Doctor 022

Biochemistry

Sheet no.

Analysis of gene expression and RNA levels

Writer: AL-Razi Node team

Corrector: AL-Razi Node team

Doctor: Dr Mamoun, Dr Diala

ANALYSIS OF GENE EXPRESSION (Gene transcription, converted to RNA) AND DNA LEVELS

Overview of the concepts discussed:

- Techniques to analyze the level of gene expression.
- Comparisons between two samples in terms of: gene activity, level of expression and amount of DNA.

RNA LEVELS:

Basic methods: Northern blotting, in situ hybridization.

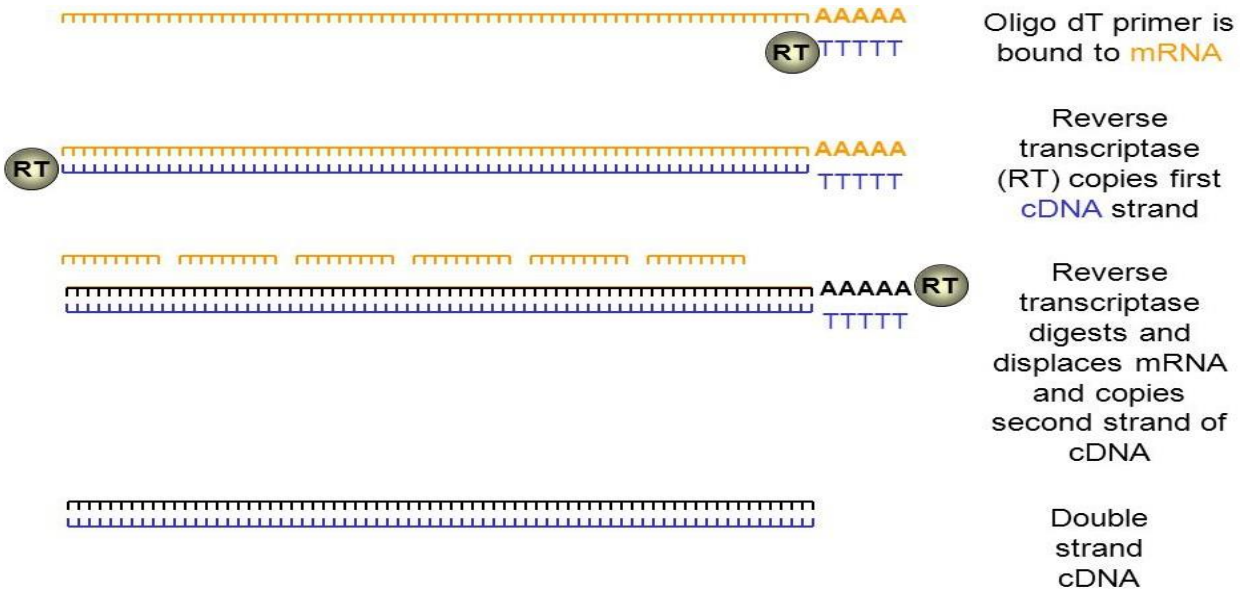
Advanced methods: real-time PCR, DNA microarray

Very advanced methods: RNA-seq.

FIRST MAKE A COMPLEMENTARY DNA (CDNA):

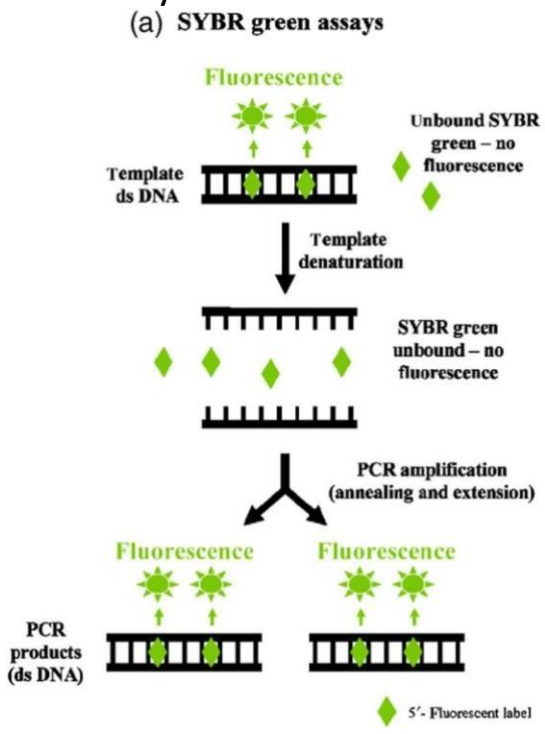
- Real-time PCR basically we amplify the DNA, the same time DNA is amplified, I am quantifying how much amplification there is in the site, and so, can we do the same for RNA??
- Whenever we want to analyze RNA using advanced and very advanced methods, we need to convert RNA to a double-stranded DNA using the enzyme reverse transcriptase and we call the DNA cDND
- The idea is the genes are expressed at different levels(different activities)some genes are expressed at high level(really active) transcription to thousands RNA molecules per second, in other hand, there are genes (the cells need a little amount of proteins) make ten copies per second, there is other genes their expression are zero like insulin, the nerve cells make insulin (the expression is zero in nerve cells and liver cells) the insulin is expressed in pancreatic cells this is tissue specific gene expression, other genes are needed all time(by the cell) in high amount like actin, tubulin, metabolic enzymes(phosphate dehydrogenase, hexokinase)

- Firstly, we need a primer (oligo dT primer) which binds specifically to the poly A tail of the mRNA.
- Then, reverse transcriptase can start synthesizing the cDNA.
- After the enzyme extends the first strand of DNA, it degrades the mRNA strand and synthesizes the second strand.
- And voila, the double-strand cDNA is produced!



THE QUANTITATIVE REVERSE TRANSCRIPTASE REAL-TIME QPCR OF MRNA.

- The basic idea here is that the amount of cDNA exactly reflects the amount of mRNA.
- **Another way of relative quantitation of RNA expression is by converting RNA into cDNA followed by PCR in the presence of SYBR green.**
- The cDNA can be amplified using primers just as RNAs (the same process mentioned in the previous lecture).
- And the amount of DNA can be quantified according to the moment when the signal was detected.

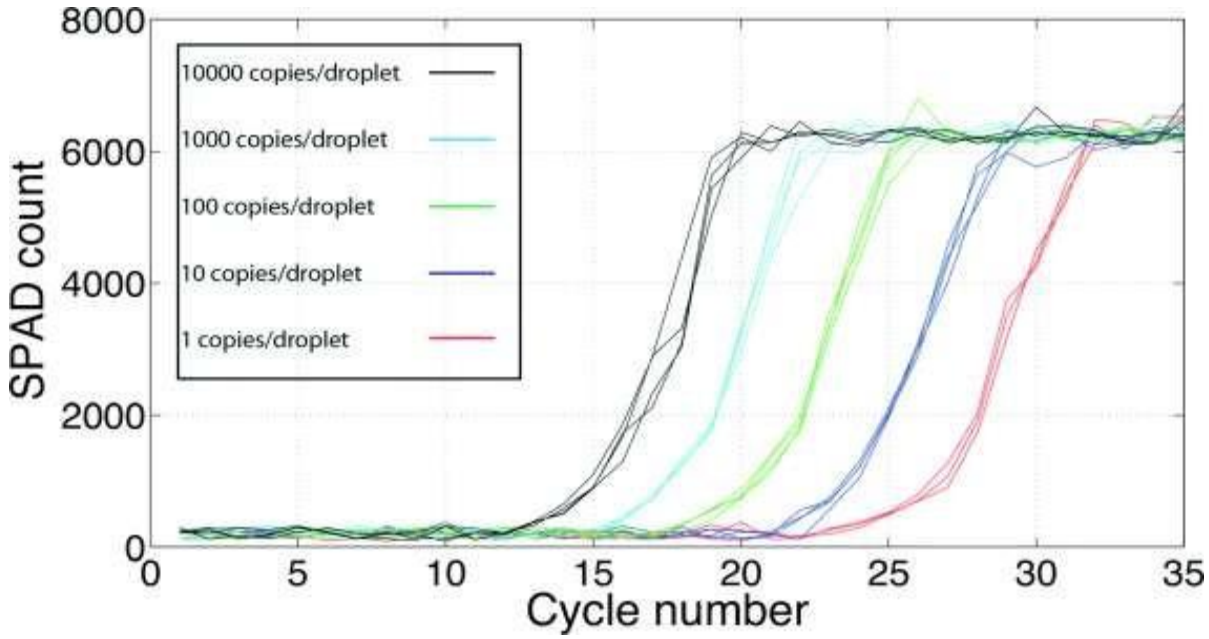


- **The higher the amount of RNA**
- **(cDNA),the sooner it is detected.**

- We can use SYBR green as a signal to detect DNA or taqman qPCR.
- If I start with ten thousands

RNA molecules converted to
cDNA (10000Cdna)
1:1 ratio

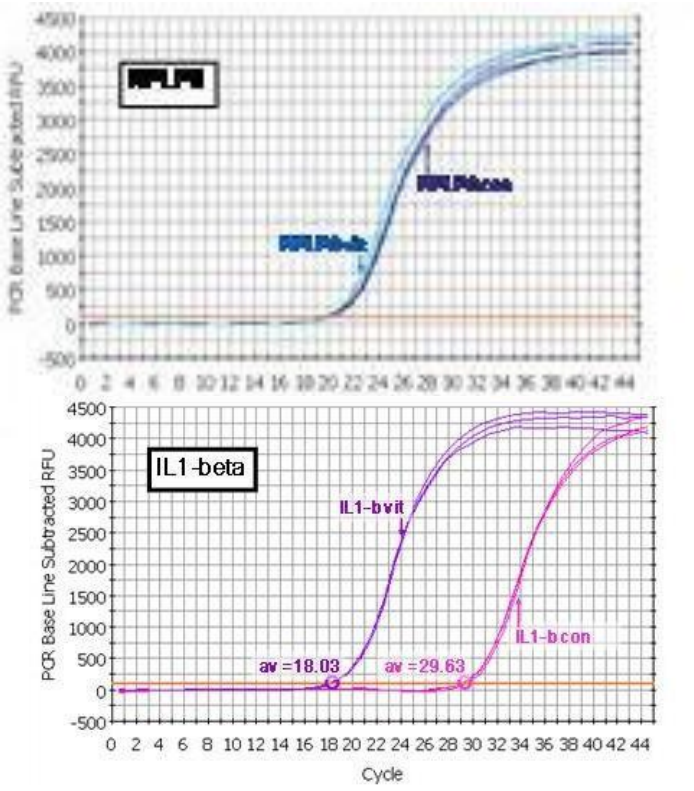
- Remember: the more molecules you start with , the earlier the signal is detected.



We have different samples (amplified)
CT (threshold cycle) for the black is 13, why there is no detection earlier??
The device is not sensitive enough (there is a low amount of cDNA)
And that's why corona virus test needs 2-3 hours; the sample was prepared and amplified while in bioterrorism we can't wait that long, we have to know if there is virus or not in five minutes.

COMPARISON BETWEEN TWO SAMPLE

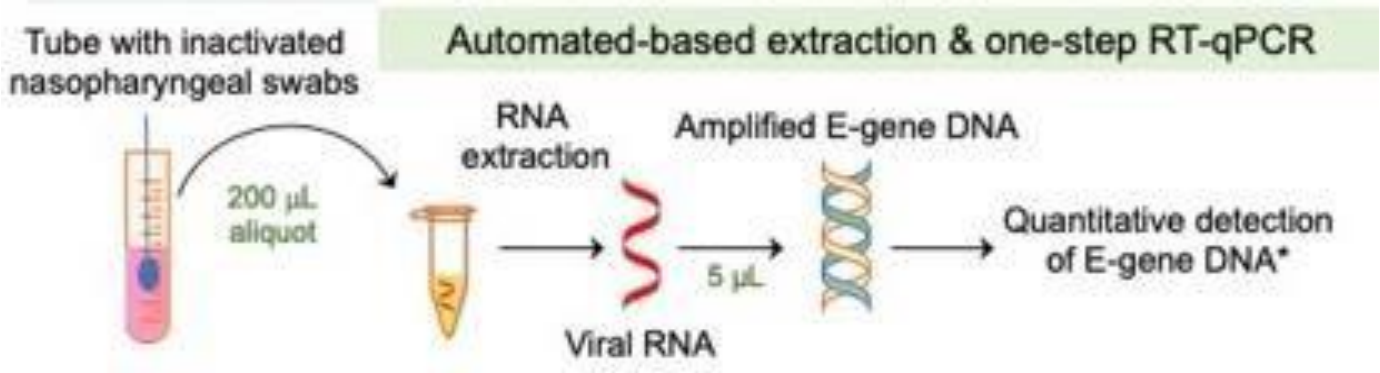
- When we compare two things, we need a constant agent which is the HOUSEKEEPING GENE, so let's say someone comes (his hand shaking) to make PCR for two samples (normal, cancer), while he prepare he takes a specific amount of the starting material of the sample (10 microliters) the second sample he takes (20 microliters) then he makes PCR, so which sample will be detect first?? 20 microliters will be detect first ;the starting material is double this is technical error so the best way is to look to a constant gene (HOUSEKEEPING GENE).
- Insulin is not a HOUSEKEEPING GENE
- REMEMBER: housekeeping is a gene whose expression does NOT change.
- Examples of housekeeping genes: Histone genes, Myosin genes, Actin genes, tubulin, etc.
- Notice that the two plots are overlapping which implies unaltered expression.
- If the expression of HOUSEKEEPING GENE on the same cycle for the two samples that's mean there is no error, but if one versus the other there is error.
- We should have control (no technical error).
- Here we are looking at a gene of interest (cytokine produced by inflammatory cells (immune cells)).
- The detection of one sample (purple) is before than the other one (pink).



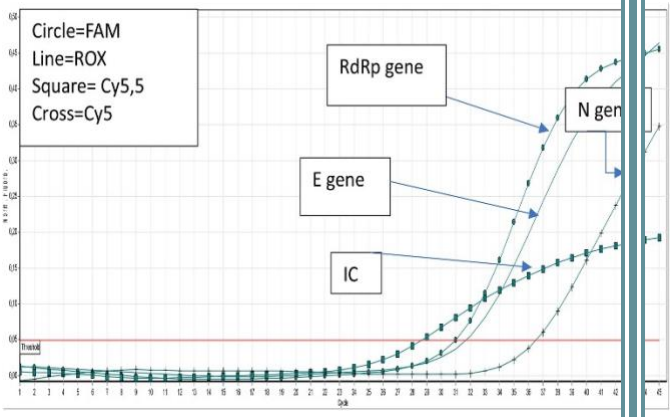
- So, there is an increased expression in the sample represented by the purple plot, and we are sure that this is true since the housekeeping gene is expressed in equal amounts in both samples.

DETECTION OF SARS-CO-2

- qPCR is used to detect corona virus, the genome for the corona virus is RNA.
- You definitely remember the way of collecting samples during the pandemic.
- The ugly method of collecting samples from the nasalpharyngeal swabs.
- These swabs are being inserted into a tube that contains a buffer.
- In this buffer, you have lysis of the sample→ release of the genetic material of SARS-CO-2→ RNA is converted to cDNA→ amplification of the Corona virus gene.



There is an amplification of 3 viral genes,just to confirm that this sample contains SARS-Co-2 (the 3 genes arespecific for the virus).

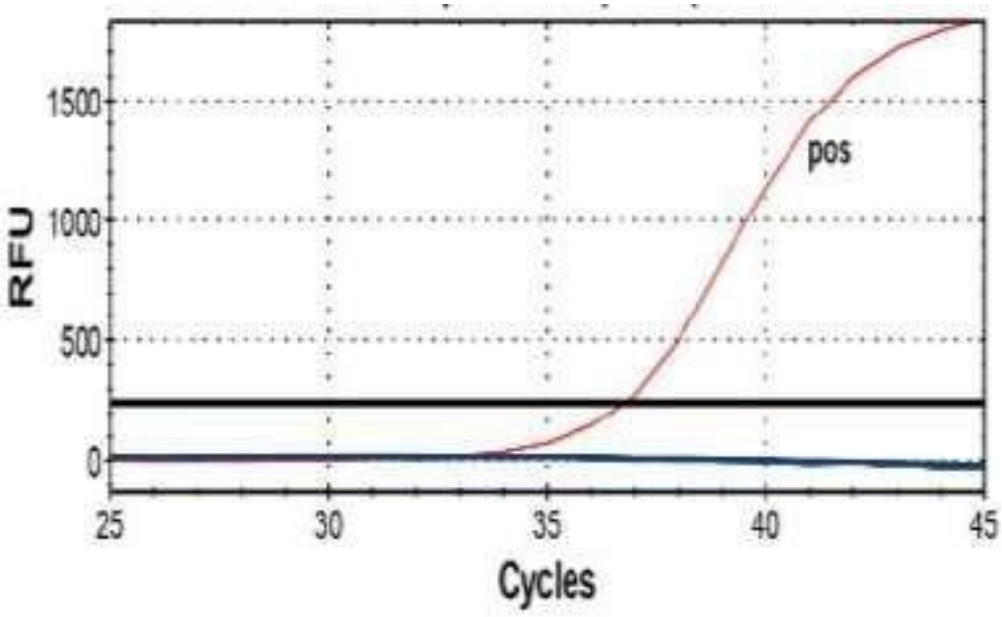


- The three signals had been detected above the threshold, which means it is +ve for SARS-Co-2.
- The internal control is IC, which is basically amplification of a human RNA tha means human cDNA, to make sure that the sample wascollected properly. (a signal means a YES)

- If there is no amplification of corona virus we can't say the test is negative unless we make control for human RNA
- The red plot refers to the control and the sample was collected properly. (there is a signal)

Positive for IC

But the viral (blue is negative)

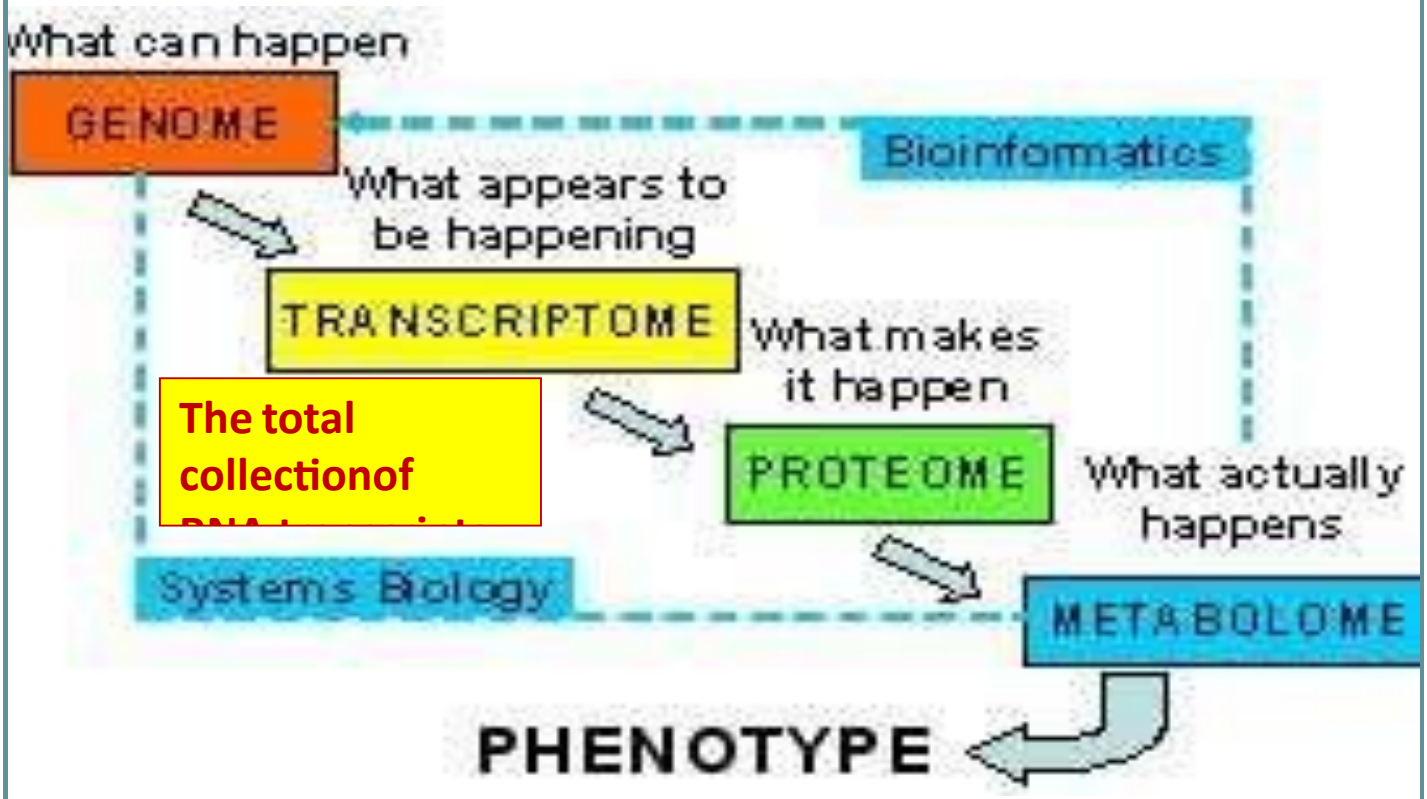


- If there is amplification of human RNA, no amplification for viral RNA-cDNA so this is a negative test.
- This figure implies a -ve result

- This is an invalid result since the control's signal was not detected.
- A technical problem has occurred.

THE SCIENCE OF -OMICS

- The figure shows how science developed.
- We are concerned with the science of transcriptomics.



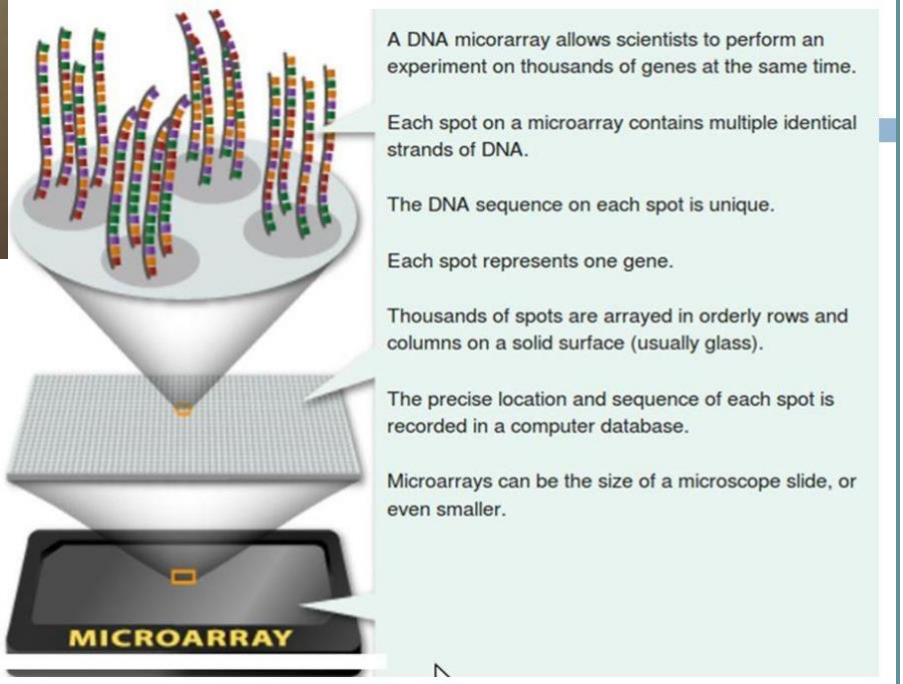
STUDYING THE TRANSCRIPTOME

- One such method in studying transcriptomes is DNA microarrays, whichallow the analysis of the RNA products of thousands of genes all at once.
- By examining the expression of so many genes simultaneously, we canunderstand gene expression patterns in physiological and pathological states.

DNA MICROARRAYS

- **DNA microarrays are solid surfaces (glass microscope slides or chips) spotted with up to tens of thousands of DNA fragments in an area the size of a fingernail.**
- Different DNA probes (not labelled) are arrayed on this chip.
- **The exact sequence and position of every DNA fragment on the array is known.**
- **http://learn.genetics.utah.edu/content/labs/microarray/**
- **http://www.sumanasinc.com/webcontent/animations/content/dnachips.html**

The black is the array



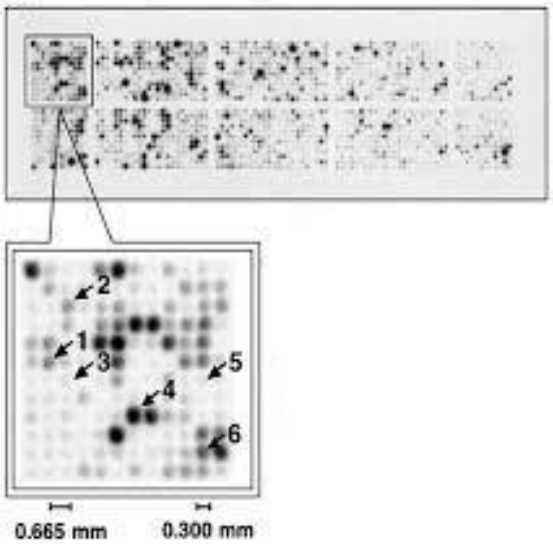
This is a zoom in the chip.

At each position we have probes (short DNA that is complementary to the Sequence of one of the gene) that is Connect at specific location

Maybe we have million probes for different genes that arrayed on this little ship

THIS IS HOW IT IS DONE

- **This is done for a single sample using radioactively labeled cDNA.**
- **We take mRNA from the cells (tumor, normal) being studied isfirst extracted and converted to cDNA.**
- **The cDNA is labeled with a radioactive probe (radioactive phosphorus).**
- **The microarray is incubated with the labeled cDNA sample for hybridization to occur.**



- **If a gene is expressed, then the cDNA Will exist and bind to a specific complementary DNA fragment on themicroarray (probes).**
- **Binding can be detected since the cDNA is labeled and expression isdetermined.**
- Each spot contains identical probes for a specific gene (the probes are complementary for the DNA of a specific gene).
- Remark that we know what each spot represents.
- The more cDNA bound to the probe in a certain spot, the stronger signal you would get.
- The more active the gene is→ more mRNA produced→ more cDNA→ more binding→ stronger signal. (notice spot #4)
- So, the DNA microarray tells you about gene activity of all genes in a cell at a certain time point.
- If there is gene expression we have RNA, cDNA, signal.

COMPARATIVE EXPRESSION

Scientists didn't like working with radioactivity; because it only gives you one color, which is black.in addition they couldn't do comparison between 2 different samples.

- **In order to compare expression of genes two different samples, the cDNA molecules are fluorescently labeled with different colors (green and red) and added to the array.**

So here we are comparing the genes that are **differentially** expressed.

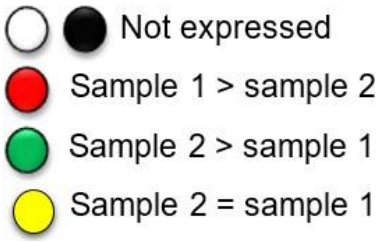
For example, if we compare a cancerous gene with a normal gene, we can get information about the biology of cancer or the cause of it.

For example, **oncogenes:** a gene that is highly expressed in cancer and it makes cells cancerous.

Or a gene that is down expressed in cancer cells in comparison to normal cells, so this tells us that this gene could be a tumor suppressor gene (a gene that prevents theformation of cancer).

- **An increase in the amount of a RNA molecule in one sample versus the other is reflected by an increase the amount of produced cDNA and an increase in fluorescence in the bound spot.**

So we take the mRNA molecules from 2 different cells let's say cell 1 and 2, and we convert the mRNA to cDNA and we label these DNA molecules with different colors, let's say cDNA from cell 1 with red fluorescence and from sample 2 with green fluorescence, then we mix them, then we add them to the slide or chip.

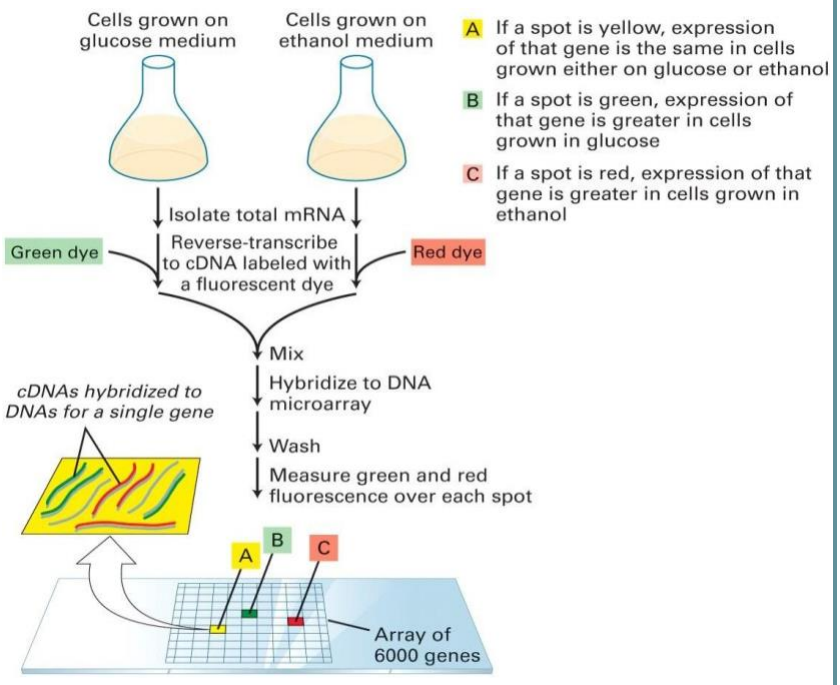


There would be competition in binding of the green cDNA and red cDNA to the probes in each cluster, so if you have equal amounts of them in a cluster the probes will bind to them equally, if you have more green cDNA (more expression in sample 2 compared to sample 1) in this spot you will have more green fluorescence than red fluorescence and vice versa.

However, if there is no expression of the gene in either sample, it means there is no mRNA, NO cDNA and no signal coming out of the cluster, you can look at the first cluster in the figure, and notice that there are probes but there is nothing bound to them, so there is no signal.

Note if there are equal amounts of green and red the computer will generate an artificial color of its own, in this example it is yellow.

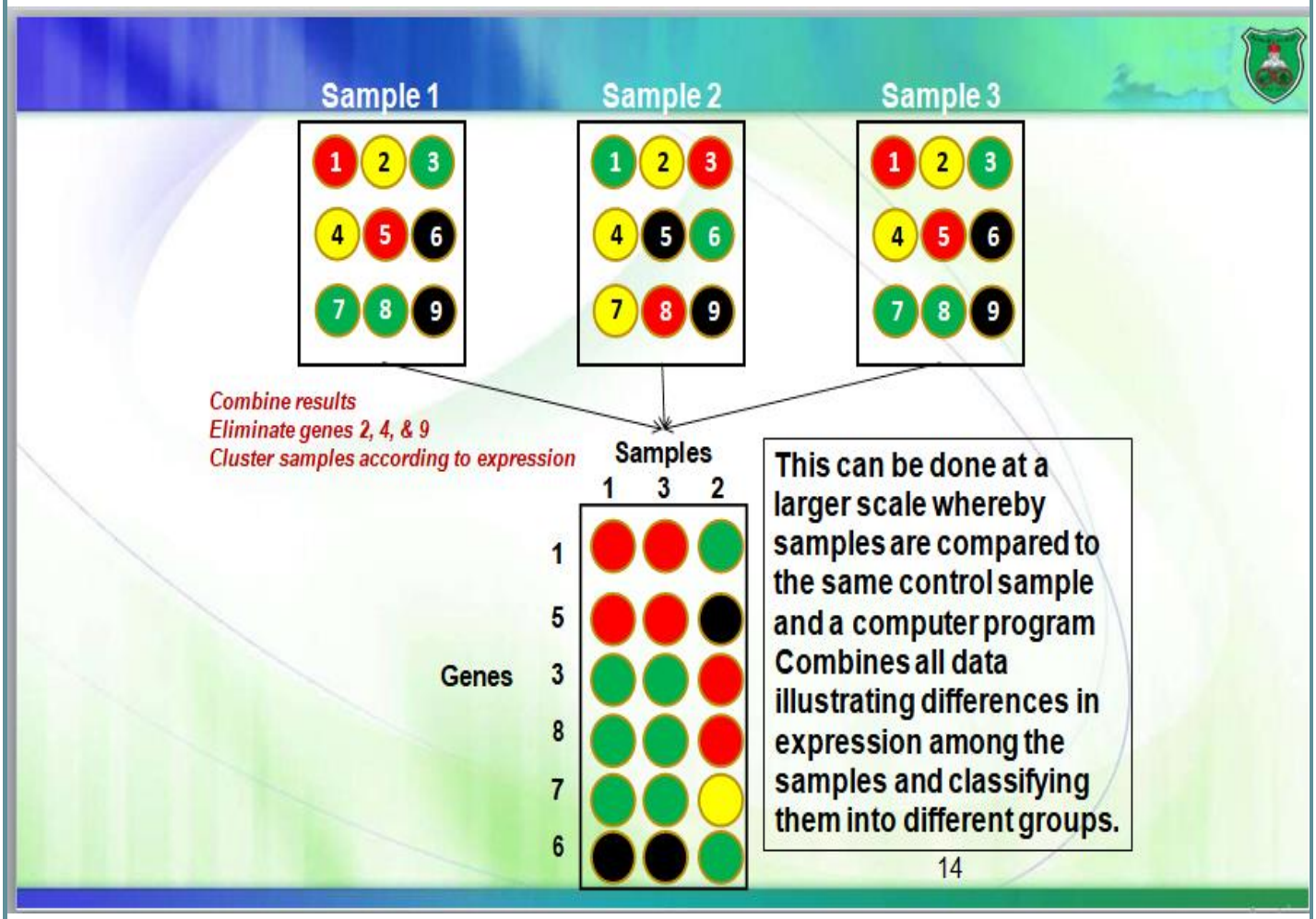
Look at the dots bellow the figure in the previous page to know their meaning.



In this experiment shown in figure we have bacteria grown in the presence of glucose and another one grown in the presence of ethanol.

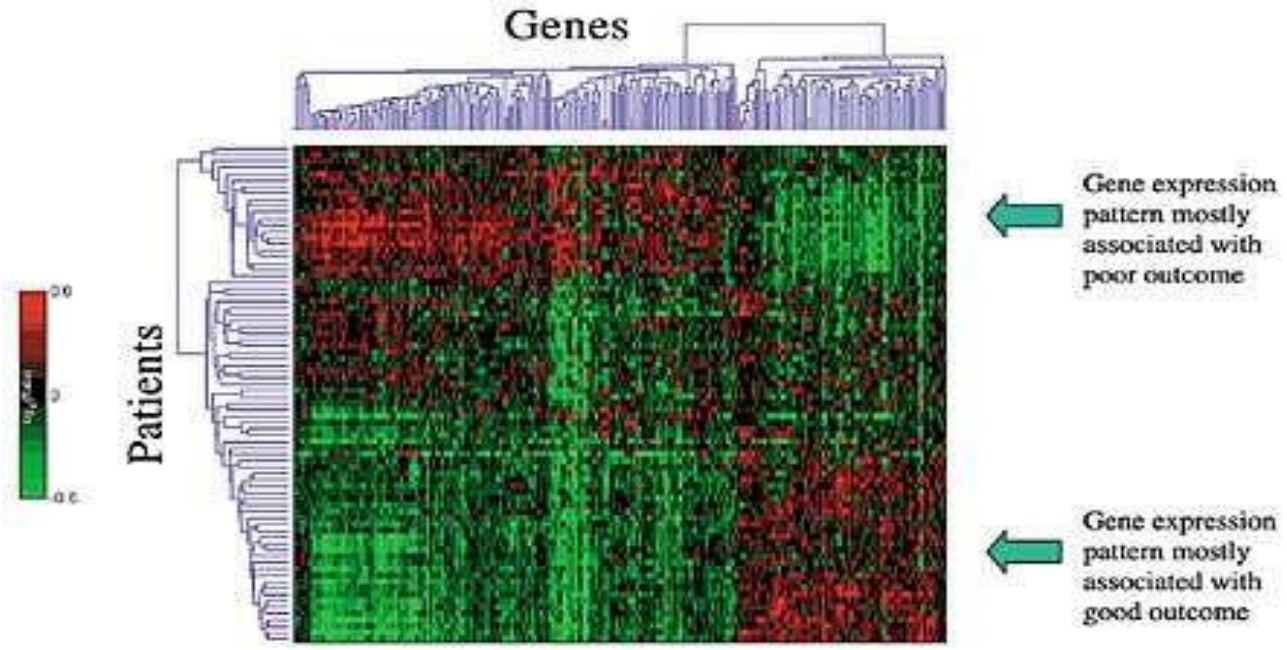
And we isolate the 2 messenger RNAs, and then theyare reverse transcribed intocDNA and labeled with different fluorescent dyes, Then they are mixed, and then added to the microarray, then we wash toremove the unbound material, then we detect the signal.

Again you can get red or green or yellow fluorescence, or you can get a cluster with no signal, which means there is no expression in either one of these samples.



- We need informatics of course we are not going to analyze thousands of spots at the same time.
- We need comparative measurement of the fluorescence that is emitted from each cluster; because we're going to have thousands of spots at the same time.
- So we can have a look at the samples and compare them with each other, for example in sample 1 cluster 1 has a red color, while in sample 2 it has a green color.
- So what the computer does, is he says forget about clusters 2, 4, and nine; because they are not informative; there is no change of expression (the color remained the same), and there is no signal from cluster 9 in all three samples.
- But all the other genes or clusters are variable (they are informative).
- And by comparing samples 1 and 3 we realize that they are identical to each other, however in sample in number 2 the pattern of expression is quite different. For example, in sample number 2 gene 1 is under expressed, it's down regulated. -The doctor said: please note the terminology that I'm using-.
- So the computer makes the comparison between the samples and defines the important genes.

DNA MICROARRAYS AND BREAST CANCER



Here we have a real test with real results.

So a sample is taken from a lot of patients, and we are looking at the expression pattern in breast cancer.

Notice that in the first group of patients we have over expression if the first 2 genes and under expression in the last one, while in the last group of patients the opposite is true.

But what does this study help us with? Simply we notice that the first group of patients has a poor outcome, they are not responsive to a particular treatment or cancer is aggressive or deadly.

The other group of patients has a good outcome; the treatment is working well.

So next time we take samples from a patient, we look at gene expression and based on the pattern of gene expression, we can predict if the patient will have good outcome or poor outcome, if the patient would have a poor outcome then we hit the cancer aggressively with treatment, but if the outcome would be good, then why should we make him suffer by giving him a treatment that is not necessary.

RNA SEQUENCING (RNA-SEQ)

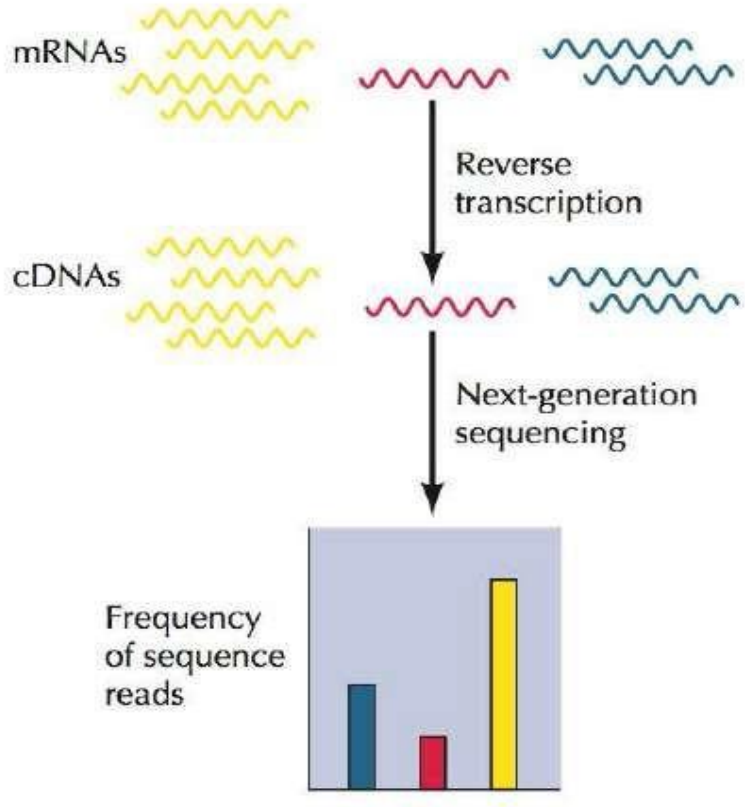
- **Cellular RNA is reverse transcribed to cDNAs**(the amount of cDNA reflects the amount of RNA so if we have 10

  Molecules of RNA I would have 10 molecules ofcDNA, and so on), **which are subjected to next-generation sequencing.**

We sequence all cDNA molecules, so what the machine does is that it will tell us the sequences of cDNA, thus we can know the sequences of RNA, and it tells us the levels or numbers of these molecules.

- **The relative amount of each cDNA (mRNA) is indicated by the frequency at which its sequence is represented in the total number of sequences read.**
- **So I can know the sequence of the gene that expressed and how many times it expressed (how active).**

So in our example we have more of the yellow cDNA, LOOK AT THE FIGURE AND SEE THE FREQUENCIES.



RNA-SEQ VS. MICROARRAY

❖ **RNA-seq can be used to**

So RNA-seq has advantages over microarrays in transcripts for example.

➢ **Characterize novel transcripts**

In RNA-seq we can sequence any transcript whether we know it or not, however, in microarrays we design probes that are specific to known transcripts.

➢ **Identify splicing variants** (we can have piece that are identical, or in other words spliced in the same way, these piece can bind and form one cluster, but using this technique we will be able to identify each one of the fragments in the cluster).

➢ **Profile the expression levels of all transcripts**

So in RNA-seq we can know the expression levels of all genes or all transcripts, in microarrays however; we only look at the known genes with known genomic sequences.

❖ **Microarrays are limited to detect transcripts corresponding to known genomic sequences** (probes).

This is how we were able to know that 75% of the human genome is transcribed.

The end

It doesn't matter what others are doing, it matters what you are doing

" خِتَامُهُ مِسْكٌ "