Genetics & molecular biology

◉ Sheet ○ Slide

Number:	13
Done by:	LEEN
Corrected by:	RAZAN
Doctor:	Ma’moun Ahram

Transcription related techniques

In the last lecture we start talking about techniques used to study gene expression

The first technique is affinity chromatography that study what transcription factor bind to certain **DNA sequence**. If I want to know a certain transcription factor to what DNA sequences bind to, we use **chromatin immunoprecipitation**.

Q: Is expression of these genes different or not?

The first technique that answer this question is **NOUTHERN BLOTTING**

In northern blotting what we do is:

1- **Separate mRNA molecules in a gel**. Why?

Because I have large genes (size of exons) and we have small genes as well. So large genes will produce large mRNA and small genes will produce small mRNA.

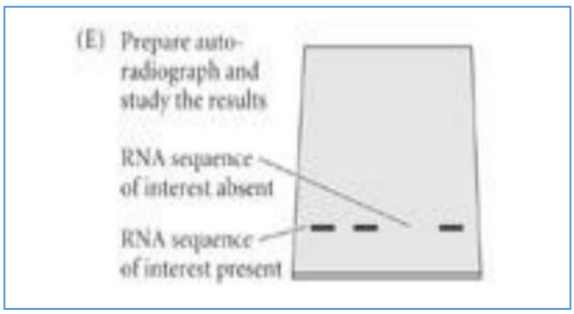
Side note:
There is 3 different type of blotting:
southern >> DNA.
northern >> RNA.
western >> proteins.

I have 20,000 genes, let's talk that 5,000 **genes** are expressed at a certain time point at a certain cell type→I will have 5,000 different **RNA molecules** with different sizes (range may be 100-10,000 nucleotide moles).

2-We transfer them to a membrane

3-And then we add a specific **probe**, cause each gene has its unique sequence that's different than other genes.
So the probe will bind to the complementary RNA molecule to it

In the right picture, we have the results of 4 samples. On the top is the large molecular weight RNA molecules, and in the bottom we have small RNA molecules, What's this results mean?

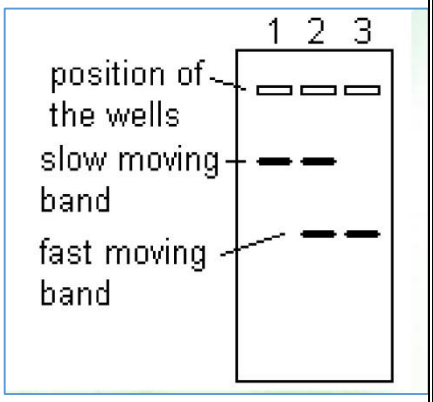


- Same RNA molecule (same size & bind to the same probe)
- 3 out of 4 samples have this RNA molecule
- Sample number 3, there is no transcription to this gene

On the picture no.2, we take samples from different tissues:
1- we have tissue specific gene expression
2-the amount of transcription is different (thickness is different).

So, northern blotting tells us is there **expression** or not, the **amount** of transcription and what is the **size** of mRNA (if present).

On this picture, we have 3 samples, we add a probe to these samples.
The 1st give one signal (lets say **4k** base RNA), 2nd give 2 signals (**4&2** k base) RNA molecule, the 3rd give one signal (**2k** base) RNA. What does that mean?
(There is no one answer)



-it may be **<u>alternative splicing</u>**. If it happens in a certain way, sample 1 give one RNA product, sample 3 give one other RNA product and sample 2 gives 2RNA products (need both forms).

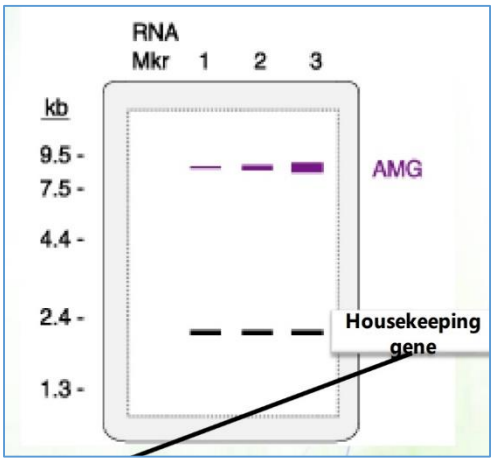
-sample 1 **homozygous** mutations**,** sample 2 **heterozygous**, sample 3 **homozygous** for a certain variant.

-if I have more than one **<u>Polyadenylation</u>** site**,** it will produce different RNA products, one stopped at the 1st site, the other stopped at the second site

-sample could be **contaminated.**

-**<u>probe may not be specific</u>,** cause we have gene family (like histones, 2A/2B), so the same probe may detect histones 2A&2B because they have highly similar sequences (this is assumption) . This is called **<u>non-specific hybridisation</u>**.

In this experiment, we study if estrogen affect the expression of certain gene (AMG) or not. The 1st is untreated, we add 1mmolar estrogen to sample 2 and 10 mmolar to sample 3
And then I see what's the expression of this gene



When we add the hormone, AMG transcription increase (because the band become thicker),I can conclude that estrogen induces expression of gene AMG.

House keeping gene: is a gene that is needed by all cells an expression of this genes does not change with different conditions (like: actin, tubulin, histones) -equal amount of RNA-
On the experiment, the intensity of house keeping gene is identical, that's mean that the change that I see in AMG expression is real
That's what northern blotting can tell you!! 😁

In situ hybridization (in situ=in particular place)

The idea of in situ hybridisation is to know in which type of cells genes got expressed in a tissue sections.
If I want to study RNA expression in a tissue, I can use **northern blotting** to know if there is over expression of a gene (AMG for example) in a patient who have the disease.
And we use **IN SITU HYBRIDISATION** to know which type of cells in a tissue are the cells that are expressing that gene.

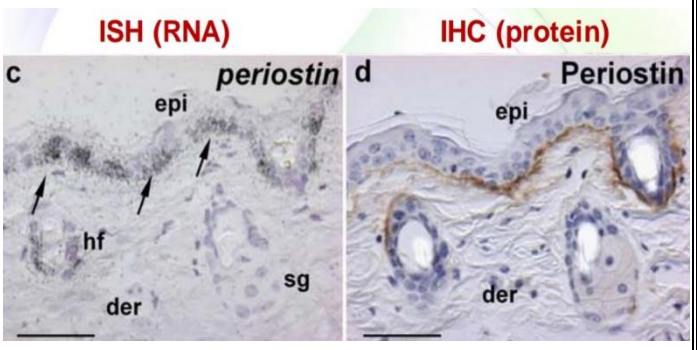
Can we determine that certain gene expressed in a certain type of cells based in immunohistochemistry (protein stain)?

It may be wrong. Because the gene may be expressed in one cell, then the protein leaves immediately and enters other type of cells.

So, if we have a tissue section with different types of cells -like differentiated and undifferentiated cells- What we do is adding the probe to the tissue

The **probe** will bind to the **RNA** that exist in that cell → it will give a **signal** → that cell is the cell that express this particular gene.

In this tissue, when we use immunohistochemistry (IHC), protein appears in basement membrane (outside the cells), so I cannot know which cells express the gene, how would I know?



In situ hybridisation(ISH)→the signal is from undifferentiated cells they express the gene, and then the protein leaves these cells to the basement membrane

*In situ hybridisation is significant in knowing from which cell type this protein come from.

What is DNA library?

A library can be created for DNA fragments just like book libraries.

I have DNA fragment in the freezer, I want to study this DNA fragment, all information are present on a computer in really organized pattern, it will tell you the exact location of DNA fragment that you're interested in, you take the tube which contains DNA fragment of interest, and that's it! ☻

There are 2 types of DNA library, the 1st is genomic DNA library, and the second is cDNA library (C=complimentary)

1-Genomic DNA library

The idea, we take the hall genome, cut it to smaller fragment. We place each fragment into a plasmid (a vector like plasmid) and then we insert these plasmids to the bacteria, each bacterial cell will have one plasmid, and then I make identification, what's the fragment that is present in each bacteria, I take this bacteria and put it in the freezer, when I want to study this DNA fragment, all I want to do is go to the freezer, take a little sample of this bacteria, I grow it, take the plasmid out, cut it with restrictions endonuclease, And then DNA fragment is ready to study it.

What kind of DNA sequences exist in genomic DNA library (consensus sequences)?

Promoters, restriction sites, origin of replication, DNA that doesn't get transcribed, genes, promoter proximal elements, termination signals, microsatellite, telomeres, centromeres, VNTRs, STRs, enhancers, silencers, everything!!!

2-cDNA library (Complimentary DNA)

it is a DNA that is made from mRNA

I need reverse transcriptase to convert mRNA to cDNA

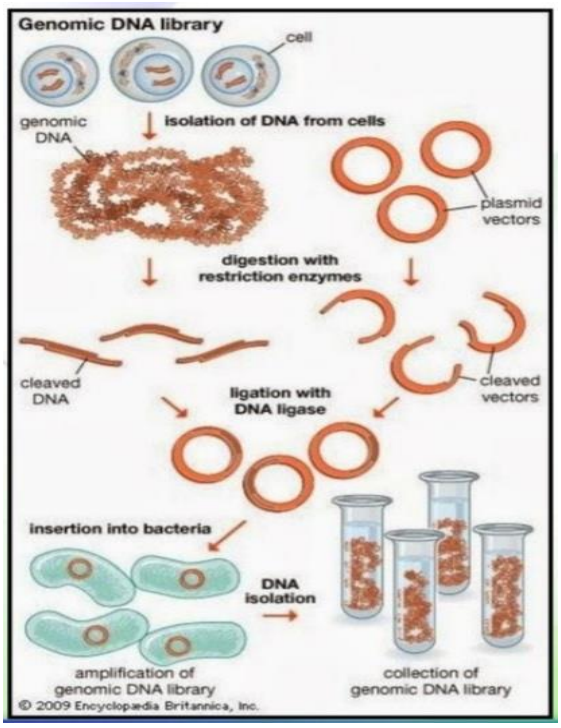
What kind of DNA sequences exist in cDNA library?

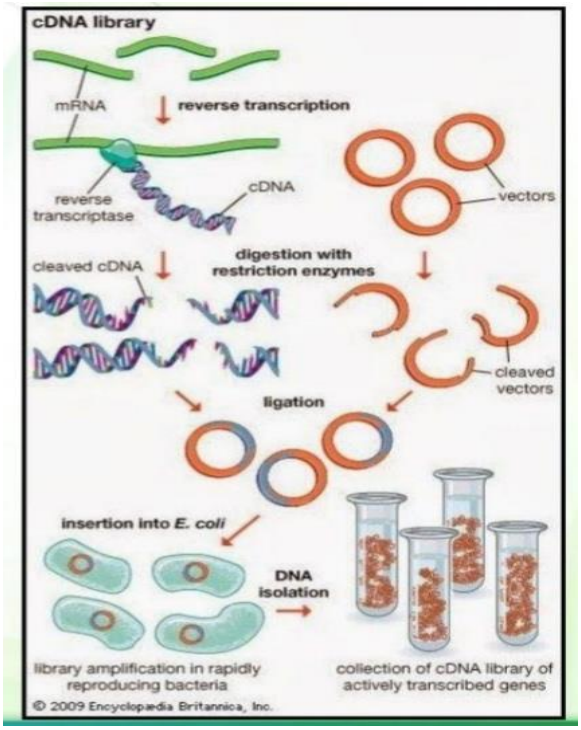
Exons of genes

What's about protein coding sequences and untranslated sequences??? Also, these 2 sequences exist in cDNA library.

-If I make genome library from skin cells and the liver cells to one person they Will be identical (the same genome)

-cDNA library to the skin vs liver: they would not be identical. Why? Cause they transcribe different genes.





Quantitative Real-time PCR from mRNA

Using **SYBR green.** A small molecule that bind to dsDNA and gives a signal when binds to dsDNA. So, the more DNA you have, larger the signal is.

Can I do the same with RNA? Definitely (but now, we use it to know the starting material of mRNA). What do I do?

Take RNA molecules→ convert them to cDNA and then amplify them (real time PCR).

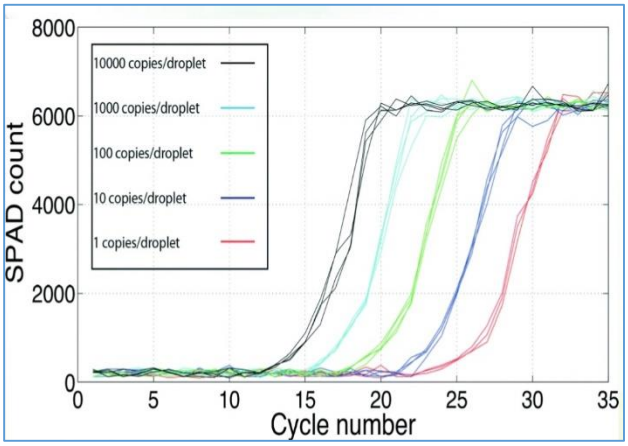
Can it be quantitative? Definitely it can be quantitative. If gene is expressed in that cell and produces at a certain point 10 RNA molecules, and at a different condition, the gene expression is really high reached 1000 mRNA molecules

10 RNA→convert to 10 cDNA

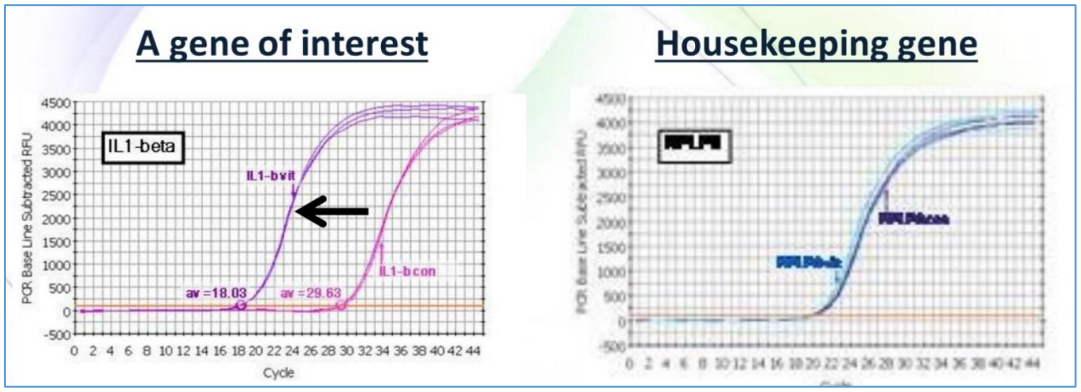
1000 RNA→convert to 1000 cDNA

And then I can do amplification using real time PCR (or using SYBR green).

In the figure, when we reached cycle 12, the 1st sample (black) start producing signals, the green start at 18, red at 22 and so on. That's mean the starting material of mRNA is different. In the black sample, the gene expression is higher than other samples, the red has the least amount of gene expression (least mRNA expressed) for a particular gene. How I does amplification only for that gene? Primer is specific (PCR).



-To ensure that we put same amounts of DNA in all samples, we use housekeeping gene in the same experiment.



Note that the signal starts in cycle 18 in one sample, and in cycle 29 in the other sample (low expression in the red sample compared with the black). And the house keeping gene is almost identical. That's indicate that starting material is the same in the two samples. the two samples. Remember, housekeeping genes express the same amount of mRNA in all cells. So, if the 2 curves were not identical, this indicates that I used different amount of DNA in different samples and the results are fake.

-Omics

when the scientist study:

One gene.... Genetics

All genes ... genomics

Gene expression (mRNA)... Transcriptomics

Protein... proteomics (proteom=total collection of protein in the cell)

All is new technologies permit to study all proteins, genes, RNA at the same time

***We would focus on studying **transcriptome**... To study the expressed genes all at the same time.

We will discuss one technique that is **DNA microarray** (Array= مصفوفة)

It's really small, contains thousands of spots, each spot contains probe (multiple molecule) for a specific gene.

I know what is the gene that each spot is specific to. Really, I can have probes for all 21 thousands genes in this small chip!! Actually, we can put 300,000 probes, amazing!!

What is the idea? If I take RNA molecule from the cell→convert it to cDNA→then labeling the cDNA and put it on the array→it will bind to where there is a probe specific to it → the spot where the probe exist will produce a signal (from labelled cDNA).

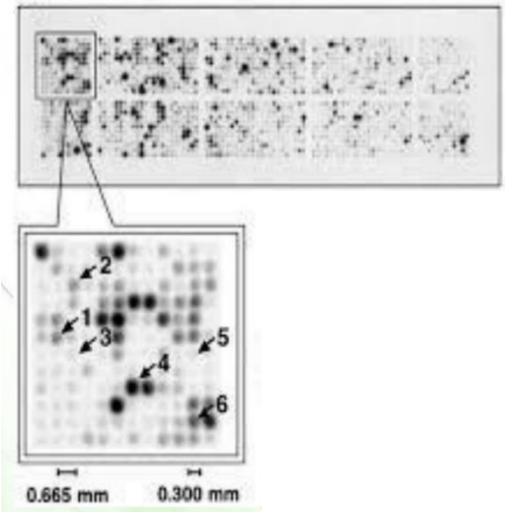
So, if the spot produce signal→cDNA present → mRNA present→ that's mean the gene is expressed.

If there is no signal→ the gene is not expressed.

If the signal is intense→ transcription is high.

If the signal is low→ transcription is low.

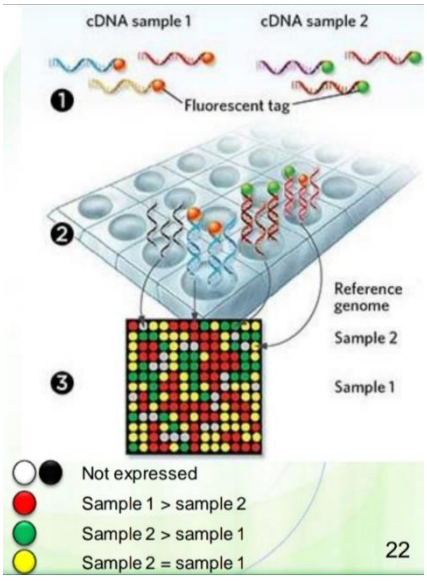
This is array, look at the square.



The gene that is represented be the white spot (no signal) is not expressed.

The black spot is represent high signal, so the gene is highly expressed.

So I can indicate what genes are expressed by each cell, specifically. Also we can find out Which genes are induced (expression is high), which are not highly expressed, and which are not expressed at all. This is what we call **molecular profiling**. Basically, it is what gene are expressed for this cell at particular time of point at particular condition.



The gene could be mutated and the mRNA weren't produced. So what we do is something called **comparative hybridisation**. I take RNA from sample 1, convert it to cDNA, and the same with sample 2. I label cDNA with florescent red tag, and the second with green, then I pool them together, and then put them on the array.

We said that a spot contains one thousand probes of same type to a particular gene.

These cDNA will bind with probes (hybridisation).

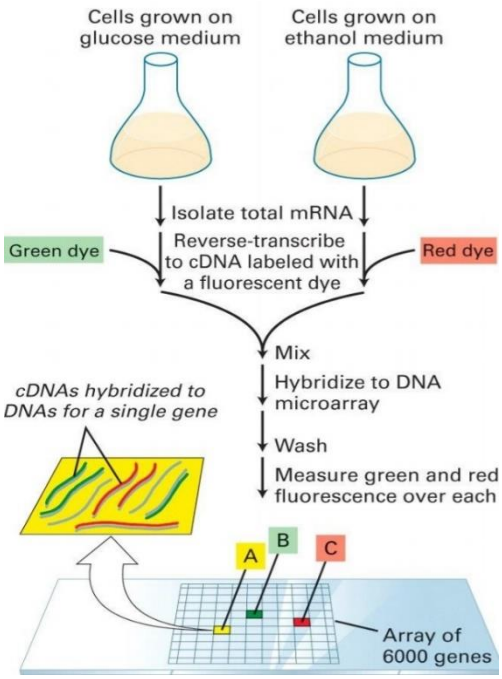
-If cDNA is not present in these two samples, the spot won't produce a signal, it will be dark (no light).

-If the expression is equal, it is a matter of competition. Let's say that we have 100 cDNA molecule of certain gene in each sample, when we put them together, hybridisation takes place and will be equal as well. In this case, the computer will give yellow colours (yellow=identical).

But if sample 1 has more cDNA than sample 2, then sample 1 bind more→ red signal stronger than green, the computer will present red spot. The green spot is the opposite to the red (sample2 >sample1).

So how can we benefit from that, we can take healthy and disease cells (normal vs. tumor) and then see gene expression, the gene that is increased could be oncogene, and the gene that is decreased could tumor suppressor gene. So I am studying gene expression to all these genes at the same time.

Other example, I study how glucose and ethanol do modifications on gene expression, so I can know what genes that's affected by presence of ethanol & glucose, depend on what's upregulated and what's downregulated, what's over expressed and what's under expressed.

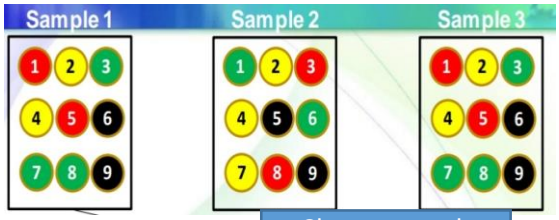


A If a spot is yellow, expression of that gene is the same in cells grown either on glucose or ethanol

B If a spot is green, expression of that gene is greater in cells grown in glucose

C If a spot is red, expression of that gene is greater in cells grown in ethanol

These are three samples from three patients that have breast cancer, I gave them a treatment. Two of them responded to a treatment, and one did not respond, why?



Because of gene expression (it's true they all have breast cancer, but different type of breast cancer).

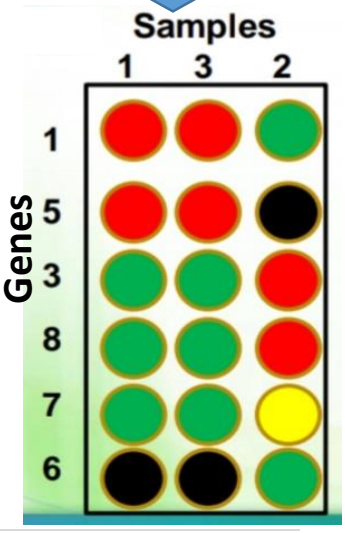
So, I take RNA from sample1& 2&3 and put them with the same control. The results:

-In sample number one: gene 1 is over expressed in patient's cells gene 2 is normally expressed gene 3 is under expressed

-In sample number two: gene 1 is under expressed in patient's cells gene 2 is normally expressed gene 3 is over expressed

-sample 3 is the same as sample 1

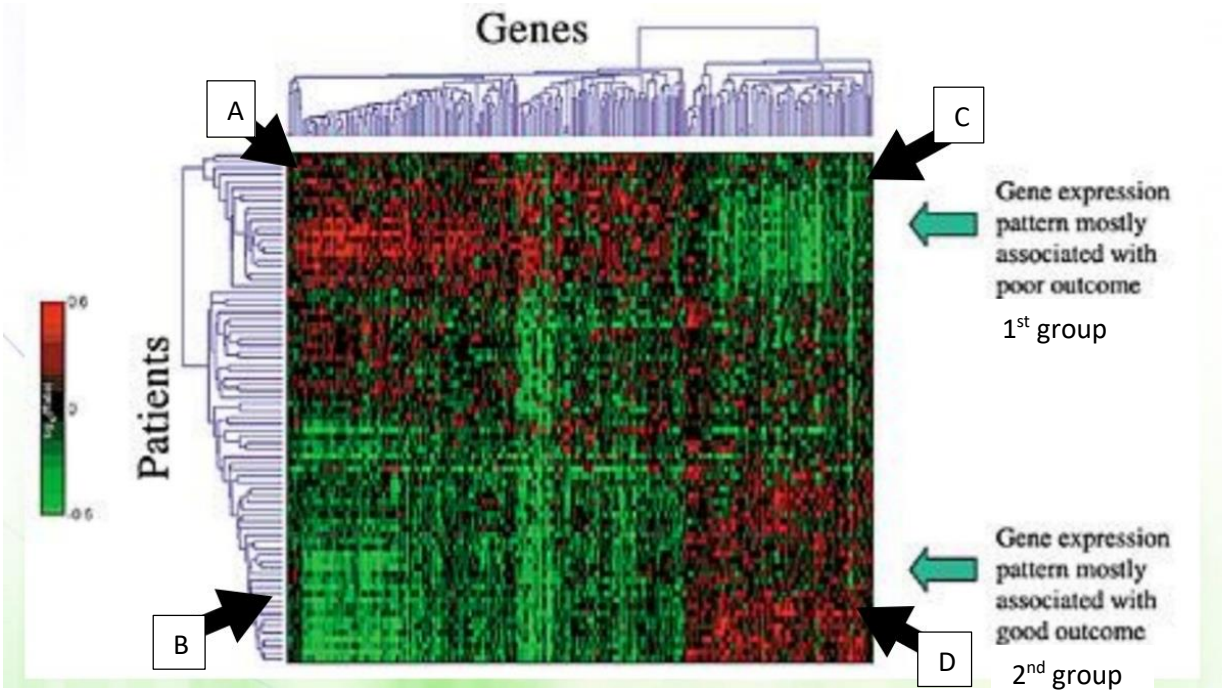
-Cluster samples according to expression
-Eliminate samples 2, 4, 9



I can't study the expression of 20,000 genes at the same time, so the computer helps me out. It helps me by making clustering.

The computer put sample 3 next to 1 because they have similarities (similar expression profile) and sample 2 is totally different.

Now, we will conclude the sheet with this picture:



Humans have 20,000 genes. Are they all important? No. so the computer deleted the genes that is not expressed (the black spots), and remove genes with equal expression between different samples (yellow).

Look at part (A), these patients have **over expression** for these genes. Look at part (B), these are other patients who have **under expression** to the same genes.

Other genes on the first patients group is **under expressed** (part C), and on the second patient group is **over expressed** (part D), the two patient groups have totally different molecular profile.

What is special about the 1st and 2nd groups of patients?

The 1st have poor outcome, the 2nd have good outcome. In 1st group, the disease was really aggressive. So I can make prediction by making molecular profile to patient's cancer.

If the profile was like 1st group, I can predict poor outcome, so I'm going to be aggressive on the treatment, and if it was like 2nd group, I'm going to be light on the treatment because she has good outcome prediction, cause chemotherapy treatment is really harsh, and it's better to reduce chemotherapy, and this is what was call personalized medicine (now it is called recession medicine). Before the treatment, I would know what is the molecular profile of the patient and of his disease to give him a proper treatment.

WISH YOU BEST OF LUCK ^^