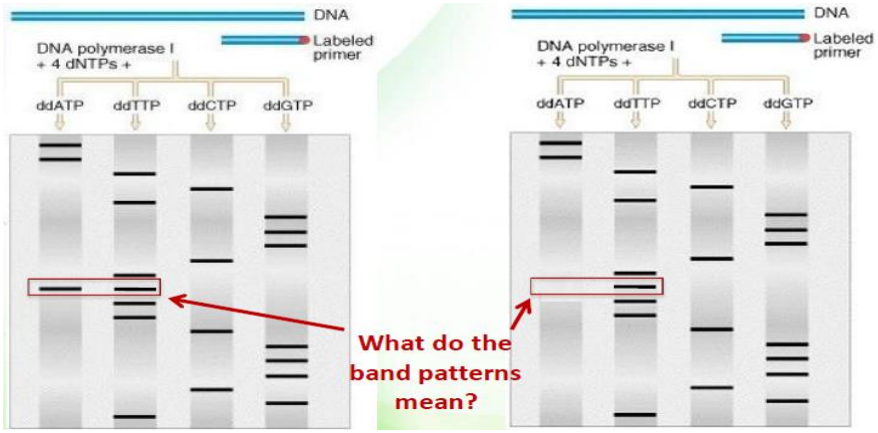
Instead of using labeled radioactive primers, we use each ddNTP labeled with a unique fluorescent tag. We used the same mechanism, but we let the computer look into the colors & do the sequencing.

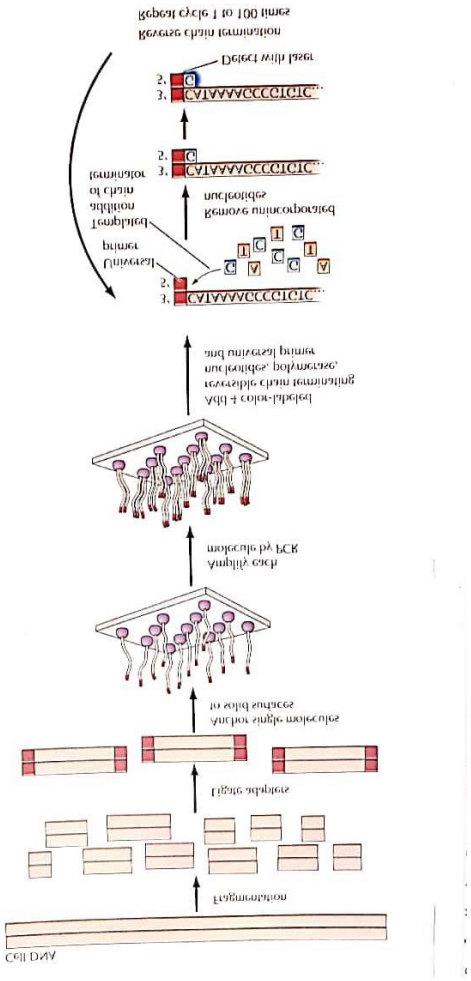
However, these mechanisms are dangerous, expensive& time consuming so we use instead of them the modern methods:

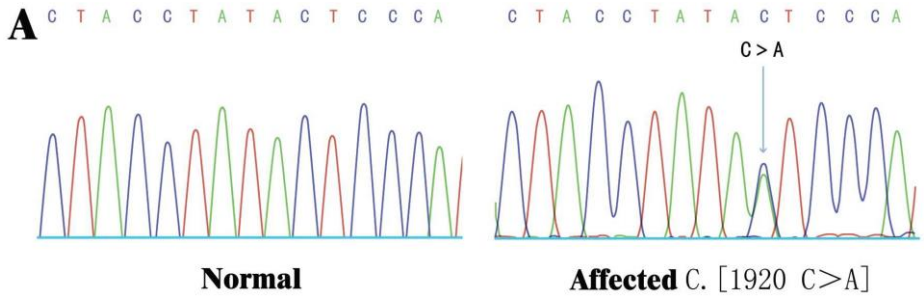
Next generation sequencing: we turn DNA into fragments then we attach the fragments to adapters in order to attach them to primers& platforms, after that we add special nucleotides that are activated (by light source) & it gives fluorescent signal when they attach & 2nd one can't attach until the 1st one is activated, by analyzing the colors we know the sequence.

What about band patterns? What do they mean?

1- If there is an indication that there is 2 different nucleotides on the same position on different chromosome, we say this person has heterozygosis or has mutation on one of the chromosomes.
2- If there is no indication, we say the person has homozygosis or has mutation on both chromosomes







Polymerase Chain Reaction (PCR)

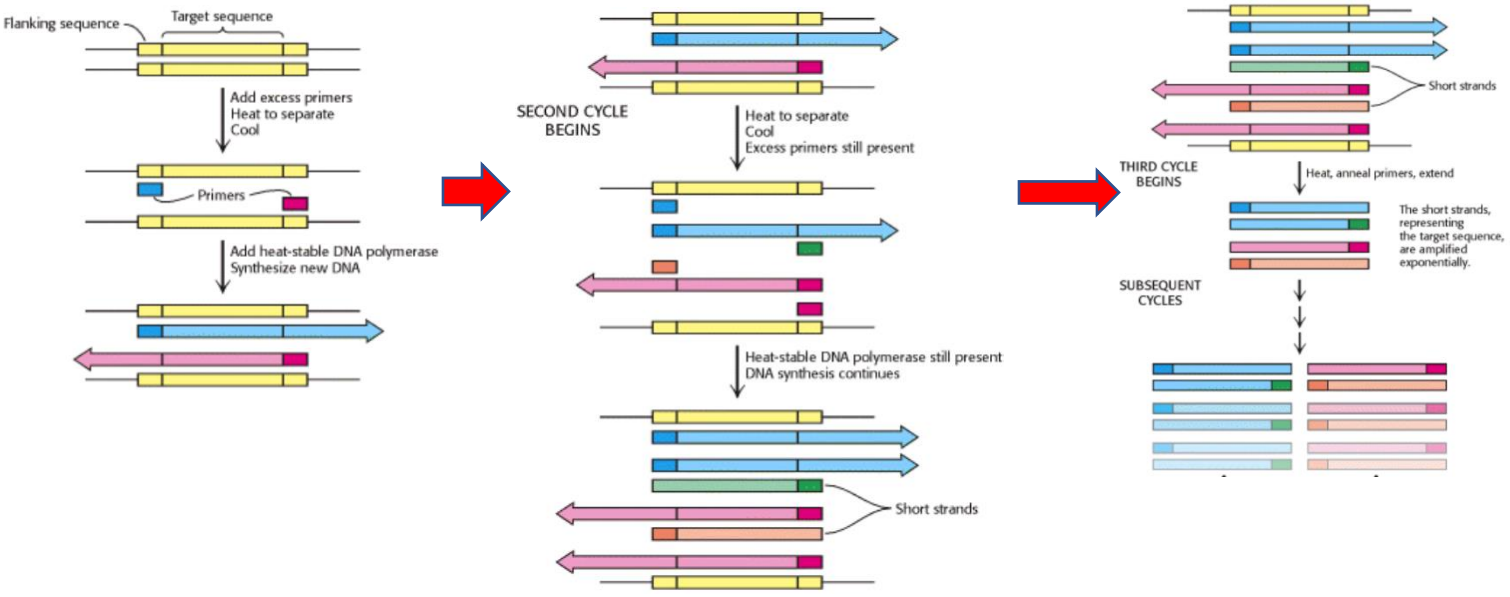
Amplifying selected region of a genome, effectively "purifying" this DNA away from the remainder of the genome. (Really fast specific cloning, catalyzed by DNA polymerase)

Components: DNA template, pair of primers, dNTPs, heat-stable DNA polymerase.

This reaction have cycles(25-30), 3 steps:

1- **Denaturation** (at 95°C): DNA is denatured into single stranded molecules.
2- **Reannealing** (50°C to 70°C): the primers anneal to the DNA.
3- **DNA synthesis** (at 72°C): optimal for the polymerase

Heat-stable DNA polymerases: such as (Taq DNA polymerase) is obtained from a thermophilic bacterium, **Thermus aquaticus**. Can be tolerant to temperature up to 95°



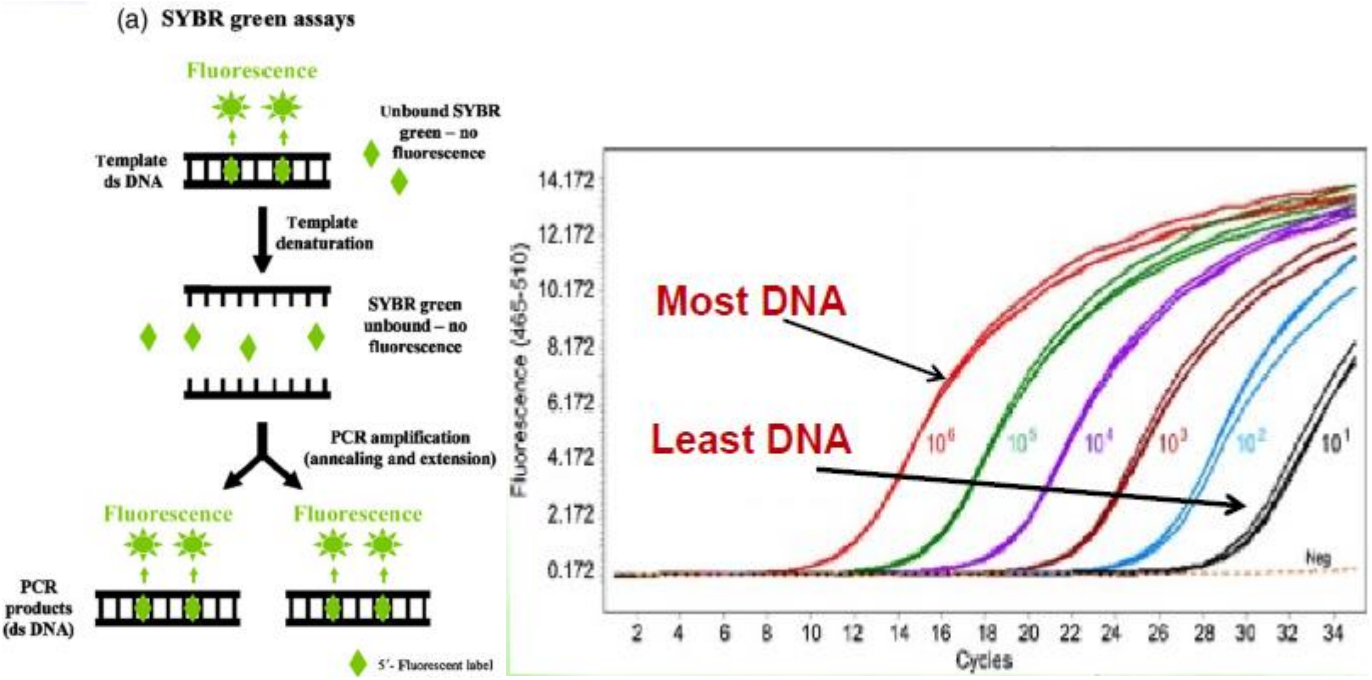
- After 30 cycles, there will be over 250 million short products, after the reaction we detect the presence of DNA by gel electrophoresis.
- The specificity depends on the primers (we need 2 to surround the gene of interest)
- We prevent imperfect hybridization by rising the temperature
- We can take advantage of it to find homology which indicates gene families

Uses of PCR:

1- Discovery of gene families
2- Disease diagnosis (we don't see deleted genes& we note that there is shortened gene)
3- Paternity and criminal cases
4- Viral and bacterial load: the quantity of virus in a given volume (by **qPCR)**

What is Quantitative PCR (qPCR)?

We use SYBR green particles that bind to double-stranded DNA and fluoresces (at 3rd step) only when bound→ the amount of fluorescence represents the amount of DNA → if w are interested in identifying viral or bacterial DNA, we can detect there load, that is : the amount of DNA .



The higher the amount of DNA in the sample, the sooner it is detected (less cycles)

	Initial amount	1st cycle	2nd cycle	3th cycle	4th cycle	5th cycle
Red sample	1000	2000	4000	8000	16000	32000
Green sample	100	200	400	800	1600	3200

(After 5th cycle we detect signal from the red sample not the green sample) due to the fact that the initial concentration of red sample is far greater than that of the green sample and subsequently, we will be able to detect the signal sooner (less amount of cycles).

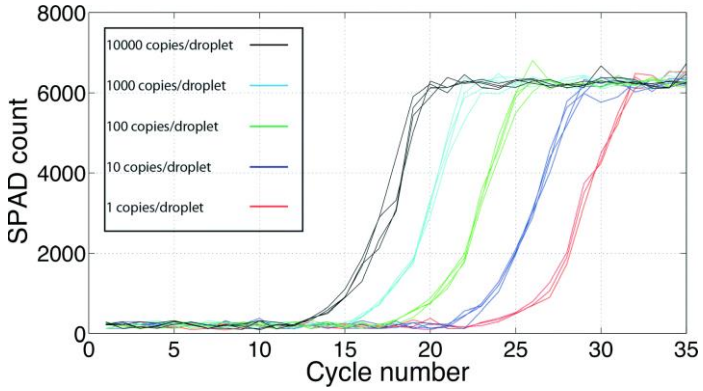
How to know RNA level ?

- **Basic (old) methods**: northern blotting (thickness of the band) & in situ hybridization (we know relative expression& in what tissue section).
- **Advance methods**: real-time PCR, DNA microarray
- **Very advanced methods**: RNA-seq

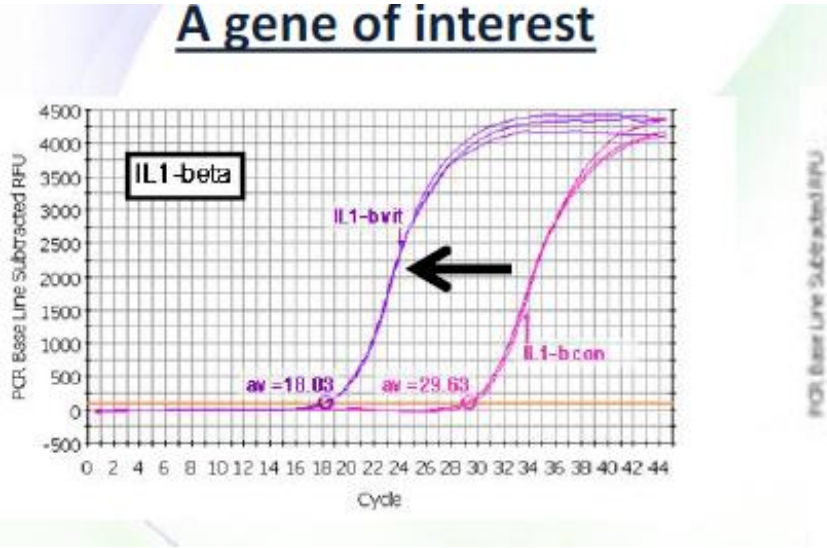
We will talk about the advance methods, but in order to do them we need to convert m-RNA to cDNA (by **reverse transcriptase enzyme & oligo-dT primer which attach to poly-A tail**)

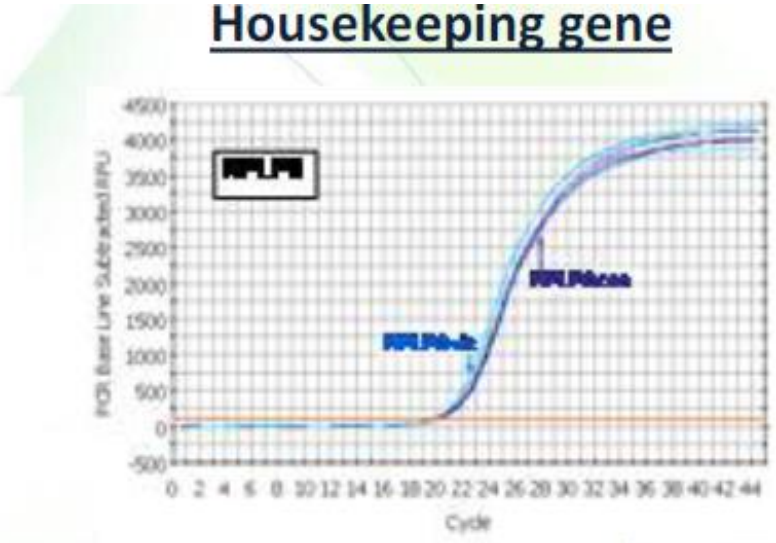
RT-PCR (can be quantitative): the same mechanism but here we use it on cDNA in order to know RNA level (all this instead of using DNA in qPCR)

the higher RNA in the sample, the higher cDNA, the sooner it is detected.

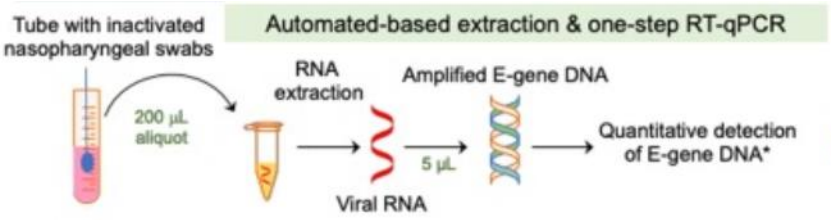


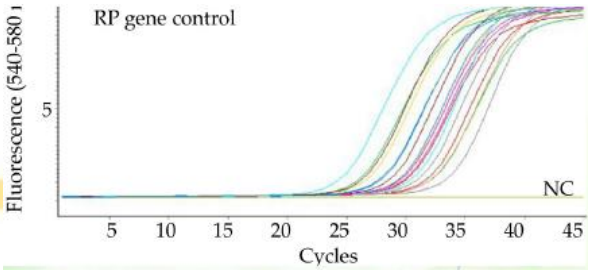
How can we sure that the same amount of starting material? By analyzing the levels of a house keeping gene test. (A gene that is expressed at the same rate in all types of cells regardless of the conditions)





We use this technique for detection of SARS-Co-2. How?? (Remember the genetic material for this virus is RNA)





(X)-omics: studying total collection of (X).

(We can define the transcriptomes as the studying of all active genes (can be transcribed) in a cell at the same time)

How can we study the expression of all genes at the same time?

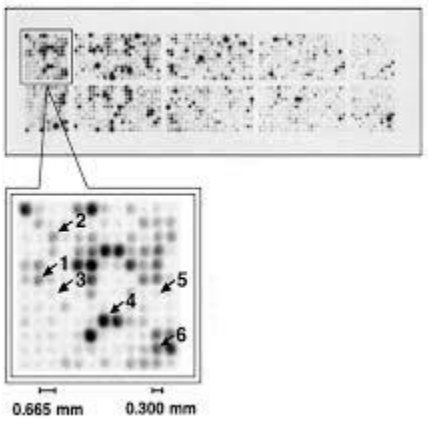
1- DNA microarray: (array= sequence of something)

We use tens of thousands of spots in which each spot has multiple probes(not labeled, identical & specific for the gene)

We isolate m-RNAs→ convert them for cDNA→ label the cDNA→ the cDNA tries to bind to a complementary probe specific for one of these spots (specific for a given gene).

(Intensity of the signal represents the degree of expression)

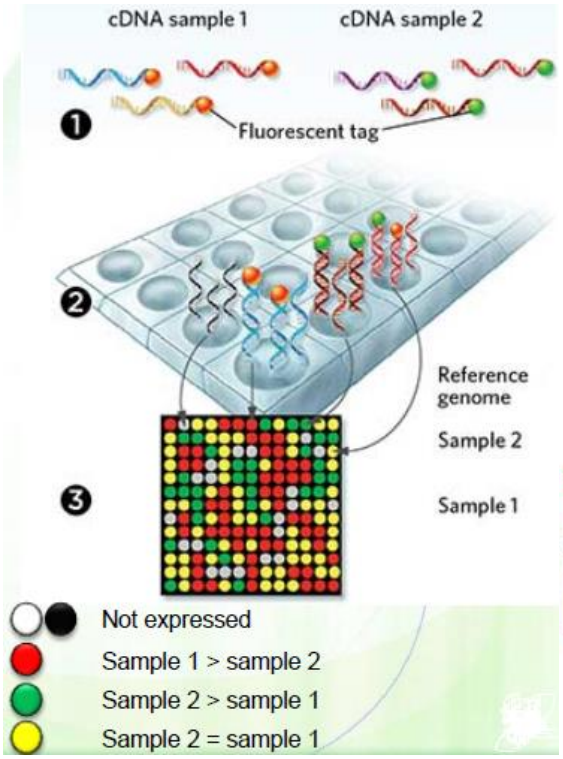
We can be more precise and do **comparative expression**

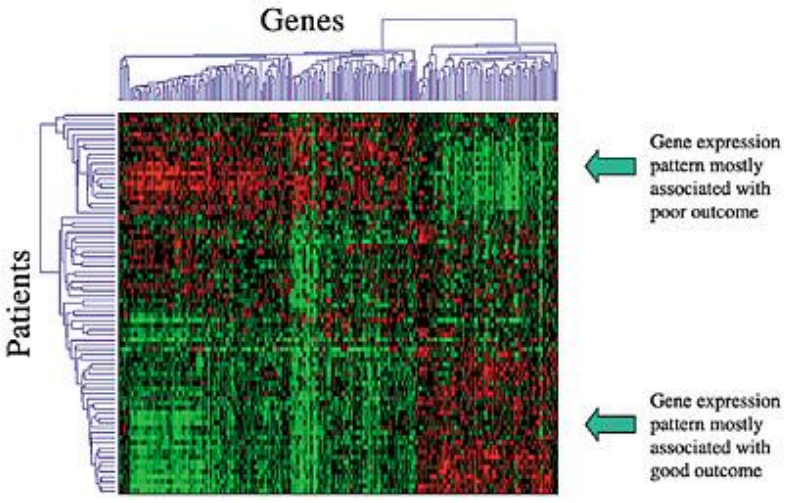


It is done with 2 samples (control vs other)

We do the same things, but **we label cDNA with fluorescence** then they will bind & we compare

Computer eliminate non-significance results& cluster the results according to similarities.





Variation in gene expression gives variation in the outcomes of treatment, so we can create molecular profile & give treatment to the patient based on it.

2- RNA-sequencing:

We use this technique with any type of RNA not necessarily m-RNA and it doesn't require previous knowledge about the sequence of nucleotides within these RNA-molecules.

Mechanism of work:- Cellular RNA (all types) is reverse-transcribed into cDNA by using a RT, followed by performing **NGS-(Next generation sequencing)**.

It is possible to identify the sequence & relative amount of each cDNA (mRNA) which is indicated by the frequency at which the cDNA is sequenced in NGS.

Uses of this mechanism:

1- characterize novel transcripts
2- Identify splicing variants resulting from the same gene.
3- profile the expression levels of known transcripts.

The major difference between DNA-microarrays and RNA-sequencing:-

In **DNA-Microarrays**, we need to have a prior knowledge in regard to the sequence of cDNA we are trying to analyze. Whereas in **RNA-Sequencing**, it can be used to identify unknown genes (sequences).